Jak udělat Statistiku I na ČZU – Příklady + Teorie

www.kckurzy.cz

Vážení přátelé,
v tomto materiálu najdete soubor příkladů a teoretických otázek, podle kterých vysvětlujeme na našich lekcích příklady a teorii ke zkoušce ze Statistiky I. na ČZU. Výběr pokrývá většinu typů, které se u zkoušky objevují. Přihlásit se můžete na [www.kckurzy.cz.](www.kckurzy.cz)
Přejí příjemné učení☺.

RNDr. Marian Rybář – lektor statistiky

Rada 1: Pokud u zkoušky vůbec nevíte, který typ příkladu máte zrovna před sebou, zkuste přirovnat, kterému z níže uvedených příkladů je daný příklad podobný a většinou to bude on☺.

Rada 2: Nezapomeňte psát na konci příkladu vždy slovní odpověď, aby bylo řešení kompletní.

PŘÍKLADY

1) POPISNÁ STATISTIKA

1. Z následujících hodnot vypočtěte všechny charakteristiky polohy a variability
   10, 8, 9, 7, 7

2. Z následující tabulky četností vypočtěte všechny charakteristiky polohy a variability

Výše platu (v tis. Kč)	Počet pracovníků
10	30
20	10
100	2

3. Z následující tabulky četností vypočtěte všechny charakteristiky polohy a variability

Výše platu (v tis. Kč)	Počet pracovníků
méně než 15	30
15 - 25	10
25-35	5
více než 35	2

4. Máte k dispozici neúplnou tabulku dat o počtu uzavřených sňatků v určitém roce roztříděných podle věku nevěsty (18-24let):

Věk nevěsty	Absolutní četnost	Relativní četnost v %	Kumulativní absolutní četnost	Kumulativní relativní četnost
18	432			
19		4,176		
20			13 405	29,9341
21	10 501			
22	9 116			
23		14,986		
24				
celkem	44 782			

Doplň data do tabulky a odpověz:

a) průměrný věk nevěsty je ……. let
b) nejčastěji se dívky vdávají v ….. letech
c) zhruba 50% nevěst je mladších než …. let
d) variabilita stáří nevěst je ….. %

Výsledky:

1. $\bar{x}$ = 8,2 tis. Kč $\tilde{x}$ = 8 tis. Kč $\hat{x}$ = 7 tis. Kč

s^2 = 1,36 tis. Kč^2 s = 1,1662 tis. Kč V_x = 14,24 % R = 3 tis. Kč

2. $\bar{x}$ = 16,6667 tis. Kč $\tilde{x}$ = 10 tis. Kč $\hat{x}$ = 10 tis. Kč

s^2 = 365,0794 tis. Kč^2 s = 19,107 tis. Kč V_x = 114,64 % R = 90 tis. Kč

3. $\bar{x}$ = 15,5319 tis. Kč $\tilde{x}$ = 10 tis. Kč $\hat{x}$ = 10 tis. Kč

s^2 = 71,5256 tis. Kč^2 s = 8,4573 tis. Kč V_x = 54,4512 % R = 30 tis. Kč

4.

Věk nevěsty	Absolutní četnost	Relativní četnost v %	Kumulativní absolutní četnost	Kumulativní relativní četnost
18	432	0,9647	432	0,9647
19	1870	4,176	2302	5,1407
20	11103	24,7934	13 405	29,9341
21	10 501	23,44915	23906	53,38325
22	9 116	20,3564	33022	73,7396
23	6711	14,986	39733	88,7256
24	5049	11,2764	44782	100,000
celkem	44 782	100,000		

a) 21,4811 b) 20 c) 21 d) V = 6,7238% (S = 1,445)

2) INTERVALY SPOLEHLIVOSTI

1. V balírně bylo zabaleno 100 000 balíčků rýže. Kontrolní vážení 20 balíčků přineslo následující výsledky: 1004, 1003, 1002, 1011, 998, 1000, 997, 998, 1000, 1002, 1003, 1001, 995, 1004, 999, 1000, 1001, 1002, 997, 1000. Vypočtěte bodový a intervalový odhad průměrné hmotnosti balíčku zabalené dodávky s 95% spolehlivostí. Kolik balíčků bychom museli zvážit, chceme-li provést odhad s přesnosti 1 g.

2. Z 500 respondentů bylo 24 voličů ODS. Sestrojte 95% interval spolehlivosti pro procento voličů ODS. Kolik bychom museli mít respondentů, abychom se s 95% spolehlivosti dopustili chyby odhadu nejvýše 1 %.

3. Vypočtěte 99% interval spolehlivosti pro rozptyl a směrodatnou odchylku hmotnosti balíčků rýže pomoci údajů z příkladu č. 1.

4. Při hodnocení variability hmotnosti brojlerů bylo zváženo 60 kuřat ve stáří 8 týdnů. Vypočtený rozptyl byl $s^2 = 0{,}28\ kg^2$. Stanovte 95% interval spolehlivosti pro rozptyl a směrodatnou odchylku hmotnosti kuřat.

5. Z celkového vzorku 2500 konzerv bylo vybráno 30 a z nich bylo 6 vadných. Po výběru už se konzervy nevracely. Stanovte 95% interval spolehlivosti pro podíl vadných konzerv.

Výsledky:

1. a) $\bar{x} = 1000{,}85$ g
   $\Delta = 1{,}6017$
   P $(999{,}248 < \mu < 1002{,}4517) = 95\%$
   Průměrná hmotnost balíčku rýže bude s 95% spolehlivosti v intervalu (999,248; 1002,402).
   b) n = 52 balíčků

2. a) $\Delta = 0{,}0187$
   P $(0{,}0293 < p < 0{,}0667) = 95\%$
   Podíl voličů ODS bude s 95% spolehlivostí v intervalu (2,93%; 6,67%).
   b) n = 1756 respondentů.

3. P $(5{,}768 < \sigma^2 < 32{,}517) = 99\%$
   Rozptyl balíčků rýže bude s 99% spolehlivosti v intervalu (5,7681; 32,517).
   P $(2{,}4017 < \sigma < 5{,}7023) = 99\%$
   Směrodatná odchylka balíčků rýže bude s 99% spolehlivosti v intervalu (2,4017; 5,7023).

4. P $(0{,}2012 < \sigma^2 < 0{,}4165) = 95\%$
   S 95% spolehlivostí rozptyl hmotnosti kuřat bude v intervalu (0,2001; 0,4165).
   P $(0{,}4485 < \sigma < 0{,}6453) = 95\%$
   S 95% spolehlivostí směrodatná odchylka hmotnosti kuřat bude v intervalu (0,4485; 0,6453).

5. $\Delta = 0{,}1423$
   P $(0{,}0577 < p < 0{,}3423) = 95\%$

3) TESTOVÁNÍ HYPOTÉZ

1. Výrobce aut udává, že průměrná spotřeba je 7 l/100 km. Otestujte, zda výrobce udává pravdu, nebo se spotřeba liší od udávaných 7 l. Hladina významnosti je 5 %. Bylo vybráno 12 vozidel a byly zjištěny hodnoty: 7,2; 6,8; 8,1; 6,9; 7,5; 7; 6,5; 7,3; 7,2; 7,4; 7,7; 7,1.

2. Výrobce udává, že cca 2 % jeho výrobků jsou vadná. Z 500 náhodně odebraných výrobků se zjistilo, že 24 je vadných. Liší se procento vadných výrobků od udávaných 2 %?

3. V první pokusné skupině bylo 120 kuřat a zjištěná směrodatná odchylka byla s_{01} = 2,1 g. Ve druhé skupině bylo 80 kuřat a s_{02} = 2,3 g. Otestujte, zda se variabilita dvou souborů liší, nebo je stejná.

4. V následující tabulce jsou uvedené hodnoty IQ náhodně vybraných 12 mužů a žen. Otestujte, zda se statisticky významně liší **rozptyly (variabilita)** IQ mužů a žen.

Muži	113	115	109	100	107	97	118	135	125	121	117	116
Ženy	107	131	128	118	111	114	112	102	100	101	115	119

5. V následující tabulce jsou uvedené hodnoty IQ náhodně vybraných 12 mužů a žen. Otestujte, zda se statisticky významně liší **průměrné IQ (střední hodnota IQ)** mužů a žen.

Muži	113	115	109	100	107	97	118	135	125	121	117	116
Ženy	107	131	128	118	111	114	112	102	100	101	115	119

6. V následující tabulce jsou uvedené hodnoty výkonnosti dělníků při různých teplotách. Otestujte, zda výkonnost dělníků při teplotě 19°C a 23°C je stejná nebo se liší.

Dělník č.	19°C	23°C
1	52	50
2	58	51
3	59	60
4	65	61
5	71	58
6	55	52
7	59	60
8	63	59
9	67	61
10	68	57
11	70	64
12	63	64
13	51	50
14	56	51
15	62	53

7. Z 250 výrobků vyrobených starým způsobem bylo 18 vadných. Z 200 výrobků vyrobených novým způsobem bylo 86 vadných. Posuďte na 5% hladině významnosti, zda se podíl vadných výrobků u staré a nové technologie liší, nebo je stejný.

Výsledky:

1. Nezamítám H_0 (t = 1,836)
2. Zamítám H_0 (u = 4,4721)
3. Nezamítám H_0 (F = 1,204)
4. Nezamítám H_0 (F = 1,10026)
5. F-test: nezamítám H_0; t-test: nezamítám H_0 (t = 0,3002), $\overline{x1}$ = 114,4167, $\overline{x2}$ = 113,1667, S_1 = 10,4399, S_2 = 9,9529
6. Zamítám H_0 , t = 4,0921, $\bar{d}$ = 4,5333, S_d = 4,2906
7. Zamítám H_0 (u = -8,9521)

4) ANOVA

1. V následující tabulce jsou uvedené počty napsaných slov na klávesnicích A, B, C. Otestujte, zda se průměrné počty napsaných slov na těchto klávesnicích statisticky významně liší, nebo jsou stejné. Pokud se průměrné počty napsaných slov významně liší, zjistěte, které z dvojic jsou odlišné.

A	77	71	74	67
B	67	62	63	57
C	63	59	59	54

Výsledky:

1. Zamítám H_0 (F = 12,0843)

5) REGRESE A KORELACE

1. Zkouška z určitého předmětu má písemnou a ústní část. Maximální počet bodů v písemné části je 30, v ústní části je 20. U 15 posluchačů byly zjištěny následující výsledky:

Počet bodů v písemné části	Počet bodů v ústní části
25	15
26	18
28	20
14	10
5	2
8	8
19	13
29	15
30	17
15	11
18	12
7	6
9	2
15	10
3	5

a) Vypočtěte rovnice sdružených regresních přímek.
b) Jaký počet bodů u ústní části zkoušky můžeme očekávat u studenta, který v písemné části zkoušky dosáhl 24 bodů?
c) Jak velké změny v ústní části (v průměru) vyvolá snížení počtu bodů u písemné části zkoušky o 3 body?
d) V ústní části zkoušky dosáhl student 10 bodů. Kolik bodů můžeme očekávat z písemné části zkoušky?
e) Interpretujte regresní parametry a_{yx} a b_{yx}
f) Určete sílu závislosti a otestujte statistickou významnost korelačního koeficientu r.
g) Z kolika procent jsou změny v ústním hodnocení vyvolány změnami v písemné části zkoušky?
h) Popište závislost počtu bodů v ústní části na počtu bodů v písemné části pomoci kvadratické funkce (pomocí regresní paraboly)

2. Při prodeji ojetých automobilů stejné značky byla sledována závislost ceny automobilu (tis. Kč) na počtu ujetých km v tisících. Máme k dispozici následující výsledky:
   $\bar{x} = 54{,}78$ $\bar{y} = 35{,}69$ $s_x = 28{,}06$ $s_y = 20{,}74$ $b_{yx} = -0{,}356$
   a) Určete lineární regresní rovnici popisující daný vztah.
   b) Vyjádřete sílu této závislosti.
   c) Jakou cenu můžeme očekávat u automobilu této značky, který má najeto 90 tisíc km?
   d) K jaké změně ceny povede zvýšení počtu najetých km o 10 tisíc?

3. Vyjádřete hodnotu koeficientu korelace, znáte-li:
   $b_{yx} = -0{,}745$ $b_{xy} = -0{,}826$

4. Pro závislost hodnoty produkce na výši investic u souboru 10 vybraných strojírenských podniků vyjádřete hodnotu Spearmanova koeficientu pořadové korelace

Produkce (mil. Kč)	Investice (tis. Kč)
6,28	142
5,86	138
6,42	165
5,00	112
6,48	152
6,39	148
6,31	142
6,20	124
6,51	172
6,52	169

Výsledky:

1. a) y = 1,4614 + 0,5661.x x = -0,0504 + 1,5351.y
   - b) y = 15,0478
   - c) Δy = -1,6983
   - d) x = 15,3006
   - e) ayx = průsečík s osou x, je to hodnota y pro x = 0, byx = směrnice přímky, říká o kolik jednotek se průměrně zvýší hodnota y, pokud se hodnota x zvýší o jednu jednotku
   - f) r_{xy} = 0,9322; silná přímá závislost, t = 9,2711, zamítáme H0
   - g) R^2 = 86,8997 %
   - h) Y = 0,5887 + 0,707.x – 0,004138.x^2

2. a) y = 55,19 – 0,356x
   b) r_{yx} = - 0,481; střední nepřímá závislost
   c) y = 23,15
   d) Δy = -3,56

3. $r_{yx} = -0{,}78$
4. $r_S = 0{,}96$

y_i	p_i	x_i	q_i	$d_i = p_i - q_i$	d_i^2
6,28	7	142	6,5	0,5	0,25
5,86	9	138	8	1	1
6,42	4	165	3	1	1
5,00	10	112	10	0	0
6,48	3	152	4	-1	1
6,39	5	148	5	0	0
6,31	6	142	6,5	-0,5	0,25
6,20	8	124	9	-1	1
6,51	2	172	1	1	1
6,52	1	169	2	-1	1
Σ	**x**	**x**	**x**	**x**	**6,5**

6) NEPARAMETRICKÉ TESTY

1. Ověřte, zda teplotní rozdíly mezi dvěma stanovišti A a B, které jsou od sebe vzdáleny 200 m, jsou statisticky významné. Měření bylo prováděno každé dvě hodiny po dobu jednoho dne stejným typem teploměru. K výpočtu použijte Wilcoxonův – Whiteův test a dvouvýběrový Wilcoxonův test.

Stanoviště	Teplota v °C											
A	4,3	4,2	4,2	5,0	5,3	8,7	7,6	6,7	5,1	4,9	4,5	4,3
B	4,4	4,0	4,1	5,2	5,4	8,8	7,4	6,8	5,0	4,7	4,7	4,5

2. U 12 studentů bylo zjišťováno, zda výsledky zápočtového testu a zkouškového testu z předmětu matematická statistika spolu souvisí. Zjistěte na základě údajů uvedených v následující tabulce, zda existuje na 5% hladině významnosti průkazný rozdíl mezi výsledky zápočtového a zkouškového testu. Z výpočtu použijte Wilcoxonův test a znaménkový test.

Test	Počet bodů											
zápočtový	24	13	22	19	17	12	8	25	18	21	21	15
zkouškový	21	19	24	22	14	23	17	22	14	18	21	20

3. Při zhodnocení cenové úrovně jablek ve Středočeském kraji byly ve třech typech obchod§ sledovány ceny konzumních jablek (v Kč/kg) v jednotlivých čtvrtletích určitého roku. Posuďte na hladině významnosti α = 0,05 pomoci vhodného neparametrického testu, zda existuje statisticky významný rozdíl v průměrné cenové hladině konzumních jablek mezi jednotlivými typy obchodů.

Typ obchodů	Průměrné ceny jablek v jednotlivých čtvrtletích roku (Kč/kg)			
	I. čtvrtletí	II. čtvrtletí	III. čtvrtletí	IV. čtvrtletí
Hypermarket	29,43	29,58	26,75	28,83
Supermarket	31,14	32,19	27,56	29,23
Maloprodejny	32,03	33,98	28,78	32,02

Výsledky:

1. Wilcoxonův – Whiteův test: nezamítám H_0 ; dvouvýběrový Wilcoxonův test: nezamítám H_0.
2. Wilcoxonův test: nezamítám H_0 ; znaménkový test: nezamítám H_0.
3. Nezamítám $H_{0.}$

TEORIE

Níže jsou uvedeny některé typové ukázky teoretických otázek pro lepší pochopení, jak se při hledání správné odpovědi rozhodovat.

POPISNÁ STATISTIKA

Modus řadíme mezi:

*a) míry polohy

b) míry variability

c) míry šikmosti

d) míry špičatosti

e) ani jedna odpověď není správná

Průměrná absolutní odchylka patří mezi:

a) míry polohy

*b) míry variability

c) míry šikmosti

d) míry špičatosti

e) ani jedna odpověď není správná

Co patří mezi relativní míry variability:

a) směrodatná odchylka

*b) variační koeficient

c) rozptyl

d) průměrná absolutní odchylka

e) ani jedna odpověď není správná

Jakou informaci nám poskytují charakteristiky variability?

měří rozptýlení hodnot příslušného souboru, tzn. určují přibližné rozmezí, v němž se jednotky statistického souboru vyskytují

Medián je definován jako:

a) rozptýlenost hodnot

*b) prostřední hodnota řady pozorování

c) robustní charakteristika variability

d) nejčetnější hodnota znaku

e) ani jedna odpověď není správná

Uveďte situace (minimálně 2), kdy může medián popsat polohu statistického souboru lépe než průměr.

a) pokud se v souboru vyskytuje nějaká extrémně vysoká či nízká hodnota

e) pokud je soubor sešikmený hodně doprava či doleva **(ukázka viz doučování)**

Popsat podrobně co znamená směrodatná odchylka

Směrodatná odchylka popisuje míru rozptýlení hodnot souboru. Je to odmocnina z rozptylu. Graficky je možno směrodatnou odchylku vyjádřit pomocí pravidla tří sigma **(obrázek viz doučování)**

Pro která pravděpodobnostní rozdělení je jejich střední hodnota rovna mediánu a zároveň modu? Vysvětlete a uveďte příklady.

Pro symetrická (např. Normální rozdělení)

Co se stane s průměrem, rozptylem, směrodatnou odchylkou, mediánem a rozpětím statistického souboru, jestliže každá hodnota statistického souboru se:

a) zvětší o čtyři – průměr a medián se zvětší o čtyři; rozptyl; směrodatná odchylka a rozpětí statistického souboru se nezmění

b) zvětší dvakrát - průměr a medián se zdvojnásobí, rozptyl se zvýší čtyřikrát; směrodatná odchylka a rozpětí statistického souboru se zvýší dvakrát

Jaké znáte druhy grafů

koláčové, sloupcové(polygon, histogram), krabicový graf (Box plot),bodové, spojnicové **(obrázky a vysvětlení viz doučování)**

INTERVALY SPOLEHLIVOSTI (BODOVÝ A INTERVALOVÝ ODHAD)

Lze zvětšením rozsahu výběru ovlivnit šíři intervalu spolehlivosti pro odhad střední hodnoty ZS?

a) na šířku intervalu to nemá vliv

*b) ano, interval spolehlivosti bude užší a přesnější

c) bude nutné zvolit jiný typ rozdělení

d) ano, interval spolehlivosti bude širší

e) ani jedna odpověď není správná

Velikost přípustné chyby odhadu při odhadu průměru základního souboru nezávisí na:

a) rozptylu

b) počtu pozorování

*c) průměru

d) přesnosti odhadu

e) ani jedna odpověď není správná

Při konstrukci intervalových odhadů platí, že čím je větší spolehlivost 1-α odhadu, tím je přesnost odhadu:

a) větší

*b) menší

c) stejná

d) nelze rozhodnout

e) ani jedna odpověď správně

Mezi základní vlastnosti bodového odhadu řadíme:

nestrannost odhadu – použitý odhad nenadhodnocuje ani nepodhodnocuje opravdovou hodnotu charakteristiky ZS, kterou se snažíme odhadnout
konzistentnost odhadu - s rostoucím rozsahem výběru se bude daný odhad stále více blížit charakteristice ZS, kterou odhadujeme
vydatnost odhadu - nejvydatnější (nejpřesnější) je ten odhad, která má nejmenší rozptyl
postačujícnost odhadu – odhad je postačující, pokud shrnuje všechny informace, které poskytuje daný VS

Bodový odhad je konzistentní, pokud:

a) nevede k systematickým chybám

b) má ze všech odhadů nejmenší rozptyl

*c) s rostoucím rozsahem výběru klesá pravděpodobnost chyby

d) s rostoucím rozsahem výběru roste pravděpodobnost chyby

e) ani jedna odpověď není správná

Mezi základní vlastnosti bodového odhadu řadíme:

a) podhodnocování

*b) nestrannost

c) velký rozptyl

d) nadhodnocování

e) ani jedna odpověď není správná

Bodový odhad je nestranný, pokud:

*a) nevede k systematickým chybám

b) má ze všech odhadů nejmenší rozptyl

c) má ze všech odhadů největší rozptyl

d) s rostoucím rozsahem výběru roste pravděpodobnost chyby

e) ani jedna odpověď není správná

Bodový odhad je postačující, pokud:

a) nevede k systematickým chybám

b) má ze všech odhadů nejmenší rozptyl

c) s rostoucím rozsahem výběru klesá pravděpodobnost chyby

d) s rostoucím rozsahem výběru roste pravděpodobnost chyby

*e) ani jedna odpověď není správná (postačující – obsahuje všechny informace o charakteristice ZS)

Jak lze definovat pojem "přesnost odhadu"?

Je to maximální chyba, které se při odhadu s danou spolehlivostí 1-alfa dopustíme

TESTOVÁNÍ HYPOTÉZ

Welchův test se používá v případě:

a) dvou závislých souborů s rozdílnými rozptyly

b) dvou nezávislých souborů se shodnými rozptyly

c) dvou závislých souborů se shodnými rozptyly

*d) dvou nezávislých souborů s rozdílnými rozptyly

e) ani jedna odpověď není správná

Popsat postup statistického testování hypotéz:

- sestavení hypotéz (H0, H1)
- výpočet příslušného testového kritéria
- porovnání testového kritéria s tabulkovou kritickou hodnotou na příslušné hladině významnosti alfa a zjištění zda testové kritérium padne do kritického oboru (zamítáme H0) či oboru přijetí (nezamítáme H0)
- slovní odpověď

Podle jakého kritéria se rozdělují statistické testy na jednovýběrové a dvouvýběrové?

podle počtu výběrových souborů

Co je to testové kritérium?

vzorec, kterým se testuje správnost nulové hypotézy **(podrobně viz doučování)**

V testování statistických hypotéz se hovoří o chybě 1. druhu. Jak je tato chyba definována?

Je to pravděpodobnost zamítnutí H0, i když je správná. Někdy nazýváme hladina významnosti alfa. Nejčastěji nabývá hodnoty α =0,05 (někdy α =0,01 nebo α =0,10). **(vysvětlení viz doučování)**

Chyba 2. druhu při testování je:

*a) nezamítnutí H0, i když není správná **(vysvětlení viz doučování)**

b) vyjádření síly testu

c) zamítnutí H0, i když je správná

d) číselné vyjádření hladiny významnosti

e) ani jedna odpověď není správná

Definujte pojem kritický obor

Interval testových kritérií svědčících ve prospěch **alternativní** hypotézy. Podmnožina prostoru všech možných výsledků testových kritérií .

Definujte pojem obor přijetí

Interval testových kritérií svědčících ve prospěch **nulové** hypotézy. Podmnožina prostoru všech možných výsledků testových kritérií .

Pokud chceme ověřit významnost shody dvou průměrů a víme, že soubory jsou závislé, jaký test použijeme?

a) dvouvýběrový t-test se stejnými rozptyly

b) dvouvýběrový t-test s různými rozptyly (Welchův test)

*c) párový test

d) F-test pro shodu rozptylů

e) ani jedna odpověď není správná

Tvrzení, že neexistuje rozdíl mezi dvěma nebo více charakteristikami, se nazývá:

*a) nulová hypotéza

b) alternativní hypotéza

c) oboustranná hypotéza

c) jednostranná hypotéza

e) ani jedna odpověď není správná

Jaký závěr učiníme, jestliže při testování hodnota testového kritéria padne do kritického oboru?

a) nezamítneme nulovou hypotézu

*b) zamítneme nulovou hypotézu

c) nelze určit

c) musíme použít jiný test

e) ani jedna odpověď není správná

Jaký závěr učiníme, jestliže při testování hodnota testového kritéria nepadne do oboru přijetí?

a) nezamítneme nulovou hypotézu

*b) zamítneme nulovou hypotézu

c) nelze určit

c) musíme použít jiný test

e) ani jedna odpověď není správná

Jaké testové kritérium se používá pro testování shody dvou relativních četností:

a) F kritérium

b) t kritérium

*c) u kritérium

d) chí-kvadrát kritérium

e) ani jedna odpověď není správná

ANOVA

Pokud chceme ověřit významnost shody tří a více průměrů, jaký test použijeme?

a) dvouvýběrový t-test se stejnými rozptyly

b) dvouvýběrový t-test s různými rozptyly (Welchův test)

*c) ANOVA test

d) F-test pro shodu rozptylů

e) ani jedna odpověď není správná

Analýza rozptylu (metoda ANOVA) se používá pro:

a) testování shody tří a více rozptylů

b) testování shody dvou rozptylů

*c) testování shody tří a více průměrů

d) testování shody dvou průměrů

e) ani jedna odpověď není správná

REGRESE A KORELACE

Korelační analýza se zabývá:

a) předpoklady o datech

b) jednostrannými závislostmi

c) systematickými vlastnostmi reziduí

*d) vzájemnými závislostmi s důrazem na intenzitu vztahu

e) ani jedna odpověď není správná

Jak lze interpretovat regresní koeficient a jakých hodnot nabývá?

udává, o kolik jednotek se v průměru změní závisle proměnná, změní-li se nezávisle proměnná o jednotku. Je to směrnice příslušné regresní přímky. Nabývá hodnot (-∞, ∞).

Dvě proměnné jsou korelovány tehdy, jestliže:

a) se jedná o nezávislé proměnné

b) určité hodnoty jedné proměnné se nevyskytují společně s určitými hodnotami druhé proměnné

c) určité hodnoty jedné proměnné nemají tendenci se vyskytovat společně s určitými hodnotami druhé proměnné

*d) určité hodnoty jedné proměnné mají tendenci se vyskytovat společně s určitými hodnotami druhé proměnné

e) ani jedna odpověď není správná

Kdy hovoříme o sdružených regresních přímkách?

pokud vedle závislosti Y na X zkoumáme zároveň závislost X na Y, hledáme druhou regresní přímku

a souhrnně tak hovoříme o sdružených regresních přímkách

Jaký cíl sleduje regresní analýza?

nalezení vhodného modelu, určení průběhu a tvaru studované závislosti

Jakými mírami měříme sílu závislosti mezi kvantitativními znaky a jakých hodnot nabývají?

Pearsonův korelační koeficient, Spearmanův pořaďový korelační koeficient; <-1;1>

Pokud jsou statistické znaky nekorelovány, pak korelační koeficient nabude hodnoty či hodnot:

a) r náleží <0;1>

*b) r = 0

c) r = 1

d) r náleží <-1;1>

e) ani jedna odpověď není správná

Jaké znáte metody pro získání odhadů parametrů regresních funkcí lineárních v parametrech?

a) Metoda nejmenších čtverců (MNČ) – **(obrázek a vysvětlení viz doučování)** - samotný výpočet probíhá následně pomocí: soustavy normálních rovnic, výpočetními vzorci, pomocí kalkulačky, pomocí software např. Excel
b) Metoda maximální věrohodnosti

Jaké odhady lze provést na základě rovnice regresní přímky?

interpolace, extrapolace **(vysvětlení viz doučování)**

STATISTICKÉ ZJIŠŤOVÁNÍ

Definujte pojem statistika

Věda o sběru a zpracování **hromadných dat**, nezabývá se tedy např. jedincem, podnikem apod, ale počítá s celým souborem prvků (např. skupina jedinců, skupina podniků)

Rozlišujeme dvě úrovně statistiky:

a) **popisná statistika** – zabývá se elementárními (jednoduchými) charakteristikami souboru např. průměry, mediány, směrodatné odchylky, grafy, tabulky

matematická (induktivní) statistika – moderní, pokročilejší – zabývá se složitějšími charakteristikami souboru

a) teorie odhadu (bodové a intervalové odhady)
b) testování statistických hypotéz

Definujte pojem statistický soubor

Konečná neprázdná množina prvků (statistických jednotek), které mají z daného hlediska podobné vlastnosti.

Rozlišujeme dva typy statistických souborů:

a) **základní** (populační) – obsahuje všechny jednotky dané populace (např. všichni muži na zemi) – tzv. úplné zjišťování
b) **výběrový** – pouze část základního souboru, který je určitým způsobem vybrán ze základního souboru (např. 10 vybraných mužů) – tzv. neúplné zjišťování

Pracujeme většinou s výběrovými soubory, práce je jednodušší a méně finančně náročná => zevšeobecnění úvah z výběrového souboru na celý soubor.

Definujte pojem statistická jednotka

Jednotlivý prvek statistického souboru (např. jeden muž)

Co jsou to statistické znaky a jak je dělíme

Vlastnosti statistických jednotek, které na nich měříme (např. výška daného muže).

Mohou být povahy:

a) **kvantitativní** – vyjádřené čísly
    1. *spojité* – znaky nabývají všech reálných hodnot (např. výška muže)
    2. *nespojité* (diskrétní) – znaky nabývají pouze celočíselných hodnot (např. počet dětí daného muže)
b) **kvalitativní** – vyjádřené slovně
    1. *alternativní* – pouze 2 možnosti (např. pohlaví – muž, žena)
    2. *množné* - více možností (např. vzdělání – zš, sš, vš)

Definujte pojem rozsah statistického souboru

Počet statistických jednotek daného souboru, značí se **n.**

Jaké znáte základní etapy statistických prací

statistické zjišťování (šetření, sběr) - získávání neznámých informací o znacích statistických jednotek , výsledkem statistického zjišťování jsou neuspořádané údaje

statistické zpracování (statistická analýza) - souhrn postupů, které by měly zpřehlednit a uspořádat neuspořádané údaje získané jako výsledek statistického zjišťování a analyzovat, co z daných získaných údajů plyne

Definujte pojem náhodný výběr

Výběr, který spočívá v tom, že o tom, která jednotka bude zařazena do výběrového souboru rozhoduje náhoda. Je lepší než záměrný výběr.

Definujte pojem záměrný výběr

Výběr, který spočívá v tom, že o tom, která jednotka bude zařazena do výběrového souboru nerozhoduje náhoda, ale sběratel dat. Na základě tohoto výběru nelze pořizovat objektivně přesné odhady.

Jaké znáte techniky náhodného výběru?

a) **losování** - nejjednodušší technika (z osudí, klobouku), lístky („opora výběru") musí být všechny naprosto stejné
b) **mechanický (systematický) výběr** - vybíráme každou n-tou jednotku z náhodně uspořádané posloupnosti. Např. z abecedního seznamu vybereme každého 10. studenta.
c) tabulka náhodných čísel, generátor náhodných čísel
d) **oblastní výběr** - nejprve náhodně vybereme oblasti a potom budeme náhodně vybírat jednotky ve vybrané oblasti
e) **výběr vícestupňový** - vybíráme náhodně na více stupních. Např. škola: 1. stupeň – zda je v Praze, 2. stupeň – konkrétní škola, 3. stupeň – fakulta.

Definujte pojem metoda základního masivu

Metoda, která se používá, pokud se soubor skládá z několika velkých (rozhodujících) jednotek a z mnoha malých (nerozhodujících či doplňkových) jednotek. Stačí vybrat pouze rozhodující jednotky (např. významná krajská města, zanedbáme menší města).

Definujte pojem reprezentativnost výběru

výběr je reprezentativní, když má každý prvek základního souboru stejnou pravděpodobnost, že bude vybrán (např. použitím náhodného výběru).

Pokud při zahrnování jednotek do výběrového souboru spolurozhodují různá logická hlediska a názory jedná se o:

a) anketu

b) metodu základního masivu

*c) záměrný výběr

d) mechanický výběr

e) ani jedna odpověď není správná

NEPARAMETRICKÉ TESTY

Jaký je rozdíl mezi parametrickými a neparametrickými testy

- Parametrické – používají se v případě, že data mají normální rozdělení a neobsahují odlehlé hodnoty, počítají přímo se samotnými hodnotami souboru, testy sloužící k ověřování hypotéz, mají větší sílu než neparametrické testy. Např. Dvouvýběrový t-test, Párový t-test, ANOVA

- Neparametrické – používají se v případě, že data nemají normální rozdělení a obsahují odlehlé hodnoty, nepočítají přímo se samotnými hodnotami souboru, ale pouze s jejich pořadími, mají menší sílu než parametrické testy. Např.Wilcoxonův dvouvýběrový test, Wilcoxonův párový test, Kruskal-Wallisův test

DOPORUČENÉ PŘÍKLADY K PROCVIČENÍ ZE SKRIPT „CVIČENÍ ZE STATISTIKY 1“

6.12

6.13 a,b

8.3

8.15

9.4

9.5

9.23

10.17

10.18

10.28

11.4