Vzorová písemná zkouška – teoretická část

Úkol 1.: Popište, v čem spočívá rozdíl mezi jednovýběrovým a párovým Wilcoxonovým testem
Řešení: Jednovýběrový Wilcoxonův test slouží k testování hypotézy, že medián rozložení, z něhož pochází daný náhodný výběr, je rovna nějaké konstantě, zatímco párový Wilcoxonův test slouží k testování hypotézy, že rozdíl mediánů dvourozměrného rozložení, z něhož pochází daný dvourozměrný náhodný výběr, je roven nějaké konstantě.

Úkol 2.: Za jakých podmínek může být výběrový koeficient korelace R_{12} považován za přibližně nestranný odhad teoretického koeficientu korelace ρ?
Řešení: Pokud rozsah dvourozměrného náhodného výběru je aspoň 30.

Úkol 3.: Máme náhodný výběr $X_1, \ldots, X_n$ z normálního rozložení $N(\mu,\sigma^2)$, kde rozptyl σ^2 známe. Na hladině významnosti α testujeme H_0: $\mu = c$ proti H_1: $\mu \neq c$. Jak se nazývá uvedený test? Jakým rozložením se řídí testové kritérium, je-li nulová hypotéza pravdivá?
Řešení: Jedná se o jednovýběrový z-test. Testové kritérium se řídí rozložením $N(0,1)$.

Úkol 4.: Jaký je kritický obor pro test hypotézy o shodě středních hodnot $r \geq 3$ normálních rozložení se stejným rozptylem, pokud předpokládáme, že test provádíme na základě znalosti r nezávislých náhodných výběrů o celkovém rozsahu n a hladina významnosti je α?
Řešení: $W = \langle F_{1-\alpha}(r-1, n-r), \infty)$

Úkol 5.: Máme jednorozměrný datový soubor tvořený hodnotami $x_1, \ldots, x_n$. Jak se nazývá hodnota, která leží v intervalu $(x_{0,25} - 3q, x_{0,25} - 1{,}5q)$, kde $q = x_{0,75} - x_{0,25}$ je kvartilová odchylka.
Řešení: Jde o odlehlou hodnotu.

Každý z úkolů je hodnocen maximálně 8 body. Na teoretickou část lze získat 40 bodů.

Vzorová písemná zkouška – praktická část

Úkol 1.: 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, $S_A = S_T - S_E = 15 - 3 = 12$

$$F = \frac{S_A/(r-1)}{S_E/(n-r)} = \frac{12/4}{3/26} = 26, F_{0,95}(4,26) = 2{,}9752$$

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

Úkol 2.: Z realizace náhodného výběru rozsahu 100, který pochází z rozložení $N(\mu, \sigma^2)$, byl vypočten výběrový průměr m = 15 a výběrový rozptyl $s^2 = 36$. Najděte 95% empirický interval spolehlivosti pro neznámou střední hodnotu μ.

Řešení: $d = m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 15 - \frac{6}{\sqrt{100}} t_{0,975}(99) = 15 - \frac{6}{10} 1{,}96 = 13{,}82$

h = 16,18, tedy $13{,}82 < \mu < 16{,}18$ s pravděpodobností aspoň 0,95.

Úkol 3.: 100 náhodně vybraných pokusných osob mělo nezávisle na sobě odhadnout, kdy od daného signálu uplyne minuta. 35 osob délku 1 minuty nadhodnotilo, přičemž součet pořadí odchylek jejich odhadů od 1 minuty činil 3110. Zbylých 65 osob délku 1 minuty podhodnotilo. Lze na asymptotické hladině významnosti 0,05 zamítnout hypotézu, že polovina lidské populace délku jedné minuty nadhodnotí a polovina podhodnotí?

Řešení: Testujeme hypotézu H_0: $x_{0,50} = 0$ proti H_1: $x_{0,50} \neq 0$.

Úloha vede na jednovýběrový Wilcoxonův test, kde $S_W^+ = 3110$

Realizace asymptotické testové statistiky: $U_0 = \dfrac{S_W^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \dfrac{3110 - \frac{100\cdot 101}{4}}{\sqrt{\frac{100\cdot 101\cdot 201}{24}}} = 2{,}0114$

hodnota příslušného kvantilu = 1,96, rozhodnutí o nulové hypotéze: zamítáme na asymptotické hladině významnosti 0,05.

Úkol 4.: Pro 12 náhodně vybraných ojetých automobilů byl vypočten výběrový koeficient korelace mezi jejich stářím v měsících a počtem najetých kilometrů. Nabyl hodnoty 0,831. Za předpokladu, že data pocházejí z dvourozměrného normálního rozložení, testujte na hladině významnosti 0,05 hypotézu o nezávislosti obou veličin. Uveďte hodnotu testové statistiky a kritický obor.

Řešení: Testujeme H_0: $\rho = 0$ proti oboustranné alternativě H_1: $\rho \neq 0$ Testová statistika

$$T = \frac{R_{12}\sqrt{n-2}}{\sqrt{1-R_{12}^2}} = \frac{0{,}831\sqrt{10}}{\sqrt{1-0{,}831^2}} = 4{,}724$$. Kritický obor

$W = \left(-\infty, -t_{1-\alpha/2}(n-2)\right\rangle \cup \left\langle t_{1-\alpha/2}(n-2), \infty\right) = \left(-\infty, -2{,}2281\right\rangle \cup \left\langle 2{,}2281, \infty\right)$. Protože $T \in W$, H_0 zamítáme na hladině významnosti 0,05.

Úkol 5.: Nechť $X_1, \ldots, X_n$ je náhodný výběr z $N(\mu,\sigma^2)$, kde rozptyl známe. Jak musíme změnit rozsah výběru, aby šířka 100(1-α)% empirického intervalu spolehlivosti pro střední hodnotu klesla na polovinu?

Řešení: Označme d, h meze původního intervalu spolehlivosti a n původní rozsah náhodného výběru. Šířka intervalu spolehlivosti je $h - d = m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} - (m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}) = \frac{2o}{\sqrt{n}} u_{1-\alpha/2}$,

tedy $n \geq \frac{4\sigma^2 u_{1-\alpha/2}{}^2}{(h-d)^2}$.

Nyní označme d*, h* meze nového intervalu spolehlivosti a n* nový rozsah náhodného výběru. Požadujeme, aby $h^* - d^* = \frac{h-d}{2}$, tedy $n^* \geq \frac{4\sigma^2 u_{1-\alpha/2}{}^2}{(h^* - d^*)^2} = 4 \cdot \frac{4\sigma^2 u_{1-\alpha/2}{}^2}{(h-d)^2}$. Vidíme, že rozsah výběru se musí zvětšit 4 x.

Každý úkol je hodnocen maximálně 12 body. Na praktickou část lze získat 60 bodů. Body z praktické a teoretické části se sčítají.

Výsledné hodnocení:
(90, 100] … A, (80, 90] … B, (70, 80] … C, (60, 70] … D, (50, 60] … E, [0, 50] … F