Teoretická část zkoušky ze Statistiky II, 18.12.2014

Úkol 1.: Nechť $X_1, \ldots, X_n$ je náhodný výběr z $N(\mu,\sigma^2)$. Pomocí distribuční funkce rozložení $N(0,1)$ vyjádřete pravděpodobnost, že výběrový průměr překročí hodnotu k.

Úkol 2.: Nechť X_1, X_2, X_3, X_4 je náhodný výběr z $Rs(0,b)$, kde parametr $b > 0$ neznáme. Určete konstantu c tak, aby statistika $T = X_1 + X_2/2 + X_3/3 + cX_4$ byla nestranným odhadem parametru b.

Úkol 3.: Nechť $X_1, \ldots, X_{400}$ je náhodný výběr z $N(\mu,0{,}01)$. Je známo, že výběrový průměr se realizoval hodnotou 0,01. Na hladině významnosti 0,05 testujte hypotézu H_0: $\mu = 0$ proti pravostranné alternativě H_1: $\mu > 0$ pomocí p-hodnoty.

Úkol 4.: Dvourozměrný náhodný výběr $(X_1, Y_1), \ldots, (X_{20}, Y_{20})$ pochází z dvourozměrného normálního rozložení. Výběrový koeficient korelace se realizoval hodnotou $r_{12} = 0{,}4$. Na hladině významnosti 0,05 testujte hypotézu o nezávislosti náhodných veličin X, Y proti pravostranné alternativě. Výsledek interpretujte.

Úkol 5.: Jsou dány čtyři nezávislé náhodné výběry postupně z rozložení $N(\mu_1,\sigma^2)$, $N(\mu_2,\sigma^2)$, $N(\mu_3,\sigma^2)$, $N(\mu_4,\sigma^2)$, přičemž každý z nich má rozsah 6. Na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot, je-li známo, že skupinový součet čtverců je 46,5 a reziduální 67,5.

Praktická část zkoušky ze Statistiky II, 18.12.2014

Příklad 1.: (15 bodů) Výkon 18 gymnastek byl ohodnocen stanovením jejich pořadí od nejlepší (pořadí 1) po nejslabší (pořadí 18). V hodnocené skupině bylo 11 žákyň trenérky A a 7 žákyň trenérky B. V tabulce je uvedeno pořadí žákyň obou trenérek:

A	1	4	5	7	8	10	11	13	14	16	17
B	2	3	6	9	12	15	18				

Na hladině významnosti 0,05 testujte hypotézu, že výukové metody obou trenérek jsou stejně účinné proti oboustranné alternativě.

Příklad 2.: (15 bodů) Náhodně bylo vybráno třicet vozidel stejné kategorie. U nich byl zjišťován: výkon (veličina X_1, v kW), maximální rychlost (veličina X_2, v km/hod.), spotřeba (veličina X_3, v l/100 km) a cena (veličina Y, v tisících Kč). Předpokládáme, že cena závisí na výkonu, maximální rychlosti a spotřebě lineárně.

Máte k dispozici výstupní tabulku vícenásobné lineární regrese ze systému STATISTICA:

	Výsledky regrese se závislou proměnnou : Y (Tabulka1) R= ,90483644 R2= ,81872899 Upravené R2= ,79781310 F(3,26)=39,144 p<,00000 Směrod. chyba odhadu : 62,812					
N=30	b*	Sm.chyba z b*	b	Sm.chyba z b	t(26)	p-hodn.
Abs.člen			116,7074	198,4404	0,58812	0,561524
X1	0,833407	0,215908	4,8601	1,2591	3,86002	0,000673
X2	0,328018	0,200519	2,2255	1,3605	1,63584	0,113923
X3	-0,356197	0,126022	-55,2524	19,5483	-2,82646	0,008930

a) Napište regresní rovnici vyjadřující závislost ceny na výkonu, maximální rychlosti a spotřebě.
b) Z kolika procent je variabilita ceny vysvětlena tímto lineárním regresním modelem?
c) Je na hladině významnosti 0,05 dostačující model konstanty? Rozhodnutí zdůvodněte.
d) Na kterých regresorech cena nezávisí na hladině významnosti 0,05? Rozhodnutí zdůvodněte.

Příklad 3.: (15 bodů) 30 náhodně vybraných absolventů určité vysoké školy bylo dotázáno, zda pracují v řídicí funkci a zda absolvovali školu s vyznamenáním. Ze 14 respondentů, kteří prospěli s vyznamenáním, 8 pracuje v řídicí funkci a ze 16 respondentů, kteří pouze prospěli, pracují 3 v řídicí funkci. Na asymptotické hladině významnosti 0,05 testujte hypotézu, že práce v řídicí funkci a způsob ukončení dané vysoké školy jsou nezávislé.
a) Sestavte kontingenční tabulku;
b) ověřte splnění podmínek dobré aproximace;
c) vypočtěte realizaci testové statistiky;
d) stanovte kritický obor a na dané hladině významnosti rozhodněte o nulové hypotéze.

Příklad 4.: (15 bodů) 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 μ_1, μ_2, σ_1^2, σ_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.

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