

MODUL 2: JAK PRACOVAT S INFORMACEMI

Garant: Ing. Alena Kopfová

Cíl modulu:

Cílem semináře je naučit studenty orientovat se v základních statistických pojmech, jež jim pomohou nejen při studiu ekonomie, ale i při studiu jiných předmětů a v běžném životě. Na semináři budou vysvětleny a procvičeny výpočty charakteristiky polohy (modus, medián, aritmetický průměr, geometrický průměr) a charakteristiky variability (variační rozpětí, průměrná odchylka, rozptyl a směrodatná odchylka). Studenti také získají základní povědomí o indexech.

Pomůcky s sebou na semináři:

kalkulačka, folie a fixy pro prezentaci, znát základní statistické pojmy, jejichž přehled je uveden níže

Doplňující literatura:

BUDÍKOVÁ, M., MIKOLÁŠ, Š., OSECKÝ, P. *Popisná statistika*. 3. vyd. Brno : Přírodovědecká fakulta MU Brno, 1998.

VÁGNER, I., MORAVANSKÝ, D. *Statistika pro posluchače právnické fakulty*. Brno : Právnická fakulta MU Brno, 1982.

ZÁKLADNÍ STATISTICKÉ POJMY¹

Hromadný jev – jev, který je charakterizován dostatečně velkým počtem výskytů; tedy nikoliv izolovaný, jednorázový.

Náhodný jev – jev, jehož výskyt nebo forma projevu není před pozorováním jednoznačně určena, tzn. v daném okamžiku může nebo nemusí nastat a je-li kvantitativně ohodnocen, může nabýt jedné z několika alternativních hodnot.

Statistická jednotka – základní prvek, na kterém je možno zkoumat konkrétní projev určitého hromadného jevu, jako přesně vymezený objekt pozorování. Předpokládá se, že tyto prvky mají podstatné vlastnosti z hlediska účelu společné. Statistickou jednotkou může být např. člověk, podnik, domácnost, událost (narození atd.), činnost lidí apod.

Při každé konkrétní statistické analýze je třeba přesně vymežit statistickou jednotku z hlediska věcného, prostorového a časového. **Věcné vymezení** znamená určení, co v daném případě je a co není statistickou jednotkou. **Prostorové vymezení** znamená určení území, na něž se bude zkoumání vztahovat. **Časové vymezení** znamená určení okamžiku nebo časového období, po které bude statistická jednotka zahrnuta do zkoumání.

Statistický soubor – souhrn statistických jednotek, který podléhá danému zkoumání. Vzhledem k tomu, že statistika zkoumá hromadné jevy, musí mít každý statistický soubor dostatečný rozsah (velikost), aby bylo možno vyvozovat závěry o podstatných společných zákonitostech a vývojových tendencích sledovaných hromadných jevů. Soubory malého rozsahu, u nichž se může výrazně projevit působení náhodných vlivů, jsou z tohoto pohledu pro statistické zkoumání málo významné.

Základní soubor – statistický soubor, který je předmětem našeho zkoumání a o němž chceme činit příslušné závěry. Získání úplné znalosti o sledované vlastnosti základního souboru je možné jen přešetřením všech statistických jednotek, z nichž se základní soubor skládá (provedením vyčerpávajícího zjišťování).

Výběrový soubor – část základního souboru, který bezprostředně podporuje zkoumání a z něhož usuzujeme na vlastnosti základního souboru. Skutečnost, že se v mnoha případech omezujeme na výběrové soubory, je dána pracností, nákladností a někdy také praktickou nerealizovatelností vyčerpávajícího zjišťování.

Statistický znak – charakteristika sloužící jako postižitelný nebo měřitelný výraz vlastnosti statistické jednotky. Někdy může být totožný se zkoumanou vlastností, jindy je volba znaků obtížnější, protože některé vlastnosti nelze charakterizovat jediným statistickým znakem. Obdobně jako statistické jednotky můžeme třídit i statistické znaky. Mluvíme pak o znacích prostorových, časových a věcných. Věcné znaky se dále dělí na kvalitativní, které nelze přímo vyjádřit číslem (např. druh zaměstnání, studijní obor) a kvantitativní = číselné (např. počet dopravních nehod za měsíc leden v jihočeském kraji).

Podle počtu variant (obměn), jichž může sledovaný znak nabýt, rozdělujeme dále kvalitativní znaky na alternativní, které se mohou vyskytovat jen ve dvou variantách (pohlaví mužské a

¹ VÁGNER, I., MORAVANSKÝ, D. *Statistika pro posluchače právnické fakulty*. Brno : Právnická fakulta MU Brno, 1982.

ženské) a množné (povolání, národnost). Kvantitativní znaky pak členíme na spojité a nespojité (diskrétní). Spojité znaky mohou nabýt libovolných reálných hodnot z daného intervalu (vyjádřené v celých i zlomkových hodnotách, např. kilogramy), nespojité pak jen určitých, zpravidla celočíselných, hodnot (počet zaměstnanců, kusů výrobků). Je-li to užitečné, pohlížíme často na v podstatě diskrétní veličinu (výše mzdy v Kč) rovněž jako na spojitou.

U časově vymezených znaků rozlišujeme dvě základní kategorie – znaky určené časovým okamžikem (počet zaměstnanců k 31.1.2006) a znaky určené časovým intervalem (počet dopravních nehod za poslední čtvrtletí roku 2005).

Náhodná veličina – proměnná, přiřazující výsledku pozorování konkrétní hodnotu statistického znaku z množiny hodnot, kterými tento znak může být ohodnocen. Přitom není předem známo přesně (pouze s určitou pravděpodobností), kterou z těchto hodnot proměnná konkrétně nabude – obecně může nabýt hodnoty znaku jakékoliv jednotky základního souboru.

Podle toho, jaký statistický znak náhodná veličina charakterizuje, rozdělujeme náhodné veličiny na spojité a diskrétní. Např. při vrhu dokonalou kostkou může nabýt diskrétní náhodná veličina některé z hodnot 1, 2, 3, 4, 5, 6, každé s pravděpodobností 1/6.

Četnost – počet výskytů dané hodnoty statistického znaku odpovídající počtu realizací určitého hromadného jevu. Rozlišujeme četnosti absolutní (úhrnný počet výskytů dané hodnoty), relativní (podíl počtu výskytů dané hodnoty a počtu všech pozorovaných hodnot). Četnost kumulativní (vzestupná, sestupná), pak udává úhrn výskytů všech hodnot menších, větších nebo rovných dané hodnotě.

STATISTICKÉ CHARAKTERISTIKY (MÍRY) SOUBORU

Pro hodnocení rozdělení četností:

- charakteristiky polohy
- charakteristiky variability
- charakteristiky šikmosti (symetrie)
- charakteristiky špičatosti (koncentrace)

Charakteristiky polohy

Často nazývané středními hodnotami – číslo, které jednoduše zastupuje (reprezentuje, nahrazuje) jednotlivé hodnoty uvažovaného statistického znaku, udává polohu daného rozdělení četností a charakterizuje obecnou velikost jevu v daném souboru.

Aritmetický průměr

Konkrétní výpočet je dán charakterem příslušného souboru:

a) Prostý aritmetický průměr – u souboru, v němž se hodnoty znaku neopakují

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

n ... počet prvků souboru

x_i .. hodnota znaku x u i -tého prvku souboru

b) Vážený aritmetický průměr – u souboru s vícenásobnou četností hodnotu znaku

$$\bar{x} = \frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i}$$

k ... počet skupin se stejnou hodnotou

x_i ... hodnota znaku x v i -té skupině

n_i ... četnost i -té skupiny

c) U souboru s intervalovým rozdělením četností se stanoví aritmetický průměr jako vážený aritmetický průměr, přičemž reprezentuje každý interval hodnota středu intervalu.

Medián

Medián je prostřední neboli centrální hodnota znaku. Je to ta hodnota, která rozděluje soubor na dvě stejné části. Pro určení mediánu je nutné statistickou řadu nejprve uspořádat podle velikosti hodnoty znaku.

Pokud má statistická řada lichý počet prvků, pak je mediánem prostřední prvek řady.

U statistické řady se sudým počtem prvků se medián určí jako aritmetický průměr z nejvyšší hodnoty dolní poloviny a z nejnižší hodnoty horní poloviny uspořádané řady.

Modus

Modus je nejčastější hodnota statistického znaku. Je to tedy hodnota, která se v daném souboru vyskytuje nejčastěji.

Geometrický průměr

Geometrický průměr je v souboru o n prvcích n -tá odmocnina ze součinu jednotlivých kladných hodnot sledovaného znaku.

a) Prostý geometrický průměr

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^n x_i}$$

n ... rozsah souboru

x_i ... hodnota i -tého prvku souboru

\prod ... symbol pro součin

b) Vážený geometrický průměr

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^k x_i n_i}$$

n ... rozsah souboru $\left[n = \sum_{i=1}^k n_i \right]$

k ... počet skupin

x_i ... hodnota znaku x v i -té skupině

n_i ... četnost i -té skupiny

Charakteristiky variability

Tedy charakteristiky měnlivosti či kolísavosti hodnot, jichž sledovaný znak v souboru nabývá.

Variační rozpětí

Variační rozpětí je rozdíl mezi nejvyšší a nejnižší hodnotou sledovaného znaku v statistickém souboru.

$$R_x = x_{\max} - x_{\min}$$

Samo o sobě není vhodnou mírou variability, neboť nic neříká o rozptýlení pozorovaných hodnot kolem střední hodnoty.

Průměrná odchylka

Průměrná odchylka je aritmetický průměr z absolutních hodnot odchylek znaku od některé střední hodnoty, nejčastěji od aritmetického průměru, což je dále předpokládáno.

a) v prostém tvaru

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

n ... rozsah souboru

x_i ... hodnota i -tého prvku

\bar{x} ... aritmetický průměr

b) ve váženém tvaru

$$\bar{d} = \frac{\sum_{i=1}^k |x_i - \bar{x}| \cdot n_i}{\sum_{i=1}^k n_i}$$

k ... počet skupin

x_i ... hodnota znaku x v i -té skupině

\bar{x} ... aritmetický průměr

n_i ... četnost i -té skupiny

Rozptyl a směrodatná odchylka

Obě tyto míry jsou svojí konstrukcí velmi příbuzné, mají univerzální význam a nejčastěji se používají.

Rozptyl je aritmetický průměr ze čtverců odchylek jednotlivých hodnot od aritmetického průměru.

a) v prostém tvaru

$$\text{var } x = \frac{\sum_{i=1}^n |x_i - \bar{x}|^2}{n}$$

n ... rozsah souboru

x_i ... hodnota n -tého prvku

\bar{x} ... aritmetický průměr souboru

b) ve váženém tvaru

$$\text{var } x = \frac{\sum_{i=1}^k |x_i - \bar{x}|^2 \cdot n_i}{\sum_{i=1}^k n_i}$$

k ... počet skupin

x_i ... hodnota znaku x v i -té skupině

\bar{x} ... aritmetický průměr souboru

n_i ... četnost i -té skupiny

Stanovení rozptylu podle výše uvedených výrazů je u rozsáhlejších souborů poměrně pracné, proto vhodné stanovit rozptyl např. podle tzv. výpočtového vzorce:

$$\text{var } x = \overline{x^2} - \bar{x}^2$$

... $\overline{x^2}$ označuje průměr ze čtyř čtverců hodnot znaku x a vypočítá se jako kvadratický průměr umocněný na druhou:

$$\overline{x^2} = \frac{\sum_{i=1}^k x_i^2 \cdot n_i}{\sum_{i=1}^k n_i}$$

... \bar{x}^2 je čtverec aritmetického průměru a určí se podle výrazu

$$\bar{x}^2 = \left(\frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i} \right)^2$$

Směrodatná odchylka je druhá odmocnina rozptylu

$$s = \sqrt{\text{var } x}$$

Směrodatná odchylka tedy odstraňuje neobjektivnost pro srovnávání, kterou vytváří rozptyl tím, že z důvodu zamezení rušení kladných a záporných odchylek od aritmetického průměru pracuje se čtverci odchylek.

Použití směrodatné odchylky je ze všech charakteristik variability nejčastější, zvláště pro možnosti srovnatelnosti kvantitativního vyjádření s pozorovanými hodnotami.

INDEXY

Uplatnění indexů v praxi je velmi rozsáhlé. Existuje velké množství indexů, které lze třídit podle hlediska:

1. zda se srovnávají veličiny stejnorodé nebo různorodé

- indexy individuální srovnávají veličiny stejnorodé (bezprostředně srovnatelné)
- indexy souhrnné srovnávají veličiny různorodé (např. celkový objem produkce různých výrobků)

2. zda se přímo srovnávají veličiny extenzivní nebo intenzivní

- indexy objemové (množství) srovnávají veličiny extenzivní, které charakterizují rozsah, množství, objem určitého ukazatele

- indexy úrovně (cenové) srovnávají veličiny intenzivní, které jsou vlastně podílem dvou veličin extenzivních

3. Velmi často v praxi dochází k srovnávání jevů za několik období, čímž vzniká celá indexová řada. Při konstrukci takovéto řady je možné postupovat dvojím způsobem, buďto vypočítat základní indexy a nebo řetězové indexy.

- v řadě základních indexů jsou všechny indexy vypočítány ke stejnému základnímu období
- v řadě řetězových indexů je základem vždy údaj za období bezprostředně předcházející

4. Stanovují-li se indexy v základním souboru, pak se jedná o indexy vyčerpávající. Jestliže se pracuje s výběrovým souborem, pak se jedná o indexy výběrové.

Laspeyresův index

$$I_L = \frac{\sum_{i=1}^n p_{i(1)} \cdot q_{i(0)}}{\sum_{i=1}^n p_{i(0)} \cdot q_{i(0)}}$$

Paascheův index

$$I_P = \frac{\sum_{i=1}^n p_{i(1)} \cdot q_{i(1)}}{\sum_{i=1}^n p_{i(0)} \cdot q_{i(1)}}$$

$p_{i(1)}$... jednotková cena i-tého výrobku v běžném období

$q_{i(1)}$... spotřebované množství i-tého výrobku v běžném období

$p_{i(0)}$... jednotková cena i-tého výrobku v základním období

$q_{i(0)}$... spotřebované množství i-tého výrobku v základním období

Fisherův index

$$I_F = \sqrt{I_L \cdot I_P}$$