

Matematická statistika

Co je statistika

Statistika je věda o získávání a zpracování informace

Teorie pravděpodobnosti na základě znalosti chování určité náhodné veličiny určuje pravděpodobnost určitého výsledku (náhodného pokusu)

Matematická statistika na základě dat hledá vlastnosti náhodné veličiny

Náhodná veličina

- číselné vyjádření výsledku [náhodného jevu](#) – projevu kvantitativního nebo kvalitativního znaku, který je předmětem zkoumání

2 typy náhodných veličin:

1. **Diskrétní (nespojité) náhodná veličina** – taková, která může nabývat pouze jednotlivých [hodnot](#) (celých čísel) z konečného nebo nekonečného intervalu, tzn. může se měnit jen po skocích.
2. **Spojité náhodná veličina** – taková, která může nabývat všech hodnot z konečného nebo nekonečného intervalu, tzn. může se měnit spojitě bez skoků.

Rozdělení četností náhodné veličiny

- **absolutní četnost** – hodnota, která vyjadřuje, kolik hodnot se v dané skupině vyskytuje.
- **relativní četnost** – hodnota, která vyjadřuje poměr (%) výskytu hodnot v dané skupině k celkovému počtu hodnot ve všech skupinách.

Co dělá statistika?

- Zkoumáme složitý systém, jehož funkci nelze jednoduše pochopit nebo popsat
- Systém se za stejných nebo podobných podmínek může projevovat odlišným způsobem

Přístup k řešení

- Získáme pozorování (data); vyjádříme je jako matematické objekty (čísla, vektory, funkce, . . .)
- Pozorování považujeme za náhodná
- Stanovíme pravděpodobnostní model pro tato pozorování
- V rámci modelu přesně zformulujeme problém, který chceme řešit
- Data a model použijeme k vyřešení daného problému

Druhy úloh

- Odhady parametrů
- Výpočet číselných charakteristik sledovaného systému
- Testování hypotéz
- Ověřování pravdivostní hodnoty výroků o chování systému
- Predikce předpovědi chování systému ve specifických podmínkách
- Optimalizace, hledání optimálních rozhodnutí pro řízení systému

Data a jejich získání

- Data především mohou být úplná
- Častěji máme k dispozici jen jejich podmnožinu, zvanou ve statistice výběr výběrový soubor
- Počet objektů v této podmnožině se označuje n a nazývá rozsah výběru
- Postupy získávání výběru zkoumá teorie výběru, která se zabývá mimo jiné tím, zda je výběr reprezentativní
- Není-li výběr reprezentativní, vzniká systematická chyba
- V takovém případě je k zobecnění potřeba přistupovat s velkou opatrností.

Statistický soubor a statistická jednotka

- Množinu všech předmětů pozorování shromážděných na základě toho, že mají společné vlastnost - **statistický soubor**
- Jednotlivé prvky této množiny se nazývají **prvky (elementy) statistického souboru**
- Počet všech prvků statistického souboru se nazývá **rozsah souboru N** .
- Soubor, který je předmětem zkoumání, se nazývá **základní soubor**
- Často nelze nebo není účelné provést zkoumání všech statistických jednotek tohoto základního souboru. Základní soubor pak zkoumáme pomocí statistických jednotek, které z něj byly urči tím způsobem vybrány a které tvoří tak zvaný **výběrový soubor**

Statistický znak

- Statistický znak je vyjádřením určité vlastnosti statistických prvků sledovaného statistického souboru
- Slouží k charakterizování sledovaného hromadného jevu, vlastnosti daného statistického souboru
- Kvalitativní hodnoty
- Kvantitativní hodnoty
- Statistické třídění – rozčlenění souboru do menších skupin (průnik musí být prázdná množina, musí se roztrždit všechny hodnoty znaku)

Četnost

- Počet prvků, které patří do k -té třídy se nazývá četnost (absolutní četnost) prvků v k -té třídě
- Podíl počet prvků, které patří do k -té třídy/rozsah statistického souboru se nazývá relativní četnost prvků v k -té třídě

Charakteristiky statistického souboru

- Statistickými charakteristikami nazýváme čísla, která podávají stručnou základní informaci o uvažovaném statistickém souboru z různých hledisek
- Jeli předmětem našeho zájmu jediný kvantitativní znak, jde o charakteristiku úrovně (polohy) a charakteristiku variability (proměnnosti, rozptýlení)

Charakteristika úrovně (polohy)

- Čísla, která charakterizují průměrnou hodnotu sledovaného kvantitativního znaku
- Aritmetický průměr, medián, modus, harmonický průměr, geometrický průměr

Charakteristiky úrovně (polohy)

- Aritmetický průměr – součet hodnot dělený jejich počtem
- Medián – prvky seřazené podle velikosti
 - je-li počet prvků liché číslo, prostřední prvek medián
 - Je-li počet prvků sudé číslo, průměr dvou prostředních medián
- Modus – prvek s největší četností

Charakteristiky úrovně (polohy)

- Harmonický průměr

$$\bar{x} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

- Geometrický průměr

$$\bar{x} = \sqrt[n]{x_1 x_2 \dots x_n}$$

Charakteristika variability (proměnnosti, rozptýlení)

- Variační rozpětí – rozdíl mezi největší a nemější hodnotou znaku
- Průměrná odchylka

$$\bar{d} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}$$

- Relativní průměrná odchylka – podíl průměrné odchylky a aritmetického průměru

- Rozptyl
$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

- Směrodatná odchylka – druhá odmocnina z rozptylu
- Variační koeficient – podíl směrodatné odchylky a aritmetického průměru násobeno stem (vyjádřeno v procentech)

Statická závislost

- Dvourozměrný statistický soubor – statistický soubor, u kterého nás zajímá vztah dvou statistických znaků (známka na vysvědčení, známka z pololetní písemky)
- Nejjednodušší závislost – lineární závislost, těsnost se měří korelačním koeficientem, nabývá hodnot od -1 do 1
- 1 přímá závislost, -1 nepřímá závislost, o lineární závislost není pro aproximaci vhodná

Vzorce

- Korelační koeficient
- Kovariance

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

$$s_{xy} = \frac{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}{n}$$

$$r_{xy} = \frac{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum (y_i - \bar{y})^2}}$$

Normální (nebo Gaussovo) rozdělení pravděpodobnosti

