Řešené příklady na zkoušku ze Statistiky 2

Příklad 1.: Je známo, že roční přírůstek platu manažerů středně velkých firem má normální rozložení se střední hodnotou 12,2% a směrodatnou odchylkou 3,6%. Jaká je pravděpodobnost, že v náhodném výběru rozsahu 9 bude průměrný přírůstek platu nejvýše 10%? Odpověď napište celou větou.

Řešení: $X \sim N(12{,}2;\, 3{,}6^2)$, $n = 9$

$$P(M \le 10) = P\left(\frac{M - 12{,}2}{\frac{3{,}6}{3}} \le \frac{10 - 12{,}2}{\frac{3{,}6}{3}}\right) = P(U \le -1{,}83) = 1 - \Phi(1{,}83) = 1 - 0{,}9664 = 0{,}0336\ .$$

Pravděpodobnost, že v náhodném výběru rozsahu 9 bude průměrný přírůstek platu nejvýše 10%, je 0,0336.

Příklad 2.: Je dána neúplná tabulka analýzy rozptylu jednoduchého třídění. Místo otazníků doplňte chybějící čísla a na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot.

zdroj variability	součet čtverců	stupně volnosti	podíl	F_A
skupiny	?	2	?	?
reziduální	172	?	?	-
celkový	326	17	-	-

Řešení:

zdroj variability	součet čtverců	stupně volnosti	podíl	F_A
skupiny	154	2	77	6,715
reziduální	172	15	11,467	-
celkový	326	17	-	-

Protože testová statistika je větší než kvantil $F_{0,95}(2,15) = 3{,}6823$, nulovou hypotézu zamítáme na hladině významnosti 0,05.

Příklad 3.: U osmi náhodně vybraných firem poskytujících konzultace v oblasti jakosti výroby byly v roce 1993 zjištěny počty zaměstnanců (náhodná veličina X) a roční obraty (náhodná veličina Y, v miliónech Kč), jak je uvedeno v tabulce:

Číslo firmy	1	2	3	4	5	6	7	8
X	3	5	5	8	9	11	12	15
Y	0,8	1,2	1,5	1,9	1,8	2,4	2,5	3,1

Předpokládáme, že závislost ročního obratu na počtu zaměstnanců lze popsat regresní přímkou. K dispozici jsou částečné výstupy regresní analýzy ze systému STATISTICA:

N=8	Beta	Sm.chyba beta	B	Sm.chyba B
Abs.člen			0,361207	0,121417
X	0,984798	0,070914	0,181034	0,013036

Efekt	Součet čtverců	sv	Průměr čtverců	F	Úroveň p
Regres.	3,801724	1	3,801724	192,8571	0,000009
Rezid.	0,118276	6	0,019713		
Celk.	3,920000				

a) Napište rovnici regresní přímky vyjadřující závislost Y na X. Interpretujte úsek a směrnici regresní přímky.
b) Vypočtěte index determinace a interpretujte ho.

Řešení:

ad a) y = 0,361207 + 0,181034x

Pokud firma nebude mít žádné zaměstnance (tzn., že pracují pouze majitelé), bude roční obrat asi 361 000 Kč.

Pokud se zvýší počet zaměstnanců o jednoho, vzroste roční obrat asi o 181 000 Kč.

ad c) $ID^2 = \frac{S_R}{S_T} = \frac{3{,}801724}{3{,}92} = 0{,}9698$

Znamená to, že variabilita ročního obratu je z téměř 97% vysvětlena regresní přímkou.

Příklad 4.: Časová řada obsahuje údaje o počtu zaměstnanců určité akciové společnosti v letech 1989 – 1996 vždy k 31.12.

1989	1990	1991	1992	1993	1994	1995	1996
622	627	631	635	641	641	632	625

Vypočtěte chronologický průměr této časové řady.

Řešení:

$$\bar{y} = \frac{1}{n-1}\left(\frac{y_1}{2} + \sum_{i=2}^{n-1} y_i + \frac{y_n}{2}\right) = \frac{1}{7}\left(\frac{622}{2} + 627 + 631 + 635 + 641 + 641 + 632 + \frac{625}{2}\right) = 632{,}9$$

Příklad 5.: Je dáno pět nezávislých náhodných výběrů o rozsazích 5, 7, 6, 8, 5, přičemž i-tý výběr pochází z rozložení $N(\mu_i,\sigma^2)$, i = 1, ..., 5. Byl vypočten celkový součet čtverců $S_T = 15$ a reziduální součet čtverců $S_E = 3$. Na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot.

Řešení: n = 5 + 7 + 6 + 8 + 5 = 31, r = 5, $S_A = S_T - S_E = 15 - 3 = 12$

$$F = \frac{S_A/(r-1)}{S_E/(n-r)} = \frac{12/4}{3/26} = 26,\ F_{0,95}(4,26) = 2{,}7426$$

Protože $F \geq F_{0,95}(4,26)$, H_0 zamítáme na hladině významnosti 0,05.

Příklad 6.: Je známo, že týdenní výdaje domácností na určité potravinářské zboží se řídí normálním rozložením se střední hodnotou 90 Kč a směrodatnou odchylkou 14 Kč. Jaká je pravděpodobnost překročení hranice 100 Kč pro průměrné výdaje pěti náhodně vybraných domácností?

Řešení: $X \sim N(90; 14^2)$, n = 5

$$P(M > 100) = 1 - P\left(\frac{M-90}{\frac{14}{\sqrt{5}}} \leq \frac{100-90}{\frac{14}{\sqrt{5}}}\right) = 1 - P(U \leq 1{,}5972) = 1 - \Phi(1{,}6) = 1 - 0{,}9452 = 0{,}0548$$

Hledaná pravděpodobnost je tedy 5,48%.

Příklad 7.: V dílně pracuje 15 dělníků, u nichž byl zjištěn počet směn odpracovaných za měsíc (proměnná X) a počet zhotovených výrobků (proměnná Y).

X: 20 21 18 17 20 18 19 21 20 14 16 19 21 15 15
Y: 92 93 83 80 91 85 82 98 90 60 73 86 96 64 81

Předpokládáme, že závislost počtu zhotovených výrobků na počtu směn lze popsat regresní přímkou. K dispozici jsou částečné výstupy regresní analýzy ze systému STATISTICA:

	Výsledky regrese se závislou proměnnou : Y (Smeny.sta) R= ,92718009 R2= ,85966293 Upravené R2= ,84886777 F(1,13)=79,634 p<,00000 Směrod. chyba odhadu : 4,2834					
N=15	Beta	Sm.chyba beta	B	Sm.chyba B	t(13)	Úroveň p
Abs.člen			5,010135	8,875949	0,564462	0,582049
X	0,927180	0,103900	4,302365	0,482123	8,923795	0,000001

a) Napište rovnici regresní přímky a interpretujte její směrnici.
b) Na hladině významnosti 0,01 testujte hypotézu o nevýznamnosti úseku regresní přímky a nevýznamnosti směrnice regresní přímky.
c) Najděte regresní odhad počtu zhotovených výrobků pro 16 směn.
d) Z kolika procent je variabilita počtu zhotovených výrobků vysvětlena regresní přímkou?
Řešení:
ad a) Regresní přímka má rovnici y = 5,01 + 4,3x.
Když se počet odpracovaných směn zvýší o 1, počet vyrobených výrobků se v průměru zvýší o 4,3.
ad b) Na hladině významnosti 0,01 nezamítáme hypotézu, že regresní přímka prochází počátkem a zamítáme hypotézu, že počet výrobků nezávisí na počtu odpracovaných směn.
ad c) Predikovaná hodnota počtu výrobků pro 16 směn je 73,85.
ad d) Regresní přímka závislosti počtu výrobků na počtu odpracovaných směn vystihuje variabilitu počtu výrobků téměř z 86%.

Příklad 8.: Medián počtu výrobků vyrobených za směnu starým strojem je 80. Do provozu byl uveden nový stroj a ve 12 náhodně vybraných dnech byly sledovány počty výrobků za směnu: 75 85 92 80 94 90 91 76 88 82 96 83. Pomocí Wilcoxonova testu testujte na hladině významnosti 0,05 hypotézu, že výkon starého a nového stroje se neliší.
Řešení: Na hladině významnosti 0,05 testujeme H_0: $x_{0,50} = 80$ proti H_1: $x_{0,50} \neq 80$.

$\lvert Y_i \rvert$	5	**5**	**12**	0	**14**	**10**	**11**	4	**8**	**2**	**16**	**3**
R_i	4,5	4,5	9		10	7	8	3	6	1	11	2

$S_W^+ = 4,5 + 9 + 10 + 7 + 8 + 6 + 1 + 11 + 2 = 58,5$, $S_W^- = 4,5 + 3 = 7,5$, testová statistika = $\min\{58,5; 7,5\} = 7,5$, kritický obor: $W = \langle 0,10 \rangle$. Protože testová statistika se realizuje v kritickém oboru, H_0 zamítáme na hladině významnosti 0,05.

Příklad 9.: Měřením délky deseti válečků byly získány hodnoty (v mm): 5,38 5,36 5,35 5,40 5,41 5,34 5,29 5,43 5,42 5,32. Pro úsporu času máte uveden aritmetický průměr m =

5,37 mm a směrodatnou odchylku s = 0,044 mm. Těchto deset hodnot považujeme za realizace náhodného výběru rozsahu 10 z normálního rozložení $N(\mu, \sigma^2)$.

a) Sestrojte 99% interval spolehlivosti pro neznámou střední hodnotu μ.

b) Sestrojte 99% interval spolehlivosti pro neznámou směrodatnou odchylku σ.

Řešení:

ad a)

$$d = m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 5,37 - \frac{0,044}{\sqrt{10}} t_{0,995}(9) = 5,37 - \frac{0,044}{\sqrt{10}} 3,25 = 5,3248$$

$$h = m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 5,37 + \frac{0,044}{\sqrt{10}} t_{0,995}(9) = 5,37 + \frac{0,044}{\sqrt{10}} 3,25 = 5,4152$$

5,3248 mm < μ < 5,4152 mm s pravděpodobností aspoň 0,99

ad b)

$$d = \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}} = \sqrt{\frac{9 \cdot 0,044^2}{\chi^2_{0,995}(9)}} = \frac{3 \cdot 0,044}{\sqrt{23,589}} = 0,0272$$

$$h = \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}} = \sqrt{\frac{9 \cdot 0,044^2}{\chi^2_{0,005}(9)}} = \frac{3 \cdot 0,044}{\sqrt{1,735}} = 0,1002$$

0,0272 mm < σ < 0,1002 mm s pravděpodobností aspoň 0,99.

Příklad 10.: Získali jsme náhodný výběr rozsahu 18 z dvourozměrného rozložení, jímž se řídí náhodný vektor (X,Y). Je známo, že náhodné veličiny X a Y jsou ordinálního typu a že součet kvadrátů odchylek pořadí $\sum_{i=1}^{18} (R_i - Q_i)^2 = 502$. Na hladině významnosti 0,05 testujte hypotézu, že náhodné veličiny X a Y jsou pořadově nezávislé proti oboustranné alternativě.

Řešení:

Na hladině významnosti 0,05 testujeme H_0: X, Y jsou pořadově nezávislé náhodné veličiny proti oboustranné alternativě H_1: X, Y jsou pořadově závislé náhodné veličiny.

Vypočteme realizaci testové statistiky:

$$r_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (R_i - Q_i)^2 = 1 - \frac{6}{18(18^2-1)} 502 = \frac{2802}{5814} = 0,4819$$

V tabulkách najdeme kritickou hodnotu: $r_{S,0,95}(18) = 0,4716$. Protože 0,4819 > 0,4716, nulovou hypotézu zamítáme na hladině významnosti 0,05.

Příklad 11.: Je dán náhodný výběr rozsahu 6800 z dvourozměrného diskrétního rozložení, přičemž veličina X nabývá tří variant a veličina Y nabývá čtyř variant. Testová statistika pro test nezávislosti veličin X a Y nabyla hodnoty 1073,5.

a) Lze na asymptotické hladině významnosti 0,05 zamítnout hypotézu o nezávislosti veličin X a Y?

b) Jaký je Cramérův koeficient a jak ho lze interpretovat?

Řešení:

ad a) $W = \langle \chi^2_{0,95}((3-1)(4-1)), \infty) = \langle \chi^2_{0,95}(6), \infty) = \langle 12,592, \infty)$. Protože testová statistika se realizuje v kritickém oboru, nulovou hypotézu zamítáme na asymptotické hladině významnosti 0,05.

ad b) $V = \sqrt{\frac{K}{n(m-1)}} = \sqrt{\frac{1073,5}{6800 \cdot (3-1)}} = \sqrt{\frac{1073,5}{13600}} = 0,281$. Mezi veličinami X a Y existuje jenom slabá závislost.

Příklad 12.: Pro náhodný výběr (X_i, Y_i) , i = 1, ..., 10 z dvourozměrného normálního rozložení byl vypočten výběrový koeficient korelace –0,9325. Na hladině významnosti 0,01testujte hypotézu o nezávislosti veličin X, Y proti levostranné alternativě.

Řešení: Testová statistika $T = \dfrac{R_{12}\sqrt{n-2}}{\sqrt{1-R_{12}^2}} = \dfrac{-0{,}9325\sqrt{10-2}}{\sqrt{1-(-0{,}9325)^2}} = -7{,}3027$, kritický obor pro levostrannou alternativu $W = (-\infty, -t_{1-\alpha}(n-2)\rangle = (-\infty, -t_{0,99}(8)\rangle = (-\infty, -2{,}8965\rangle$. Protože testová statistika se realizuje v kritickém oboru, nulovou hypotézu zamítáme na hladině významnosti 0,01 ve prospěch levostranné alternativy.

Příklad 13.: Jsou dány dva nezávislé náhodné výběry o rozsazích $n_1 = 12$, $n_2 = 10$, první pochází z rozložení $N(\mu_1, \sigma_1^2)$, druhý z rozložení $N(\mu_2, \sigma_2^2)$, kde parametry $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ neznáme. Byly vypočteny realizace výběrových průměrů: $m_1 = 0{,}4832$, $m_2 = 0{,}5769$ a výběrových rozptylů: $s_1^2 = 2{,}3516$, $s_2^2 = 2{,}1268$.

a) Na hladině významnosti 0,1 testujte hypotézu, že neznámé rozptyly σ_1^2 a σ_2^2 jsou shodné proti oboustranné alternativě. Test proveďte pomocí intervalu spolehlivosti.
b) Na hladině významnosti 0,1 testujte hypotézu, že neznámé střední hodnoty μ_1 a μ_2 jsou shodné proti oboustranné alternativě. Test proveďte pomocí kritického oboru.

Řešení:

ad a) Testujeme H_0: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$ proti H_1: $\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$

$$d = \frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1, n_2-1)} = \frac{2{,}3516/2{,}1268}{F_{0,95}(11,9)} = \frac{2{,}3516/2{,}1268}{3{,}1025} = 0{,}3564$$

$$h = \frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1, n_2-1)} = \frac{2{,}3516/2{,}1268}{F_{0,05}(11,9)} = \frac{2{,}3516/2{,}1268}{1/F_{0,95}(9,11)} = \frac{2{,}3516/2{,}1268}{1/2{,}8962} = 3{,}2023$$

Protože číslo 1 leží v intervalu (0,3564; 3,2023), H_0 nezamítáme na hladině významnosti 0,1.

ad b) Testujeme H_0: $\mu_1 - \mu_2 = 0$ proti H_1: $\mu_1 - \mu_2 \neq 0$. Vypočteme vážený průměr výběrových rozptylů: $s_*^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \dfrac{11 \cdot 2{,}3516 + 9 \cdot 2{,}1268}{20} = 2{,}2504$. Dále stanovíme realizaci testové statistiky: $t_0 = \dfrac{m_1 - m_2 - c}{s_*\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} = \dfrac{0{,}4832 - 0{,}5769}{\sqrt{2{,}2504}\sqrt{\dfrac{1}{12}+\dfrac{1}{10}}} = -0{,}1459$ a kritický obor:

$$W = (-\infty, -t_{1-\alpha/2}(n_1+n_2-2)\rangle \cup \langle t_{1-\alpha/2}(n_1+n_2-2), \infty) = (-\infty, -t_{0,95}(20)\rangle \cup \langle t_{0,95}(20), \infty) =$$
$$= (-\infty, -1{,}7247\rangle \cup \langle 1{,}7247, \infty)$$

Protože testová statistika se nerealizuje v kritickém oboru, nulovou hypotézu nezamítáme na hladině významnosti 0,1.