Vzorová písemná zkouška z předmětu Aplikovaná statistika I

Úkol 1.: Pro kontingenční tabulku 3 x 3, kterou jsme vytvořili na základě náhodného výběru $(X_1, Y_1), \ldots, (X_{858}, Y_{858})$, byla spočtena testová statistika K = 464 pro test nezávislosti veličin X, Y. Určete Cramérův koeficient.
Řešení: Cramérův koeficient = 0,52

Úkol 2.: U 200 náhodně vybraných absolventů jisté VŠ, mezi nimiž bylo 73 žen, bylo zjišťováno, zda pracují v řídicí funkci. V řídicí funkci pracuje 57 mužů a 24 žen. Vypočtěte a interpretujte podíl šancí mužů a žen na práci v řídicí funkci.
Řešení: Sestavíme čtyřpolní kontingenční tabulku simultánních absolutních četností:

Práce v ŘF	pohlaví respondenta		$n_{j.}$
	muž	žena	
ano	57	24	81
ne	70	49	119
$n_{.k}$	127	73	200

$OR = \frac{ad}{bc} = \frac{57 \cdot 49}{24 \cdot 70} = 1{,}66$ což znamená, že muži mají 1,66x vyšší šanci na práci v řídicí funkci než ženy.

Úkol 3.: Je dána neúplná tabulka ANOVA. Místo otazníků doplňte chybějící čísla.

zdroj variability	součet čtverců	stupně volnosti	podíl	F_A
skupiny	?	2	?	?
reziduální	16,033	?	?	-
celkový	17,301	35	-	-

Řešení:

zdroj variability	součet čtverců	stupně volnosti	podíl	F_A
skupiny	**1,268**	2	**0,634**	**1,304**
reziduální	16,033	**33**	**0,486**	-
celkový	17,301	35	-	-

Úkol 4.: 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. Uveďte hodnotu testové statistiky, kritický obor a rozhodnutí o nulové hypotéze.

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:

$$r_S = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i - Q_i)^2 = 1 - \frac{6}{18(18^2-1)}502 = \frac{467}{969} = 0{,}4819$$

V tabulkách najdeme kritickou hodnotu: $r_{S,0,975}(18) = 0,4716$. Protože 0,4819 > 0,4716, nulovou hypotézu zamítáme na hladině významnosti 0,05.

Úkol 5.: Náhodný výběr rozsahu 400 pochází z normálního rozložení s neznámou střední hodnotou μ a rozptylem $\sigma^2 = 0{,}01$. Výběrový průměr nabyl hodnoty 0,01. Jaká je hodnota testové statistiky pro test hypotézy H_0: $\mu = 0$ proti alternativě H_1: $\mu \neq 0$?

Odpověď: $t_0 = \dfrac{m - c}{\frac{\sigma}{\sqrt{n}}} = \dfrac{0{,}01 - 0}{\frac{0{,}1}{\sqrt{400}}} = 2$

Úkol 6.: Pomocí mediánového testu testujeme na hladině významnosti 0,01 hypotézu, že 7 nezávislých náhodných výběrů, které mají všechny rozsah 11, pochází z téhož rozložení. Stanovte kritický obor pro test této hypotézy.
Řešení: $W = \langle \chi^2_{1-\alpha}(r-1), \infty) = \langle \chi^2_{0,99}(6), \infty) = \langle 16{,}812, \infty)$

Úkol 7.: Pro náhodný výběr (X_i, Y_i) , i = 1, ..., 27 z dvourozměrného normálního rozložení byl vypočten výběrový koeficient korelace 0,77. Na hladině významnosti 0,01 testujte hypotézu o nezávislosti veličin X, Y proti pravostranné alternativě.
Řešení: Testová statistika $T = \dfrac{R_{12}\sqrt{n-2}}{\sqrt{1 - R_{12}^2}} = \dfrac{0{,}77\sqrt{27-2}}{\sqrt{1 - 0{,}77^2}} = 6{,}034$, kritický obor pro pravostrannou alternativu $W = \langle t_{1-\alpha}(n-2), \infty) = \langle t_{0,99}(25), \infty) = \langle 2{,}462, \infty)$. Protože testová statistika se realizuje v kritickém oboru, nulovou hypotézu zamítáme na hladině významnosti 0,01 ve prospěch levostranné alternativy.

Úkol 8.: V 500 po sobě jdoucích zápisech v matrice narozených je 266 chlapců a 234 dívek. Lze na asymptotické hladině významnosti 0,1 zamítnout hypotézu, že narození dívky a chlapce je stejně pravděpodobné?
Řešení: Zavedeme náhodný výběr $X_1, \dots, X_{500}$, přičemž $X_i = 1$, když i-té narozené dítě je chlapec a $X_i = 0$, když i-té narozené dítě je dívka, i = 1, .., 500. Tyto náhodné veličiny tvoří náhodný výběr z rozložení $A(\vartheta)$. Na asymptotické hladině významnosti 0,1 testujeme hypotézu H_0: $\vartheta = 0{,}5$ proti H_1: $\vartheta \neq 0{,}5$.

Ověření podmínky dobré aproximace: $500 \cdot \dfrac{266}{500} \cdot \dfrac{234}{500} = 124{,}488$, což je větší než 9, podmínka je splněna.

Realizace testové statistiky: $t_0 = \dfrac{m - c}{\sqrt{\frac{c(1-c)}{n}}} = \dfrac{0{,}532 - 0{,}5}{\sqrt{\frac{0{,}5(1-0{,}5)}{500}}} = 1{,}4311$

Kritický obor: $W = (-\infty, -u_{0,95}\rangle \cup \langle u_{0,95}, \infty) = (-\infty, -1{,}28155\rangle \cup \langle 1{,}28155, \infty)$

Protože testová statistika se realizuje v kritickém oboru, nulovou hypotézu zamítáme na asymptotické hladině významnosti 0,1. S rizikem omylu nejvýše 0,1 jsme tedy prokázali, že pravděpodobnost narození chlapce a dívky se liší.

Upozornění: Každý úkol je hodnocen maximálně 7 body. Pro úspěšné absolvování písemné části zkoušky je nutné získat aspoň 30 bodů.