

# Jak vytváří statistika obrazy světa a života. I. díl

---

Část I. [odst. 1,1-1,6, 2,1-2,4]

In: Jaroslav Janko (author): Jak vytváří statistika obrazy světa a života. I. díl. (Czech). Praha: Jednota českých matematiků a fyziků, 1942. pp. 6–16.

Persistent URL: <http://dml.cz/dmlcz/403049>

## **Terms of use:**

© Jednota českých matematiků a fyziků

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://dml.cz>

## ČÁST I.

**(1,1) Hromadné pozorování je praktickou cestou k poznávání.**

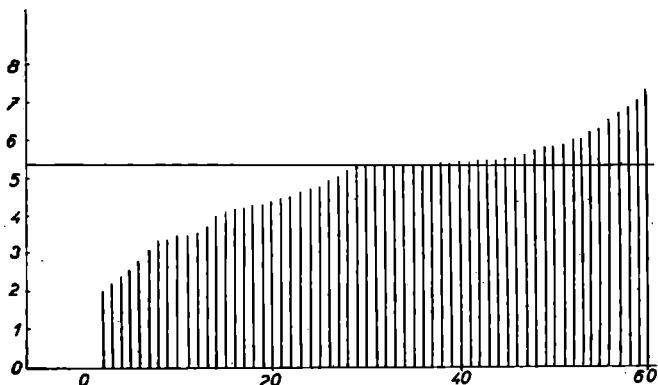
Chce-li majitel nákladního automobilu určité výrobní značky a typu znáti dobu života tohoto svého vozu, musí míti možnost změřit časový interval od okamžiku, kdy byl vůz dán jako hotový výrobek do provozu, do okamžiku skončení jeho poslední jízdy. Tento časový interval bude pro každý jednotlivý vůz téhož typu jiný, takže rozmanitost výsledných čísel nám nedá přesnou odpověď na další otázku, jaká je doba života vozu tohoto typu vůbec.

Podle získané zkušenosti přiřazujeme ke každé radioaktivní substancii určité číslo (konstantu) znamenající poměrnou část atomů, které se rozpadnou v nějakém daném čase. Tento rozpad se pozoruje pomocí fluoreskujícího stínítka. Vezme-li se na př. pro radium za jednotku času den, je zmíněné číslo přibližně jedna miliontina. Určité množství radioaktivní substance klesá tedy a zredukuje se na polovinu po době úměrné zmíněné konstantě, která je pro tu substancii charakteristická. Tato pravidelnost se projevuje vzhledem k tomu, že i nejmenší částička substance obsahuje veliký počet atomů a lze ji potvrditi tím lépe, čím je pozorován vyšší počet rozpadů. Můžeme-li pozorovati přístrojem jen malý zlomek těchto rozpadů, pozorujeme přibližnou stálost. Nemůžeme však dáti přesnou odpověď na otázku, kolik atomů se rozpadne v příští minutě, neboť ten počet se mění; ale je tu jistá pravidelnost t. zv. statistická. Pro ni hledáme vyjádření.

Utrhneme-li nějaké množství vyspělých listů košatého stromu a změříme jejich délku, nebudou všechny stejné; mezi nejkratším a nejdelším bude třeba značný rozdíl. Opět bychom ne mohli dáti určitou odpověď na otázku, jak dlouhý

list má dotyčný strom. Proto hledáme způsob, jakým bychom mohli dát odpověď, která by jasně vyjádřila výsledky našich měření.

Seřadíme je tedy nejprve podle naměřené délky od nejmenšího do největšího a zobrazíme tuto posloupnost úsečkami v obr. 1 stejně od sebe vzdálenými. Shledáme, že konce



Obr. 1. Délka 60 listů s vyznačeným mediánem.

úseček dávají určitý průběh, zvláště kdyby byly spojeny, který svědčí tomu, že změny délek vytvářejí sled naznačující jistou pravidelnost. Chceme se přesvědčit zda tento obraz bude jiný, vezmeme-li v úvahu jiné množství listů téhož stromu. Provedeme tedy s druhým množstvím též postup a shledáme, že průběh v této posloupnosti je zcela podobný; opět úsečky kolem prostřední, která má skoro touž délku jako v prvním případě vykazují od ní malé odchylky a směrem k oběma krajům je menší počet větších odchylek. Prostřední člen této posloupnosti se nazývá medián a lze jej při lichém počtu členů  $2k + 1$  určit jako  $k + 1$  člen od kraje, při sudém počtu  $2k$  jsou dva prostřední členy ( $k$ -tý od každého

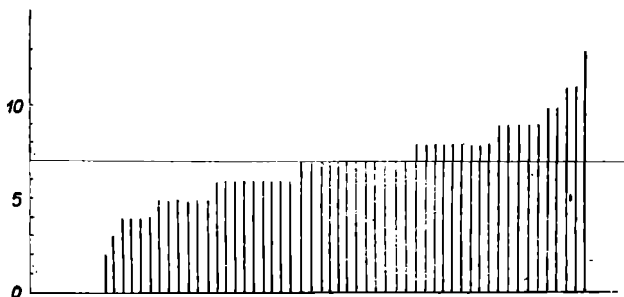
kraje) a hodnotu mediánu určíme jako jejich průměr.  
 V našem zobrazení je jeho velikost vyznačena rovnoběžkou s osou základní. Tak se přesvědčíme, zvláště kdybychom tento postup opakovali vícekrát, že velikost mediánu a zobrazený tvar průběhu délek je pro měřený druh předmětů příznačný čili charakteristický. Ale velmi podobné tvary bychom dostali, kdybychom si znázornili výsledky měření proměnlivého znaku jiných předmětů, neboť jsme svůj příklad zvolili zcela nahodile. Jsme pak vedeni k domněnce, že za těmito pozorovanými jevy je nějaká všeobecná pravidelnost, která nám může pomoci k zjednodušení představ o velikosti hodnot znaku nějakého druhu předmětů.

Zvolený příklad dává tušiti jisté společné podmínky, za nichž vyrůstají listy téhož stromu a vytvářejí jejich délku, takže jsme tu ochotni očekávat nějakou zákonitost. Pokročíme proto dále ve své úvaze, budeme-li pozorovati nějaký jev, který patří mezi t. zv. náhodné, jako je házení kostkou nebo mincí. Hodíme do výše čtrnáct stejných mincí a po dopadu stanovíme počet rubů. Provedeme to třeba 201krát a výsledky, které jsme dostali jsou tyto:

7, 9, 7, 10, 6, 7, 5, 6, 7, 7, 9, 5, 7, 9, 8, 10, 7, 7, 6, 5,  
 8, 8, 8, 7, 8, 11, 8, 9, 5, 5, 7, 6, 4, 11, 6, 8, 10, 2, 5, 7,  
 7, 8, 3, 6, 8, 6, 7, 9, 8, 8, 6, 8, 8, 5, 7, 7, 4, 7, 9, 10,  
 5, 7, 8, 4, 7, 10, 11, 7, 5, 7, 6, 8, 8, 8, 9, 10, 7, 9, 7, 8,  
 8, 6, 7, 7, 7, 7, 5, 5, 7, 6, 7, 8, 9, 5, 7, 4, 5, 8, 7, 4,  
 4, 5, 5, 8, 7, 11, 7, 9, 5, 7, 8, 9, 8, 4, 10, 5, 5, 9, 4, 6,  
 7, 8, 7, 4, 9, 7, 13, 6, 4, 7, 6, 6, 9, 6, 7, 4, 6, 9, 6, 7,  
 6, 5, 6, 11, 6, 4, 5, 8, 7, 6, 10, 4, 9, 6, 8, 4, 8, 8, 7, 9,  
 9, 8, 10, 7, 7, 5, 6, 6, 7, 8, 6, 6, 8, 8, 11, 3, 4, 5, 7, 4,  
 6, 9, 4, 6, 6, 7, 8, 7, 4, 6, 5, 11, 6, 8, 11, 6, 3, 7, 6, 9, 7.

Tak jako jsme seřadili listy podle délky, tak seřadíme nyní pokusy se čtrnácti mincemi podle počtu rubů a výsledek si zase zobrazíme (obr. 2). Každá úsečka kromě krajních případů zastupuje přibližně čtyři výsledky. Dostá-

váme podobný tvar jako v dřívějším případě, takže přicházíme k myšlence, že by nám snad bylo usnadněno zkoumání takových případů, kdybychom odvodili teorii vysvětlující výsledky pokusů s jevy náhodnými. Tuto myšlenku budeme sledovati později.



Obr. 2. Počet rubů na 14 mincích.

(1,2) **Hromadný jev.** Nyní si především dobře všimneme, že k získání určitých poznatků nám nestačí pozorování, po případě měření jednoho předmětu (jedince) určitého druhu, nýbrž je třeba nahromaditi pozorování většího množství předmětů téhož druhu (prvků) čili nastoupiti cestu hromadného pozorování, což je vlastní metoda statistiky. Z jevu pozorovaného na jednotlivých předmětech téhož druhu se skládá jev hromadný, který vzniká působením určitého výseku všeobecných podmínek dění. Hromadný jev lze tedy pozorovati na souboru množství prvků odpovídajících určitému pojmu.

(1,3) **Statistický soubor.** Abychom sestrojili potřebný statistický soubor odpovídající určitému pojmu (člověk, dům, podnik, motorové vozidlo, strom, jablonoňový list, úmrtí, sňatek), musíme definovat statistickou jednotku. Vymezíme tudíž nejprve věcně statistický soubor tím, že vytkneme podstatné znaky, které musí míti každý prvek, který má

býti zahrnut do souboru. Tak přejdeme od skutečných předmětů nebo jevů vnějšího světa k myšlenkovému předmětu (statistické jednotce) tím, že vytkneme znaky, jež považujeme s hlediska cíle šetření za podstatné; uijeme tedy logického postupu zvaného abstrakce.

(1,4) Statistická jednotka. Statistická jednotka vykazuje znaky a) shodné, obsažené v obsahu pojmu, které jsou společné všem prvkům, jež budou pojaty do souboru;

b) vyšetřované, které se u některých prvků vyskytují a u druhých nikoliv (znaky alternativní) nebo se u nich vyskytují v různém stupni (znaky kvalitativní nebo kvantitativní);

c) ostatní, které jsou jednak postižitelné, jednak nepostižitelné a mohou být všem prvkům souboru společné, ale nemusí.

Definice statistické jednotky musí obsahovati znaky shodné, na jejichž základě je sestrojen zkoumaný statistický soubor, který je vzhledem k nim stejnorodý (homogenní) [1]. Není však stejnorodý vzhledem k vyšetřovaným znakům a k některým ostatním znakům. Tak na př. soubor lidí není stejnorodý vzhledem k znaku (alternativnímu) pohlaví nebo vzhledem k znaku (kvantitativnímu) věk. Můžeme však zkoumaný statistický soubor roztržiti podle některého vyšetřovaného znaku a tak z něho odvoditi nové částečné soubory, jejichž prvky mají kromě shodných znaků původního souboru ještě jeden nebo několik dalších shodných znaků (muži třicetiletí, ženy dvacetileté). Tyto soubory jsou stejnorodější než původní soubor, neboť množina shodných znaků je u nich větší. Jsou tedy různé stupně stejnorodosti čili formální homogeneity statistického souboru, jejíž mez je dána především cílem dotyčného statistického šetření.

(1,5) Statistické číslo. Statistika se snaží udati, jako jeden z prvních výsledků statistického šetření, kolik prvků je zahrnuto definicí statistické jednotky čili jaký je rozsah souboru odpovídajícího určitému pojmu. Tak vzniká ze

statistických jednotek první statistické číslo. Tato statistická čísla osvětlují nejdůležitější skutečnosti lidského života ve společnosti a státě, jakož i poměry v přírodě; patří tudíž k základnímu stavu lidského vědění.

Pro hodnocení významu statistických čísel je důležité, abychom měli na paměti:

1. Prvky statistického souboru jsou vzájemně vázány pojmovým společenstvím, které je vyznačeno určitým stupněm stejnorodosti; homogenita je tedy proměnná.

2. Kromě pojmového společenství nemají prvky souboru vzájemné jiné vazby; jsou tedy na sobě nezávislé. Mezi jednotlivými prvky není strukturních vztahů, které však jsou mezi soubory (jednak mezi částečnými navzájem, jednak mezi nimi a původním).

3. Cím je stupeň homogenity vyšší, tím je zpravidla rozsah souboru menší.

(1,6) Statistika. Můžeme nyní také říci, co rozumíme statistikou, abychom vyznačili zorný úhel dalších výkladů.

Statistika je věda, jejímž předmětem je statistický soubor. Je to věda empirická, jejíž jednotnost je založena jednotným způsobem hromadného čili kolektivního pozorování vztahujícího se na množství předmětů nebo událostí.

Jako každá věda empirická má jednak úkol povahy materiální, který plní ve své části popisné a používá k tomu své techniky šetření a zpracování pozorovaných dat, jednak úkol povahy logické, který plní v části teoretické. Tato část musí odvozovati prvky pro své základy ze zkušenosti, aby logická konstrukce měla význam praktický. Spojení částí abstraktní se skutečností tvoří věta o stálosti statistických četností, odvozená ze zkušenosti, již se budeme později blíže zabývat.

Můžeme rozeznávat dva obecné typy hromadného pozorování.

a) První typ je představován posloupností čísel  $x_1, x_2, \dots, x_r$ , která jsou výsledky pozorování resp. měření téhož

znaku na  $r$  předmětech či událostech téhož druhu; jedná-li se o znak alternativní, pak zjišťujeme u každého prvku toliko má-li tento znak nebo nemá a označujeme obyčejně jeho přítomnost číslicí 1 a nepřítomnost číslicí 0, takže výsledkem pozorování je sled jedniček a nul (na př. 010011101....).

b) Druhý typ je rovněž představen posloupností čísel  $x_1, x_2, \dots, x_r$ , která však jsou výsledky  $r$  měření téhož znaku na jednom předmětu. Tyto číselné údaje shromažďujeme, chceme-li si opatřit pro určitou ověřovací metodu odhad chyby měření.

Přesto, že metody zpracování výsledků obou typů hromadného pozorování jsou obdobné, zabývá se druhým typem zvláště teorie chyb [7]. Při našich úvahách budeme mít na mysli první typ hromadného pozorování.

## (2,1) Technika statistického šetření a výsledek jeho v nashromážděných datech.

Číselné údaje získané statistickým šetřením čili hromadným pozorováním, nazýváme také statistická data nebo statistický materiál jakožto souhrn všech záznamů o prvcích zahrnutých do souboru; záznamy musejí býti metodicky bezvadné, aby bylo účelno vyvozovati z nich další úsudky pomocí statistické teorie, jejíž pevné základy tvoří metody, o nichž budeme dále jednati. Proto musí odpovídati logickým podkladům, vyloženým v předcházejícím oddílu, vyspělá technika statistického šetření, a to nejen rozsáhlého, jakým je sčítání lidu nebo sčítání závodů, statistika zahraničního obchodu, statistika mezd a pod., nýbrž i menšího rozsahu jako při studiu souboru nějakého počtu lebek, nebo nevelkých vzorků předmětů výroby průmyslové či zemědělské. Cesta získávání statistických čísel bývala prvním zdrojem „statistických hříchů“. Musí proto odborně a přesně připsati souhrn všech definic a předpisů, jakožto rovinné zrcadlo, v němž se má hromadný jev číselně zobraziti, aby pokud možno nic nedefovalo.



Každému statistickému šetření musí předcházeti jeho příprava. Má být formulován účel šetření čili stanoven úkol nebo problém, který má být dotyčným šetřením řešen nebo objasněn. (Na př. zjištění velikosti a složení obyvatelstva podle řady hledisk.)

**(2,2) Plán šetření povahy logické.** S hlediska účelu šetření musí být vypracován plán šetření jednak pro získání (sbírání), jednak pro zpracování statistického materiálu. Plán šetření musí obsahovati:

1. Vymezení statistické jednotky a to věcné, prostorové a časové. Na př. má-li se zjistiti při sčítání lidu soubor přítomného obyvatelstva na určitém území, je to obyvatel na určitém územním prostoru v kritickém okamžiku (třeba o půlnoci na 1. prosince) přítomný. Znaky shodné mají být udány tak, aby bylo možno v každém konkrétním případě rozhodnouti o tom, patří-li do souboru či nikoliv. Časové vymezení je dáno určitým okamžikem při zkoumání t. zv. jevů trvalých, a určitým časovým intervalem u jevů okamžitých (porod, sňatek, úmrtí,  $\alpha$ -částice vysílané radioaktivním zářením). Věcné vymezení se setkává s obtížemi v tom, že pojmy, jak je dává život, nejsou vždy jasné a jednoznačné (nezaměstnaný) a pojmy, které dává věda nebo právní řád, nejsou vždy všeobecně přijímány, ale bývají často sporné (reálná mzda). Proto musí statistik k určitému účelu sestrojiti často sám definici jednotky k určení pojmu. To je možné při provádění statistiky primérní, t. j. statistiky určené jen k poznání pozorovaných hromadných jevů. Musí tedy definovati pojmy podle možnosti tak jak to vyžadují příslušné vědní obory, technika, národní hospodářství, a pod.; kde však to vyžaduje účel šetření, může se odchýliti a vlastním zásahem některé skupiny zahrnouti či vyloučiti. Při zpracování musí ovšem dbáti toho, aby dal příslušné vysvětlení o pojmovém vymezení a tím také správný význam získaným číslům. Tak na př. při sčítání domů je nutno určití, je-li jednotkou dům ve smyslu jednotky stavebně

technické nebo administrativní a v definici musí být řešeno, jak se zařadí nouzové kolonie bez popisných čísel, weekendové chaty, obývané baráky a pod. Koná-li se šetření za účelem fiskálním, musí tvořit základ definice jednotky statistické, definice zákona o dani domovní. Je zřejmo, že v každém případě bude okruh zahrnutých případů, a tedy soubor, jiný. Tam, kde má být statisticky zpracován materiál, který byl původně určen jinému účelu (na př. správnímu), vzniká statistika sekundární. Zde na rozdíl od primární statistiky je statistik vázán materiálem, jehož pojmy musí převzít a nezná vlivu na jejich vytváření; musí je jen jasně uvést, aby každý statistický spotřebitel tyto základní skutečnosti znal a musil jich dbát při posuzování čísel a porovnávání s výsledky jiných statistik.

2. Stanovení vyšetřovaných znaků (na př. pohlaví, věk, povolání, národnost). Rozhodným je splnění účelu šetření, který někdy sdružuje řadu zájmů. Při statistice požárů na př. má věda hospodářská hlavně zájem na škodách vzniklých národním hospodářství, požární pojišťovny potřebují zjistit pro výměru pojistné prémie četnost požárů a řadu vlastností poškozených předmětů, aby mohly sestojit nebezpečenské třídy.

3. Určení, které částečné soubory jest odvoditi z původního souboru. To vyžaduje stanovení podrobností o tom, podle kterých znaků má být původní soubor roztržěn a v jakých kombinacích.

4. Rozhodnutí, která statistická čísla mají být vypočtena (absolutní, relativní, různé charakteristiky, s nimiž se dále seznámíme) a do jakých tabulek budou seřazena a publikována.

(2,3) Plán organizačně technický pro sbírání a zpracování materiálu. Tato rozhodnutí povahy logické jsou pak doprovázena opatřeními organizačně technickými pro sbírání a zpracování statistického materiálu. Podle statistické jednotky a vyšetřovaných znaků jakož i hledisek celého zpracování se vypracuje dotazník nebo sčítací arch, v němž

jsou položeny všechny potřebné otázky co nejjasněji se zřetelem k osobám, které je budou vyplňovat, aby odpovědi byly snadné a správné. Obyčejně se připojuje také návod k vyplňování formulářů čili instrukce. Šetření se provádí buď individuálními lístky, t. j. pro každý prvek souboru, nebo hromadnými, tedy sběrnými listinami pro celé skupiny (domácnost, bytová strana), které nabývají značného rozšíření vzhledem k zpracování pomocí elektrických strojů, poněvadž usnadňují běžné vyznačování znaků podle předem vypracovaných klasifikačních schemat a dírkování, kterým se přenášejí znaky na tuhé štítky vhodné k třídění. Třídění se může prováděti při malém rozsahu souboru pomocí čárkování, lepení známek z útržkového bloku nebo vkládáním sčítacích lístků čili štítků, při velkém rozsahu pomocí strojů [3].

Musí býti stanovena dále úplná organizace pro způsob, jakým se provede sbírání a zpracování. Rozdělí se vhodným místům a orgánům úkoly rozeslání dotazníků, vyplnění, revise, sbírání a kontroly buď s hlediska statistického zpracování centrálního, při němž se veškeren materiál shromáždí v jednom místě nebo decentralisovaného v několika místech. Moderní potřeby vyžadují při rozsáhlejších šetřeních množství znakových kombinací. To vyžaduje složitého zpracování pomocí vyspělé techniky, které může býti provedeno jen dobře vyzbrojenými statistickými centrály, jež musejí všechen původní materiál do svých rukou převzít a přesvědčiti se o jeho úplnosti. Nespadá do rámce našich úvah o obecných metodách výklad podrobností technického procesu, který musí býti vždy přizpůsobován konkrétnímu šetření a studován z učebnic k tomu cíli sepsaných jako [2] a [3]: do přípravy plánu určitého šetření patří také studium metody a techniky užití při takovém šetření snad již dříve nebo jinde a posouditi jak dobře vedla k zamýšlenému cíli.

(2,4) **Plán publikační.** Vyložili jsme si metodický postup, jímž vznikají statistická čísla, která se pak jako výsledky statistického šetření po přezkoumání číselné správ-

nosti a vnitřní shody sestavují do číselných přehledů čili tabulek nejprve soustředných. Z nich se podle publikačního plánu odvozují hlavní tabulky podrobné, pak přehledné, srovnávací a jiné. Pro formální úpravu jednotlivých tabulek, jež jsou vyznačeny nápisem, hlavičkou seřazující logicky sloupce a legendou popisující obsah řádků, platí důležité zásady; jejichž splnění vyžaduje posláním tabelárních přehledů [2]. Výsledky rozsáhlých šetření statistických úřadů jsou publikovány v publikacích jednak periodických, jednak v obsáhlejších dílech pramenných [3]. Vyskytuje se také nový systém „mikroskopických archivů“. Podle něho se obsáhlé tabulky, jichž je však třeba k nějakému účelu často užívat, fotografují na velmi malou plošku, takže se sto čtverečních jednotek tabulky psané nebo tištěné objeví na třech čtverečních jednotkách filmu. Tabulky tohoto mikroskopického archivu se promítají, aby je bylo možno čísti. Takový postup urychlí některé práce užívající tohoto materiálu a chrání jej před opotřebením.

Ke konci tohoto odstavce třeba ještě zvláště zdůraznit, že jakmile vyjádříme rozsah souboru určitým číslem, které se vztahuje na prvky s jistou množinou společných znaků, zmocní se potom tohoto čísla matematika. Při jejích operacích zmizel empiricko-statistický význam tohoto čísla. Kdyby se jednalo na př. o pojem „hodiny“ definovaný jako „zařízení k měření času“, jehož obsah je tedy dán dvěma znaky, zahrnoval by hodiny sluneční, přesypací, kyvadlové věžní, pokojové, kapesní, náramkové ve všech rozmanitých druzích atd. Tomuto pojmu odpovídá jistý statistický soubor, jehož rozsah je dán statistickým číslem, složeným ze statistických jednotek, které mají dva znaky společné, v ostatních pak ponechávají velikou rozmanitost; a tyto statistické jednotky se tu sčítají. Zpracování matematické přiblíží k takovému číslu jako k ryzímu číslu, kde mezi jednotkami není vůbec rozdílu. Proto musí statistik při rozboru výsledků na konec viděti vždy za číslem jasně jeho podklad empiricko-statistický.