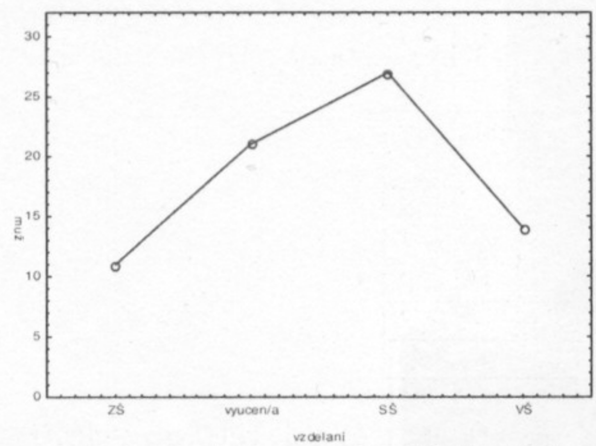
Řešení písemné zkoušky z předmětu Aplikovaná statistika I, 19.1.2016 (pro antropology)

Příklad 1.: 154 účastníků dotazníkového šetření, z toho 73 mužů, odpovídalo mj. na otázku, jaké je jejich nejvyšší dosažené vzdělání. Výsledky máme v tabulce:

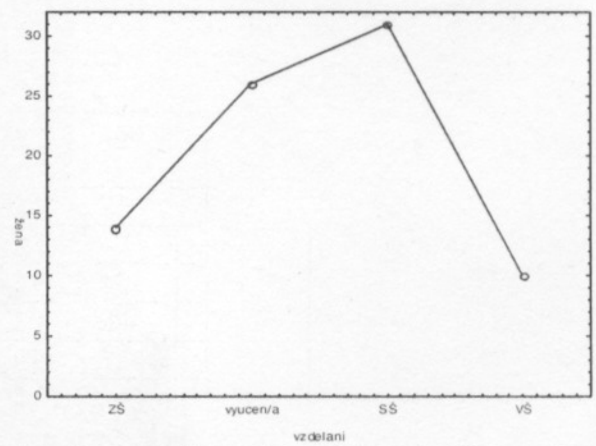
vzdělání	pohlaví		celkem
	muž	žena	
ZŠ	11	14	25
Vyučen/a	21	26	47
SŠ	27	31	58
VŠ	14	10	24
Celkem	73	81	154

a) Doplňte chybějící čísla do tabulky.
b) Kolik % žen má VŠ vzdělání? (10/81 = 12,3 %)
c) Kolik % vyučených účastníků šetření jsou muži? (21/47 = 44,7 %)
d) Kolik % účastníků šetření má aspoň SŠ vzdělání? (130/154 = 84,4 %)
e) Kolik % účastníků šetření je žen se základním vzděláním? (14/154 = 9 %)
f) Nakreslete polygon četností pro varianty vzdělání pro muže a pro ženy.

Pro muže:



Pro ženy:



Příklad 2.: Je dán datový soubor 12 1,1 6,3 3,9 11 5,8 2,5 8 4,1 2 9,5 6,6 1,7 3,4 4,9 3 10,3 2,2 5,4 15,5. Stanovte medián a první a devátý decil datového souboru.

Řešení:

Datový soubor uspořádáme vzestupně podle velikosti: 1,1 1,7 2 2,2 2,5 3 3,4 3,9 4,1 4,9 5,4 5,8 6,3 6,6 8 9,5 10,3 11 12 15,5

První decil je průměr 2. a 3. uspořádané hodnoty, tedy $x_{0,10} = (1,7 + 2)/2 = 1,85$

Devátý decil je průměr 18. a 19. uspořádané hodnoty, tedy $x_{0,90} = (11 + 12)/2 = 11,5$

Medián je průměr 10. a 11. uspořádané hodnoty, tedy $x_{0,50} = (4,9 + 5,4)/2 = 5,15$

Příklad 3.: Při provozu balicího automatu vznikají během směny náhodné poruchy, které se řídí rozložením Po(2). Jaká je pravděpodobnost, že během směny dojde k aspoň jedné poruše?

Řešení: X – počet poruch během směny, X ~ Po(2),

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \frac{2^0}{0!}e^{-2} = 0,8647.$$

Příklad 4.: Automat na kávu je seřízen tak, že plní šálky po 200 ml kávy se směrodatnou odchylkou 15 ml. Předpokládáme, že množství kávy v šálku se řídí normálním rozložením. Kolik procent šálků bude obsahovat méně než 224 ml kávy?

Řešení:

X – množství kávy v šálku, X ~ N(200, 225)

$$P(X < 224) = P\left(\frac{X-200}{15} < \frac{224-200}{15}\right) = P(U < 1{,}6) = \Phi(1{,}6) = 0{,}9452$$

Asi 94,5% šálků bude obsahovat méně než 224 ml kávy.

Příklad 5.: Stanovte následující kvantily: $u_{0,025}$, $\chi^2_{0,995}(19)$, $t_{0,01}(9)$, $F_{0,025}(3,9)$.

Řešení: $u_{0,025} = -1{,}96$, $\chi^2_{0,995}(19) = 38{,}852$, $t_{0,01}(9) = -2{,}8214$,

$$F_{0.025}(3,9) = \frac{1}{F_{0,975}(9,3)} = \frac{1}{14{,}4731} = 0{,}069$$

Příklad 6.: Z realizace náhodného výběru rozsahu 9, který pochází z rozložení $N(\mu, \sigma^2)$, byl vypočten výběrový průměr m = 15 a výběrový rozptyl $s^2 = 36$. Najděte 90% empirický interval spolehlivosti pro neznámou střední hodnotu μ.

Řešení: $d = m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 15 - \frac{6}{\sqrt{9}} t_{0,95}(8) = 15 - 2 \cdot 1{,}8595 = 11{,}281$

h = 18,719, tedy $11{,}281 < \mu < 18{,}719$ s pravděpodobností aspoň 0,9.

Příklad 7.: 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 8.: Do firmy bylo dodáno 2000 součástek. Při podrobném průzkumu bylo zjištěno, že 11 z nich je zmetkových. Lze na asymptotické hladině významnosti 0,05 zamítnout hypotézu, že pravděpodobnost výskytu zmetku je nevýše 0,1 %?

Řešení:

Zavedeme náhodné veličiny $X_1, ..., X_{2000}$, přičemž $X_i = 1$, když i-tá součástka byla zmetek a $X_i = 0$ jinak, i = 1, ..., 2000. Tyto náhodné veličiny tvoří náhodný výběr z rozložení $A(\vartheta)$.

Testujeme hypotézu H_0: $\vartheta \leq 0{,}001$ proti alternativě H_1: $\vartheta > 0{,}001$.

Známe: $n = 2000$, $m = \frac{11}{2000} = 0{,}055$, $c = 0{,}001$, $\alpha = 0{,}05$, $u_{1-\alpha} = u_{0,95} = 1{,}64$

Ověření podmínky $n\vartheta(1-\vartheta) > 9$: parametr ϑ neznáme, musíme ho nahradit výběrovým průměrem. Pak 2000.0,0045.0,9955 = 10,9395 > 9.

a) Testování pomocí kritického oboru:

Realizace testového kritéria: $t_0 = \dfrac{m-c}{\sqrt{\dfrac{c\cdot(1-c)}{n}}} = \dfrac{0{,}055-0{,}001}{\sqrt{\dfrac{0{,}001\cdot 0{,}999}{2000}}} = 6{,}3671.$

Kritický obor: $W = \langle u_{0,95}, \infty) = \langle 1{,}64, \infty)$. Protože $6{,}3671 \in W$, H_0 zamítáme na asymptotické hladině významnosti 0,05.

b) Testování pomocí intervalu spolehlivosti

$$d = m - \sqrt{\frac{m(1-m)}{n}}u_{1-\alpha} = 0{,}0055 - \sqrt{\frac{0{,}0055\cdot 0{,}9945}{2000}}1{,}64 = 0{,}0028$$

Protože číslo $c = 0{,}001$ neleží v intervalu $\langle 0{,}0028;\infty)$, H_0 zamítáme na asymptotické hladině významnosti 0,05.

c) Testování pomocí p-hodnoty

Protože testujeme nulovou hypotézu proti pravostranné alternativě, vypočteme p-hodnotu podle vzorce: $p = 1 - \Phi(6{,}3671) = 9{,}63*10^{-11}$.

Protože vypočtená p-hodnota je menší než hladina významnosti 0,05, H_0 zamítáme na asymptotické hladině významnosti 0,05.