

Základy statistiky

Zpracování pokusných dat

Praktické příklady

Kristina Somerlíková



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



**OP Vzdělávání
pro konkurenceschopnost**

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Data v biologii

- Znak nebo skupina znaků popisuje přírodní jevy, úlohou výzkumníka je vybrat takovou skupinu znaků, které charakterizují šetřený přírodní jev nejlépe. Biometrika v této volbě nemůže většinou pomoci, statistická analýza však může zpětně posoudit, s jakou spolehlivostí podchycují měřené znaky zkoumaný přírodní jev a vybrat z množiny sledovaných znaků skupinu nejvhodnější k popisu jevů.
- Věcný význam dat může být kvantifikován v různé míře a podle toho rozeznáváme různé typy znaků:
 - nominální: jsme schopni věcně interpretovat jen rovnost $x_1 = x_2$ popř. nerovnost $x_1 \neq x_2$ dvou hodnot, protože hodnoty zde představují pouhé smluvené kódy kvalitativních pojmenování.
 - ordinální: kromě rovnosti jsme schopni interpretovat i nerovnost $x_1 > x_2$ popř. $x_1 < x_2$, protože jejich hodnoty vyjadřují nějaké vzestupné nebo sestupné uspořádání intenzity zkoumané vlastnosti.
 - intervalové: kromě rovnosti a uspořádání jsem schopni věcně interpretovat i rozdíl $x_1 - x_2$ dvou hodnot, protože stejný interval mezi jednou dvojicí hodnot a jinou dvojicí hodnot u nich vyjadřuje i stejný rozdíl v intenzitě zkoumané vlastnosti.
 - poměrové: navíc jsme schopni interpretovat i podíl $\frac{x_1}{x_2}$ dvou hodnot, je i stejný podíl v extenzitě zkoumané vlastnosti.

Intervalové a poměrové znaky se společně nazývají kardinálními, protože jsou spojené s existencí měrné jednotky.



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Užití statistických metod je různě adekvátní u různých typů znaků:

Nominální znaky – různé typy třídění, kontingenční tabulky

Ordinální znaky – pořadové statistiky (kvantily, medián, koeficienty pořadové korelace)

Intervalové znaky – většina statistických metod (aritm. průměr, korelační koeficient)

Poměrové znaky – geometrický průměr, ...

Základní pravidlo: Data znaku na vyšším stupni kvantifikace lze zpracovat metodami určenými pro nižší stupeň kvantifikace, avšak za cenu ztráty původní informace. Opak (zpracování dat znaku na nižším stupni kvantifikace metodami určenými pro stupeň vyšší) hrozí zanášením statistikovy libovůle do konečných výsledků.



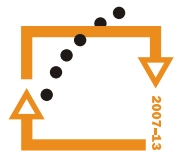
evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Základní pojmy

- Experimentální jednotka – objekt, na kterém provádíme šetření (velikost snůšky jednotlivých plemen slepic – exp. jednotkou je slepice, nikoliv vejce).
- Znáhodnění – náhodný výběr ze studované populace je základem pro jakýkoliv hodnotný statistický test (osoba zcela neznalá podstaty experimentu, náhodný klíč výběru, výběr naslepo).
- Populace – soubor pokusných jednotek, na kterých provádíme potřebná měření.



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



**OP Vzdělávání
pro konkurenceschopnost**

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Přesnost a opakovatelnost

- Přesnost bývá definována pomocí rozdílu měřené či počítané hodnoty od hodnoty přesné.
- Opakovatelnost charakterizuje schopnost pozorovatele obdržet stejné hodnoty opakovaným měřením stejného objektu.

Přesnost lze zlepšit opakovaným měřením stejnou nebo jinou metodou. Opakované měření pomůže vyloučit hrubé chyby ve čtení, zápisu či přenosu dat. Znaky s velkou proměnlivostí stačí měřit méně přesně než znaky méně proměnlivé.

Př. Neseřizené, ale citlivé váhy mohou dávat přesné, ale neopakovatelné výsledky, zatímco necitlivé, ale dobře seřizené, váhy mohou dávat opakovatelné, ale nepřesné výsledky.



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Soukromý zemědělec vlastní stádo mléčného skotu tří různých plemen různého stáří. Jeho hlavním produktem je mléko, vede si denní záznamy o produkci jednotlivých krav.

1. Navrhnete tabulku rozdělení četností z uvedených dat. Dopočítejte relativní četnost a kumulativní četnosti. Grafické zobrazení četností.
2. Naleznete významné hodnoty variační řady. Analýza struktury. Sestrojení Lorenzovy koncentrační křivky.
3. Vypočítejte z uvedených dat charakteristiky obecné úrovně a charakteristiky variability. Pracujte s daty tříděnými i netříděnými.
4. Výpočet regresní úlohy. Výpočet indexu korelace. Grafické znázornění regresní funkce.
5. Výpočet sdružených regresních přímek a korelačního koeficientu. Grafické znázornění přímek.
6. Měření závislosti slovních znaků. Výpočet koeficientů kontingence a asociace.
7. Střední a přípustná chyba výběru, stanovení rozsahu výběrového souboru.
8. Výpočet konfidenčních intervalů pro střední hodnotu, rozptyl a směrodatnou odchylku, jejich grafické zobrazení.
9. Testování homogenity rozptylu, t – testy: testování významnosti rozdílu dvou středních hodnot u nezávislých i závislých souborů.
10. Jednofaktorová a vícefaktorová analýza rozptylu.
11. Metody následného testování.



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



**OP Vzdělávání
pro konkurenceschopnost**

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Bylo náhodně vybráno 62 krav tří plemen různého věku ve druhé fázi laktace.
Hodnoty denního nádoje jsou:

Denní nádoj u náhodně vybraných dojnic (l)				
35	40	29	29	19
38	43	30	31	18
42	41	32	31	
40	41	31	32	
44	40	28	29	
36	40	32	20	
32	41	32	19	
35	39	28	18	
34	44	29	17	
39	42	30	19	
39	43	27	19	
38	41	30	16	
40	38	36	17	
41	37	29	21	
42	41	30	15	



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



**OP Vzdělávání
pro konkurenceschopnost**

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Postup při variačním třídění

1. Hodnoty uspořádat vzestupně
2. Seskupení do intervalů stejné šířky
 - do 100 prvků 6 - 9 intervalů
 - do 500 prvků 10 - 15 intervalů
 - nad 500 prvků 16 – 20 intervalů
3. Šířka intervalu $h=R/k$ $R = X_{max} - X_{min}$ (variační rozpětí)
4. Hranice intervalů musí být vyznačeny jednoznačně, první a poslední interval může být určen pouze jednou hranicí (otevřený interval) – zachycení extrémních hodnot
5. Třídění (čárkovací metodou) **//// //**
6. Střed intervalu - průměr dolní a horní hranice intervalu (nezávislý na hodnotách zařazených jednotek), formálně nahrazuje individuální hodnoty patřící do určité třídy, reprezentuje úroveň třídy
7. Zjistíme absolutní četnost jednotlivých intervalů n_i
8. Vypočteme relativní četnosti p_i , [%],
9. Kumulativní četnosti k_{n_i} , k_{p_i}
10. Graf
 - histogram četností (sloupkový graf)
 - úsečkový (ve středech intervalů)
 - polygon četností (spojnicový - spojuje vrcholy úsečkového grafu)



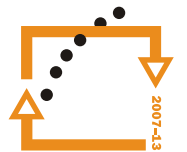
evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Významné hodnoty variační řady

1. uspořádat řadu
2. roztrždit řadu
3. minimální hodnota
4. maximální hodnota (extrémní hodnoty, chyby měření)
5. typická – modální hodnota - hodnota s největší četností (modální třída)
6. kvantily - rozdělují řadu v určitém poměru četností
 - a) dolní kvartil $\times 0,25$
 - b) prostřední kvartil $\times 0,50$ (medián)
 - c) horní kvartil $\times 0,75$
 - kvartily – rozdělují soubor na 4 části (jsou 3)
 - oktily – rozdělují soubor na 8 částí (je jich 7)
 - decily – na 10 částí (9)
 - sedecily – na 16 částí (15)
 - percentily – na 100 částí (99)



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



**OP Vzdělávání
pro konkurenceschopnost**

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Charakteristiky obecné úrovně a variability

- Aritmetický průměr
- Medián
- Modus
- Rozptyl
- Směrodatná odchylka
- Variační koeficient

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

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

$$s^2_x = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s^2_x = \frac{1}{n} \cdot \sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i$$

$$s_x = \sqrt{s^2_x}$$

$$v_x = \frac{s_x}{\bar{x}} \cdot 100$$



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Jednoduchá regrese

- vystihuje průběh závislosti hodnot znaku Y na znaku X
- X ... nezávisle proměnná, vysvětlující proměnná
- Y ... závisle proměnná, vysvětlovaná proměnná

1 – rozměrný znak – přímka

2 – rozměrný znak – rovina

Předpoklad: známe tvar závislosti, který můžeme určit:

1. z bodového diagramu
2. logickým rozborem situace

Regresní funkce:

Předpoklad: Funkce je lineární v parametrech:

$f_0(x) \dots f_m(x) = \text{regresory}$

$b_0 \dots b_m = \text{regresní parametry} - \text{určujeme METODOU NEJMENŠÍCH ČTVERCŮ}$

Regresní funkce je tedy funkcí $m+1$ neznámých parametrů b_0, b_1, \dots, b_m , jejíž hodnoty musíme nalézt tak, aby bylo splněno kritérium nejmenších čtverců: .

Extrém této funkce najdeme tak, že najdeme první parciální derivace postupně podle všech $m+1$ neznámých parametrů, položíme je rovny nule a vzniklou soustavu lineární normálních rovnic řešíme.



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ