

JEDNOROZMĚRNÁ POPISNÁ STATISTIKA – 1. ČÁST

Pojem statistika lze chápat v zásadě ve třech pojetích:

1. číselné nebo slovní údaje (data) a jejich souhrny o hromadných jevech
2. činnost spočívající ve sběru, zpracování a vyhodnocování dat o hromadných jevech
3. teoretická disciplína (věda), která zkoumá zákonitosti hromadných jevů, resp. souhrn vědeckých metod sběru, zpracování a analyzování dat.

Základní statistické pojmy

Hromadné jevy

- takové skutečnosti, které se vyskytují mnohokrát a mohou se znovu opakovat
- jevy, které se vyskytují v masovém měřítku u velkého počtu prvků (běžně alespoň 30)
- pouze v případě hromadnosti lze předpokládat, že to, co je ve zkoumaných vlastnostech prvků podstatné a pravidelné převáží nad tím, co je individuální, náhodné atd.

Statistický soubor

- množina prvků s přesně stanovenými shodnými vlastnostmi (např. množina osob, organizací, atd.).

Statistická jednotka

- prvek statistického souboru
- individuální nositel vlastností daného statistického souboru.

Rozsah statistického souboru

- počet jednotek statistického souboru (symbolické značení – n , N).

Existují dvě možnosti přístupu ke statistickému souboru, jejich chápání je přitom relativní a závisí na konkrétní situaci.

❖ **Základní soubor (populace):** statistický soubor všech jednotek, které jsou předmětem zkoumání, obvykle velmi rozsáhlý, rozsah značíme N .

❖ **Výběrový soubor (výběr):** vzorek ze základního souboru, pořízený tak, že se určitým způsobem vyberou pouze některé jednotky, rozsah značíme n .

Statistický znak

- označení (odraz) určité vlastnosti, kterou má v té či oné míře každá jednotka daného statistického souboru
- u souboru osob např. věk, váha, výška, atd.

Hodnota statistického znaku (pozorování)

- míra dané vlastnosti (statistického znaku) u každé jednotky statistického souboru.

Počet hodnot (pozorování) = rozsah souboru (n).

Obměna (varianta) statistického znaku

- hodnota ve smyslu vyjádření různého stupně dané vlastnosti.

Počet variant (k) \leq rozsah souboru (n).

Statistický znak shodný: v daném statistickém souboru nabývá pouze jedné varianty.

Statistický znak proměnný: v daném statistickém souboru nabývá více než jedné varianty. Ekvivalentní označení: **statistická proměnná**.

Druhy proměnných

Klasifikace proměnných může být prováděna z několika různých hledisek. Správné určení druhu proměnných je nezbytné pro volbu adekvátních metod jejich zpracování a analýzy.

1. Způsob vyjádření hodnot proměnné

- **slovní (kategoriální, alfabetické, kvalitativní):** jsou vyjádřeny slovy
- **číselné (numerické, kvantitativní):** jsou vyjádřeny čísly.

2. Typ vztahů mezi obměnami a hodnotami proměnné

- **nominální (jmenné, názvové):** slovní proměnné, jejichž obměny nelze hierarchicky uspořádat, tzn., že nelze jednoznačně stanovit, která je nižší a která vyšší. O jejich obměnách lze pouze konstatovat, zda jsou stejné nebo různé. *Např.: pohlaví, jméno, rodinný stav, atd.*
- **ordinální (pořadové):** slovní i číselné proměnné, jejichž obměny lze jednoznačně seřadit od nejnižší k nejvyšší nebo obráceně. Jejich obměny lze porovnávat rozdílem, ale ne podílem. *Např.: nejvyšší dokončené vzdělání, hodnosti v armádě, pořadí v soutěži, atd.*
- **metrické (měřitelné):** vždy číselné, jsou udány v určitých měrných jednotkách – vyjadřují tedy velikost měřených vlastností. Nabývají jak kladných, tak nekladných hodnot. Lze změřit o kolik je jedna obměna větší či menší než druhá. Obměny lze porovnávat rozdílem, někdy také podílem (ne vždy – pokud jsou některé obměny záporné či nulové, není to možné). *Např.: teplota vzduchu, zisk podniku, atd.*
- **kardinální (stěžejní):** ty metrické proměnné, které nabývají pouze kladných hodnot, jejich obměny lze porovnávat jak rozdílem, tak podílem. Je tedy možno změřit, o kolik měrných jednotek je jedna obměna větší či menší než druhá a také kolikrát je jedna obměna větší či menší než druhá. *Např.: věk, váha, výška, atd.*

3. Počet variant, kterých proměnné nabývají

- **alternativní:** nabývají pouze dvou obměn
- **množné:** nabývají více než dvou obměn.

4. Počet hodnot, kterých proměnné nabývají

- **diskrétní (nespojité):** nabývají spočetně mnoha hodnot z konečného či nekonečného intervalu. Např.: počet dětí v rodině, atd.
- **spojité (kontinuální):** nabývají všech hodnot z konečného či nekonečného intervalu. Např.: spotřeba elektrické energie, příjem, atd.

Ke statistickému zkoumání je třeba mít k dispozici statistická data (údaje, pozorování), která získáváme prostřednictvím statistických šetření. Šetření v sociálně-ekonomické oblasti přitom vykazují určitá specifika, takže data z nich pocházející je třeba zpracovávat za pomoci vhodných statistických metod a postupů.

Zpracování dat

- hodnoty proměnných = data = údaje = pozorování
- většinou jde o větší množství údajů, které jsou více či méně nepřehledné
- prvním krokem při zpracování dat je proto jejich zpřehlednění (setřídění)
- zpřehlednění provádíme za pomoci *statistických tabulek a grafů*
- cílem tohoto procesu je, aby vynikly charakteristické rysy a zákonitosti souboru.

Statistické tabulky

Tabulka prostého rozdělení četností

- je vhodná při zpracování diskrétní proměnné s nepříliš velkým počtem obměn.

Tabulka 2.1

Obměna proměnné x_i	Četnost		Kumulativní četnost	
	absolutní n_i	relativní p_i	absolutní	relativní
x_1	n_1	p_1	n_1	p_1
x_2	n_2	p_2	$n_1 + n_2$	$p_1 + p_2$
.
.
.
x_k	n_k	p_k	n	1
Celkem	n	1	×	×

$$p_i = \frac{n_i}{\sum_{i=1}^k n_i} = \frac{n_i}{n} ; \quad \sum_{i=1}^k n_i = n ; \quad \sum_{i=1}^k p_i = 1$$

Tabulka intervalového rozdělení četností

- je vhodná v případě zpracování diskrétní proměnné s mnoha obměnami nebo spojitě proměnné
- klíčovým problémem je určení optimálního počet intervalů (k), na které rozdělíme variační rozpětí (R)
- k tomu slouží různá pravidla (např. Sturgesovo pravidlo: $k \approx 1 + 3,3 \log_{10} n$)
- při výpočtech lze pak každý interval zastoupit jeho středem
- výsledky takovýchto výpočtů jsou samozřejmě pouze přibližné.

Grafy

Další možností zpřehlednění statistických údajů je jejich grafické znázornění. Existuje mnoho druhů grafů, vždy je třeba vybrat takový, který odpovídá charakteru dat.

Polygon četností

- spojnicový graf
- vhodný pro znázornění prostého rozdělení četností.

Histogram četností

- sloupkový graf
- vhodný pro znázornění intervalového rozdělení četností.

Výsečový graf (piechart)

- vhodný pro znázornění rozdělení četností nominální proměnné.

Sloupkový graf (barchart)

- vhodný pro znázornění rozdělení četností nominální proměnné.

Krabicový graf

- vhodný pro znázornění extrémních hodnot a kvartilů
- užitečný pro srovnání určitého numerického znaku ve dvou či více souborech.