

## Varianta S3

**Teoretické otázky:**

- 1) K čemu slouží index determinace? Jakých hodnot může nabýt?
- 2) U kterých z uvedených funkcí nelze k odhadu parametrů použít metodu nejmenších čtverců: hyperbola, mocninná funkce, parabola, exponenciála? Názor vysvětlete!
- 3) Pro která pravděpodobnostní rozdělení je jejich střední hodnota rovna mediánu a zároveň modu? Vysvětlete a uveďte alespoň 2 příklady!
- 4) Jak budete postupovat, jestliže chcete zjistit, jak se na změně průměrné prodejní ceny určité komodity podílely změny cen u jednotlivých prodejců a jak změna struktury prodeje podle jednotlivých prodejců?
- 5) Vysvětlete pojem kvantil. Jaké skupiny kvantilů znáte? Blíže je charakterizujte!

**Příklady:**

- 1) V následující tabulce jsou údaje o vývozu ekologických solárních zařízení v mld. Kč stálých cen:

Rok	1998	1999	2000	2001	2002	2003
Vývoz v mld. Kč	495	537,7	620,3	651,8	716,2	713,5

Charakterizujte vývoj vývozu solárních zařízení pomocí řetězových a bazických indexů (1998 = 1) a vypočítejte průměrný koeficient růstu a průměrný absolutní přírůstek vývozu za období 1998 – 2003. Výsledky interpretujte (u bazických a řetězových indexů stačí interpretace vždy 1 z nich)! 3 BODY

- 2) Vypočítejte celkový rozptyl, máme – li k dispozici následující údaje:

Skupina	$\bar{x}_i$	$n_i$	$s_i$
1	5	10	2,0
2	4	15	0,8
3	2	25	0,5
<b>Celkem</b>	x	50	x

2 BODY

- 3) Mezi zákazníky Domu knih bylo náhodně vybráno 400 mužů a žen za účelem ověření hypotézy o nezávislosti preferovaného literárního žánru a pohlaví. Byly získány tyto údaje:

Pohlaví zákazníka	Literární žánr					Součet
	sci – fi	milostné romány	vzdělávací literatura	válečné romány	historické romány	
Muž	29	9	25	40	37	140
Žena	14	60	102	28	56	260
<b>Součet</b>	43	69	127	68	93	400

Na hladině významnosti  $\alpha = 0,01$  ověřte hypotézu o nezávislosti literárního žánru a pohlaví zákazníka. Případně změřte vhodnou charakteristikou sílu závislosti. 2 BODY

Celkový počet bodů: 12

Minimální počet bodů pro úspěšné složení zkoušky: 7,25

Minimální počet bodů z teorie: 1,75

Minimální počet bodů z příkladů: 2,5

Každá teoretická otázka je za 1 bod.

## Teoretické otázky – řešení:

- 1) Slouží k měření těsnosti závislosti číselných proměnných, jejichž vztah je popsán pomocí nějaké matematické funkce. Zároveň je jedním z kritérií pro posouzení kvality vybrané regresní funkce. Nabývá hodnot z intervalu  $<0, 1>$ .
- 2) Není to možné u mocninné funkce a exponenciály, protože to jsou funkce nelineární v parametrech. Aplikace metody nejmenších čtverců u takových funkcí může vést k soustavě nelineárních rovnic, které nemusí mít jednoznačné řešení, nebo u nich nedokážeme parametry vyjádřit pomocí vhodných výpočtových vzorců.
- 3) Pro symetrická rozdělení. Př. normální, normované normální a Studentovo rozdělení.
- 4) Index proměnlivého složení, jímž charakterizujeme změnu průměrné ceny dané komodity, rozložíme např. pomocí metody postupných změn na dva dílčí indexy – index stálého složení a index struktury. Index stálého složení vyjadřuje, jak se na změně průměrné ceny podílely změny cen u jednotlivých prodejců, a index struktury vyjadřuje, jak se na změně průměrné ceny podílela změna struktury prodeje.
- 5) Kvantil je hodnota, která rozděluje uspořádaný statistický soubor na určitý počet stejně obsazených částí. Př. kvartily – tři kvantily ( $\tilde{x}_{25}, \tilde{x}, \tilde{x}_{75}$ ), které rozdělují neklesající řadu hodnot na 4 stejně četné části. Decily – 9 kvantilů ( $\tilde{x}_{10}, \tilde{x}_{20}, \tilde{x}_{30}, \tilde{x}_{40}, \tilde{x}, \tilde{x}_{60}, \tilde{x}_{70}, \tilde{x}_{80}, \tilde{x}_{90}$ ), které rozdělují uspořádaný statistický soubor na 10 stejně četných částí. Atd.

## Příklady – řešení:

1)

Rok	1998	1999	2000	2001	2002	2003
Vývoz v mld. Kč	495	537,7	620,3	651,8	716,2	713,5
Řetěz. index	•	1,0863	1,1536	1,0508	1,0988	0,9962
Baz. index (1 = 1998)	1	1,0863	1,2531	1,3168	1,4469	1,4414

Řetězové indexy jsou obecně definovány jako podíl hodnoty ukazatele v běžném a základním období, tj. pro rok 1998 nelze hodnotu řetězového indexu zjistit z poskytnutých dat, protože rok 1997, se kterým bychom měli srovnávat, není k dispozici. Pro rok 1999 počítáme podle:

$$i_{1999/1998} = \frac{u_{1999}}{u_{1998}} = \frac{537,7}{495} = 1,0863$$

$$\text{Dále pak: } i_{2000/1999} = \frac{u_{2000}}{u_{1999}} = \frac{620,3}{537,7} = 1,1536 \text{ atd.}$$

*Interpretace:* V roce 2000 vzrostl vývoz solárních zařízení oproti roku 1999 o 15,36 %.

Bazické indexy jsou definovány jako podíl hodnoty ukazatele v běžném a bazickém období. Zde je bazickým obdobím rok 1998, proto je u něj hodnota 1. Další index vypočítáme jako:

$$i_{1999/1998} = \frac{u_{1999}}{u_{1998}} = \frac{537,7}{495} = 1,0863$$

Dále pak:

$$i_{2000/1998} = \frac{u_{2000}}{u_{1998}} = \frac{620,3}{495} = 1,2531$$

$$i_{2001/1998} = \frac{u_{2001}}{u_{1998}} = \frac{651,8}{495} = 1,3168 \text{ atd.}$$

*Interpretace:* V roce 2001 vzrostl vývoz solárních zařízení oproti roku 1998 o 31,68 %.

Průměrný absolutní přírůstek je možné vypočítat podle:

$$\bar{\Delta} = \frac{y_n - y_1}{n - 1}$$

$$\bar{\Delta} = \frac{y_n - y_1}{n - 1} = \frac{713,5 - 495}{6 - 1} = 43,7$$

*Interpretace:* Vývoz solárních zařízení v letech 1998-2003 v průměru vzrostl o 43,7 mld. Kč.

Průměrný koeficient růstu je možné vypočítat podle:

$$\bar{k} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

$$\bar{k} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[5]{\frac{713,5}{495}} = 1,076$$

*Interpretace:* Vývoz solárních zařízení v letech 1998-2003 v průměru vzrostl o 7,6 % (NEBO vzrostl 1,076krát).

- 2) Vzhledem k tomu, že máme soubor rozdělený do dílčích částí, ve kterých jsou nám známy dílčí charakteristiky, použijeme k řešení větu o rozkladu rozptylu.

Skupina	$\bar{x}_i$	$n_i$	$s_i$	$\bar{x}_i n_i$	$\bar{x}_i^2 n_i$	$s_i^2$	$s_i^2 n_i$
1	5	10	2,0	50	250	4,00	40,00
2	4	15	0,8	60	240	0,64	9,60
3	2	25	0,5	50	100	0,25	6,25
<b>Celkem</b>	x	50	x	160	590	x	55,85

Pomocné výpočty jsou uvedeny v tabulce modrou barvou.

$$s_x^2 = \bar{s}^2 + s_{\bar{x}_i}^2$$

$$s_{\bar{x}_i}^2 = \frac{\sum_{i=1}^k \bar{x}_i^2 n_i}{\sum_{i=1}^k n_i} - \left( \frac{\sum_{i=1}^k \bar{x}_i n_i}{\sum_{i=1}^k n_i} \right)^2$$

$$\bar{s}^2 = \frac{\sum_{i=1}^k s_i^2 n_i}{\sum_{i=1}^k n_i}$$

$$s_{\bar{x}_i}^2 = \frac{\sum_{i=1}^k \bar{x}_i^2 n_i}{\sum_{i=1}^k n_i} - \left( \frac{\sum_{i=1}^k \bar{x}_i n_i}{\sum_{i=1}^k n_i} \right)^2 = \frac{590}{50} - \left( \frac{160}{50} \right)^2 = 1,56$$

$$\bar{s}^2 = \frac{\sum_{i=1}^k s_i^2 n_i}{\sum_{i=1}^k n_i} = \frac{55,85}{50} = 1,117$$

$$s_x^2 = 1,56 + 1,117 = 2,677$$

Celkový rozptyl má hodnotu 2,677.

- 3) Vhodnou metodou pro zkoumání závislosti preferovaného literárního žánru na pohlaví čtenáře je test nezávislosti kategoriálních znaků, neboť obě proměnné jsou nominální. Je doporučeno řešit prostřednictvím statistického programu. Prvním úkolem je ověřit předpoklady pro použití dané metody, tj. v každém poličku kontingenční tabulky musí být teoretické četnosti  $n'_{ij} \geq 5$ . V programu STATGRAPHICS Centurion 18 získáme tabulku teoretických četností:

**Frequency Table**

	sci-fi	cervena_knih	vzdel_lit	valec_romany	hist_romany	Row Total
muz	29	9	25	40	37	140
	15,05	24,15	44,45	23,80	32,55	35,00%
zena	14	60	102	28	56	260
	27,95	44,85	82,55	44,20	60,45	65,00%
Column Total	43	69	127	68	93	400
	10,75%	17,25%	31,75%	17,00%	23,25%	100,00%

Cell contents:

Observed frequency

Expected frequency

Teoretické četnosti jsou zvýrazněny modrou barvou. Vidíme, že všechny hodnoty převyšují hodnotu 5, proto je předpoklad pro použití testu nezávislosti kategoriálních dat splněn.

$H_0$ : preferovaný literární žánr nezávisí na pohlaví zákazníka

$H_1$ : non  $H_0$

$$G = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}}$$

$$W \equiv \{G; G > \chi_{1-\alpha}^2[(r-1)(s-1)]\}$$

$$W \equiv \{G; G > \chi_{99}^2(4)\}$$

$$W \equiv \{G; G > 13,277\}$$

$$G = 65,508$$

$G \in W \Rightarrow$  zamítáme  $H_0$ , přijímáme  $H_1$ .

Na hladině významnosti 1 % jsme prokázali, že preferovaný literární žánr závisí na pohlaví zákazníka.

Pozn.: Pokud budeme celý příklad řešit v programu, závěr testu provedeme na základě P-Value:

#### Tests of Independence

Test	Statistic	Df	P-Value
Chi-Square	65,508	4	0,0000

P-Value = 0,0 <  $\alpha$ , proto zamítáme  $H_0$ , přijímáme  $H_1$ .

Vzhledem k tomu, že závislost mezi proměnnými byla prokázána, má smysl změřit její sílu.

K tomu využijeme např. Cramérův koeficient kontingence:

$$C_c = \sqrt{\frac{G}{n \cdot h}} = \sqrt{\frac{65,508}{400 \cdot 1}} = 0,405$$

Závislost mezi preferovaným literárním žánrem a pohlavím zákazníka je spíše menší.