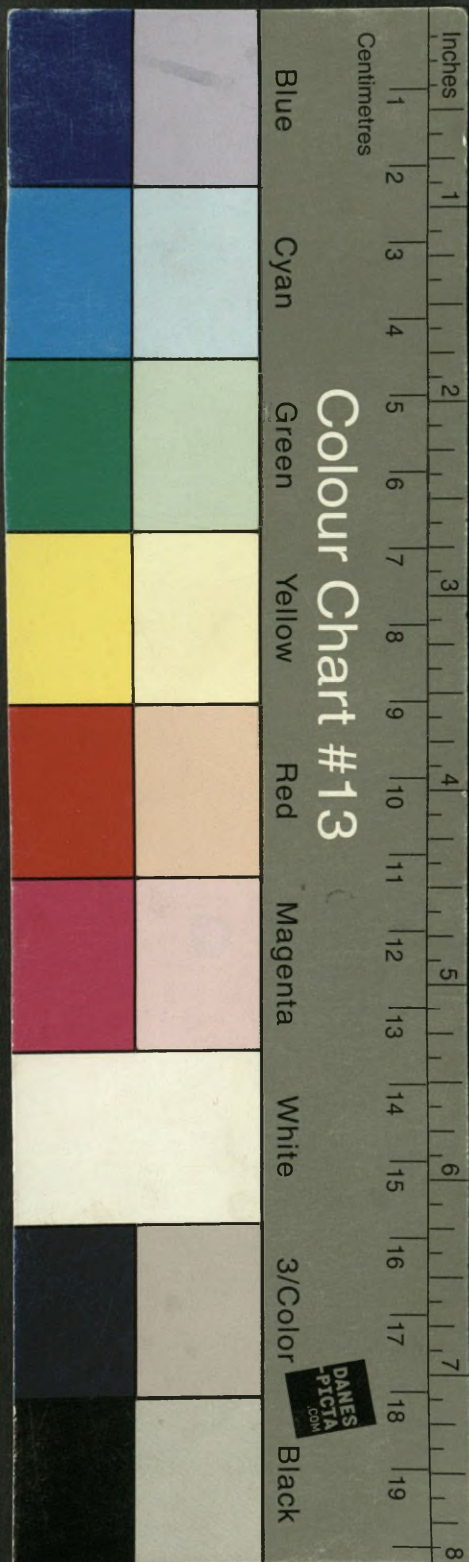
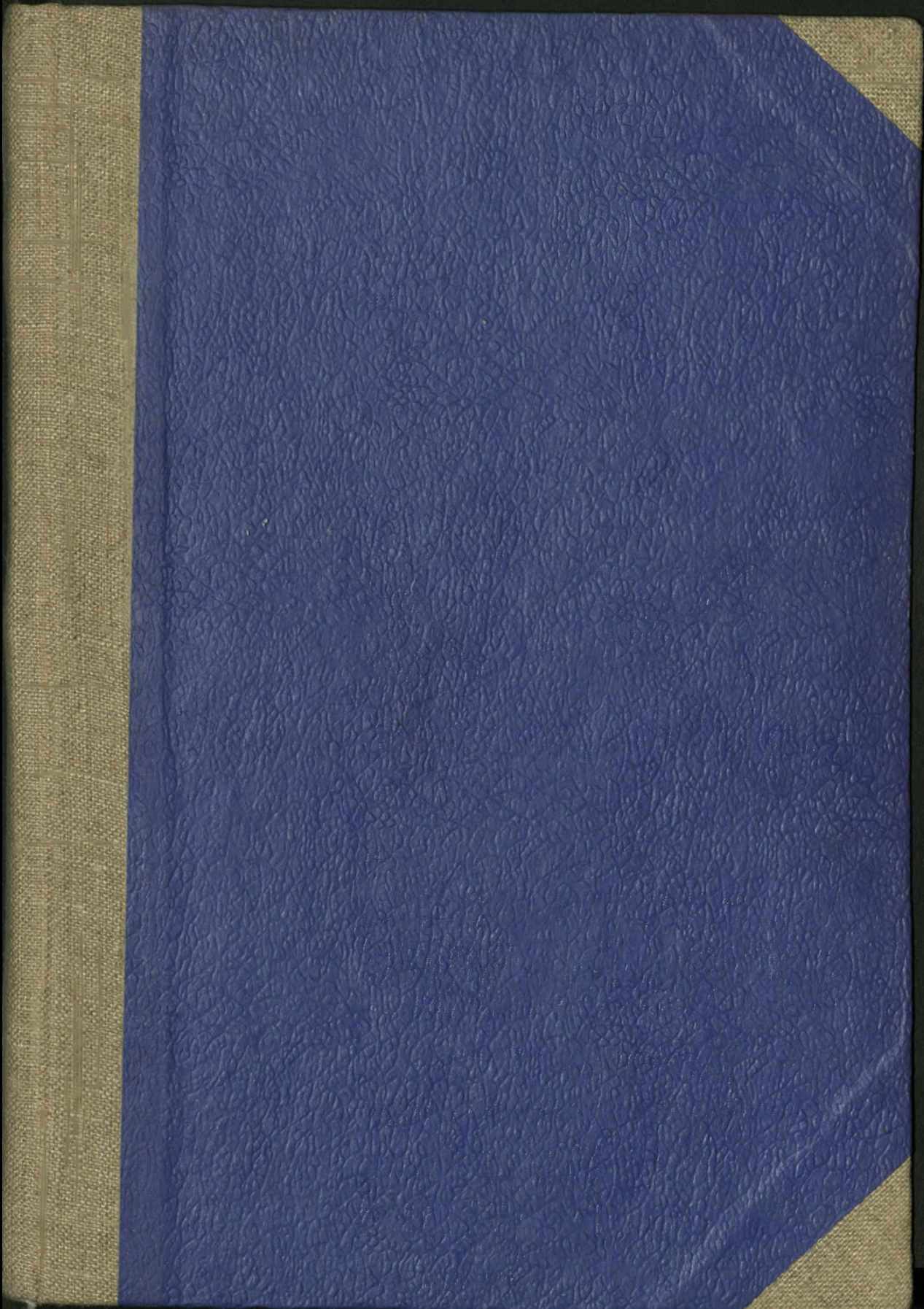


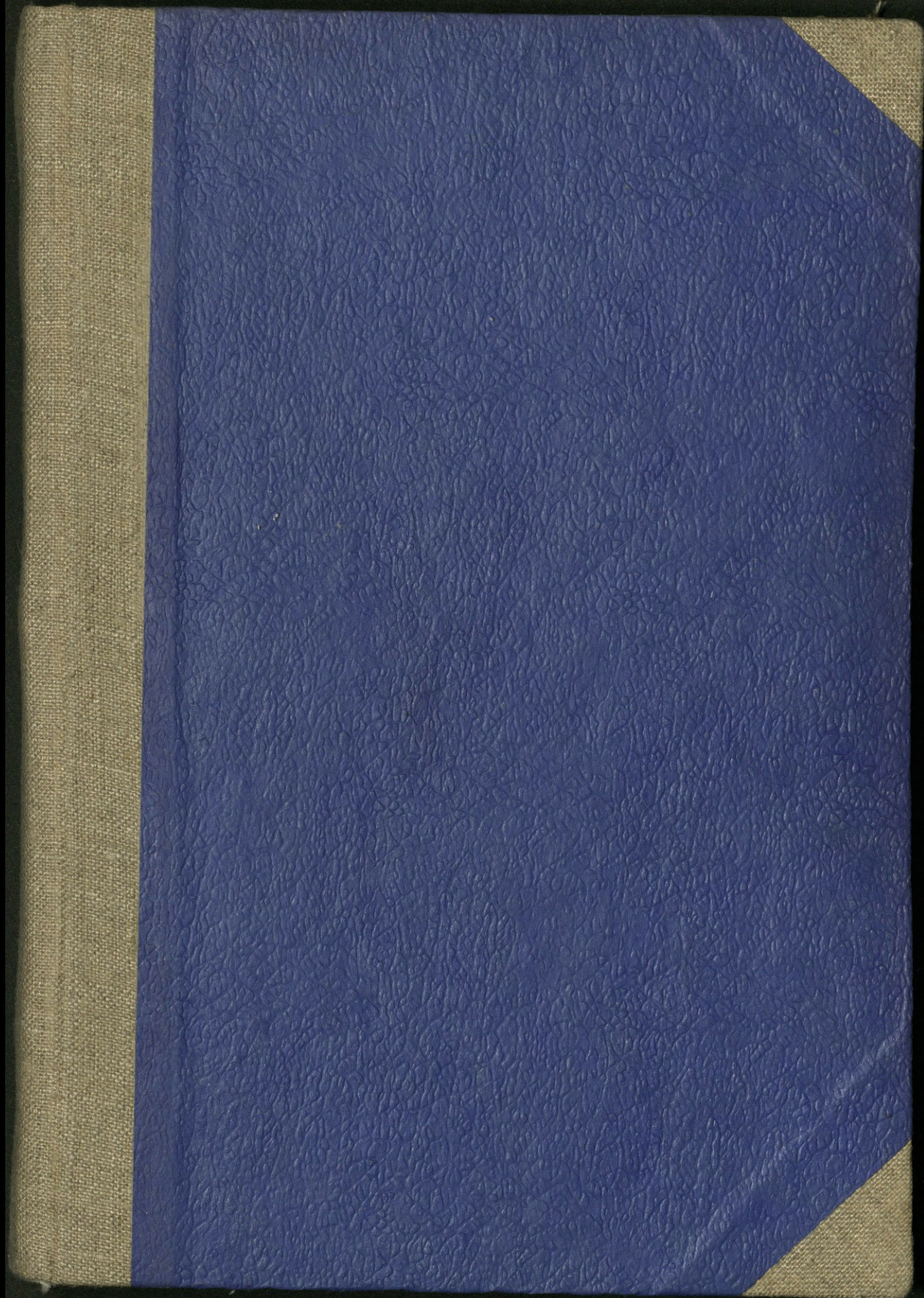
Grey Scale #13



DANES-PICTA .COM

A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19





BIBLIOTEKA UNIwersYTETU LUBELSKIEGO

STATYSTYKA

TEORJA METODY STATYSTYCZNEJ

OPRACOWAŁ

LEON WAŚCISZAKOWSKI



LUBLIN, 1930.

SKŁAD GŁÓWNY: LUBLIN, UNIwersYTET

BIBLIOTHÈQUE DE L'UNIVERSITÉ DE LUBLIN.

LA STATISTIQUE

LA THÉORIE DE LA MÉTHODE STATISTIQUE

PAR

LÉON WAŚCISZAKOWSKI



LUBLIN, 1930.
UNIVERSITÉ DE LUBLIN.

STATYSTYKA

BIBLIOTEKA UNIWERSYTETU LUBELSKIEGO

LEON WAŚCISZAKOWSKI

DOC. UNIW. JAGIELLOŃSKIEGO

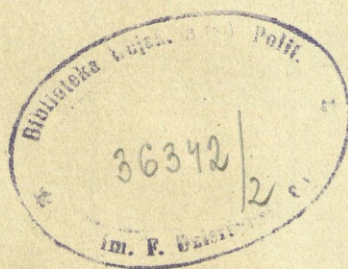
PROF. ZWYCZ. KAT. UNIW. LUBELSKIEGO

STATYSTYKA

TEORJA METODY STATYSTYCZNEJ

LUBLIN, 1930.

SKŁAD GŁÓWNY: LUBLIN, UNIWERSYTET



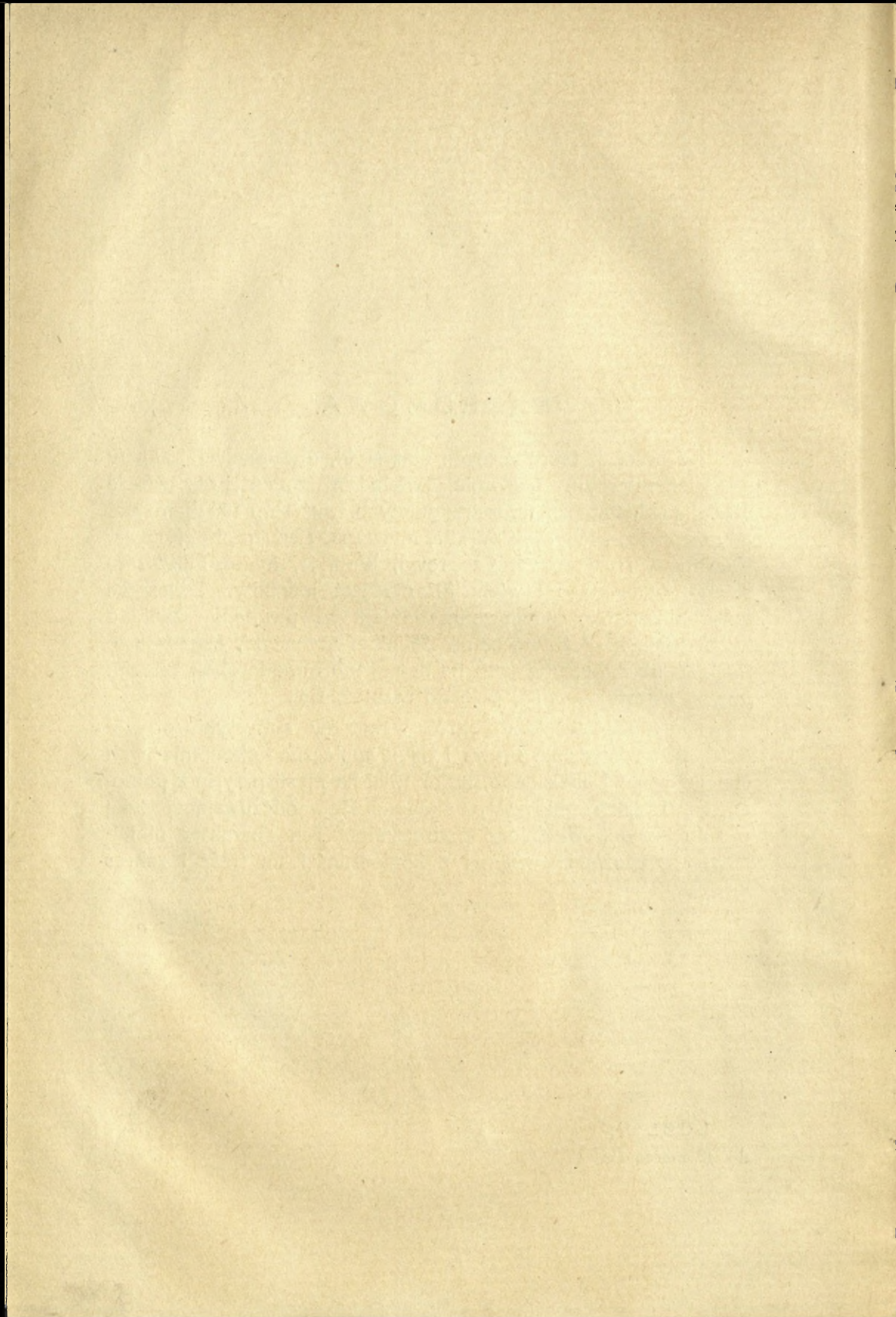
PRZEDMOWA.

W pracach teoretycznych statystyki, utrzymanych choćby na elementarnym poziomie, wiele zagadnień jest jeszcze dzisiaj niedostatecznie rozstrzygniętych. Już sam układ materiału w pracach Mayr'a, Conrada, Meitzena, Tischera, Kabłukowa, Kaufmana, Ancyferowa, Czuprowa, Wobłyja, Pilata, Moszczeńskiego, Orzęckiego i wielu in. nie jest jednolity. Zwłaszcza autor niniejszego opracowania odbiegł od pewnego szablonu w rozdziale II. Pozatem sama definicja statystyki, zagadnienie poglądu na statystykę jako na naukę lub metodę obok szeregu innych kwestyj — dają pole do rozbieżności.

Ogłaszając drukiem skrót wykładów statystyki, prowadzonych na Wydziale Prawa i nauk społeczno - ekonomicznych Uniwersytetu Lubelskiego, autor miał na myśli przyjść z pomocą w studjach młodzieży akademickiej, odczuwającej lukę w zakresie odpowiedniego podręcznika. Ten charakter niniejszego opracowania wpłynął z konieczności na treść i zakres książki.

Autor.

LUBLIN,
dn. 19 marca 1930 r.



I. Statystyka jako nauka i jako metoda.

----- Nazwa statystyki. -----	Wyraz „statystyka” na oznaczenie nauki i metody, zwanej statystyczną, jest znacznie późniejszego pochodzenia, aniżeli sam przedmiot badania. Źródłosłów tego słowa jest dwojakiego pochodzenia: statystykę wyprowadzamy bądź od łacińskiego <i>status</i> , co oznaczało pierwotnie stan lub położenie, bądź od włoskiego <i>statista</i> , którym określano męża stanu, polityka, poświęcającego się pracom publicznym. Istotnie, w pewnym okresie rozwoju statystyki, naukę tę pojmowano jako naukę o państwie, niezbędną przedewszystkiem dla ludzi, działających na polu państwowem. Jednak, jak zaznaczyliśmy, statystyka była znana znacznie wcześniej, aniżeli nazwa, którą jej — w perspektywie rozwoju nauki — stosunkowo niedawno nadano.
--	---

Pewien zasób wiedzy statystycznej oraz system czynności praktycznych, związanych z dochodzeniami statystycznymi, były już znane w czasach zamierzchłych. Posiadamy świadectwa historyczne, które stwierdzają, że już w starożytności były wykonywane pewne prace praktyczne, nieraz na szeroką skalę zakrojone, które należy nazwać statystycznymi. Tak, w starożytnych Chinach, Indjach, Persji, Egipcie i t. d. zbierane były dane ludnościowe, których posiadaniem przez ówczesnych zarządców państwa możemy sobie wytłumaczyć przeprowadzone reformy społeczne. W Chinach już w r. 2238 przed erą chrześcijańską był przeprowadzony spis ludności. Reformy społeczne naprz. Solona w Grecji, byłyby niemożliwe do wykonania, gdyby nie posiadano drobiazgowych danych o demograficznym składzie

ludności, co mogło być osiągnięte na podstawie przeprowadzonych badań ludnościowych. Dokonywane różnemi czasy w państwach Wschodu i Europy prace statystyczne miały na celu zbieranie wiadomości o stanie ludności, o jej strukturze, o sile podatkowej i zbrojnej mieszkańców i t. p., co nosiło na sobie cechy badań powszechnych, masowych, stanowiących i dzisiaj przeważnie treść statystyki, jako nauki i metody.

Na katedrę uniwersytecką statystykę jako odrębny przedmiot wykładów wprowadzono dopiero w Niemczech w XVII w. Pionierem statystyki, jako nauki uniwersyteckiej, był Conring, który wykłady tego przedmiotu rozpoczął w r. 1660 na uniwersytecie w Helmstedt, poświęcając jej pracę zatytułowaną „Notitia rerum publicarum”. Ale nazwa „statystyki” ukazuje się jeszcze później. Twórcą jej jest również Niemiec Gottfried Achenvall, który nową gałąź wiedzy, jej pożytek i prawo istnienia w systemie uniwersyteckich wykładów obronił w publicznej dyssertacji („Dissertatio de notitia rerum publicarum Academiae vindicata”), zdobywając dla siebie możliwość wykładania tego przedmiotu w uniwersytecie w Getyndze. Achenvall’a Niemcy nazywają ojcem statystyki. Dzięki niemu oraz jego uczniom statystyka już w owym czasie zdobywa sobie szoroką popularność i przy nowej nazwie nazawsze pozostaje. Jednak samo pojęcie statystyki, zarówno za Achenvall’a, jak i za czasów starożytności, t. j. w okresie empirycznych badań statystycznych, jest z punktu widzenia stanowiska dzisiejszej nauki — tylko przejściowem.

Statystyka w starożytności.

Jako praktyczna umiejętność zdobywania wiadomości liczbowych, dotyczących potrzeb państwa, statystyka była znana w starożytnych Chinach, Indjach, Persji, Egipcie, Palestynie, Grecji i Rzymie. Herodot podaje, że jeden z władców starożytnego Egiptu podzielił swoje państwo na równe części i rozdarował ziemię między poddanych, zobowiązując ich do pewnych opłat pieniężnych z tytułu użytkowania ziemi. Ażeby osiągnąć powyższy cel fiskalny, władca ten musiał posiadać do-

kładne wiadomości o liczbie ludności, o jej podziale na produkcyjną i nieprodukcyjną, o liczebności rodzin i t. p. To wszystko mogło być stwierdzone przez spisy ludności, stanowiące czynności par excellence statystyczne.

Żydzi po wyjściu z ziemi egipskiej, w ciągu 40 lat błąkania się po pustyni, byli kilkakrotnie spisywani przez Mojżesza i Aarona, a później i przez kapłana Eleazara przy przejściu przez Jordan do Ziemi obiecanej. Jedna z 5 ksiąg Mojżesza, mianowicie czwarta, nosi nazwę „księgi liczb” (Numeri). Zawiera ona „policzenie” Żydów w zakresie powinności wojskowej i obowiązku płacenia podatków. Mojżesz i Aaron „zgromadzili wszystką rzeszę ludności i spisywali ich według rodów i domów, i rodzin, i głów, i imion każdego, od dwudziestego roku i wyżej”. Spis ludności u Żydów odbył się także za czasów króla Dawida. Spis ten trwał 9 miesięcy i 20 dni.

W dziejach starożytnej Grecji znajdujemy dowody historyczne dokonywania spisów ludności i w następstwie tego — podział jej na klasy społeczne. Za Likurga, Solona, Peryklesa i in. odbywały się w Grecji poważne operacje statystyczne. W „Polityce” Arystotelesa mamy opis 157 państw i miast greckich, scharakteryzowanych przez Arystotelesa bardzo szczegółowo i dokładnie ze strony zarządu państwowego, stanu finansów, policji, nauki i sztuki. Późniejsi państwowoznawcy — Conring, Achenvall, Schlözer i in. — znajdują w Arystotelesie jak gdyby swego najznakomitszego poprzednika. Potrzeby — głównie skarbu i wojska (rejestracja mężczyzn zdolnych do noszenia broni) oraz — reformy społeczne były w starożytności temi pobudkami, które zmuszały państwo do liczenia ludności i jej poszczególnych warstw.

W starożytnym państwie rzymskim za Serwjusza Tuljusza obywatele Rzymu obowiązani byli wobec upoważnionych do tego osób składać zeznania, dotyczące imienia, wieku i in. szczegółów personalnych, oraz legitymować się z posiadanego majątku — pod przysięgą na zabezpieczenie dokładności — wraz z podaniem otrzymywanego dochodu. Był to t. zw. cenzus rzymski, a osoby które go przeprowadzały, nosiły miano cen-

zorów. Z początku cenzus obejmował tylko ludność, posiadającą pełne prawa obywatelskie, czasem został rozszerzony i na niewolników, stanowiąc źródło wiadomości, niezbędnych dla właściwego rozłożenia ciężarów państwowych i podatków. Marek Aureljusz zarządził rejestrację urodzeń, żądając od każdego obywatela Rzymu w razie przyjścia na świat dziecka do składania zeznania z podaniem imienia nowonarodzonego. Daje to początek statystyce ruchu ludności. Znakomity prawnik rzymski tej epoki Ulpianus, posiadając materiały o ruchu ludności, pierwszy układa tablicę śmiertelności, wykazującą prawdopodobieństwo długości życia dla poszczególnych klas wieku. Dzisiejsze tablice śmiertelności oparte na teorii prawdopodobieństwa w tablicach Ulpianusa znajdują swój niejako pierwowzór. Wraz z rozwojem państwa rzymskiego rośnie znaczenie statystyki administracyjnej. Wiadomości nagromadzają się. Powstaje konieczność ogłaszania ich w formie zbiorów statystycznych, wśród których znany jest „Breviarium Augusti”, stanowiący rodzaj perjuryka Rzymu, zawierający daty o sile zbrojnej państwa, o dochodach, o stanie skarbu i t. p. Wraz z upadkiem silnego organizmu państwowego upada i związana z nim statystyka administracyjna, co cechuje cały okres wieków średnich aż do czasów Conringa i Achenvall'a.

Statystyka w wiekach średnich.

W średniowieczu, przy słabej organizacji państwa, znajdujemy tylko bardzo nieliczne dowody wykonywania prac statystycznych. Jedną z takich prac jest „Breviarium fiscalium” Karola Wielkiego, sporządzone w r. 812, stanowiące opis dóbr królewskich. Pomieszczono tam dane, dotyczące domów mieszkalnych, znajdujących się w dobrach królewskich, budynków gospodarskich, ziemi uprawnej, zbiorów zbóż, ziarna siewnego, inwentarza żywego, drobiu, zapasów żywności i t. p. Oprócz tego spisowi podlegali chłopci poddani oraz posiadany przez nich majątek nieruchomy, ilość produktów, niezbędnych dla ich przeżywania, wreszcie ilość drzew owocowych i różnych roślin.

Za Wilhelma Zdobywcy w Anglii w celu zabezpieczenia dochodów królewskich został wydany „Doomesday - book”

(r. 1086)), zbiór wiadomości statystycznych, dotyczących topografii Anglii, jej urządzenia państwowego, dochodów króla i całej osiadłej ludności. Komisarze królewscy zbierali dane pod przysięgą i przeprowadzili spis — z małymi wyjątkami — na całym terytorjum ówczesnej Anglii. Większe majątki były opisane bardzo szczegółowo: oprócz imion właścicieli, notowano ile oni posiadają ziemi, zdolnej do uprawy i uprawianej, ile lasów, łąk, pastwisk, młynów, stawów rybnych, ile ziemi oddają w dzierżawę, ile posiadają inwentarza, ile wart pieniądze majątek i t. p. Te prace katastralne przeprowadzono między rokiem 1083 — 86, wyniki spisano na pergaminie w 2 tomach. Za przykładem Wilhelma Zdobywcy mieli iść jakoby i niektórzy inni późniejsi władcy Anglii w przeprowadzaniu podobnych prac statystycznych.

Nieco inne, na odmiennych zasadach oparte, były próby opracowań statystycznych republik włoskich z wieków średnich. Najgłośniejsza z nich republika Wenecji na mocy dekretów z końca XIII stulecia ustaliła porządek, że każdy, kto był posłany za granicę bądź w charakterze posła, bądź wysłannika handlowego, obowiązany był po powrocie do kraju do złożenia relacji z odbytej podróży z podkreśleniem spostrzeżeń nad politycznym, społecznym i fizycznym ustrojem zwiedzanego państwa. Na początku wieku XV zażądano, aby te i tem podobne sprawozdania (naprz. sprawozdania administratorów poszczególnych prowincyj republiki) były składane na piśmie. Te sprawozdania nosiły nazwę „relationi” i były pod względem formy i zasobu wiadomości — w swoim czasie doskonałe.

Znany jest następujący fakt z życia Wenecji. W r. 1421 Wenecja połączywszy się militarnie z Florencją, nosiła się z zamiarem wypowiedzenia wojny Medjolanowi. Jednakże zamiary speliły na niczem. Gdy na radzie jeden z członków rady, opierając się na „relationi” dowiódł, że Wenecja znaczną część swych produktów zbywa w Lombardji, zapęły militarne ostygły, bowiem stało się jasnem, że wojna Wenecji

z Medjolanem, byłaby to wojna producenta z konsumentem. Cyfry były tak przekonywujące, że udaremniły wojnę.

W innych krajach tej epoki odnajdujemy dowody, świadczące o zbieraniu dat statystycznych. W wieku XVI Filip II, król hiszpański, zarządza ułożenie zbioru wiadomości statystycznych na podstawie kwestjonariuszów, skierowanych do zarządzających poszczególnymi obwodami administracyjnymi królestwa. Za Henryka IV dzieje się podobnie we Francji w zakresie opisu środków skarbu i warunków obrony państwa w relacjach ministra de Sully. Na początku XVI stulecia Sobór Trydencki w r. 1503 nakazuje prowadzenie ksiąg kościelnych celem rejestrowania chrztów, ślubów i pogrzebów (urodzeń, małżeństw i zgonów). Obowiązek prowadzenia ksiąg metrycznych ustala porządkiem prawnym w tymże wieku Henryk VIII dla Anglii (r. 1538).

Pomimo częstokrotnych, w licznych krajach przedsiębiornych, obliczeń statystycznych stan wiadomości statystycznych w wiekach średnich był niski. Tak, naprz. w zakresie najważniejszego zagadnienia — liczby ludności — cyfry podawane przez różnych autorów różniły się pomiędzy sobą nieraz o 50% dla jednych i tych samych krajów, dając smutne świadectwo statystycznej ścisłości.

Na wiek XVI i XVII przypadada szereg prac statystycznych opisowych, zawierających słowną charakterystykę krajów i państw ówczesnej Europy, głównie ze stanowiska ich ustroju politycznego. Znane są również jeszcze wcześniejsze opracowania *Piccolomini*, późniejszego papieża Piusa II, i *Macchiavelli*, podające podobną treść. W połowie XVI w. (r. 1562) ukazuje się praca Franciszka *Sansovino*, w której znajdujemy opis 20 i kilku państw, w tej liczbie i Polski na 9-tym miejscu. Była to znakomita praca ówczesnej epoki, uważana przez wielu za pierwsze naukowe opracowanie statystyczne, aczkolwiek w układzie materiału daleka jeszcze od jednolitości i systematyczności, gdyż oprócz opisu ustrojów państwowych zawiera ona wiadomości historyczne, ekonomiczne, religijne etc. Nieco później (r. 1589) pisze Giovanni *Botero* „le relationi universali etc.”,

cieszące się niezwykle uznaniem u współczesnych, gdyż w ciągu półwieku doczekało się 12-tu wydań. *Relationi* zawiera opis Europy, Azji, Afryki pod względem terytorjum, ustroju politycznego i religii.

Prace o zakresie i charakterze, jak powyżej, ukazują się w tym czasie również we Francji i Holandji. Między innemi państwami europejskiemi Polskę opisał Jean *de Loet* w pracy statystycznej, którą wydano w Amsterdamie (de Laet zmarł w roku 1649).

Ze stanowiska krytyki prace te niewiele się różnią od wzorów włoskich, które naśladowały.

Statystyka w Niemczech w XVII i XVIII w. Znaczny krok naprzód czynią dopiero prace statystyczne XVII wieku w Niemczech. Na czoło uczonych statystyków ówczesnych wysuwa się tu znakomity Hermann *Conring* (ur. 1606 — zm. 1681), który w okresie studjów uniwersyteckich pracował nad teologią, medycyną i filozofją i początkowo piastował katedrę medycyny na uniwersytecie w Helmstedt, z czasem dopiero objął katedrę polityki w tymże uniwersytecie. Cel nowej nauki, którą nazwał „*Notitia rerum publicarum*” widział w tem, ażeby dać niezbędne wiadomości dla mężów stanu, powołanych do prowadzenia spraw państwowych, ucząc ich sztuki zarządzania temi sprawami. Jest więc statystyka *Conringa* nauką o państwie, państwoznawstwem. Z punktu widzenia metodologicznego jest to nauka opisowa. Najważniejszą częścią tego opisu jest charakterystyka ludności, bardzo szczegółowa. Delej następuje opisanie terytorjum państwa, tak pod względem warunków naturalnych, jak wielkości i granic politycznych. Dalej jeszcze znajdujemy charakterystykę ustroju państwowego i zarządu państwem. Punktem zainteresowania badawczego tak pojętej pracy statystycznej — jest państwo.

Pod względem zasadniczego pojęcia statystyki niczem nie różni się od *Conringa* późniejszy badacz wielkiej miary i teoretyk statystyki *Achenvall* (ur. 1719, zm. 1772).

Wydana przez niego praca w r. 1749 p. t. „*Abriss der neuesten Staatswissenschaft der heutigen vornehmsten euro-*

päisichen Reiche und Republiken" uważana była swego czasu za całkowite ujęcie pojęcia i programu statystyki. Tamże pierwszy raz zostało użyte słowo „statystyka”. Przedmiotem statystyki według Achenvalla jest opisanie ustrojów państwowych ze strony pewnych ważniejszych i ciekawszych właściwości tych ustrojów, t. zw. osobliwości państwowych. Temi osobliwościami (Merkwürdigkeiten) są rzeczy, dotyczące przede wszystkim kraju i ludzi. Geograficzne położenie kraju głównie z politycznego punktu widzenia, jego wytwórczość, a dalej jego mieszkańcy pod względem ilości, podziału według płci, wieku, zawodowych czynności, stosunku do państwa, organizacja władzy, urzędów i t. p. — oto treść tego opisu. Pozostaje więc statystyka w ujęciu Achenvalla w dalszym ciągu państwowznawstwem, podobnie jak była za czasów Conringa. Achenvall uczynił z niej naukę bardzo popularną, wszczepił przekonanie, że jest ona niezbędną dla tych wszystkich, którzy interesują się życiem państwa. Nowa nauka zyskała sobie trwale fundamenty rozwoju.

Achenvall stał się twórcą szkoły statystycznej, która przez długi czas w Niemczech była panującą. W r. 1804 następca Achenvalla Schlözer wydaje pracę pod tytułem „Teorja statystyki”, która zasadniczo nie wnosi nic nowego do nauki. Sam Schlözer zatracą poczucie granicy między historją i statystyką, twierdząc, że statystyka jest zatrzymaną w ruchu historją, a historja jest statystyką w ruchu, przyczyniając się do tego, jak zresztą uczynili to i niektórzy późniejsi statystycy, że zakres badań statystycznych zamiast krystalizować się coraz bardziej, począł rozpyływać się w nieokreślonym kierunku, wkraczając w pokrewne gałęzie wiedzy. Tak, *Büsching* w swojej „Erdbeschreibung” traktował statystykę, niemal jak dział geografji. Jego liczne prace statystyczne stoją na granicy między statystyką, a geografją polityczną. Posiłkuje się *Büsching* w nich w przeciwieństwie do Achenvalla metodą porównawczą, systematyzując materiał statystyczny nie według państw, jak to czynił Achenvall, lecz według składowych czynników opisu, jak terytorjum, ludność i t. d. Jednak i *Büsching* nie wybiega

w treści opisu statystycznego poza granice Achenvall'owskich osobliwości.

Statystyka tablicowa.

Nieco odrębności nie tyle co do treści, ile co do formy wnoszą prace ówczesnego duńskiego statystyka

Anchersena. W r. 1741 Anchersen wydał pracę statystyczną, która obejmowała najważniejsze zagadnienia państwowe, ujęte w szeregi cyfr, a te z kolei w tablice statystyczne. On jest twórcą t. zw. „tablicowej”, albo „tabellarycznej” statystyki, która kładła wyłączny nacisk na układanie materiału statystycznego w tablice, przedstawiające zbiór uszeregowanych cyfr. Zadaniem tej statystyki było przedstawienie wszystkiego tego w osobliwościach państwowych, co dało się ilościowo wymierzyć i wyrazić cyfrą, głównie czynników materialnych państwa; opis był uważany za zbędny.

Między zwolennikami starej szkoły Achenvalla, a statystykami tablicowymi zawrzała gorąca walka o wyższość opisowej i liczbowej metody. Mimo zaciętości był to jednak spór jałowy, dotyczący nie treści, lecz formy opracowania statystycznego. W związku jednak ze sporem poczęło się budzić u statystyków zrozumienie wagi samej metody opracowania, co pociągnęło za sobą wydanie licznych prac, któreby można nazwać teoretycznymi, dotyczyły bowiem one zagadnienia, jak opracowywać dane statystyczne. Do takich prac w pierwszym rzędzie winny być zaliczone: wspomniana wyżej „teoria statystyki” Schlözera oraz książka Niemanna, wydana w r. 1807, dotycząca tego samego przedmiotu.

Arytmetycy polityczni.

Na innej podstawie oparte były pierwsze prace statystyczne, zbliżone do współczesnych, które powstały jed-

nak nie w Niemczech, a w Anglii. Mamy tu na myśli prace angielskich „arytmetyków politycznych”, twórców takiego kierunku w statystyce, który miał na celu zastosowanie doświadczenia także w naukach społecznych. W tym czasie, kiedy Conring rozpoczął swe wykłady „Notitia rerum publicarum” w Niemczech, John *Graunt* wystąpił w r. 1662 z pracą nad statystyką ludności

Londynu w okresie poprzedzającego 50-cio lecia, oparta na zapisach urodzeń i zgonów, która pierwszy raz w historii naszej nauki stwierdziła istnienie prawidłowości w demografii. Graunt wykazał na podstawie powyższego materiału, że liczba urodzeń chłopców stale przewyższa liczbę urodzin dziewcząt, że śmiertelność ludzka do 6 lat stanowi 36% liczby wszystkich zgonów, że ludność Londynu powiększa się tylko dzięki napływowi obcej ludności, gdyż liczba zgonów w Londynie przewyższa liczbę urodzeń i t. p. Wywody Graunta, dotyczące, jak widzimy, dziedziny ruchu ludności, naogół biorąc, szły w kierunku odkrycia istotnych praw, rządzących życiem fizycznym społeczeństwa. Nazwę „arytmetyka polityczna” wprowadził do nauki jeden z późniejszych statystyków tej szkoły *Petty* szeregiem prac, opublikowanych ku końcowi 17-go stulecia.

Szkoła Conringa i Achenwalla oparta była głównie na opisie; stąd słuszną jej nazwa statystyki opisowej lub historycznej; szkoła angielska wymagała ścisłości matematycznej w opracowaniu liczbowego materiału za pomocą przeważnie arytmetycznych obliczeń i dążyła do rozpoznania praw naturalnych w życiu społecznym i politycznym, stąd jej nazwa arytmetyki politycznej. Szkoła statystyki tego kierunku w swym dalszym rozwoju prawie całkowicie ześrodkowała się na opracowaniu tematów z zakresu ruchu ludności, ponieważ ta dziedzina zjawisk społecznych, dzięki prowadzonym systematycznie zapisom małżeństw, urodzeń i zgonów dostarczała ścisłego materiału liczbowego.

Jednakże sam *Petty* w odróżnieniu od *Graunta* przenika w inne dziedziny badań, mianowicie stosuje nową metodę do zagadnień gospodarczych. Jego to zasługą jest praca porównawcza nad rozwojem życia gospodarczego Francji i Anglii, w której wykazał wyższość ekonomiczną wysp brytyjskich nad kontynentem europejskim dzięki warunkom geograficznym pierwszych. Na tej podstawie, że Anglia ma nieograniczony dostęp do morza, że wewnątrz lądu angielskiego bogate jest w złoża mineralne, *Petty* przepowiedział przodujące stanowisko Anglii na przyszłość w handlu europejskim i przewagę jej nad

Francją w zakresie uprzemysłowienia. Historia w niedługim czasie potwierdziła trafność jego spostrzeżeń.

Petty w swoich dociekaniach statystycznych posługuje się cyframi, które skrzętnie zbiera z dziedziny życia państwowego ówczesnych narodów, nie zdradza jednak zbytniego pietyzmu dla cyfr, bowiem częstokroć dowolnie niemi operuje. Wszelako w arytmetyce politycznej należy mu się ta zasługa, że świadomie wytyka drogi nowej nauce, opierając ją na liczbach *ilości, miary i wagi* zamiast, jak się wyraża, używania „nieokreślonych i zbytecznych słów oraz intelektualnych dowodów”. W pracach Petty oceniamy zapoczątkowanie właściwej metody statystycznej.

Z imieniem jednego z następców Petty, mianowicie *Halley'a* związane jest powstanie pierwszej, nie licząc próby Ulpianusa, tablicy śmiertelności. Zjawisko tak powszechne, codzienne, jak śmiertelność, zwracało uwagę na siebie wielu arytmetyków politycznych, państwowznawców, a nawet matematyków ze względu na wielkie znaczenie jego w życiu narodów. Wielu też uczonych tego czasu dążyło do odkrycia praw, rządzących tem zjawiskiem, przynajmniej w zakresie najogólniejszym, jako to: jak wielka jest śmiertelność narodów, jaki jest stosunek pomiędzy ilością ludzi zmarłych a pozostających przy życiu i t. p. To ostatnie pojęcie nazywali średnim życiem — *vie moyenne* — dążąc do odnalezienia takiej liczby, któraby wyrażała przeciętną długość życia ludzkiego. Jeżeli stosunek żyjących do zmarłych wyrażał się naprz. liczbami 1000:300, to iloraz 33,3 miał oznaczać przeciętną długość życia.*)

*) Stosunek liczby wypadków śmierci do liczby ludności żyjącej nazywali współcześni statystycy średnim życiem — *vie moyenne* — na następującej zasadzie. W krajach z ludnością nieruchomą, t. j. takich, w których nie ma przyrostu ludności, istotnie średnia śmiertelność równa się średniej długości życia. Gdyby dla wszystkich zmarłych obliczyć liczbę przeżytych lat i podzielić sumę przez liczbę zmarłych, to przeciętnie każdy żyłby w naszym przykładzie 33,3 lata.

Naprz. w wieku lat 5 zmarło 20

„ „ 6 „ 30

W przysłanym mu materiale statystycznym z Wrocławia przez Neumanna Halley posiadał dwie liczby: liczbę urodzeń i liczbę zgonów za 5-letni okres czasu od 1687 do 1691 roku. Posiadając te 2 liczby Halley potrafił ułożyć tablicę śmiertelności metodą, którą od jego imienia nazwali później Halley'owską. Według tych danych urodzonych było 6.193, a zmarłych 5,869. Przeciętnie na każdy rok wymienionego okresu padało urodzonych 1238 osób, zmarłych — 1174 osoby, a zatem przyrost wynosił 64. Na podstawie swoich własnych spostrzeżeń Halley przyszedł do wniosku, że z liczby urodzonych w pierwszym roku zmarło 348 osób, a w ciągu następnych 5 lat — 198, t. j. łącznie 546 osób, czyli zostało przy życiu po upływie 6 lat z liczby urodzonych w pierwszym roku 692. Poczynając od 7-go roku śmiertelność się zmniejsza: a więc w 7-ym roku umiera 11 osób, w dziewiątym 6 i t. d. W ten sposób dla każdego wieku Halley obliczył stosunek śmiertelności, tworząc t. zw. tablicę śmiertelności.

Tablicą śmiertelności Halley'a pomimo jej teoretycznych braków posługiwały się przez długi czas towarzystwa ubezpieczeń.

Znaczne udoskonalenia do tablic śmiertelności na poziomie ówczesnej wiedzy wprowadził *Kerseboom*, rozporządzający obszerniejszym i bardziej ścisłym materiałem statystycznym, dotyczącym rent dożywotnich (rok 1740).

Dzisiejsze tablice śmiertelności układane są na podstawie prawdopodobieństwa zgonu, obliczonego dla wszystkich lat życia ludzkiego (do 100 lat). Na zasadzie prawdopodobieństwa zgonu oblicza się liczbę pozostałych przy życiu po upływie każdego roku po kolei, poczynając od 0 wieku do 100 lat oraz średnie trwanie (długość) życia dla każdego wieku.

w wieku lat 7 zmarło 25

„ „ 8 „ 22

Suma przeżytych lat wynosi $(5.20+6.30+7.25+8.22) = 631$, podzielona przez liczbę zmarłych 97, daje średnią długość życia 6,5 lat. Oczywiście jest to średnia długość życia zmarłych, a nie tych którzy zostali przy życiu. (Fiedercowicz. Istoria i teoria statistiki, str. 123).

Johann
Süssmilch.

Podczas gdy arytmetycy polityczni szerzyli poza granicami Niemiec nowy kierunek statystyki, w Niemczech w r. 1741 ukazuje się dzieło *Süssmilcha*, pastora, p. t. „die göttliche Ordnung etc.”, które robi znakomity krok naprzód w nauce. Süssmilch zainteresował się zjawiskami życia ludzkiego, jak urodzenia, małżeństwa i śmierci ze stanowiska porządku (prawidłowości) jaki istnieje w tych zjawiskach. Nazwał ten porządek „boskim”, t. j. pochodzącym z Wyższej woli, który jednakże potwierdzają cyfry statystyczne, dotyczące ruchu ludności. Treść książki doskonale ilustruje pełny tytuł, jak następuje: „Porządek boski w zmianach rodzaju ludzkiego, t. j. zasadniczy dowód Opatrzności Bożej w stosunku do rodzaju ludzkiego z porównania urodzonych i zmarłych, wstępujących w związki małżeńskie i rodzących się, a w szczególności ze stałego stosunku rodzących się chłopczyków i dziewczynek i t. d.”

Süssmilch pierwszy ze statystyków nie tylko stwierdził ale i dowiódł ze stanowiska faktycznego materiału cyfrowego, że w stosunkach ludzkich, dotyczących życia i śmierci, związków małżeńskich, proporcji narodzin, płci etc. istnieje prawidłowość, t. j. że istnieją pewne prawa, rządzące temi zjawiskami. Z tego względu zasługi w jego nauce są ogromne, gdyż w zjawiskach pozornie dowolnych, jak zjawiska ruchu ludności, stwierdził obecność regularności, porządku, wyraźnie występujących w cyfrach statystycznych. Zakres badań statystycznych został rozszerzony przez niego na zjawiska, dotyczące biologicznych zjawisk człowieka w zrozumieniu głębszem, aniżeli u arytmetyków politycznych, bowiem materiał statystyczny podlegał opracowaniu w celu wysnucia uogólnienia t. j. wniosków naukowych. Dzieło Süssmilcha wyprzedziło prace współczesnych państwowznawców i arytmetyków politycznych pod względem metody i obszaru zagadnienia: podczas gdy arytmetycy polityczni dotyczyli poszczególnych monograficznych kwestji z zakresu teorii ludności, Süssmilch pragnął ogarnąć całość naukowego zjawiska, przy zastosowaniu indukcyjnej metody badania, możliwie obiektywnie przeprowadzonej. Die göttliche Ordnung przez

późniejszych badaczy zostało ocenione jako pierwsza praca z zakresu teorii ludności. Naukowe znaczenie prac Süssmilcha jest wielkie.

Statystyka w XIX w. Quetelet.

Od czasu pracy Süssmilcha, w swoim czasie niedostatecznie ocenionej i zostającej bez naśladowców, dopiero w połowie XIX stulecia statystyka wstępuje na tory dalszego znakomitgo rozwoju, t. j. rozszerzenia przedmiotu badań, pogłębienia i skierowania we właściwe łożysko, a to za sprawą genialnego uczonego belgijskiego Adolfa *Quetelet'a*, którego wpływ naukowy sięgnął daleko poza granice statystyki. Quetelet jest twórcą matematycznego kierunku w statystyce. On sam był z wykształcenia matematykiem, fizykiem i astronomem, a posiadając przygotowanie przyrodnicze, przeniósł metodę badań przyrodniczych w dziedzinę badań statystycznych. Przedmiotem badań Queteleta stał się człowiek jako osobnik, naród i ludzkość ze stanowiska średnich właściwości, czyli przeciętnego typu. Statystykę Quetelet pragnął podnieść na wyżyny ścisłej nauki w tem znaczeniu, w jakim rozumiemy nauki przyrodnicze, np. fizykę. Dociekania nad naturą człowieka pojedynczego, budową społeczeństwa, czyli układem społecznym Quetelet zawarł w dziele p. t. „*Sur L'homme et le developpement de ses facultés ou essai de Physique sociale*” (1835 r.), t. j. „O człowieku i rozwoju jego zdolności czyli próba fizyki społecznej”. Swoją statystykę Quetelet nazwał fizyką społeczną, t. j. nauką ścisłą, posiadającą prawa na podobieństwo praw fizycznych, dokładnych i nienaruszalnych, odnoszących się do rozwoju człowieka i społeczeństwa. W badaniach swych Quetelet stosuje teorię średnich wielkości do spostrzeżeń nad człowiekiem, tworząc system teorii człowieka przeciętnego (*l'homme moyen*). Ów *l'homme moyen* jest to typ człowieka, dokoła którego wahają się w przeciwnych kierunkach wszyscy ludzie danego narodu. Elementami poznania właściwości takiego człowieka są: wiek, wzrost, wysokość głowy, obwód czaszki, szerokość piersi, wielkość ręki, nogi, obwód piersi, waga, tętno,

oddech i t. p. Elementy poznania fizycznego człowieka winny być uzupełnione badaniami w kierunku jego moralnych i politycznych własności.

Quetelet zdaje sobie sprawę, że w celu posiadania ścisłych wiadomości o człowieku, należy mieć liczne spostrzeżenia. Wartość licznych spostrzeżeń Quetelet na przykładzie tłumaczy w ten sposób. Jeżeli, powiada, nakreślić kredą na tablicy koło, to na bardzo bliskiej odległości dostrzegać będziemy poszczególne punkty koła, oddzielone różnymi odległościami, lecz wystarcza oddalić wzrok na dostateczną przestrzeń ażeby poszczególne punkty zwały się w jeden wizerunek. Tak, poszczególne zjawiska w życiu społecznym mogą czynić wrażenie dowolnych wypadków, lecz połączone w masie tworzyć będą pewne prawidłowości, zupełnie wyraźnie występujące. W tem uogólnieniu został wytknięty kierunek metodologiczny dla statystyki.

Na podstawie licznie wielkich badań Quetelet 'dalej stwierdza istnienie prawa przyczynowości. To, co jest w tej chwili, a więc teraźniejszość — jest przyczyną przyszłości, tak jak przeszłość jest przyczyną teraźniejszości. W układzie społecznym, w zjawiskach, dotyczących człowieka, Quetelet 'żąda doszukiwania się przyczyn, powodujących te zjawiska. Postępuje więc odmiennie, jak Süssmilch. Podczas, gdy Süssmilch *à priori* przypuszcza istnienie boskiego porządku w zjawiskach społecznych i tylko znajduje potwierdzenie w cyfrach statystycznych, że podobny porządek boski istnieje, Quetelet stara się rozważać pytanie: dlaczego tak jest, jak wskazuje nam obserwacja rzeczywistości? Owo — dlaczego — tłumaczy się prawem przyczynowości. Obserwowanie faktów rzeczywistości naprowadza nas na stwierdzenie przyczyn, powodujących te zjawiska. Najlepiej to tłumaczy matematyk Laplace, współczesny Queteletowi, autor słynnej „Próby filozoficznej o prawdopodobieństwie” (r. 1814). Mając urnę, wypełnioną białymi i czarnymi kulami, czynimy doświadczenia nad wyjmowaniem z urny kulami, notując, ile kul czarnych, a ile białych wyciągnęliśmy. Przy małej liczbie doświadczeń, rezultaty będą chaotyczne, lecz w miarę powiększania liczby doświadczeń, będą coraz

bardziej odpowiadały ustosunkowaniu kul w urnie. To znaczy, że mając fakty, ustalone na drodze masowego badania (ilość wyciągniętych kul z urny) możemy sądzić o przyczynach, które je spowodowały (ustosunkowanie kul w urnie). Innymi słowy, statystyka daje nam możliwość sążenia ze skutków (zaobserwowane fakty) o przyczynach, powodujących te skutki.

Quetelet odróżnia dwa rodzaje przyczyn: stałe i przypadkowe. Te pierwsze działają w określonym kierunku, te drugie wywołują odchylenia od typu przeciętnego i działają w rozbieżnych kierunkach. Przy dostatecznie wielkiej ilości doświadczeń przyczyny przypadkowe, czyli jak je Quetelet nazywa, perturbacyjne, wzajemnie zobojętniają się, gdyż suma odchyłeń przypadkowych w jednym kierunku, winna być równą, jak wykazuje doświadczenie, sumie odchyłeń przypadkowych w drugim kierunku. Rezultat pozostanie pod wpływem przyczyn stałych. Tak, przy doświadczeniach z wyjmowaniem kul, ilość kul w urnie jest przyczyną stałą, rezultat doświadczeń będzie odpowiadał liczebnemu ustosunkowaniu się kul w urnie, natomiast przyczyny przypadkowe (chwilowy układ kul w urnie, nieprawidłowe ruchy ręki i t. p.) mogą wpłynąć na pewne wahania w rezultatach, ale tylko przy małej liczbie doświadczeń. Wyniki ogólne są powodowane t. zw. prawem liczb wielkich. Uwzględniając powyższe, wypływa, jako wniosek, naturalne dążenie statystyki do spostrzeżeń masowych, wykluczających oddziaływanie przyczyn perturbacyjnych.

Quetelet nie poprzestaje na stwierdzeniu praw, rządzących fizycznymi własnościami człowieka. Szuka tych praw i w dziedzinie moralnej i umysłowej człowieka. Jego to wielką zasługą jest rozszerzenie badań statystycznych na zjawiska moralne.

Badania nad przestępczością we Francji doprowadziły Quetelet'a do stwierdzenia istniejącego prawa, rządzącego pozornie dowolnymi czynami, a jednak zadziwiająco stałymi, w dziedzinie moralności. To stwierdzenie pozwoliło mu wygłosić zdanie, uderzające swym pesymizmem: „istnieje budżet, wypłacany z przerażającą ścisłością, jest to budżet więzień,

galer i rusztowań". Człowiek, według Quetelea, obarczony jest wrodzoną skłonnością do przestępstw — *penchant au crime* — które sprawia, że pewna ilość przestępstw będzie spełniona w społeczeństwie mimo woli człowieka. Przestępstwo jest niejako pobocznym produktem działalności człowieka, jest daniną, wypłaconą na ołtarzu konieczności. Stąd pojęcie winy musi być zreformowane, pojęcie wolnej woli człowieka poddane krytycznej ocenie.

Znaczenie prac Quetelet'a w dziedzinie statystyki moralności bynajmniej nie może być mierzone miarą powyższych wniosków pesymistycznych, szybko się bowiem okazało, że Quetelet rozporządzał zbyt małym materiałem statystycznym i znaczył krótkiego okresu czasu, ażeby wnioski mogły utrzymać się wobec późniejszych badań, dotyczących zagadnienia wolnej woli w związku z przestępczością (G. Mayr). Ale próba odnalezienia praw dla świata moralnego pozostanie nazawsze świadectwem geniuszu Queteleta. Liczne późniejsze prace statystyków niemieckich, jak Wagnera, Drobischa, Oettingena, Mayra i in. rzuciły bardzo wiele nowego światła, za pośrednictwem badań statystycznych, na te zagadnienia.

Oprócz wyżej wspomnianych zasług Quetelet'a dotyczących metodologii statystyki, jego teorii przeciętnych, statystyki moralności i t. p., należy dorzucić zasługi jego na polu statystyki administracyjnej, w szczególności organizacji spisów ludności. W r. 1846 odbył się w Belgji w koncepcji Quetelet'a pierwszy jednodniowy spis ludności, będący wzorem późniejszych jednodniowych spisów ludności w całej Europie. Quetelet był również inicjatorem i pierwszym organizatorem t. zw. międzynarodowych Kongresów statystycznych, które dla statystyki, jej terminologii, ujednostajnienia metod zbierania i opracowania materiału statystycznego, porównania statystycznego i t. p. zrobiły bardzo wiele. Mimo, że z licznych teorii Queteleta niektóre prawie zupełnie upadły, jak teoria przeciętnego człowieka, teoria budżetu przestępstw i t. p. stanowisko Quetelet'a w nauce jest pierwszorzędne i winien być on uważany

za najzakomitszego przedstawiciela tej gałęzi wiedzy w czasach najnowszych.

Wybitna naukowa indywidualność Quetelet'a, poparta szeregiem śmiałych prac, pociągnęła za sobą plejadę epigonów. Jedni z nich, jak Wagner, Drobisch i in., podjęli prace z dziedziny moralności, doprowadzając myśli Quetelet'a do krańcowych stanowisk (Wagner w dziedzinie zjawisk moralności i zagadnienia wolnej woli), inni jak np. Lexis, Bowley, Yule w kierunku rozwoju statystyki matematycznej, zapoczątkowanej przez Queteleta, jeszcze inni, jak Mayr, w kierunku uczynienia ze statystyki nauki o społeczeństwie.

Pojęcie statystyki.

Krótki rzut oka na przebieg historycznego rozwoju statystyki pozwala nam stwierdzić, że statystyka, jako gałąź wiedzy, poczynając od czasów najdawniejszych, a kończąc na Quetelet'ie i Mayr'ze przeszła ogromną ewolucję. Z empirycznych czynności statystycznych dosięga szczytów ścisłej nauki, posiłkującej się swoistą metodą ilościową. Coraz wyraźniej przedmiotem badań statystycznych staje się człowiek, już nie pojedynczo wzięty, jak nawet za czasów Quetelet'a, ale człowiek wzięty w skupieniu jako masa społeczna. Bawarski statystyk Jerzy *Mayr* w licznych doskonałych pracach statystycznych przeprowadza dowód, że statystyka jest nauką o społeczeństwie w zakresie pewnych cech, które są cechami zbiorowemi ludzkości, dostępne do odkrycia i usystematyzowania w naukowej klasyfikacji na drodze masowych ilościowych badań. Ten znakomity uczony nowszych czasów w pracy swej p. t. „Statistik und die Gesellschaftslehre” dał doskonały i niemal skończony system statystyki w powyższym pojęciu. Statystyka, zdaniem jego, jest to nauka społeczna, której przedmiotem badania są masy społeczne w zakresie zjawisk zbiorowych. Posiłkuje się ona odrębną, opartą na indukcji, metodą, zwaną statystyczną. Mayr traktuje statystykę jako naukę, która posiada własny przedmiot i własną metodę.

Atoli stanowisko Mayr'a nie jest powszechne. Inny kierunek statystyki, którego typowym przedstawicielem nazwiemy

Meitzen'a, odrzuca zarówno achenvallowską ideę państwowznawstwa, jako nauki mającej za cel opisanie osobliwości państwowych, jak również i teorię Mayr'a i Schnapper-Arndt'a, wychodząc z założenia, że statystyka jest tylko metodą badania, bynajmniej nie wyłączną w badaniach nad zjawiskami zbiorowymi społeczeństwa, lecz przydatną do zastosowania i w innych działach wiedzy ludzkiej, np. w przyrodoznawstwie, prawie, historii, językoznawstwie i t. p.

Spór o to, czy statystyka ma odrębny przedmiot badania, czy jest tylko metodą naukową — pozostawimy na stronie, poprzestając na zaznaczeniu zagadnienia.

Prof. *Czuprow* zwraca uwagę, że współcześnie pod wyrażeniem „statystyka” rozumie się równorzędnie trzy pojęcia: przedewszystkiem naukę opisową, która bada w określonych warunkach miejsca i czasu zjawiska masowe życia społecznego; z kolei — teorię specjalnej metody badania stosunków pomiędzy rzeczami w sferze wolnych przyczynionych związków, wreszcie nazwą statystyki obejmują wszelkie konkretne zastosowania liczbowego ujęcia tematu w tych lub innych zagadnieniach.

Tischer w swej najnowszej pracy (r. 1929) wymienia aż 7 znaczeń słowa „statystyka”, z których na pierwszym miejscu stawia określenie, dane przez Mayr'a, na drugim określenie statystyki jako metody, a dalej idą określenia mniej-szej wagi.

Podział statystyki
jako nauki.

Biorąc pod uwagę przedmiot w statystyce i uzględniając zakres materialny zjawisk, odróżniamy: statystykę *ludności*, zwaną demografią, lub demologją; statystykę *ekonomiczną*, obejmującą statystykę rolnictwa, przemysłu, handlu, finansów etc.; statystykę *administracyjną* (urządzenia statystyczne, organizacja statystyki państwowej, komunalnej i t. p.) i wreszcie statystykę *fizykalną* (zastosowanie metod statystycznych do biologji, antropologji, meteorologji i in. nauk przyrodniczych). Z powyższych działów, niewyczerpujących całkowicie materialnej statystyki, najlepiej opracowana jest demografia.

Metoda statystyczna.

Przedstawiwszy ewolucję pojęcia statystyki w rozwoju historycznym, za trzymujemy się na bliższej charakterystyce metody statystycznej, jako bezpośrednim celu naszej pracy. Jedną z najważniejszych metod poznania, jako takiego, stosowaną przede wszystkim w wielkim odłamie wiedzy ludzkiej — w przyrodoznawstwie, jest obserwacja, czyli spostrzeganie. Spostrzeganie stosujemy w botanice, zoologii, fizyce, chemii, meteorologii, geologii, mineralogii i t. d., ale nie tylko; stosujemy je również do zjawisk zbiorowych życia człowieka, które zawierają: demografja, ekonomja, socjologja, historja i t. d. Nauki wspomniane w szerokim zakresie posługują się spostrzeganiem, ustalającem bądź pojedyncze fakty, bądź grupy faktów. Poznanie rzeczywistości dotyczy zwykle albo wszechstronnych cech jednego przedmiotu, albo tylko pewnych, niektórych cech, należących do bardzo wielu przedmiotów danej zbiorowości. Jeżeli przyrodnik bada okaz rośliny, to opisuje możliwie wyczerpująco wszystkie cechy, które są dla danego okazu charakterystyczne; są to cechy główne, wyróżniające. Właściwość tych cech będzie taka, że dla wszystkich okazów czy osobników tego samego gatunku obecność ich jest niezbędna. Np., trzykomórkowe serce u żab, ilość pręcików w rodzinie Cruciferae, skład chemiczny wody H_2O i t. d. Te cechy wyróżniające, będące dla przyrodników podstawą naukowej klasyfikacji, są cechami gatunkowymi i stanowią głównie przedmiot badań przyrodniczych. Obok tych cech istnieją cechy inne, charakterystyczne dla części zbiorowości danego gatunku, lecz niejako ukryte, indywidualne, dające się zauważyć przy badaniu wielkich mas pojedynczych przedmiotów, a występujące w pojedynczych okazach w formie mało pozornie znaczących zboczeń, czy odchyień od przeciętnego typu. Te cechy indywidualne, jak wzrost, waga, tętno i t. p. człowieka, jak barwa lub układ plamek na skrzydle motyla, jak ilość poszczególnych listków na złożonym liściu akacji (*Robinia pseudoaccacia*) i t. p. mogą również być poddane spostrzeganiu i służyć za przedmiot naukowego badania. Wa-

runkiem podobnego spostrzegania, istotnym i nieodzownym, będzie wielka ilość przedmiotów lub osobników, u których wspomiane cechy mają być liczone. W pojedynczych wypadkach obserwacja cech indywidualnych nie daje żadnego materiału do naukowego uogólnienia i stanowi mało znaczący szczegół, przeciwnie zliczenie cech indywidualnych u wielkiej liczby osobników może pozwolić na zauważenie istnienia pewnej prawidłowości w występowaniu tych cech lub wypadków. Przedmiotem obserwacji, którą nazywamy statystyczną, będą nie cechy gatunkowe w przyrodzie, o których wspominaliśmy wyżej, a które przeważnie interesują przyrodników a obojętne są dla statystyków, a te drugie — indywidualne, zmienne, pozornie drobne i nic nie znaczące. Jeżeli zaś chodzi o zjawiska społeczne, jak ruch ludności (urodzenia, zgony, małżeństwa), jak przestępstwa, jak stosunki gospodarcze i t. p., to obserwacji statystycznej będą podlegały te cechy, które można nazwać zbiorowymi, t. j. obejmującymi większą liczbę osobników, lub wypadków, lecz charakteryzujące się dowolnością.

Cechy gatunkowe są stałe. Poza bardzo nielicznymi wypadkami, świadczącymi o jakiejś anomalji, o patologicznym objawie, cecha gatunkowa towarzyszy gatunkowi wszędzie, bez względu na szerokość geograficzną, jest obecną u wszystkich osobników danego rodzaju.

Cechy indywidualne natomiast są zmienne. Lecz pozornie tylko ukazują się one u tych lub innych osobników bez planu, chaotycznie. Dopiero u większej liczby osobników można zauważyć, że i one należą do cech charakterystycznych, na pierwszy rzut oka tylko dowolnych i niemających znaczenia, lecz w gruncie rzeczy istotnych.

Już z tego samego wynika, że obserwując cechy zmienne należy liczyć ilość wypadków, w których one występują, że więc pierwszym warunkiem odkrycia ich znaczenia jest pewna znaczna ilość spostrzeżeń. Pojedyncze, jednostkowe spostrzeżenia nie posiadają wartości.*)

Obserwacja statystyczna w pierwszym rzędzie jest obserwacją ilościową. Winna ona obejmować zbiorowości, masy, w których dopiero można dostrzec cechy zbiorowe. Cechy zbiorowe z kolei mogą pozwolić po naukowem zbadaniu, t. j. po opracowaniu statystycznym, na ustalenie zbiorowych prawidłowości, a te ostatnie mogą nam wskazać na zbiorowe współzależności.

Niema dzisiaj dziedziny wiedzy ludzkiej, gdzieby nie można było stosować badania zbiorowości. Człowiek, pojedynczo wzięty, podlega badaniu statystycznemu ze względu na cechy indywidualne, któremi obficie jest wyposażony; człowiek w skupieniu, jako ludność danego kraju, kontynentu, czy całego świata — ujawnia cechy zbiorowe, zmienne, wspólne większej lub bardzo dużej ilości osobników; człowiek, wzięty jako część pewnej specjalnej zbiorowości, np. ludzie jednego zawodu,

*) Jako ilustrację powyższej tezy t. j. że niektóre zjawiska, jeżeli je pojedynczo bierzemy, nie posiadają wartości, lecz w masie posiadają wartość, gdyż ujawniają ukrytą od oka prawidłowość w swem występowaniu, można przytoczyć szereg potocznych przykładów. Weźmy fakt trafilania się wypadków wysyłania listów bez adresu. Korespondencja bez adresu jest zjawiskiem możliwem, lecz absolutnie dowolnem i stosunkowo dość rzadkiem. Przy małej liczbie spostrzeżeń w tym kierunku nie można wyobrazić sobie jakiegokolwiek prawidłowości. Lecz na 1.000.000 korespondencyj takich wypadków zanotowano w Rosji przedwoj. 26. Ten stosunek powtarzał się z podziwu godną prawidłowością w ciągu szeregu kolejnych lat, o czem wspomniamy na Innem miejscu.

Ilość dorożek niezajętych, odchodzących z dworca kolejowego w Berlinie, była dla każdego poszczególnego dnia zupełnie dowolną, lecz po podsumowaniu dla całego roku okazała się niemal — z roku na rok — wielkością stałą dla każdego z istniejących dworców kolejowych.

Moneta opatrzona orłem i reszką, padając zupełnie dowolnie, daje orła lub reszkę. Przy małej liczbie doświadczeń nie można dostrzec żadnej prawidłowości co do wyników doświadczenia. *Jevons* wykonał 10.000 rzutów z 10 monetami i stwierdził, że pozornie dowolny rezultat padania monet przy dużej liczbie doświadczeń podlega określonemu prawu.

Numerы wygranych loteryjnych tylko przy wielkiej liczbie losów w połowie kończą się na cyfry parzyste, a w połowie — na nieparzyste. O tych zjawiskach wspomniamy szczegółowo na Innem miejscu.

ludzie obarczeni jakimś charakterystycznym brakiem fizycznym, np. kalectwem, posiadają pewne cechy zmienne, charakterystyczne dla tej właśnie zbiorowości. Także świat organiczny i nieorganiczny poza sferą cech gatunkowych posiada cechy zmienne, indywidualne, dostępne badaniu na drodze masowej obserwacji. Przyrodnik Weissman powiada, że niema motyla, któryby we wszystkich szczegółach był podobny do drugiego motyla tego samego gatunku. Tak jest w istocie. Niema liścia Robinii pseudoaccacii, któryby był identyczny z drugim liściem tegoż drzewa.

Otoczający nas świat zjawisk jest więc światem nieobliczalnie wielkiej ilości zmiennych form. Stanowią one pole badania naukowego.

Zmienność, panująca w otaczającym nas świecie zjawisk, odznacza się zasadniczą właściwością, która czyni ją przydatną do badania naukowego: jest nią ciągłość.

Leibnitz zauważa, że przyroda nie robi skoków. Jeżeli lekarz notuje temperaturę chorego co godzinę, to czyni to w tym przypuszczeniu, że w ciągu tej godziny nie było nadzwyczajnych skoków temperatury. Wiara w ciągłość zjawisk jest podstawą badania zmienności.

Z tego co wyżej powiedzieliśmy, wynika, że badanie statystyczne, oparte na spostrzeganiu, jest badaniem ilości i stosunków, zachodzących między temi ilościami, w odróżnieniu od badania jakościowego, opartego na opisaniu.

Ażeby zrozumieć znaczenie czynnika ilości w porównaniu z jakością, przypomnijmy sobie przykład z drgania fal eteru. Przy pewnem natężeniu drgania fal powietrznych doznajemy wrażenia dźwięku; przy zwiększonym natężeniu wrażenie jakościowe dźwięku przetwarza się we wrażenie wzrokowe światła; tak, ilość jest w stanie zmienić jakość zjawiska.

Ilość statystyczna nie jest identyczna z ilością arytmetyczną, czy jakkolwiek inną matematyczną. Jest to cecha, zbliżona naturą do cechy jakościowej.

Zmienność cech wywołana jest przyczynami. Już Quetelet odróżniał dwa rodzaju przyczyn, o czem wspominaliśmy wyżej.

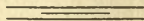
Jedne z nich działają stale i w określonym kierunku i wywołują określone zmiany w zjawiskach; drugie są perturbacyjne, zmienne, przypadkowe, działające bezplanowo, chaotycznie, bez z góry obranego kierunku. Te ostatnie przyczyny wywierają wpływ na zjawiska równomiernie w dwu przeciwnych sobie kierunkach: w kierunku ujemnym i dodatnim. Ponieważ nie można odnaleźć założenia, wedle którego przyczyna przypadkowa działałaby w jednym kierunku, np. dodatnim, przyjmujemy, że suma wpływów, pochodzących z przyczyn przypadkowych jest jednakowa, jak w jednym, tak i w drugim kierunku, dlatego wpływy przyczyn przypadkowych wzajemnie niwelują się. Ażeby nastąpiło wzajemne zbilansowanie się przyczyn, trzeba tylko, aby liczba wypadków była dostatecznie wielka. Błędy przy pomiarach, o ile pochodzą z przyczyn przypadkowych, przy wielkiej liczbie pomiarów wzajemnie bilansują się: rezultat jest bliski prawdy.

Poisson nazwał powyższą prawidłowość, wynikającą z wzajemnego bilansowania się wpływów przyczyn przypadkowych, występującą przy wielkiej liczbie spostrzeżeń, *prawem liczb wielkich*. Prawo liczb wielkich wywołuje taki stan rzeczy, że rezultat badania ilościowego, statystycznego, stosuje się do działania przyczyn stałych, uchylając działanie przyczyn ubocznych, przypadkowych. Powyższe zostało stwierdzone empirycznie na przykładzie wyjmowania kul z urny, podrzucania monet etc. Matematyczna teoria prawdopodobieństwa potwierdziła powyższe wywody. Z zasadami tej teorii zapoznamy się we właściwym miejscu.

Z tego cośmy wyżej powiedzieli, wynika, że metoda statystyczna jest metodą wielkich ilości. Jest to metoda notowania faktów, wypadków, cech we wszystkich tych dziedzinach zjawisk, w których znajduje przejaw miara, waga, liczba — innymi słowy wymiary ilościowe — przytem charakter tych zjawisk winien mieć cechy zbiorowości, zmienności, nietypowości. Takie zjawiska nazwiemy statystycznymi, zbiorowymi.

Statystyka więc, jest to metoda wielkich ilości w dziedzinie badania zjawisk zbiorowych.

Należy dodać, że użycie metody wielkich ilości w obecnym rozwoju nauk jest powszechne i coraz większe na wszystkich polach badania, nie wyłączając dziedziny fizycznej.



II. Zastosowanie statystyki.

Wyczerpać przykłady zastosowania statystyki jest niemożliwe. Jak już zaznaczyliśmy, metodę wielkich ilości użyto już niemal do wszystkich działów wiedzy ludzkiej. Długo czas przeciwstawiano zjawiska przyrodnicze zjawiskom społecznym, jako tej dziedzinie, w której stosowanie metody statystycznej jest naturalne. Ale i w dziedzinie nauk przyrodniczych statystyka okazała się bardzo cennym, a częstokroć wprost niezastąpionym narzędziem badania naukowego. Z chwilą, kiedy zwrócono uwagę na cechy zmienne świata przyrodniczego, kiedy zrozumiano, jak mówi Ostwald, że niema kryształu górskiego, któryby był podobny do drugiego we wszystkich szczegółach, albo rozstworu soli identycznej z drugą o jednym i tym samym składzie chemicznym, okazało się koniecznością zastosować metodę ilościowego badania także w naukach przyrodniczych. Ma to na celu wykrycie związków przyczynowych między zjawiskami przyrodniczymi powstałymi wskutek działania wielokrotnych przyczyn. Wiele tajników przyrody odkryto na tej drodze, ściślej biorąc, objaśniono wiele zjawisk, które do tej pory pozostawały poza kołem znanych przyczyn.

Demografia.

Najobszerniejszą dziedziną zastosowania statystyki są nauki społeczne, w szczególności demografia, czyli nauka o ludności. Ten dział statystyki słusznie można nazwać klasyczną statystyką, gdyż była ona kolebką innych nauk statystycznych, jest najsystematyczniej opracowana i doprowadzona w dziele J. Mayra p. t. „Statystyka i nauka o społeczeństwie” do stopnia dosko-

nałości. Wystarczy otworzyć pierwszy lepszy rocznik statystyki, aby znaleźć zgrupowane liczby, dotyczące stanu ludności. W pierwszym rozdziale niniejszej książki wspominaliśmy o spisach ludności już w czasach zamierzchłych; wspominaliśmy o stanie wiadomości o ludności u Egipcjan, Żydów, Greków i Rzymian; w pracach „relationi” i in. z wieków średnich napotykamy dane ludnościowe; badacze tacy jak Graunt, Halley i Süssmilch podejmują już zadania konstrukcyjne z dziedziny badań ludnościowych w celu wykrycia prawidłowości, istniejących w pozornie dowolnych zjawiskach ruchu ludności; Quetelet buduje system fizyki socjalnej z centralnem pojęciem człowieka przeciętnego, jako pewnego ideału ludzkiego i t. d. Cała historia statystyki rozwija się na terenie zastosowania metody statystycznej do badań nad skupieniami ludzkimi. W tej dziedzinie uczyniono spostrzeżenie, które jest jednym z najstalszych uogólnień statystycznych, godnym noszenia nazwy prawidłowości statystycznej, a jest niem stwierdzenie, że *liczba urodzeń chłopczyków* jest większa od liczby urodzeń dziewczynek, wahając się koło stosunku 106:100.

Statystyka ludności ustaliła spostrzeżenie, że w krajach Zachodu w ogólnej masie ludności jest *więcej kobiet, aniżeli mężczyzn*, pomimo odmiennego stosunku przy urodzeniu. U G. Mayr'a znajdujemy szeregi, dotyczące ludności Francji w okresie większej części XIX stulecia:

<i>Lata</i>	<i>mężczyzn %</i>	<i>kobiet %</i>	<i>różnica</i>
1806	49.17	50.83	1.66
1821	48.57	51.43	2.86
1831	48.97	51.03	2.06
1841	49.38	50.62	1.24
1851	49.73	50.27	0.54
1861	49.87	50.13	0.26
1866	49.95	50.05	0.10
1872	49.81	50.19	0.38

Z powyższej tablicy statystycznej widać, że największa różnica w liczebnej sile płci przypada we Francji na lata po

wielkich wojnach napoleońskich, po tem następuje prawie całkowite wyrównanie. Szereg innych badań nad tem zagadnieniem przyprowadził G. Mayr'a do przekonania, że, jeżeli chodzi o Europę, liczebności płci są prawie identyczne, że więc istnieje wielkie *prawo równowagi płci*. Inne zagadnienie - zagadnienie podziału ludności na grupy wieku—podaje statystyka spisów ludności w Niemczech według Mayr'a.

Udział procentowy grup ludności państwa
niemieckiego w latach:

<i>Grupy wieku.</i>	<i>1871</i>	<i>1875</i>	<i>1880</i>	<i>1885</i>	<i>1890</i>
do 5 lat	12.83	13,38	13.87	13,08	13,01
5—10 „	11,25	11.23	11.02	11.81	11.19
10—15 „	10.39	10.26	10.33	10.62	10.95
15—20 „	9.39	9.47	9.33	9.45	9.32
20—25 „	8.63	8.31	8.56	8.45	8.61
25—30 „	7.82	7.61	7.32	7.57	7.58
30—40 „	13.31	13.37	12.99	12.65	12.76
40—50 „	10.64	10.29	10.36	10.55	10.38
50—60 „	8.35	8.42	7.97	7.70	7.83
60—70 „	5.20	5.14	5.32	5.46	5.20
70—80 „	2.09	2.15	2.14	2.22	2.36
80 lat i wyżej	0.36	0.37	0.39	0.43	0.42

Statystyka ludności rzuca ciekawe światło na *odrębność cywilizacji Wschodu i Zachodu*. Gęstość zaludnienia narodów Zachodu, jak wiadomo, pochodzi głównie ze skupień miejskich, przeciwnie ludność Wschodu skupia się nie po miastach, jak w Europie i Ameryce, a po wsiach. Nie nowe wprowadzie, na starych cyfrach oparte, natomiast bardzo wymowne zestawienie szeregów gęstości zaludnienia i $\%$ ludności miejskiej w krajach Europy Zachodniej i Indjach azjatyckich znajdujemy u Woblyj'a *).

*) Pryw. doc. K. G. Woblyj. Statistika. Kijów 1909. Str. 258.

	gęstość załudnienia	% ludn. miejskiej	
Belgja i Niderlandy	200	33,5	Zachód
Niemcy	104	28,2	
Austro-Węgry przedwoj.	72	14,0	
Rosja europejska	19	9,8	
Bengal	190	4,2	Wschód
Madras	102	4,8	
Pendżab	65	7,8	
Bombay	53	8,4	

Mamy tutaj, zauważa prof. Wobłyj, szeregi równoległe $\frac{0/0}{0.0}$ ludności miejskiej — malejące dla Europy, dla Indji rosnące. W Indjach — wielkiej gęstości załudnienia towarzyszy mały procent ludności miejskiej i odwrotnie; widzimy tutaj szeregi statystyczne t. zw. rozbieżne.

Braki fizyczne ludności, notowane przez statystykę, są to: ślepota, głuchoniemota i choroby umysłowe. Nasz spis z r. 1921 uwzględniał kalectwo, t. j. ludzi bez rąk, bez nóg etc. U Conrad'a*) znajdujemy szeregi ślepych według grup wieku, z których widać, że ten nader ciężki brak fizyczny potęguje się u ludzi z wiekiem. W Prusach w r. 1905 na 10.000 ludności było ślepych:

Wiek	mężczyzni	kobiety	razem
— 5	0.8	0.7	0.8
5—10	1.4	1.0	1.2
10—15	2.4	1.8	3.1
15—20	3.1	2.1	2.6
20—30	3.3	2.3	2.8
30—40	4.5	3.4	3.9
40—50	8.0	5.3	6.6
50—60	12.8	9.0	10.8
60—70	21.5	18.3	19.7
wyżej 70	54.6	55.9	54.4

*) Grundriss zum Studium der politischen Oekonomie von prof. dr. I. Conrad, IV Tell. Statistik. Jena 1918 Str. 115.

Zbędne byłoby powoływanie się na dalsze przykłady z tej dziedziny. Znaczenie, potęga cyfr statystycznych w demografii jest widoczna. Metoda wielkich ilości daje możliwość odkrycia w masie faktów prawidłowości, działających w sferze zjawisk ludnościowych; ona rzuca światło tam, gdzie inną drogą dotrzeć nie można, uchyla rąbek tajemnicy w stosunkach zawikłanych, chaotycznych i ciemnych.

Jako przykład badań statystycznych w biologii przytoczymy dane bułgarskiego uczonego Bachmetjewa i polską pracę Sypniewskiego.

Biologia.

Bachmetjew *) do zagadnień biologicznych (botaniki, zoologii, entomologii etc.) zastosował metodę liczbową, którą nazwał metodą statystyczno-analityczną. Zaczerpnijmy dwa przykłady: jeden z entomologii, drugi z botaniki.

Ażeby zrozumieć zasady metody Bachmetjewa, pamiętać należy że w dziedzinie zoologii i botaniki istnieją gatunki i odmiany. Np., motyl *Aporja crataegi* jest gatunkiem *crataegi*, ale pod wpływem zewnętrznych warunków, głównie wilgoci i temperatury, w niektórych okolicach odchyła się od charakterystycznych cech typu i staje się odmianą, wśród których znane są dotychczas *Aporia crataegi* ab. *alepica*, ab. *marginata*, ab. *augusta*, ab. *melana* i in., mniej lub więcej różniące się między sobą ubarwieniem i t. p. Tak samo w dziedzinie botaniki znanych jest wiele odmian pszenicy, żyta i t. d., odchylających się mniej lub więcej wyglądem i właściwościami od gatunku. Sam gatunek nie stanowi jednak czegoś stałego i niezmiennego. Cechy jego wykazują zmienność w pewnych granicach, różnych dla różnych gatunków. Ta zmienność nie necessarily musi pochodzić od warunków zewnętrznych, a może mieścić się w indywidualnych właściwościach egzemplarza danego gatunku, mających wewnętrzną przyczynę. Tak więc, osobniki jednego

*) M. Bachmetjew. *Statistiko-analiticeskij metod i jewo znaczenie w sielskom chozajstwie*. Sofja. Odb. z „*Sielskaho chozajstwa i llesowodstwa*”, t. 233, str. 491—514.

i tego samego gatunku mniej lub więcej różnią się od siebie. Ta zmienność podlega—mówi Bachmetjew—jednemu ogólnemu prawu. Weźmy konkretny przykład.

Mamy, dajmy na to, do czynienia z rodziną pszczoł gatunku *Apis mellifera*. W jednym ulu, jak wiadomo, znajduje się królowa, kilka tysięcy robotnic i kilkaset trutniów. Wszystkie pszczoły mają na tylnych skrzydłach osobliwe haczyki, trudne do rozpoznania, a tembardziej do liczenia gołym okiem. Można dokonać obliczania przez szkło powiększające lub mikroskop, oddzielnie prowadząc rachunek dla prawych i lewych skrzydeł, gdyż ilość tych haczyków na jednych i na drugich nie jest jednakowa (cecha zmienna). Bierzemy około 100 trutniów w celu obliczenia tej cechy morfologicznej. Rezultaty:

liczba haczyków	ilość skrzydeł	
n	prawych	lewych
	f	f
17	2	0
18	3	7
19	9	10
20	12	13
21	16	20
22	21	25
23	19	11
24	7	6
25	4	2
26	0	2
	<u>93</u>	<u>96</u>

Pierwszy szereg oznacza ilość haczyków na skrzydełkach od 17 do 26; drugi i trzeci szereg oznacza ilość skrzydełek, na których znaleziono haczyki w określonej ilości; dwa ostatnie szeregi nazywamy liczebnością albo frekwencją.

Z powyższej tablicy statystycznej widać, że z powiększeniem się n liczebność stopniowo się powiększa, dochodząc do szczytu—maximum—przy $n=22$, a potem znowu się zmniejsza. Gdyby wziąć nie 100, a 200, 300 i t. d. trutniów z tego samego ula i pochodzących tak samo od 4-letniej królowej, to zno-

wu otrzymalibyśmy podobną zależność f od n , t. j. szereg, posiadający jeden szczyt liczebności.

Już Quetelet znalazł, że liczba szczytów liczebności oznacza, z ilu odmian (ras) składa się dany materiał autropologiczny, opracowywany metodą statystyczno-analityczną. Później jednak okazało się, że ta reguła jest istotną tylko dla materiału, który składa się z indywiduów, rozwijających się partenogenetycznie. Partenogeneza czyli dziewicoródstwo ma wtedy miejsce, kiedy z niezapłodnionego jajka rozwija się dane indywiduum. Tak np., z niezapłodnionych jaj pszczoł rozwijają się trutnie. Trutnie, wzięte z jednego ula jak wyżej, posiadają po jednym szczycie liczebności na prawem i lewym skrzydle, co się zgadza z teorią Queteleta, gdyż trutnie te należały do tej samej rasy, mianowicie bułgarskiej. Przejdźmy do przykładu z botaniki.

Wiadomo, że *Robinia pseudoacacia*, pospolicie zwana akacją, może wyrosnąć tylko z posadzonego nasienia, a nie z gałązki w ziemię zatkniętej. Takie drzewo stanowi produkt dwóch biologicznych pierwiastków: męskiego i żeńskiego. Należałoby oczekiwać — wobec dwupłciowości tego drzewa — że badanie morfologicznych elementów metodą statystyczno — analityczną wykaże dwa szczyty liczebności. Liczenie listeczków na złożonym liściu akacji dało następujące rezultaty:

ilość listeczków	ilość liści złożonych	
	strona prawa	strona lewa
n	f	f
1	—	1
2	2	9
3	8	8
4	24	19
5	37	30
6	23	26
7	40	35
8	51	56
9	13	13
10	1	1

Zauważamy, że istotnie jak na prawej, tak na lewej stronie złożonego liścia akacji znajdują się po dwa szczyty liczebności.

Bachmetjew zmieszał trutnie, otrzymane z dwóch źródeł, a więc różniące się od siebie pewną odmiennością form. Analiza statystyczna wykazała w szeregach liczebności dwa szczyty liczebności. W wyniku swych badań Bachmetjew przyszedł do wniosku, że ilość szczytów liczebności odpowiada liczbie form lub ras danej mieszaniny w tym wypadku, kiedy badane osobniki rozwijają się partenogenetycznie (albo w drodze samoopylecia).

J. Sypniewski *) przeprowadził badania metodą statystyczną nad odmianami ziemniaków, pszenic i buraków, czyniąc pomiary dotyczące: 1) wielkości szparek oddechowych 2) powierzchni komórek naskórka 3) średnicy komórek mezofilu i 4) ilości szparek oddechowych na jednostce powierzchni liścia. Przytaczamy następującą tablicę:

№ roślin	długość szparek odd.		ilość szparek odd. na 1 mm.	
	na górnej stronie liścia a.	na dolnej stronie liścia b.	na górnej stronie liścia c.	na dolnej stronie liścia d.
1	39.34	32.00	74	283
2	37.32	32.90	70	286
3	34.94	31.02	113	434
4	34.72	27.44	120	408
5	34.66	29.68	77	370
6	33.88	28.84	86	403
7	31.86	25.76	132	480
8	31.08	26.18	96	463
9	29.68	25.90	74	396

Elementarne opracowanie powyższych szeregów (wyprowadzenie średniej arytmetycznej z każdych 3 kolejnych klas) daje u Sypniewskiego następującą tablicę:

*) J. Sypniewski. Ustosunkowanie wielkości elementów anatomicznych różnych odmian ziemniaków, pszenic i buraków. Petersburg r. 1913.

Klasy	wielkości:			
	a	b	c	d
I	37.20	31.97	86	334
II	34.42	28.65	94	394
III	30.87	25.95	101	446

Otrzymana tablica, po przeprowadzonym, jak wyżej, opracowaniu statystycznym, jest przejrzysta i pozwala na wysnucie kilku wniosków, z których najważniejsze: 1) długość szparek oddechowych na górnej stronie liścia jest większa, aniżeli na dolnej i 2) ilość szparek oddechowych na górnej stronie liścia jest kilkakrotnie (4.2) mniejsza, aniżeli na dolnej.

W pracy swej ponadto J. Sypniewski stosuje obliczanie granic błędów pomiarowych metodą, przyjętą w statystyce.

Meteorologia.

Obserwacje meteorologiczne, t. j. notowania temperatury powietrza, wilgoci, opadów atmosferycznych etc., noszą znamiona liczb statystycznych, a przeto nadają się do opracowania statystycznego. Wśród licznych zagadnień, które mogą być rozstrząsane na tej drodze, jedno polega na określeniu, jak długo należy prowadzić notowania na stacji meteorologicznej, ażeby, dajmy na to, średniej rocznej temperatury, wyprowadzona z liczb obserwacji, była dostatecznie stała, t. j. różniła się od wieloletniej średnicy nie więcej, jak żądana dokładność — 0.5° , 0.3° , 0.1° i t. d. Podobną pracę wykonał *M. Pankow**) dla miasta Puław, skąd zaczerpujemy przykład:

Średnia miesięczna temperatura Puław za okres 38 lat od 1872 do 1909 roku była następująca:

styczeń	— $3^{\circ},1$	lipiec	$18^{\circ},8$
luty	— $2^{\circ},1$	sierpień	$17^{\circ},9$
marzec	$1^{\circ},7$	wrzesień	$13^{\circ},9$
kwiecień	$7^{\circ},9$	październik	$8^{\circ},5$
maj	$13^{\circ},9$	listopad	$2^{\circ},6$
czerwiec	$17^{\circ},6$	grudzień	— $1^{\circ},8$

*) *M. Pankow*. Klimaticzeskij oczerk Nowej Aleksandryi. Petrograd 1915

Średnia temperatura zimy wynosiła: — 2^o,3, wiosny 7^o,8, lata 18^o,1, jesieni 8^o,3, całego roku 8^o,0.

Ażeby sądzić o zmienności temperatury dla poszczególnych lat i dla poszczególnych miesięcy roku, należy obliczyć wahanie temperatury około temperatury średniej dla lat i miesięcy, inaczej—znaleźć statystyczne odchylenie od średniej arytmetycznej temperatur roku i poszczególnych miesięcy. W danym wypadku wystarczy arytmetyczne odchylenie*). Tablica arytmetycznego odchylenia temperatur miasta Puław za okres 38-letni (1872 — 1909):

styczeń	$\pm 1^{\circ},89$	lipiec	$\pm 1^{\circ},10$
luty	$\pm 1^{\circ},87$	sierpień	$\pm 1^{\circ},00$
marzec	$\pm 1^{\circ},80$	wrzesień	$\pm 1^{\circ},28$
kwiecień	$\pm 1^{\circ},60$	październik	$\pm 1^{\circ},68$
maj	$\pm 1^{\circ},44$	listopad	$\pm 1^{\circ},80$
czerwiec	$\pm 1^{\circ},26$	grudzień	$\pm 1^{\circ},98$
cały rok $\pm 0^{\circ},53$.			

Z tej tablicy wynika, że np. średnia roczna temperatura Puław pojedynczego roku może się różnić od wieloletniej średnicy na wielkość $\pm 0^{\circ},53$. Max. odchylenia pada na grudzień, min. — na sierpień.

Arytmetyczne odchylenie daje możność obliczyć t. zw. błąd prawdopodobny**) otrzymanych miesięcznych średnich według wzoru Gaussa:

$$F = 0,6745 \sqrt{\frac{\sum v^2}{n(n-1)}}$$

w którym v^2 oznacza sumę kwadratów odchyień, a n — okres, z którego wyprowadzono średnie. Jeżeli n nie jest mniejsze od 20 (jak w danym wypadku) to można posługiwać się wzorem uproszczonym

$$F = 1,1955 \frac{\sum v}{n \sqrt{2n-1}}$$

*) Metody obliczania odchylenia arytmetycznego będą podane w następnych rozdziałach.

**) Jak wyżej.

w którym ułamek $\frac{v}{n}$ jest arytmetycznem odchyleniem, które oznaczmy przez V, wobec czego

$$F = 1,1955 \frac{V}{\sqrt{2n-1}}$$

Wzór powyższy pozwoli nam obliczyć błąd prawdopodobny dla średnich miesięcznych notowań temperatury, który dla poszczególnych miesięcy będzie następujący:

styczeń	$\pm 0,27$	lipiec	$\pm 0,15$
luty	$\pm 0,26$	sierpień	$\pm 0,14$
marzec	$\pm 0,25$	wrzesień	$\pm 0,18$
kwiecień	$\pm 0,22$	październik	$\pm 0,23$
maj	$\pm 0,20$	listopad	$\pm 0,25$
czerwiec	$\pm 0,17$	grudzień	$\pm 0,28$
cały rok $\pm 0,07$.			

Jak widać, największe prawdopodobieństwo, według notowań za okres lat 38, mają w Puławach temperatury notowane dla całego roku i dla letnich miesięcy. Przy pomocy tych cyfr łatwo określić, ile lat należy prowadzić obserwacje meteorologiczne w Puławach, ażeby błąd prawdopodobny zmniejszył się do 0,1, t. j. rozstrzygnąć pytanie postawione na początku.

Ponieważ błąd prawdopodobny jest odwrotnie proporcjonalny do pierwiastka kwadratowego z liczby lat obserwacji, to

$$\frac{F \text{ (bł. prawd. otrzymany)}}{0,1 \text{ (bł. prawd. żądany)}} = \sqrt{\frac{x}{n}} \quad \text{(liczba lat obserwacji)}$$

x — jest poszukiwaną liczbą lat obserwacji.

Zgodnie z tym wzorem, ażeby otrzymać temperatury średnie z dokładnością do 0,1, należy w Puławach prowadzić obserwacje w ciągu:

dla stycznia	262 lat	dla lipca	85 lat
„ lutego	250 „	„ sierpnia	71 „
„ marca	231 „	„ września	117 „
„ kwietnia	180 „	„ października	195 „
„ maja	152 „	„ listopada	231 „
„ czerwca	110 „	„ grudnia	290 „

i dla całego roku — 18 lat.

Jak widać, 38-letni okres obserwacji meteorologicznej w Puławach jest tylko dostateczny dla wyprowadzenia dokładnej rocznej temperatury (z dokładnością do 0^o,1.), dla poszczególnych miesięcy jest zgoła niewystarczający.

Rolnictwo.

W tej dziedzinie zanotujemy zagadnienia: stałości plonów, doświadczeń polowych i intensywności rolniczej według prac Oboleńskiego, dr. M. Górskiego i L. Waściszakowskiego, opracowywane metodą statystyczną.

Obolenski *) badał plony zbóż w południowej Rosji, zadając sobie między innymi pytanie, które zboża są stałsze w plonach: jare, czy ozime? Cyfry plonów wziął dla lat 1905—12. Odchylenie arytmetyczne w liczbach % % od plonu średniego dało szeregi następujące:

Gubernie	Zboża ozime	zboża jare
Obw. Wojska Dońsk.	36.8	22.1
Jekaterynosławska	18.1	11.8
Taurydzka	13.0	9.6
Chersońska	36.8	6.8
Charkowska	16.7	20.5
Połtawska	18.5	11.6
Woroneska	21.4	34.0
Kurska	14.1	14.1
Czernihowska	20.4	14.6
Przeciętna dla całego rejonu	14.5	13.1

Analiza powyższa pozwala autorowi cytowanej pracy zakończyć z wielkim pożytkiem dla poznania rzeczywistości, że wahanie plonów zbóż jarych jest mniejsze od wahanja plonów zbóż ozimych (w warunkach danego miejsca).

Prof. Marjan *Górski* daje próbę zastosowania metody statystycznej do obliczeń wyników doświadczeń rolniczych. Zagadnienie, na które w tej chwili chcemy się powołać, pole-

*) *Urożaj chlebow w Jużnoj Rossii*, str. 41, r. 1915.

ga na tem, że powszechnie przyjęta ilość doświadczeń z plonami odmian zbóż jest zbyt mała, ażeby błąd prawdopodobny, obliczony z małej ilości spostrzeżeń (3 — 4 t. zw. poletka w doświadczeniach polowych równoległych), był wielkością zupełnie pewną. Np., błąd prawdopodobny obliczony z 4 równoległych (identycznych) doświadczeń jest obarczony błędem prawdopodobnym wynoszącym $\frac{1}{3}$ jego wielkości, czyli błąd prawdopodobny błędu prawdopodobnego jest za duży. Dr. Górski proponuje używanie błędu prawdopodobnego, obliczonego ze wszystkich poletek, wchodzących w skład doświadczenia polowego. W ten sposób obliczony „błąd prawdopodobny doświadczenia polowego” jest już wielkością dość dokładną.*)

Plony żyta w kg. z 1 morga:

	bez nawozu	kainit	super- fosfat	toma- syna	mąka kostna odklejona	kainit i su- perfosfat	kainit i to- małyna	kainit i mąka kostna odklejona
poletka równoległe	237	294	336	336	449	506	598	500
	181	316	481	495	442	701	647	630
	325	462	467	519	481	589	682	634
	244	304	505	615	551	557	654	647
	332	520						
	330	523						
średnia arytmetyczna	283	403	447	491	481	588	645	602

Prof. Górski zamiast obliczenia błędu prawdopodobnego spostrzeżeń w granicach poszczególnych poletek równoległych, proponuje wziąć wszystkie poletka razem.

Najpierw oblicza różnice między plonami średnimi a poszczególnymi i wyraża te różnice w procentach wartości średnich. Tym sposobem otrzymuje szeregi:

*) „Błąd prawdopodobny doświadczenia polowego” napisał M. Górski. 1919 r.

Stosunek odchy- lenia *) średniej	Stosunek odchy- lenia do śred- niej	Stosunek odchy- lenia do śred- niej	Stosunek odchy- lenia do śred- niej	Stosunek odchy- lenia do śred- niej	Stosunek odchy- lenia do śred- niej	Stosunek odchy- lenia do śred- niej	Stosunek odchy- lenia do śred- niej	Stosunek odchy- lenia do śred- niej	Stosunek odchy- lenia do śred- niej
- 46 16	- 109 27	- 111 25	- 154 31	- 32 7	- 82 14	- 47 7	- 102 17		
- 102 36	- 87 22	+ 34 8	+ 4 1	- 39 8	+ 113 19	+ 2 0	+ 29 5		
+ 42 15	+ 59 15	+ 20 5	+ 28 6	± 0 0	± 0 0	+ 37 6	+ 32 5		
- 39 14	- 99 25	+ 58 13	+ 124 25	+ 70 15	- 31 5	+ 9 1	+ 45 7		
+ 49 17	+ 117 29								
+ 97 34	+ 120 30								

Dalsze postępowanie dotyczy już tylko opracowania cyfr drugich szeregów. Jest to przejście od cyfr odchylenia do cyfr stosunku procentowego odchylenia do średniej arytmetycznej, t. j. do cyfr względnych, które mogą być połączone jako takie w jeden szereg. Obliczenie błędu prawdopodobnego tego szeregu daje wielkość, którą prof. Górski nazywa „błędem prawdopodobnym doświadczenia połowego” i oznacza symbolem p.

$$p = 0.674 \sqrt{\frac{\sum v^2}{n-1}}$$

gdzie n oznacza ilość spostrzeżeń, a v—stosunek procent. odchylen od średniej arytmetycznej.

*) Suma odchylen dodatnich wina się równać sumie odchylen ujemnych, czego uczkolwiek prof. dr. Górski nie uwzględnił. Prawdopodobnie jednak są to błędy zecerzkie.

W ten sposób:

v	v ²	v	v ²
- 16	156	+ 6	36
- 36	1296	+ 25	625
+ 15	225	- 7	49
- 14	196	- 8	64
+ 17	289	+ 0	0
+ 34	1156	+ 15	225
- 27	729	- 14	196
- 22	484	- 19	361
+ 15	225	+ 0	0
- 25	625	- 15	225
+ 29	841	- 7	49
+ 30	900	+ 0	0
- 25	625	+ 6	36
+ 8	64	+ 1	1
+ 5	25	- 17	289
+ 13	169	+ 5	25
- 31	961	+ 5	25
+ 1	1	+ 7	49

$$\Sigma v^2 = 11122, \text{ zaś } n = 36$$

$$p = 0.674 \sqrt{\frac{11122}{36-1}} = \pm 12\%$$

Innego zagadnienia dotknął *autor* niniejszej pracy.*) Badań statystyczne obszaru pod zasiewami wykazały, że uprawa roślin, wymagających dużego nakładu robót ręcznych, czyli t. zw. roślin pracopojemnościowych, jest zwiększona tam, gdzie jest większa gęstość zaludnienia. Średnia arytmetyczna procentu ziemi uprawnej dla ziemniaków była w latach 1910-12 w Królestwie Kongresowym 18,6. Jeżeli z 84 powiatów przedwojennej Kongresówki utworzyć 2 grupy, z których jedna będzie posiadała mniej powierzchni ziemniaków jak średnia arytmetyczna

*) Leon Waściszakowski. Ewolucja produkcji rolnej etc. Wyd-ctwo Biura Prac Kongresowych. Warszawa 1919 r.

dla całego Królestwa, a druga więcej, to w zestawieniu z gęstością zaludnienia ogólnego i rolniczego otrzymamy szeregi, będące między sobą w związku przyczynowym.

grupy powiatów	obszar ziemniaków	gęstość zaludnienia	
		ogólnego	rolniczego
40	< 18.6%	92.1	79.3
44	> 18.6%	130.3	91.7

Widzimy tutaj próbę ustalenia związku *między powierzchnią uprawy roślin pracopojemnościowych, a gęstością zaludnienia*.

Ubezpieczenia.

Ustalone liczby statystyczne posiadają szczególnie wielkie znaczenie nie tylko teoretyczne, ale i praktyczne w dziedzinie rokowań ubezpieczeniowych. Działalność towarzystw asekuracyjnych opiera się na poznanych przez dochodzenie statystyczne stałych stosunkach nieszczęśliwych wypadków. *Schwittau* *) przytacza cyfry, dotyczące nieszczęśliwych wypadków w Niemczech:

na 1000 ubezpieczonych przypada

	1898	1899	1900	1901	1902	1903	1904	1905	1906
wypadków śmierci	0,43	0,43	0,45	0,44	0,41	0,42	0,43	0,43	0,43
całkowitego uszkodz.	0,06	0,07	0,07	0,08	0,07	0,08	0,08	0,08	0,07
częściowego „	2,62	2,75	2,70	2,88	2,90	2,99	3,15	3,14	2,93

Przykład powyższy daje wymowne świadectwo t. zw. stałości cyfr statystycznych, na której w pewnej mierze opiera się współczesny mechanizm życia społecznego. Nic dziwnego, że w poszukiwaniu odkrycia statystycznych szeregów tego rodzaju poświęcali swe siły najlepsi statystycy, poczynając od XVII stulecia. Słusznie zauważa *A. Czuprow*, że zagadnienie tych liczb jest osią, dookoła której obraca się myśl statystyczna.

*) *G. Schwittau. Wwedenje w ekonomiczeskiju statistiku*, str. 88. Petersburg 1910.

Zagadnienia
moralne.

Od czasów Quetelet'a statystyka zagadnień moralnych rozwinęła się w obszerny i interesujący dział nauki. Wchodzą tu kwestje religji, pożycia małżeńskiego, rozwodów, nielegalnych urodzeń, prostytucji, alkoholizmu, samobójstw, przestępstw etc. Statystyka stara się nie tylko dać wierny obraz tego co jest, co się składa na życiowe zjawiska społeczeństw, ale dąży do uchylenia rąbka tajemnicy, dlaczego tak jest, jakie są przyczyny zjawisk? *Schnapper - Arndt**) przytacza zestawienie szeregów *kradzieży* i cen na zboże na podstawie statystyki niemieckiej za okres od 1882 do 1898 roku (szeregi zbieżne):

Lata	Cena na zboże	Kradzieże
1882	152.3	250
1883	144.7	241
1884	143.3	231
1885	140.0	214
1886	130.6	210
1887	120.9	198
1888	134.5	194
1889	155.5	211
1890	170.0	206
1891	211.2	216
1892	176.3	236
1893	133.7	202
1894	117.8	198
1895	119.8	192
1896	118.8	184
1897	113.0	189
1898	146.3	191

Tamże u Schapper - Arndta**) znajdujemy statystykę motywów takiego ponurego zjawiska, jakim jest w społeczeństwie *samobójstwo*. Według statystyki państwa japońskiego na 1000 wypadków samobójstwa w Tokio powody samobójstw dzieliły się tak:

*) Dr. Schnapper-Arndt. *Socialstatistik*. Leipzig 1908. Str. 618.

**) *Ibid*, str. 595.

	u mężczyzn	u kobiet	razem
choroby duchowe	497	518	505
„revers de fortune et misère”	158	83	128
choroby	131	138	134
miłość	37	85	56
wyrzuty sumienia i wstyd	43	28	37
niesnaski rodzinne	19	49	31
troska o przyszłość	8	10	9
straty majątkowe i długi	17	4	12
inne przyczyny	37	48	41
nieznane	53	37	47
	1000	1000	1000

I w tej dziedzinie współczynniki statystyczne są stałe, jak widać z innego jeszcze przykładu u tegoż autora *):

na 100.000 mieszkańców liczba samobójstw
w Niemczech (1894—1904)

1894	—	21,7
1896	—	20,6
1898	—	19,9
1900	—	20,3
1902	—	21,4
1904	—	21,0

Ekonomja.

Wielkie zastosowanie statystyka znajduje w ekonomji. Śmiało można powiedzieć, że niema zagadnienia ekonomicznego, któreby mogło być bądź oświetlone, bądź rozwiązane bez ujęcia liczbowego. Statystyka jest podstawą dla badań ekonomicznych. Spisy gospodarcze dostarczają obszernego i źródłowego materiału statystycznego, który jedynie miarodajnie może zobrazować stan ekonomiczny kraju. Zatrzymamy się tylko na kilku potocznych przykładach.

*) Ibid., str. 580.

Określenie wielkości *rozwoju ekonomicznego* dokonywamy za pomocą cyfr indeksowych. Porównanie stanu ekonomicznego z okresów chronologicznych otrzymuje się w ten sposób, że jeden okres, przyjęty za podstawę porównania, oznaczamy przez 100, drugi wyrażamy w pewnej cyfrze stosunkowej. Naprz.

Rozwój transportu wszechświatowego *).

	r. 1926	1913 = 100
długość linii kolejowych	1.220.300 klm.	111
ilość samochodów	27.500.000	1400
przewóz okrętowy w tonnach w r. 1927	65.200.000	133

Rozwój zużycia nawozów sztucznych *).

	1926 r.	1913=100
Azot (N)	1,136.007 tonn	173
potas (K ₂ O)	1.515.500 „	136
Superfosfat	12.741.000 „	113

Rozwój hodowli *).

	r. 1926	1913 = 100
koni	106.300.000	95
bydła rog.	579.400.000	109
świń	232.500.000	89
owiec	596.600.000	98

Zbiory zbóż *).

	r. 1926	1913=100
pszenicy	130.000.000 tonn	106
żyta	43.400.000 „	96

Drugi szereg tworzą cyfry indeksowe, wskazujące na progress, wzgl. regress zjawisk ekonomicznych.

Spisy gospodarcze, w szczególności rolnicze, rzucają światło na kwestję autarcji, czyli *samowystarczalności* państwa z zakre-

*) Les forces économiques du monde. 1927, str. 9.

su produkcji rolnej. Niestłuchane znaczenie tego zagadnienia aż nadto dobitnie podkreśliła wojna ubiegła. W latach 1909—1913 spożycie na głowę ludności wynosiło w Rosji Europejskiej: żyta — 9.1 puda, pszenicy—6.0 pud., owsa—5.6 pud., jęczmienia — 2.3 pud., razem 23.0 pudy. Jeżeli wielkość spożycia na głowę ludności zestawić z wysokością plonów z desiatiny, to otrzymamy wielkość obszaru pod zasiewami na głowę ludności niezbędną w celu zaspokojenia potrzeb wyżywienia. Ponieważ według danych Centr. Kom. Statystycznego wysokość plonów żyta wynosiła 53 pudy z 1 desiatiny, pszenicy — 49 pud., jęczmienia — 59 pud. i owsa — 55 pud., to normalny obszar pod zasiewami, niezbędny do zaspokojenia potrzeb ludności na 1 głowę, wyniesie w desiatinach:

żyta	pszenicy	jęczmienia	owsa	łącznie
0.17	0.12	0.04	0.10	0.43

Według zaś spisu rolniczego z r. 1916 *) zarejestrowano obszar pod zasiewami na głowę ludności:

0.20	0.15	0.04	0.12	0.55
------	------	------	------	------

Spis rolniczy dowiódł tym sposobem, uwzględniając plony przeciętne, że obszar pod zasiewami w r. 1916 przewyższał mniej więcej o 20—25% wielkość obszaru, niezbędnego dla zaspokojenia potrzeb wyżywienia ludności. Możliwość eksportu zbóż dla Rosji przedwojennej była oczywista. I rzeczywiście Rosja praktykowała wywóz zagranicę w dość znacznych rozmiarach. W pięcioleciu 1904 — 1908 wywieziono 464.153 tys. pud. żyta, pszenicy, jęczmienia i owsa; w pięcioleciu, poprzedzającym wojnę bezpośrednio, t. j. w latach 1909 — 1913, wywóz wymienionych zbóż stanowił 592.192.000 pudów rocznie.**)

Pracę syntetyczną, niezmiernie doniosłego znaczenia, dotyczącą całokształtu naszego życia gospodarczego, t. j. górnictwa, przemysłu i rolnictwa, z punktu widzenia *lokalizacji* form

*) Предварительные итоги всероссийской сельскохозяйственной переписи 1916 года, стр. XXX.

**) Материалы к продовольственному плану. Производство, потребление и потребление хлеба в России, ст. VII. r. 1916.

produkcji na terenie Rzeczypospolitej — wykonali Jan *Piekalkiewicz* i Stanisław *Rutkowski* p. t. „Określi gospodarcze Polski”, opartą na materiałach statystycznych*). W wyniku tej pracy autorowie otrzymali w formie syntezy 22 okręgi gospodarcze czysto rolnicze. Podobne prace, co dopiero zapoczątkowane w Polsce, jako czysto konstrukcyjne, są całkowicie uzależnione od wiarygodności i dokładności materiału statystycznego, przy zastosowaniu odpowiedniej metody opracowania.

Majątek narodowy oblicza Dr. Bohdan *Dederko***). Posiłkując się liczbami statystycznymi ilości: ziemi bez gruntów leśnych, lasów i gruntów leśnych, inwentarzy żywych, martwych, budynków, przedsiębiorstw przemysłowych i górnictwa, kolei żelaznych, dróg bitych, samochodów, ruchomości mieszkaniowych, kruszców i t. d. i mnożąc te ilości przez odpowiednie współczynniki wartości pieniężnej, otrzymuje jako końcowy rezultat 137.463.100.000 złotych majątku narodowego w Polsce. Na 1 mieszkańca czyni to 517 dolarów. Polska wśród narodów świata zajmuje pod tym względem 14-te miejsce.

Zagadnienie wartości pieniężnej *rocznej produkcji surowców* podaje Hipolit *Gliwic****) w następującym zestawieniu. W porządku wartości w miliardach franków złotych rocznej produkcji światowej kolejno idą z przeniłą na czele:

pszenica	24.0
ziemniaki	23.0
kukurydza	18.7
ryż	17.5
węgiel kamienny	16.1
drzewo użytkowe	15.0
cukier	10.8

*) Jan Piekalkiewicz i Stanisław Zdzisław Rutkowski. Określi gospodarcze Polski. Kwartalnik Statystyczny. Rok 1927. Tom IV. Zeszyt 3.

**) Zbiór prac ekonomiczno-społecznych. Prace zakładu polityki ekonomicznej S. G. G. W. pod kierunkiem prof. Władysława Grabskiego. Warszawa, 1930, str. 140 i 141.

***) Hipolit Gliwic. Podstawy ekonomiki światowej. Tom I. Sprawa surowców. Warszawa, 1926, str. 34.

owies	9.1
bawełna	8.6
surowiec żelaza	7.8
żyto	5.8
ropa naftowa	4.2
mięso	3.1
tytoń	2.8 i t. d.

Alfred *Maschall*, *) profesor Uniwersytetu w Cambridge, w swych „Zasadach ekonomiki” bardzo chętnie powołuje się na dowód statystyczny. Naprz., w zagadnieniu *budżetów rodzinnych* przytacza tablicę Dr. Engela, dotyczącą spożycia klasy niższej, średniej i robotniczej w Saksonji w r. 1857.

Rodzaj wydatków	WYDATKI CAŁEJ RODZINY		
	Robotnik z dochodem od 45—60 funtów sterl. rocznie	Robotnik z dochodem od 90—120 funtów sterl. rocznie	Osoba klasy średniej z doch. od 150—200f.st. rocznie
	%	%	%
1. Pożywienie	62	55	50
2. Ubranie	16	18	18
3. Mieszkanie	12	12	12
4. Światło i opał	5	5	5
5. Kształcenie	2	3.5	5.5
6. Pomoc prawna	1	2	3
7. Pomoc lekarska	1	2	3
8. Rozrywki i przyjemności	1	2.5	3.5

Z tej tablicy wynika, że w klasach zamożniejszych większy % dochodu rocznego idzie na rozrywki, na leczenie, na kształcenie, natomiast stosunkowo mniejszy — na pożywienie. Poznanie struktury budżetów rodzinnych jest niezmiernie ważnym czynnikiem w polityce płac, społecznej etc.

*) A. Marschall. *Zasady ekonomiki*. Z oryginału angielskiego przełożył prof. Dr. Czesław Znamierowski. Tom I. Warszawa, 1925, str. 112.

Zagadnienia *handlu zagranicznego* ilustruje następująca tablicą statystyczną *Kwiatkowski* *).

I — Przywóz do Polski w ‰-ach.

	1924	1925	1926	I półr. 1927
z państw sąsiadujących z Polską	42.4	39.7	30.7	38.8
z innych państw europejskich	39.0	39.3	42.7	39.6
z państw po za europejskich	18.6	21.0	26.6	21.6

II — Wywóz z Polski w ‰-ach.

	1924	1925	1926	I półr. 1927
do państw sąsiadujących z Polską	60.8	63.1	42.4	46.7
do innych państw europejskich	36.8	34.6	55.3	49.7
do państw po za europejskich	2.4	2.3	2.3	3.6

Dodać należy, że największy handel eksportowy i importowy Polska prowadzi z Niemcami **).

Emigracja.

Statystyka emigracyjna należy do najtrudniejszych działów tej nauki ze względu na specyficznie trudne warunki uzyskania materiału statystycznego. Tylko poważna analiza naukowa może w tej dziedzinie zapewnić wiarogodne cyfrowe dane. A jednak zagadnienie emigracji jest tak doniosłe, że państwa zmuszone są robić w tym kierunku rozliczne dociekania.

Na 100 emigrantów z Polski wypadało w roku ***)

	1901	1902	1903	1904	1905
Wychodźców z Rosji	49.18	44.98	45.70	45.83	47.41
„ z Niemiec	4.10	4.66	6.30	6.95	4.02
„ z Austrii	46.72	50.36	48.00	47.32	48.57

*) Eugenjusz Kwiatkowski, *Postęp gospodarczy Polski*. Wyd. II. Warszawa 1928.

**) patrz: *Rynki zbytu*. Wyd-ctwo Państw. Instyt. eksportowego, Warszawa 1928.

***) Dr. Kazimierz Władysław Kumaniecki. *Studja z zakresu statystyki wędrowek*. Kraków, 1909, str. 75.

Widzimy, że przed wojną prawie cała emigracja polska obejmowała zabór rosyjski i austrijski, rozdzielając się na oba niemal równomiernie.

Emigrujący Polacy w Stanach Zjedn. A. P. według zawodów *)

W roku	Posiadało wyższe wykształcenie	czyli na 100 osób	Rzemieślników było	na 100 osób	Robotników rolnych, fabrycznych i wyrobników	na 100 osób	Bez zajęcia (w tym kobiety - dzieci do 14 lat)	na 100 osób	Ogółem
1901	39	0.09	1797	4.12	26.285	60.26	15.496	35.53	43.617
1902	41	0.06	2971	4.27	52.198	74.98	14.410	20.69	69.620
1903	50	0.06	3715	4.51	62.646	76.08	15.932	19.35	82.346
1904	136	0.20	3811	4.62	48.981	72.29	14.829	21.89	67.757
1905	160	0.16	5056	4.92	78.889	77.01	18.332	17.89	102.437

Widać z powyższego, że emigracja polska do Ameryki Północnej była przeważnie wychodźstwem ludności wiejskiej, szukającej w Stanach Zjednoczonych zarobku przedewszystkiem w charakterze wyrobników i posługaczy, częściowo przy robotach ziemnych, rolnych lub przemysłowych.

Pod względem oświaty emigracja polska i ruska do Stanów Zjedn. przedstawiała następujący obraz.**)

Na 100 wychodźców w wieku ponad 14 lat życia

W roku	<i>polskich</i>		<i>ruskich</i>	
	tylko czyta	nie czyta i nie pisze	tylko czyta	nie czyta i nie pisze
1901	2.38	35.14	0.89	52.31
1902	1.86	36.57	0.67	49.17
1903	1.76	30.35	0.42	49.01
1904	2.49	33.29	0.89	57.98
1905	3.80	35.93	0.01	61.63

Uderzająca jest różnica między ludnością polską i ruską. Pierwsza wysyłała na emigrację żywioły przeważnie oświecone,

*) Ibid., str. 81.

**) Ibid., str. 82.

bo około $\frac{2}{3}$ ogółu emigracji polskiej składało się z osób, umiających czytać i pisać, kiedy wśród ruskiej — połowa nie umiała czytać i pisać, a niekiedy więcej, bo w r. 1905 — 61.63%.

Dr. *Caro* podaje, iż Galicja przed wojną w ciągu 40 lat straciła przez emigrację 858.579 ludzi, co w stosunku do ogółu ludności czyni 10%. Ponieważ straty przez emigrację dzieliły się między Polaków i Rusinów w stosunku 70.63% i 29.37%, dr *Caro* konkluduje, że teoretycznie 400 lat podobnej emigracji wystarczałoby, ażeby Galicję zamienić na kraj ruski *).

Medycyna.

Dr. *Ciągliński* **) zastosował metodę statystyczną do badań klinicznych, zaś dr. *Maciejewski* ***) — do rozstrzygnięcia zagadnienia korzyści użycia t. zw. obrotu profilaktycznego przy porodach. Interesujące dociekania dr. Maciejewskiego jednocześnie dowodzą, że metoda *różnicy* jest nie tylko możliwa w badaniu statystycznym, lecz konieczna — przy zachowaniu pewnych warunków, o których mówimy niżej, powołując się na doświadczenia dr. Maciejewskiego.

Na tych kilku przykładach zastosowania metody statystycznej w różnych gałęziach wiedzy — porzeczajemy.

*) Dr. Leopold Caro. *Emigracja i polityka emigracyjna* r. 1914, str 28.

**) Dr. A. Ciałliński. *Zasady statystyki szpitalnej*. Mies. Gl. Urz. Stat. Warszawa, 1920.

***) Dr. med. Kazimierz Maciejewski. *Podręcznik statystyki, Teoria statystyki*. Warszawa, 1925.

III. Zbieranie materiału statystycznego.

Jedną z najważniejszych czynności statystycznych jest zbieranie materiału statystycznego. Żadne późniejsze, najbardziej precyzyjne opracowanie statystyczne nie jest w stanie naprawić błędów pierwiastkowego badania. O ile przy zbieraniu materiału statystycznego zakradł się błąd zasadniczy, rezultaty badania będą mało wiarogodne i prowadzące do fałszywych wniosków. Dlatego, przystępując do zbierania materiału statystycznego, należy dbać, aby uzyskany materiał był, o ile to jest możliwe w danych warunkach, najwiarogodniejszy, najprawdopodobniej zebrany.

Zbieranie materiału statystycznego opiera się na spostrzeganiu, czyli obserwacji statystycznej. Statystyk ma zazwyczaj do czynienia: albo z materiałem już zebrany i dostarczony mu w formie surowej publikacji lub notatek, albo sam przystępuje do zbierania materiału. W tym wypadku winien on materiał sam zebrać, technicznie opracować, pogrupować i opracować naukowo. Dla statystyka jest ciekawszy ten drugi wypadek i praca jego w tym razie jest cenniejsza, ale bardzo często są zdarza, że statystyk rozporządza już zebrany materiałem, który ma opracować. W obu wypadkach winien się kierować pewnymi regułami, wypracowanymi przez praktykę i teorię statystyczną. W żadnym wypadku jednak statystyk nie może lekceważyć materiału statystycznego. O ileby raczej można było w pewnej mierze zgodzić się na pewną nieścisłość opracowania (naprz., w t. zw. obliczeniach przedwstępnych), o tyle nie można wyobrazić sobie pracy wartościowej, opartej

na wadliwie uzyskanym materiale statystycznym. Wszelkie, daleko idące, metody matematycznego opracowania tylko wówczas mogą posiadać wartość istotną, kiedy spostrzeganie statystyczne były przeprowadzone bez zarzutu, z zachowaniem wszystkich możliwych ostrożności, zabezpieczających wiarogodność materiału statystycznego.

Spostrzeganie.

Spostrzeganie statystyczne może obejmować wszystkie bez wyjątku zjawiska danej zbiorowości, naprz. wszystkich mieszkańców kraju przy spisie ludności, albo tylko pewną część danej zbiorowości, naprz., plony zbóż z części przestrzeni. W pierwszym wypadku spostrzeganie nazywamy badaniem wyczerpującem, w drugim — badaniem częściowym, niecałkowitem. Zazwyczaj wyczerpujące badania, o ile dotyczą one wielkich zbiorowości, jak ludność kraju, stan gospodarczy etc., przeprowadzane są przez państwo przy użyciu znacznych środków materialnych; badania zbliżone do wyczerpujących, a więc częściowe, niecałkowite — mogą być przedsięwzięte przez poszczególnych badaczy. Do badań wyczerpujących zaliczamy przedewszystkiem spisy, niekiedy nazywane popisami (spisy albo popisy ludności, koni, gospodarcze etc.) oraz rejestrację bieżącą (zapisy urodzeń, zgonów, notowanie tych lub innych faktów etc.), do badań niecałkowitych — orientację liczbową, szacowanie, ankietę, monografię, badanie reprezentacyjne (dobieranie). Zaczniemy od metod badania niecałkowitego.

Orientacja liczbowa.

Orientacja liczbowa — według terminologii Mayr'a — albo *notowanie* ma miejsce wtedy, kiedy poprzestajemy na pewnej dowolnej liczbie spostrzeżeń, nie określając całkowitej ich ilości. Ażeby, naprz. wiedzieć, jakie są na rynku ceny towarów, nie potrzeba brać wszystkich wypadków dokonanych transakcyj, ale wystarcza notowanie pewnej ilości tych transakcyj. Tak samo postępujemy z notowaniem płacy zarobkowej, plonów etc. Niema potrzeby znać wszystkich wypadków cen kupna — sprzedaży,

dlatego że, wychodząc z założenia o istnieniu prawidłowości w życiu społecznym, wnosimy, że cena danego towaru, ustalona w kilkudziesięciu wypadkach, nie odbiega zasadniczo od cen spotykanych powszechnie lub tylko najczęściej. Nie potrzeba również ustalać wysokości plonu z każdego poszczególnego ha, gdyż opierając się na równości warunków naturalnych danej okolicy przy uwzględnieniu stopnia jej rozwoju ekonomicznego, można wnosić o plonach w okolicy z próbnych przestrzeni (kilku lub kilkudziesięciu ha).

Wszelako wartość orientacji liczbowej jest wyższa, gdy podajemy liczbę wypadków, z których została oznaczona jej wielkość, oraz granice wahania czynników składowych (cen, płac, plonów i t. p.).

..... Szacowanie albo *oznaczenie przybliżone* często jest konieczne ze względów praktycznych, gdy bezpośrednie badanie statystyczne jest niemożliwe. Naprz., na zasadzie wiadomego % przyrostu ludności obliczamy, ile ludności liczył kraj przed 50 laty. Przyjmujemy przytem, że proc. przyrostu przez cały ten czas był jednakowy. Lub też na zasadzie podziału ludności według wieku lub innej jakiej cechy w jednym kraju, określamy przypuszczalnie taki sam podział ludności w innym kraju. O ile przy szacowaniu ma miejsce uwzględnienie warunków, wywierających wpływ na zmianę zjawisk, mamy do czynienia z *interpolacją*.

..... Ankieta (enquête) jest to taki rodzaj badania statystycznego, który sprowadza się do odpowiedzi na zapytania. Zapytujemy w formie ustnej lub pisemnej osoby zainteresowane lub kompetentne o wiadomości, dotyczące badanego przez nas zjawiska, i zapisujemy odpowiedzi. Zbiór tych odpowiedzi tworzy ankietę. Jest to badanie statystyczne bardzo rozpowszechnione ze względu na dostępność przeprowadzenia. Naprzód układamy kwestionariusz, zawierający pytania, a potem rozsyłamy go z prośbą o wypełnienie.

W badaniu ankietowym spotykamy się przedewszystkiem

z 2-ma wypadkami. Albo zapytujemy tylko osoby kompetentne, ekspertów, co myślą, jaka jest ich opinia o interesującym nas fakcie, który badamy; będą to wiadomości tylko od grona określonych osób; jako jedynie pewne czy prawdopodobne; albo pytania kierujemy do dużej ilości osób zainteresowanych, chcąc zebrać możliwie najwięcej materiału bez względu na to, czy materiał ten będzie posiadał taką samą wartość jak w pierwszym wypadku. Stosownie do powyższego odróżniamy ankiety, zawierające sądy kompetentnych osób, co ma miejsce w praktyce komisji parlamentarnych, przedewszystkiem angielskich, i dlatego nazywamy takie ankiety — ankietami typu angielskiego; i dalej — ankiety, które mają na celu nie zbieranie wyłącznie opinii, lecz faktycznych danych. Te ostatnie nazywamy ankietami amerykańskiego typu. Charakterystyczna ich cecha polega na masowości danych. Praktyka wykazuje, że odpowiedzi na ankietę otrzymuje się tylko częściowo, pewien procent; liczba ich jest wypadkowa. Najczęściej zaledwie połowa, niekiedy 60 proc. i więcej, rozesłanych kwestjonariuszów powraca do wysyłającego po wypełnieniu. Często jednak zdarza się, że liczba odpowiedzi jest niedostateczna. Tak w r. 1895 jedno z amerykańskich biur pracy rozesłało 15000 kwestjonariuszów, otrzymało odpowiedzi 1448, co czyni około 10%. Nieco więcej uzyskano odpowiedzi na ankietę w sprawie stosunków robotniczo - rolnych w Królestwie Kongresowem, opracowaną w następstwie przez Wł. Grabskiego, bo 14%. Natomiast niekiedy rezultat ankiety bywa dobry. Naprz., materiały, ilustrujące strukturę społeczną stanu zawodowego (opublikowane w Roczniku pracowniczych związków zawodowych w Polsce w r. 1926) zostały zebrane staraniem Ministerstwa Pracy i Opieki Społecznej również w drodze ankiety, na którą nadesłano 73.5% odpowiedzi, opracowano zaś 63.2% odpowiedzi. Widzimy z powyższego, że ankietą mimo jej rozpowszechnienia jako środka dochodzenia statystycznego jest często dosyć zawodna, a przytem nie przestaje być tylko częściowem, niecałkowitem badaniem statystycznym i z tego względu nie jest w stanie zastąpić badania wyczerpującego.

Metoda monograficzna.

Do poprzedniej grupy metod częściowego badania statystycznego należy również metoda monograficzna, należąca w gruncie rzeczy do najstarszych metod w dziedzinie badań społecznych wogóle. Zyskuje ona na rozpowszechnieniu i uznaniu dopiero jednakże w XIX stuleciu, kiedy do badań statystycznych zostaje wprowadzona przez Le-Play, skąd niekiedy nazywamy ją metodą Le-Play'a. W r. 1855 Le-Play wydaje doskonałą pracę p. t. „Les ouvriers européens”, w której znajdują się opisy monograficzne życia niższych klas społecznych. Specjalne uwzględnienie znalazły tu badania nad budżetami rodzinnymi. W Niemczech metodą tą w zastosowaniu do robotniczej ludności i jej budżetów domowych pracował Engel.

Metoda monograficzna polega na tem, że zamiast całkowitej masy zjawisk statystyk obserwuje zjawiska jedynie *typowe*, t. j. takie, które uogólniają w mniej lub więcej doskonałym stopniu przeciętne cechy danej zbiorowości. Poza kołem spostrzegania pozostają wszystkie te zjawiska, które posiadają jakieś cechy odrębne, rzadziej spotykane, czyli nietypowe dla danego środowiska. Metoda ta czyni możliwem wniknąć bardzo szczegółowo w istotę badanego zjawiska przez możliwie wszechstronne oświetlenie za pomocą odpowiednich faktów, podlegających dochodzeniu statystycznemu. Badanie monograficzne porusza tyle kwestyj, ile w żadnym wypadku badanie masowe poruszyć nie jest w stanie. Jeżeli przytem analiza naukowa dotyczy nie jednego wyłącznie obiektu typowego (ściśła monografia), lecz pewnej znacznej liczby obiektów typowych (np. kilkudziesięciu lub kilkuset budżetów rodzin, gospodarstw rolnych i t. p.), to badanie takie ma wszelkie cechy statystycznego. Od metody monograficznej w statystyce należy odróżniać monografię wogóle, jako środek poznania w innych naukach, bowiem w tym wypadku mamy do czynienia z pojedynczym przedmiotem, mającym samodzielne i niezależne znaczenie.

Trudności metody monograficznej polegają na tem, że wydawanie sądu, czy dane zjawisko jest typowe w danej zbior-

rowości, czy nie—jest bardzo subiektywne. Tego rodzaju orzeczenie kwalifikuje à priori zjawiska bez dostatecznej ich znajomości, która wszak powstaje dopiero po ukończeniu i opracowaniu badania. Sądy zaś, oparte tylko na ogólnych wiadomościach o danej zbiorowości, mogą być niepewne. Tem nie mniej jednak żadne inne częściowe badanie statystyczne nie da nam tak wszechstronnego materiału obserwacyjnego, jak monograficzne.

Metoda dobierania,
czyli Kiaera.

Na granicy między częściowym badaniem statystycznym, a wyczerpującym — znajduje się metoda badania wprowadzana do nauki około r. 1890 przez norweskiego statystyka *Kiaera*, która posiada cechy ściśle statystyczne, aczkolwiek ogranicza się tylko częściowym, z góry jednak określonym, materiałem statystycznym. Wzbudza ona szczególne zainteresowanie wśród statystyków, gdyż jej znaczenie naukowe i zastosowanie przy odpowiednim udoskonaleniu teoretycznym i praktycznym winno być ogromne.

Metoda dobierania polega według Kiaera na tem, że z całej zbiorowości zjawisk wybieramy tylko niektóre zjawiska, odpowiadające pewnym warunkom. Liczba tych zjawisk winna być dostatecznie wielka, ale może stanowić tylko pewną określoną część całości, np. $\frac{1}{5}$, $\frac{1}{10}$, $\frac{1}{20}$ i t. d. Charakter dobieranych zjawisk ma być taki, ażeby łącznie stanowiły one podobieństwo całości. Winny one tworzyć niejako minjaturę całości. Jest to więc metoda fragmentów, dobieranych pod specjalnym kątem widzenia. Wszystkie składowe części tego badania, a więc zarówno zjawiska typowe, jak i krańcowe, winny być w takim ustosunkowaniu w granicach przeprowadzonego badania, w jakim znajdują się w rzeczywistości. Jeżeli w jakiejś zbiorowości zjawiska typowe stanowią $\frac{8}{4}$ wszystkich jednostek obserwacji, a krańcowe — $\frac{1}{4}$, to stosując metodę Kiaera i biorąc naprz., 8% zjawisk z całej zbiorowości, obliczamy, ażeby w wybranym fragmencie typowe zjawiska stanowiły 6% całej sumy zjawisk, a nietypowe 2%, czyli znajdowały się

we wzajemnem ustosunkowaniu jak $\frac{8}{4}$ do $\frac{1}{4}$. Metoda Kiaera ma na celu przeprowadzenie dochodzenia statystycznego nad zbiorowością w drodze obserwacji nad specjalnie dobraną jej częścią, dostatecznie wielką i mieszczącą w sobie proporcjonalnie wszystkie cechy, charakterystyczne dla całej zbiorowości.

W r. 1895 przeprowadzono badania w Norwegji nad 127 gminami wiejskimi i 23 miastami, t. j. badaniem objęto, zamiast całości, 3.3% wszystkich dorosłych mężczyzn mieszkańców miast i 1.6% całej męskiej ludności okręgów wiejskich, razem około 10.800 osób. W roku 1900 poddano badaniu również w Norwegji 5% wszystkich gospodarstw wiejskich i 10% prywatnych gospodarstw w miastach. Profesor *Czuprow*, o czym niżej, zastosował tę metodę do przeprowadzonych badań w gub. Moskiewskiej w 500 wsiach (na ogólną liczbę 5220 wsi), t. j. w granicach 10.4%.

Widzimy więc, że przy metodzie Kiaera poddajemy badaniu określoną część masy statystycznej.

Z powyższego określenia omawianej metody wynika, że opracowanie planu badania, t. j. ustalenie jaka część zbiorowości ma podlegać badaniu, jakie składowe czynniki mają się złożyć na minjatiurę całości — jest jedną z najważniejszych czynności statystyka. Od tego planu zależy powodzenie całego dochodzenia statystycznego.

Jest zupełnie zrozumiałe, że w kraju dzikim pod względem statystycznym, t. j. takim, w którym uprzednio nie było żadnych badań statystycznych, metoda Kiaera nie może być zastosowana. Warunek dobierania poszczególnych zjawisk ze zbiorowości ogólnej w celu skompletowania zmniejszonego obrazu całości wymaga nieodzownie znajomości przedmiotu badania, która tworzy się na podstawie przedewszystkiem wyczerpującego badania. Dla udatnego stosowania metody Kiaera muszą być w rozporządzeniu w pierwszym rzędzie materiały statystyczne, pochodzące ze spisów, bądź z innych źródeł dochodzenia statystycznego.

Określenie ilości zjawisk dobieranych z danej zbiorowości nie powinno być dokonywane na ślepo. Nie chodzi tu o me-

chaniczne zmniejszenie liczby spostrzeżeń, lecz o taką ich ilość, któraby była dostateczna dla wiernego odtworzenia stosunków, panujących w całości, a przytem stanowiła minimum. Powiększanie liczby spostrzeżeń ponad minimum nie jest potrzebne, a to ze względu na to, że t. zw. błąd prawdopodobny średniej zmniejsza się w stosunku odwrotnie proporcjonalnym do pierwiastka kwadratowego z liczby dokonanych spostrzeżeń. Wiarygodność wyników badania, o ile osiągnięto wspomniane minimum spostrzeżeń, niezbędne dla dostatecznego stopnia reprezentowania całości, tylko w nieznacznym stopniu zależy od zwiększenia ilości spostrzeżeń. Przy zwiększeniu ilości obserwacji 4, 9, 25, 100 razy więcej, wiarygodność rezultatów wzrasta w stopniu $\sqrt{4}$, $\sqrt{9}$, $\sqrt{25}$, i $\sqrt{100}$, t. j. o 2, 3, 5 i 10 razy. Możemy zatem poprzestać na osiągniętej wiarygodności przy zastosowaniu minimum spostrzeżeń zgodnie z duchem teorii Kiaera, samo zaś oznaczenie minimum najlepiej dokonać na gruncie rachunku prawdopodobieństwa.

Dobieranie spostrzeżeń nie jest łatwe i wymaga zachowania tego warunku, ażeby jednostki spostrzegania były reprezentacyjnymi jednostkami dla odtworzenia obrazu całości oraz ażeby weszły do badania w takim ustosunkowaniu, w jakim znajdują się poszczególne zjawiska w rzeczywistości. Tem trudniejsze będzie dobieranie, o ile masę statystyczną stanowią będą zjawiska bardziej różnorodne, względnie wyróżniające się cechami o silnej skali wahan. Dlatego prawidłowiej jest stosować metodę Kiaera w tych wypadkach, kiedy materiał obserwacyjny jest bardziej jednolity, gdyż wówczas poprzestać można na mniejszej liczbie spostrzeżeń i wyniki badania będą wiarygodniejsze. Jeżeli poszczególne zjawiska masy ujawniają duże odchylenia od przeciętnego typu, to liczba spostrzeżeń musi być znacznie większa i dobór zjawisk bardzo staranny. W tych wypadkach najlepiej całą masę podzielić na kilka jednolitych części i w granicach każdej części przeprowadzić dobieranie zjawisk.

Poniżej umieszczony przykład wyjaśni, czy metoda Kiaera może dać dobre wyniki. Prof. Czuprow, rozporządzając mater-

jałem statystycznym z 5220 wsi gubernji Moskiewskiej, uzyskanym w drodze powszechnego spisu, dokonanego przez ziemstwo, losowaniem (a więc w danym wypadku na ślepo) wydzielił 500 wsi, w których podsumował wyniki badań, celem porównania z wynikami badań przeprowadzonych na drodze spisu dla całej masy statystycznej, t. j. 5220 wsi. Jak widzimy, prof. Czuprow nie zastosował nawet warunku reprezentacyjności zjawisk i ich należytego ustosunkowania, a wręcz użył mechanicznego dobierania. Oto rezultaty:

	dla 500 wsi wybranych przez losowanie	dla 5220 wsi według spisu ziemskiego
Liczba wsi, w których mężczyźni byli piśmienni	47.5%	47.6%
Liczba wsi, w których piśmienni mężczyźni sta- nowili mniej jak połowę wszystkich mężczyzn .	55%	53%
Liczba wsi, w których piśmienni mężczyźni sta- nowili mniej jak $\frac{1}{4}$ wszystkich mężczyzn .	4.2%	4.1%
Liczba wsi, w których piśmienni mężczyźni sta- nowili więcej jak $\frac{3}{4}$ wszystkich mężczyzn .	1.2%	1%

Spostrzegamy, iż daty statystyczne, podane przez prof. Czuprowa, są niemal identyczne z datami spisu, co mówi na korzyść metody dobierania. Metoda ta ma szczególne znaczenie w badaniach naukowych — tam, gdzie mniej zależy na cyfrach absolutnych, a więcej na względnych, t. j. na ujawnieniu ustosunkowania zjawisk, na stosunku między rzeczami, który interesuje nauki społeczne, a w szczególności statystykę.

Natomiast spór o to, czy metoda Kiaera z racji swej ekonomiczności ma wyższość nad metodą spisów powszechnych należy uważać za bezprzedmiotowy. Metoda Kiaera nie może

zastąpić metody badania wyczerpującego, która jedynie dostarcza pewnych wyników, o ile została zastosowana prawidłowo, jak metoda badania wyczerpującego nie może zastąpić metody Kiaera. Ta ostatnia dąży do otrzymywania dat statystycznych bardziej szczegółowych, o znaczeniu głównie porównawczem, naukowem, podczas gdy pierwsza daje zazwyczaj schemat stosunków masowy, uproszczony, przynosząc odpowiedzi wprawdzie na najgłówniejsze, ale nieliczne pytania. Metodę Kiaera nie należy traktować równorzędnie z metodą badania wyczerpującego, ale raczej jako znakomite, pod względem naukowym częstokroć niezastąpione, uzupełnienie.

Dodać do tego należy, że statystyka zamyka się w kole liczb, przytem liczb dostatecznie wielkich, a te mogą być uzyskane jedynie na podstawie masowego badania.

Warunki badania
wyczerpującego.

Przeprowadzenie dochodzenia statystycznego w formie wyczerpującej wymaga przedewszystkiem opracowania planu organizacji badania. Ażeby plan organizacyjny był należycie wykonany, należy odpowiedzieć na pytania: *co* badać? *kiedy* badać? *gdzie* badać? *jak* badać? Te cztery pytania należy zresztą mieć na względzie nie tylko przy badaniu wyczerpującem, ale wogóle przy każdym dochodzeniu statystycznym.

Pierwsze pytanie, dotyczące tego, co badać, pozornie nie zasługuje na większą uwagę, gdyż badacz, przystępujący do pracy, zdaje sobie z tego sprawę, o co mu chodzi, co chciałby wyświecić, jakie stosunki chciałby ustalić. Jednak tak nie jest. W zależności od zakresu i dziedziny badania, czy chodzi o ludność, gospodarstwo społeczne, higienę, przestępstwa i t. d. — badacz winien być gruntownie obeznany z przedmiotem badania, t. j. rozumieć istotę zagadnień odpowiednich gałęzi wiedzy, jak demografia, ekonomja społeczna, medycyna, prawo karne i t. p. Statystyk, ma przed sobą zadanie nie tylko badać to, co już było badane, ale odkrywać rzeczy nowe, nieznanne, związki przyczynowe dotychczas niespostrzegane, musi zatem

posiadać wiadomości z danej gałęzi wiedzy na poziomie ostatnich zdobyczy, rozumieć istotę nowych zagadnień i w zależności od tego ułożyć program badania. Słowem, statystyk winien posiadać gruntowną znajomość przedmiotu badania, a przede wszystkim najdokładniej poznać to, co przed nim w tej dziedzinie na drodze podobnych badań zrobiono, ażeby nie odkrywać nanowo Ameryki.

Do programu organizacji należy również, jak wspominaliśmy, odpowiedź, kiedy należy przeprowadzić badanie. Praktyka szeregu państw w zakresie naprz. spisów ludności, wykazała, że spisy ludności są najodpowiedniejsze w określonych porach roku — zależnie od mniejszego mechanicznego ruchu ludności, przyzwyczajenia i t. p., a takimi są, ogólnie biorąc, miesiące zimowe. Odpowiedź na pytanie trzecie — gdzie badać? — wymaga dokładnej znajomości terenu i warunków, na które natrafi dochodzenie statystyczne ze względu na mogący mieć wpływ układ polityczny, społeczny, gospodarczy i t. p.

Wreszcie odpowiedź na pytanie czwarte — jak badać — uzależniona jest od takich warunków, jak cel badania, szybkość przeprowadzenia, pożądana dokładność, rozporządzanie środkami materialnymi, siłami technicznymi i t. d.

Można i należy przytem posługiwać się starymi planami, wykorzystać stare programy — zwłaszcza gdy chodzi o materiał perjodyczny, wymagający porównania z datami uprzednich dochodzeń. Nie należy jednak zapominać, że stare plany, stare programy mogą być użyte w charakterze tylko pomocniczym, gdyż życie jest wiecznie czemś nowem, wiecznie młodem, zmieniającem formy i kształty już zastygłe, i dlatego badanie statystyczne wymaga zawsze nowych planów i nowych idei. Życie społeczne idzie nieprzerwanym burzliwym prądem naprzód, w odpowiedniku do tego winny kształtować się i nasze pojęcia o niem, co z kolei wymaga nowych form badania.

Zszeregowanie pytań, na które szukamy odpowiedzi, stwarza kwestionariusz badania, czyli arkusz spisowy, inaczej zwany kartą rejestracyjną. Sformułowanie pytań czyli ułożenie arkusza spisowego należy do najważniejszych czynności wstępnych do

chodzenia statystycznego. Już Quetelet, wielki organizator jednodniowych spisów ludności, zwrócił uwagę na wymagania, którym winien zadość uczynić prawidłowo ułożony arkusz spisowy.

Pierwsza reguła Queteleta polega na tem, aby w karcie rejestracyjnej umieszczać tylko niezbędne pytania i tylko takie, na które można otrzymać odpowiedź. Jeżeli chodzi naprz., o spis ludności, to liczba pytań wogóle nie może być wielka. Badanie wyczerpujące, jakim jest spis, ma na celu objęcie spostrzeganiem wszystkich przedmiotów danej kategorii, może więc uwzględnić tylko te cechy, przedmioty lub rysy charakterystyczne, które są wspólne wszystkim jednostkom badanym, natomiast winno się unikać rejestrowania cech, których nie można uogólnić, jako zbyt specjalnych, lub zbyt indywidualnych. Do karty rejestracyjnej należy włączać tylko te cechy badanych przedmiotów czy osobników, które najistotniej oddają treść zjawiska, a z tego względu są najniezbędniejsze. Wszelako nie należy zapominać o zdaniu prof. A. Kaufmana, który twierdzi, że w statystyce niema szczegółów, że wszystko jest jednakowo ważne i potrzebne.

Badania wyczerpujące z punktu widzenia pierwszej reguły Queteleta jest notowaniem powszechnem, głównem, usuwającym szereg mniej ważnych cech badanej zbiorowości, skupiającą uwagę w niewielkiem promieniu, a więc dającą kontury ogólne, małoźróźniczkowane. W tem się mieści jego różnica z badaniem ankietowem, monograficznym, fragmentarycznym (Kiaera) i t. p., które pozwalają badać zjawiska społeczne bardziej szczegółowo, indywidualnie.

Druga reguła dotyczy wymagania, aby arkusz spisowy nie zawierał pytań, które mogą wzbudzać obawy lub podejrzenia wśród ludności. Ludność kraju, w którym powszechne spisy są rzadkością, przy niskim stopniu ogólnej kultury zachowuje się zazwyczaj wobec spisu nieufnie. Pochodzi to stąd, że wielokrotnie w historii przedsiębrano dociekania statystyczne li tylko z powodu ułatwienia w ściąganiu podatków lub poboru wojskowego. W obawie przed ciężarami podatkowymi lub da-

ninami ludność, nie znająca tradycji spisów powszechnych, niechętnie składa zeznania, lub podaje je fałszywie, ukrywając prawdziwy stan rzeczy. Statystyk, zdając sobie sprawę z trudności podobnego rodzaju, winien umieszczać w karcie rejestracyjnej tylko te pytania, które nie wzbudzają wątpliwości. Jednakże nie zawsze się to udaje.*)

Trzecia reguła dotyczy warunku, aby pytania arkusza spisowego były dla wszystkich zrozumiałe. Nie powinno być zatem pytań o znaczeniu dowolnem, dwumyślnem lub nieokreślonem. Przyjęte jest obok kwestjonariusza na tym samym arkuszu spisowym umieszczać *instrukcję* pouczającą, jak należy rozumieć poszczególne punkty kwestjonariusza. Jeżeli względy techniczne nie pozwalają na umieszczenie instrukcji na tym samym arkuszu, to instrukcję należy oddzielnie wydrukować i dołączyć do materiałów spisowych, aby każdy komisarz spisowy miał ją pod ręką. Jeżeli kwestjonariusze skierowane są bezpośrednio do ludności, jak to się często zdarza w krajach o wysokiej kulturze, celem wypełnienia ich przez samą ludność, t. j. w wypadku samoliczenia się, należy schemat pytajnika ułożyć w ten sposób, aby ci, co odpowiadają, nie mieli za dużo pola do szerokiej interpretacji pytań, co zawsze prowadzi do nieporozumień i omyłek. Naprz., w sprawie elementarnego wykształcenia odpowiedni jest taki formularz:

czyta i pisze	tylko czyta	nie czyta nie pisze
1.	2.	3.

*) Nasz spis ludności z r. 1921, połączony z częściowym spiszem gospodarczym, zastrzegając wykorzystanie dat statystycznych dla celów ubocznych (podatkowych lub innych), pomimo to ludność celu spisu wskutek braku przyzwyczajenia nie rozumiała. W okręgu spisowym lubelskim miał miejsce następujący fakt. Komisarz spisowy przybył do domu podmiejskiego włościanka, w momencie, kiedy w czasie obiadu podawano na stół gotowaną kurę. Komisarza zaproszono, aby przekąsił, ale dodał: „niech pan już tej kury nie liczy”. Przypuszczano, bowiem, że spis miał na celu nałożenie podatku od posiadania.

Przekreślenie cyfry na tym formularzu daje nam odpowiedź, nie budzącą żadnych wątpliwości, czy odpowiadający rozumiał znaczenie pytania. Gdybyśmy umieścili tylko dwie rubryki: „piśmienny i „niepiśmienny”, takie wątpliwości mogłyby mieć miejsce.

Wreszcie czwarta reguła Quetelet'a stawia wymaganie, ażeby jedno pytanie arkusza spisowego kontrolowały drugie, co nazywamy *kontrolą wewnętrzną*. Naprz., liczba pługów w gospodarstwie rolnem może kontrolować liczbę hektarów ziemi uprawnej, cyfra wieku przesądza nieraz kwestję stanu rodzinnego (dziecko nie może być, naprz., wdowcem) i t. p.

Instrukcja, umieszczona bezpośrednio na arkuszu spisowym, lub załączona jest niezbędna. Winna ona wyjaśnić, jak należy pojmować pytania zawarte w karcie rejestracyjnej. Należy w niej szczegółowo podać, naprz., jak rozumieć stopień pokrewieństwa, stan rodzinny, wykształcenie, religję, przynależność państwową, język ojczysty, wiek, zawód. Rubryka „zawód” wymaga szczególnie należytego oświetlenia, ponieważ zawodów jest bardzo wiele (spis niemiecki z r. 1895 wykazał przeszło 10.000 zawodów), a prócz tego obok zajęcia głównego istnieją zajęcia poboczne, dodatkowe. Obecnie naprz., przyjmuje się, że to zajęcie uważa się za zawód, na które sam zapytywany wskaże.

Samo zapisywanie odpowiedzi posiada wiele stron ciekawych i zależy od wielu warunków, aby uzyskany materiał był istotnie wartościowy. Zapisywanie dokonywane jest przez specjalny personel statystyczny, t. zw. *komisarzy spisowych*, albo rejestrantów, których zadanie polega na wręczeniu i odebraniu arkuszy spisowych, a częstokroć i zapelnieniu ich, o ile ludność ze względu na niski stopień kultury nie jest w stanie sama tego dokonać.

Źródłem wiadomości statystycznych przy badaniu wyczerpującem jest ogół ludności kraju. Zeznaniom ludności towarzyszy szereg okoliczności, które należy omówić.

Odróżniamy odpowiedzi bałamutne, niewiarogodne, częściowo kłamliwe, powstałe wskutek nieuświadomienia lub niezna-

jomości rzeczy ze strony ludności — od odpowiedzi *tendencyjnie fałszywych* wywołanych obawą, strachem, ambicją, wstydem i t. d. Powodowana ukrytymi pobudkami ludność składa fałszywe zeznania w przypuszczeniu, że w ten sposób broni się przed zamierzonym uciskiem władz. Statystyka żywot swój wie, że od zapotrzebowań państwa pod względem wiadomości skarbowych i wojskowych. Te zadania statystyki przetrwały w świadomości ludu i są wystarczające, by wzniecić nieufność, zwłaszcza jeżeli rejestracji dokonywują organa rządowe. Dużą rolę odgrywają również psychologiczne momenty, jak naprz., wstyd do przyznania się, że zawodowo uprawia się złodziejstwo, prostytutkę i t. p.

Okoliczności powyższe są źródłem błędów w statystyce — dość pospolitem, ale nie jedynym. Obok nich zdarzają się błędy. pochodzące z *przypadku*, ale jako takie nie są niebezpieczne, ponieważ przy podsumowaniu wyników dążą one do wyrównania. Często również wkradają się do zeznań ludności błędy, pochodzące z *zapominania*. Ten rodzaj błędów również nie jest niebezpieczny, ponieważ zapominaniu ulegają zwykle fakty drobne, bez znaczenia, natomiast zdarzenia większe przeważnie mocno tkwią w pamięci.

Pospolitym błędem w zeznaniach ludności jest brak ścisłości cyfr, wynikający z pociągu do zaokrąglania liczb. Ten pociąg jest charakterystyczny i przejawia się w zeznaniach majątkowych, ekonomicznych, wieku i t. p. Zeznający, o ile ma podać straty majątkowe cyfrą naprz., 1167, zaokrągla ją i podaje 1200. W statystyce ludności, z podziałem według wieku, ludzi w wieku lat, zakończonych na 5 lub 0, jest więcej, aniżeli lat przyległych. Tak ludzi 40-letnich jest więcej, jak 41-letnich, mimo, że ci 41-letni rok wcześniej byli liczniejsi od 40-letnich w obecnym roku. Błąd ten powstaje niewątpliwie wskutek zaokrąglania i może być usunięty przez odpowiednie grupowanie za pomocą podziału liczb zaokrąglonych równomiernie pomiędzy sąsiednie klasy.

Również podobnej kategorii jest błąd, zdarzający się często, gdy chodzi o wiek starców. Mianowicie starcy mają

skłonność do powiększania dat swego wieku. Ósmdziesięcioletnich, dziewięćdziesięcioletnich i stuletnich starców mamy według wykazów statystycznych więcej, aniżeli ich jest w rzeczywistości. Pochodzi to skutkiem chęci staruszków podkreślenia swej żywotności, a także z braku pamięci. Z tego względu statystyka starców jest mało wiarygodna, o ile opiera się na zeznaniach, a nie na dokumentach.

Z powyżej przytoczonego o warunkach zbierania zeznań wnioskujemy, że pola do nadużyć prawdy w statystyce jest bardzo dużo. Jest nawet tak dużo, że początkującego statystyka ogarnia przerażenie. Znakomity Mayr jednakże powiada: „początkujący statystyk odkrywa wielką liczbę źródeł błędów i zaczyna powątpiewać o wiarygodności materiału statystycznego; przy głębszej analizie bardzo często przychodzi jednak do wniosku, że źródła błędów są przecenione i że ostateczny rezultat jest dość stały”.

Istotnie, są warunki, które zmniejszają niebezpieczeństwo bałamutnych, kłamliwych i tendencyjnych zeznań w badaniu masowym. Do takich warunków należą: oświata i kultura ludności, przyzwyczajenie do badań statystycznych, dobór i wykształcenie personelu. Pierwszy warunek nie wymaga jakichkolwiek wyjaśnień, co do drugiego — praktyka państw współczesnych wykazuje, że gdzie badania statystyczne o charakterze powszechności otoczone są dostateczną powagą i opieką państwa oraz prowadzą do celów, niemających nic wspólnego z wyzyskiem ludności, tam w szybkim stosunkowo czasie stają się one popularne. Wybór personelu statystycznego posiada wielkie znaczenie ze względu na konieczność wzbudzenia zaufania u ludności. W państwie rosyjskim przed wojną szereg badań statystycznych przeprowadzano za pośrednictwem znieuawidzonej przez społeczeństwo policji. Nie należy się dziwić, że ludność stroniła od rzetelnych świadczeń statystycznych i że badania chybiały celu. Nawet statystyka rosyjska zbierana przez inspektorów fabrycznych w zakresie życia przemysłowego i robotniczego ze względu na nieufność, jaka panowała ze strony robotniczej do inspektoratu fabrycznego, mimo dość

znacznego cenzusu wykształceniowego samych inspektorów fabrycznych — nie stała na wysokości zadania, mając zasadnicze braki pod względem wiarygodności.

Ażeby przeciwdziałać błędom, które mogą przeniknąć do statystyki skutkiem przypadkowości lub złośliwej tendencji, zaleca się stosowanie kontroli zewnętrznej, która polega na zasięganiu odpowiedzi w odpowiednich warunkach u różnych grup społecznych. Znajduje to zwłaszcza zastosowanie wówczas, gdy dany kompleks zjawisk społecznych może być oświetlony z dwu lub więcej źródeł społecznych. Gdy chodzi naprz., o badanie sprawy przemysłu w związku z zagadnieniem robotnictwem nie wystarcza czerpanie dat statystycznych tylko od przemysłowców, lecz należy je skontrolować datami, zaczerpniętymi bezpośrednio od warstwy robotniczej, lub w odpowiednim wypadku vice versa. Kontrola zewnętrzna daje możliwość zestawienia dwu niejako krańcowych odpowiedzi i wypośredkowania sądu, niepozostającego w rażącym przeciwieństwie do rzeczywistości.

Rejestracja bieżąca.

Wyczerpujące badanie statystyczne dzieli się zasadniczo na dwa rodzaje. Jeżeli spostrzeganie statystyczne skierowane jest na określenie stanu elementów społecznych w przekroju jednego momentu (np., dnia, jak przy jednodniowych spisach ludności), mamy wówczas do czynienia ze spisem. Spisy mogą być różnorakie w zależności od tego, co poddajemy spostrzeganiu. Mamy więc spisy ludności, zwierząt domowych, w szczególności koni dla celów wojskowych, mieszkań, przemysłowe, rolnicze, dzieci w wieku szkolnym itp.

Jeżeli natomiast spostrzeganie statystyczne, zamiast uchwycenia określonych momentów, trwa nieprzerwanie, jest ciągle i ma na celu notowanie zmian, powstających w zjawiskach społecznych — mamy wówczas do czynienia z rejestracją bieżącą.

Rejestracja bieżąca obejmuje zjawiska płynne, niejako codzienne. Z tego wynika, że organa, dokonywujące notowań,

winni być stałe, a nie czasowe, jak to ma miejsce w wypadku spisów. W każdym państwie prowadzone są takie notowania przez organa administracji równoległe do czynności służbowych. Wiele z tych notowań niemal bezpośrednio styka się z przejawami życia społecznego, a przeto wchodzi w zakres statystyki, jako to: notowania metrykalne (urodzeń, małżeństw, zgonów), sezonowej i stałej emigracji, sanitarne, przestępczości, handlu zagranicznego, poczt i telegrafów, ruchu towarów (kolejowe i okrętowe) etc.

Z dziedzin społecznych, podlegających spostrzeganiu statystycznemu za pomocą rejestracji bieżącej, najważniejsza jest dziedzina ruchu naturalnego ludności. Zapisywanie faktów urodzeń, małżeństw i zgonów jest niezbędne nie tylko z punktu widzenia potrzeb demografii w dziale przyrostu ludności, który to przyrost ludności stanowi dla każdego państwa rzecz pierwszorzędного znaczenia, ale i w innych celach — dla prawa prywatnego i publicznego, które nawiązują skutki prawne do do tych faktów, stwierdzonych przez notowania. Rzecz jest charakterystyczna, że państwo znacznie później przyszło do zrozumienia znaczenia tych zapisów, aniżeli kościół chrześcijański, który pierwszy wkroczył na drogę prowadzenia zapisów metrykalnych w obrębie swego wykonania, t. j. przez organa kościelne. Wkrótce uznało państwo te zapisy za akta publiczne i z biegiem czasu zainteresowało się systemem prowadzenia ksiąg metrykalnych, a także w niektórych wypadkach, jak we Francji, powołało do życia własne urzędy stanu cywilnego, niezależne od duchowieństwa.

Spisy ludności.

Ze spisów — najważniejsze są spisy ludności. W czasach nowożytnych spisy ludności należą w życiu państwowym niemal do zjawisk codziennych. W Anglii, poczynając od roku 1801, co dziesięć lat odbywa się spis ludności; w Stanach Zjednoczonych Ameryki Północnej również co lat 10; w Niemczech co lat 5; podobnie rzecz się ma ze spisami ludności w innych krajach europejskich. Spis ludności jest zarządzeniem stałym, perjo-

dycznie się powtarzającym, coraz bardziej udoskonalaniem i rzeczywiście dającym niesłychanie cenne wskazówki dla administracji państwowej.

Jest to jedyny rodzaj badań statystycznych, który był znany w czasach zamierzchłej starożytności. Jeszcze, jak już wspominaliśmy, nie było nazwy statystyki, która weszła w użycie dopiero w czasach najnowszych, a już konieczność zbierania dat statystycznych była natyle gwałtowna, że pod rozmaitemi postaciami badania statystyczne były przeprowadzane. Tak Konfucjusz daje świadectwo, że w XXII stuleciu przed naszą erą był spis ludności w Chinach, połączony ze zbieraniem wiadomości o dochodowości gruntów, ruchu handlowym i t. p. Herodot rozporządza cyframi z czasów Darjusza i Kserksesa; w starożytnym Egipcie znany był kataster, spisy ludności i prowadzone były akta stanu cywilnego; Żydzi przywiązywali wielką wagę do statystyki, czego dowodem może służyć „księga liczb”; Grecja za czasów Likurga i Solona przeprowadza takie reformy społeczne, które musiały mieć oparcie w dokładnych danych statystycznych, dotyczących podziału na klasy i stanu majątkowego ludności; Serwusz Tulljusz w r. 550 przed Narodzeniem Chrystusa wprowadza szczególny rodzaj badań ludnościowych pod nazwą cenzusu itd., itd.

Z biegiem czasu, od zamierzchłej ery państw starożytnych — aż do czasów najnowszych, kiedy spisy ludności stały się zjawiskiem codziennym, absolutnie niezastąpionym przez jakiegokolwiek inne badania masowe, staje się coraz widoczniejsza ewolucja teoretycznych założeń spisów oraz ich praktycznego wykonania. Z nieudolnych rejestracji, częstokroć na fałszywych podstawach opartych, dążących tu i owdzie do celów diametralnie sprzecznych, prawie zawsze wywoływanych skrajnym utylitaryzmem w kierunku wyciskania podatków, wyrasta *społeczny typ spisu ludności*, wolny od tendencji politycznych i nie mający nic wspólnego z bezpośrednim obciążeniem ludności ciężarami podatkowymi. Współczesne spisy ludności mają głównie na celu badanie masy społecznej drogą ilościowego ustalenia poszczególnych grup według wieku, stanu, wyznania,

płci, zawodu i t. p. dla naukowego zrozumienia, czym jest dane społeczeństwo i jakie są jego potrzeby. Przeprowadza się je w interesie poznania tych potrzeb i ich zadowolenia, lecz nie dla obciążenia ludności nowymi podatkami, bo do tego drogi są inne.

Nad stroną teoretyczną spisów pracowano głównie w wieku XIX. Znane są w tym kierunku prace kongresów statystycznych, których zapoczątkodawcą był nieśmiertelny Quetelet, inicjator pierwszego powszechnego jednodniowego spisu ludności w Belgji w r. 1846. Sprawę spisów ludności omawiano na kongresach w Brukselli, Londynie, Berlinie i Petersburgu. Już wówczas zwrócono uwagę na najważniejsze momenty spisu, a mianowicie: kto podlega spisowi, kiedy spisy winny się odbywać, w jakich odstępach czasu i jak winny być dokonywane. W zakresie pierwszego pytania już kongres brukselski w roku 1853 ustalił, że spisywać należy ludność faktyczną (*population de fait*), t. j. obecną w chwili spisu, bez względu na to, czy w danej miejscowości chwilowo się znajduje, czy stale prawnie do danego miejsca jest przypisaną. Poza tem ustalono, że spisy winny być imienne i odbywać się co lat 10 w tym miesiącu, w którym ludność jest najmniej ruchliwa (w grudniu). W Europie — Belgja, Holandja, Szwecja, Norwegja, Włochy, Hiszpanja, Austria i Węgry spisują ludność 31 grudnia. Niemcy i Szwecja — 1 grudnia; Danja — 1 lutego; Portugalja — 1 listopada; Anglja spisuje ludność na wiosnę, jak również Francja. Termin spisu winien być jaknajkrótszy, t. j. jednodniowy. Gdyby jednak spis nie mógł być ukończony w ciągu jednego dnia, ze względów natury technicznej, wszystkie dane muszą się stosować do jednego. ustalonego z góry, dnia. Jest to wymaganie najistotniejsze. Spis winien dać przekrój w jednym jakimś momencie. Zazwyczaj przyjmuje się godzinę 12 w nocy na dzień spisu. W ten sposób dla całego terytorjum państwa dane statystyczne są jednoczesne, przez co teoretycznie unika się błędów przy porównaniu. Jest to zrozumiałe, że podobne zdjęcie momentalne, utrwalające stan faktyczny w określonym terminie, nie należy do rzeczy łatwych do przeprowadzenia.

Kongres statystyczny międzynarodowy w Petersburgu w r. 1872 specjalnie interesował się zagadnieniem klasyfikacji danych, mających wejść do spisu w każdym państwie, a dotyczących dziedziny demograficznej. Są one następujące: 1 — imię i nazwisko, 2 — płeć, 3 — wiek, 4 — stosunek do głowy rodziny i gospodarstwa, 5 — stan cywilny i małżeński, 6 — zajęcie lub stan, 7 — wyznanie, 8 — język w użyciu, 9 — umiejętność czytania i pisania, 10 — pochodzenie, mielsce urodzenia, 11 — narodowość, 12 — zwykły pobyt i rodzaj pobytu w miejscu, gdzie odbywa się spis, 13 — ślepi, 14 — głuchoniemi, 15 — nienormalni umysłowo. Wszystkie inne wiadomości kongres uważa za dodatkowe, mogące interesować poszczególne państwa, lecz nie stanowiące schematu międzynarodowego.

Od czasu zebrania się ostatniego Kongresu statystycznego ubiegło zgorą lat 50. Przez ten czas poszła naprzód zarówno wiedza statystyczna, jak i wzrosło doświadczenie. Poszczególne państwa nigdy, zresztą, ściśle nie stosowały się do wymagań kongresów statystycznych, zachowywano tylko główne warunki, stanowiące ośnowę spisów ludności, pozatem traktowano rzeczy indywidualnie, w zależności od miejsca i czasu. To jednak, co dziś uważane jest za warunek sine qua non poprawnego spisu, nie zostało zdobyte bez wysiłków i prób, często chybionych. Współczesny spis ludności, oparty na ustalonych zasadach teorii i praktyki, jest wyrazem wielowiekowej pracy narodów. Wiele prób zmarnowano, wiele sposobów użyto, zanim osiągnięto rzeczywiście pewne podstawy do tego, aby spis ludności mógł dać pożyteczne i wiarogodne wiadomości.

We Francji pierwsze próby spisów wykonano za czasów słynnego Colberta, nie zostały one jednak uwieńczone pomyślnym skutkiem. Podobny los spotkał i próby Vauban, koło roku 1700. W późniejszym czasie Necker obliczył ludność Francji, ale nie bezpośrednio, tylko przyjmując jedno urodzenie na 25.5 mieszkańców. Godnem jest uwagi, że w tym czasie Lavoisier, kiedy mu powierzono obliczenie przestrzeni uprawnych, porzestał na obliczeniu plugów, i dopiero tą drogą — pośrednio

— rozwiązał zadanie. Dekret z dnia 16 stycznia 1822 r. ustanawia 5-letnie spisy ludności, ale pierwsze obliczenie wykonane było w r. 1831. Nawet spis w r. 1856 wykonany był jeszcze przy pomocy urzędników podatkowych, co naturalnie wpłynęło ujemnie na rezultaty spisu. Dopiero spis ludności we Francji w r. 1861 został przeprowadzony z wielką dokładnością, z zastosowaniem postanowień i zaleceń Kongresu Statystycznego.

W *Prusach* obliczenia ludności wprowadzono już w r. 1683 za elektora Fryderyka Wilhelma. Poczynając od r. 1748 obliczano ludność Prus corocznie, przedstawiając Fryderykowi II wykazy ludności, któremi król bardzo się interesował dzięki pracom Süssmilcha. Wskutek dojścia do skutku w r. 1834 związku celnego niemieckiego, miały się odbywać co 3 lata spisy ludności w celu rozkładu dochodu między państwami, wchodzącymi w skład związku. Dopiero od r. 1861, dzięki pracom Eugela, spisy ludności oparte zostały na zasadach pewnych.

W *Austrji* reskrypty cesarskie 1753 i 1754 roku zarządzały spis ludności, ale były to spisy niedokładne. Pierwszy spis ogólny doszedł do skutku w całym państwie dopiero w r. 1850, ale także był niedokładny; z tego względu opracowano nowy projekt, który wszedł w życie w r. 1857.

W *Rosji*, do czasu wojny, jedyny spis ludności odbył się w r. 1897. Historia tego spisu ujawnia nieudolność pod względem fachowym i skrajną tendencję polityczną. Przedtem surogatami spisu były t. zw. *rewizje*, biorące początek od roku 1710 (ukaz Piotra Wielkiego), ale najlepszą charakterystyką tych badań statystycznych będzie to, że dopiero w r. 1858 zaczęto notować płeć żeńską, zajmując się dotąd tylko ludnością męską, opłacając podatek pogłówny. Jeżeli w rewizjach notowano liczbę kobiet, to tylko „dla samej ciekawości”.

Spisy w dawnej Rzeczypospolitej.

I w dawnej Rzeczypospolitej, jak wszędzie, sprawa liczenia ludności nie stała na wysokości zadania. Jak stwierdza Kleczyński, badania nad przeszłością tego rodzaju nie są łatwe. Główne trudności leżą w zmiennej konstrukcji

urzędów i w przechodzeniu akt z archiwum do archiwum. Dodać do tego należy rozbiory państwa polskiego i niszczenie zabytków przez zaborców, tudzież wywożenie zbiorów do Moskwy i Petersburga.

W Polsce administracja była słaba. Nie należy się przeto *a priori* spodziewać dobrej statystyki. Gdy jednak kiedyś zarząd skarbowy był zcentralizowany, a poborcy podatkowi obowiązani byli składać przed Sejmem odpowiednie wykazy, liczba włościan i łańów mogła być z mniejszą, lub większą dokładnością ustalona. Prof. *Pawiński* w dziele „Polska XVI wieku pod względem topograficzno-statystycznym” odtworzył ludność Wielko- i Małopolski z końca XVI wieku.

Z chwilą, gdy system skarbowy został zdecentralizowany, liczbę ludności do podatku podymnego oznaczano pod przysięgą, bez liczenia. Zaprowadzone w r. 1717 pogłównne chrześcijańskie wybierano również nie według liczenia, a rozkładano na województwa, stosownie do porozumienia między niemi. Dopiero Sejm Czteroletni wprowadza zmianę w zakresie stanu statystyki, a także już nawet po pierwszym rozbiore widać zainteresowanie się władz liczbą ludności i jej życiem gospodarczem.

W dawnej Rzeczypospolitej statystykę ludności, zresztą jak podówczas w całej Europie, traktowano jedynie z punktu widzenia potrzeb skarbowych. Począwszy od r. 1564 w uniwersałach poborowych i konstytucjach jest ciągle mowa o pogłównem żydowskiem. Spisy przeprowadzali sami zainteresowani, stwierdzając pod przysięgą liczbę głów gminy żydowskiej. Spis z r. 1768 przyniósł podatku, licząc po 2 złote od osoby, w Koronie 859,312 złp., na Litwie — 315.297 złp., t. j. gdyby spis był zupełny, liczba Żydów wynosiłaby 587,305 głów.

Pierwszy rozbiór Polski zbudził z odrętwienia naród. Departament Policji w Radzie Nieustającej zarządził w roku 1777 spis ludności miast z podziałem na chrześcian i żydów, osiadłych i luźnych. Prof. *Korzon* odnalazł księgę z tych czasów p. t. „Stan miast królewskich koronnych z r. 1777”. W aktach

archiwum krakowskiego odnaleziono „Popis gospodarczy i wszystkich ludzi tak męskich, jak i białej płci na kwartale rzeźniczym w miesiącu wrześniu 1773 r. spisany”. Wykaz ten sporządzono kamienicami, a w każdej kamienicy podani zostali gospodarz i lokatorzy, a w każdej rodzinie ilość osób.

Kraków dokonał kilku takich spisów. W r. 1790 odbył się t. zw. spis jeneralny oraz w tymże roku sporządzono tabelę ludności w mieście Kazimierzu, Stradomiu i Podzamczu. W r. 1795 był również w Krakowie spis 4 kwartałów i zamku.

Z tego samego czasu w archiwum konsystorskiem krakowskiem znajduje się spis ludności djecezji krakowskiej, przeprowadzony w r. 1786 z polecenia prymasa Poniatowskiego. Spis odróżnia katolików, dyssydentów i żydów — mężczyzn, kobiet, dzieci od 7 roku życia, przystępujących do komunji św., i do lat 7. Daty statystyczne są zebrane z 561 parafij, z 208 miast, z 3668 wsi i 274 kolonij, łącznie z 4150 miejscowości. Opracował go ks. *Siemieński*. Podobny sumarjusz znalazł *Kleczyński* i dla djecezji Płockiej.

Jednym z ciekawszych dokumentów jest spis Warszawy i Pragi z r. 1787. Spis był *jednodniowym*, obowiązek przeprowadzenia go włożony był na właścicieli domów. Ludność Warszawy wyniosła 89448 osób. Spis ten nosi charakter gospodarczego.

Podczas Sejmu Czteroletniego duże zasługi dla statystyki położył poseł bractawski — *Moszyński*. Obliczał on ludność na podstawie lustracji dymów, przyjmując po 6 osób na 1 dym. Sejm na wniosek Moszyńskiego 22 czerwca 1789 r. zlecił przeprowadzić nowe obliczenie dymów: był to pierwszy ogólny dla całego państwa spis ludności. Wzory formularza z pytaniami były 4 typów: dla wsi i miast, dla chrześcian i Żydów. Niestety, spis ten obejmował tylko ludność podatkową, ale niecałkowitą. Wogóle, do chwili obecnej spis ten nie jest analitycznie potraktowany i zbadany. Być może, jak świadczy *Kleczyński*, dałby on możność odtworzenia stanu ludności Rzeczypospolitej z r. 1789.

Spis ludności
z r. 1921.

Zarządzony na dzień 30 września
1921 roku spis ludności Rzeczypos-
politej Polskiej zawiera orócz dat

demograficznych dane, dotyczące budynków, mieszkań, inwentarza żywego i gospodarstw wiejskich. Był to więc spis nie wyłącznie ludnościowy, lecz połączony ze spisem gospodarczym, co jak na pierwszą próbę spisu jednodniowego w niepodległej Polsce stanowiło program za obszerny. Koniec września, jako termin spisu, wybrano ze względu li tylko na jaknajrychlejsze przeprowadzenie spisu bez oglądania się na wadliwość tego terminu (święta żydowskie i okres kopania kartofli). Spis przeprowadzono w warunkach ciężkich: istniała wielka trudność w zakresie przejazdów dla komisarzy spisowych, srożyły się choroby zakaźne, które sprawiły, że kilku komisarzy spisowych zapadło na tyfus, stosunki gminne były nieuregulowane; z wyjątkiem zaboru pruskiego, gdzie ludność była obeznana ze spisami, czynności spisowe wzbudzały wielką nieufność u ludności, a na terenach narodowościowo-mieszanych powodowały przeszkody ze względów politycznych. Napięcie nieufności wobec spisu, jako operacji nieznannej, o celach rzekomo nieprzyjaznych, powodowało w nielicznych wprawdzie okolicach składanie notorycznie fałszywych zeznań. W niektórych także okolicach, jak n.p. w wielu gminach Kongresówki i Ziemi Wschodnich wystąpił charakterystyczny objaw nieznajomości przez włościanina liczby posiadanych jednostek ziemi. Powodowało to zrozumiałe niedokładności. Zdarzały się wypadki odmowy zeznań, a także oporu, w kilku miejscowościach włościanie przekonani, że spis ma posłużyć do celów podatkowych, ukrywali żywy inwentarz po lasach, albo po odległych pastwiskach, a nieraz zamykali mieszkania i znikali podczas spisu — w przypuszczeniu, że rząd dokonywa spisu celem poboru młodzieży do służby wojskowej, nałożenia podatków i przeprowadzenia rekwizycji. Jest rzeczą wielce charakterystyczną, że przebieg spisu odbył się znacznie spokojniej na Ziemiach Wschodnich, aniżeli w Małopolsce Wschodniej i b. Kongresówce.

Technika przeprowadzenia czynności spisowych może być w krótkości wyjaśniona na przykładzie spisu m. Lublina. Naczelnny komisarz spisowy*) po zaznajomieniu się dokładnem z instrukcjami spisowemi odbył podróż po peryferjach miasta w celu zapoznania się szczegółowego z granicami i mniej znanymi dzielnicami miasta. Okazało się, że szereg ulic m. Lublina nie posiadał nazw, które trzeba było naprędce utworzyć i zanotować na planie. Stwierdzono nieprawidłową i fałszywą numerację domów, wobec czego zarządziliśmy nową numerację domów i placów specjalnie przeznaczoną dla użytku statystycznego, ażeby agenci spisowi popełniać omyłek lub dokonać opuszczeń nie mogli. Z chwilą, kiedy każdy dom, każdy plac został zarejestrowany, kiedy pomocnicy komisarza w liczbie dwóch zapoznali się najdokładniej z planami miasta, przystąpiono do studjów nad aglomeracją ludności miejskiej, badając stopień skupienia na poszczególnych ulicach — na podstawie dawniejszych badań statystycznych, jako to spisu z r. 1916, przeprowadzonego przez władze okupacyjne, oraz wykazów aprowizacyjnych, które kilkakrotnie w tym czasie były sporządzane sposobem administracyjnym. W wyniku tych studjów podzielono miasto, zarówno centrum, jak i rozległe peryferja, na 330 okręgów spisowych z równomiernem podziałem ludności między okręgi, z równoczesnem sporządzeniem planów tych okręgów z załączonemi wykazami ulic, placów, domów z nową numeracją — w celu wręczenia ich agentom spisowym. Zaczem nastąpiło werbowanie personelu i wyszkalanie go za pomocą cyklu systematycznych wykładów. Wydano do ludności miasta odezwę za podpisem prezydenta miasta i naczelnego komisarza spisowego, wyjaśniającą cele i zadania spisu, w celu zjednania sobie przychylności mieszkańców i wytworzenia życzliwego nastroju. Jednocześnie w prasie miejscowej opublikowano szereg źródłowych artykułów na temat spisów ludności, a na tydzień przed spisem w 10 punktach miasta urządzono popularne odczyty dla szerokich warstw

*) autor niniejszej pracy.

ludności w celach informacyjnych i propagandowych. W dzień spisu naczelny komisarz spisowy zorganizował lotną inspekcję dla przekonania się, czy przebieg spisu jest normalny.

Po dokonaniu spisu, w kilka dni potem, zarządzono za pośrednictwem organów policyjnych sprawdzenie, czy w każdym domu był agent spisowy. Stwierdzono, iż na przedmieściu jeden domek został przez niedbalstwo, czy omyłkę opuszczony, wobec czego dokonano w nim spisu.

Dodać należy, że agenci spisowi rekrutowali się z personelu nauczycielskiego, głównie szkół powszechnych, studentów oraz młodzieży szkół średnich wyższych klas, która tłumnie stanęła do apelu. Dzięki doborowi sił spisowych ludność odniosła się z pełnem zaufaniem do czynności spisowych, chętnie i wiarogodnie zeznawała, dzięki czemu nie było potrzeby powtórzenia spisu, jak to się zdarzyło w kilku innych dużych miastach.

Analizując warunki spisu z r. 1921 oraz charakter wyników jego, prof. Ludwik *Krzywicki* stwierdza, co następuje. Wiarogodność wykazanych dat statystycznych stosownie do dziedziny badania (ludzie, ziemia, inwentarz żywy) zmniejsza się w określonym kierunku. Najbliższe prawdy są liczby, dotyczące ludności, ale i tutaj bardzo niedokładna jest rubryka wieku. Liczby, dotyczące ilości gruntów, a tembardziej podziału ich na użytki, są bardzo niepewne. Najmniej wiarogodne są liczby, dotyczące zwierząt gospodarskich, a szczególnie drobiu.

Spis dzieci z r. 1926. W celu zrealizowania planu *powszechnego nauczania* władze Ministerstwa W. R. i O. P. zorganizowały w czerwcu 1926 r. spis 13 roczników dzieci, t. j. 7 roczników szkolnych i 6 przedszkolnych, który dostarczył materiałów statystycznych dla preliminowania etatów nauczycielskich, budowy szkół powszechnych, jej sfinansowania oraz dla reorganizacji sieci szkolnej. Wyniki spisu dzieci w zakresie potrzeb szkolnictwa powszechnego

zostały w następstwie opracowane przez Marjana *Falskiego* i opublikowane.

Organizacja spisu była następująca. Wszystkie powiaty Rzeczypospolitej zostały podzielone przez inspektorów szkolnych na rejony instruktorskie. Na każdy rejon instruktorski wyznaczono jednego instruktora spisowego i jednego zastępcę z pośród [nauczycieli szkół powszechnych. Z kolei rejony zostały podzielone na okręgi spisowe, z których każdy otrzymał komisarza spisowego z grona nauczycieli. Materiał składał się z arkuszy spisowych, map i wykazów miejscowości. Spis dzieci powinien być rozpoczęty w dniu, wyznaczonym przez władze szkolne i ukończony w ciągu nie więcej niż kilku dni. Zarówno mapka okręgu spisowego, jak i wykaz miejscowości mogły być być niedokładne, komisarz spisowy obowiązany był przeprowadzić odpowiednie sprostowania.

Wypełnione karty spisowe winny były być podsumowane i wraz z poprawionymi mapkami okręgu doręczone instruktorowi rejonowemu natychmiast po dokonaniu spisu. Instruktor rejonowy porządkował otrzymane arkusze spisowe, korygował mapkę rejonów oraz wykaz miejscowości i cały materiał statystyczny wręczał inspektorowi szkolnemu, który z kolei przy pomocy instruktorów rejonowych sporządzał zestawienie powiatowe.

Omawiany spis dzieci odbył się w miesiącu czerwcu 1926 r. zaraz w pierwszym tygodniu miesiąca i trwał zazwyczaj do 3 dni. Spis wykazał 98.758 miejscowości na całym obszarze Rzeczypospolitej z liczbą dzieci urodzonych od 1913 do 1925 roku, t. j. w obrębie 13 roczników — 8.112.954, z czego na miasto przypada 1.759.246, a na wieś — 6.353.708 dzieci.

Spisy gospodarcze.

We wszystkich tych państwach, gdzie niema normalnego, perjodycznego powtarzanego, dobrze zorganizowanego i na ustalonych zasadach teoretycznych opartego spisu gospodarczego, stan rzeczywistości gospodarczej przedstawia się mętnie, niewiarogodnie, nieraz z gruntu fałszywie.

Znakomity Inama *Sternegg* w przedmowie do wydania pierwszego spisu gospodarczego w Austrii powiada: „celem praktycznego wykonania zadań polityki ekonomicznej i społecznej jest niezbędna najdokładniejsza analiza poszczególnych składników organizmu narodowo-gospodarczego zarówno ze strony ilościowej, jak i jakościowej; do takiej analizy ogólny spis zawodowy i przedsiębiorstw jest niezbędną przesłanką”. Atoli organizacja spisów gospodarczych nie należy do rzeczy łatwych, za dowód czego służyć mogą pierwsze spisy gospodarcze w Niemczech, dokonane w latach 1875 i 1882, a poniekąd nawet w r. 1895. Drugi i trzeci spis mają właściwie charakter spisów zawodowych, a nie spisów przedsiębiorstw przemysłowych, którymi zajął się prawie wyłącznie spis z r. 1875. Krytyka uznała niemal wszystkie te trzy spisy, a zwłaszcza pierwszy za nieudane.

Pierwsze spisy przemysłowe, wykonane we Francji i Belgji, aczkolwiek zakreślone mniej obszerne, jak spisy niemieckie, również nie dały pożądaných rezultatów. Większość państw europejskich poprzestaje na zbieraniu corocznych danych, obawiając się trudności, z którymi jest połączona statystyka przemysłu na drodze spisu. Dane te pochodzą bądź bezpośrednio od przedsiębiorców za pośrednictwem organów administracji, bądź od izb przemysłowo-handlowych, lub tym podobnych instytucyj, bądź od inspektoratów fabrycznych. Trudności spisów przemysłowych połączone są z rozproszeniem przemysłu po terytorjum państwa, z różnorodnością form przemysłu, nastręczającą przeszkody do jednolitej nomenklatury, z tendencją utajania prawdy ze strony sfer gospodarczych.

Godnem jest uwagi, że w r. 1928 władze w Polsce, pragnąc znaleźć oświecenie dla spraw przemysłu, zarządziły badanie ankietowe, publikując jego wyniki w formie oddzielnych książek, poświęconych poszczególnym gałęziom przemysłu, jako sprawozdania Komisji ankietowej przy Komitecie Ekonomicznym Rady Ministrów.

Od spisów przemysłowych zasadniczo różnią się spisy zawodowe, których przeprowadzenie nastęrcza znaczne prze-

szkody z powodów podobnych, jak przy spisach przemysłowych. Trudne do rozwiązania zagadnienie jednolitości nomenklatury zawodów, przy tak znacznej, jak wiadomo, ich liczbie, zaproponował międzynarodowy Instytut statystyczny na sesji w Chicago rozstrzygnąć w ten sposób, żeby utworzyć 3 nomenklatury. Pierwsza zawiera 61 grupę. Następna, stanowiąc bardziej szczegółowe rozbięcie pierwszej, składa się z 206 grup. Wreszcie ostatnia, rozbudowana w podobny sposób, łączy 499 zawodów.

Zbieranie statystyki rolniczej.

Podstawą dobrej statystyki rolniczej jest właściwie kataster, obejmujący obok innych danych także pomiary własności gruntowej, jej liczebność oraz dochodowość, zawierający w ten sposób większą ilość dat statystycznych, które co prawda wymagają systematycznych poprawek w miarę dokonywujących się zmian w ustroju własności i produkcji rolnej, ale stanowią niezaprzeczenie najlepsze źródło wiadomości rolniczych.

Już pierwszy kongres statystyczny, obradujący w Brukseli w r. 1853, wyraził życzenie, aby w każdym państwie był dokonany pomiar własności gruntowej i uchwalił nawet zasady ogólne dla urządzenia katastru w państwach, nie mających go. Ale słusznie powiedział prof. *Radziszewski* w mowie sejmowej w czerwcu 1921 roku, że kataster należy do bardzo skomplikowanych operacji państwowych, przytaczając na dowód, że zarządzenia we Francji w celu przeprowadzenia katastru, wydane za Napoleona I, zostały dopiero ostatecznie wykonane i zakończone za Napoleona III. Korzystanie więc z katastru nie zawsze może mieć miejsce, ponieważ wskutek trudności technicznych i finansowych kataster nie we wszystkich państwach został przeprowadzony i nie prędko będzie urządzony tam, gdzie go nie ma, jak np. w Polsce.

Dlatego muszą istnieć i inne źródła wiadomości rolniczo-statystycznych, pozwalające państwu oprzeć gospodarkę narodową na pewnych podstawach znajomości życia ekonomicznego,

bez których jakakolwiek rozumna polityka ekonomiczna byłaby niemożliwa.

Sprawa ta była niejednokrotnie poruszana na międzynarodowych Kongresach Statystycznych. Na Kongresie w Brukseli obok kwestji spisu ludności, katastru, o czym wyżej wspominaliśmy, emigracji, statystyki przemysłowej etc. żywo interesowano się i poddano pod obrady kwestję spisów rolniczych. Statystyka rolna była omawiana potem na kongresach w Paryżu w r. 1855, w Wiedniu w r. 1857 (punkt VI obrad dotyczył statystyki własności ziemskiej), w Londynie, w Berlinie w r. 1863 (podział własności gruntowej według 12 grup użytkowania), we Florencji i w Budapeszcie w r. 1876. Ten ostatni kongres zastanawiał się nad źródłami statystyki rolniczej i przyszedł do wniosku, że statystyka rolnicza winna się składać z 2 zasadniczych działów: 1) ze spisów, ponawianych w odstępach 10-letnich i 2) z wiadomości, zbieranych corocznie. Do pierwszego winien należeć opis rolniczy całego państwa, do drugiego: a) ceny produktów rolniczych, b) zbiory, c) ceny transportu.

Kwestji nie ulega, że spostrzeżenia z zakresu statystyki rolniczej posiadają różny stopień trwałości: gdy jedne z nich ulegają zasadniczym zmianom dopiero wciągu kilku lat, inne, ażeby dawały prawdziwy obraz rzeczywistości, muszą być corocznie dokonywane. Jeżeli porównamy dwa następujące rodzaje danych: podział na użytki rolne i wysokość plonów, to pierwsze będą należały do kategorii perjodycznych danych, powtarzanych w długich odstępach czasu, drugie — do corocznej obserwacji, ze względu na zmienność i wahania, jakim w krótkim przeciągu czasu ulegają. Gdy podział na użytki rolne wywołany jest pewnemi, trwale działającemi czynnikami ewolucji gospodarstwa wiejskiego, zarówno natury przyrodniczej, jak i ekonomicznej, wysokość plonów waha się z roku na rok w zależności od efektywnych warunków wegetacji, w rzędzie których dotychczas wybitną rolę odgrywają pomyślne lub niepomyślne wpływy atmosferyczne każdego poszczególnego roku. Wyczucie różnicy w naturze i trwałości zjawisk rolniczo-

statystycznych wysuwa na czoło zagadnienie użycia tej lub innej metody statystycznej do masowych badań rolniczych. Źródła statystyki rolniczej, różniąc się, jak na powyższym przykładzie, co do konieczności i potrzeby częstego odnawiania danych, mają również różnice i w celowości zastosowania tych lub innych metod badania; gdy jedne opierają się na spisach, drugie są rezultatem posługiwania się w tej lub innej formie—ankietą. Gdy pierwsze dają nam materiał we właściwym znaczeniu słowa — statystyczny, drugie — jedynie zbliżony do ściśle statystycznego.

Do pierwszej kategorii winny należeć: 1) dane, dotyczące statystyki agrarnej, t. j. stosunki ilościowe, cechujące strukturę oraz zmiany w posiadaniu ziemi (statystyka większej i mniejszej własności), 2) dane o ilości przedsiębiorstw rolnych, 3) oznaczenie rozmiaru typów i charakteru użytkowania rolniczego, 4) spis zwierząt domowych, 5) systemy nawożenia, płacy roboczej, ilość zatrudnionych rąk roboczych stale i czasowo, inwentarz martwy i t. p., 6) obszar pod zasiewami, przedstawiający gdzie, ile i czym dokonano obsiewu gruntów ornych.

Do drugiej kategorii zaliczamy: 1) statystykę plonów i od niej pochodną — zbiorów, 2) statystykę cen, 3) statystykę zapasów i in.

Powszechne badania rolnicze w Polsce.

Pierwsze wiadomości z zakresu rolnictwa w niepodległej Polsce dostarczył spis ludności z dn. 30 września 1921 r., bowiem spis ten połączony był ze spisem „gospodarstw rolnych, ogrodowych i leśnych”. W ciągu tegoż roku dokonane było specjalne badanie ankietowe (na zasadzie rozporządzenia Rady Ministrów z dn. 20 września 1920 r.), obejmujące statystykę wielkiej własności rolnej, określonej we wspomnianym rozporządzeniu na 50 ha i wyżej. Badanie nosiło charakter przymusowy, jednak nie było przeprowadzone przez specjalnych agentów spisowych sposobem ekspedycyjnym, lecz odpowiednie formularze, otrzymane za pośrednictwem sta-

rostw, musieli wypełnić sami właściciele, względnie ich zastępcy. Rezultaty tego badania zostały opublikowane p. t. „Wielka własność rolna”, w którym to wydawnictwie znajdujemy opracowanie pytań, dotyczących narodowości właścicieli, użytkowania ziemi, dzierżawy parcelowej, inwentarza żywego, maszyn rolniczych, przemysłu rolnego i służby. Inne zagadnienia, które zawierał arkusz spisowy, nie zostały opracowane z powodu różnych wadliwości zebranego materiału statystycznego.

Do stałych corocznych publikacji Głównego Urzędu Statystycznego należą, między innymi, „Zasiewy i zbiory” i „Ceny ziemi” w opracowaniu *E. Szturm de Sztrema*. Są to badania ankietowe, wypełniane przez korespondentów rolnych Gł. Urz. Statystycznego.

Badania ilości inwentarza żywego były przeprowadzane dwa razy: pierwszym razem łącznie z powszechnym spisem ludności w r. 1921, drugim razem — na specjalne zarządzenie Rady ministrów z dnia 17-go października 1927 roku. Drugi spis został przeprowadzony od 1 do grudnia 1927 roku według stanu z dnia 30 listopada tegoż roku — samodzielnie bez wszelkich innych głównych czy dodatkowych spisów. Na podstawie rozporządzenia władzami spisowymi były starostwa i w miastach liczących powyżej 25.000 mieszkańców magistraty tych miast. Wykonanie spisu należało do władz gminnych, które do technicznego spisywania użyły sołtysów, osoby z pośród personelu zarządu gminnego oraz osoby prywatne, zaangażowane w tym celu. Wyniki zostały ogłoszone drukiem już w r. 1928 p. t. „Inwentarz żywy w Polsce” w opracowaniu *E. Szturm de Sztrema*.

Powszechny Spis rolniczy.

Już w r. 1924 na sesji Międzynarodowego Instytutu Rolnictwa w Rzymie poruszona była sprawa *powszechnego spisu rolniczego* mającego się odbyć jednocześnie we wszystkich państwach. Po stosownych studjach i przygotowaniach nad projektem tego spisu na sesji w r. 1928 zapadła uchwała, ażeby spis był dokonany na obu półkulach w ten

sposób, iżby na półkuli północnej dotyczył upraw, zebranych w roku kalendarzowym 1929, zaś na półkuli południowej — upraw, zebranych w przeciągu roku od 1 lipca 1929 do 30 czerwca 1930 roku. Przygotowany projekt formularza spisowego zawiera następujące punkty: 1) imię i nazwisko kierownika gospodarstwa i jego adres, 2) obszar gospodarstwa z podziałem na użytki rolne, 3) ręce robocze w gospodarstwie z uwzględnieniem liczby godzin pracy, płac i mieszkań, 4) uprawy i zbiory i 5) inwentarz żywy. Ustalono zasadę, że każde państwo może przeprowadzić spis w tym okresie, który jest dlań najbardziej dogodnym.

Obok spisu gospodarstw rolnych projektuje się spis lasów, mający na celu wyjaśnienie zupełnie nieznanymi zagadnień ekonomiki leśnictwa, jako to podziału lasów według rodzajów drzew, według wieku, użytkowania leśnego, produkcji i t. p.

Główny Urząd Statystyczny w Polsce uważał za pożądane ze względu na konieczności państwowe, dotyczące podniesienia produkcji rolnej, reformy agrarnej, prawidłowej gospodarki leśnej, handlu zagranicznego wytworami rolnictwa, wyżywienia ludności i t. d. urządzenie spisu w Polsce już w r. 1929, co aczkolwiek nie nastąpiło.

Instytucje zbierania materiałów statystycznych

Statystyka administracyjna.

w żadnym państwie nie było właściwych instytucyj zbierania materiałów statystycznych. Bodaj pierwszą instytucją tego rodzaju była *Komisja Tabel*, powołana do życia w Szwecji w r. 1756, nie stanowiąca jednak urzędu statystycznego we właściwym znaczeniu słowa. We Francji pierwsze biuro statystyczne powstało przy Ministerstwie spraw wewnętrznych w roku 1801. Od tego czasu datuje się rozwój urzędów statystycznych w całej Europie jako instytucyj poświęconych zbieraniu, opracowywaniu i opublikowywaniu różnorodnych materiałów sta-

Przez bardzo długi okres dziejów, mimo, że rosło wciąż zapotrzebowanie na materiał statystyczny ze względu na konieczności państwowe, aż do początków XIX stulecia

tystycznych. Całokształt urzędów statystycznych, służących państwu w charakterze agend, pracujących nad materiałami statystycznymi, nosi nazwę statystyki administracyjnej.

Mayr odróżnia dwa rodzaje *statystyki administracyjnej*: pierwszą czyli *główną* (primäre), której poświęcone są specjalne urzędy statystyczne, i *wtórną* (secundäre), która powstaje współzrędnie z właściwymi czynnościami niektórych innych urzędów, stanowiąc dla nich produkt poboczny, jak np. na kolejach (statystyka kolejowa), w urzędach celnych (statystyka cel), na poczcie (statystyka pocztowa) i t. p.

W celu skoordynowania prac statystycznych w całym państwie, Francja od roku 1885 powołała do życia „Conseil superieur de statistique”, czyli wyższą *radę statystyczną*, jako rodzaj nadbudowy w ustroju urzędów statystycznych. Podobna organizacja okazała się bardzo celowa i została wprowadzona we wszystkich państwach. Do rad statystycznych należy kierownictwo pracami statystycznymi, do biur — wykonanie.

Praktyka państw europejskich wypracowała kilka typów statystyki administracyjnej. Jeżeli wszystkie prace statystyczne ześrodkowane są w jednym głównym biurze statystycznym, mamy wówczas do czynienia z systemem *centralistycznym*, który stanowi typ niemiecki. W przeciwieństwie do niego mamy typ angielski, który charakteryzuje się niezależnością poszczególnych biur, urządzonych na zasadzie *decentralizacji*. Francuska statystyka administracyjna zajmuje miejsce pośrednie między systemem centralistycznym i decentralistycznym, bowiem francuskie biura statystyczne urządzone w poszczególnych gałęziach administracji państwowej prowadzą samodzielne badania, mając obowiązek odsyłania materiałów statystycznych do głównego biura, które zajmuje się podsumowaniem nadesłanych materiałów. Najbardziej celowym jest typ niemiecki, który zabezpiecza stosowanie jednolitej metody badania na terenie całego państwa.

W Polsce centralną instytucją zbierania, opracowania, i opublikowania materiałów statystycznych jest *Główny Urząd Statystyczny* z siedzibą w Warszawie, do którego należy kie-

rownictwo wszelkimi pracami statystycznymi w państwie, mającemi charakter ogólny. Główny Urząd Statystyczny publikuje: 1) „Rocznik Statystyki Rzeczypospolitej Polskiej”, 2) „Kwartalnik statystyczny”, 3) „Wiadomości statystyczne”, 4) „Statystykę pracy”, 5) „Handel zagraniczny Rzplitej Polskiej”, 6) „Statystykę Polski”, obejmującą opracowania z różnych dziedzin, jak wyniki spisu z r. 1921, statystykę wyborów do Sejmu i Senatu i t. p. i 7) „Skorowidz miejscowości Rzplitej”. Na czele Urzędu od chwili powstania stał prof. dr. Józef *Buzek*. Do naczelnego kierownictwa pracami statystycznymi, głównie w formie omawiania i akceptowania projektów badań, powołana została Główna Rada Statystyczna.

Obok centralnych urzędów statystycznych w poszczególnych państwach, a także i w Polsce, istnieją *lokalne urzędy statystyczne*, głównie miejskie, funkcjonujące przy zarządach miast pod postacią wydziałów lub miejskich biur statystycznych. Można by je nazwać niższemi urzędami statystycznymi.

Cechą charakterystyczną dla wieku XIX jest wzrost miast, zwłaszcza wielkich. Zagadnienia społeczne, związane z życiem tych miast, wymagają obserwacji statystycznej, racjonalnie prowadzonej, która by dała możność oprzeć umiejętność zarządzania miastem na podstawach faktycznych. Pierwsze biuro statystyczne miejskie powstało na terenie Europy w Paryżu w r. 1829, nielicząc pierwszych prób rejestracji ruchu ludności, datujących się jeszcze z czasów Colberta (r. 1670). Niewielkie biura statystyczne w miastach powstawały w drugiej połowie ubiegłego stulecia. W Polsce posiadamy biura w Warszawie, Łodzi, Krakowie, Lwowie, Poznaniu, Lublinie i w kilku innych miastach.

Zadania miejskich biur statystycznych są bardzo różnorodne. Dotyczą one przejawów życia fizycznego ludności (urodzenia, małżeństwa, zgony), życia gospodarczego (dowóz produktów spożywczych, ceny), urządzeń miejskich (rzeźnie, elektrownie, gazownie), mieszkań, kultury (teatry, czytelnie, widowiska), pożarów, wypadków, spostrzeżeń nad życiem przyrody (notowania meteorologiczne) i t. d. Duże zasługi położyły miejskie

biura statystyczne w sprawie badań nad przyczynami zgonów. Badania te powstały we Francji w r. 1855. Wobec trudności ujednostajnienia nazw chorób *Bertillon* zaproponował międzynarodową nomenklaturę, obejmującą przy pełnem rozwinięciu 161 nazw chorób, przy średnim — 99 nazw chorób, przy skróconem — 44 nazwy. Lubelski miesięcznik statystyczny uwzględnia 38 przyczyn zgonów.

Niższe urzędy statystyczne terytorjalne (miejskie, wojewódzkie i t. p.), z pośród których w Polsce istnieją tylko miejskie, winny mieć na celu według dr. *Horowicza*: 1) organizację służby statystycznej wykonawczej przy operacjach statystycznych ogólnopństwowych, 2) kontrolę nad rejestracją statystyczną w urzędach państwowych na swoim terytorjum, 3) sporządzanie opracowań statystycznych dla celów administracji ogólnej, pracy społecznej, samorządów, jak również na prośbę organizacji społecznych, 4) dochodzenia statystyczne specjalne, dotyczące swego terytorjum i t. d. w porozumieniu i pod kontrolą Główn. Urz. Statyst.

Instytucje statystyki międzynarodowej.

Równocześnie z szybkim rozwojem statystyki w XIX stuleciu rychło zrozumiano potrzebę ujednostajnienia pracy statystycznej na terenie międzynarodowym. Zaczęto dokładnie zdawać sobie sprawę z korzyści, wypływających z porównania wyników operacji statystycznych w różnych krajach tak dla teorii, jak i praktyki statystycznej. Cel powyższy może być osiągnięty, jeżeli nastąpi: a) zbieranie jednakowych wiadomości, b) drogą jednakowych metod, c) opracowanie i publikowanie materiałów według jednakowych schematów.

Pierwszym, który zwrócił uwagę na potrzebę porównania międzynarodowego na podstawie cyfr statystycznych był Quetelet po zwiedzeniu wystawy wszechświatowej w Londynie w r. 1851, kiedy zgromadzone materiały statystyczne z różnych krajów nie dały się porównać wskutek braku jednolitości w zbieraniu i publikowaniu danych. On to zaproponował zwo-

lanie przedstawicieli biur statystycznych różnych państw na kongres statystyczny dla omówienia zasad uzgodnienia powszechnych operacyj statystycznych. Pierwszy *kongres* odbył się w Brukseli w r. 1853, na który przybyło 170 osób. Następny paryski odbył się w r. 1855, wiedeński — w r. 1857, londyński w r. 1860, berliński — w r. 1860, florencki — w r. 1867, haski — w r. 1869, petersburski — w r. 1872 i budapestzteński w r. 1876. Ostatni projektowany w r. 1880 już nie odbył się.

Zagadnienia, poruszane na Kongresach, należą do wszystkich działów statystyki. Można śmiało powiedzieć, że nie było kwestji statystycznej, poczynając od statystyki spisów ludności, rolnictwa, przemysłu etc., a kończąc na statystyce literatury, meteorologii, hydrografji i t. p., któraby nie podlegała obradom owych 9-ciu sesyj Kongresu. I w tem leży wielkie znaczenie owych międzynarodowych zjazdów, którymi statystyka, można powiedzieć, wyprzedziła inne działy nauki. Porozumiewanie się statystyków między sobą w zakresie najżywoźniejszych zagadnień statystyki pod kątem widzenia potrzeby uzgodnienia terminologii, metod zbierania, opracowywania i publikowania materiałów statystycznych — nie mogło pozostać bez wpływu na rozwój nauki i praktyki, aczkolwiek Sesje Kongresu międzynarodowego nie pozbawione były wad organizacji, które wreszcie zniszczyły piękną instytucję współżycia narodów na polu naukowem. Do tych wad należały: przeludnienie (udział brało przeważnie od 300 do 700 osób), przewaga teoretyków i nieobowiązkowość uchwał. Państwa przeważnie nie stosowały się do uchwał Kongresu, jako nieraz zbyt oderwanych od życia i trudnych do urzeczywistnienia, wreszcie zubożniały do spraw Kongresu, na skutek czego żywot jego został przerwany, czemu nie zapobiegło utworzenie nawet „komisji stałej”, wyłonionej na kongresie petersburskim w celu czuwania nad wypełnieniem treści uchwał Kongresu.

Jednakże idea statystyki międzynarodowej nie została ostatecznie pogrzebana; odrodziła się w następnych latach w kongresach higjeny i demografji, które w sekcji swej demo-

graficznej rozstrząsały zagadnienia statystyki ludności, a następnie w powołanym do życia z inicjatywy Say i Neymanna-Spallarta *Międzynarodowym Instytucie Statystycznym*. Pierwsza sesja Instytutu odbyła się w Rzymie w r. 1887. Organizacja Instytutu skierowana jest na rozwijanie nauki i praktyki statystycznej, głównie w zakresie statystyki administracyjnej, za pomocą następujących środków, istotnie pierwszorzędного znaczenia: wprowadzania jednolitości w metodach zbierania i opracowania danych statystycznych, zwracania uwagi państw na konieczność reform w statystyce, utrzymywania stałych stosunków w pracach statystycznych na terenie międzynarodowym i popularyzowania w społeczeństwach idei statystyki. Skład Instytutu tworzy grono 150 osób z pośród uczonych i statystyków administracyjnych całego świata. Na czele stoi biuro z prezydentem, dwoma zastępcami, skarbnikiem i jeneralnym sekretarzem, przytem siedziba Instytutu znajduje się tam, gdzie zamieszkuje sekretarz jeneralny, którym przez długi czas był Bodio (Rzym). Pierwszym prezydentem został wybrany kierownik statystyki państwa austro-węgierskiego (do r. 1908) słynny Inama-Sternegg. Prace drukowane są we własnem wydawnictwie p. t. „Bulletin de l'Institut international de statistique”.

Na gruncie tej samej idei międzynarodowej nieco później bo dopiero w r. 1905, z inicjatywy Lubin i przy czynnem poparciu króla włoskiego powstał *Międzynarodowy Instytut Rolniczy* z siedzibą w Rzymie. Zadanie jego polega na zbieraniu i publikowaniu danych statystycznych o stanie rolnictwa wszechświatowego. Na ostatniej sesji Instytutu w roku 1928 uchwalono dokonanie powszechnego spisu rolniczego we wszystkich krajach.

W r. 1875 w Bernie szwajcarskiem powstało biuro wszechświatowego związku pocztowego, które ma na celu rejestrowanie *ruchu pocztowego*. Tamże od r. 1868 mieści się międzynarodowe biuro związku *telegrafów i telefonów*, uwzględniające statystykę ruchu zawodowego.

Od r. 1901 istnieje w Bazylei międzynarodowe biuro *pracy*, interesujące się specjalnie ilością bezrobotnych we wszystkich

krajach. Dziś statystyka pracy rozwinęła się w obszerną gałąź wiedzy, wymagającą oddzielnego traktowania.

Z chwilą powstania po wojnie europejskiej *Ligi Narodów* utworzona została przy niej, między innymi, sekcja ekonomiczna, która prowadzi w specjalnym zakresie statystykę ruchu pocztowego, telegraficznego, handlu zagranicznego etc.

IV. Grupowanie i techniczne opracowanie materiału statystycznego.

Po dokonaniem dochodzenia statystycznym w wyniku zastosowania tej lub innej metody materiał statystyczny pod postacią arkuszy spisowych, kart rejestracyjnych, wypełnionych kwestionariuszy etc. nagromadza się w biurze statystycznym nieraz w olbrzymiej ilości. Tak, na spis przeprowadzony w r. 1891 w Indiach Wschodnich zużyto 290.000 kg. formularzy spisowych, z czego przeważająca część powróciła po wypełnieniu do biur statystycznych. Nasz spis z r. 1921 przyniósł ok. 6.000.000 arkuszy spisowych.

Z chwilą, kiedy materiał statystyczny dotrze do biura i zalegnie w postaci mniejszej lub większej masy zapisanego papieru, pierwszym zadaniem statystyka, zanim przystąpi do opracowania, jest ocenienie stopnia wiarygodności uzyskanego materiału, co nazywamy krytyką.

Krytyka i kontrola materiału statystycznego

Materiał statystyczny winien być wiarygodny, kompletny i należycie wypełniony.

Braki pod względem wiarygodności powstają głównie z tych przyczyn, o których wspominaliśmy przy spisach ludności, a mianowicie wówczas, gdy ludność nie ma dostatecznego zaufania do przeprowadzanego badania statystycznego na skutek obaw lub podejrzeń wytwarzających pesymistyczny stosunek do czynności rejestrantów. Często także pochodzi z tendencyjności przeprowadzanych badań (np., zaliczenie uni-

tów podczas rosyjskiego spisu ludności z r. 1897 do wyznania prawosławnego). Zanim więc przystąpimy do opracowania, winniśmy dokładnie zdać sobie sprawę z okoliczności, w jakich odbywało się dochodzenie statystyczne, t. j. czy były zachowane warunki, mogące zapewnić wiarogodność materiałowi statystycznemu.

Brak ścisłości jest powodowany niejasnem lub niewłaściwem formułowaniem pytań; jak słusznie podkreśla Schwittau, nawet jednakowe słowa mają niejednakową treść w statystyce, tem bardziej należy dbać o to, ażeby pytania były należycie rozumiane. Przeglądanie arkuszy spisowych musi wyrobić pojęcie, czy formularze są właściwie, z dostateczną ścisłością wypełnione.

Wreszcie zostaje kwestja kompletności, czyli zupełności materiału statystycznego. Należy pilnie baczyć, czy dane dochodzenie statystyczne objęło wszystkie jednostki obserwacyjne. Jeżeli nie, co winna wykazać kontrola, mamy do czynienia z niekompletnością, którą pojęciowo dzielimy na *ekstensywną* i *intensywną*. Ekstensywna niekompletność ma miejsce wówczas, gdy liczba dokonanych obserwacji jest niepełna, gdy są luki w zebranych materiale pod względem ilościowym. Intensywna niekompletność tworzy się dzięki lukom w samych arkuszach spisowych, bądź przez brak ścisłości i akuratności przy ich wypełnieniu, bądź przez opuszczenie w arkuszach spisowych niektórych, niezbędnych z punktu widzenia opisu statystycznego, cech.

Oba braki dają się usunąć przez dodatkowe badania, lub zupełnie nowe, powtórnie wykonane.

Czasami mamy do czynienia ze zjawiskiem wręcz odwrotnem — z *nadkompletnością* materiału wskutek dwukrotnie lub więcej przeprowadzonej rejestracji jednego i tego samego obiektu. Znajdujemy o tem wzmiankę w wydawnictwie Gł. Urzędu Statyst. „Wielka własność rolna” w odniesieniu do niektórych majątków, zapisanych — raz w miejscu, gdzie one się znajdowały, i drugi raz — w miejscu, gdzie zamieszkiwał właściciel. W wielkiem badaniu statystycznym, dokonywanem

w okresie wojny nad polską ludnością wygnańczą, rozproszoną po całym terytorjum Rosji, bardzo często spotykaliśmy się ze zjawiskiem dwukrotnego i więcej rejestrowania wygnańców ze względu na zmianę miejsca zamieszkania przez nich. Takie dublety należało w następstwie z ogólnego materiału wyławiać.

Jasnym się staje z powyższego, że surowy materiał statystyczny z chwilą, kiedy znajdzie się w biurze statystycznym, należy poddać kontroli, gdyż według obrazowego powiedzenia Mayr'a jest on podobny w pierwotnym stadium do kupy kamieni, z których później ma powstać estetyczny gmach.

Za przykładem G. Mayra zadajemy sobie pytanie: jakie znaczenie może mieć kontrola przeprowadzana nad materiałem zbieranym w biurze statystycznym? Przykład poniższy z ilustruje to w należyтым stopniu. Przeprowadzając metodą rejestracji bieżącej badania nad ludnością wygnańczą w Rosji czasu wielkiej wojny*), sprawdzaliśmy każdą nadesłaną kartę rejestracyjną, odsyłając rejestrantom karty źle wypełnione. W miarę jak praca podobnego rodzaju postępowała naprzód liczba zdyskwalifikowanych kart znakomicie zmniejszała się. I tak:

z 1-go tysiąca nadesłanych kart	odesłano	złych	322 t.j.	32.2%
" 2-go	"	"	218	" 21.8,"
" 3-go	"	"	183	" 18.3,"
" 4-go	"	"	134	" 13.4,"
" 5-go	"	"	100	" 10.0,"
" 6-go	"	"	23	" 2.3,"
" 10-go	"	"	19	" 1.9,"
" 12-go	"	"	6	" 0.6,"

Na odcinku badania, o którym wspominamy, ogółem z liczby 21267 wypełnionych kart rejestracyjnych odrzuciliśmy 6.2% do czasu otrzymywania materiału bez zarzutu.

Ilustracja powyższa zarazem mówi o tem, że gdzie tylko warunki spostrzegania pozwalają na to, należy bezwzględnie dążyć do przeprowadzenia ścisłej kontroli przy zachowaniu warunków: 1) ażeby kontrola przeprowadzana była możliwie

*) autor był kierownikiem Biura Statystycznego dla badań nad ludnością wygnańczą w Petersburgu.

szybko i 2) przez jedną osobę. Co się tyczy tego ostatniego warunku, to wymaganie to tłumaczy się przez fakt różnorodnego pojmowania przez poszczególnych ludzi nawet najprostrzych wypadków, a wiadomo, że każda rejestracja nastęrcza wiele powodów do rozbieżności. Bardziej jest przeto celowem, gdy materiał statystyczny surowy jest sądzony pod jednym kątem widzenia, o ile tylko jest to fizycznie wykonalne.

Grupowanie.

Segregowanie materiału statystycznego rozpoczynamy zazwyczaj od grupowania według momentu *przestrzeni*. Przy terytorjalnych badaniach materiał nadsyłany bywa już w gotowym podziale według jednostek administracyjnych. W niektórych badaniach, jak naprz. rolniczych, grupowanie materiału według geograficznego rozmieszczenia jest szczególnie niezbędne ze względu na to, że szereg cech, dotyczących życia rolniczego, jest specyficznie związanych z czynnikami miejsca. Dlatego moment geograficzny uważamy za generalny moment grupowania. Jest godnem zaznaczenia, że jednostki geograficzne winny być możliwie równe dla osiągnięcia możności porównania statystycznego, ale przy opracowaniach tylko niekiedy się to udaje, ponieważ najczęściej zmuszeni jesteśmy stosować się do podziału administracyjnego, (według województw, powiatów, gmin), który pospolicie nie odpowiada temu warunkowi.

Drugim momentem grupowania, bardzo dogodnym ze względu na możność idealnej równopodzielności, jest *czas*. Grupowanie według czasu (lat, miesiące, tygodni etc.) jest charakterystycznym dla tych badań statystycznych, które mają na celu stwierdzenie zjawisk w ich ruchu dynamicznym. Takimi zjawiskami są przedewszystkiem zjawiska ruchu ludności (urodzenia, małżeństwa, zgony), przestępstwa, wypadki. ceny i t. d.

Dalsze segregowanie materiału statystycznego polega na łączeniu w grupy według poszczególnych cech lub kombinacji cech. Naprz., ludność można podzielić według płci, wieku, wyznania, narodowości, zawodów i t. d. Gospodarstwa miejskie można grupować według obszarów rolnych, liczby inwentarzy, pól, dochodowości i t. p. Łączenie w grupy według kombi-

nacji cech zazwyczaj ma na celu stwierdzenie zależności między poszczególnymi zjawiskami, czyli odkrycie przyczynowości.

Technika opracowania.

Materiał statystyczny dowolnie pogrupowany przy kolejnym stadium pracy statystycznej należy obliczyć.

Liczenie jest zasadniczą funkcją czynności statystycznych. Przystępujemy do niego wybierając jeden z następujących sposobów. Na pierwszym miejscu wymieniamy kreskowanie.

Kreskowanie polega na tem, że każda jednostka obserwacyjna, zanotowana w materiale statystycznym, otrzymuje przy obliczeniu jedną kreskę na formularzu, podzielonym na tyle rubryk, ile odmiennych jednostek obliczamy. Kreski mogą być pionowe lub poziome, czemu można nadać specjalny sens, naprz., mężczyzn notujemy kreskami pionowymi, kobiety — poziomymi, lub odwrotnie. Cztery zanotowane kreski dla łatwości podsumowania zazwyczaj pszekreślamy piątą kreską z kolei, czego aczkolwiek nie uwzględniliśmy na wzorze ze względów technicznych. Naprzykład:

FORMULARZ A.

kawaler panna	żonaty zameżna	wdowiec wdowa
	—	
28 kawalerów 34 panny	34 żonatych 31 mężatka	17 wdowców 14 wdów

FORMULARZ B.

mężczyźni				kobiety				dzieci do lat 15 włącznie			
1	2	3	4	1	2	3	4	1	2	3	4
6	2×2= = 4	3×4= = 12	8×4= = 32	—	7×2= = 14	5×3= = 15	2×4= = 8	12×1 = 12	11×2 = 22	9×3= = 22	7×4= = 28

Posługując się wzorem B, należy pod daną cyfrą umieścić liczbę arkuszy spisowych, w których dana jednostka obserwacji powtarza się tyle razy, na ile opiewa cyfra. Jeżeli w arkuszu spisowym, obejmujących rodzinę, zanotowane są 3 kobiety umieszczamy w rubryce „kobiety” jedną kreskę pod cyfrą 3, jeżeli w tymże arkuszu zanotowano 4 dzieci, umieszczamy jedną kreskę w rubryce „dzieci” pod cyfrą 4 i t. d. Następnie rezultat obliczenia na formularzu mnożymy odpowiednio przez 1, 2, 3 i 4 i sumujemy, jeżeli chodzi o liczbę osób, umieszczonych na całym formularzu. Tym sposobem obliczenie postępuje niezwykle szybko, unika się przytem nagromadzenia olbrzymiej ilości kresek, jak na formularzu A.

Najistotniejszy brak systemu kreskowania polega na tem, że każda obserwacja oznacza się przez jeden i ten sam znak, przez co łatwo umieścić go w rubryce nieodpowiedniej, zwłaszcza gdy formularz podzielony jest na kilka rubryk. Po dokonaniu zapisania znaku w niewłaściwej przedziałce niepodobna naprawić błędu, o ile się tego natychmiast nie spostrzeże. Drugi wielki brak tej metody wynika z powolności pracy. Engel zaobserwował, że system kartkowy w porównaniu z kreskowaniem daje oszczędności na czasie 53%.

Nasze doświadczenia co do ścisłości obliczeń systemem kreskowania wykazują, że dość znaczną ścisłość można osiągnąć przy zastosowaniu *kontrolującego formularza*. Oto jego wygląd.

Wygnańcy Polacy rozproszeni w Rosji.

A.	Liczba osób w rodzinie	1	2	3	4	5	6	7	8	9	Łącznie
	Liczba rodzin	I	III	IIII IIII II	IIII IIII IIII II	IIII IIII I	IIII III	III	II		
		1	6	36	68	55	48	21	16	0	251
B.	Zawody	do lat 15 włącz.			od 16 do 55 lat wł.			od 56 lat wieku			Łącznie
	rolnicy	IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII			IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII			IIII IIII IIII IIII II			121
	rzemie- ślnicy	IIII III IIII IIII			IIII IIII IIII II IIII			IIII IIII II			53
	robotnicy	IIII II IIII IIII			IIII IIII IIII IIII			IIII IIII			46
	zawody wolne	IIII IIII III			IIII IIII II			IIII I			31
	Łącznie	97			104			50			251

Taki formularz dzielimy na dwie części. W jednej części stosujemy podział masy według jednej cechy, w drugiej — według cech innych. Suma jednostek z jednej części formula-

rza winna się równać sumie jednostek z drugiej części, bowiem jest to jedna i ta sama masa rozbita na części składowe według różnych cech. Formularz taki stosowaliśmy przy obliczeniach masy wygnańczej podczas wojny w Petersburgu.

Dla przykładu podaliśmy podział tylko na 4 zawody w rzeczywistości nasz formularz w części B zawierał 11 rubryk.

Formularz w części A podaje podział masy ludnościowej na rodziny jednoosobowe, dwuosobowe, trzyosobowe i t. d. z możliwością obliczenia ogólnej liczby osób; w części B znajdujemy podział tychże osób według zawodów i grup wieku i ogólna liczba winna być ta sama. O ile ostateczne rezultaty nie zgadzają się, należy obliczenie powtórzyć.

Możność sprawdzenia wyników obliczenia ma i tę dobrą stronę, że personel statystyczny pomocniczy, nawiasem mówiąc, niezdradzający zazwyczaj entuzjazmu do prac statystycznych, szybko nagina się do wymagań ścisłości.

System kreskowania mimo swej dostępności (wymaga drukowanych formularzy i jako narzędzia ołówek) ma ograniczone pole zastosowania. Posługujemy się nim bądź, gdy badanie jest nieduże, bądź, gdy do tego zmuszają środki materialne. Wszystkie inne bowiem systemy opracowania są kosztowniejsze.

System kartkowy jest znacznie doskonalszym środkiem opracowania. Polega on na tem, że każda jednostka obserwacji otrzymuje dla siebie oddzielną kartkę, na której są odnotowane charakteryzujące ją cechy. Mamy zatem tyle kartek, ile jest jednostek spostrzegania. Dla kilku rodzajów jednostek (n.p. w badaniach ludnościowych dla mężczyzn, kobiet i dzieci do określonego wieku) można zastosować kartki w kilku kolorach w tym celu, aby każdy rodzaj jednostek był zapisany na kartach odrębnego koloru. Przy obliczeniach ustrzega to od pomyłek i zapewnia znaczne ułatwienia.

Najłatwiej zastosować system kartkowy w tym wypadku, gdy materiał surowy został zebrany na kartach indywidualnych, przystąpienie do obliczeń nie napotyka wówczas żadnych

trudności. Natomiast, gdy mamy do czynienia z kartami rodzinnymi, z wykazami zbiorowymi, listami — powstaje konieczność przerzucenia materiału statystycznego na kartki indywidualne, co pociąga za sobą pracę przygotowawczą.

Wykonanie podobnej pracy jest dość kłopotliwe i zmusza mieć na względzie: a) czy rozporządzamy dostateczną liczbą rąk roboczych, względnie środków materialnych, b) czy możemy dopełnić skrupulatnego sprawdzania prawidłowości kopjowania, bowiem doświadczenie uczy, że nawet najlepsi przepisywacze nie są pozbawieni, lekko mówiąc, roztargnienia.

Kartki *indywidualne*, na które ma być przerzucony materiał pierwotny, winny być ułożone w ten sposób, ażeby proces przepisywania był jak najbardziej ułatwiony.

Nazwisko								№					
Imię													
Imię ojca													
Wiek								№					
S T A N			W Y K S Z T A Ł C E N I E					W Y Z N A N I E			P o d d a ń s t w o		
1	2	3	4	5	6	7	8	9	10	11	12	13	
Zawód przed opuszczeniem kraju							Gdzie są obecnie członkowie rodziny						
			14	15	16	17	18	19	20	21	22	23	
Sposób opuszczenia kraju							Majątek rodzinny (ojca rodziny)						
Zapytyw. rodziny			24	25	26								
Całej wsi lub miasteczka			24	25	26								
Kierunek i sposób podróży							Pomoc udzielona rodzinie zapisanego						
							27	28	29	30	31	32	33
Data o przybyciu na miejsce obecnego zamieszkania							Notatka o zajęciu						

W naszych doświadczeniach posługiwaliśmy się oznaczeniami numerowymi, jak to widać z podanego wzoru. Przekreślenie numeru na drukowanej kartce wymagające dla tej czynności min. czasu, służyło za odpowiedź, gdyż każdy numer posiadał w stosunku do materiału pierwotnego umówione znaczenie. Naprz., na załączonym wzorze numer 1 oznaczał: żonaty, mężatka; numer 2 — kawaler, panna; numer 3 — wdowiec, wdowa. Zamiast więc pisać słowami „wdowiec” wystarczyło przekreślić cyfrę 3, co zabierało mniej czasu.

W porównaniu z metodą kreskowania system kartkowy ma tę przewagę, że czyni dostępnymi opracowania bardzo szczegółowe, skomplikowane, prowadzące do łączenia cech w najrozmaitszych kombinacjach, bez potrzeby ograniczenia się z góry zakreślonymi rubrykami formularza. Przytem liczenie kartkami jest bardzo proste. Materiał rozkłada się na grupy i grupki według cech i notuje się, ile w każdej grupce znajduje się jednostek. Omyłki są prawie niemożliwe, gdyż w każdej chwili daną grupę można przeliczyć na nowo.

Zalety systemu kartkowego sprowadzają się do danych następujących: a) zmniejszenie formatu i wagi pierwotnego materiału, czyli zwiększenie zdolności uruchomienia go, b) ułatwienie w manipulacjach przez zastosowanie różnobarwnych kartek, c) ekonomja czasu, resp. ekonomja na personelu roboczym, zwłaszcza t. zw. stałym, d) osiągnięcie maksymalnej ścisłości wskutek możliwości każdorazowego sprawdzania czynności nieprawidłowych i wreszcie, co najważniejsze e) dowolne kombinowanie cech.

System maszynowy. Próby zastąpienia obliczeń ręcznych maszynowymi są bardzo zrozumiałe i pożądane. Wyobrażamy sobie, że genjusz ludzki wynajdzie w przyszłości taki aparat, który pozwoli na notowanie mechaniczne, drogą systemu zegarów licznikowych, tego wszystkiego, co da się odczytać na karcie rejestracyjnej. Będzie to zapewne rodzaj aparatu Morse'ego, na którym da się wypukać mowę statystyczną.

Po raz pierwszy maszyna do liczenia była zastosowana w Ameryce w r. 1890 jako maszyna elektryczna systemu *Ho-*

lerita. Konstrukcja tej maszyny w zasadzie polega na tem, że prąd elektryczny przechodzi przez rtęć, w której zanurzone są specjalnie dziurkowane karteczki. Rtęć przenika, a z nią prąd elektryczny, tylko w tych miejscach, gdzie są zrobione otworki. Otworki zaś wycina się na karteczkach stosownie do liczonych cech danej obserwacji. Zegary elektryczne poruszane prądem dopełniają liczenia.

Nasz Główny Urząd Statystyczny przy opracowaniu pierwszego powszechnego spisu ludności r. 1921 posługiwał się maszynami systemu *Powers'a*. Maszyny te składają się z dziurkarek, segregatorów i tabulatorów. Są to więc trzy rodzaje maszyn. Przystępując do opracowania materiałów statystycznych systemem maszynowym należy naprzód wszelkie pojęcia słowne, zawarte w materiale, zastąpić liczbami - symbolami na zasadzie opracowanego klucza. W dalszym ciągu mamy już do czynienia tylko z liczbami, eliminując z pracy jakiekolwiek pojęcia słowne. Zadanie maszyny dziurkarki polega na przeniesieniu ujednolitych danych z materiału na specjalne karteczki maszynowe, określonych rozmiarów, zadrukowane poziomo i pionowo liczbami porządkowymi od 0 do 9 i podzielone na kilkanaście rubryk, opatrzonych w nagłówku nazwami, stosownie do treści materiału statystycznego. Przeniesienie odbywa się za pomocą maszynowego wybijania otworków na karcie w odpowiednich miejscach przez pracownika, odczytującego arkusz spisowy. Po wydziurkowaniu i skontrolowaniu karty maszynowe idą do opracowania na t. zw. segregatorach za pomocą wielce skomplikowanych manipulacyj. Uzupełnienie opracowania stanowi nie mniej skomplikowane tabulowanie na t. zw. tabulatorach, polegające na sumowaniu liczb, dziurkowanych na kartkach maszynowych.

Wszystkie czynności maszyn *Poweos'a* wymagają wielkiej czujności w celu usunięcia niesprawnego działania. Jeżeli ten warunek jest zachowany, kontrola jest ustawiczna i sumienna, opracowanie może być dokonane zupełnie bez błędu.

Przedwstępne opracowanie.

Przy dużych rozmiarach badania statystycznego, jak w wypadku spisów powszechnych, właściwe opracowanie materiału surowego wymaga znacznej straty czasu. A tymczasem życie domaga się odpowiedzi na palące pytania niemal natychmiast. Przyjęte jest w celu zadośćuczynienia tej potrzebie dawać do publicznego użytku opracowania tymczasowe, przedwstępne — do czasu opublikowania pełnych wyników. Tak, po dokonanim w r. 1891 wielkim spisie ludności w Indjach Wschodnich, dotyczącym 250.000.000 ludzi, przedwstępne obliczenie zostało dokonane w przeciągu 5 tygodni i w stosunku do ostatecznego rezultatu dało błąd w wielkości zaledwie 0.6%. Znakomity statystyk niemiecki Engel w celu możliwie szybkiego dojścia do przybliżonych wyników stosował ważenie materiału surowego pewnymi grupami, przyjmując na jednostkę wagi określoną ilość obiektów obserwacji.

Po dokonanim w r. 1916 spisie rolniczym Rosja opublikowała „przedwstępne wyniki”.

Główny Urząd Statystyczny, przeprowadzając spis ludności w r. 1921, w celu najrychlejszego opublikowania tymczasowych danych zastosował następującą metodę. Każdy komisarz spisowy winien był zestawić arkusz zbiorowy; na zasadzie tych arkuszy powstawały zestawienia według gmin; gminy obowiązane były przesyłać zestawienia do powiatów, te z kolei przygotowały zestawienia powiatowe, przesyłane pocztą do Warszawy, a w zakresie najważniejszego zagadnienia: liczby ludności — drogą telegraficzną. Na zasadzie telegraficznych doniesień można było podać ogólną cyfrę ludności w państwie w ciągu kilku dni od chwili przeprowadzenia spisu.

W następstwie zebrania całkowitego materiału surowego w Warszawie przystąpiono do właściwego opracowania, trwającego kilka lat.

Tablice.

Do rozklasyfikowania cyfr, otrzymanych z opracowania statystycznego, służą tablice, które zdaniem *Meitzena* winny zawierać

niejako powtórzenie ogólnego planu opracowania. Służą one zarazem do odczytywania materiału statystycznego, wymagając umiejętnego układu i pomysłowości. Źle zbudowana tablica jest jednocześnie złem użyciem materiału statystycznego.

Rozróżniamy tablice zwykłe i skojarzone (kombinacyjne). Pierwsze powstają wówczas, gdy w rubrykach tablicy umieszczamy wyrazy ilościowe pojedynczych cech; drugie — gdy kojarzymy cechy między sobą. W pierwszym wypadku otrzymujemy w tablicy tyle rubryk, ile cech posiada przedmiot, suma rubryk równa się zatem sumie cech. W drugim wypadku suma rubryk równa się iloczynowi ilości cech. Naprz., jeżeli w tablicy zwykłej z dziedziny demografii dane o płci dzielimy na 2 grupy, dane o wieku — na 6, dane o stanie rodzinnym — na 4 i dane wyznaniowe — na 3, to suma rubryk równa się 15 (czyli $2 + 6 + 4 + 3 = 15$).

Ludność gminy $x = 1000$ osób.

płeć		w i e k						stan rodzinny				wyznanie		
mężczyzn	kobiet	0-10	11-20	21-30	31-40	41-50	51-60 i wyżej	kawaler, panna	żonaty zamężna	wdowiec wdowa	rozwiędz- ły rozwiędz-a	rz- kat.	ewang.	prawosł.
498	502	241	202	161	127	103	166	635	301	64	—	893	15	92

Gdybyśmy te same dane chcieli rozklasyfikować w tablicy skojarzonej, to suma rubryk w konsekwencji wyniosłaby 144 (czyli $2 \times 6 \times 4 \times 3 = 144$). Mielibyśmy wówczas kombinacje płci z wiekiem, płci ze stanem rodzinnym, wieku ze stanem rodzinnym etc.

Na schemacie podanym na stronie następnej, widzimy 4 stopnie kojarzenia cech.

Pierwszy stopień uwzględnia 3 cechy: rzymskich katolików, ewangelików, prawosławnych. Suma rubryk 3.

TABLICA SKOJARZONA.

r z y m s k o - k a t o l i c y											ewan- gelicy
kawaler, panna						żonaty, zamężna	wdowiec, wdowa				
wiek											
0-10	11-20	21-30	31-40	41-50	51-60 i wyżej						
mężczyźni	kobiety	mężczyźni	kobiety	mężczyźni	kobiety	mężczyźni	kobiety				

Stopień drugi uwzględnia podział każdej poprzedniej cechy na 4 nowe: kawaler (panna), żonaty (zamężna), wdowiec (wdowa) i rozwiedziony (a), niewykazane na schemacie ze względu na brak osób powyższej kategorii w gminie X. Suma rubryk $12 \times 6 = 72$.

Wreszcie czwarty stopień uwzględnia podział według płci. Suma rubryk $72 \times 2 = 144$. Czyli suma rubryk, jakieśmy to wyżej zaznaczyli, równa się iloczynowi ilości cech.

Słusznie zauważa prof. A. Kaufman, że kojarzenie cech w tablicy nie powinno iść dalej 4-go stopnia, ponieważ technika statystyczna nie jest w stanie uporać się z setkami i tysiącami rubryk, do których trafiłaby zaledwie nikła część notowanych jednostek, usuwająca możliwość przejawienia się wpływu prawa liczb wielkich.

Z pomiędzy cech, które mają być umieszczone w tablicy, należy wybierać najważniejsze. Rzadko się zdarza, ażeby opracowanie statystyczne wyzyskało do najdrobniejszych szczegółów materiał statystyczny, dlatego poprzestajemy na rzeczach naj-

istotniejszych, w każdym wypadku przedewszystkiem zależnie od celu, do którego dąży opracowanie.

Tablice zawierają nagłówki słowne poziome i pionowe, z których pierwsze umieszczone są z boku tablicy, drugie — u góry. Co umieszczać w każdym danym wypadku w linii poziomej, a co w pionowej — rozstrzyga indywidualne zrozumienie statystyka.

Najogólniej biorąc, linje poziome odpowiadają na pytanie: *kto i co?*, czyli są podmiotowe, natomiast pionowe odpowiadają — *jaka cecha?*, czyli są przedmiotowe. W liniach poziomych umieszczamy zazwyczaj nagłówki, zawierające nazwiska osób (w spisach), nazwy miejscowości, powiatów i t. p. W pionowych — stan rodzinny, wiek, zawód, wyznanie etc., czyli charakteryzujemy szczegółowo osobę, przedmiot, teren, umieszczone w linii poziomej (o podmiocie wydajemy to lub inne orzeczenie).

Tablica statystyczna winna być przejrzysta, łatwa do czytania. Jeżeli ułożenie uzyskanego materiału statystycznego w jedną główną tablicę jest niemożliwe z racji przytoczonych wyżej zastrzeżeń Kaufmana, należy tablicę kombinacyjną rozbić na 2, 3 lub więcej prostszych tablic.

Godzi się pamiętać, że tylko w najprostszych wypadkach tablica statystyczna może pozostać bez objaśnienia słownego. Z reguły każda tablica winna być komentowana z podkreśleniem tego, co jest w niej najistotniejszego.

V. Najprostsze wyobrażenia graficzne.

W historii statystyki był okres, kiedy uważano za wystarczające dla pracy statystycznej ułożenie materiału w odpowiednie tablice statystyczne bez podania opisu i jakichkolwiek rysunków graficznych. Dziś uważamy za warunek konieczny poprawnej pracy statystycznej, aby tablice były objaśnione, czyli omówione, i oprócz tego opracowane graficznie. Nietylko w pracach popularnych, w których *zobrazowanie* materiału statystycznego siłą rzeczy odgrywa rolę pierwszorzędną, a zrozumiałość i przejrzystość stanowi główny warunek celowości pracy, ale opracowanie graficzne należy do bardzo ważnych czynności statystycznych także w pracach naukowych. Wyobrażenia graficzne mają na celu nietylko poglądowe przedstawienie stosunku między liczbami, ale także stosunków pod względem geograficznym wówczas, gdy statystycznie opracowujemy mapę geograficzną dla wykazania związku między terenem a faktami statystycznymi. Zastosowanie wyobrażeń graficznych wpływa pozatem jeszcze z t. zw. wstrętu szerokich mas publiczności do suchych cyfr statystycznych, wynikającego z powszechnego braku umiejętności „czytania” tablic statystycznych. Istotnie, tekst dołączony do tablic, wzory graficzne — są wielką pomocą w zrozumieniu treści cyfr statystycznych, pozwalają szybko zorientować się w charakterze zależności między zjawiskami, ilustrują stosunki liczbowe bardzo dosadnie.

Jednak poglądowe przedstawienie wyników opracowania statystycznego nie jest wyłącznym celem metody graficznej. Jej cele są głębsze: umysł człowieka nie ogranicza się prostym unaocznieniem stanu faktycznego zjawisk, dąży on do zrozu-

mienia związku między zjawiskami i ustalenia ich prawidłowości, do wykrycia przyczynowej zależności jednych zjawisk od drugich. W tej złożonej pracy badawczej *analiza graficzna* jest ważnym środkiem poznania.

Zasadniczo odróżniamy wyobrażenia graficzne, które mają na celu uplastycznienie stosunku między wielkościami statystycznymi w formie punktów, linii, figur, płaszczyzn i t. d. i te nazywamy djagramami lub wykresami i także wyobrażenia geometryczne w połączeniu z mapą geograficzną, które nazywamy kartogramami.

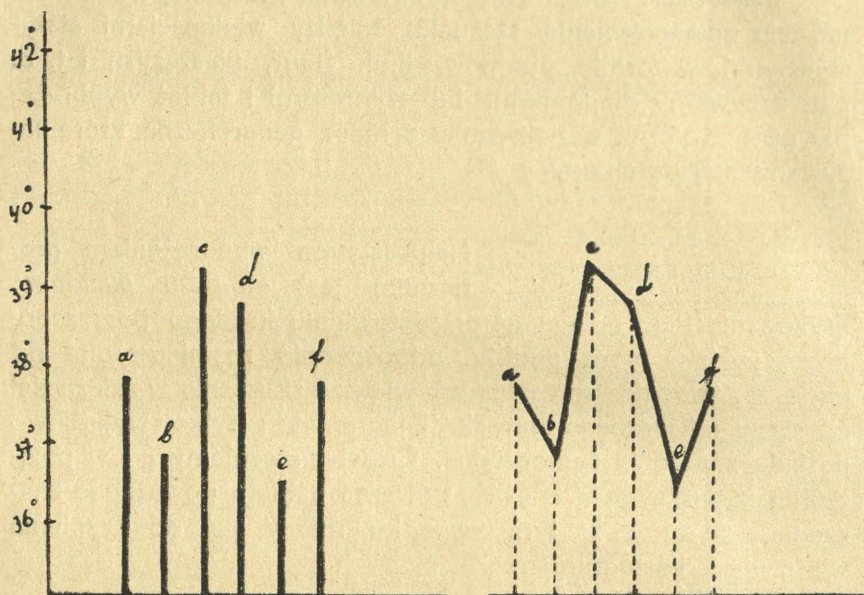
Djagramy.

Najprostszym wyobrażeniem graficznym jest *djagram punktowy*. Wykonanie jego polega na przedstawieniu każdego poszczególnego zjawiska przez punkty, odpowiednio rozmieszczone na papierze, tworzące mniejsze lub większe skupienia w zależności od stanu faktycznego zjawisk. Djagram ten tylko wyjątkowo posiada znaczenie samodzielne. Częściej stosujemy go w połączeniu z siatką geograficzną, wówczas nabiera poważnego znaczenia, wszakże w tym wypadku mamy już do czynienia właściwie z kartogramem.

Powszechnie stosowanym w opracowaniach statystycznych jest *djagram linjowy*. Nietylko statystyka, ale i szereg innych nauk, jak medycyna, biologia, meteorologia i t. p. posilkuja się z pożytkiem tym djagramem. W najogólniejszej formie djagram linjowy jest to linja, odpowiadająca w sposób warunkowy wielkości liczby statystycznej. Ponieważ metoda graficzna w większości wypadków dąży do przedstawienia porównania między liczbami, geometryczne wyobrażenia liczb podaje ona najczęściej w postaci kilku lub kilkunastu linii poziomych lub pionowych, sprowadzonych do wspólnej osi. Wspólnej osi nadajemy przeważnie położenie poziome, linjom — pionowe, łącząc je w punktach górnych linją łamaną, nazywaną pospolicie „krzywą”, aczkolwiek o krzywej w matematycznym znaczeniu słowa niema tu mowy.

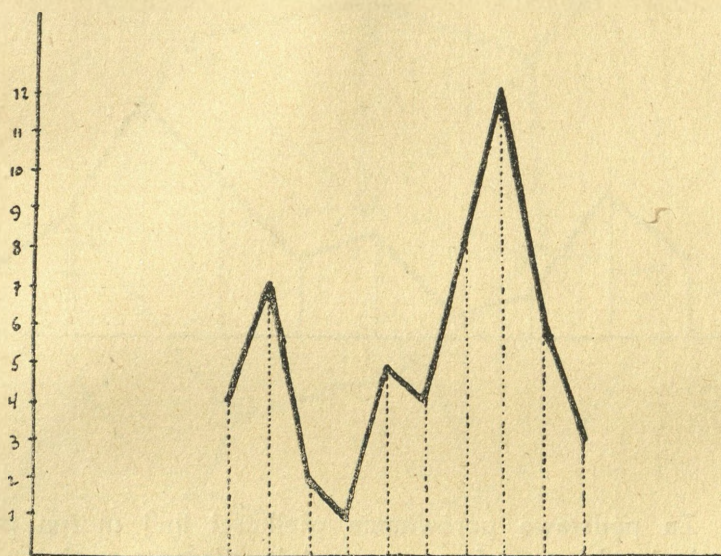
PRZEBIEG TEMPERATURY CHOREGO.

Djagram *linjowy*.



Brak wewnętrznej zależności między szerokością podstawy a wysokością pionów w djagramach linjowych powoduje dowolność wrażenia wzrokowego w tym wypadku, gdy dla jednych i tych samych wartości liczbowych obieramy różne podstawy, lub różnej wielkości piony. W zależności od obranej podstawy i wielkości pionów djagram może robić wrażenie ostrokątnego lub płaskiego. Powiedzmy, że mamy liczby szeregu 4, 7, 2, 1, 5, 4, 8, 12, 6, 3 — które chcemy wyobrazić graficznie, przyjmując dowolną podstawę.

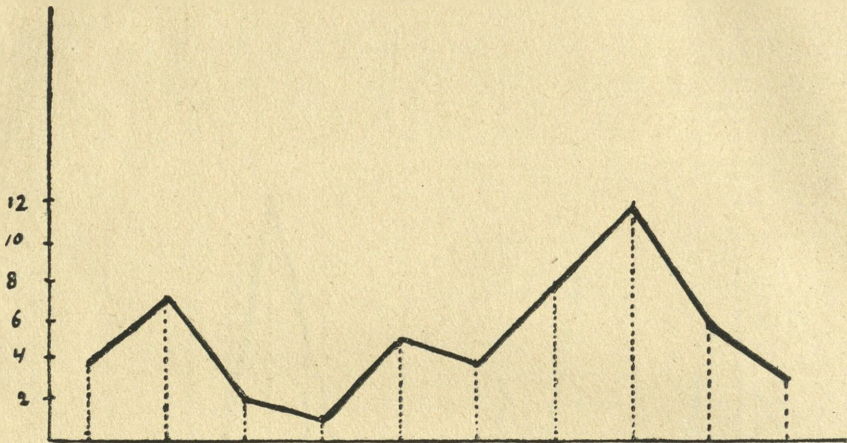
DIAGRAM LINJOWY
stromy.



Jeżeli w powyższym ddiagramie obierzemy podstawę dwa razy większą, a wysokości pionów dwa razy mniejsze, to otrzymamy ddiagram płaski, wywierający zupełnie inne wrażenie wzrokowe. Brak zależności między szerokością podstawy a wysokością pionów stanowi cechę ujemną ddiagramów linjowych.

DIAGRAM LINJOWY

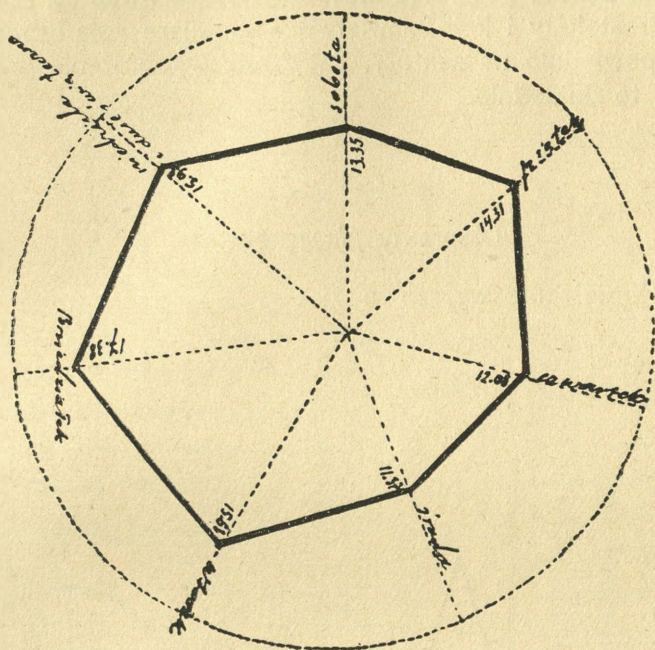
plaski.



Za podstawę porównania wielkości linii można przyjąć promień okręgu, wówczas powstaje *djagram w okręgu*, bardzo pożyteczny wówczas, gdy chodzi o wyobrażenie faktów, zachodzących w równych odstępach czasu. Fakty w tygodniu, zdarzenia w ciągu roku według kolejnych miesięcy — doskonale nadają się do ilustracji podobnym djagramem. Gdy mamy do przedstawienia śmiertelność według miesięcy roku, odmierzymy ilość zgonów na 12 promieniach okręgu; w tym wypadku miesiąc styczeń wyobrażony jest na djagramie równie blisko miesiąca lutego jak i grudnia, co daje miarę porównania dla sąsiednich okresów czasu. Tak samo blisko siebie będą położone dni w tygodniu, o ile naprz. ilość wypadków w tygodniu wyobrazimy na djagramie okręgowym.

Przykład.*)

Nieszczęśliwe wypadki w Monachjum w l. 1894 i 1895.



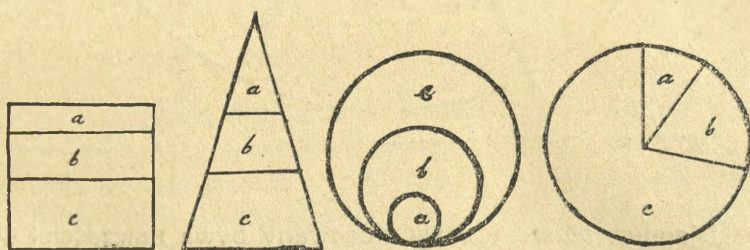
Dajagram *plaszczyznowy* stosowany bywa najczęściej wówczas, gdy mamy zamiar wyobrazić całość podzieloną na te lub inne części (naprz. masę socjalną na poszczególne grupy socjalne). Zasadnicza forma tego dajagramu jest to kwadrat, podzielony na prostokąty. Stosujemy jednak także trójkąty, koła i t. p. Niektóre zjawiska nadają się bardziej do wyobrażenia przez ten, a nie inny dajagram. Podział ludności według

*) Dr. G. Schnapper - Arndt. *Sozialstatistik*. Leipzig 1908, s. 563.

stopni wykształcenia (wyższe, średnie, niższe) może być bardzo dobrze wyobrażony przez trójkąt, którego szczyt zajmuje wyższe wykształcenie, środek — średnie, a podstawa — powszechne wykształcenie, będące istotnie podstawą oświecenia ludności. Ale także dobrze powyższe stosunki wyraża kwadrat, podzielony na prostokąty i koło, zawierające mniejsze koła lub podzielone odpowiednio na sektory. Załączone wyobrażenia graficzne ilustrują to dokładnie.

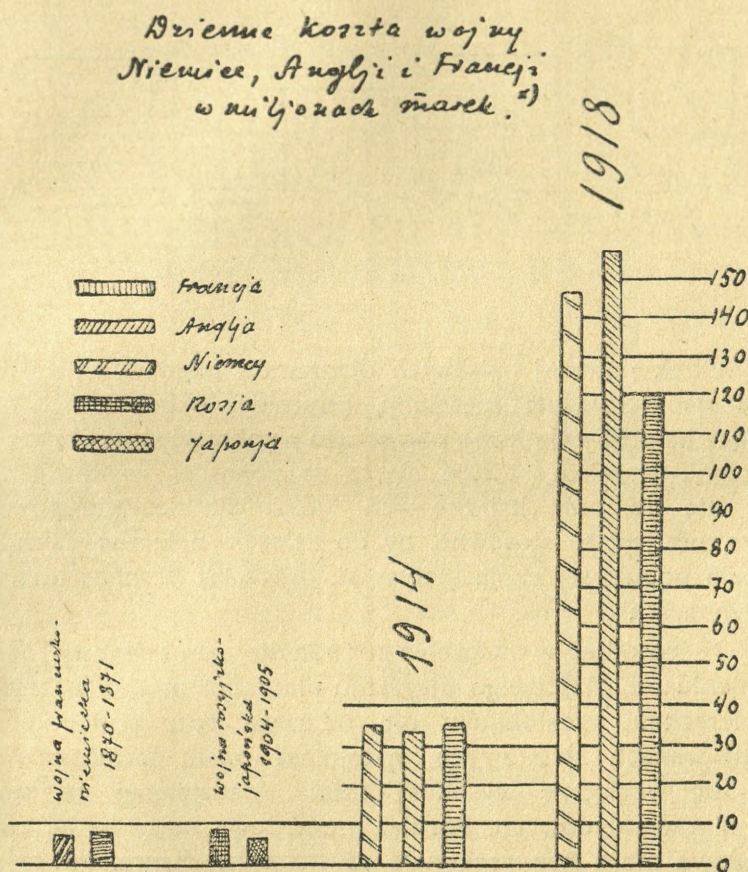
Diagramy płaszczyznowe.

Wykształcenie wyższe a
 „ średnie b
 „ niższe c



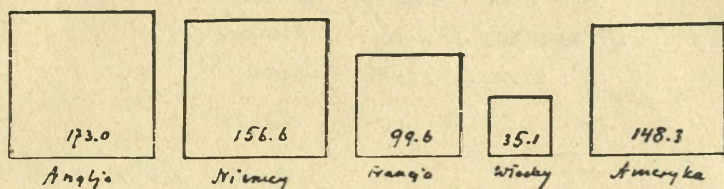
Przy posilkowaniu się diagramami płaszczyznowymi w formie kwadratów należy uwzględnić właściwość, że boki kwadratów, które mają wyobrażać liczby, są proporcjonalne nie do liczb samych, lecz do ich pierwiastków kwadratowych.

Zarówno w dajagramach linjowych, jak i płaszczyznowych, w celu osiągnięcia większej poglądowości stosujemy kolorowanie lub kreskowanie zależnie od naszej pomysłowości lub artystycznego smaku. Zamiast dajagramów linjowych posiłkujemy się b. często dajagramami słupkowemi, które w tym wypadku kreskujemy lub kolorujemy.



*) Die Welt In Zahlen v. Wl. Woytinsky. Sechstes Buch, Tafel 8, 1927. W oryginalne dajagram utrzymany w kolorach.

Diagram kosztów wojennych w złocie
Anglii, Niemiec, Francji i Stanów Zjedn. Ameryki Półn.
ku końcowi 1919 roku.*)
Liczby opiewają miljardy marek.



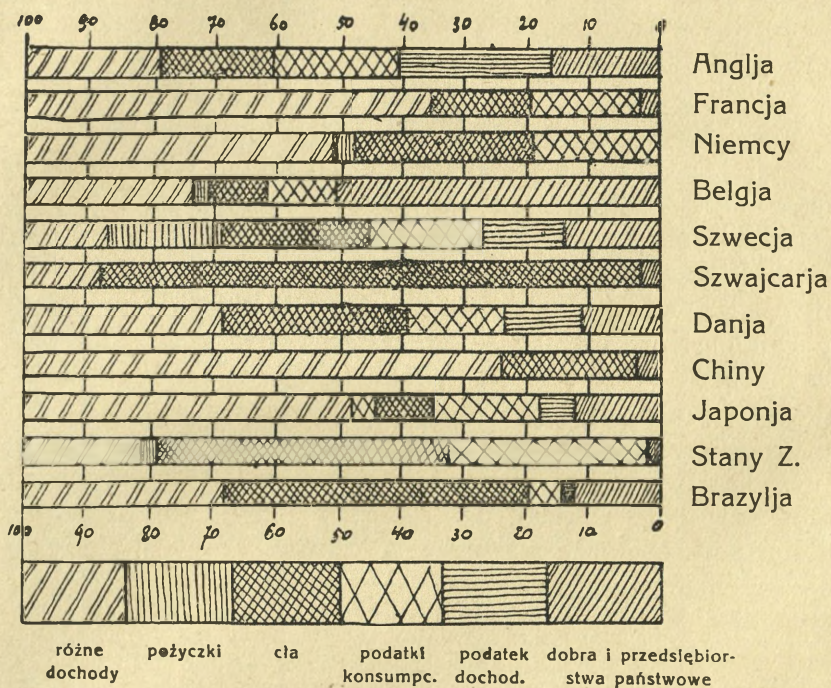
Obok powyżej opisanych djagramów niekiedy są w użyciu ciała geometryczne: sześciiany, piramidy, ostrosłupy i t. p., mające na celu wyjątkowo poglądowo wyobrazić stosunki między liczbami, wszakże stosuje się to do pokazów, wystaw i t. d., lecz nie do prac książkowych, gdyż jak złośliwie powiada Mayr, nie jest to stosowne, by do książek dołączać „skrzynki w celu popularyzowania statystyki, naprz. za pomocą kubików drewnianych”.

Nieco odrębne znaczenie posiada djagram taśmowy. W odróżnieniu od właściwego djagramu płaszczyznowego (p. wyżej), powierzchnia prostokątów nie odgrywa roli, podobnie jak w djagramach słupkowych, natomiast miarą porównania jest długość, względnie wysokość taśmy, podzielonej na szereg czynników składowych. Podział procentowy masy na jej części składowe może być doskonale zilustrowany djagramem taśmowym.

*) Ibid., tablica 8.

Przykład.

PODZIAŁ PROCENTOWY DOCHODÓW niektórych państw w r. 1913*)



Zastosowanie w pracy statystycznej wyobrażeń graficznych tego lub innego typu najczęściej podyktowane jest naturą zjawisk. Umiejętność i doświadczenie statystyka wskażą mu wybór najodpowiedniejszego djagramu. Najodpowiedniejszym djagramem będzie ten, który najtrafniej, a zarazem w sposób

*) Ibid., Tablica 3. W oryginale djagram utrzymany w kolorach.

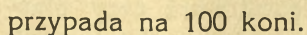
najprostszy oddaje rysunkowo wszystkie właściwości stosunków liczbowych. Dlatego zalecanie z góry jednego jakiegoś rodzaju djagramu w celu posiłkowania się nim nie jest możliwe. Można raczej mówić o urozmaiceniu djagramów, gdyż jednostajność rysunków czyni nużące wrażenie na czytelniku i osłabia zainteresowanie się danymi statystycznymi. Ale nawet najprostsze wyobrażenia graficzne dają możliwość stworzenia tylu kombinacji, że jednostajność rysunków może być uniknięta.

Kartogramy.

Wiele więcej niż tablica statystyczna zawiera materiału naukowego kartogram. Jest to mapa geograficzna, na której wyobrażone są stosunki liczbowe. Ze względu na to, że kartogram daje obraz związku stosunków liczbowych z warunkami miejsca (obszaru, gleby, klimatu, zaludnienia i t. p.) jeszcze bardziej, niż djagram, ułatwia uchwycenie zależności między zjawiskami. Wyobrażenia geograficzno-statystyczne, o ile materiał nadaje się do tego, winny być zawsze stosowane, jako cenna metoda opracowania, dająca poważne rezultaty. Zadanie kartogramu polega na tym, ażeby wyświetlić stosunek między danym zjawiskiem społecznym, jako funkcją, a momentem przestrzeni, jako czynnikiem, zapomocą liczb względnych. Niekiedy stosuje się na mapach statystycznych także wyobrażenie liczb absolutnych, ale te pod względem wartości naukowo - statystycznej posiadają mniejsze znaczenie. Ułożenie kartogramu wymaga uszeregowania liczb względnych od max do min, charakteryzujących natężenie zjawiska w związku ze zmianami geograficznymi. Ponieważ nie można wszystkich wahań danych stosunków liczbowych uzmysłowić na kartogramie, czynimy uogólnienie, łącząc cyfry szeregu w grupy, możliwie nieliczne (najczęściej 5 — 10 grup). W zależności od charakteru liczb i spostrzeganego zjawiska przy tworzeniu grup postępujemy rozmaicie, wszelako zazwyczaj różnicę między max i min dzielimy przez liczbę grup (klas), którą uważamy dla danego wypadku za wystarczającą. Ale podział obszaru statystycznego może być i inny. Za podstawę podziału bierzemy średnią arytmetyczną. Do pierwszych

W celu umieszczenia klas szeregu na mapie przyjmujemy warunkowe oznaczenia klas umówionymi znakami. Temi znakami mogą być różne kolory, lub odcienie jednego koloru, jak również kreski, w tej lub innej kombinacji zastosowane.

Liczba koni w wieku 4 — 16 lat
powyżej 148 cm.*)



125

Posiłkowanie się kolorowaniem lub kreskowaniem winno mieć na względzie wywołanie wrażenia, odpowiadającego treści liczb, wyobrażanych na kartogramie.

Kartogram wykonany w jednym kolorze lub cieniowany kreskami stosownie do zmian w natężeniu zjawiska najlepiej uzmysłowi to, co pragnie osiągnąć kartografja statystyczna.

Do bardzo ważnych warunków prawidłowego ułożenia kartogramu należy to, ażeby masa statystyczna rozbita była na najmniejsze (o ile na to pozwala materiał liczbowy i logika przeprowadzanego badania) jednostki obszarowe, któremi zazwyczaj są mniejsze lub większe jednostki administracyjne. Wybór wielkości jednostki obszarowej zależy od charakteru zjawisk i konieczności mniej lub więcej szczegółowego ich wyobrażenia. W jednym wypadku kraj dzielimy na województwa, w innym na powiaty lub nawet na gminy.

Posługiwanie się podziałem administracyjnym obok dużych zalet, głównie z praktycznego punktu widzenia, posiada poważne wady ze względu na niejednorodność jednostek administracyjnych pod względem warunków naturalnych, społeczno-ekonomicznych i kulturalnych w obrębie tych samych granic.

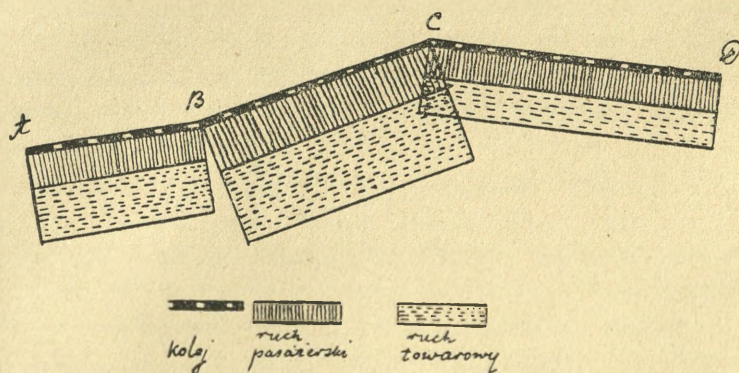
Prof. E. Romer, układając cenny swój „Atlas Polski”, ze względu na zaznaczone wady podziału administracyjnego zastosował inną metodę wykreślenia kartogramów, opartą na zasadzie interpolacji. Przyjmował mianowicie, że wartość statystyczna powiatu odnosi się do punktu, określonego przez położenie miasta powiatowego, i jeżeli pomiędzy sąsiednimi powiatami zachodziła różnica w intensywności wyobrażanego zjawiska, to przeprowadzał między nimi linje w odległości proporcjonalnej do różnic wartości obu punktów. Powstawały linje, bynajmniej nie identyczne z granicami administracyjnymi, lecz łączące punkty o równej wartości zjawisk, nazwane przez Romera *izarytmami*. Posiłkowanie się izarytmami przy układaniu kartogramów daje możność wyobrażenia zjawisk, odnośnie rozmieszczenia w przestrzeni, bardziej odpowiadającego rzeczywistości, aniżeli wówczas, gdy to rozmieszczenie stosuje się do obszarów administracyjnych.

Poruszone zagadnienia dotyczą kartogramów *właściwych*. Oprócz tego odróżniamy kartogramy punktowe, kartogramy linjowe i wstęgowe.

Kartogram punktowy jest to mapa geograficzna, na której w kształcie punktów umieszczamy spostrzegane zjawiska (np. miasta na karcie geograficznej). Jeżeli na mapie geograficznej mamy rozmieszczone szkoły, ochronki, urzędy pocztowe, fabryki i t. p. za pomocą umówionych znaków geometrycznych, to mamy przed sobą kartogram punktowy.

Kartogram linjowy są to wyobrażenia na mapie geograficznej dróg, kolei, rzek, kanałów i t. p.

Kartogram wstęgowy wyobraża najczęściej stan ruchu na liniach kolejowych (przewóz pasażerski i towarowy). W celu ułożenia kartogramu wstęgowego rysujemy wzdłuż linii kolejowej prostokąty, odpowiadające wielkości ruchu kolejowego od stacji do stacji, jak na przykładzie.



Niemna możliwości wymienić na tem miejscu wszystkich będących w użyciu kartogramów, których układanie zależy od technicznych warunków wykonania i pomysłowości opracowującego materiał statystyczny. W tych wypadkach, gdy chodzi

o szybkie ogłoszenie prac drukiem, a wykonanie kartogramów nie może opóźnić wydania, są w użyciu kartogramy *schematyczne*, polegające na tem, że zamiast ścisłego przedstawiania okręgów administracyjnych w kartogramach właściwych, umieszczamy prostokąty, pokryte odpowiednią barwą lub jakimikolwiek innemi umówionemi znakami, przygotowane w drukarni, podobnie jak czcionki. Ułożenie kartogramu odbywa się w tym wypadku w taki sam prosty sposób, jak liter w drukowanej książce.

Od najprostszych wyobrażeń graficznych należy odróżniać *krzywe* we właściwym znaczeniu słowa, które stanowią wyobrażenia funkcji matematycznych.

VI. Pojęcie wielkości statystycznych.

Opracowanie szczegółowe.

Obliczenie materiału statystycznego, ułożenie go w tablice — stanowią niezmiernie ważne czynności pracy statystycznej, ale są tylko wstępem do dalszej analizy, która nosi nazwę opracowania szczegółowego.

Opracowanie szczegółowe ma na celu zapoznanie nas z istotą rozpatrywanych zjawisk przez rozłożenie masy statystycznej — zbiorowości — na czynniki prostsze, dostępnejsze naszemu zrozumieniu. Posługujemy się w tym celu pewnymi wielkościami statystycznymi, których znaczenie wyjaśnia nam teoria metody wielkich ilości wskazująca jednocześnie sposoby ich poszukiwania. Temi wielkościami są w najogólniejszym znaczeniu słowa t. zw. liczby średnie, których znaczenie dla statystyki najdobitniej określa *Bowley*, nazywając statystykę nauką o liczbach średnich.

Pojęcie liczby średniej zgóry zakłada istnienie szeregu liczb, dla którego dana liczba będzie albo faktyczną, albo abstrakcyjną średnią. Istotnie, w każdym badaniu statystycznym mamy do czynienia z szeregami liczb. Tablice statystyczne są niczem innym, jak usystematyzowanymi, ujętymi w rubryki szeregami liczb, stanowiącymi charakterystykę ilościową, t. j. statystyczną różnych cech. Opracowanie szczegółowe polega na opracowaniu szeregów.

Szeregi.

Już najprostsze uporządkowanie szeregu według wartości liczb nosi znamiona opracowania. Jeżeli ułożymy liczby od najmniejszej

do największej, otrzymamy szereg rosnący i przy odwrotnym kierunku liczb — od max do min — szereg malejący. Dwa szeregi, porównywane ze sobą, z których każdy jest rosnącym albo malejącym równolegle, tworzą szeregi *zbieżne*. Dwa szeregi, z których jeden jest rosnącym, a drugi — malejącym, tworzą szeregi *rozbieżne*. Szeregiem *szczegółowym* nazywamy szereg, obejmujący wszystkie oznaczenia pojedyncze. Jeżeli jednak niektóre oznaczenia pojedyncze łączymy w równe grupy, notując ile w każdej grupie połączyliśmy oznaczeń, otrzymamy szereg *rozdzielczy*. Grupy połączonych oznaczeń nazywamy klasami, obszar klasy nazywamy przedziałem klasy, a ilość pojedynczych oznaczeń, wchodzących w granice klasy, nazywamy liczebnością albo frekwencją. Naprz.

Szereg *szczegółowy*.

Szereg *rozdzielczy*.
Szwajcarzy w Zarychu.
Szerokości głowy.*)

	klasy	liczebności
322	136.5 — 140.5	3
375	140.5 — 144.5	21
394	144.5 — 148.5	22
415	148.5 — 152.5	18
426	152.5 — 156.5	29
453	156.5 — 160.5	15
495	160.5 — 164.5	7
512	164.5 — 168.5	2
424	168.5 — 172.5	1

Jeżeli w szeregu rozdzielczym *Czekanowskiego* połączymy po 3 klasy w jedną, otrzymamy szereg przekształcony, bardziej przejrzysty.

klasy	liczebności
136.5— 148.5	46
148.5— 160.5	62
160.5— 172.5	10

*) Jan Czekanowski. Zarys metod statystycznych do antropologii. R. 1913, str. 12.

Ponieważ klasy, jako takie, nie nadają się do działań arytmetycznych (np. do mnożenia przez liczebności), zastępujemy ich wartość przez *wartość środkową* przedziału klasy, która jest średnią arytmetyczną granicy dolnej i górnej każdej klasy.

Opracowanie szeregów według prof. *Kabłukowa* stanowi istotną treść opracowania statystycznego, gdyż „całkowita technika opracowania ma na celu uszeregowanie zjawisk”.

Zanim przejdziemy w niniejszym rozdziale do *elementarnego* opracowania szeregów, omówimy szczegółowo znaczenie niektórych wartości statystycznych, stanowiących narzędzia tego opracowania.

Średnia arytmetyczna. Już Pitagorasowi znane były trzy wielkości matematyczne: średnia arytmetyczna, średnia geometryczna i średnia harmoniczna. Wszystkie trzy wielkości otrzymujemy z proporcji: arytmetyczną z proporcji $a - b = b - c$, geometryczną — $a : b = b : c$, harmoniczną — $a : b = a - b : b - c$, skąd odpowiednie wzory

$$b = \frac{a + c}{2}$$

$$b = \sqrt[3]{a \cdot c}$$

$$b = \frac{2 a c}{a + c}$$

Jeżeli $a = 2$ i $c = 6$, to średnia arytmetyczna $= \frac{2 + 6}{2} = 4$;

geometryczna $= \sqrt[3]{2 \cdot 6} = 3,464...$ i harmoniczna $= \frac{2(2 \cdot 6)}{2 + 6} = 3$.

Średnią arytmetyczną będziemy oznaczali przez symbol *sa*, średnią geometryczną — przez *sg* i śr. harmoniczną — przez *sh*. Zastosowanie *średniej arytmetycznej* w statystyce z tego względu nie wymaga bliższego wyjaśnienia, że jest to najpopularniejsza ze wszystkich wielkości średnich. Zastosowanie *średniej geometrycznej* może być wyjaśnione na przykładzie przyrostu naturalnego ludności. Ludność, jak wiadomo, powiększa się

w postępie geometrycznym, wobec tego i średnia dla oznaczenia przyrostu winna być geometryczną. Jeżeli ludność w ciągu 100 lat powiększyła się dwukrotnie, t. j. o 100%, to roczny przyrost stanowi nie 1%, a $\sqrt[100]{2}$, czyli 0,717, t.j. mniej więcej 0,7% w stosunku rocznym.

Zastosowanie *średniej harmonicznej* widać na przykładzie określenia wartości spadku pieniądza, co niekiedy zdarza się przy wahaniami nieustabilizowanej waluty. Przy spadku wartości pieniądza powiększa się cena towarów, jednakże w innej proporcji. Jeżeli skutek zmiany wartości pieniądza ceny towarów w pewnym okresie czasu podnoszą się w proporcji 100 , 125 , 150 ;

to ilość towarów, które można nabyć za poprzednią sumę pieniędzy, zmniejsza się w proporcji 100 , 80 , 66,66 .

W pierwszym wypadku (spadek wartości pieniądza) mamy przed sobą postęp arytmetyczny, w drugim (zmniejszenie siły kupna = ilości nabytych towarów) — harmoniczny.

Zastosowanie tej lub innej średniej zależy od charakteru badanego zjawiska, jednak największe pole zastosowania ma przed sobą średnia arytmetyczna ze względu na łatwość jej otrzymywania oraz na niewielkie różnice ilościowe, zachodzące między wymienionymi trzema rodzajami średnich w pewnych warunkach.

Średnia arytmetyczna posiada charakter uniwersalny w opracowaniach statystycznych; pomimo tego jednak według trafnego spostrzeżenia *Liésse'a* ma ona tę wadę, że zbyt brutalnie niweluje zjawiska, zacierając odcienie, istniejące pomiędzy poszczególnymi zjawiskami. Tem nie mniej dla wyrażenia jedną liczbą wartości szeregu jest ona niezastąpiona.

Już Quetelet odróżniał dwa rodzaje średniej arytmetycznej: średnią arytmetyczną obiektywną albo *typową* i średnią arytmetyczną subiektywną albo *abstrakcyjną*. Jedna powstaje w tym wypadku, kiedy mamy do czynienia z wynikami pomiarów jednego i tego samego fizycznego obiektu; druga —

kiedy pomiary obejmują nie jeden, lecz kilka, kilkanaście lub kilkadziesiąt obiektów tego samego gatunku. Jeżeli, naprz., w pracowni chemicznej ważymy na czułych wagach chemicznych jakiś związek soli, to z reguły dokonywamy ważenia kilkakrotnie w celu wypośrodkowania średniej arytmetycznej z kilku pomiarów, gdyż niedokładności naszych instrumentów, niedoskonałości naszego aparatu wzrokowego, nieprzewidziane ruchy ręki, poruszenia gmachu w skutek wstrząśnienia od ruchu ulicznego i t. p. przyczyny — wpływają na niedokładność pojedynczych pomiarów. Średnia arytmetyczna w tym wypadku będzie obiektywną.

Jeżeli mamy do czynienia z licznymi przedmiotami, naprz., pragniemy obliczyć średnią wysokość domów, stojących przy danej ulicy, to w tym wypadku czynimy pomiary każdego poszczególnego domu, pomiary sumujemy i dzielimy przez ilość domów. Ta średnia arytmetyczna będzie abstrakcyjną, ponieważ zdarzyć się może, że ani jeden dom nie będzie posiadał obliczonej wysokości. Również średni wiek ludności z obliczenia nie będzie odpowiadał wiekowi większości mieszkańców kraju. Mamy tu do czynienia z abstrakcją, z czemś nieistniejącym realnie, lecz nie mniej cennem z punktu widzenia zasady uogólnienia.

Odróżniamy dalej dwa inne rodzaje średnich arytmetycznych: średnią arytmetyczną *prostą* i średnią arytmetyczną *ważoną*. Weźmy przykład. Na targu sprzedawano ziemniaki po 8, 7 i 6 złotych za quintal. Przeciętna cena na zasadzie średniej arytmetycznej prostej równa się $8 + 7 + 6$ dzielone przez 3, t. j.: $sa = 7$ złotych. Inaczej atoli sprawa się przedstawi, jeżeli uwzględnimy ilość transakcyj, zawartych za tę lub inną cenę. Przypuśćmy, że sprzedano 1000 q. po 8 złotych za 1 q., 100 po 7 zł. i 15 po 6 zł. W takim razie średnia arytmetyczna ceny, obliczona z uwzględnieniem *wagi* poszczególnych cen (ważniejsza jest ta cena, za którą dokonano większej ilości transakcyj), czyli średnia arytmetyczna ważona równać się będzie: $8 \cdot 1000 + 7 \cdot 100 + 6 \cdot 15 = 8790$, dzielone przez ilość transakcyj $1000 + 100 + 15 = 1115$, t. j. $sa = 7,88$ zł. Jak wi-

dzimy, średnia arytmetyczna prosta jest w podobnych wypadkach obliczeń mniej dokładna, jak średnia ważona.

Wszelako częstokroć, mając możność wyprowadzenia średniej arytmetycznej ważonej, poprzestajemy na średniej zwykłej. Ma to miejsce wówczas, gdy liczba obserwacji jest niejednakowa z kilku terenów (ceny z kilku różnych targów, plony zbóż z różnych okolic i t. p.). Prof. R. *Orzęcki* opinuje w tej materji jak następuje: „wagę należy brać pod uwagę w tych wypadkach, kiedy wszystkie pojedyncze dane stosują się do jednego i tegoż zjawiska, gdy obejmują całkowity teren lub całkowity okres czasu bez wyjątku; natomiast, o ile dane zjawisko znajduje różnorodny wyraz w poszczególnych częściach terenu lub okresach czasu, to owe poszczególne wyrazy są równowartościowe z punktu widzenia ogólnej średniej, niezależnie od liczby obserwacji, na zasadzie których zostały wyprowadzone, i z tego względu składają się na średnią liczbę z jednakową wagą”.

Kardynalnym warunkiem prawidłowego zastosowania średnich jest 1) dostateczna liczba dokonanych obserwacji i 2) jednolitość liczb, wchodzących do szeregu, czyli w jednym i drugim wypadku dokładna znajomość danej masy statystycznej. Istnieje elementarna metoda przekonania się, czy oba warunki są zachowane. Co do pierwszego tworzymy z danej ilości obserwacji 2 równe grupy i o ile średnie arytmetyczne obydwóch grup są do siebie bliskie, mamy do czynienia z dostatecznie wielką liczbą obserwacji; co do drugiego — sprawdzamy, czy poszczególne człony szeregu nie odbiegają zbyt od wartości średniej; różnica nie powinna przekraczać połowy wartości średniej, w przeciwnym wypadku można takie człony z szeregu wyłączyć, jako niejednolite. Jednak, mając do czynienia z liczbami niejednolitemi lepiej jest zamiast wyłączenia stosować wyprowadzenie kilku średnich o charakterze samodzielnym z części jednego szeregu (np., zamiast ceny rocznej można podać kilka przeciętnych cen z poszczególnych okresów roku).

Średnia geometryczna. Średnia geometryczna znacznie rzadziej bywa stosowana w statystyce, aniżeli średnia arytmetyczna, jednakże są warunki, w których średnia geometryczna, jak widzieliśmy, lepiej charakteryzuje cechy zbiorowe, aniżeli ta druga. Dotyczy to w szczególności takich szeregów, w których są znaczne nierówności. Prof. *Moszczeński* przytacza obliczenie równoległe dla średniej arytmetycznej i średniej geometrycznej dwóch szeregów, z których jeden jest jednolity, drugi zaś zawiera ogromne odchylenie jednego wyrazu od pozostałych.

Szereg A	Szereg B
230	230
222	222
234	234
215	2660
Średnia aryt. = 225.25	Średnia aryt. = 836.5
średnia geom. = 225.13	„ geom. = 422.2

Jak widzimy, w szeregu B różnica pomiędzy obiema wielkościami średnich jest bardzo duża.

Podobny przykład podaje J. <i>Piekałkiewicz</i> za Bowley'em.									
80	80	20	20	w I szeregu	śr. aryt.	120,	śr. geom.	113	
60	80	20	80						
100	100	80	80	„ II	„	„	„	146,	„
324	80	100							
	120	100	„ III	„	„	„	„	73,	„
	120	120							
		160	„ IV	„	„	„	„	198,	„
		324							
		972							

W szeregu ostatnim różnice między obiema wielkościami są znaczne. Średnia geometryczna jest mniej wrażliwa na rozbieżność liczb w szeregu aniżeli średnia arytmetyczna, co stwierdzają powyższe przykłady.

Ogólny wzór dla poszukiwania średniej geometrycznej jest taki:

$$sg = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdot \dots \cdot X_n}$$

Stąd

$$\log sg = \frac{1}{n} (\log X_1 + \log X_2 + \log X_3 + \dots + \log X_n)$$

Równanie powyższe zarazem stwierdza, że logarytm średniej geometrycznej równa się średniej arytmetycznej sumy logarytmów poszczególnych wyrazów szeregu.

Srednia harmoniczna.

Od średniej geometrycznej jeszcze rzadsze zastosowanie znajduje w statystyce średnia harmoniczna. Niekiedy za jej pomocą obliczamy średnią wielkość, gdy zamiast ceny towaru mamy podaną ilość towaru, otrzymaną za jednostkę pieniądza. Prof. Moszczeński przytacza następujący przykład *Czubera*. Za 1 markę sprzedawano na targu jajek:

w styczniu	9 sztuk	w lipcu	13 sztuk
„ lutym	11 „	„ sierpniu	11 „
„ marcu	15 „	„ wrześniu	10 „
„ kwietniu	16 „	„ październ.	9 „
„ maju	15 „	„ listopadzie	6 „
„ czerwcu	14 „	„ grudniu	7 „

Razem za 12 marek sprzedano 136 jajek, czyli za 1 markę według średniej arytmetycznej wypada 11,33 jajek.

Zastosowanie średniej harmonicznej może mieć miejsce na tej zasadzie, że odwrotność średniej harmonicznej równa się średniej arytmetycznej sumy odwrotności poszczególnych wyrazów szeregu. Czyli

$$\frac{1}{sh} = \frac{1}{12} \left(\frac{1}{9} + \frac{1}{11} + \frac{1}{15} + \frac{1}{16} + \frac{1}{15} + \frac{1}{14} + \frac{1}{13} + \frac{1}{11} + \frac{1}{10} + \frac{1}{9} + \frac{1}{6} + \frac{1}{7} \right) = 0.0949.$$

Wobec czego

$$sh = \frac{1}{0.0949} = 10.54$$

Za 1 markę wypada z obliczenia według średniej harmonicznej 10.54 jajek.

Odchylenia. Każda z wymienionych średnich jest tylko najprostszą charakterystyką zbiorowości, ale bynajmniej nie dostateczną. Mayr podaje przykład, kiedy dwa szeregi przy jednakowej wartości średniej arytmetycznej składają się z najzupełniej różnorodnych elementów. W ciągu 10 kolejnych lat liczba urodzeń wynosiła

w mieście A	w mieście B
4.800	2.000
4.900	3.000
5.000	4.000
5.100	5.000
5.200	6.000
5.100	7.000
5.000	8.000
4.900	9.000
4.950	4.000
5.050	2.000

W jednym i drugim wypadku przeciętna roczna urodzeń równa się 5.000, ale szeregi są różne. Ażeby bliżej scharakteryzować szeregi należy obliczyć *odchylenia* poszczególnych liczb szeregu od średniej arytmetycznej w celu otrzymania rozpiętości wahania około średniej arytmetycznej. Otrzymamy zatem:

A	odchylenia	B	odchylenia
4.800	— 200	2000	— 3000
4.900	— 100	3000	— 2000
5.000	0	4000	— 1000
5.100	+ 100	5000	0
5.200	+ 200	6000	+ 1000
5.100	+ 100	7000	+ 2000
5.000	0	8000	+ 3000
4.900	— 100	9000	+ 4000
4.950	— 50	4000	— 1000
5.050	+ 50	2000	— 3000

Suma odchyleń (nie uwzględniając znaku) równa się w mieście A liczbie 900; średnia arytmetyczna tych odchyleń 90.

W mieście B suma odchyień wynosi 20.000; średnia arytmetyczna odchyień 2000.

Porównanie otrzymanych odchyień wykazuje dobitnie różnice istniejące w budowie obu szeregów.

Mayr proponuje, ażeby średnie arytmetyczne odchyień, które nazywamy *odchyleniami arytmetycznymi* lub przeciętnymi, przedstawiać w stosunku do średniej arytmetycznej szeregów. Otrzymamy liczby względne: dla szeregu A — 1.8%, dla szeregu B — 40%; wobec tego wzór średniej arytmetycznej dla A będzie $5000^{1.8}$, dla B — 5000^{40} . Uwzględniając w każdym szeregu *max* i *min* poszczególnych wyrazów, wzory średnich arytmetycznych będą następujące:

$$\begin{aligned} \text{dla A} &= 5000^{1.8} \frac{\text{max } 5050}{\text{min } 4800} \\ \text{dla B} &= 5000^{40} \frac{\text{max } 9000}{\text{min } 2000} \end{aligned}$$

Tak zbudowane wzory dla średnich arytmetycznych ułatwiają zrozumienie, że oba szeregi przy jednakowej wielkości średniej posiadają strukturę zasadniczo różną.

Odchylenie arytmetyczne jest jedną z najważniejszych wielkości statystycznych. Cechy zbiorowe zjawisk, t. j. te cechy, które poddajemy opracowaniu statystycznemu, są z reguły cechami, ulegającymi indywidualnym wahaniom. Poszczególne części składowe zbiorowości odchylają się od przeciętnego typu, tworząc skalę wahania zjawiska. Odnalezienie przeciętnego odchylenia jest ważnym pierwiastkiem analizy naukowej.

Weźmy dla przykładu własności fizyczne człowieka. Wzrost, waga, siła, objętość klatki piersiowej i t. d. są to cechy zbiorowe masy ludzkiej, oscylujące między pewnymi granicami wielkości pomiarowych. Średni wzrost, średnia waga, średnia objętość klatki piersiowej — charakteryzują masę ogólnikowo. Ażeby poznać masę bliżej, należy zbadać *w jakich granicach* mieszczą się odchylenia czynników składowych masy, jakie jest przeciętne odchylenie masy, t. j. dojść do zrozumienia budowy masy, co jest tembardziej konieczne, gdy mamy do czynienia z porównaniem dwu i więcej mas między sobą.

Odchylenie arytmetyczne posiada dwie ważne własności. Pierwsze twierdzenie: *suma* algebraiczna (t. j. z uwzględnieniem dodatnich i ujemnych znaków) *odchyień* od średniej arytmetycznej równa się 0. Oznacza to, że odchylenia ujemne równe są w sumie odchyleniom dodatnim. Jeżeli mamy szereg złożony z wyrazów $a \ b \ c \ d \ \dots \dots \dots x$, to średnia arytmetyczna

tego szeregu $= \frac{a + b + c + d + \dots + x}{n}$, gdzie n ozna-

cza ilość wyrazów w szeregu. Odchylenie arytmetyczne będzie

$$\text{dla } a \quad a - \frac{a + b + c + d + \dots + x}{n}$$

$$\text{dla } b \quad b - \frac{a + b + c + d + \dots + x}{n}$$

$$\text{dla } c \quad c - \frac{a + b + c + d + \dots + x}{n}$$

$$\text{dla } d \quad d - \frac{a + b + c + d + \dots + x}{n}$$

$$\begin{array}{cccc} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array}$$

$$\text{dla } x \quad x - \frac{a + b + c + d + \dots + x}{n}$$

Sumując odchylenia, zauważamy, że dodajemy wyrazy szeregu

$(a + b + c + d + \dots + x)$ a odejmujemy n razy $\left(\frac{a + b + c + d + \dots + x}{n}\right)$

t. j.

$$\text{suma odchyień} = (a + b + c + d + \dots + x) - \frac{n(a + b + c + d + \dots + x)}{n} =$$

$$= (a + b + c + d + \dots + x) - (a + b + c + d + \dots + x) = 0.$$

Drugie twierdzenie: *suma kwadratów odchyień* od średniej arytmetycznej jest wielkością najmniejszą, czyli jest mniejsza od sumy kwadratów odchyień od jakiegokolwiek innego wyrazu szeregu, lub innej jakiej wielkości, choćby bardzo mało różniącej się od średniej arytmetycznej. Wyjaśniamy to na przykładzie:

Szereg	odch. od śr. arytm.	kwadraty odchyień	odch. od 1-go wyrazu	kwadraty odchyień
4	— 1	1	0	0
7	+ 2	4	+ 3	9
3	— 2	4	— 1	1
6	+ 1	1	+ 2	4
		<u>10</u>		<u>14</u>
od 2-go wyr.		kwadraty odchyień	od 3-go wyr.	kwadraty odchyień
— 3		9	+ 1	1
0		0	+ 4	16
— 4		16	0	0
— 1		1	+ 3	9
		<u>26</u>		<u>26</u>

Suma kwadratów odchyień kolejno wynosi 10, 14, 26, 26, z czego widać w formie przykładu, że suma kwadratów odchyień od średniej arytmetycznej jest *najmniejsza*.

Oprócz odchylenia arytmetycznego stosujemy jako rzecz nieuniknioną w opracowaniach statystycznych *odchylenie kwadratowe*, które *Czekanowski* nazywa *średniem* w odróżnieniu od przeciętnego (arytmetycznego). Odchylenie kwadratowe otrzymujemy jako pierwiastek ze średniej arytmetycznej kwadratów odchyień. Powszechne użycie odchylenia kwadratowego pomimo nieco więcej skomplikowanego sposobu obliczania tłumaczy się cennymi zaletami tej wielkości statystycznej, która jest bardziej czułym narzędziem miary. Podnoszenie odchyień do kwadratu pozwala na uwypuklenie znaczenia większych odchyień, które w stosunku do odchyień mniejszych zyskują na wadze. Odchylenie wartości 2, podniesione do drugiej potęgi, daje rezultat 4; odchylenie wartości 9 w tych samych warunkach daje 81. Rozpiętość stosunku między liczbami 4:81 jest większa, aniżeli pomiędzy liczbami 2:9, przeto odchylenia większe zostają wprowadzone do sumy kwadratów z większą stosunkowo wagą.

Niekiedy, oprócz wymienionych, stosujemy odchylenie *ćwiartkowe*, jako miarę prostą i łatwą do obliczania, o której wspomnimy na właściwym miejscu.

Wskaźnik zmienności.

$$v = \frac{\delta}{Sd} \cdot 100$$

Przykład *).

Średni wzrost mężczyzny dorosłego w Anglii $A_1 = 1.74$ metry, $\delta = \pm 0.06$ m.

Średni wzrost dziecka 9-letniego w Anglii $A_2 = 1.18$ m.,
 $\delta = + 0.05$ m.

Jak widzimy, odchylenie wzrostu mężczyzn dorosłych w Anglii jest większe jak u dzieci, z czego wnosićby można, że wzrost dorosłych mężczyzn jest bardziej zmienny. Analiza za pomocą wskaźnika zmienności mówi, że jest inaczej. Mianowicie

$$V_1 = \frac{0.06}{1.74} \cdot 100 = 3.5$$

$$V_2 = \frac{0.05}{1.18} \cdot 100 = 4.2$$

Z tego wynika, że wzrost dzieci 9-letnich w Anglii jest bardziej zmienny, jak wzrost dorosłych mężczyzn.

Wielkość topologiczna, inaczej zwana *medjaną*, lub *wartością środkową*, jest to taka wielkość statystyczna, która przepoławia szereg na dwie równe części, z których jedna zawiera wielkości mniejsze od niej, drugą większe — w wypadku szeregu rosnącego, zatem średnia topologiczna leży pośrodku. W szeregu

*) Prof. St. Moszczeński. Podstawy organizacji gospodarstw wlejskich. Cz. I. Metody statystyczne. Warszawa 1924, str. 110.

7, 8, 9, 10, 11, 12, 13, 14, 15 — medianą będzie 11, które rozgranicza dwie równe części szeregu. W tym wypadku, kiedy ilość liczb w szeregu jest parzysta, średnią topologiczną będzie przeciętna z dwu liczb środkowych.

Średnia topologiczna znajduje dość częste zastosowanie w opracowaniu statystycznym, gdy ku temu nadaje się charakter zjawisk. Gdy naprz., w dziedzinie psychologii doświadczalnej badamy zdolności uczniów klasy, należy ich uszeregować według zdolności — od najmniej zdolnego do najzdolniejszego; osobnik, stojący w połowie szeregu będzie przedstawiał typ przeciętny. Albo, gdy badamy przeciętny typ dochodu społecznego — szeregujemy dochody według skali rozpiętości od najmniejszych do największych; typ przeciętnego dochodu znajduje się pośrodku. W tym wypadku średnia topologiczna jest prawidłowsza od średniej arytmetycznej, gdyż wyobraźmy sobie, że do szeregu dochodów wprowadziliśmy dochód miliardera. Średnia topologiczna przesunie się tylko o jedno miejsce, natomiast średnia arytmetyczna ulegnie w danym wypadku poważnej zmianie.

Wielkość modalna, inaczej zwana *modus*, lub *średnia normalna, najczęstsza*. Jest to taki rodzaj wielkości statystycznej, który w szeregu powtarza się najczęściej. Spotykamy się z modusem w statystyce cen, plonów, płac zarobkowych etc. Gdy zapytujemy o przeciętną cenę towaru, kupiec podaje tę cenę, za którą sprzedał największą ilość towaru. Gdy zapytujemy o przeciętny urodzaj, rolnik wskazuje na urodzaj najczęściej spotykany w okolicy. Jest to więc taka przeciętna, która spotyka się najczęściej. Wielkość modalna ma poważne znaczenie w opracowaniu statystycznym, bowiem najlepiej charakteryzuje typ zjawiska. Zasadnicza różnica ze średnią arytmetyczną polega na tem, że krańcowe notowania (np., ceny maksymalne i minimalne) nie posiadają wpływu na wartość wielkości modalnej, podczas gdy średnia arytmetyczna jest uzależniona od wszystkich składników szeregu.

Liczby
względne.

Jeżeli w szeregu mamy n spostrzeżeń, a jakiś wypadek lub cecha występuje w nim a razy, to stosunek a/n nazywamy *liczbą względną*, która wskazuje na stosunkową cz stość powtarzania się poszczególnego zjawiska w danej zbiorowości. Prof. A. Kaufman odróżnia dwa rodzaje liczb względnych: jedne wyrażają ustosunkowanie części zbiorowości do całej zbiorowości, lub jednej części do drugiej — w tej samej zbiorowości, są to ustosunkowania *ekstensywne*; drugie wyrażają stosunek zjawisk jednej kategorii do zjawisk innej kategorii. Naprz., ustosunkowaniem ekstensywnym będzie stosunek jednej części ludności do ogółu ludności, lub jednej części do drugiej (stosunek liczebny mężczyzn do kobiet). Ustosunkowaniem *intensywnym* będzie porównanie liczby mieszkańców kraju z obszarem zaludnienia (gęstość zaludnienia), porównanie śmiertelności z liczbą ogółu mieszkańców i t. p.

Liczbom względnym dają pospolicie miano *współczynników statystycznych*. Mówimy o współczynniku śmiertelności, o współczynniku urodzeń, małżeństw, samobójstw, spożycia etc.

Współczynniki obliczamy według procentów, pro mille, na 100.000 i t. d. Ilość zgonów obliczamy na 1000 ludności; mówimy współczynnik śmiertelności wynosi 22, to znaczy na 1000 ludzi wypada 22 wypadków śmierci; w r. 1891 w Anglii na 100.000 ludzi przypadało 75 osób ślepych, we Francji — 84 — to znaczy współczynnik ślepoty w Anglii wynosi 75, we Francji 84.

Index numbers.

Do liczb względnych należą *liczby wskaźnicze*, z angielska zwane index numbers. Są to wielkości, które określają zmianę wartości zjawisk w stosunku do przyjętej podstawy. Posiłkujemy się wskaźnikami w zagadnieniach kosztów utrzymania, kosztów produkcji, zmiany wartości pieniądza etc. Za okres podstawowy przyjmujemy zazwyczaj okres kilkuletni poprzedzający, dajmy na to ubiegłe 10-lecie, z którym porównujemy zjawiska bieżące. Można ograniczyć się okresem nawet jednego

roku, jeżeli rok ten jest przełomowym, jak naprz. ostatni przed wojną rok 1913, po którym wskutek działań wojny nastąpiły zmiany we wszystkich dziedzinach społeczno - gospodarczych. Czy w pierwszym, czy w drugim wypadku okres podstawowy jest nieruchomy. Wszelako nie jest to warunek niezbędny. Można stosować okresy podstawowe zmienne w ten sposób, że każda wartość z nowego okresu wyrażona jest w stosunku do poprzedniego, a ta z kolei w stosunku do jeszcze wcześniejszego i t. d. Otrzymamy w wyniku łańcuch liczb wskaźniczych, dla których podstawą porównania jest liczba poprzednia. Takie liczby nazywamy *wskaźnikami łańcuchowymi*. Znaczenie wskaźników łańcuchowych w praktyce statystycznej prof. *Lipiński* ocenia w ten sposób: „ważne towary o charakterze sezonowym, powinny być uwzględniane przy obliczaniu poziomu cen, gdyż wywierają na ten poziom znaczny wpływ, tymczasem wychodząc z pewnego oddalonego okresu bazy, musimy opierać się stale na tej samej i niezmiennej liczbie artykułów”.

Przy obliczaniu liczb wskaźniczych nie wystarcza ograniczyć się zmianami, dotyczącymi jednego jakiegoś zjawiska, należy brać pod uwagę grupę zjawisk lub całe kompleksy zjawisk. Tak naprz., dla określenia zmian wartości pieniądza posilkujemy się wahaniem cen niejednego towaru, lecz całej grupy towarów, z których jedne mają większe, inne mniejsze znaczenie dla gospodarstwa narodowego, a w zależności od tego mają większe lub mniejsze znaczenie dla formowania liczb wskaźniczych. Powstaje konieczność ustalenia wagi poszczególnych składników, określających liczby wskaźnicze. Przykład*): mamy trzy towary — zboże, węgiel kamienny i żelazo, których cenę w roku podstawowym oznaczamy poszczególnie przez 100. W roku następnym niech będzie cena względna tych towarów 105, 102 i 104. O ile wagę w spożyciu ogólnokrajowym zboża przyjmijemy za 10, węgla za 4 i żelaza za 5, to w takim razie liczba wskaźnicza równać się będzie

*) Kablukow. Statistika, str. 178, r. 1915.

$$\frac{105 \cdot 10 + 102 \cdot 4 + 104 \cdot 5}{10 + 4 + 5} = 104.1, \text{ czyli ceny wzrosły}$$

przeciętnie w stosunku 100 do 104.1, a wartość pieniądza spadła. Powyższa cyfra indeksowa nosi miano *wskaźnika ważonego*.

Określenie wagi zjawiska jest subiektywne poniekąd i nastrocza znaczne trudności przy ustaleniu. Dlatego, o ile ilość wybranych do porównania zjawisk jest dostatecznie wielka (100 lub więcej), obliczanie wskaźników prostych, czyli nieważonych, można uważać za rzecz wystarczającą. Wskaźniki proste mało różnią się od wskaźników ważonych.

Współczynnik współzależności.

Dwa szeregi, porównywane ze sobą, o ile dotyczą określonych cech danej zbiorowości, mogą wykazać, że albo ze zmianą liczb w jednym szeregu równolegle zmieniają się liczby drugiego szeregu, albo odwrotnie, liczby w szeregach zmieniają się we wprost przeciwnych kierunkach, albo też, że pomiędzy zmianą liczb w jednym szeregu, a zmianą liczb w drugim szeregu nie ma żadnego związku. Badamy zależność szeregów od siebie w tym wypadku, gdy między nimi może zachodzić związek przyczynowy. Jeżeli jeden szereg, wyrażający zmienne cechy zjawiska, ma się tak do drugiego szeregu jak przyczyna do skutku, to powstaje potrzeba określenia ścisłej miary zależności. Taką miarą jest *współczynnik korelacji*, czyli współczynnik *współzależności*, stanowiący miarę matematyczną zależności między uszeregowanymi zjawiskami z jednoczesnem wskazaniem na charakter tej zależności. Zależność jest prosta, kiedy dwa szeregi ulegają zmianom w jednakowym kierunku, czyli są zbieżne; zależność jest odwrotna — gdy szeregi są rozbieżne. Jeżeli matematyczna miara zależności zbliża się do ± 0 , związku przyczynowego między zjawiskami, reprezentowanymi przez szeregi, nie ma wcale.

Elementarne opracowanie szeregów.

Zapoznawszy się w wielkościami statystycznymi, z kolei omówimy zagadnienie opracowania szeregów.

Węglug jakiego kryterjum odbywa się zestawianie, porównywanie szeregów? Naprzód — według cechy jakościowej. Gdy mamy do czynienia naprz., z wysokością plonów, czynimy zestawienie szeregu plonów z różnemi gatunkami gleby. Staramy się skonstatować metodą ilościową związek między wysokością plonu a jakością gleby. Gdy mowa jest o śmiertelności w mieście X, zestawiamy liczbę wypadków śmierci z dzielnicami miasta według ich położenia topograficznego i t. p. W tych wypadkach wiążemy przyczynowo ilość faktów z podmiotową oceną warunków, w jakich zjawiska przebieg mają.

Właściwem jednak zadaniem analizy statystycznej jest porównanie szeregów, złożonych z cech ilościowych. Zestawiamy więc ilość wypadków śmierci w mieście X z wysokością opłat mieszkaniowych, z ilością metrów kwadratowych powietrza, przypadających na 1 mieszkańca, z ilością osób, zajmujących 1 pokój; zestawiamy ilość popełnionych w kraju przestępstw z wysokością cen na zboże itd., itd.

Otrzymane szeregi analizujemy. Prof. Kaufman utrzymuje, że nawet najbardziej subtelna analiza matematyczna nie wyklucza pożytku i konieczności zastosowania „*elementarnych środków opracowania*”.

1. Do elementarnych środków opracowania szeregów należy w pierwszym rzędzie posilkowanie się średnią arytmetyczną, wyprowadzoną z części szeregów, w celu uczynienia ich więcej przejrzystymi. W 2 szeregach dzielimy liczby na grupy, dla każdej grupy określamy średnią arytmetyczną.

I.	II.		
18	22		
22	35		
35	18	Po przekształceniu:	
15	27	I. szereg	II. szereg
25	20	25	25
37	34	26	27
30	40	24	30
22	15	25	32
20	35		

15	42
40	30
20	24

W pierwszym szeregu, jak to widać po przekształceniu, niema wyraźnej tendencji w zmianach liczb; natomiast drugi szereg jest — ogólnie biorąc — rosnący.

2. Często się zdarza, że mamy przed sobą szeregi, między którymi trudno dopatrzyć się związku z pierwszego rzutu oka, naprz. (Kaufman):

18,	21,	15,	23,	17,	27,	31,	36,	28
35,	47,	34,	37,	32,	49,	48,	53,	42

Kaufman proponuje w takim wypadku ponumerować liczby szeregów cyframi rzymskimi, poczynając od najmniejszej, kończąc największą. Szeregi przybiorą następujący wygląd:

III,	IV,	I,	V,	II,	VI,	VIII,	IX,	VII
III,	VI,	II,	IV,	I,	VIII,	VII,	IX,	V

Cyfry rzymskie pierwszego szeregu ułożymy po kolei od I do IX, wówczas w drugim szeregu otrzymamy odpowiedniki:

I,	II,	III	IV,	V,	VI	VII,	VIII,	IX
II,	I,	III	VI,	IV,	VIII	V,	VII,	IX

Na zasadzie numeracji stwierdzamy, że liczbom mniejszym pierwszego szeregu odpowiadają liczby mniejsze drugiego szeregu, liczbom większym — większe, t. j. sądząc na podstawie wielkości liczb oba szeregi mają dużo cech wspólnych.

3. Szeregi opracowujemy, sprowadzając wartości poszczególnych wyrazów do wspólnej podstawy, naprz. do sumy ich, przyjmowanej za 100, 1000 i t. p.

Liczba 4002 samobójców (Oettingen) dzieli się według grup wieku:

Klasy	Liczby absolutne	Po sprowadzeniu do sumy = 1000
poniżej lat 16	28 osób	7
od 16—20	175 "	44
" 21—30	693 "	173
" 31—40	787 "	197

od 41—50	879 osób	219
„ 51—60	666 „	167
„ 61—70	483 „	121
„ 71—80	233 „	56
powyżej lat 80	58 „	16

Suma 4002 osoby

4. Analogicznie przeprowadzamy przekształcenie szeregu według max. Za podstawę bierzemy największą liczbę szeregu, przyjmując ją za 100, 1000 etc., i odpowiednio do tego wyrażamy poszczególne wyrazy.

Obszar seradelli w $\frac{0,0}{100}$ w r. 1901 Szereg przekształcony
w b. Król. Kongresowem

Ziemia Lubelska	—	Ziemia Kaliska	100
„ Radomska	0,11	„ Piotrkowska	63,1
„ Kielecka	0,14	„ Płocka	47,4
„ Piotrkowska	1,20	„ Warszawska	26,3
„ Kaliska	1,90	„ Siedlecka	10,5
„ Warszawska	0,50	„ Kielecka	7,3
„ Płocka	0,90	„ Radomska	5,7
„ Łomżyńska	0,10	„ Łomżyńska	5,2
„ Siedlecka	0,20	„ Lubelska	—
„ Suwalska	—	„ Suwalska	—

5. Opracowanie według min.

We Francji w okresach 5-letnich popełniano przeciętnie następujące ilości samobójstw:

Lata	Ilości samobójstw	Szereg przekształcony
1826 — 30	1739	1000
1831 — 35	2263	1301
1836 — 40	2573	1712
1841 — 45	3951	2272
1846 — 50	3446	1981

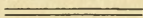
6. Jeżeli szereg złożony jest z liczb niemających ustalonego kierunku, który mają liczby rosnące lub malejące, to zasługuje na uwagę sposób obierania podstawy ze średniej arytmetycznej, oznaczanej przez 100, 1000 i t. p.

W wyżej przytoczonym szeregu ilości samobójstw we Francji przyjmujemy średnią arytmetyczną za 100, wówczas przekształcony szereg będzie się składał z liczb następujących:

62.2
80.0
92.7
141.3
123.3

Liczby powyższe świadczą, że ilość samobójstw w pierwszych 3 okresach była mniejsza od przeciętnej za cały okres, w następnych dwóch wzniosła się ponad poziom przeciętny — w określonym stosunku.

Opracowanie szeregów sposobem elementarnym winno być stosowane tam, gdzie ze względu na charakter zjawisk może być uważane za wystarczające. Nie należy go w każdym bądź razie lekceważyć. Natomiast w innych wypadkach, gdy chodzi o głębszą analizę zjawisk, zwłaszcza dotyczącą związków przyczynowych w skomplikowanych zbiorowościach, należy sięgnąć po bardziej subtelne narzędzia pracy, których dostarcza choćby elementarna matematyka.



VII. Znaczenie teorii prawdopodobieństwa dla metody wielkich ilości.

Zjawiska gatunkowe
i indywidualne.

Zastanawiając się nad charakterem badania statystycznego, zauważamy, że cechy poszczególnych przed-

miotów, ustalone przez obserwację, same w sobie nie stanowią celu obserwacji, dopiero dostatecznie wielka zbiorowość cech indywidualnych staje się przedmiotem badania. Fakty statystyczne są ważne dla naukowego poznania nie jako takie, poszczególnie, lecz jako masa, posiadająca odrębną charakterystykę, wyrażoną cechami zbiorowymi. Pod tym względem daje się wyraźnie przeprowadzić granicę pomiędzy badaniami przyrodniczymi i statystycznymi. W badaniach przyrodniczych chodzi głównie o ustalenie cech gatunkowych przedmiotu. Tak naprz., botanik, badając roślinę, zastanawia się nad cechami, które są wspólne wszystkim osobnikom tego samego gatunku. Zoolog, konstatujać budowę serca u żaby, uogólnia, że wszystkie żaby posiadają serce trzykomórkowe, podobnie jak anatomicy stwierdza, że wszyscy ludzie posiadają serce czterekomórkowe. W tych wypadkach mamy do czynienia z jednostkowym przedmiotem badania, ale wynik badania jest ogólny dla całego rodzaju danych zjawisk. Teoretycznie biorąc, poza wypadkami patologii roślinnej lub zwierzęcej, stosunek liczby wypadków opatrzonych określoną cechą, winien odpowiadać i pokrywać się z liczbą wypadków spostrzeżonych, t. j. stosunek m (liczba wypadków z określoną cechą) do m^1 (liczba spostrzeżeń) winien równać się jedności ($\frac{m}{m^1} = 1$).

Zjawisko
statystyczne.

W badaniu statystycznym taki stosunek jest teoretycznie wykluczony, ponieważ statystyk bada cechy indywidualne, t. j. takie, które u jednych osobników danej masy są, u innych ich nie ma. Dlatego liczba ogólna wypadków spostrzeżonych (m_1) będzie zawsze większa od liczby wypadków, opatrzonych określoną cechą (m), czyli stosunek

$\frac{m}{m_1} < 1$. Widzimy, że wzór liczby statystycznej jest to ułamek prawidłowy, w którym licznik odpowiada liczbie wypadków, posiadających określoną cechę, a mianownik przedstawia liczbę wszystkich wypadków.

Weźmy dla przykładu zbiorowisko ludzi, a w niem ludzi, obarczonych jakimś brakiem fizycznym, naprz. ślepotą. Jasnym jest, że liczba ludzi ślepych (m) jest mniejsza od liczby wszyst-

kich ludzi (m_1), czyli $\frac{m}{m_1} < 1$.

Stosunek $\frac{m}{m_1} < 1$ jest matematycznym wyrażeniem prawdopodobieństwa, a zatem zjawisko statystyczne zawsze opiera się na prawdopodobieństwie. Zjawisko statystyczne tem się różni od zjawiska rzeczywistego, że jest tylko prawdopodobne. Jeżeli stwierdzamy fakt statystyczny, że przyrost naturalny ludności w Polsce wynosi 14 promille, to przyrost jakiegokolwiek z najbliższych lat prawdopodobnie wyniesie 14, ale może być mniejszy lub większy od 14.

Teoria prawdopodobieństwa posiada dla badań statystycznych pierwszorzędne znaczenie, aczkolwiek najzupełniej winniśmy zdawać sobie z tego sprawę, że dziedziny badań ściśle statystycznych i dziedziny teorii prawdopodobieństwa są dziedzinami różnemi: gdy pierwsza dotyczy zjawisk naturalnych, t. j. życiowych, druga — dotyczy zjawisk sztucznych, stworzonych przez człowieka w formie koncepcyj oderwanej natury (np., gry loteryjne, hazardowne, losowe etc.).

Prawdopodobieństwo. Punktem wyjścia w teorii prawdopodobieństwa jest określenie prawdopodobieństwa przez Laplace'a. Prawdopodobieństwo zdarzenia jest stosunkiem liczby przypadków sprzyjających zdarzeniu do liczby przypadków możliwych. Przypadki przytem winny być jednakowo możliwe i wzajemnie wykluczające się.

Przewidywanie nastąpienia zdarzenia w dziedzinie wielu nauk odgrywa pierwszorzędną rolę, gdyż umysł ludzki skłonny jest wychodzić z pewnych założeń, w myśl których całokształt warunków sprzyja lub nie sprzyja danemu zdarzeniu. Jedną z najogólniejszych takich apriorycznych umów jest wiara w ciągłość, bez której to wiary żadna nauka nie utrzymałaby się. Już Leibnitz stwierdził, że naprz. zjawiska przyrodnicze są ciągle.

Przy określeniu prawdopodobieństwa założyliśmy, że przypadki winny być jednakowo możliwe i wzajemnie wykluczające się. Jeżeli w urnie posiadamy 2 kule, czarną i białą, to przy wyjmowaniu kul winien być zachowany warunek, ażeby wyciągnięcie białej lub czarnej kuli było jednakowo ułatwione (jednakowe rozmiary i jednakowe rozmieszczenie), oraz wyciągnięcie jednej wykluczało możliwość jednoczesnego wyciągnięcia drugiej, co istotnie następuje. Albo: przy podrzucaniu monety mamy do czynienia z przypadkami jednakowo możliwymi, kiedy moneta pada do góry orłem lub reszką, i wzajemnie wykluczającymi się, t. j. kiedy upadnie reszką, to jednocześnie nie może upaść orłem.

**Rachunek
prawdopodobieństwa.**

Każde wogóle zdarzenie może być pewne, albo też ewentualne, prawdopodobne.

Jakie jest prawdopodobieństwo, że przy rzucie moneta upadnie reszką? Ponieważ możliwy jest przypadek jeden (orzeł lub reszka), a przypadków wszystkich jest 2, to prawdopodobieństwo jednego z tych przypadków równa się $\frac{1}{2}$. Jeżeli w urnie posiadamy 3 kule różnokolorowe, to prawdopodobieństwo wyciągnięcia kuli białej będzie się równać $\frac{1}{3}$, t. j. w myśl definicji prawdopodobieństwa — równać się będzie stosunkowi

możliwych przypadków (jedna kula biała) do wszystkich przypadków (3 kule: biała, czerwona, czarna). Tak samo w loterii: przy 1 wygranej na 1000 biletów prawdopodobieństwo wygranej równa się $\frac{1}{1000}$.

Widzimy, że prawdopodobieństwo zdarzenia równa się sumie sprzyjających przypadków, dzielonej przez liczbę wszystkich przypadków, a suma prawdopodobieństw wzajemnie się wykluczających zdarzeń równa się 1, t. j. jeżeli prawdopodobieństwo p jednego zdarzenia R (naprz., reszki) równa się $\frac{1}{2}$ i q drugiego zdarzenia O (naprz., orła) równa się $\frac{1}{2}$, to $p + q = 1$. Z tego wynika, że jeżeli prawdopodobieństwo nastąpienia zdarzenia równa się p , to prawdopodobieństwo przeciwnie, t. j. nienastąpienia zdarzenia, czyli q , będzie się równać:

$$q = 1 - p.$$

Przy 2 zdarzeniach R i O (naprz. przy podrzucaniu 2 monet) z prawdopodobieństwami $p = \frac{1}{2}$ i $q = \frac{1}{2}$ w kolejno po sobie następujących doświadczeniach otrzymamy następujące kombinacje:

albo obie reszki . . . $R R$
 jedna reszka, jeden orzeł $R O$
 jeden orzeł, jedna reszka $O R$
 oba orły $O O$.

Ponieważ wszystkich jednakowo możliwych kombinacji otrzymaliśmy 4, to prawdopodobieństwo każdej równa się $\frac{1}{4}$. Biorąc pod uwagę, że prawdopodobieństwo prostego zdarzenia R lub O wynosi $\frac{1}{2}$, zauważamy, że

$$\frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$$

czyli, że prawdopodobieństwo przypadku złożonego równa się iloczynowi prawdopodobieństw przypadków prostych.

Przy podrzucaniu 3 monet prawdopodobieństwo zdarzenia, że wszystkie monety upadną reszką do góry, t. j. prawdopodobieństwo kombinacji $R R R$ równa się

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8},$$

ponieważ możliwe są następujące kombinacje:

R R R
 R R O R O R O R R
 O O R O R O R O O
 O O O

Wszystkich kombinacyj mamy 8, kombinacja R R R jest 1, a zatem prawdopodobieństwo tej kombinacji wyniesie $\frac{1}{8}$.

Zaobserwowany powyżej wynik nosi miano reguły *mnożenia* prawdopodobieństw. Mając przed sobą świeżo stosowaną talję kart, zadajemy sobie pytanie, jakie jest prawdopodobieństwo, że ciągnąc jedną kartę na chybi trafi wyciągniemy asa? To prawdopodobieństwo, równa się $\frac{1}{13}$, albowiem na 52 karty

mamy asów 4, t. j. $\frac{4}{52} = \frac{1}{13}$. Gdybyśmy, tasując nanowo karty, po raz drugi chcieli wyciągnąć asa, to prawdopodobieństwo takiego zdarzenia równa się $\frac{1}{169}$ ($= \frac{1}{13} \cdot \frac{1}{13}$).

Jeżeli w przykładzie z 3 monetami nie będziemy zwracali uwagi na szereg zdarzeń R i O w przypadkach, kiedy oba zdarzenia występują wspólnie, a tylko na liczbę powtórzeń, to przypadków, gdzie mamy 2 reszki i 1 orła, naliczymy 3, a mianowicie

R R O R O R O R R, t. j.

prawdopodobieństwo takiego połączonego przypadku wyniesie $\frac{3}{8}$. Biorąc pod uwagę, że prawdopodobieństwo każdego z tych 3 przypadków oddzielnie stanowi $\frac{1}{8}$, zauważamy, że

$$\frac{3}{8} = \frac{1}{8} + \frac{1}{8} + \frac{1}{8}.$$

Wynik ten nosi nazwę reguły *dodawania* prawdopodobieństw.

Wogóle prawdopodobieństwo różnych skojarzeń przypadków wyraża się przez poszczególne człony rozwiniętego dwumianu Newtona. Jeżeli prawdopodobieństwo dwóch zdarzeń równa się dla każdego z nich $\frac{1}{2}$, to kombinacje zdarzeń przy dwóch kolejnych doświadczeniach będą następujące: $(\frac{1}{2} + \frac{1}{2})^2$, czyli $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$; przy trzech doświadczeniach $(\frac{1}{2} + \frac{1}{2})^3$, t. j. $\frac{1}{8}$, $\frac{3}{8}$, $\frac{3}{8}$, $\frac{1}{8}$ i t. d.

Liczba zdarzeń odpowiada współczynnikom rozwiniętego binomu. Współczynniki dla dwumianu 2-go stopnia są następujące

• • • • •

Prawo liczb wielkich.

Prawo wielkich liczb jest wynikiem rozważań czysto teoretycznych i tylko przy pewnych warunkach wyrokować może

i w dziedzinie zjawisk świata rzeczywistego, t. j. w statystyce. Poza granicami doświadczenia, kiedy mamy do czynienia z faktami zaistnienia zdarzeń w naturalnych warunkach rzeczywistości, zastosowanie prawa liczb wielkich napotyka na trudności ze względu na to, że wogóle nie są nam znane wszystkie warunki zdarzeń wobec złożoności zjawisk w świecie rzeczywistym oraz wielkich trudności wyodrębnienia działających w nim czynników, a tudzież nie jest nam znane a priori prawdopodobieństwo nastąpienia zdarzeń. Tem nie mniej jednak, wszyscy teoretycy, poza nielicznymi wyjątkami, są zdania, że zastosowanie prawa liczb wielkich w statystyce przy zachowaniu warunków analizy naukowej, polegającej tutaj na umiejętności uchwycenia analogii między warunkami naturalnymi nastąpienia zdarzeń a warunkami matematycznej koncepcji, może mieć miejsce i posiada duże znaczenie.

Doświadczenia, wykonane w warunkach koncepcyj matematycznych, przekonywują nas, że w tej dziedzinie prawo liczb wielkich daje się rzeczywiście pozytywnie stwierdzić. W naszym seminarjum*) przeprowadzono również podobne doświadczenia, o których wspomnimy niżej.

Prawo liczb wielkich było ustalone i sprawdzone na zasadzie doświadczeń z kulami. „Wyobraźmy sobie — powiada Laplace — urnę, zawierającą czarne i białe kule. Zaczynamy wyjmować kule z urny i kłaść je z powrotem”. Z początku tej operacji stosunek kul będzie b. nieprawidłowy, ale w miarę zwiększania się liczby doświadczeń, nieprawidłowości te znikają i liczba wyciągniętych białych i czarnych kul odpowiada stosunkowi ich liczby w urnie.

Quectelet powtórzył doświadczenia z kulami 4096 razy, wyjmując kule z urny, w której było pół na pół czarnych i białych. W rezultacie otrzymał 2066 białych kul i 2030 czarnych kul, czyli 50.4% białych i 49.6% czarnych (prawdopodobieństwo — $\frac{1}{2}$).

*) Seminarjum statystyczno-ekonomiczne przy katedrze statystyki i polityki ekonomicznej Kat. Uniw. Lub. pod kierunkiem autora.

Buffon na 2048 orłów przy podrzucaniu monet otrzymał 1992 reszki, czyli na 100 rzutów 51 orłów i 49 reszek.

Westergaard zrobił 10.000 doświadczeń z białymi i czarnymi kulami, które znajdowały się w urnie w stosunku $1/2 : 1/2$, i otrzymał 5011 kul jednego koloru i 4989 kul drugiego koloru, co czyni 50.1% i 49.9%. Przytem *Westergaard* prowadził obliczenie grupami po 100, po 200 i po 2000. Przy obliczaniu grupami po 100 wyjęć otrzymywał następujące wahania

od 66 czarnych kul do 34 białych

od 37 " " do 63 białych kul, czyli sto-

sunek wahań był, jak 2:1 i 1:2.

Odchylenie od przeciętnej wynosiło 32% w pierwszym przypadku (16:50).

Odchylenie przy obliczeniach większymi grupami zmniejszało się:

17 — 11 — 11 — 9 — 10 — 9 — 9 — 8%

Przy obliczeniach po 2000 odchylenie wynosiło 4%.

Przy całkowitem podsumowaniu wszystkich 10.000 wypadków odchylenie wyniosło 0,22%.

Tenże *Westergaard* obliczył, że przy losowaniu kopenhaskiej loterii otrzymano 3640 razy cyfry od 1 do 45 i 3635 razy cyfry od 45 do 90, czyli 50,34% i 49,66%.

W naszym seminarjum *) wykonano doświadczenie z numerami, wygrywającymi w ciągnięciach V klasy Polskiej Loterii Państwowej, które odbyły się od 6 września do 13 października 1928 r. w 31 dniach. Obliczono metodą kreskowania numery, zakończone od 0 do 50 i od 51 do 100. Numerów wygrywających w pierwszej grupie okazało się 30714, w drugiej 30786, czyli 49.94% i 50.06%. Średnia arytmetyczna ilości numerów wygrywających w ciągu jednego dnia wynosiła w pierwszej grupie 990,8, w drugiej 993,1.

Stosunek procentowy numerów I i II grupy — 49.94% i 50.06% — ze względu na znacznie większe rozmiary doświad-

* Doświadczenie przeprowadziły pp. studentki Bleńkówna, Jarceka i Sikorska.

czenia, jest bliższy stosunku $1/2 : 1/2$, aniżeli u Westergaarda.

W naszym seminarjum wykonano również doświadczenia z rzutami monet, lecz ze względu na stosunkowo niewielką liczbę rzutów wyniki zbliżają się jedynie do wywodów teoretycznych, wszakże stanowią ilustrację dostateczną dla wykazania, że w warunkach zajścia zdarzeń jednakowo możliwych i wzajemnie wykluczających się liczby przypadków odpowiadają współczynnikom rozwiniętego dwumianu. Wzięto 7 monet i wykonano 1024 rzuty. Kombinacje orła i reszki powtórzyły się następującą liczbę razy:

Szereg z doświadczenia		Szereg teoretyczny
7	— same orły	8
47	— 6 orłów i 1 reszka	56
167	— 5 „ i 2 „	168
272	— 4 „ i 3 „	280
302	— 3 „ i 4 „	280
174	— 2 „ i 5 „	168
53	— 1 „ i 6 „	56
2	— same reszki	8
1024		1024

Prawdopodobieństwo orła $\frac{1}{2}$ i reszki $\frac{1}{2}$. Rozwinięty dwumian $\left(\frac{1}{2} + \frac{1}{2}\right)^7$ daje współczynniki

1 7 21 35 35 21 7 1

Jeżeli w szeregu otrzymanym z doświadczenia przeprowadzimy dzielenie przez wspólny dzielnik 7, to otrzymamy szereg zbliżony do poprzedniego, a mianowicie:

1 $6\frac{5}{7}$ $23\frac{6}{7}$ $38\frac{6}{7}$ $43\frac{1}{7}$ $24\frac{6}{7}$ $7\frac{4}{7}$ $\frac{2}{7}$

Jevons przeprowadził doświadczenie z 10 monetami, wykonywując 2048 rzutów. Według $\left(\frac{1}{2} + \frac{1}{2}\right)^{10}$ współczynniki winny być:

1 10 45 120 210 252 210 120 45 10 1

Otrzymano zaś:

2 17½ 65 126 185½ 244½ 199 115 51 18 ½
przy zastosowaniu dzielnika 2, w rzeczywistości bowiem cyfry otrzymane były 2 razy większe.

Jak widzimy, ilość kombinacyj, praktycznie otrzymanych mimo niewielkiej liczby doświadczeń zbliża się do teoretycznej.

W dziedzinie społecznej liczby wielkie na zasadzie omawianego prawa pozwalają na odkrycie ciekawych prawidłowości. Weźmy jeden przykład. Listy, wrzucone do skrzynki pocztowej bez podania adresu lub bez wskazania nazwy miejscowości w masie korespondencji stanowiły (Rosja):

Rok	Liczba listów	listy bez adresu	bez nazwy miejscowości	na 1.000.000 korespondencyj
1906	983.000.000	26.112	28.749	27 29
1907	1.076.000.000	26.977	26.523	27 25
1908	1.214.000.000	35.315	26.112	27 21
1909	1.357.000.000	33.643	28.445	25 21
1910	1.507.000.000	40.101	33.513	27 24

Prawo liczb
mających.

Obserwacja statystyczna przy zachowaniu prawa liczb wielkich doprowadza do uchwycenia stałych stosunków pomiędzy zjawiskami. Stąd wyciągano wniosek, że stałość stosunków w określonej liczbie zjawisk dla danego czasu i miejsca zależy całkowicie od ilości przeprowadzonych spostrzeżeń. Jednak rozszerzenie podstaw badania statystycznego, zastosowanie go do najbardziej różnorodnych dziedzin zjawisk pozwoliło zrozumieć, że stałość stosunków nie tylko zależy od wielkiej liczby obserwacji, lecz i od *charakteru zjawisk* oraz warunków, które je stwarzają. W dziedzinie naprz. demografji, ustalona przewaga liczebna osobników płci żeńskiej wykazuje zmienność w zależności od miejsca i czasu. Są kraje, które posiadają wybitną przewagę liczby mężczyzn. Niewątpliwie jest to w związku z warunkami, które dane zjawisko stwarzają. Należy przypuszczać tu działanie warunków rzeczywistości, sprzyjających zjawisku w jednym kierunku.

Okoliczność powyższa wywołała konieczność *badania środowiska* i poznania warunków, towarzyszących zjawiskom, gdyż ten lub inny stopień stałości uzależniony jest od nich.

W związku z tem współczesny uczony *Bortkiewicz* wprowadził do nauki pojęcie *prawa liczb małych*, znanego również pod nazwą prawa Bortkiewicza, które opiewa, że zjawiska bardzo rzadkie, zdarzenia osobliwe, występujące częstokroć w znikomej liczbie, ujawniają wielką stałość.

Jeżeli porównamy zjawiska *częste* ze zjawiskami *rzadkimi*, to pierwsze powstają pod działaniem warunków masowych, bardzo różnorodnych, dających najprzeróżniejsze kombinacje skojarzeń, a stąd możliwe są większe odchylenia od typu średniego, od wielkości średniej, mogą być więc raczej mniej stałe; przeciwnie — zjawiska rzadkie pochodzą od warunków spotykanych rzadko, ale zbliżonych do siebie, jednolitych, podobnych, stąd i stałość tych zjawisk może być większa.

Zrozumienie przytoczonej okoliczności czyni koniecznem nie tylko zbieranie wielkich liczb faktów, lecz rozróżnienie charakteru zjawisk i warunków wytwarzających je, t. j. uwzględnienie prawa liczb małych.

----- Wzór Poissona. -----	Wielka liczba doświadczeń, wielka liczba spostrzeżeń — stanowią podstawę dla prawidłowego wnioskowania statystycznego, bowiem, jak widzieliśmy, tylko wielkie ilości, a przynajmniej dostatecznie wielkie ilości — czynią możliwem przejawienie się właściwego ustosunkowania się zdarzeń, odpowiadającego ich prawdopodobieństwu. Liczbowe stosunki między zdarzeniami, sprowadzone do stosunków ich prawdopodobieństw, nabierają cech stałości. Oznaczenie stałości zjawisk posiada doniosłe znaczenie w statystyce.
--	---

W celu określenia *stałości* zjawisk zaobserwowanych posługujemy się wzorem Poissona: $\pm 2 \sqrt{\frac{2 mn}{(m+n)^3}}$, gdzie *m* — oznacza liczbę jednych wypadków, *n* — liczbę drugich. U *Jan-sona* znajdujemy następującą ilustrację zastosowania tego

wzoru. Z liczby 904.594 urodzeń chłopczyków i dziewczynek, na chłopczyków przypada 468,151, na dziewczynki — 436,443. Ze stosunku tych liczb wynika, że prawdopodobieństwo urodzeń chłopczyka równa się $\frac{468151}{904594} = 0,5175$. Zadajemy sobie

pytanie, czy liczba powyższa ma cechy stałości, a więc czy liczba obserwacji była dostatecznie wielka?

Posiłkując się wzorem otrzymujemy:

$$\pm 2 \sqrt{\frac{2 \times 468151 \times 436443}{(468151 + 436443)^3}} = \pm 0,0015, \text{ które należy dodać}$$

i odjąć od liczby prawdopodobieństwa w celu otrzymania górnej i dolnej granicy tegoż. W danym wypadku otrzymamy liczby $(0,5175 + 0,0015) = 0,5190$ i $(0,5175 - 0,0015) = 0,5160$, jako górną i dolną granicę. Prawdopodobieństwo urodzeń chłopca waha się w granicach od 0,5190 do 0,5160. Ponieważ granice wahania są nieduże, wnioskujemy, że liczba spostrzeżeń była dostatecznie wielka.

Prawdopodobieństwo a posteriori.

Wyniki doświadczeń przy dostatecznie wielkiej liczbie doświadczeń mogą być przewidziane, jeżeli na-
przód podane jest prawdopodobieństwo zdarzeń. Liczne do-
wody mamy na to, że przy dostatecznie wielkiem (m_1) do-
świadczeń stosunek liczby (m) przypadków stania się zdarzenia
R do liczby wszystkich zdarzeń (m_1) równa się prawdopodobo-
bieństwu zdarzenia R lub też bardzo mało się od niego różni.
Ale i odwrotnie: sądząc z wyników doświadczenia, które dało
m razy zdarzenie R przy dostatecznie wielkiem m_1 można być
pewnym, że stosunek $\frac{m}{m_1}$ jest prawdopodobieństwem zdarzenia

R lub mało się od niego różni. Dzięki temu można a posteriori,
na zasadzie wyników m_1 doświadczeń, oznaczyć wielkość
prawdopodobieństwa lub dostateczne jego przybliżenie. „Zdarza
się często — mówi *Poincaré* w swej „Nauce i hipotezie” —
że zamiast przewidywania jakiegoś faktu na podstawie mniej
lub więcej niedoskonałej znajomości prawa, które nim rządzi,

znamy właśnie fakty i usiłujemy odgadnąć prawo; że zamiast wyprowadzenia skutków z przyczyn chce się wyprowadzić przyczyny ze skutków. Są to tak zwane zagadnienia o prawdopodobieństwie przyczyn, najbardziej interesujące ze względu na swe zastosowanie naukowe". Prawdopodobieństwo a posteriori, określone na podstawie wyniku doświadczeń, daje możliwość zastosowania rachunku prawdopodobieństwa do zjawisk rzeczywistych, t. j. do tych, z którymi ma do czynienia statystyka. Statystyka bowiem zajmuje się spostrzeganiem faktów, a ze stosunku faktów można dopiero wnosić o przyczynach, wywołujących te fakty.

Teoria błędów.

Powszechne zastosowanie znajduje teoria prawdopodobieństwa w obliczeniach błędów spostrzeżeń. Skorośmy powiedzieli, że ze skutków chcemy sądzić o prawdopodobieństwie przyczyny, to należy mieć na względzie, że skutki składają się z prawdziwej wartości oraz z błędów obserwacji. Błędy odróżniamy systematyczne oraz przypadkowe. Pierwsze powstają zazwyczaj wskutek jakiejś organicznej, istotnej wady spostrzegania, drugie — wskutek przypadkowej niedokładności. Te ostatnie nazywamy błędami *elementarnymi*. Błąd elementarny posiada kilka cech charakterystycznych, które wyróżniają go z pośród innych kategorii błędów. Przede wszystkim, ze względu na jego przypadkowość, w naszym umyśle tkwi przeświadczenie, że daje on odchylenie pomiaru, względnie wyniku obserwacji od wartości prawdziwej w jedną lub drugą stronę z jednakowym prawdopodobieństwem; niema dobrego założenia do przypuszczenia, że błąd elementarny, jako przypadkowy, będzie działał w jakimś określonym kierunku co jest dostatecznie charakterystyczne dla błędu *systematycznego*, odchylającego wynik obserwacji w jedną stronę. Dalej, błąd elementarny należy do kategorii błędów nadzwyczaj małych, powstających wskutek niedających się usunąć niedokładności naszych zmysłów, metod spostrzegania, instrumentów — w przeciwstawieniu do t. zw. *grubych* błędów,

prawie zawsze możliwych do usunięcia. Wreszcie błąd elementarny powstaje od każdej przyczyny niezależnie, kojarząc się z innymi błędami tej kategorii w sposób dowolny.

Im błąd elementarny jest większy (wszystko jedno jakiej wartości, czy dodatniej, czy ujemnej), tym jest mniej prawdopodobny, a to prawdopodobieństwo zmniejsza się w miarę powiększania się błędu. Błąd wpływa na odchylenie pomiaru od przeciętnej wartości (średniej arytmetycznej); im większe jest odchylenie, tem mniej jest ono prawdopodobne, a zatem rzadziej występuje, i odwrotnie, im mniejsze jest odchylenie, tem ono częściej się spotyka. Jest to *prawo błędów*.

Przy jednakowem prawdopodobieństwie odchylenia zarówno w stronę dodatnią, jak i ujemną, szereg odchyleń będzie zupełnie symetryczny, a w skutek tego średnia arytmetyczna będzie wolna od błędu; przy nierównowartem prawdopodobieństwie błędów dodatnich i ujemnych — szereg odchyleń staje się mniej lub więcej asymetryczny, a wskutek tego średnia arytmetyczna będzie obciążona błędem.

Rzeczywista wartość średniej, od której odchylenia podlegają prawu błędów, oznacza z wielkiem przybliżeniem wielkość, istniejącą w naturze. Jednak odnalezienie takiej średniej wymaga, ażeby dane były jednolite, metody spostrzegania najzupełniej jednakowe oraz ilość spostrzeżeń dostatecznie wielka, a wszystkie te warunki nie tak często mogą być zachowane w statystyce. Im bardziej różnorodne są dane liczbowe, im większa ilość przyczyn warunkuje badane zjawisko, tem trudniej otrzymać średnią arytmetyczną, od której wahania ulegałyby prawu błędów.

Błąd elementarny można traktować nietylko jako błąd pomiarów jednej i tej samej wielkości fizycznej, ale również jako odchylenie, leżące w naturze samej wielkości, a więc dotyczące wielu przedmiotów tego samego rodzaju. I tutaj, o ile działają przyczyny zmienne, niestałe, powodujące błędy elementarne, obserwujemy podporządkowanie się spostrzeżeń prawu błędów. Quetelet przeprowadził obliczenie pomiarów

objętości piersi u żołnierzy, poddanych przeglądowi komisji sanitarnej, z następującym rezultatem w 1516 wypadkach:

Objętość piersi w ca- lach ang.	Ilość pomiarów	Liczby na 1000 otrzy- mane z do- świadczenia	Liczby na 1000 otrzy- mane z wy- liczenia	Różnica tych liczb
28	2	1	1	0
29	4	3	3	0
30	17	11	11	0
31	55	36	32	+ 4
32	102	67	69	— 2
33	180	119	121	— 2
34	242	160	170	— 10
35	310	204	190	+ 14
36	251	166	169	— 3
37	181	119	120	— 1
38	103	68	68	0
39	42	28	31	— 3
40	19	13	11	+ 2
41	6	4	3	+ 1
42	2	1	1	0
Razem	1516	1000	1000	— 21 + 21

Średnia objętość piersi równa się około 35 cali. Widzimy, że największa liczba spostrzeżeń przypada na wielkość średnią oraz, że im większe jest odchylenie od średniej, tem mniejsza jest ilość przypadków i odwrotnie, największa ilość przypadków grupuje się dookoła średniej, przytem spostrzeganie sprawdza się przez rachunek prawdopodobieństwa (równoległość i niemal identyczność szeregow praktycznego i teoretycznego).

Prawo błędów przejawia się w warunkach oddziaływania tylko przyczyn przypadkowych. W przykładzie zacytowanym do takich przyczyn będą należały: niedoświadczenie osób, do-

konywujących pomiary, niedoskonałość instrumentów, użytych do pomiarów, mimowolne zmiany w pozycji ciała osób, podlegających pomiarowi i t. p. Już naprz., samo oddychanie może znacznie wpłynąć na wielkość objętości piersi, nie mówiąc o szeregu innych przyczyn, jak niedbała postawa ciała i t. p.

Jeżeli jednak do przyczyn przypadkowych wkładnie się jakaś *stała* przyczyna, działająca w jednym kierunku, naprz. chęć zmniejszenia wzrostu w celu uniknięcia poboru do wojska, to w szeregu liczb pomiarów nie będzie całkowitej symetrii, i liczby uzyskane z spostrzeżenia, będą znacznie odbiegały od liczb teoretycznych.

Nadzieja matematyczna.

Pojęcie nadziei matematycznej praktycznie powstało w sferze stosunków handlowych i finansowych, a ściśle jest związane z teorią prawdopodobieństwa. Jeżeli naprz., kupiec pożyczył drugiemu kupcowi pewną sumę pieniędzy, lecz zwrot jej nie jest pewny, lecz tylko prawdopodobny, to iloczyn tej sumy przez prawdopodobieństwo jej otrzymania stanowi *nadzieję matematyczną*. Przypuśćmy, kupiec daje pieniądze do Ameryki w kwocie 100.000 zł., lecz przewóz jej jest niebezpieczny i prawdopodobieństwo dowiezienia tych pieniędzy równa się $\frac{3}{4}$ pewności, to można kupić tę sumę za 75.000 zł. ($\frac{3}{4} \times 100.000 = 75.000$).

Na zasadzie nadziei matematycznej obliczamy ceny biletów loteryjnych. Jeżeli mamy loteryję, złożoną z 2000 biletów, a wygrana wynosi 10.000 zł., to prawdopodobieństwo wygrania wynosi $\frac{1}{2000}$, czyli nadzieja matematyczna równa się $\frac{1}{2000} \cdot 10.000 = 5$ zł. Jest to t. zw. „sprawiedliwa” cena za 1 bilet.

Zasada nadziei matematycznej ma wielkie praktyczne znaczenie dla obliczeń pożyczek państwowych (premijówek), loterii oraz ubezpieczeń.

Nadzieję matematyczną można uogólnić na większą liczbę oczekiwanych zysków lub strat. Jeżeli naprz., w pewnym przedsięwzięciu lub grze losowej $X_1, X_2, X_3, \dots, X_n$ są zyski i straty albo wygrane i przegrane oczekiwane z prawdopodo-

bieństwami $p_1, p_2, p_3 \dots p_n$, to nadzieją matematyczną oczekiwanego zysku (lub straty) nazywamy sumę iloczynów wartości X przez ich prawdopodobieństwa, czyli

$$X_1 p_1 + X_2 p_2 + X_3 p_3 + \dots + X_n p_n.$$

Prawdopodobieństwo statystyczne.

Zaznaczyliśmy wyżej różnicę między warunkami koncepcji umysłowej zajścia lub niezajścia zdarzeń, w których przejawia się prawdopodobieństwo matematyczne, a warunkami życia, stanowiącemi dziedzinę prawdopodobieństwa statystycznego. Ażeby zjawiska życiowe mogły stanowić podłoże do gry zdarzeń, podobnej losowaniu, warunki zajścia zdarzeń muszą być zbliżone. Ustosunkowanie się zdarzeń będzie posiadało cechy stałości, charakterystyczne dla gier losowych, podporządkowujących się rachunkowi prawdopodobieństwa, kiedy nastąpi zachowanie szeregu okoliczności, z których najważniejsze są: dostateczna masowość (liczebność) zjawisk i losowość zjawisk, t. j. ich niezależność wzajemna.

Liczne doświadczenia ze zjawiskami zbiorowemi ustaliły, że tylko wielka liczba tych doświadczeń pozwala ustalić pewną niezmienność, czyli stałość stosunków, panujących w sferze tych zjawisk. Niezawsze mamy do czynienia z danemi obserwacji całej masy. Częstość poprzestajemy na zbadaniu określonej jej części. Jedną z metod dochodzenia statystycznego, mianowicie metoda Kiaera, jak już nam wiadomo, polega na badaniu części masy w celach wnioskowania o całej masie. Ażeby w badanej części masy mogło znaleźć przejaw prawdopodobieństwo statystyczne, należy, ażeby ona była dostatecznie wielka, gdyż w przeciwnym wypadku stałego ustosunkowania się zdarzeń nie zaobserwujemy. Oznaczenie ilości nie może być, jak z tego wynika, wzięte na ślepo, lecz odpowiadać charakterowi badanej masy i prawdopodobieństwu statystycznemu, t. j. takiemu prawdopodobieństwu, które zabezpiecza stałość stosunków.

Zjawiska winny być niezależne, t. j. samoistnie obok siebie stojące. O ile w grach losowych, np. w kombinacjach wy-

ciągnąć kul z urny, mamy zagwarantowaną niezależność poszczególnych zdarzeń, jednakową możliwość i równowartość ich, jako przyjęte a priori założenie naszych doświadczeń, o tyle w zjawiskach zbiorowych życia absolutne uniezależnienie zdarzeń wzajemnie od siebie jest trudne do urzeczywistnienia. Badamy np. jednostki gospodarcze. Czy możemy przypuścić, że istnienie jednych nie jest uzależnione od wpływu drugich? Dlatego przewidywanie nastąpienia zdarzeń na podstawie prawdopodobieństwa statystycznego jest znacznie trudniejsze, aniżeli na podstawie prawdopodobieństwa matematycznego, a to ze względu na charakter badanych zjawisk.

Metody obserwacji zjawisk, dla których istnieje prawdopodobieństwo matematyczne, są z natury rzeczy ujednolajnione (doświadczenia z kulami w urnie, z rzutami monet i t. p.), czego nie można powiedzieć o metodach spostrzegania zjawisk statystycznych. Praktyka badań statystycznych daje wiele przykładów stosowania odmiennych metod zbierania materiału statystycznego wbrew kategorycznemu żądaniu teorii. Centralizacja urzędów statystycznych ma na celu ujednolajnienie planów badania, przeprowadzenia dochodzeń i opracowania. Jest to warunek niezbędny do tego, ażeby między zjawiskami gry losowej a zjawiskami zbiorowymi istniało podobieństwo.

Teoria prawdopodobieństwa w zastosowaniu do zagadnień statystycznych polega na sprowadzeniu zdarzeń jednego i tego samego rodzaju do pewnej liczby wypadków jednakowo możliwych z jednoczesnem określeniem tych wypadków, które sprzyjają danemu zdarzeniu. O zajściu danego zjawiska my sądzimy według liczby sprzyjających wypadków dla danego zjawiska, zwracając uwagę, po której stronie jest więcej szans: po stronie sprzyjających czy niesprzyjających szans dla danego zdarzenia. Sąd więc nasz o prawdopodobieństwie zdarzenia jest względny; decyduje o nim wyobrażenie o przewadze dodatnich lub ujemnych warunków. Dlatego wszystkie prawdy, osiągnięte na podstawie rozumowań o prawdopodobieństwie wypadków, nie są prawdami bezwzględными, a tylko zbliżają się do prawd powszechnych.

Prawdy o charakterze względnie dowolnych uogólnień należy uważać za szczebel do poznania prawd powszechnych jako wynik tymczasowego naszego stanu poznania. Nie mniej jednak są one ściśle naukowe w specjalnych dziedzinach badania. Temi są nauki o społeczeństwie. Tutaj wystarcza nam, że „większość” ludzi zachowuje się w ten, a nie inny sposób, że „większość” ludzi posiada te, a nie inne cechy. Wskutek wielostronności i różnorodności zjawisk w sferze życia ludzkiego, a stąd płynącej trudności badania, poprzestajemy na uogólnieniach w myśl prawdopodobieństwa, które to uogólnienia zasługują na całkowite zaufanie i pod względem wartości naukowej nie ustępują wynikom i innym dziedzin wiedzy. W naukach społecznych dostępne jest dla nas osiągnięcie dostatecznego przybliżenia do prawdy, wynikającego z charakteru i treści badanych zjawisk. Teoria prawdopodobieństwa pozwala zmierzyć ilościowo wartość naukową tego przybliżenia, t. j. ścisłość naukową, oraz wielkość błędu.

Zagadnienie
wypadku.

Odpowiedź na pytanie — co to jest wypadek? — znajdujemy również w teorii prawdopodobieństwa.

Z punktu widzenia tej ostatniej, według Czubera, wypadek to są „te zmienne warunki, które wobec ich zmienności, są niedostępne dla naszego poznania”, a zjawisko przypadkowe jest to takie, którego „zajście nie może być z pewnością oznaczone na zasadzie znanych nam warunków lub znanych przesłanek”.

Mówiąc „wypadek”, nie odrzucamy istnienia związków przyczynowych, lecz uznajemy, że związek ten albo niedostępny jest naszemu poznaniu, albo dalej jak do prostego skonstatowania dotrzeć nie możemy. Winniśmy odróżniać: 1) spotkanie się dwóch niezależnych od siebie zjawisk i 2) zetknięcie się sił dowolnych, przypadkowych z jednej strony z kompleksem sił ogólnego charakteru z drugiej strony.

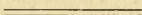
Z pierwszego punktu widzenia, według Wildenbanda, wypadek jest to zetknięcie się faktów, pozbawionych jakiegokol-

wiek łączności ze sobą. Naprz., śmierć Curie pod kołami wozu ciężarowego. Nie mamy możliwości objaśnienia przyczyny jednoczesnego zajścia faktów: udawania się do laboratorium prof. Curie i najechania na niego wozu ładownego.

Inny przykład daje nam pojęcie o działaniu przyczyny ogólnej w kombinacji z przyczynami dowolnymi. Pod wpływem podrożenia cen na zboże zwiększają się liczby przestępstw. Przestępstwa spełniają nie wszyscy ludzie jakiejś jednej kategorii (np., ludzie najbiedniejsi w społeczeństwie), lecz pewne indywidua pod wpływem jeszcze dodatkowych przyczyn. Nieznajomość tych przyczyn stwarza fakt wypadku, mówimy o wypadku kradzieży, morderstwa, samobójstwa i t. p.

Wypadek, jako taki, w przyrodzie nie istnieje, Istota jego polega na tem, że „zbiorowość warunków dostępnych dla naszego poznania nie ma niezbędnego po sobie następstwa albo nienastępstwa zjawiska (Kries)”.

Wypadki mogą, rzecz prosta, interesować statystyków, ale tylko jako zjawiska zbiorowe, indywidualnie znaczenia nie posiadają.



VIII. Opracowanie naukowe.

Analiza naukowa.

Cykl opracowania statystycznego nie kończy się na zebraniu materiału statystycznego, na technicznym jego opracowaniu oraz poszukiwaniu miar statystycznych, bowiem niezbędnym ogniwem w pracy statystycznej jest wnioskowanie, t. j. *analiza naukowa*. Analiza dat statystycznych służy do zrozumienia współzależności między zjawiskami, do odkrycia ich wzajemnego przyczynowego związku, do stwierdzenia stopnia stałości zjawisk, dalej do t. zw. porównania statystycznego i wreszcie do stworzenia uogólnienia, zastępującego w statystyce „prawa” innych, bardziej ścisłych nauk. Stosowanie tych lub innych wzorów matematycznych i obliczeń obraca się w granicach techniki i metodyki opracowania, wszakże nie stanowi rdzenia dociekania statystycznego. Słusznie powiada *Poincaré* w swej „Nauce i hipotezie”, że to, do czego dociera nauka „nie są to rzeczy same, jak sądzą naiwni dogmatycy, lecz *stosunki między rzeczami*; poza temi stosunkami niema rzeczywistości poznawalnej”. Konieczna jest zatem umiejętność i odpowiedni kierunek wnioskowania. Naukowe opracowanie materiału statystycznego winno dać możliwość przeniknięcia w dziedzinę istotnych stosunków, panujących między zjawiskami, ustalić działanie tych lub innych przyczyn, warunkujących zjawiska. Jedna liczba stała, powiada inny badacz, więcej jest warta, aniżeli steki literatury. Stwierdzone przez statystykę fakty stanowią prawa empiryczne zdarzeń, które w pewnych dziedzinach życia społecznego są bardzo cenne.

Naprz., w dziedzinie ruchu ludności ustalono większą liczbę urodzeń chłopców, aniżeli dziewcząt, ustalono dla danego miejsca i czasu względnie stałe współczynniki urodzeń, zgonów, małżeństw, przestępstw, ludzi nienormalnych, obarczonych brakami fizycznymi, określono współczynniki spożycia cukru, wódki i t. p. Liczby *stałe* są niezbędną charakterystyką zjawisk masowych.

Wszakże nie każda liczba statystyczna, otrzymana w wyniku przeprowadzonego badania, może pretendować do roli prawa empirycznego. Określenie jej „stałości”, czyli naukowego znaczenia, a zarazem powszechnego znaczenia — może dać analiza naukowa, poparta należytem wnioskowaniem.

Metody logiki	W statystyce, jak i we wszystkich innych naukach, posługujemy się ogólnie przyjętymi prawami myślenia, t. j. prawami logiki. Metody indukcji, mające zastosowanie w rozumowaniu statystycznym, są następujące: metoda zgodności, metoda różnicy, metoda równoległych zmian i metoda reszty. Ze względu na charakter zjawisk statystycznych, grupowanych przy opracowaniu statystycznym w szeregi wzajemnie od siebie uzależnione, najczęstszem zastosowaniem cieszy się metoda zmian równoległych, aczkolwiek znaczenie naukowe posiadają i inne przytoczone metody.
--	---

Metoda zgodności	Jeżeli dwa lub więcej przypadków pewnego zjawiska mają tylko jedną okoliczność wspólną, to okoliczność ta, w której jedynie wszystkie przypadki są zgodne, stanowi przyczynę danego zjawiska (Mill).
---	---

Jeżeli posiadamy następujący szereg faktów:

- S_1 (ABCD) jest P
- S_2 (AKIL) jest P
- S_3 (AMNO) jest P
- S_4 (ATQR) jest P i t. d.,

to rozpatrując te fakty we wzajemnej zależności, widzimy, że tylko składnik A jest stale obecny we wszystkich wypadkach, gdzie spotykamy P. Orzekamy, że A jest powiązane z P pewnym związkiem koniecznym.

----- Metoda różnicy. -----	Tenże <i>Mill</i> metodę różnicy formuluje tak: „jeżeli przypadek, w którym dane zjawisko występuje, i przypadek, w którym ono nie występuje, są zgodne we wszystkich okolicznościach z wyjątkiem jednej, która zjawia się tylko w pierwszym przypadku, to okoliczność ta, stanowiąca jedyną różnicę obydwóch wypadków, jest albo skutkiem, albo przyczyną, albo konieczną częścią przyczyny danego zjawiska”.
---	---

Przypuśmy, że

ABCDE jest P BCDE nie jest P A jest P
--

Wnioskujemy, że A jest P dlatego, że ani B, ani C. ani D, ani E nie mogą być połączone związkiem koniecznym z P wobec nieobecności jego w drugim sądzie.

Metoda różnicy znajduje powszechne zastosowanie w badaniach przyrodniczych, t. j. tam, gdzie może mieć miejsce dowolne doświadczenie. Istota doświadczenia przyrodniczego polega na tem, że obserwując wynik działania pewnego kompleksu sił w dwóch wypadkach, staramy się w jednym z nich wyłączyć działanie tej siły, której wpływ pragniemy poznać. Jest to eliminowanie tego warunku, który mierzymy. Naprz., przy sztucznem żywieniu roślin w okolicznościach dowolnego eksperymentu mając na celu zbadanie wartości azotu, jako środka pokarmowego dla roślin, zakładamy doświadczenie, w którym rośliny w jednym wypadku otrzymują wszystkie składniki odżywcze, w drugim zaś wszystkie — oprócz azotu. W obu wypadkach oprócz różnicy w dawce azotu wszystkie inne warunki powinny być, w myśl logicznego założenia, jednakowe.

Ta równowartość wszystkich innych warunków, oprócz badanego, jest trudna do osiągnięcia w okolicznościach badania przyrodniczego, a jeszcze trudniejsza w badaniach życia społecznego. Nawet życie przyrody, choćby to miały być martwe kryształ górskie, nie jest tak jednostajne, jak to się wydaje początkującemu przyrodnikowi. Pszczoła do pszczoły nie jest podobna, nie jest podobny motyl do motyla, zarówno jak nie podobny liść do liścia. Tembardziej jest różnorodne, indywidualne życie społeczne, życie ludzkie. Izolowanie w życiu społecznym jakiegoś jednego czynnika przy warunku zachowania jednakowych i niezmiennych wszystkich innych czynników jest zgola niemożliwością. Dlatego ściśle doświadczenie, jeżeli niekiedy mogłoby znaleźć zastosowanie w badaniach życia społecznego, musiałaby mieć z natury rzeczy niezmiernie ograniczone pole zastosowania. Dlatego też i metoda różnicy w badaniu statystycznym, jeżeli nie jest całkowicie wykluczona, może mieć zastosowanie bardzo ograniczone.

Na możliwość i celowość posiłkowania się metodą różnicy w statystyce w nowszej literaturze polskiej kładzie szczególnie nacisk dr. *Maciejewski* w swej „teorii statystyki”.

Dr. *Maciejewski* najpierw zauważa, jak stwierdziliśmy wyżej, że w doświadczeniu fizycznym również jest niemożliwe wytworzenie absolutnej tożsamości warunków, podobnie jak w doświadczeniu społecznym. Dalej uważa on za mylny ten pogląd, jakoby nie można było znaleźć w życiu socjalnym dwóch grup identycznych, gdyż, zdaniem jego, prawidłowości statystyczne dowodzą wręcz przeciwnego; wreszcie powołuje się na przykład rosyjskiego ministra komunikacji, który w dziedzinie taryf kolejowych zastosował doświadczenie, wprowadzając podwyższenie taryfy biletów kolejowych tytułem próby na okres dwuletni. Czas doświadczalny był dostateczny, ażeby wykazać, że próba prowadziła do ujemnych wyników. Opierając się na powyższem rozumowaniu dr. *Maciejewski* kategorycznie stwierdza, że „wszystkie dowodzenia przeciwników stosowania sposobu różnicy w statystyce dają się najzupełniej obalić”.

Sam dr. Maciejewski przeprowadzał badania statystyczne metodą różnicy w dziedzinie medycyny ginekologicznej z dobrym skutkiem.

W celu możliwości wykorzystania wszystkich zalet tej metody oraz jako warunek jej zastosowania dr. Maciejewski wprowadza do nauki pojęcie grup statystycznych, które dzieli na równomierne i nierównomierne. Pojęcie to wyjaśnia na przykładzie pracy lekarzy, zwłaszcza chirurgów, którzy często „zestawiają grupę bardzo ciężkich przypadków z grupą mniej ciężkich albo łatwych przypadków”. Jak widzimy dr. Maciejewski ma na myśli zdarzenia różnej wagi, zaznaczając jednocześnie, że niema grup bezwzględnie równomiernych. „Pomimo to — pisze on na str. 72 cyt. pracy — znaczenie równomiernych grup jest bardzo wielkie, albowiem one służą nam jako miara wszystkich grup zestawianych, a więc takich, które w rzeczywistości istnieją. Wedle tej miary wszystkie grupy zestawiane mogą być podzielone na 3 kategorie: na grupy, które mniej lub więcej zbliżają się do równomiernych, na grupy nieokreślone pod względem równomierności i na grupy nierównomierne pomiędzy sobą”.

Tamże znajdujemy następującą charakterystykę „grup”: „Do grup I kategorii odnoszą się grupy, obejmujące cały materiał statystyczny, które nazywają się grupami pełnymi, takimi są naprz., grupy ludności całego państwa albo grupy jakiegokolwiek miasta wielkiego z kilku sąsiednich okresów czasu”.

„Grupy pełne często zawierają pewien nadmiar przypadków, jakim naprz., jest nadmiar cudzoziemców w grupach ludności. Odejmując od grup pełnych podobne nadmiary otrzymamy grupy, które nazywam grupami całkowitemi. Rzecz oczywista, że grupy całkowite będą bardziej równomierne, niż grupy pełne”.

„Gdy cały materiał statystyczny podzielimy wedle cechy, która ściśle się ustanawia, naprz., grupy ludności wedle płci, wieku albo zajęcia, wtedy otrzymamy grupy mniejszej objęto-

ści, które przy dostatecznej ilości przypadków również będą zbliżać się do grup równomiernych. Jeśli od jednej kategorii takich grup odejmiemy rzadkie przypadki, które zmniejszają równomierność grup, wtedy otrzymamy grupy bardziej równomierne, niż grupy poprzednie. Otóż wszystkie wyżej wskazane grupy zaliczamy do grup I-ej kategorii na zasadzie ich budowy analogicznej”.

„Do grup tejże kategorii odnosimy również grupy, które stale dają prawie jednakowe wyniki, t. j. grupy, wytwarzające prawidłowości statystyczne. Na tej zasadzie przyszedłem do wniosku, że grupy żołnierskie bardzo łatwo wytwarzają grupy równomierne, a więc bardzo naginają się do badania statystycznego związków przyczynowych”.

„Wreszcie do grup I kategorii odnosimy grupy, w których czynniki, bardziej silnie działające na wyniki ogólne, są sobie równe albo równoważą się pomiędzy sobą”.

„Oto sposoby, które dają nam możliwość rozpoznawać grupy, zbliżające się do równomiernych. Rzecz oczywista, że grupy, które bardzo różnią się ogólną budową, albo dają bardzo różne wyniki ogólne albo znowu grupy, w których czynniki analogiczne ani zbliżają się ku sobie, ani równoważą się drugimi czynnikami, muszą być traktowane jako grupy nierównomierne. Takimi są najczęściej grupy operacyj chirurgicznych; albo znowu grupy, z których jedna przedstawia sobą grupę całkowitą, zaś druga stanowi drobną część takiej grupy”.

„Grupy, o których nie mamy możliwości sądzić, że zbliżają się do równomiernych, albo, że są nierównomierne pomiędzy sobą, muszą być traktowane, jako nieokreślone pod względem równomierności”.

Teoria „grup statystycznych” pozwoliła dr. Maciejewskiemu na zastosowanie metody różnicy w opracowaniu wyników operacyj chirurgiczno - ginekologicznych w zakresie badania statystycznego, przytem w wypadku zestawienia grup równomiernych dr. Maciejewski uważa wnioski za najzupełniej pewne, w wypadku, jeśli zestawiamy grupy, które zbliżają się do równomiernych, to możemy wyprowadzać wnioski prawdopo-

dobne. Jeśli przy posilkowaniu się metodą różnicy grupy nierównomierne bywają porównywane jako grupy równomierne albo jako grupy, zbliżające się do równomiernych, to wówczas wnioski zawsze będą fałszywe.

Wogóle dr. Maciejewski uważa, „że w badaniu statystycznym związków przyczynowych predominuje sposób różnicy (str. 81)”.

Metoda zmian
równoległych.

Określenie tej metody jest następujące: wszelkie zjawisko, które zmienia się w określony sposób, kiedy inne zjawisko ulega odpowiedniej zmianie, jest albo przyczyną tego zjawiska, albo jego skutkiem, albo wreszcie jest z niem połączone jakimkolwiek pośrednim związkiem przyczynowym.

Jeżeli przy zmianie kompleksu ABC na kompleks $A_1 BC$ zauważamy, że i odpowiadający pierwszemu kompleksowi kompleks abc ulega zmianie na kompleks $a_1 bc$, to możemy być pewni, że A jest przyczyną a . Naprz., badając warunki, od których uzależniony jest ruch przestępstw, popełnionych w różnych okresach czasu przez niektóre jednostki w społeczeństwie, porównywujemy odnośne liczby statystyczne z liczbami, dotyczącymi ruchu cen na zboże i znajdując, że oba szeregi liczb zmieniają się w jednakowym kierunku, czyli są zbieżne, t. j. liczba przestępstw wzrasta jednocześnie ze wzrostem cen na zboże i zmniejsza się przy ich spadku, konstatujemy, że pomiędzy wspomnianymi szeregami faktów istnieje związek przyczynowy.

Metoda
reszty

Niech abc jest wynikiem kompleksu ABC i niech zostało ustalone metodą różnicy, że a jest wynikiem lub zostało wywołane przez A , b przez B , to w takim razie możemy być pewni, że c zostało wywołane przez C . Reguła więc tej metody jest następująca: należy ze zbiorowości zja-

wisk odjąć tę jej część, która według poprzednich badań jest wynikiem pewnych znanych czynników; reszta będzie wywołana działaniem pozostałych czynników.

Spostrzeganie statystyczne

podobnie jak doświadczenia w naukach przyrodniczych posiłkuje się indukcją. W gruncie rzeczy nie ma różnicy między obserwacją statystyczną a doświadczeniem w tym znaczeniu, że doświadczenie zastosowane do zjawisk negatywnych, nietypowych, musiałoby być *wielokrotnie powtarzane*, by wnioski, płynące z doświadczenia mogłyby być uogólniane. Obserwacja statystyczna jest to właśnie notowanie „doświadczeń”, wywołanych nie wolą ludzką, lecz samoistnych, niezależnych od woli człowieka, w sferze zjawisk zmiennego charakteru, a właściwie tej strony zjawisk, która z reguły nie posiada cech gatunkowości. Statystyka, jako metoda, opiera się na indukcji, posiłkuje się bowiem rozumowaniem indukcyjnym do zorientowania się w systematycznie nagromadzonych faktach. Fakty winny być gromadzone i grupowane według metod statystyki, gdyż w przeciwnym wypadku myślenie indukcyjne nie wiele by pomogło w zrozumieniu związków przyczynowych między zjawiskami.

Zadania opracowania naukowego.

Posiłkując się tą lub inną metodą wnioskowania, stosując te lub inne obliczenia rachunkowe, oparte na elementarnej lub wyższej matematyce, pragniemy przez metodę statystyczną osiągnąć następujące trzy zadania opracowania naukowego: 1) stwierdzić *trwałość* stosunków w dziedzinie spostrzeganych zjawisk, wyrażającą się w stałych cyfrach statystycznych, 2) przeprowadzić *porównanie statystyczne* między spostrzeganiami faktami lub masami społecznymi i 3) sformułować *uogólnienie statystyczne*, mające wszelkie cechy prawa empirycznego. Te konkretne zadania narzucają się przede wszystkim jako cel każdego poszukiwania statystycznego. Natomiast samo pytanie: *dlaczego?*, t. j. dlaczego naprz., dzieje się

w życiu społecznym tak, a nie inaczej — dotąd zostaje w statystyce bez odpowiedzi. Jak dotąd, istota rzeczy jest dla nas niepoznawalna, nauka bada jedynie stosunki między rzeczami.

Stałość stosunków.

Opracowanie materiału statystycznego, dotyczące danej masy społecznej, według metod grupowania statystycznego, doprowadza do odkrycia pewnej *stałości* liczb statystycznych. Owa stałość liczb statystycznych, posiadająca poważną wartość naukową, jest względna i nie może być porównana z niezmiennością stosunków w innych, naprz. przyrodniczych gałęziach wiedzy ludzkiej. Jeżeli naprz., dla porównania sięgniemy do chemii, to w razie, gdy jest mowa o składzie chemicznym wody — H_2O , rozumiemy, że skład cząsteczki wody jest stały, niezmienny, t. j. na cząsteczkę wody składają się 2 atomy wodoru i 1 atom tlenu — bez względu na warunki otaczające. Względność stałości liczb statystycznych polega na tem, że spostrzegamy ją nie zawsze i nie wszędzie, a tylko dla pewnych okresów czasu i w granicach pewnych określonych warunków miejsca. Powyższa okoliczność stwarza fakt, że stałe liczby statystyczne ulegają wahaniom już to oscylując w pewnych względnie niewielkich granicach, już to zdradzając ruch postępowy w tym lub innym kierunku.

Stałość stosunków została udowodniona przedewszystkiem w najważniejszej dziedzinie statystyki — w demografii. Liczby statystyczne, t. zw. współczynniki urodzeń, zgonów, małżeństw, przyrostu naturalnego ludności, ludzi obarczonych brakami fizycznymi, jak to: ślepych, głuchych, nienormalnych umysłowo i t. p. są liczbami stałymi. Badając dane statystyki kryminalnej Quetelet był uderzony niezmiennością liczb przestępstw, wprowadzając do nauki, zresztą przejściowo, swoje słynne penchant au crime. Statystyka dostarczyła nauce dowodów, że istnieją *prawidłowości społeczne* i że świat zjawisk życiowych człowieka podlega surowym prawom na wzór tych praw, które działają w sferze zjawisk fizycznych.

Ale nawet dziedziny bardzo już dowolnych, indywidualnych zjawisk za pomocą opracowania statystycznego wykazują zadziwiające prawidłowości. *Schwittau* przytacza według Statistical Abstract liczby spożycia herbaty i napojów alkoholowych w Anglii za lata 1891 — 1900:

	% spożycia na głowę ludności									
	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900
Piwa	30.2	29.8	29.6	29.5	29.6	30.8	31.3	31.8	32.6	31.6
Wódki	1.03	1.03	0.98	0.97	1.00	1.02	1.03	1.04	1.09	1.12
Wina	0.39	0.38	0.37	0.35	0.37	0.40	0.39	0.41	0.41	0.38
Herbaty	5.35	5.43	5.40	5.51	5.65	5.75	5.79	5.83	5.95	6.07

Podobnych przykładów możnaby przytoczyć bardzo dużo z każdej dziedziny badania statystycznego. Nawet dziedziny wprost egzotycznie dowolnych zjawisk, tak dowolnych, jak opadanie liści z drzewa, wykazują charakterystyczne prawidłowości. Naprz., z 2 dworców kolejowych w Berlinie odjechała następująca liczba dorożek:

Rok	1900	1901	1902
I dworzec	5105	5738	5945
II dworzec	1352	1306	1341

Liczba pustych dorożek, odchodzących z głównego dworca w Berlinie bez pasażerów, była w tych latach prawie jednakowa.

Liczba zastawionych zegarków w berlińskim lombardzie:

Rok	1899	1900	1901	1902	1903
% wśród innych przedmiotów	16,40	16,02	16,71	16,75	17,12
przeciętna cena zegarka w markach	21,40	21,50	21,15	21,01	21,26

Stołość liczb statystycznych występuje tutaj, jak widzimy, z podziwu godną prawidłowością.

Stwierdzenie stołości stosunków w zjawiskach społecznych posiada zrozumiałe znaczenie praktyczne. Ubezpieczenia społeczne i prywatne opierają się na ustalonych liczbach wy-

padków śmierci, kalectw, niezdolności pracy i t. d. Organizacja szpitalnictwa i pomocy lekarskiej opiera się na liczbie zachorowań i ruchu śmiertelności. Organizacja szkolnictwa — na liczbie dzieci, wymagających nauki i t. p. Uogólniając, można śmiało powiedzieć, że cały gmach współczesnej budowy społecznej i ekonomicznej opiera się na podobnej stałości stosunków.

Zagadnienie stałości liczb statystycznych sięga głębokich teoretycznych rozważań. Sam fakt zaobserwowanego zjawiska stałości winien znaleźć teoretyczne uzasadnienie. Najwięcej szans uznania w charakterze naukowej podstawy dla określenia warunków stałości liczb statystycznych posiada teoria prawdopodobieństwa.

Porównanie statystyczne.

Porównanie statystyczne dokonywa się za pomocą t. zw. wielkości statystycznych, czyli miar, które metoda statystyczna uczy poszukiwać i celowo zastosowywać. Zjawiska w celu ich zrozumienia winny być mierzone i porównane ze sobą. Tylko przez porównanie serji dat statystycznych przychodzimy do wniosku, czy są one stałe czy nie.

W statystyce mamy do czynienia z liczbami absolutnymi, jako z pierwotnym materiałem statystycznym, które również posiadają samodzielne znaczenie naukowe; jednakże dla analizy naukowej znacznie większe znaczenie mają liczby średnie i liczby względne, których poszukiwanie i umiejętność zastosowania podaje nam właśnie teoria statystyki. Liczby średnie i liczby względne, jako miary porównania statystycznego, w rozwoju nauki ulegają udoskonaleniu i różniczkowaniu celem coraz bardziej właściwego ich zastosowania stosownie do natury zjawisk badanych.

Arsenał liczb, którymi rozporządza statystyka, jest już w tej chwili dość bogaty. Na czele kroczy, jak i przedtem, średnia arytmetyczna.

Wartość średniej arytmetycznej, jako miary porównania statystycznego, jest dostatecznie uznana choćby przez fakt najwię-

kszego rozpowszechnienia we wszelkich opracowaniach naukowych. Słusznie jednakże powiada Lièsse, że średniej arytmetycznej nie tylko się używa, ale i nadużywa. Słabe strony wewnętrznej treści średniej arytmetycznej polegają na jej abstrakcyjności i częstym odbieganiu od rzeczywistości. Oderwany charakter średniej arytmetycznej powstaje wskutek procesu zamiany szeregu rozbieżnych dat statystycznych jedną liczbą, stanowiącą wartość szeregu w przecięciu bez uwzględniania różnic rozpięcia między datami, t. j. przez zastosowanie średniej arytmetycznej do grup pełnych. Wskutek tego powstaje również odbieganie jej od rzeczywistości. Jeżeli naprz., do pojęcia przeciętnego dochodu społecznego gminy X zastosujemy średnią arytmetyczną, to tylko w tym wypadku da ona prawdziwe wyobrażenie rzeczywistości, jeżeli dochód społeczny gminy X stanowi grupę jednolitą. Wystarczy bowiem wprowadzić do grona obywateli gminy X miliardera z jego nieproporcjonalnie wielkim w stosunku do innych obywateli dochodem, ażeby dochód przeciętny przesunął się ku wyższej granicy. W tym wypadku odpowiedniejsze będzie zastosowanie, jako miary porównania, średniej topologicznej, która tutaj będzie znacznie bliższa rzeczywistości, aniżeli średnia arytmetyczna.

W zależności od natury zjawisk lepsze wyniki, aniżeli średnia arytmetyczna, daje również średnia modalna. Weźmy dla przykładu zjawisko cen targowych. Ażeby otrzymać przeciętną cenę towaru X na danym terytorjum ekonomicznym, jako średnią arytmetyczną należałoby wziąć notowania wszystkich zawartych transakcji, uwzględnić cenę każdej transakcji i podzielić sumę obroconych w ten sposób pieniędzy przez ilość towaru. Ale bliższą rzeczywistości będzie cena towaru X za którą dokonano największej liczby transakcji. W stosunkach życiowych tę właśnie cenę kupiec uważa za przeciętną.

Podobnie rzecz się ma z zastosowaniem średniej geometrycznej, średniej harmonicznej i t. p. O zastosowaniu tej lub innej miary porównania statystycznego decyduje *natura* zjawisk.

Porównanie statystyczne bardzo często daje się przeprowadzić za pomocą liczb względnych, które nazywamy współ-

czynnikami statystycznymi. Jeżeli wiemy, że na 1000 ludności rodzi się 28 jak w Polsce, to stosunek $\frac{28}{1000}$, czyli 0,028 nazywamy współczynnikiem urodzeń. W podobny sposób otrzymujemy współczynniki śmiertelności, małżeństw, samobójców etc.

Współczynniki statystyczne stanowią zwykłą miarę porównania statystycznego w zjawiskach społecznych. Cechą charakterystyczną jest ich względna stałość. Jeżeli naprz., ludność kraju wzrasta, a razem z nią wzrasta liczba zmarłych, to stosunek zmarłych do liczby ludności zmienia się tylko nieznacznie, o ile warunki ogólne pozostają mniej więcej te same. Powyższa względna stałość współczynników statystycznych staje się bardzo cenna przy badaniach porównawczych z różnych terytoriów. Ponadto nieznaczne postępowe zmiany współczynników statystycznych pozwalają na ustalenie kierunku zmian społecznych, które za pomocą badań statystycznych mogą być w ten sposób przewidziane na dłuższy okres czasu naprzód. Przez porównanie statystyczne odkrywamy *tendencje* rozwoju społecznego.

Charakter porównania występuje ze szczególną siłą w index numbers, czyli we wskaźnikach. Liczby wskaźnicze proste lub łańcuchowe stają się niezastąpione w badaniach statystycznych nad życiem ekonomicznym, w szczególności w ustaleniu konjunktur gospodarczych. Porównanie staje się tutaj głównym narzędziem analizy naukowej.

Jak widzimy statystyczne porównanie polega na tem, ażeby zjawiska złożone, masowe sprowadzić do niewielkich liczb, nadających się do zestawienia i ocenienia. O kompleksach zjawisk sądzimy z ich najprostszych charakterystyk. Każda dziedzina życia społeczno - ekonomicznego ma swoje urobione pojęcia porównawcze. Przez analizę naukową staramy się liczby wielkie, masy, zbiorowości — rozłożyć na pojęcia prostsze, doprowadzając je do stopnia maksymalnej i celowej porównawczości.

Uogólnienie, czyli prawo statystyczne.

Opracowanie naukowe prowadzi
poza to do tego, aby z masy
faktów, pojedynczych, konkretnych

zjawisk odnaleźć wszystko to, co jest danym zjawiskom wspólne, co za pośrednictwem jakiejś wspólnej cechy łączy te zjawiska w jedną zbiorowość mniej lub więcej jednolitą. Statystyka bardziej, niż inne nauki, jest nauką *uogólniającą*, pomimo, że uogólnienia te nie posiadają ścisłości i dokładności, które charakteryzują naprz., nauki przyrodnicze. Uogólnienia nauk przyrodniczych, z natury rzeczy ścisłych, stanowią *prawa* ogólne, powszechne, zawsze i wszędzie jednakowo działające. Zaś prawa, wynikające z uogólnień statystycznych, są prawami faktów, które posiadają mniejszy lub większy zakres pewności, są prawdopodobne, ale nie absolutnie pewne, jako że warunkowane okolicznościami miejsca i czasu. Prawa statystyczne dopiero w określonych warunkach, o ile takowe nastąpią, mogą pozyskać znaczenie praw powszechnych. Ale takich „praw” statystyka zna dotąd niewiele.

Najczęściej powołują się, jako na prawo statystyczne, na fakt nadwyżki urodzeń chłopczyków w stosunku do liczby urodzeń dziewczynek. Do tej samej kategorii prawidłowych zjawisk społecznych zaliczają również porządek wymierania ludności według grup wieku; dalej prawidłowości spożycia, znane pod nazwą „prawa Engela” i t. p. Jednakże fakty, stwierdzone statystycznie, zależą zbyt wyraźnie od warunków miejsca i czasu, dlatego możemy je kwalifikować jedynie jako „*prawidłowości społeczne*”, dalekie od powszechności praw przyrody.

Prawa przyrody odznaczają się prostotą, niezależnością od miejsca i czasu i powszechnością, t. j. zastosowaniem w podobnych wypadkach. Prawidłowości społeczne, stanowiące treść uogólnienia statystycznego, są bardzo skomplikowane i posiadają wagę dla tego miejsca i czasu, dla których zostały ustalone. Niezbędnym warunkiem ujawnienia ich jest wielka liczba spostrzeżeń, przeprowadzona nad zjawiskami zbiorowymi.

Wyższość praw przyrody nad „prawami” statystycznymi rzuca się sama przez się w oczy. Jeżeli jednak zastanowić się nad „niezmiennością” praw, rządzących światem przyrody, nad wspaniałą formą takich uogólnień, jako prawo ciężenia, prawo zachowania energii, prawo ustosunkowania w rozwoju i t. d., które wznoszą się jak wielkie kolumny w gmachu wiedzy ludzkiej, to na tych kolumnach dostrzeżemy rysy, przekonywające nas, że poziom naszej myśli naukowej we wszystkich dziedzinach wiedzy ludzkiej nie jest jeszcze dostateczny, ażebyśmy mogli „niezmiennosc” praw uważać za ostateczną. Kiedy genjusz *Mendelejewa* dał światu tablicę pierwiastków, pierwiastki te uważane były za „niezmienne”. Postęp nauki dowiódł fałszywości tego twierdzenia, jako zasady. Podobnie i niektóre inne niezmiennie prawa fizyczne uległy właściwшему zrozumieniu, wykazując, że pochod ludzkości ku zdobyciu prawdy nie jest zakończony.

Wiara w istnienie niewzruszonego układu społecznego pozwoliła w połowie XIX najwybitniejszemu statystykowi nowszych czasów Quetelet'owi na pojmowanie statystyki jako „fizyki społecznej”. Sądził on, że życie społeczne rządzi się takimi samymi żelaznymi prawami, jak świat fizyczny. Ale już dawno zauważono, że liczby statystyczne dają zrozumienie, jak jest rządzony świat, ale światem nie rządzą. Tem nie mniej naukowe zadania statystyki są takie same, jak innych, bardziej ścisłych nauk. Statystyka nie tylko konstatuje fakty, jako prawa empiryczne, ale dąży do wyjaśnienia związku przyczynowego między zjawiskami, t. j. do wyjaśnienia zależności koniecznej między określonymi warunkami a określonymi zdarzeniami.

Warunkowość działania prawa statystycznego pochodzi od natury badanych zjawisk. Statystyka bada nie pojedyncze przedmioty, indywidualne zjawiska, lecz masy, zbiorowości, społeczeństwa. Tutaj wystarcza wiedzieć, że *w większości* wypadków następuje to, a nie inne zjawisko. I dlatego wartość uogólnienia statystycznego bynajmniej nie jest mniejsza od

wartości prawa fizycznego. Ponadto ważnem jeszcze jest i to, że statystyka, dążąc do uogólnienia naukowego, pozwala poznawać prawdy dotąd nieznane i w stosunku do takich zjawisk, które tylko metodą wielkich ilości poznać można.

IX. Miary statystyczne.

Charakterystyka miar.

Dane statystyczne powstają przez liczenie. Liczymy bądź różne wielkości fizyczne, względnie określone cechy różnych wielkości fizycznych, bądź stopnie wartości cechy jednej wielkości fizycznej. W pierwszym wypadku powstają dane, stanowiące statystykę cech, w drugim wypadku mamy do czynienia ze statystyką zmiennych. Metody liczenia, stosowane jak w jednej, tak w drugiej, są w wielu wypadkach wspólne.

W statystyce zmiennych, przy analizie dat, powstaje konieczność zastąpienia szeregów pojęciami prostszymi, stanowiącymi cechy zbiorowe szeregów. Cecha zbiorowa jest to wyrażenie proste, uogólniające w jednym wyrazie liczbowym wartość zmiennej. Przez cechę zbiorową dokonywamy mierzenia wartości szeregu; jest więc ona wielkością lub miarą statystyczną, którą należy przy analizie poszukiwać.

Zmierzyć szereg zmiennej znaczy odnaleźć jej wartość, co nazywamy określeniem miary *położenia*, zmierzyć rodzaj i stopień zmienności zmiennej — znaczy znaleźć miarę dyspersji, czyli *rozsiania*.

Do miar położenia zaliczamy znane nam wielkości średnie: średnią arytmetyczną, średnią geometryczną, średnią harmoniczną, średnią topologiczną i średnią modalną. Najprostsze zaś miary rozsiania stanowią: odchylenie arytmetyczne, odchylenie kwadratyczne i odchylenie ćwiartkowe.

Do innych miar statystycznych należeć będą miary współzależności, czyli korelacji, oraz miary błędów, stanowiące,

ogólnie biorąc, t. zw. rachunek wyrównawczy. Zajmiemy się kolejno metodami *najprostrze*mi poszukiwania miar statystycznych, mając na uwadze cel podobnego poszukiwania, wskazany powyżej.

Wielkości średnie. Średnia arytmetyczna, którą ozna-
Średnia arytmetyczna. czamy przez sa , jest to suma

 wyrazów szeregu, dzielona przez
 ich liczbę. Jeżeli mamy do czynienia z szeregiem, w którym
 poszczególne wyrazy posiadają różną liczebność, obliczanie
 średniej arytmetycznej polega na przemnożeniu każdego wyrazu
 przez jego liczebność, podsumowaniu iloczynów i podzielenie
 otrzymanej sumy przez sumę liczebności. Podobnie postępujemy
 z szeregami, w których przeprowadzono podział na klasy.
 Aby obliczyć średnią arytmetyczną szeregu rozdzielczego
 mnożymy wielkość każdej klasy przez jej liczebność, a dalej
 postępujemy jak w poprzednim wypadku. Przykład.

Obszar gospodarstw w morgach	wielkość klasy	liczba gospodarstw	
v	v_1	f	$v_1 f$
0 — 12	6	83	488
12 — 24	18	46	828
24 — 36	30	34	1020
36 — 48	42	26	1092
48 — 60	54	11	594
		<u>200</u>	<u>4032</u>

$$sa = \frac{4032}{200} = 20.16$$

Wybór przedziału klasowego nie jest bez znaczenia. Po-
 żądaniem jest, aby przedziały klasowe były równe, jak to
 widzimy na naszym przykładzie. Podział jednak na nierówne
 klasy może być wywołany naturą zjawisk i celem do jakiego
 dążymy przy opracowaniu. Podobne znaczenie posiada również
 wielkość przedziału klasowego. W tym wypadku rozmieszczanie

liczebności w szeregu, czyli szereg liczebności zmienia się i wpływa na wyniki obliczenia. Przykład:*)

Rozmieszczenie liczby małżeństw na 10.000 ludności prawosławnej w 50 guberniach Rosji z r. 1903 według różnych przedziałów klasowych:

v	f	v	f
— 56	1	— 56	1
56 — 66	3	56 — 71	6
66 — 76	4	71 — 86	9
76 — 86	8	86 — 101	21
86 — 96	14	101 — 116	13
96 — 106	17		
106 — 116	3		

Średnia arytmetyczna w pierwszym wypadku = 89,8, w drugim = 90,2, przy ścisłym obliczeniu winna równać się 89,5.

Wybór przedziału klasowego nie jest również bez znaczenia przy wyobrażeniach graficznych szeregu rozdzielczego. Wielkości klas nanosimy w odpowiedniej skali na osi odciętych, a następnie wykreślamy rzędne, równe liczebnościom, przyjmując, że wszystkie spostrzeżenia zawarte w klasie mają średnią wartość przedziałów klasowych. Linja łamana, łącząca punkty szczytowe, tworzy dżagram linjowy, zwany *wielobokiem liczebności*. Wygląd wieloboku liczebności, jak to się łatwo przekonać, zależy od wyboru wielkości przedziału klasowego, gdyż kilkuszczytowe wieloboki liczebności przy odpowiednim doborze podziału klasowego można zamienić na jednoszczytowe.

Przy obliczaniu średniej arytmetycznej metodą przedziałów klasowych otrzymujemy rezultat o przybliżonej wartości, który w użyciu praktycznym posiada dostateczną dokładność, o ile wzięte przedziały klasowe nie są zbyt wielkie.

*) R. Orzechki. *Uczebnik matematycznej statystyki*, S. - Peterburg, 1914, str. 52. W oryginale pierwszy szereg liczebności podano błędnie, co tutaj prostujemy.

Metoda
momentów.

Zauważamy, że mnożenie wielkości klas przez liczebności przy skomplikowanych szeregach rozdzielczych

może być kłopotliwe. Dlatego obliczanie średniej arytmetycznej staramy się uprościć *metodą Pearsona*, zwaną metodą momentów. Obieramy dowolnie wielkość jakiegokolwiek przedziału klasowego, najczęściej jednego ze środkowych szeregu, za wartość wyjściową i oznaczamy ją przez O . Zaczynając od wartości wyjściowej oznaczamy odległości klas przez liczby początkowe całkowite 1, 2, 3... Te odległości, które od wartości wyjściowej idą w górę oznaczamy przez liczby ujemne, idące w dół — przez dodatnie. Wówczas wielkości klasy zostaną zamienione przez liczby całkowite początkowe, odpowiadające odległościom klasy od wartości wyjściowej. Każdą wielkość klasy, sprowadzoną do liczb małych, mnożymy przez odpowiednią jej liczebność, iloczyny dodajemy i dzielimy przez sumę liczebności. Otrzymujemy w ten sposób średnią arytmetyczną odległości, która nosi nazwę pierwszego momentu, czyli wielkości potęgowej pierwszego stopnia. Na tej odległości od wartości wyjściowej winna znajdować się poszukiwana średnia arytmetyczna. Ponieważ odległość między wielkościami klas przyjęliśmy za 1, a w rzeczywistości równa się ona przedziałowi klasowemu, winniśmy pierwszy moment pomnożyć przez przedział klasowy i otrzymaną wielkość dodać lub odjąć (w zależności od znaku) od wartości wyjściowej. W wyniku otrzymamy średnią arytmetyczną.

v klasy	v_1 wielkość klasy	f liczebność	z odległość od wartości wyjściowej	$f \cdot z$
0—12	6	83	— 2	— 166
12—24	18	46	— 1	— 46
24—36	30	34	0	
36—48	42	26	1	26
48—60	54	11	2	22
		200		— 164

$$\text{Pierwszy moment } z_1 = \frac{-164}{200} = -0.82; \quad sa =$$

$$= v_0 + z_1 \cdot i \text{ (przedział klasy)} = 30 + (-0.82 \cdot 12) = \\ = 30 - 9.84 = 20.16.$$

Jak widzimy średnia arytmetyczna obliczona metodą momentów jest identyczna z otrzymaną poprzednio za pomocą metody przedziałów klasowych.

Jeśli w przytoczonym przykładzie każdy iloczyn $f \times z$ pomnożyć jeszcze raz przez z , czyli w iloczynie $f \times z$ zamiast z wziąć z^2 , to przeprowadzając obliczenie, analogiczne do poprzedniego, otrzymamy z_2 , czyli drugi moment. Jeżeli z zastąpić przez z^3 , to otrzymamy z_3 , czyli trzeci moment i t. d.

Metoda dodawania.	Metoda dodawania, zwana inaczej metodą Czubera, opiera się na tej zasadzie, że wynik działania arytmetycznego jest jednakowy, czy jakąś wartość mnożymy, czy dodajemy do siebie tyle razy, ile jedność zawiera w sobie mnożnik (reguła mnożenia). Przy obliczeniach, liczebności kolejno dodajemy do siebie aż do przedziału klasowego, w którym zawiera się wartość wyjściowa, postępując jednakowo dla obydwóch części szeregu, podzielonego wartością wyjściową na 2 odcinki. Sumy liczebności z kolei sumujemy. Z otrzymanych sum otrzymujemy różnicę, którą dzielimy przez sumę liczebności i mnożymy przez wielkość przedziału klasowego. Rezultat będzie stanowił różnicę między średnią arytmetyczną a wartością wyjściową. Bierzemy ten sam przykład.
----------------------	--

v	v_1	f	suma liczebności kolejno dodawanych
0—12	6	83	83
12—24	18	46	129
24—36	30	34	0
36—48	42	26	37
48—60	54	11	11
		200	

Dodajemy sumy liczebności poniżej wartości wyjściowej ($83 + 129 = 212$) i powyżej wartości wyjściowej ($11 + 27 = 48$), traktując pierwszą jako ujemną, drugą jako dodatnią, dzielimy rezultat przez sumę liczebności szeregu (200) i mnożymy przez wielkość przedziału klasowego (12). Otrzymujemy

$$\frac{-212 + 48}{200} \cdot 12 = \frac{-164}{200} \cdot 12 = -0,82 \cdot 12 = -9,84.$$

A zatem

$$sa = 30 - 9,84 = 20,16.$$

Dla wyjaśnienia praktycznego zastosowania tej metody przytoczymy jeszcze jeden przykład, zaczerpnięty z pracy prof. Moszczeńskiego.*)

Obliczenie średniej arytmetycznej metodą dodawania z pomiarów długości ziarn fasoli.

Długość ziarn fasoli w mm.	Liczba ziarn w klasowych przedziałach	Sumy liczebności oznaczeń w przedziałach
17,5	5	5
18,5	13	18
19,5	38	56
20,5	41	97
21,5	95	192
22,5	124	316
23,5	152	468
		— 1152
24,5	134	
25,5	129	398
26,5	100	269
27,5	0	169
28,5	45	99
29,5	38	54

*) Stefan Moszczeński, prof. Praca cytowa, str. 73.

30.5	7	16
31.5	7	9
32.5	2	2
		+ 1016

Wartość wyjściowa = 24.5.

Różnica między wartością wyjściową a średnią arytmetyczną =

$$= \frac{-1152 + 1016}{1000} \cdot 1 = -0.136 \quad sa = 24,5 - 0,136 = 24,36.$$

Wielkość topologiczna
czyli mediana.

Medianą, jak już nam wiadomo, nazywamy wartość środkową szeregu, która dzieli szereg na dwie równe części. Z tego względu obliczenie mediany w szeregu szczegółowym nie nastręcza trudności, gdyż w szeregu nieparzystym jest ona centralnym wyrazem, powyżej i poniżej którego leży jednakowa ilość wyrazów, a w szeregu parzystym równa się średniej arytmetycznej dwóch wyrazów środkowych. Inaczej nieco przedstawia się obliczanie mediany z szeregów rozdzielczych. Jeżeli chodzi o obliczenie z przybliżoną dokładnością postępujemy w myśl następującego rozumowania. Dodajemy kolejno liczebności szeregu nie dochodząc do takiej sumy, która przekracza połowę sumy liczebności. W następnym przedziale klasowym winna mieścić się mediana. Określamy w ten sposób granice, między którymi zawarta jest mediana, przytem pierwsza liczba wskazanego przedziału będzie dolną granicą mediany, ostatnia — górną. Ażeby znaleźć położenie mediany od początku przedziału klasowego, należy znaleźć jaka część przedziału klasowego przypada na jednostkę liczebności. W tym celu dzielimy wartość przedziału klasowego przez liczebność, odpowiadającą tej klasie, wynik mnożymy przez liczbę jedności, których brakowało od otrzymanej sumy liczebności do $\frac{1}{2}$ wszystkich liczebności. Rezultat stanowi proporcjonalną część przedziału klasowego, którą należy dodać do początku przedziału klasowego.

Przykład zapożyczamy od prof. Orzęckiego.*)

v	f	sf
— 56	1	1
56 — 66	3	4
66 — 76	4	8
76 — 86	8	16
86 — 96	14	
96 — 106	17	
106 — 116	3	
	50	

Odnajdujemy połowę sumy liczebności: $\frac{1}{2} \cdot 50 = 25$;
 $25 - 16 = 9$, t. j. do połowy sumy liczebności brak 9 jedności.
 Mediana znajduje się w przedziale 86 — 96. W tym przedziale
 mamy 14 liczebności. Na 1 przedziału klasowego przypada
 $\frac{10}{14}$ liczebności, na 9 jedności — $\frac{10}{14} \cdot 9 = 6.4$
 Mediana $st = 86 + 6.4 = 92.4$.

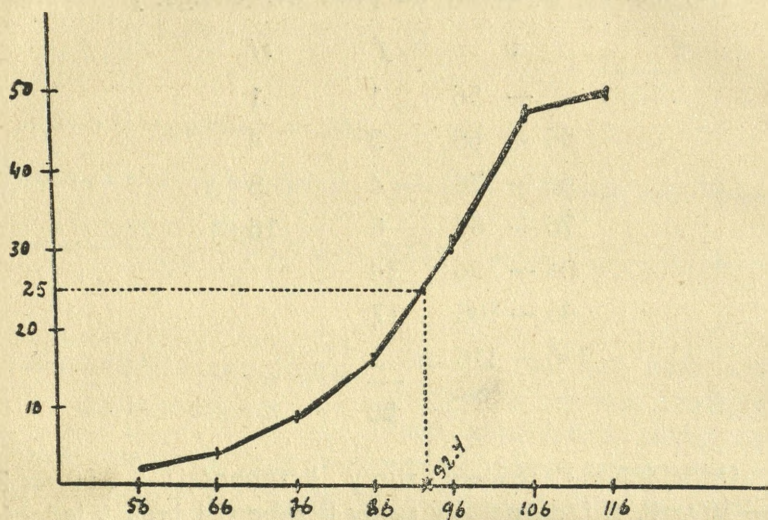
Metoda
graficzna.

Prof. Orzęcki podaje graficzną me-
 todę poszukiwania mediany, posił-
 kując się tym samym przykładem.

Na oś odciętych nanosimy przedziały klasowe; sumy liczebno-
 ści wyobrażamy jako rzędne w końcowym punkcie każdego
 przedziału. Jeżeli na osi rzędnych weźmiemy punkt odpowia-
 dający $\frac{1}{2}$ sumy liczebności i z tego punktu przeprowadzimy
 linię równoległą od osi odciętych, to linia ta natrafi na linię
 łamaną, łączącą szczyty rzędnych. Z punktu przecięcia należy
 opuścić prostopadłą na oś odciętych, która wskaże nam war-
 tość mediany.

*) R. Orzęcki. Praca cytowana, str. 58.

Poszukiwanie mediany metodą graficzną.



Modus.

Modus jest to wielkość najczęstsza, inaczej zwana wielkością normalną.

W szeregach szczegółowych poszukiwanie modusa nie nastrocza trudności. Jest to ten wyraz szeregu, któremu odpowiada max. liczebności. Wszelako dla szeregów rozdzielczych, które najczęściej spotykamy w opracowaniach statystycznych, takie proste obliczenie wielkości modalnej byłoby zbyt uproszczone. Nie wystarcza tutaj wyznaczenie wielkości przedziału klasowego, posiadającego max. liczebności, gdyż jak wiemy wybór przedziału klasowego jest dowolny i powoduje zmienność grupowania szeregu liczebności. Zbyt małe przedziały klasowe powodują powstanie dwu lub więcej szczytów nieraz o tej samej największości; zbyt wielkie przedziały klasowe zacierają znowu szczyty, co również nie jest pożądaną.

Elementarny sposób poszukiwania wielkości modalnej, oznaczamy ją przez sm , wyprowadza się ze związku między wielkością modalną a średnią arytmetyczną i medianą dla szeregów łagodnie asymetrycznych, z którymi najczęściej się

spotykamy. W wypadkach szeregów idealnie symetrycznych wszystkie trzy wielkości pokrywają się wzajemnie. W miarę jak szereg zniekształca się, wielkości omawiane rozchodzą się, zajmując różne położenie w stosunku do siebie. Mianowicie, mediana leży na jednej trzeciej odległości od średniej arytmetycznej w kierunku wielkości modalnej.

$$sm = sa - 3(sa - st)$$

Inna metoda poszukiwanna sm polega na przekształcaniu szeregu liczebności w celu znalezienia najwięcej charakterystycznych szczytów, które osiągamy przez kolejne sumowanie liczebności sąsiednich klas po dwie, zaczynając od pierwszej, potem po dwie z opuszczeniem pierwszej, potem po trzy, zaczynając od pierwszej, potem po trzy z opuszczeniem pierwszej, z opuszczeniem dwóch pierwszych i t. d. Przykład.*)

ν	f	Kolejne sumy liczebności			
5	48	}	100	}	156
6	52				
7	56	}	116	}	168
8	60				
9	62	}	122	}	182
10	60				
11	58	}	118	}	180
12	56				
13	63	}	114	}	177
14	60				
15	48	}	123	}	179
16	40				
17	32	}	88	}	148
		}	72	}	120

Największy szczyt sumy liczebności otrzymaliśmy dla klas 8, 9 i 10. Wśród tych klas największą liczebność posiada klasa 9. Wnosimy, że 9 jest wartością modalną, mimo, że w szeregu,

*) Prof. Stefan Moszczeński. Praca cyt., str. 87.

liczebności f największą liczebność posiada klasa 13, ale koło tej klasy grupują się mniejsze liczebności.

Modus, podobnie jak i średnia arytmetyczna, zależy od wszystkich wyrazów szeregu, lecz staje się wyobrażeniem dominującego typu wielkości; jest on niejako najbardziej typową wielkością. Z tego względu związek jego z pozostałym obszarem szeregu jest dość luźny. Większą zależność od liczby wyrazów szeregu wykazuje mediana, która całkowicie jest niezależna od wartości poszczególnych wyrazów, lecz tylko od ich liczby. Dlatego w szeregach bardzo rozszaniach, o wielkim obszarze i dużych różnicach między poszczególnymi wyrazami — zupełnie właściwem staje się użycie wielkości topologicznej, bardziel właściwe, aniżeli średniej arytmetycznej lub modalnej.

Miary rozszania

Odchylenie arytmetyczne.

Miary położenia dają bardzo dużo dla zrozumienia liczb statystycznych, nagromadzonych w szeregu; jednakże charakterystyka zbiorowości, względnie *obszaru* zmiennej

byłaby niecałkowita, gdybyśmy nie mogli przeniknąć w wewnętrzny układ szeregu, w rodzaj rozszania, względnie grupowania się wyrazów dokoła jednej z miar położenia. Za miarę położenia, jako ośi rozszania, przyjmujemy w opracowaniach statystycznych zazwyczaj średnią arytmetyczną, aczkolwiek możemy wziąć jakąkolwiek inną przeciętną, naprz. wartość środkową. Różnice pomiędzy średnią arytmetyczną a poszczególnymi wyrazami szeregu tworzą odchylenia, które sumowane bez uwzględnienia znaków i podzielone przez liczbę składników dają *odchylenie arytmetyczne*, zwane również odchyleniem przeciętnem.

W obliczeniach dla szeregów rozdzielczych posługujemy się wartością wyjściową. Wychodząc od wartości wyjściowej, liczymy odchylenia poniżej środka przedziału i powyżej środka przedziału, w którego granicach mieści się średnia arytmetyczna. Ze względu na to, że odchylenia liczymy od wartości wyjściowej, a nie od średniej arytmetycznej, winniśmy sprowa-

dzić je do średniej arytmetycznej przez uwzględnienie różnicy między średnią arytmetyczną a wartością wyjściową, powtórzoną tyle razy, ile wynosi różnica członów szeregu liczonych poniżej i powyżej środka. Jeżeli przez d oznaczymy różnicę między średnią arytmetyczną a wartością wyjściową, przez N_1 ilość członów szeregu poniżej, a przez N_2 — powyżej środka, to do sumy odchyłeń należy dodać N_1 razy d i odjąć N_2d , czyli $(N_1 - N_2)d$. Rezultat dzielimy przez sumę liczebności. Ponieważ jednostką pomiaru uczyniliśmy przedział klasowy, przeto ażeby przejść do jednostek bezwzględnych, należy wziąć pod uwagę wielkość przedziału klasowego. Przykład.*)

środek wartości przedziału klasowego	liczebność	odchylenie od war- tości wyjściowej	
v	f	z	fz
1	18	— 5	90
1.5	48	— 4	192
2	72	— 3	216
2.5	89	— 2	178
3	100	— 1	100
			— 776
3.5	90	0	
4	75	+ 1	75
4.5	60	+ 2	120
5	40	+ 3	120
5.5	21	+ 4	84
6	11	+ 5	55
6.5	5	+ 6	30
7	1	+ 7	7
7.5	1	+ 8	8
8	—	+ 9	—
8.5	1	+ 10	10
	632		+ 509

*) G. Udny Yule. Wstęp do teorii statystyki. Przekład Z. Limanowskiego z 2-go wyd. ang. str. 132 i 173.

Średnia arytmetyczna $\bar{x} = 3.29$; $d = -0.42$ ($-776 + 509 = -267$; $d = \frac{-267}{632} = -0.42$). Suma odchyłeń ujemnych 776, suma odchyłeń dodatnich 509, razem 1285. Ponieważ $N_1 = 327$, a $N_2 = 305$, powtórzona ($N_1 - N_2$) razy różnica d równać się będzie $d(N_1 - N_2) = -0.42(327 - 305) = -0.42 \times 22 = -9.2$ przeto

suma odchyłeń od średniej arytmetycznej $= 1285 - 9.2 = 1275.8$. Stąd odchylenie arytmetyczne $= \frac{1275.8}{632} = 2.019$ przedziałów klasowych. Ponieważ przedział klasowy równa się 0.5, to odchylenie arytmetyczne równać się będzie $2.019 \times 0.5 = 1.01$.

Odchylenie kwadratowe.	Odchylenie kwadratowe otrzymujemy podnosząc poszczególne odchylenia członów szeregu od średniej arytmetycznej do drugiej potęgi i obliczając pierwiastek kwadratowy ze średniej arytmetycznej kwadratów odchyłeń. Jest to więc pierwiastek kwadratowy ze średniej arytmetycznej kwadratów odchyłeń.
---	--

Jeżeli przez x oznaczymy odchylenie od średniej arytmetycznej, przez f — liczebność, przez d — odchylenie arytmetyczne i δ — odchylenie kwadratyczne, to odpowiednie wzory będą następujące:

$$d = \frac{\sum fx}{\sum f} \quad \text{ i } \quad \delta = \sqrt{\frac{\sum fx^2}{\sum f}}$$

Obliczenia odchylenia kwadratycznego dokonywamy niezwyczajnie łatwo, mając pierwszy i drugi moment w systemie momentów Pearsona, gdyż

$$\delta^2 = Z_2 - Z_1^2.$$

Podobnie łatwo jest otrzymać odchylenie kwadratyczne, znajdując odchylenie arytmetyczne, bowiem

$$d = 0.797885 \delta.$$

Natomiast jeżeli chcemy wyliczyć odchylenie kwadratyczne z szeregów szczegółowych bezpośrednio, to o ile materiału

statystyczny składa się z większej ilości oznaczeń, proces wyliczania jest stosunkowo dość kłopotliwy. Przykład.

Obszar gospodarstw chłopskich w morgach.

v	x	x^2	v	x	x^2
0.5	— 4.4	19.36	5.3	+ 0.4	0.16
0.8	— 4.1	16.81	5.5	+ 0.6	0.36
1.2	— 3.7	13.69	5.9	+ 1.0	1.00
1.7	— 3.2	10.24	5.9	+ 1.0	1.00
2.1	— 2.8	7.84	5.9	+ 1.0	1.00
3.5	— 1.4	1.96	6.1	+ 1.2	1.44
3.8	— 1.1	1.21	6.2	+ 1.3	1.69
3.2	— 1.7	2.89	6.5	+ 1.6	2.56
3.3	— 1.6	2.56	7.4	+ 2.5	6.25
3.6	— 1.3	1.69	7.5	+ 2.6	6.76
3.8	— 1.1	1.21	7.5	+ 2.6	6.76
4.0	— 0.9	0.81	8.4	+ 3.5	12.25
4.6	— 0.3	0.09	8.8	+ 3.9	15.21
4.6	— 0.3	0.09	9.7	+ 4.8	22.04
4.8	— 0.1	0.01			
4.9	0	0			

$$Sf = 30, Sfx = 147, sa = 147:30 = 4.9, Sfx^2 = 158.94,$$

$$\frac{Sfx^2}{Sf} = \frac{158.95}{30} = 5.298 \quad \delta = \sqrt{5.298} = 2.3.$$

Wobec uciążliwości podobnych wyliczeń zwracamy się do sposobu uproszczonego za pomocą wartości wyjściowej z szeregu rozdzielczego.

Obieramy wartość środka jednego z przedziałów klasowych jako wartość wyjściową, od której mierzymy odchylenia. Odchylenia te w stosunku do odchylen od średniej arytmetycznej x będą się różniły o pewną wielkość d , równając się w ogólnej formie $x + d$, nazwijmy je przez w

$$w = x + d$$

Jeżeli obie części równania podniesiemy do kwadratu, to

$$w^2 = x^2 + 2 \times d + d^2$$

Sumując kwadraty odchyłeń, otrzymujemy

$$Sfw^2 = Sfx^2 + 2dSfx + Sfd^2$$

Suma odchyłeń od średniej arytmetycznej Sfx równa się jak wiadomo 0, wobec tego i $2dSfx = 0$, czyli

$$Sfw^2 = Sfx^2 + Sfd^2, \text{ a dalej}$$

$$\frac{Sfw^2}{Sf} = \frac{Sfx^2}{Sf} + d^2$$

Ponieważ

$\frac{Sfx^2}{Sf}$ jest to kwadrat odchylenia średniego, czyli δ^2

$$\frac{Sfw^2}{Sf} = \delta^2 + d^2$$

Wyrażenie $\frac{Sfw^2}{Sf}$ nazwijmy przez s^2 . Jest to średnia arytmetyczna kwadratów odchyłeń od wartości wyjściowej.

$$s^2 = \delta^2 + d^2 \text{ skąd}$$

$$\delta = \sqrt{s^2 - d^2}$$

Przykład poprzedni:

Obszar gospodarstw chłopskich w morgach, rozklasyfikowany w szeregu rozdzielczym.

v	f	z	fz	fz^2
0 — 1	2	— 4	— 8	32
1 — 2	2	— 3	— 6	18
2 — 3	1	— 2	— 2	4
3 — 4	6	— 1	— 6	6
4 — 5	5	0	— 22	
5 — 6	5	+ 1	+ 5	5
6 — 7	3	+ 2	+ 6	12
7 — 8	3	+ 3	+ 9	27
8 — 9	2	+ 4	+ 8	32
9 — 10	1	+ 5	+ 5	25
	30		+ 33	161

$$d = \frac{-22 + 33}{30} = + \frac{11}{30}; d^2 = \frac{121}{900}$$

$$s^2 = \frac{161}{30} = 5 \frac{11}{30}, \delta = \sqrt{5 \frac{11}{30} - \frac{121}{900}} = 2.27.$$

Przytoczony wzór dla poszukiwania odchylenia kwadratowego metodą uproszczoną jest obarczony błędem, zresztą bez praktycznego znaczenia, pochodzącym stąd, że w szeregu rozdzielczym przy obliczeniach przyjmujemy wszystkie spostrzeżenia w obrębie klasy za równe wartości środka przedziału klasowego, a więc za równe sobie. Ale tak nie jest. Im większy jest przedział klasowy, tym w obrębie jednej klasy różnice pomiędzy poszczególnymi spostrzeżeniami mogą być większe a stąd i błąd wzoru będzie większy. W celu usunięcia tego błędu Pearson i inni proponują wprowadzenie poprawki, która ze swej strony posiada również charakter przybliżonego obliczenia, wobec czego niema większej potrzeby zastanawiania się nad nią.

Wielkość odchylenia kwadratowego jest powszechnie używana jako dogodna, łatwa do obliczania miara rozsiania. Jest ona odwrotnie proporcjonalna do pierwiastka kwadratowego

z liczby spostrzeżeń, jak wynika z wzoru $\delta = \sqrt{\frac{Sfx^2}{Sf}}$. Ponie-

waż średnia arytmetyczna jest wynikiem serji spostrzeżeń, przeto ścisłość jej powinna być większa jak poszczególnego spostrzeżenia w stosunku $\sqrt{n}$, gdzie n oznacza liczbę spostrzeżeń, a kwadratyczne odchylenie średniej arytmetycznej sa równa się

$$\delta_{sa} = \frac{\delta}{\sqrt{n}}$$

Jeżeli dla jakiejś wielkości jednej odchylenie kwadratyczne wynosi δ_1 , a dla drugiej δ_2 , to dla sumy lub różnicy tych wielkości odchylenie kwadratyczne wyniesie

$$\delta = \sqrt{\delta_1^2 + \delta_2^2}$$

Jeżeli odchylenie kwadratyczne dla danej wielkości równa się δ , to dla wielkości n razy większej ono równać się będzie

$n\delta$, a dla wielkości n razy mniejszej wyniesie $\frac{\delta}{n}$.

Yule czyni uwagę, że odchylenie kwadratowe winno być zawsze stosowane jako miara dyspersji, o ile jakieś istotne przyczyny nie każą użyć innej miary, zupełnie jak przy średniej arytmetycznej, której się powinno używać jako miary położenia.

Jak wskazaliśmy na str. 141, procentowy stosunek odchylenia kwadratowego do średniej arytmetycznej nosi nazwę wskaźnika lub współczynnika zmienności.

Odchylenie ćwiartkowe.	Odchylenie ćwiartkowe, oznaczane symbolem Q, jest najprostszą miarą dyspersji. Mając do czynienia z szeregiem szczegółowym, dzielimy liczbę spostrzeżeń na 4 części, otrzymujemy 3 wyrazy dzielące szereg w równych częściach. Pierwszy wyraz Q_1 jest takim wyrazem, poniżej którego mamy $\frac{1}{4}$ spostrzeżeń a powyżej $\frac{3}{4}$. Drugi wyraz Q_2 jest wielkością środkową szeregu, dostatecznie nam znaną. Trzeci wyraz Q_3 ma przed sobą $\frac{3}{4}$ spostrzeżeń, po za sobą $\frac{1}{4}$.
-----------------------------------	--

$$Q = \frac{Q_3 - Q_1}{2}$$

Przykład.*)

Zarobki robotników
Okręgi w szylingach i pensach

	s.	d.
1.	12	6
2.	13	6
3.	14	4
4.	14	6
5.	14	6
6.	14	7
7.	14	9
8.	14	9
9.	14	9
10.	14	10
11.	15	0
12.	15	0

Mamy 38 spostrzeżeń. Wobec tego $38:4 = 9,5$. Wyraz, poniżej którego znajduje się 9,5 spostrzeżeń, jest wyraz 10-ty = 14 s. 10 d. (Q_1).

W ten sam sposób $Q_3 = 16$ s. 11 d.

$$Q = \frac{16 \text{ s. } 11 \text{ d.} - 14 \text{ s. } 10 \text{ d.}}{2} = 12,5 \text{ d.}$$

Yule, praca cyt., str. 164.

13.	15	0	Między odchyleniem ćwiartkowym, a odchyleniem kwadratowym istnieje stały matematyczny związek wyrażający się stosunkiem $Q = 0.675 \delta$ czyli odchylenie ćwiartkowe równa się mniej więcej $\frac{2}{3}$ odchylenia średniego.
14.	15	0	
15.	15	0	
16.	15	0	
17.	15	3	
18.	15	4	
19.	15	6	
20.	15	6	
21.	15	8	
22.	15	9	
23.	16	0	
24.	16	0	
25.	16	3	
26.	16	3	
27.	16	4	
28.	16	6	
29.	16	11	
30.	17	0	
31.	17	0	
32.	17	1	
33.	17	6	
34.	17	8	
35.	18	6	
36.	19	8	
37.	20	3	
38.	20	9	

Miary korelacji czyli współzależności.

Korelacją nazywamy wzajemną zależność, czyli współzależność zjawisk. Jeżeli naprz., przez badanie otrzymujemy dwa szeregi liczb, z których jeden wskazuje na użycie P_2O_5 jako pokarmu roślinnego, a drugi na wysokość plonów, to pomiędzy jednym a drugim możemy dopatrywać się związku, stwierdzając, że taki związek bądź istnieje, bądź nie istnieje.

Statystyczne ustalenie korelacji jest konieczne wtedy, kiedy związek dwóch zjawisk nie da się odkryć w każdym poszczególnym wypadku, lecz w dostatecznie wielkiej zbiorowości.

Analiza naukowa dwóch zjawisk, ulegających ilościowym zmianom w czasie i przestrzeni, idzie naprzód w kierunku

wykrycia *logicznej podstawy* możliwości istnienia korelacji, a potem dopiero określenia jej stopnia. Uzasadnienie logiczne wypływa z gruntownej znajomości przedmiotu. Jeżeli matematyk, rozporządzający całym arsenalem środków matematyczno-statystycznego badania, operuje w dziedzinie naprz., statystyki lekarskiej, to może mu zbraknąć zrozumienia istoty zjawisk, czego z reguły nie będzie u lekarza, dla którego zjawiska medyczne są skojarzone logicznymi związkami.

Poza logiczną stroną zależności, o której co dopiero mówiliśmy, ważny jest stopień, czyli ścisłość zależności. Określenie stopnia zależności stanowi jedno z bardzo ważnych zadań analizy statystycznej. Wyrażają go miary statystyczne, z których najprostszymi są: *linja*, *tablice* i *współczynnik korelacji*.

Miary współzależności winny przedewszystkiem oznaczyć kierunek zależności, względnie czy zależność jest na całej linii, czy jest zmienna, czy wreszcie niema żadnej zależności. Kierunek korelacji nazywamy dodatnim, jeżeli zestawione dwa szeregi wyrażone są przez liczby zbieżne, t. j. albo jednocześnie wzrastające, albo jednocześnie malejące; ujemnym, kiedy szeregowi liczb wzrastających odpowiada szereg liczb malejących i odwrotnie. W jednym i drugim wypadku mamy zależność *prostoliniową*. Jeżeli liczby obu szeregów w obrębie pola wahań mają kierunek zmienny, to znaczy z początku, dajmy, w jednym szeregu rosną, a w drugim maleją, później idą w jednym kierunku, dalej w pierwszym maleją, a w drugim rosną, to mówimy, że zależność jest *krzywoliniowa*. Wreszcie nieraz na oko można zauważyć, że pomiędzy liczbami obu szeregów niema najmniejszej prawidłowości, wnioskujemy, że współzależność między szeregami jest minimalna, bądź nie istnieje wcale.

Rozporządzając wymiarami dwóch zmiennych, należy je uszeregować celem przedwstępного porównania, gdyż, jak wynika z powyższego, samo porównanie stanowi już dość cenną analizę. Częstokroć uszeregowanie według max lub min, według średniej arytmetycznej, podział szeregów na kilka części z wyprowadzeniem średnich arytmetycznych i porównanie tych ostatnich, oznaczanie liczb numerami rzymskimi według me-

tody Kaufmana i t. p. ujawniają istnienie korelacji między szeregami i nawet w pewnym stopniu określają ścisłość związku przyczynowego.

Wszakże powyższe metody analizy szeregów nie mogą dać matematycznej miary współzależności. Za taką uważamy współczynnik korelacji.

Współczynnik korelacji.

Analizując dwa szeregi w celu otrzymania współczynnika współzależności, zauważamy, że odchylenia szeregów od średniej arytmetycznej będą posiadały te same własności, co i szeregi. Jeżeli kierunek szeregów jest zbieżny, to odchyleniom ujemnym jednego szeregu będą odpowiadały odchylenia ujemne drugiego szeregu, odchyleniom dodatnim — dodatnie. Jeżeli kierunek szeregów jest rozbieżny, to odchyleniom ujemnym jednego szeregu będą odpowiadały odchylenia dodatnie drugiego szeregu i vice versa. W tym wypadku, kiedy dla obu szeregów średnie arytmetyczne znajdują się w jednakowym punkcie pola wahań zmiennych, otrzymamy idealną równoległość znaków odchyłeń. Będą to znaki bądź jednakowe, bądź przeciwne — w zależności od dodatniego lub ujemnego kierunku szeregów. Jeżeli spróbujemy przemnożyć parami odchylenia, to w jednym wypadku wszystkie iloczyny będą dodatnie, w drugim ujemne. Znaki iloczynów odchyłeń posiadają zatem zasadnicze znaczenie: znaki dodatnie świadczą o dodatnim, znaki ujemne — o ujemnym kierunku szeregów.

Jeżeli pomiędzy szeregami niema wzajemnej zależności, to żadnej prawidłowości w rozmieszczeniu odchyłeń nie znajdziemy.

Wielkość iloczynów odchyłeń, mnożonych równolegle parami, będzie tem większa, im częściej większym odchyleniom jednego szeregu będą odpowiadały większe odchylenia drugiego szeregu; w razie mnożenia większych odchyłeń przez mniejsze lub odwrotnie wielkość iloczynów odchyłeń będzie mniejsza i suma iloczynów odchyłeń będzie mniejsza. Dlatego

największa suma iloczynów odchyłeń będzie, gdy szeregi porównywane będą posiadały jednakową strukturę w stosunku do punktu średniej arytmetycznej, oraz znak dodatni w wypadku dodatniego kierunku korelacji i ujemny — w wypadku ujemnego. Suma ta będzie mniejsza przy krzywoliniowej zależności szeregów i najmniejszą, dążącą do 0 w tym wypadku, kiedy między szeregami zależności nie będzie.

Ze względu na to, że szeregi mogą składać się z liczb mianowanych, przytem wyrażonych w różnych miarach, przyjęte jest, w celu uniknięcia przeprowadzania działań arytmetycznych z liczbami mianowanymi, brać zamiast absolutnych wielkości odchyłeń stosunek odchyłeń arytmetycznych do odchylenia kwadratowego, zastępując cyfry bezwzględne cyframi względnymi. Suma iloczynów liczb względnych, wyrażających stosunek każdego odchylenia do odchylenia kwadratowego szeregu, dzielona przez liczbę składników, stanowi współczynnik współzależności.

Matematyczna właściwość współczynnika korelacji polega na tem, że wielkość jego oscyluje w granicach + 1 i - 1. Jeżeli wielkość współczynnika korelacji zbliża się do + 1, mamy do czynienia ze ścisłą wzajemną zależnością szeregów między sobą o kierunku dodatnim; jeżeli zbliża się do - 1 mamy do czynienia z wypadkiem silnej zależności ujemnej; jeżeli zbliża się do 0, zależność jest nadzwyczaj nikła, przy wielkości 0 niema jej wcale.

Jeżeli przez $X_1, X_2, X_3... X_n$ oznaczmy odchylenia arytmetyczne jednego szeregu, a przez $Y_1, Y_2, Y_3... Y_n$ odchylenia drugiego; przez δ_x — odchylenie kwadratowe jednego szeregu, a przez δ_y — drugiego, zaś przez n ilość członów szeregu, to współczynnik korelacji r równać się będzie:

$$r = \left(\frac{X_1 Y_1}{\delta_x \delta_y} + \frac{X_2 Y_2}{\delta_x \delta_y} + \frac{X_3 Y_3}{\delta_x \delta_y} + + \frac{X_n Y_n}{\delta_x \delta_y} \right) \cdot \frac{1}{n} = \frac{S(XY)}{n \delta_y \delta_x}.$$

Przykład.

Mamy dwa szeregi: 1, 3, 5, 7, 9 oraz 2, 6, 10, 14, 18, które wyobrażają dwie cechy związane związkiem przyczyno-

wym. Zadajemy pytanie, jak ścisły jest ten związek? Widzimy z porównania szeregów, że korelacja jest prostolinijna, a więc współczynnik współzależności winien być dodatni. Jaka jest jego wielkość?

X	Y	x	y	x/δ_x	y/δ_y	$\frac{x \cdot y}{\delta_x \cdot \delta_y}$
1	2	-4	-8	-1.40	-1.41	+1.974
3	6	-2	-4	-0.70	-0.71	+0.487
5	10	0	0	0	0	0
7	14	+2	+4	+0.70	+0.71	+0.497
9	18	+4	+8	+1.40	+1.41	+1.974
						+4.942

Śr. arytm. szeregu X = 5; $\delta_x = \sqrt{40/5} = 2.83$

Śr. arytm. szeregu Y = 10; $\delta_y = \sqrt{160/5} = 5.66$

$$r = \frac{S(xy)}{n \cdot \delta_x \cdot \delta_y} = \frac{5}{4.942} = 0.988$$

Weźmy inny przykład, gdzie dwa szeregi posiadają luźną budowę i trudno sobie na pierwszy rzut oka zdać sprawę jaki jest kierunek korelacji.

X	Y	x	y	$\frac{x}{\delta_x}$	$\frac{y}{\delta_y}$	$\frac{x \cdot y}{\delta_x \cdot \delta_y}$
4	1	+1	-2	+0.71	-0.95	-0.6745
5	3	+2	0	+1.42	0	0
2	7	-1	+4	-0.71	+1.90	-0.1349
1	2	-2	-1	-1.42	-0.48	+0.6816
3	2	0	-1	0	-0.48	0
						-0.1278

$A_x = 3; \delta_x = 1.41$

$A_y = 3; \delta_y = 2.1$

$$r = - \frac{0.1278}{6} = - 0.02576$$

Znak przy r wskazuje na ujemny kierunek korelacji. Wielkość r natomiast jest miarą ścisłości związku korelacyjnego, który w danym wypadku, jak widzimy, jest dość słaby.

Linja
korelacji.

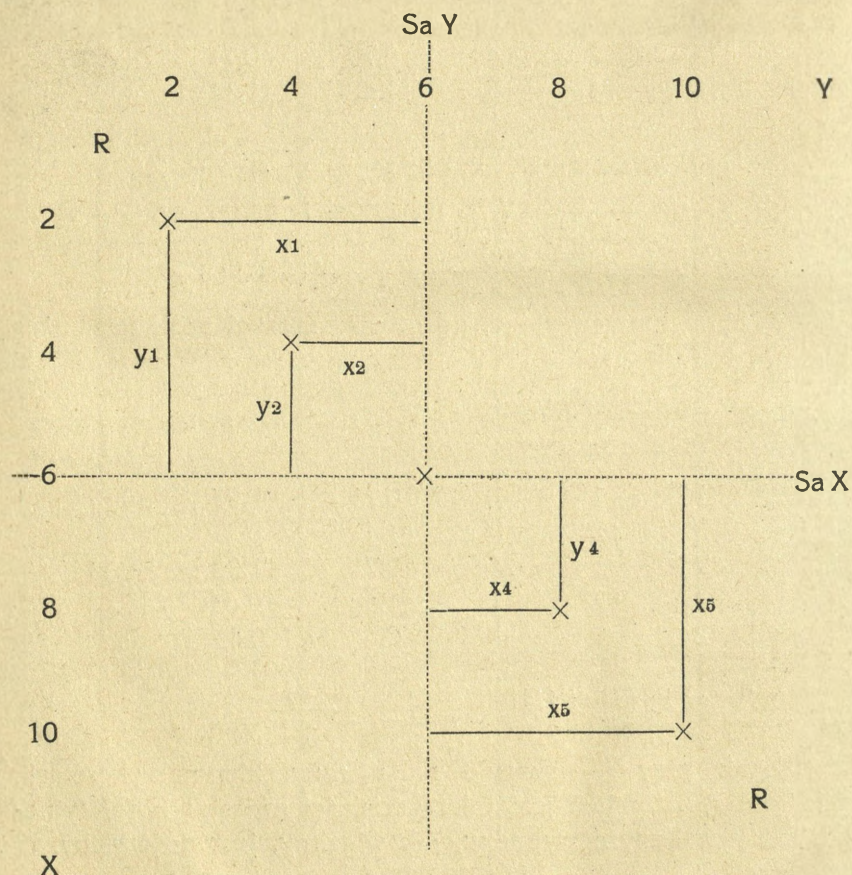
Dane liczbowe dwóch szeregów możemy wyobrazić graficznie. Weźmy przykład najprostsz, przedsta-

wiający dwa szeregi o idealnej zależności prostoliniowej. Niech dla szeregu X odchylenia od średniej arytmetycznej będą $X_1, X_2, X_3 \dots$, a dla szeregu Y odchylenia arytmetyczne od średniej arytmetycznej $Y_1, Y_2, Y_3 \dots$

Szeregi X — 2, 4, 6, 8, 10

— 2, 4, 6, 8, 10

pole układu linii odchyłeń będzie następujące:



Linie odchyłeń przecinają się wzajemnie w punktach, które oznaczamy krzyżykami. Jak widzimy z rysunku, wszy-

stkie punkty przecięcia leżą na prostej RR. Ma to miejsce w tym wypadku, kiedy szeregi są identyczne, bowiem przy całkowitej zgodności odchyłeń stosunek każdej pary odchyłeń równa się stałej wielkości, t. j. $\frac{X_1}{Y_1} = \frac{X_2}{Y_2} = \frac{X_3}{Y_3} = \dots$. Z chwilą jednak kiedy zależność między dwoma szeregami nie jest zupełna, punkty przecięcia każdej pary odchyłeń zbaczają z linii prostej to w jedną, to w drugą stronę. Otrzymujemy przez połączenie punktów linię łamaną. Należałoby przeprowadzić prostą, któraby przechodziła najbliżej wszystkich punktów przecięcia. Kreślenie takiej linii „na oko” posiada zrozumiałe wady, gdyż zawiera element subiektywny. Ażeby przeprowadzenie linii korelacji posiadało cechy obiektywnej oceny, stawiamy warunek, ażeby suma kwadratów odchyłeń rzeczywistych punktów przecięcia od zastępujących je punktów na linii RR była wielkością najmniejszą. Tę linię nazywamy *prostą regresji*.

Kąt, który tworzy linia RR z osiami rzędnych i odciętej, wskazuje na wielkość współzależności. Jeżeli RR idzie w kierunku wskazanym przez rysunek, korelacja jest całkowita i prostolinijna. Przy ujemnym kierunku korelacji linia RR zajmie położenie we wręcz przeciwnym kierunku. Przy braku korelacji linia RR jest prostopadłą do osi odciętych.

Tablice korelacji.

W praktyce statystycznej spotykamy się nie tylko z pojedynczymi wymiarami dwóch cech, kiedy łączymy parami odpowiadające sobie wielkości, lecz i z częstotliwością skojarzonych par, t. j., z liczebnością przypadków skojarzenia. Zmuszeni jesteśmy w tych wypadkach badać w jaki sposób częstotliwości jednej zmiennej łączą się z częstotliwościami drugiej zmiennej. Przytem wszystko jest jedno czy mamy do czynienia z szeregami szczegółowymi, czy rozdzielczymi. Układamy *tablicę korelacji* w ten sposób, że wymiary jednej cechy wypisujemy w nagłówku, wymiary drugiej — z boku, t. j. jedno w kierunku poziomym, drugie w kierunku pionowym. Utworzone kratki tablicy zapełniamy cyframi licze-

bności przypadków skojarzenia cech. Niech w naszym przykładzie wymiar 2 jednego szeregu spotyka się jeden raz z wymiarem 2 drugiego szeregu; niech wymiar 4 spotyka się dwa razy z wymiarem 4 i dwa razy z wymiarem 6; wymiar 6 — analogicznie siedem razy z wymiarem 6; wymiar 8 — dwa razy z wymiarem 6 i trzy razy z wymiarem 8 i wreszcie wymiar 10 — jeden raz z wymiarem 8 i dwa razy z wymiarem 10. Możemy dwa tak skojarzone szeregi ująć w tablicę korelacji podzieloną na kolumny (pionowe) i rzędy (poziome).

X \ Y	2	4	6	8	10	n_x	Y_{sa}
2	1					1	2
4		2	2			4	5
6			7			7	6
8			2	3		5	7.2
10				1	2	3	9.3
n_y	1	2	11	4	2	20	
X_{sa}	2	4	6	8.5	10		

W kolumnie n_x umieszczamy sumy liczebności szeregu X, otrzymane przez dodawanie poszczególnych liczebności w rzędzie; analogicznie w rzędzie n_y umieszczamy sumę liczebności szeregu Y. W ostatniej kolumnie wyprowadzamy wielkości średniej arytmetycznej Y_{sa} z odpowiedniego rzędu szeregu X; u dołu w rzędzie X_{sa} umieszczamy wielkości średniej arytmetycznej z odpowiednich kolumn szeregu Y.

Rozmieszczenie cyfr liczebności na tablicy korelacji posiada zasadnicze znaczenie, jeżeli przyjmiemy określony system

zapisywania szeregów. Zapisując wyrazy jednego szeregu w nagłówku tablicy, zaczynając od strony lewej, a wyrazy drugiego szeregu od góry ku dołowi, t. j. jedno po osi odciętych, drugie po osi rzędnych, winniśmy przy dodatnim kierunku zależności otrzymać przesuwanie się liczb od kąta, utworzonego przez szeregi, w stronę kąta przeciwnego. Przy ujemnym kierunku korelacji główna masa cyfr umieszcza się na linii, łączącej dwa inne kąty.

Wyrównanie błędów. Błąd przeciętny i średni.

Przeprowadzając spostrzeganie statystyczne nad jedną i tą samą wielkością fizyczną, z reguły nigdy nam się nie udaje osiągnąć prawdziwej wartości. Doświadczenie uczy, że każdy nowy pomiar przy wymaganiu idealnej ścisłości różni się od poprzedniego. Powstaje szereg liczb pomiarowych, z których każda znajduje się w pewnym przybliżeniu do wartości prawdziwej, ale nią nie jest. Jak wielkie jest to przybliżenie? Jaka jest miara tego przybliżenia? Temi zagadnieniami zajmuje się, ogólnie biorąc, rachunek wyrównawczy.

Kiedy przeprowadzaliśmy obserwację statystyczną nad wielkimi ilościami cechy zbiorowej, otrzymywaliśmy *odchylenia* od typu przeciętnego; kiedy mierzymy wielokrotnie jedną i tą samą wielkość fizyczną, otrzymujemy różne wielkości pomiarów, obciążone *błędami* w stosunku do rzeczywistej wartości. Błędem zatem nazywamy różnicę między rzeczywistą wartością a wynikiem obserwacji. Ale najczęściej rzeczywista wartość spostrzegania jest nam nieznana. Zmuszeni jesteśmy zamiast wartości prawdziwej przyjąć jakąś inną wartość zbliżoną, ale nie i dentyczną z wartością prawdziwą, i z nią porównywać wyniki spostrzegania. Różnica między wartością spostrzegania a wartością zbliżoną do wartości prawdziwej nie będzie stanowiła błędu *prawdziwego*, lecz błąd *pozorny*. Za wielkość zbliżoną do wielkości rzeczywistej przyjmujemy średnią arytmetyczną na tej zasadzie, że suma kwadratów odchyłeń od średniej arytmetycznej jest najmniejszością. Można dowieść, że średnia arytmetyczna z nieskończenie wielkiej liczby pomiarów, czyli

w granicy równa się wielkości prawdziwej. Z tego wynika, że błąd pozorny jest to różnica między wartością spostrzegania, a średnią arytmetyczną szeregu pomiarów.

Suma błędów, podzielona przez liczebność (t. j. ilość spostrzeżeń) daje błąd *przeciętny*, który jest identyczny z odchyleniem arytmetycznym, ale różni się od niego logicznym założeniem.

Podobnie jak od odchylenia arytmetycznego przechodziłiśmy do odchylenia kwadratowego, tak samo od błędu przeciętnego przechodzimy do błędu *średniego*, który różni się od odchylenia kwadratowego logicznym założeniem. W teorii błędów zastosowanie obydwóch powyższych miar przechyla się na korzyść błędu średniego, jako powszechnej miary ścisłości oznaczeń.

Obszar błędów.

Błąd średni przyjęty jako miara błędów może wskazać nam czy w granicach naszego spostrzegania popełniliśmy tylko błędy elementarne (porównaj: str. 162), czy też wkraść się do naszych pomiarów błąd systematyczny. Pojedyncze oznaczenia różniące się od wartości prawdziwej lub wartości zbliżonej do prawdziwej (średniej arytmetycznej) na wielkość błędu nie powinny wykraczać poza pewne granice. Przekroczenie pewnej dopuszczalnej granicy nastąpi, jeżeli błąd oznaczenia będzie za duży. Wielkość błędu powinna mieścić się w granicach trzechkrotnego błędu średniego, jeżeli je przekracza, należy szukać błędu grubego. Zatem 3δ stanowi *obszar* wahania błędów przypadkowych.

Jednak obszar wahania błędów jest względny. Przy wielkiej ilości spostrzeżeń obszar może być rozszerzony, a to ze względu na to, że w wielkiej liczbie przypadków jest większe prawdopodobieństwo pojawienia się błędu grubszego. Wielkość 3δ uważa się za obowiązującą przy mniej więcej 500 spostrzeżeniach. Przy 10 spostrzeżeniach za obszar wahania błędów przyjmuje się tylko $1,65\delta$, przy 10.000 spostrzeżeń — $3,9\delta$. Jeżeli posiadamy szereg liczb błędów, to należy sprawdzić za

pomocą błędu średniego czy poszczególne błędy nie wykraczają poza granice obszaru krytycznego w celu ewentualnego powtórzenia serii pomiarów.

Błąd średni
średniej arytmetycznej. Średnia arytmetyczna wielkiej ilości spostrzeżeń w granicy równa się wielkości prawdziwej. Atoli przy konkretnych obliczeniach z takim wypadkiem nie spotykamy się nigdy. Nie zawsze również spotykamy się z wypadkiem, kiedy błędy ujemne i błędy dodatnie posiadają jednakowe prawdopodobieństwo. Najczęściej szereg błędów jest asymetryczny, a wskutek tego średnia arytmetyczna będzie obciążona błędem.

Ażeby obliczyć *błąd średni średniej arytmetycznej* zwracamy się do wzoru błędu średniego (= odchyleniu kwadratowemu): $\delta = \sqrt{\frac{Sfx^2}{Sf}}$. Zauważamy, że błąd średni zmniejsza się

z powiększeniem się Sf, t. j. liczby spostrzeżeń. Jest on, jak to widać ze wzoru, odwrotnie proporcjonalny do pierwiastka kwadratowego z liczby spostrzeżeń (n). Ponieważ średnia arytmetyczna składa się z n spostrzeżeń, jest ona dokładniejsza $\sqrt{n}$ razy, aniżeli każde poszczególne spostrzeżenie, przeto błąd

średni średniej arytmetycznej $\delta_{sa} = \frac{\delta}{\sqrt{n}}$.

Błąd średni średniej arytmetycznej pozwala na obliczenie obszaru wahan, w którym mieści się prawdziwa wartość średniej arytmetycznej. Znaleziony błąd średni raz dodajemy do sa , drugi raz odejmujemy od sa , otrzymujemy górną i dolną granicę wahan prawdziwej średniej arytmetycznej.

Idąc za tokiem rozumowania, przyjdziemy do wniosku, że sam błąd średni obciążony jest błędem.

Błąd średni błędu średniego, nie przytaczając dowodu, oblicza się z wzoru: $\frac{\delta}{\sqrt{2n}}$

Błąd domniemany.

Ponieważ różnica między spostrzeżeniem, a średnią arytmetyczną jest w gruncie rzeczy błędem pozornym, a nie rzeczywistym, obliczamy błąd **domniemany**.

We wzorze na błąd średni $\left(\delta = \sqrt{\frac{\sum x^2}{n}} \right)$ w tym wypadku wprowadzamy poprawkę, pomijając dowód, mianowicie

$$\delta = \sqrt{\frac{\sum x^2}{n-1}},$$

gdzie X oznacza błąd spostrzeżenia, n — liczbę spostrzeżeń. Ponieważ błąd średni średniej arytmetycznej

$$\delta_{sa} = \frac{\delta}{\sqrt{n}}$$

przeto błąd domniemany średniej arytmetycznej równać się będzie:

$$\delta_{sa} = \sqrt{\frac{\sum x^2}{n(n-1)}}$$

Błąd prawdopodobny.

Od błędu średniego odróżniamy błąd **prawdopodobny** (oznaczamy go przez ϵ), pod którym rozumiemy taką wielkość, w obrębie której mieści się połowa wszystkich spostrzeżeń. Przy pomiarach, jak już wiemy, otrzymujemy różne wyniki, które możemy przedstawić graficznie jako rzędne na osi odciętych, tworzące powierzchnię, określoną matematyczną krzywą błędów. Stanowi ona całkowity obszar błędów. Za błąd prawdopodobny będziemy uważali taką wielkość, dla której prawdopodobieństwo, iż popełnione błędy są mniejsze od niej, wyraża się stosunkiem $\frac{1}{2}$. Połowa spostrzeżeń będzie leżała w granicach tej wielkości, druga połowa — po za jej granicami. Błąd średni jest nieco większy od błędu prawdopodobnego, gdyż obejmuje $\frac{2}{3}$ wszystkich spostrzeżeń. Z porównania wzoru na błąd średni z analogicznym wzorem na błąd prawdopodobny otrzymamy ścisły stosunek między ϵ i δ . Mianowicie $\epsilon = 0.6744897 \delta$.

W praktyce wystarcza posługiwać się relacją $\epsilon = 0.675 \delta$.

ŹRÓDŁA.

- Ancyferow: Kurs elementarnej statystyki, Charków, 1911.
 Ballod: Grundriss der Statistik. Berlin, 1911.
 Buzek: Spis ludności i zwierząt gospodarskich z dn. 30.IX.1921.
 Bachmetjew: Statistiko-analiticeskij metod i jego znaczenie „Siel-skoje Choz. i Lles.” t. 233.
 Bleicher: Statystyka. Tom. Dr S. Szulca. Warszawa, 1919.
 Bortkiewicz: O statisticzeskoj zakonomierności. „Wiest. Pr.”, 1905.
 Bortkiewicz: Das Gesetz der Kleinen Zahlen. Lipsk, 1898.
 Ciągliński: Zasady statystyki szpitalnej. Mies. Gl. Urz. Stat., 1920.
 Conrad: Grundriss zum Studium der politischen Oekonomie. IV. Teil. Statistik. Jena 1920.
 Czekanowski: Zarys metod statystycznych do antropologii. W. 1913.
 Czuber: Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung und Lebensversicherung. Lipsk, 1902.
 Czuprow (ojciec): Statistika. Posobie k lekcjam.
 Czuprow (syn): Oczerki po teorji statystyki, 1910.
 Danielewicz i Dickstein: Zarys arytmetyki politycznej. 1910.
 Faure: Elements de statistique. Paryż, 1906.
 Falski: Wyniki spisu dzieł z czerwca 1926 r: Warszawa, 1926.
 Fiedorowicz: Istoria i teorja statystyki. Odesa, 1894.
 Fortunatow: O statistikle. Moskwa, 1907.
 Gołogórski: Rachunek wyrównawczy. Poznań, 1927.
 Goslewski: Zasady rachunku prawdopodobieństwa. Warszawa, 1906.
 Górski: Stwierdzanie plonów w doświadczeniach polowych. 1918.
 Górski: Błąd prawdopodobny doświadczenia polowego. 1919.
 Grabowski: Podręcznik statystyki. Warszawa, 1917.
 Grabski: Zbiór prac ekonomiczno-rolniczych Warszawa, 1930.
 Hamburger: Prawidłowości wśród odchyleń. Warszawa, 1928.
 Haushofer: Wykład statystyki, przekład S. S i L. studentów prawa. Warszawa, 1874.
 Horowicz: Wstęp do statystyki teoretycznej. Warszawa, 1917.
 „Organizacja niższych urzędów statystycznych. Kraków 1918.
 Kabłukow: Statistika. Moskwa, 1915.
 Kaufman: Teorja i metody statystyki Petersburg, 1912.
 Kleczyński: Spisy ludności w Rzeczypospolitej Polskiej. Kraków 1892.
 „Poszukiwania spisów ludności Rzeczypospolitej Polskiej w zbiorach Moskwy, Petersburga. Wilna. Kraków, 1898.
 Krzywicki: Organizacja statystyki miejskiej. Warszawa, 1918.
 „Spis ludności i zwierząt gospodarskich z dn. 3.IX.1921.
 Kumaniecki: Studja z zakresu statystyki wędrówek. Kraków, 1909.
 Les forces économiques du monde, 1927.
 Maciejewski: Podręcznik statystyki. Teorja statystyki. Warszawa, 1925.
 Materiały k prodowolstwennomu planu. Proizwodstwo etc. chlebow w Rosii, 1916.
 Marschall: Zasady ekonomiki. Przełożył C. Znamlerowski. 1925.
 Mayr: Statistik und Gesellschaftslehre. Tübingen, 1914.
 „Zakonomierność w obščestwennoj żyzni. (przekł.) Moskwa, 1899.
 Meitzen: Geschichte, Theorie und Technik der Statistik. Stuttgart, 1903.
 Miller: O dokładności opracowania dat statystycznych przy pomocy maszyn systemu „Powers’a”. Warszawa, 1928.

- Moszczeński: Podstawy organizacji gospodarstw wiejskich. Cz. I. Metody statystyczne. Warszawa, 1924.
- Oboleński: Urożal chlebow w Jużnoj Rossii, Charków, 1915.
- Orzęcki: Uczebnik matematycznej statystyki. Petersburg, 1914.
- Pankow: Klimatycznej oczerk Nowej Aleksandrii. Petrograd, 1915.
- Pilat: Statystyka I Część ogólna i demologia. Lwów, 1923.
- Piekałkiewicz: Skróty wykładów statystyki. Opracował Henryk Dąbrowski, Lwów 1924.
- Piekałkiewicz i Rutkowski: Okręgi gospodarcze Polski. 1927.
- Predwarytelnyje Itogi wsierosijskoj sielsko choz. plerepisł 1916
- Romer: Atlas Polski. Atlas von Polen. Atlas de la Pologne.
- Rocznik pracowniczych związków zawodowych w Polsce. Rok 1926.
- Quetelet: Physique social ou essai sur le developpement des facultés de l'homme. Bruksela, Paryż, Petersburg, 1869.
- Quetelet: Układ społeczny i jego prawa. Warszawa, 1874.
- Schnapper-Arndt: Sozialstatistik. Lipsk, 1908.
- Schwittau: Wwiedienje w ekonomiczeskuju statystiku. Pet. 1910.
- Sirlnow: Oczerki po agrarnoj statystikie. Petersburg, 1911.
- Sobolew: Organizacja i metody statystyki truda. Tomsk, 1903.
- Spława-Neyman: Próba rozwiązywania niektórych zagadnień doświadczał'nictwa rolniczego za pomocą statystyki matematycznej. P. 1924
- Szturm de Sztem: Inwentarz żywy w Polsce. W., 1928.
- " Powszechny spis rolniczy. W., 1928.
- Stahl: Doktryna ilościowa w świetle metody statystycznej. „Ruch praw, ekonom. i socjol.” IV kwartał 1929 r.
- Szulc: Statystyka. Opracowanie wykładów uniwersyteckich. Warszawa, 1926.
- Sypniewski: Sootnoszenje wieliczin anatomiczeskich elementow u raznych ras kartofela, przenie i swiekłowicy. Petersburg, 1913.
- Tischer: Grundlegung der Statistik. Jena, 1929.
- Waściszakowski: Ewolucja produkcji rolnej w Król. Kong. W-ctwo Biura Prac Kogr. Warszawa, 1919.
- " O źródłach statystyki rolnej. "Sprawa rolna" zesz. 8, W., 1921.
- Weigel: Rachunek wyrównawczy. Lwów, 1923.
- Weinfeld: Tablice statystyczne Polski za r. 1924.
- Westergaard: Grundzüge der Theorie der Statistik, Jena, 1928.
- Wielka własność rolna. W two Gl. Urzędu Statystycznego.
- Wobłyj: Statistika, Kijów, 1909.
- Woytinsky: Die Welt In Zahlen. Berlin, 1927.
- Yule: Wstęp do teorii statystyki. Przełożył Z. Liłmanowski. W. 1921.
- Załęski: Teoria statystyki w zarysie. Cz. I. Warszawa 1884.
- Zweig: Cele i metody statystyki handlu zagranicznego, Warszawa 1921.

ERRATA.

Str. 207 wiersz 14 zamiast $r = \frac{5}{4.942}$ winno być $r = \frac{4.942}{5}$

" " " 25 " $r = \frac{0.1278}{6}$ " " $r = \frac{0.1278}{5}$





Biblioteka

36342/2

Naukowa