Prof. Dr. Dobroslav KREJČÍ:

K otázce ceny matematické statistiky pro statistickou praxi.

Obsahem i rázem svých dosavadních statistických prací řadím se zřejmě ke statistikům nematematickým, a v II. vydání svých „Základů statistiky“ (Praha 1923) vyslovuji se o významu t. zv. *matematické statistiky* hodně skepticky [1]). Přesto však,

[1]) Tak pravím tam na str. 48: „... část nových a nejnovějších statistiků snaží se zákony ty“ (totiž zákony ovládající hromadné zjevy sociální) „najíti pomocí vyšší matematiky... ba někteří z nich jen tuto „*matematickou statistiku*“ uznávají za jedinou pravou, vědeckou statistiku vůbec, kdežto valná většina ostatních statistiků dosud má za nebezpečno, vyvozovati pomocí abstraktních formulí matematických všeobecné zákony pro zjevy společenské, přec jen vždy příliš složité a různotvárné...“ A na str. 108, mluvě o cestách vědeckého rozboru statistického, pravím: „V novější době vzmáhá se používání vyšší matematiky při rozboru statistickém... Pomocí něho snaží se někteří dodati výsledkům statistiky rázu exaktnosti. Ale, jak právě bylo vytčeno, je vědecké statistice přísná exaktnost cizí, tak jako řadě jiných sociálních věd, následkem zvláštního rázu a složitosti jejich předmětu, zjevů sociálních vůbec, a následkem zvláštních obtíží, které se dosud staví v cestu exaktnosti a (naprosté) úplnosti pozorování hromadných zjevů sociálních zvlášť. Dosud také výsledky, k nimž tato „matematická statistika“ dospěla, buď jen namáhavě, pouze pro nemnoho zasvěcenců srozumitelně potvrzují to, co prostá statistika již dávno z výsledků svých pozorování zjistila pomocí pouhé logiky a nejprostších, každému běžných operací početních, anebo — pokud nerozumně nepřeháněji — jen opatrně naznačují některé nové vývody z téchže výsledků, jež však mají platiti teprve, potvrdí-li je nová statistická šetření a nové prosté, logické rozbory jich. Proto také velká většina statistiků dosud uznává, že ke správnému rozboru vědecko-statistickému není potřebí vyšší matematiky; používání její sice může někdy vydatně utvrditi víru ve správnost vývodů obecné statistiky nebo dáti cenné podněty pro směr dalších jejich zkoumání, ale nezřídka může také zbytečně buditi zdání přísné exaktnosti o výsledcích, které mají a mohou míti vždy jen relativní platnost, a někdy dokonce svésti na scestí... Přecpávati každý rozbor statistických dat matematikou, dokazovati zbytečně pomocí složitých matematických formulí to, co z nich každý i bez nich viděl, není rozumné... Pouze v některých zvláštních oborech, zejména v pojišťovatelství, je matematická statistika nepostrádatelna.“

když jsem r. 1924 byl presidiem našeho Státního úřadu statistického — tak jako řada jiných statistiků a národohospodářů — dotázán, které z cizích theoretických děl statistických by bylo nejspíš záhodno přeložiti do češtiny, doporučil jsem bez rozpaků matematicko-statistické dílo Yulovo. Proč asi? Nejen proto, že pro vědeckou literaturu považuji jednostrannost vždy za nedostatek a přál bych naší české literatuře, aby jako ve všech oborech vědných, tak také ve statistice mohla se vykázati díly všech různých směrů; a nejen proto, že Yule patří netoliko k nejlepším, ale i k oněm rozumným matematickým statistikům, kteří nepřepínají významu svých matematických formulí, — vždyť sám leckde upozorňuje na nebezpečí sevšeobecňování jich —; ale i proto, že jeho statistická matematika, i když se mi v mnohých částech také zdá aspoň relativně málo užitečnou nebo zbytečnou, ba někde i nebezpečnou, obsahuje přece také řadu cenných pomůcek i pro statistickou praxi.

Chci zde příkladmo poukázati aspoň na některé případy jak oné nebezpečnosti neb aspoň zbytečnosti některých formulí Yulových pro praxi tu, tak zejména praktické užitečnosti jiných. Obojí (pro obmezenost místa) prozatím jen z první, také jednodušší části Yulovy knihy [2]), totiž z části jednající o theorii znaků kvalitativních.

I.

Praktickou cenu tu mají:

1. Z kapitoly první formule, pomocí jichž si můžeme vypočísti četnost (t. j. počet případů vyskytování se v daném souboru) všech možných znaků a kombinací jich, známe-li četnosti buď

a) jen všech konečných tříd (t. j. positivních i negativních kombinací všech pozorovaných znaků dohromady) nebo

b) jen všech kladných tříd.

Tedy na př. pozorujeme-li v jistém souboru (N) tři znaky:

---

[2]) „Úvod do teorie statistiky. Napsal G. Udny Yule, C. B. E., M. A., F. R. S.... Dle VII. opraveného vydání z r. 1924 přeložili PhDr. Vladimír Novák a JUDr. Josef Mráz. Redigoval Dr. Josef Mráz." (Praha 1926.)

slepotu (A), hluchotu (B) a choromyslnost (C), stačí, když při technickém zpracování sebraného materiálu zjistíme

ad a) jen počet těch, kteří byli zároveň slepí, hluší i choromyslní (ABC)[3], pak těch, kteří byli slepí a hluší, ale nikoli choromyslní ($AB\gamma$)[4], a podobně všechny ostatní kombinace těch tří znaků: ($A\beta C$), t. j. počet slepých, kteří jsou zároveň choromyslní, ale nikoli hluší, ($A\beta\gamma$), (αBC), ($\alpha B\gamma$), ($\alpha\beta C$) a ($\alpha\beta\gamma$), t. j. počet těch, kteří nebyli ani slepí, ani hluší, ani choromyslní; a n e b o

ad b) jen celkový počet všech pozorovaných N a pak četnost jen kladných tříd, totiž jen: všech slepých (A), všech hluchých (B), všech choromyslných (C), pak slepých, kteří jsou zároveň hluší (AB) nebo choromyslní (AC), hluchých a zároveň choromyslných (BC) a těch, kteří v sobě spojují všechny ty tři znaky (ABC).

Všechny ostatní četnosti můžeme si v obou případech prostě vypočísti z těch známých čísel, aniž by bylo potřebí je zvlášť vyčítati z materiálu[5]). Jest to ovšem zkušenému statistikovi známo a každému logicky myslícímu jasno i bez matematiky, neboť každý ví na př., že když mezi 1000 obyvateli napočetl jen 20 slepců, nemusí tam vidoucí sčítat zvlášť, ale že jich tam musí být $1000 - 20 = 980$, anebo, když v jistém obyvatelstvu zjistil 70, kteří jsou jen slepí, pak 20, kteří jsou slepí a hluší, 6 slepých choromyslných a 4 slepce, kteří jsou zároveň hluší i choromyslní, tu že nemusí ještě zvlášť vyčítati z materiálu.

[3]) Závorky vyznačují zde, tak jako u Yula, četnost znaku v nich uvedeného v pozorovaném souboru (jehož četnost rovněž jako Yule značím tu písmenem N); tedy (ABC) značí počet případů, u nichž se v souboru N vyskytají všechny 3 pozorované znaky A, B i C.

[4]) Řeckými písmeny α, β, γ označuji tu podle Yula opak znaků A, B, C, tedy α = non A = vidoucí, β = non B = slyšící, γ = non C = na duchu zdraví.

[5]) Tak *ad a)* vypočteme si celkový počet slepců součtem všech těch známých četností, u nichž se vyskytuje znak slepectví A, tedy $(A) = (ABC) + (AB\gamma) + (A\beta C) + (A\beta\gamma)$; podobně vypočteme beze zvláštního zjišťování četnosti (B), (C), (AB), (AC), (BC), (α) (β) (γ), ($A\beta$), ($A\gamma$), ($B\gamma$), (αB), (αC), ($\alpha\beta$), ($\alpha\gamma$), (βC), ($\beta\gamma$) a N. Obdobně můžeme si *ad b)* ze známých 8 četností snadno vypočísti všechny ostatní z 27 vůbec zde možných: $(\alpha) = N - (A)$, $(\beta) = N - (B)$ atd. Yule neuvádí všech formulí pro ty výpočty, ale dává návod, podle něhož si je každý snadno může sestaviti. Překladatelé to ještě usnadnili svým „Přehledem vzorců“ na str. 453 a násl.

kolik tam je osob slepých vůbec, nýbrž že jich tam nemůže být než $70 + 20 + 6 + 4 = 100$. Opatrný statistik pak rád vyčíslí přímo z materiálu — pokud čas stačí — také leccos, co by si jinak mohl jen vypočísti, neboť nevelike zpravidla plus práce s tím spojené je nahrazeno kontrolou tím získanou. Ale přes to poznatek, že při třech znacích zjištění určitých 8 četností úplně postačí, bychom si vypočetli bezpečně a snadno všech 19 ostatních četností tu možných, má pro praxi nepopíratelnou cenu, nejen pro případy, kde skoupě vyměřený čas pro technické zpracování statistického materiálu nutí obmeziti je pouze na věci nejnutnější, ale i jindy: Upozorňuje, které to četnosti jsou nejdůležitější, zejména tedy, kterých kombinací nemáme při zpracování a hlavně též při publikaci výsledků opomenouti, aby si mohl každý věci znalý používatel jich snadno vypočísti též kterékoliv jiné kombinace (četnosti), jichž pro své účely, nám předem neznámé, kdykoliv bude potřebovati [6]).

2. Tytéž jednoduché formule Yulovy umožňují také některé další úsudky z daného statistického materiálu, ke kterým bychom bez nich a beze zvláštního vyčíslení z materiálu sama jen stěží dospěli. Není jich arci mnoho, a ne všechny mají praktickou cenu. Z příkladů, které shledal Yule (str. 16 a 17) — a jiný by jich sotva shledal více —, patří sem jen čtvrtý a snad pátý. Podle čtvrtého vždy, když zjistíme, že jistý znak (A) je mezi případy mající mi jiný znak (B) četněji zastoupen než mezi případy nemajícími toho jiného znaku (β), můžeme být jisti, že naopak také tento jiný znak (B) je četněji zastoupen mezi případy vykazujícími onen první znak (A) než mezi případy, jež ho nemají (α), arci jen v témže souboru, což Y. sice neuvádí, ale zřejmě předpokládá; tedy konkretně na př. zjistíme-li, že v jistém obyvatelstvu je mezi slepci více hluchých než mezi vidoucími, můžeme beze zvlášt-

[6]) K témuž účelu sice, jak Yule sám mimochodem podotýká (v § 14 na str. 13 dole), místo konečných četností stačí každá jiná soustava, obsahující týž počet (totiž 2^n) četností algebraicky nezávislých (t. j. takových, že žádná se nedá číselně odvoditi z ostatních těch četností). Ale je jistě výhodno věděti beze zkoumání té nezávislosti, že všechny konečné nebo všechny kladné četnosti tu stačí vždy a nejsnadněji.

ního zjišťování tvrditi, že v něm také naopak mezi hluchými je více slepců než mezi nehluchými. Bez těch formulí bychom sic ze zjištěné praemissy soudili, že slepci jsou náchylnější ke hluchotě než ostatní, vidoucí obyvatelstvo, ale protože slepci v celku tvoří jinou část celkového obyvatelstva nežli všichni hluší, netroufali bychom si bez nového zjištění soud ten s bezpečností obrátiti.

Podle 5. z praktických příkladů Yulových v každém souboru, kde většina případů se znakem A má zároveň znak B a většina případů s týmž znakem A má zároveň znak C, můžeme výpočtem zjistiti, kolik nejméně případů se znakem B má zároveň znak C. Tedy opět konkretně: Zjistíme-li na př., že většina slepců ve státě X trpí zároveň hluchotou a že také duševními chorobami trpí tam většina slepců (třebas arci in concreto jiná), můžeme si vypočísti, kolik hluchých musí tam při nejmenším býti zároveň na duchu chorých; bude jich nejméně tolik, oč větší je tam součet počtu slepců hluchých a slepců na duchu chorých, než celkový počet slepců [7]). Praktická cena tohoto poznatku nebude arci veliká, protože nám v praxi jen velmi zřídka stačí znalost minima, o němž nevíme, oč je skutečností převýšeno.

Ostatní příklady, uvedené Yem, jsou buď jen prostým výpočtem scházejících četností a patří tedy jen do našeho odst. I. 1., nebo jsou zcela nepraktické (jako 6. a 7., ježto v praxi se as nikdy nevyskytne jejich předpoklad, že každý z pozorovaných znaků vyskytuje se právě u polovice všech pozorovaných případů).

3. Z kapitoly II. (o souladu čili konsistenci četností) má pro praxi cenu již poznatek, že žádná třídní četnost nemůže býti záporná. Samo o sobě je to sic až směšně samozřejmé: nemůžeme přec statisticky napočísti zápornou veličinu jednotek s jakýmkoliv znakem ani bez nějakého znaku, nemůžeme zjistiti minus pět slepců, ani minus pět neslepců. Ale poznatek ten umožňuje statistikovi a každému používateli statistických výsledků uvarovati se leckterého nepříjemného omylu. Vypočte-li si totiž z daných výsledků všechny nezjištěné čet-

---

[7]) Je-li na př. z 1000 slepců 600 hluchých a 800 na duchu chorých, musí tam býti nejméně 400 hluchých zároveň choromyslnými.

nosti a pozná, že kterákoli z nich vychází se znaménkem záporným (—), má v tom důkaz, že ve výsledcích těch je nějaká chyba, ať tisková, nebo již ve vyčíslení výsledků těch samých (při technickém zpracování nebo hned při sbírání materiálu). Proto zvláště tenkrát, když při rozboru nějakých výsledků statistických dospíváme k závěrům překvapujícím, doporučuje se důtklivě, vypočísti si též všechny scházející, třeba pro náš účel nepotřebné četnosti, nemáme-li možnosti kontrolovati správnost daných výsledků jiným, ještě bezpečnějším způsobem. Jeť kontrola zde dotčená přirozeně jen slabou pomocí z nouze: i když nám vyjdou všechny četnosti kladné, není tím nikterak vyloučeno, že přes to je v pozorovaných výsledcích nějaká skrytá vada, a vyjde-li nám některá četnost záporná, máme sic jistotu, že tu je chyba, ale nevíme, ve kterém z pozorovaných čísel. Jsme však upozorněni, abychom ji hledali, a nenajdeme-li, bychom si alespoň neškodili činěním závěrů z čísel vadných.

4. Najíti takovou chybu, zejména je-li to jen chyba tisku, vzniklá na př. vypadnutím jedničky před číslem 85 (jako v příkladě, který uvádí Y. na str. 23), mohou nám snad někdy pomoci Yulovy formule podmínek souladu četností pro znaky kvalitativní. Umožňují zjistiti horní a dolní mez, mezi nimiž se musí pohybovati četnost podezřelá z chyby, aby byla v souladu s ostatními danými četnostmi. Ovšem i to je pomůcka jen z nouze: budouť meze ty zpravidla velmi široké (dolní mez při 1 znaku vždycky nula, při 2 znacích nula nebo negativní hodnota (A) + (B) — N atd.), takže menší chyba se v jejich šíři často ztratí; a i když zjistíme, že tu chyba je, nevíme, jakým určitým číslem nahraditi číslo chybné, — široké rozpětí mezi možným minimem a maximem jeho nám v praxi asi zřídka stačí[8]).

[8]) Tak v cit. příkladu Y-ově vychází pro zkoumanou četnost (BC) jako dolní mez podle formule (4 *a*) sice oněch 98, podle nichž Y. soudí, že udaná hodnota této četnosti 85 nemůže býti správná; ale podle formule (4 *b*), — v níž dlužno opraviti chybu, patrně tiskovou, takže má zníti: (AB) + (AC) — (BC) <|≦ (A) — vyjde dolní mez záporná: — 181, podle formule (4 *c*) pak jako mez horní 441 a podle formule (4 *d*) dokonce 476. I když tedy absurdní zápornou mez dolní prostě ignorujeme a z horních mezí považujeme za rozhodnou jenom mez nižší, ve skutečnosti mohlo by se tu tedy (BC) pohybovati mezi 99 a 440, i je závěr Y., že chyba se asi stala tím, že před číslem 85 v tisku vypadla jednička, hodně smělý: mohla vypadnout také dvojka nebo trojka a mohly i číslice 8 a 5 být zcela chybné.

5. Stejně jen zřídka a ne příliš velikou, ale přec jen jakousi cenu budou mít tytéž formule (*3a* až *h* a *4a* až *d*) pro praxi tenkrát, když půjde o to, zjistit aspoň přibližně — v hranicích možného maxima a minima — četnost některé kombinace, která nebyla při technickém zpracování vyčíslena a jejíž přesnou četnost nelze ani dodatečně vyčísliti, ani vypočítati z daných četností. Ovšem také zde budou hranice ty zpravidla tak široké, že sotva stačí praktické potřebě [9]); ostatně často dá se v takových případech buď zcela totéž minimum a maximum neb aspoň přibližně stejné určiti bez těch formulí, prostou logickou úvahou, jak ukazuje nejlépe samo řešení uvedených při této kapitole úloh č. 1 a 2 na str. 24 (srov. řešení jich na str. 421).

II.

Všechny ostatní vývody celé této první části Yulovy knihy, zejména tedy celé její kapitoly III. a IV. (o družnosti čili asociaci kvalitativních znaků vůbec a o dílčí družnosti jejich zvlášť), nemají po mém soudu pro statistickou praxi žádné zvláštní ceny, ba některé zdají se mi pro ni přímo nebezpečny.

1. **Nebezpečné** jsou po mém zdání na př. již formule, matematicky jistě bezvadné, které Y. uvádí v § 1 a 2 kapitoly III. (str. 26 a 27), i „důležité základní pravidlo“, jež tam Y. z nich vyvozuje: „Jsou-li znaky A a B na sobě nezávislé, pak poměrná četnost současného se vyskytnutí znaků AB je rovna součinu poměrných četností

[9]) Příklady a úlohy Y-ovy (str. 22—25) buď mají předpoklady zvlášť přizpůsobené, v praxi téměř nemožné (jako v příkl. 1.: (A) = (B) = (C) = ½ N, spolu s vysokými %), anebo — nejsou-li pouhým hraním s formulemi, při věcech naprosto samozřejmých bez nich, jako úloha 6. a 7. — dávají výsledky pro praxi svou neurčitostí nedostačující; neboť jakou asi cenu pro praxi má na př. poznatek z úlohy 4., že ze 789 hochů, u nichž zjištěna duševní tupost, nejméně 117 nemělo vad v tělesném vývoji, když nevíme, do jaké míry bylo toto minimum ve skutečnosti převýšeno? Smíme z toho činiti nějaké závěry, když přec je možno, že skutečnost byla dvakrát, třikrát neb i čtyřikrát větší? 117 je ze 789 jen asi $^1/_7$, zkušenost však učí, že těch, kdo nemají jistou vadu kombinovanou s jinou, bývá bohudík nesrovnatelně víc než ubožáků s takovou kombinací. A podle čísel, uvedených Y-em před tím z téhož pramene, bylo mezi 2308 žáky s těles. vadami jen 887 s vadami čivními a tedy 1421 bez nich. Tím víc pravděpodobně bez duševní tuposti!

jednotlivých znaků A a B." To mohlo by sváděti, aby se chtělo bez skutečného statistického zjištění pouhým výpočtem dospěti k číslům, udávajícím určitě počet případů kombinace znaků jen o sobě pozorovaných, k číslům, jež by skutečný mnohotvárný život jistě v ohromné většině případů pokáral ze lži.

Viděti to hned na prvém příkladu, jejž tu Y. sám uvádí (str. 28); potřebí si jej pouze konkretisovat. Y. tu vypočítává, že jsou-li v pozorovaných 1024 případech 144 případy se znakem A a 184 případy se znakem B, je tam právě 54 případů, vykazujících oba ty na sobě nezávislé znaky A i B.

Konkretisujme si znak A třeba na slepotu a znak B na hluchotu (tedy znaky, jimiž Y. sám předtím operuje). Zkušenost dovoluje souditi, že jsou to znaky na sobě nezávislé: Slepota sama nevyvolává hluchoty, ani naopak. Za normálních poměrů a ceteris paribus můžeme tu (jak Y. na str. 26 žádá pro znaky na sobě nezávislé) jistě předpokládati, že mezi slepci bude poměrně právě tolik hluchých jako mezi neslepci. Ale můžeme proto už býti jisti, že v **každém** souboru z 1024 případů, kde je 144 slepých a 184 hluchých, musí býti vždy právě 54 osoby zároveň hluché i slepé? Ať je to 1024 případů z jakéhokoliv prostředí a z kterékoli doby? Můžeme se spokojit v praxi vůbec jen zjištěním nekombinovaných případů slepoty a hluchoty, a počet případů kombinace jich si jen vypočísti? Nebezpečnost takového postupu je jistě i laikovi zřejma.

Ale snad by se dalo namítnouti, že tedy slepota a hluchota nejsou znaky na sobě nezávislými. K stejným výsledkům dojdeme však, aplikujeme-li toto Yulovo „základní pravidlo" na znaky, jež on sám uvádí (na začátku str. 26) jako příklady znaků na sobě nezávislých. Platí to stejně o druhé z uvedených tam dvou dvojic — t. j. o znacích pohlaví novorozence a narození v době, kdy měsíce přibývá, u kterýchžto znaků možno ostatně být aspoň právě tak, ne-li víc, než u slepoty a hluchoty v nejistotě, zda jsou opravdu na sobě zcela nezávisly — jako o dvojici prvé: abnormálně vlhkých počasí a přestupných let. Konkretisujme si to zase: Je-li v době řekněme 20 let, t. j. $20 \times 365 + 5 = 7305$ dní (N), zjištěno třeba 700 dní s počasím abnormálně vlhkým (A), a je-li v téže době celkem $5 \times 366 = 1830$ dní, patřících letům přestupným (B), můžeme tu být jisti, že v přestupných těch letech samých musilo býti právě $\frac{700 \times 1830}{7305} = 175\cdot36$ dní abnormálně vlhkých? Což jestliže právě v těch 20 letech náhodou nepadla léta abnormálně vlhká na žádné z těch pěti přestupných let? Anebo naopak, padla-li právě na všechna neb aspoň na většinu jich?

Myslím, že již je zřejmo: I kdyby vzájemná nezávislost dvou znaků byla sebe dokonalejší, čili jinými slovy: kdyby mezi těmi znaky nebylo pražádné příčinné souvislosti, bylo by v praktické statistice vždycky svrchovaně nebezpečno spolehnouti se, že kombinace jich musí se vyskytovati vždy jen v určitém poměru, čili že $\frac{(AB)}{N}$ musí se vždy rovnati $\frac{(A)}{N} \cdot \frac{(B)}{N}$. Vysvětlení je snadné: Vždyť v životě a v živé statistice nepracuje se nikdy s případy, ovládanými jen jedinou příčinou, majícími tedy příčinnou souvislost jen s jedinou jinou skutečností, jediným jiným znakem. Také Y., ač neobmezuje svou statistiku jen na pozorování zjevů sociálních, vždycky zvláště složitých, vytýká jako hlavní znak jejího předmětu „multiplicity of causes“, „množství příčin“ (srov. jeho definice statistiky na str. 6). I když tedy na hluchotu nepůsobí slepota, působí na ni u slepců — rovněž jako u neslepců — celá řada jiných příčin, a to v různých zemích, různých prostředích a různých dobách různě; vlivem těchto různých jiných příčin bude tedy přirozeně hluchota, byť i byla na slepotě samé naprosto nezávislá, vyskytovati se u zcela téhož počtu slepců v různých prostředích a dobách v počtu různém, a nikoli vždy právě v počtu stanoveném Yulovou formulí. A právě tak tomu je i při jiných znacích, na sobě sebe nezávislejších.

Proto jen v nejřidších, zcela nahodilých případech možno očekávati, že $\frac{(AB)}{N}$ bude ve skutečnosti rovno právě $\frac{(A)}{N} \cdot \frac{(B)}{N}$. Kdo by tedy ve statistické praxi spoléhal na toto „důležité základní pravidlo“, musil by se v ohromné většině případů dožíti velmi nemilého zklamání.

Zdá se ostatně, že Yule sám byl si toho vědom; proto v příkladě 2. (na téže str. 28) označuje výsledek jen za přibližný, s odkazem na §§ 7 a 8, kde sám ukazuje na příkladu s vyhazovaným penízem, že skutečnost dává často jiné poněkud výsledky než výpočet a vysvětluje to hlavně „kolísáním náhodných výběrů“. Ale přes to uvádí ono „základní pravidlo“ i matematickou formuli jeho a řadu dalších formulí z ní odvozených[10]), jakož i výsledky v příkladu 1. a 3. podle nich vypočtené, jako přesné (s jedinou jen podmínkou, že znaky A a B jsou na sobě nezávislé) —, a to je, co by mohlo zejména začátečníka v praxi velmi nebezpečně svésti.

2. Že další vývody kapitoly III. **nemají pro praxi významu,** vyplývá

a) jednak již z toho, že jsou v podstatě vybudovány právě na onom „základním pravidle", jehož nebezpečnost pro praxi jsme právě poznali a

b) jednak z toho, že k týmž výsledkům — t. j. k poznatku, zda mezi danými znaky jest či není asociace, a do jaké míry, — možno bez složitých a často dlouhých výpočtů podle formulí Yových dospěti „prostě a zřejmě" „nejpřirozenějším přirovnáním" pomocí pouhých procent nebo promille, jak Y. sám na potvrzení svých výsledků uvádí (str. 32) a často také činí.

ad a) Základní Yulovy formule asociace kladné: $(AB) > \frac{(A)(B)}{N}$ i záporné: $(AB) < \frac{(A)(B)}{N}$ vyplývají mu přímo z toho, že podle něho všude, kde znaky A a B jsou na sobě nezávislé, čili kde mezi nimi asociace není, platí o nich rovnice $(AB) = \frac{(A)(B)}{N}$. Není-li, jak jsme viděli, v praxi radno spoléhati se na tuto rovnici, není ovšem radno spoléhati se ani na matematické odvozeniny z ní[11]).

---

[10]) $(A\beta) = \frac{(A)(\beta)}{N}$ atd., $(AB)(\alpha\beta) = \frac{(A)(B)(\alpha)(\beta)}{N^2} = (\alpha B)(A\beta)$ a řadu formulí z toho odvozených, $\frac{(AB)}{(A\beta)} = \frac{(\alpha B)}{(\alpha\beta)}$ atd.

[11]) O oné první formuli kladné asociace $(AB) > \frac{(A)(B)}{N}$ praví Y. sice na str. 31, že „z teoretického stanoviska je snad nejpřirozenější", ale hned poté o první odvozenině z ní: $\frac{(AB)}{(B)} > \frac{(A)}{N}$ sám doznává, že „má-li v pozorovaném souboru velká většina předmětů nebo jednotlivců znak B, t. j., blíží-li se $\frac{(B)}{N}$ jedničce, je zřejmo, že $\frac{(AB)}{(B)}$ bude se nutně blížiti zlomku $\frac{(A)}{N}$, i kdyby rozdíl mezi $\frac{(AB)}{(B)}$ a $\frac{(A\beta)}{(\beta)}$ byl značný", a že tudíž tento „tvar přirovnání může vésti na scestí" (str. 32).

ad b) Nadto však jich není ani potřebí, neboť pouhá logická úvaha sama o sobě nebo spolu s prostým, každému snadným výpočtem několika % nebo ‰ vede nás rychleji a bezpečněji k týmž výsledkům. Samy Yulovy příklady a úlohy sem patřící jsou tu nejlepším dokladem. V příkladu nejdřív zde uvedeném (na str. 32: Asociace mezi očkováním proti choleře a uniknutím nákaze) sám uvádí, že „nejpřirozenější přirovnání" daných čísel je tu srovnání procent jednak očkovaných, jednak neočkovaných, kteří buď byli nebo nebyli zachváceni nákazou, a pokračuje: „Obojí přirovnání ukazuje prostě a zřejmě (podškrtuji já) skutečnost, že očkování a uniknutí nákaze jsou kladně sdruženy", — bez jakýchkoli matematických formulí (dodávám já); tím spíš, když na témže příkladě ukazuje sám Y., že se tu nehodí srovnání nejen podle formule $\frac{(AB)}{B} > \frac{(A)}{N}$, o níž předtím sám tvrdí, že může vésti na scestí (srov. pozn. 11), ale ani podle formule druhé: $\frac{(AB)}{(B)} > \frac{(A\beta)}{\beta}$. Podobně je tomu ve všech dalších příkladech: prostá úvaha a pár procent vede nás tu bezpečněji nejen než ony jednoduché první formule Y., ale i než hodně složité další. Jakou cenu má vůbec věděti na př., že δ, t. j. rozdíl mezi (AB) a $\frac{(A)(B)}{N}$ se rovná $\frac{1}{N}\left\{(AB)(\alpha\beta) - (\alpha B)(A\beta)\right\}$ neboli slovy: že společný rozdíl δ rovná se $^1/_{N\text{tině}}$ z rozdílu „křížových součinů" $(AB)(\alpha\beta)$ a $(\alpha B)(A\beta)$, když hned za vyvozením tohoto vzorce Y. sám upozorňuje, že „je-li N veliké, rozdíl křížových součinů může býti velmi značný, ačkoliv δ jest vskutku velmi malé", a proto radí: „rozdíl budiž srovnán s N, sice by mohla vzniknouti domněnka o vyšším stupni asociace, než který vskutku existuje" (str. 36, podškrtuji já)?!

A právě tak stačí srovnání logicky volených procent úplně k změření stupně asociace mezi dvěma znaky, aniž bychom musili znáti různé ty „míry čili koeficienty asociace", jež skoro každý matematický statistik vypočítává jinak, a o nichž Y., když byl sám sestavil dvě nové [12]), výslovně

[12]) Jeden nazývá Q a značí $= \frac{N \cdot \delta}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$, druhý $\omega = \frac{\sqrt{(AB)(\alpha\beta)} - \sqrt{(A\beta)(\alpha B)}}{\sqrt{(AB)(\alpha\beta)} + \sqrt{(A\beta)(\alpha B)}}$. Výpočet v praxi, jak viděti, není nikterak krátký, ani jednoduchý (srv. svrchu formuli pro samo δ).

praví: „Otázka, která míra se nejlépe hodí k měření asociace, jest dosud předmětem sporu“ (str. 39), spor pak zahájen prací Dav. Herona s výmluvným titulem (v překladu): „N e b e z p e č í některých vzorců, navržených za náhražky míry korelace“!

Vůbec se mi nezdá, že by bylo pro praxi nebo pro vědu ziskem, když tentýž poznatek, který se dá vyjádřiti nebo dokázati docela prostě a každému srozumitelně, vyjádří nebo dokazuje se způsobem daleko složitějším, laikovi jen těžko sledovatelným, nebo se zahaluje do řady cizích slov, učeně vypadajících výrazů nebo formulí matematického vzhledu, u nichž každý znak musí se předem vyložit a kde zpravidla pro týž pojem tvoří si různí autoři znaky různé, atd.

3. T o v š e c h n o p l a t í v p o d s t a t ě t é ž o c e l é k a p i t o l e IV., jednající o t. zv. d í l č í a s o c i a c i. Je to v podstatě jen složitý, laika tou složitostí spíš ve zmatek uvádějící než utvrzující výklad pravdy každému známé, že i při zjištění vzájemné spojitosti nebo naopak nezávislosti mezi dvěma výsledky statistickými nutno být opatrným v pronášení úsudků z toho plynoucích, protože ta spojitost nebo nezávislost může být způsobena též jinými z celé řady vlivů, na oba ty výsledky působících; a že tedy opatrný statistik nemá se spokojiti pozorováním jen jediného páru znaků, a jen v celku, nýbrž také znaků dalších, v celku i v jednotlivých zvláštních jeho částech.

Z této prosté úvahy vyplývá každému logickému pozorovateli sama sebou na př. poučka, uváděná Y. na str. 49: „Jsou-li znaky A a B na sobě nezávislé v souboru C a také v souboru γ, budou přece sdruženy v celkovém souboru N, vyjímajíc případ, že by také C bylo nezávislé buď na A nebo na B nebo na obou“ — aniž by bylo třeba dokazovati ji ze složité rovnice

$$(AB)=\frac{1}{(C)(\gamma)}\{N(AC)(BC)-(A)(C)(BC)-(B)(C)(AC)+(A)(B)(C)\}$$

a z dalších odvozenin z ní. A podobně je tomu i při ostatních poznatcích této kapitoly, ještě složitějších. Yule sám docela správně upozorňuje tu na nutnou opatrnost při užívání té které formule v praxi neb i na její nepraktičnost vůbec [13]).

[13]) Srov. na př. jeho výklady o různých „z d á n l i v ý c h“ a „k l a m n ý c h“ asociacích, zvláště na str. 50 a 51, nebo slova: „...vypočtení všech jejich hodnot pro případy, kde n jest větší než 4, je skoro n e p r o v e d i t e l n é“ (54); „tuto asociaci n e l z e p r o k á z a t i bez

III.

V celku tedy vidíme, že prostudování již prvé, celkem malé části knihy Yulovy není pro praxi statistickou bez užitku, ale ovšem také, že užitku toho nelze přeceňovati, že formulí matematicko-statistických nutno v praxi používati velmi opatrně, a že dobrý statistický rozbor — prozatím aspoň kvalitativních znaků — možno dobře dělati též beze znalosti teorií matematicko-statistických.

Z toho by se přirozeně vnucovala otázka, zda se tedy tomu, kdo nemá ani hlubšího předběžného vzdělání matematického, ani vrozených „buněk matematických", ale přes to chce, potřebuje nebo musí v praxi používati statistiky, vyplácí dosti značná námaha, jakou mu jistě působí studium i této nejpřístupnější učebnice matematicko-statistické. Ke správnému zodpovědění této otázky nutno však ještě zhodnotiti po této stránce také druhou a třetí část knihy Yulovy, mnohem důležitější a obsáhlejší než část první. O tom jindy.

Zde nutno jen ještě zdůrazniti, že vším tím, co svrchu řečeno, nemá být nikterak snad popírána veliká teoretická a vědecká cena knihy Yulovy, tím méně pak snižována rovněž veliká práce a zásluha jejích českých překladatelů. Profesor Vlad. Novák a docent Jos. Mráz svým opravdu vzorným překladem obohatili naši vědeckou literaturu o dílo nesporně velké a trvalé ceny, a spolu s prof. Fr. Nachtikalem četnými doplňky, vysvětlujícími poznámkami, návodem a řešením úloh atd. dali naší odborné literatuře knihu, v leckterém směru i lepší a zejména

---

znalosti třídní četnosti (ABC)", tedy bez prostého, nematematického vyčíslení případů kombinace všech tří pozorovaných znaků z původního materiálu; nebo „Prakticky nám však pouhá skutečnost, že se ostatní asociace dají odvoditi, málo pomůže, ledaže by se takové odvození dalo provésti jednoduše, ba takřka přímo, pouhým počítáním nazpaměť" atd. až: „Proto při úvaze, které asociace a kolik jich má býti skutečně zpracováno, rozhodují spíše praktické ohledy nežli teoretické vztahy" (56); „vztah (8) není dostatečnou podmínkou úplné nezávislosti" atd., a „nelze tvořiti nijakých závěrů bez dalších informací" (57), tedy opět bez prostého, nematematického zjišťování statistického; nebo: „z platnosti vztahu (9) nenásleduje, že všechny znaky jsou na sobě nezávislé" (58), atd.

poučnější a srozumitelnější než sám originál[14]). Moje řádky chtějí jen varovati před přeceňováním všech, i Yulových, formulí matematických pro statistickou praxi, a chrániti před zmatkem, poklesem sebedůvěry nebo snad i zoufáním nad statistikou vůbec ty, kdo nemají hlubšího vzdělání matematického, ani schopností sledovati snadno vývody matematické, a přece mají zálibu ve

[14]) Pro eventuelní druhé vydání upozornil bych jen — pokud jde o probranou zde první část knihy — na těchto několik maličkostí:

1. Podstata první Yulovy definice statistiky (str. 6 v překladu, str. 5 v orig.: „By statistics we mean quantitative data affected to a marked extent by a multiplicity of causes“) dala by se snad spíše než slovy: „...číselné údaje, jež jsou výrazem účinku *velkého* množství příčin“ vystihnouti jiným překladem, asi takto: „Statistikou rozumějí se číselné údaje o výsledcích, které jsou *ve značném rozsahu* způsobovány množstvím příčin“, nebo volněji: „... o zjevech *většího rozsahu*, vyvolaných celou řadou příčin“. Zdá se mi totiž, že slovy „to a marked extent“ nechtěl Y. zesíliti slova „multiplicity of causes“, (jež mu v definici druhé také úplně postačila sama), nýbrž spíše naznačiti, že výsledky těch mnohých příčin musí být dosti značného rozsahu, aby se daly zachytiti čísly statistickými.

2. K § 1. a 2. na str. 8 a 9 bych upozornil, že ne všechny znaky kvalitativní jsou vlastnostmi stálými a ne všechny kvantitativní proměnlivými (v § 1.) nebo proměnnými (v § 2.) v obvyklém českém slov těch smyslu. Užívati tedy obojích těch výrazů jako synonym k překladu Yulových „attributes“ a „variables“ mohlo by českého čtenáře věci neznalého splésti. Lépe používati pouze vžitých u nás výrazů z latiny vzatých (ačli se překladatelé nechtějí přidati k překladu Benešem zavedenému: znaky ličné a neličné), a eventuelně vyložiti rozdíl mezi obojími těmi znaky přesněji pod čarou.

3. V § 10. na str. 11 schází definice pojmu „řád třídy“. Když všechny ostatní použité pojmy jsou tu — jistě právem — definovány, bylo by snad záhodno vyplniti aspoň v poznámce tuto mezeru originálu, snad asi takto: „Řádem třídy nazývá se číslo, udávající počet znaků, které jsou v té třídě kombinovány. Třída jest tedy tolikátého řádu, kolik znaků se současně nalézá u všech členů té třídy.“

4. Na str. 18 v § 1. výraz „the universe of discourse“ bylo by snad spíše než slovy „svět úvah“ přeložiti slovy „pozorovaný soubor“ nebo „pozorovaný celek“.

5. Na téže str. 18 v § 1. a 2. v příkladu pro šetření o rozšíření choromyslnosti užívá Y. jakožto znaku časového prostě určení „living in 1901“, což přeloženo přesně, zajisté správně: „žijící v r. 1901.“ Při skutečném šetření by se tu arci musil časový znak čili rozhodná doba určiti jinak, určitým momentem, ne rozpětím celého roku. Aby věc neuváděla čtenáře, aspoň začátečníka, v omyl, bylo by snad dobře upozorniti na to poznámkou překladatelů.

statistice, potřebu používati jí nebo snad i své životní povolání v ní. I ti mohou dojista nejen správně a výhodně používati výsledků statistických, ale i dobře vědecky s nimi pracovati, jen když umějí správně logicky myslit a prostě sčítat, odčítat, násobit a dělit, a jen když mají dosti trpělivosti, svědomitosti a vědecké intuice. Proto však arci (jak jsem ostatně naznačil již ve svých „Základech statistiky“ na str. 109) žádnému statistikovi, který toho dovede, neuškodí, ale vždy může prospěti, používá-li — arci s týmiž vlastnostmi a opatrně — také výsledků teorií matematicko-statistických.