

- Základní pojmy
- Etapy statistického zkoumání

1.

Úvod do popisné statistiky



Cíl kapitoly

Jak jsme již zmínili, celý text studijní opory se pokusí popsat metody popisu především rozsáhlých statistických souborů. Proto než přistoupíme k těmto konkrétním metodám, musíme si nejprve vymežit základní pojmový rámec ze kterého budeme dále vycházet. První kapitola Vás proto seznámí s definicemi a obsahem pojmů jako jsou statistický soubor, statistický znak, proměnná. Stejně tak se seznámíte s možnými přístupy při vytváření statistických souborů.



Časová zátěž

2 hodiny (1. týden v říjnu)

1.1 Základní pojmy

Na pojem statistika je možno pohlížet z několika odlišných pohledů. Je možno ji chápat jako vědu, která se zabývá popisem metod, které slouží k analýze tzv. hromadných jevů. Tedy jevů, které se vyskytují ve větším měřítku u velkého počtu prvků (jedinců). Stejně tak je možno statistiku chápat jako soustavnou činnost spojenou se sběrem, tříděním a vyhodnocováním informací (označovaných jako statistické údaje). Někdy také bývá pojem statistika ztotožněn přímo s těmito údaji.

Výchozím pojmem všech pojetí je zjevně pojem **hromadného jevu**. Ne všechny číselné údaje popisující určitou skutečnost je možno považovat za statistické. K vyslovení závěrů o zkoumaných jevech je nutno, aby byla zajištěna hromadnost pozorování. Jednotlivá pozorování nikdy nevedou k získání přesného – tedy statistického – obrazu skutečnosti.

Statistická jednotka

Ve statistickém zkoumání se tedy zajímáme o ty (hromadné) jevy, které se vyskytují u velkého počtu prvků. Tyto prvky jsou ve statistice nazývány **statistickými jednotkami** a jsou považovány za výchozí nejmenší jednotku statistického pozorování.

V praktických úlohách mohou statistickými jednotkami být osoby (například počet nezaměstnaných osob, počet osob pobírajících sociální dávku apod.), organizace (počet podniků, příspěvkových organizací), věci, události, známky, hodnocení apod.

Statistický znak

Vlastnosti statistických jednotek jsou nazývány **statistickými znaky**. Statistické znaky je možno rozdělit do dvou skupin – na znaky kvantitativní a kvalitativní. **Kvantitativní** statistické znaky jsou takové, jejichž varianty (obměny) je možno vyjádřit číselně. V případě, že tyto varianty jsme schopni vyjádřit pouze slovně, hovoříme o znacích **kvalitativních**.



Je zřejmé, že zvolenou statistickou jednotku je možno charakterizovat několika znaky. Například statistickou jednotku zaměstnance neziskové organizace lze charakterizovat několika statistickými znaky. Mezi ty kvantitativní by patřily například jeho věk, průměrná hrubá mzda, stáří, délka zaměstnání

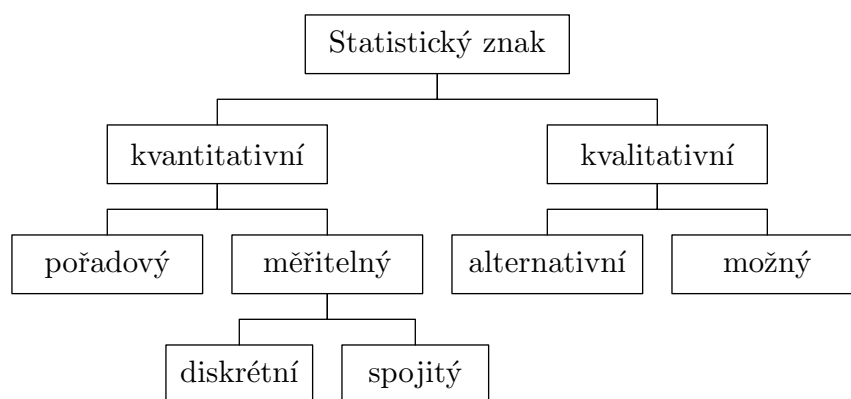
apod. Kvalitativními znaky by pak mohly být například dosažené vzdělání, pozice v organizaci či místo bydliště.

V některých pramenech bývá pojem statistického znaku nahrazován pojmem statistická proměnná (z angl. variable).

Samotné statistické znaky je možno dělit i podle jiných kritérií. Pokud může znak nabývat pouze dvou variant, hovoříme o alternativním znaku (například pohlaví pracovníka). Kvantitativní znaky se obvykle dále dělí na diskrétní a spojité. Diskrétní (nespojité) statistické znaky mohou obecně nabývat pouze některých číselných hodnot, zatímco pro spojité znaky jsou možné teoreticky jakékoliv hodnoty.

Příkladem spojitého statistického znaku je například průměrný věk obyvatel vybrané země. Diskrétním statistickým znakem jsou například výsledné známky u zkoušky z vybraného předmětu.

Výše uvedené možnosti třídění statistických znaků naznačuje obrázek 1.1.



Obrázek 1.1: Třídění statistických znaků.

Pro potřeby dalších analýz je také vhodné vymezit pojem statistického souboru. Statistický soubor je množina všech statistických jednotek, u nichž zkoumáme jednotlivé statistické znaky.

Ve výše uvedeném příkladě by statistickým souborem byl například soubor všech zaměstnanců neziskové organizace.

Statistický soubor je neprázdná množina zkoumaných objektů, které mají z daného hlediska (zkoumaných jevů) společné vlastnosti.

Pokud studujeme u každé statistické jednotky pouze jeden znak, nazýváme tento soubor **jednorozměrným**. V případě, že zkoumáme více znaků u téže jednotky, označujeme soubor za **vícerozměrný**.

Je také vhodné rozlišit pojem základní soubor, který označuje soubor všech statistických jednotek, u nichž provádíme statistické zkoumání, od pojmu výběrový soubor. Výběrový soubor je pouze výsekem základního souboru (jeho podmnožina), který pokud možno co nejlépe charakterizuje základní soubor.



Statistický soubor



Blíže se s pojmem výběrový soubor seznámíme v kapitole 5 věnované tvorbě bodových a intervalových odhadů.

1.2 Etapy statistického zkoumání

Konkrétní statistickou praxi při pořizování nejrůznějších analýz je možno rozdělit do několika na sebe navazujících etap. Tyto etapy jsou v současné době již do jisté míry standardizované a je možno je rozdělit následovně:

- etapa statistického zjišťování (statistického šetření)
- etapa statistického zpracování získaných údajů
- etapa statistického vyhodnocování (analýz)

1.2.1 Statistické zjišťování

Statistické zjišťování

Základním kamenem každé kvalitní statistiky je kvalitní datová základna. Její pořízení proto patří k nejdůležitějším, ale současně také nejobtížnějším, úkolům statistické práce. **Statistické zjišťování** tedy slouží především jako nástroj pořizování statistických údajů. Statistickým údajem je míněna jakákoli obměna (varianta) statistického znaku, kterou zjišťujeme u příslušné statistické jednotky. Proto je při zjišťování nejprve nutno stanovit okruh jednotek, u kterých budeme hodnoty statistických údajů zjišťovat. Stejně tak je nutno určit kdo bude zjišťování provádět, jakým způsobem a v jakém časovém okamžiku.

Subjekt (osoba), který poskytuje příslušné údaje o statistické jednotce je nazýván **zpravodajskou jednotkou**. Zpravodajská jednotka může, ale také nemusí, být totožná se statistickou jednotkou.



Například, budeme-li (pro potřeby státní statistiky) zjišťovat věk pracovníků určité neziskové organizace, bude statistickou jednotkou každý ze zaměstnanců této organizace. Za zpravodajskou jednotkou bude však spíše určena nezisková organizace jako celek, neboť zjišťování věku u každého zaměstnance zvlášť by bylo organizačně i finančně neefektivní.

Pro stanovení časového okamžiku statistického zjišťování je nutno rozlišit dva typy zjišťovaných údajů. V případě, že zjišťované údaje se pořizují jako charakteristika určité skutečnosti za zvolený časový interval, pak je nutno zvolit **rozhodnou dobu** zjišťování. Pokud jsou data pořizována jako stav k určitému datu (okamžiková), stanovujeme **rozhodný okamžik** zjišťování.



Zjišťujeme-li počet pracovníků v neziskové organizaci, je nutno stanovit rozhodný okamžik – například 1. leden každého roku. V případě, že nás zajímá objem prostředků vydaných touto organizací v souvislosti s jejím posláním, je nutno stanovit rozhodnou dobu – například jeden rok, či jeden měsíc.

Pro správné naplnění požadavků této etapy statistického zkoumání je také nutno stanovit, do kdy budou požadována data porížena – **dobu zjišťování**. Na začátku zjišťování je také nutno udělat rozhodnutí o **rozsahu zjišťování**, tedy zda bude zjišťování prováděno u všech jednotek (vyčerpávající zjišťování) či pouze u stanoveného vzorku (výběrové zjišťování).

Vlastních metod pořizování statistických údajů je v současné statistické praxi používána celá řada. Mezi nejvýznamnější patří například:

- přímé pozorování
- dotazování (korespondenční či expediční)
- výkaznictví
- statistické šetření (soupisy, znalecký odhad, anketa)

Přímé pozorování spočívá v přímém sledování jednotlivých statistických jednotek. Zjišťované hodnoty zkoumaných znaků jsou získávány například sčítáním, vážením, či měřením. **Dotazování** je možno provádět buď **expediční** metodou pomocí sčítacích komisařů (najatých osob, které přímo v terénu zjišťují potřebné údaje v interakci se zpravodajskými jednotkami), nebo **korespondenční** metodou pomocí formulářů (obvykle ve formě dotazníků), které zpravodajská jednotka sama vyplní. Korespondenční metoda se někdy nazývá metoda sebesčítání.

Výkaznictví je samostatným a do jisté míry nejvýznamnějším způsobem pořizování statistických údajů. Vychází z předem vytvořeného formuláře, nazývaného **výkaz**, který předkládá příslušná zpravodajská jednotka v pravidelných intervalech. Uvedená metoda je však použitelná pouze v případech, kdy existuje (například zákonem daná) možnost vymáhání těchto údajů u příslušné zpravodajské jednotky. V opačném případě hrozí, že statistické zjišťování může být neúspěšné pro neochotu zpravodajských jednotek poskytovat požadované údaje.

Statistická šetření mohou nabývat několika forem. Zjišťujeme-li skutečnosti přímo na místě a fyzicky evidujeme stav zde zjištěný, hovoříme o soupisu (censu). Zvláštní formou censu, kdy je stav zjišťován, či daná skutečnost pouze odhadována odborníky, je **znalecký odhad**. Znalecké odhady jsou využívány zejména tam, kde není možno provést přímého soupisu a kde by ani využití výběrového šetření nebylo možné.

Nejvýznamnějším soupisem, který je ve většině zemí světa prováděn státní statistikou, je populační census – sčítání lidu. Údaje jsou pro potřeby tohoto soupisu zjišťovány obvykle pomocí poučených osob (sčítacích komisařů) a jsou zjišťovány u všech statistických jednotek.



1.2.2 Statistické zpracování

Jsou-li v rámci statistického zjišťování shromážděny všechny potřebné údaje, je možno přistoupit k jejich zpracování. Obvykle je nutno nejprve tato hromadná data nějakým způsobem seřadit, případně shrnout. Jsou zde využívány nejrůznější nástroje – tabelování, statistické grafy rozdělení četností apod. (Některé z těchto tyto metod jsou popsány v následující kapitole.) Samotnému zpracování údajů však musí předcházet fáze jejich kontroly. Kontrola údajů musí být trojího typu:

- **formální** – přezkoušení správnosti provedených početních úkonů (**součtů, součinů**)
- **početní** – obvykle je zajištěna použitím výpočetní techniky

- **logická** – vychází ze znalosti věcné podstaty zkoumaných jevů. Posuzujeme, zda se získané údaje pohybují v předpokládaných mezích, či zda je překračují.



Shrnutí kapitoly

Každé statistické zkoumání se opírá o dobře vymezenou datovou základnu. Kapitola shrnuje některé z nejvýznamnějších pojmů, které jsou při vytváření této základny používány. Klíčovým je zejména pojem statistického znaku, který označuje typickou (zkoumanou) vlastnost nebo skutečnost. Statistické znaky, které jsou zjišťovány pro statistické potřeby, je možno dělit podle nejrůznějších kritérií. Pro další zkoumání je nejvýznamnější dělení na znaky kvantitativní a kvalitativní.

Druhá část kapitoly se věnuje naznačení kroků, které je při statistickém zkoumání nutno dodržet. Tyto kroky, obvykle nazývané etapy, spočívají v pravidlech pro pořizování, zpracování a vyhodnocování statistických údajů. Východiskem je etapa pořizování, která se opírá o celou řadu standardních nástrojů, jako jsou výkazy, soupisy, odhady či dotazy.



Otázky k zamyšlení

1. Nalezněte další příklady jednotlivých druhů statistických znaků (dle členění na obrázku 1.1).
2. Jaké je Vaše nejvyšší dosažené vzdělání? Pokuste se vaši odpověď zařadit do typologie statistických proměnných (znaků).