

Cvičení 11

V následujících příkladech předpokládejte, že data jsou pořízena jako náhodné výběry z populací sledujících normální rozdělení. K tomuto předpokladu však nepřistupujte nekriticky! Pokud budete přesvědčeni, že zde musíte použít jiný test, učiňte to.

1. Myšlenka analýzy rozptylu je postavena na rozkladu celkového rozptylu s_c^2 na rozptyl způsobený efektem – rozptyl mezi skupinami s_x^2 a rozptyl způsobený náhodným působením – rozptyl uvnitř tříd, tzv. reziduální rozptyl s_r^2 . Odchylka pozorování od celkové střední hodnoty $(Y_{ij} - \bar{Y}_\bullet)$ se dá rozepsat jako odchylka pozorování od průměru ve skupině $(Y_{ij} - \bar{Y}_i)$ a odchylka průměru skupiny od celkového průměru $(\bar{Y}_i - \bar{Y}_\bullet)$, tj. $(Y_{ij} - \bar{Y}_\bullet) = (Y_{ij} - \bar{Y}_i) + (\bar{Y}_i - \bar{Y}_\bullet)$. Pro součty čtverců těchto odchylek pak platí: ○T○
Těžší
Pracné

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_\bullet)^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^k n_i \cdot (\bar{Y}_i - \bar{Y}_\bullet)^2,$$

tj. $(n - 1) \cdot s_c^2 = (n - k) \cdot s_r^2 + (k - 1) \cdot s_x^2$. Dokažte, že tento vztah platí (třeba pomocí důkazu sporem).¹

2. Společnost s ručením omezeným prodávající sportovní obuv vlastní v ČR 30 obchodů. K dispozici jsou dostupné údaje o průměrném počtu kusů prodaných za měsíc (viz tabulka 1). Uvedené údaje byly získány v šesti náhodně vybraných prodejnách. Pro každou značku obuvi se jedná o jiný výběr, tj. například prodejna označená číslem jedna je jiná v případě obuvi značky Adidas a jiná v případě obuvi značky Nike apod.

Tabulka 1: Prodávánost obuvi

Značka	Prodejna					
	1	2	3	4	5	6
Adidas	25	29	28	27	30	24
Nike	22	19	23	24	30	17
Reebok	24	27	28	27	30	28

- (a) Lze předpokládat, že je variabilita prodeje obuvi sledovaných značek stejná?

¹Důkaz sporem (reductio ad absurdum) je typ logického důkazu (ano matematika i statistika jsou logické obory), ve kterém se snažíme prokázat, že předpoklad vede k nesmyslnému výsledku (ke sporu), což znamená, že předpoklad je nepravdivý, a tedy platí jeho negace. Jelikož důkaz sporem je založen na zákonu o vyloučení třetího, lze jej použít pouze v těch logických systémech, ve kterých tento zákon platí – třeba v aritmetice. Chceme-li pak dokázat, že z bezesporných tvrzení A_1, A_2, \dots, A_n plyne tvrzení B , pak se obvykle „sporem“ předpokládá, že platí tvrzení A_1, A_2, \dots, A_n a $\neg B$. Pokud zjistíme, že dojdeme k nesmyslu (např. $x^2 = -8$ pro $x \in \mathbb{R}$), pak musí dohromady platit tvrzení A_1, A_2, \dots, A_n a B , čímž byl důkaz proveden.

- (b) Zjistěte, zda zákazníci prokazatelně preferují některou značku.
- (c) Jak by se situace změnila, kdyby průměrné hodnoty prodaných kusů obuvi za měsíc pocházely pouze ze šesti prodejen tak, že zjištěné údaje za jednotlivé prodejny by byly vždy rozděleny podle značky? Otestujete stejnou hypotézu jako v 2b) i pro tuto situaci.
3. Bylo vybráno 15 studentů a ti byli rozděleni do tří skupin o stejném počtu studentů. Každý student nezávisle na ostatních měl vyřešit týž zadaný problém. U každého se zjišťoval čas v minutách potřebný k vyřešení zadaného úkolu. Všechny skupiny měly stejné podmínky, rozdíl byl jen v navození situace. V jedné skupině byli studenti pochváleni za dosavadní vynikající práci ve škole a bylo jim zdůrazněno, že předložený problém se dá snadno vyřešit. U další skupiny byly konstatovány různé nedostatky mladé generace a bylo prohlášeno, že takoví lidé mohou jen stěží vyřešit předložený problém. U poslední skupiny nebyla situace navozena žádným komentářem. Výsledky pokusu jsou uvedeny v následující tabulce (viz tabulka 2).

Tabulka 2: Časy nutné k vyřešení problému

Skupina	Čas nutný k řešení problému (min)				
A (pochvala)	2	4	5	2	3
B (konstatování nedostatků)	1	4	4	3	2
C (kontrolní skupina)	5	2	4	5	3

Měla počáteční motivace vliv na výsledný čas řešení v jednotlivých skupinách?

4. Bylo sledováno procento niklu v tavební analýze legované oceli. Analýza se prováděla u 4 pecí a u každé pece bylo odebráno 5 vzorků. Data jsou uvedena v následující tabulce (viz tabulka 3).
- (a) Má se zjistit, zda procento niklu je u všech pecí stejné, nebo zda se některé pece od sebe liší.

Tabulka 3: % Ni v tavební analýze legované oceli

1. pec	2. pec	3. pec	4. pec
4,15	4,38	4,23	4,41
4,26	4,40	4,16	4,31
4,10	4,29	4,20	4,42
4,30	4,39	4,24	4,37
4,25	4,45	4,27	4,43

- (b) Pro otestování stejné hypotézy použijte také Kruskalův-Wallisův test².

²O normalitě údajů mnoho nevíme, podle Levenova testu však nelze vyloučit homoskedasticitu, jak jste se měli možnost přesvědčit, když jste ji ověřovali pro ANOVAu.

5. Soutěž o nejlepší jakost výrobků obeslali čtyři výrobci A, B, C, D celkem 26 výrobky. Porota sestavila toto pořadí (uveden pouze původ výrobku od nejlepšího k nejhoršímu) (viz tabulka 4).

Tabulka 4: Výsledky soutěže

Pořadí	1	2	3	4	5	6	7	8	9	10	11	12	13
Výrobce	B	C	C	A	B	D	D	C	A	B	B	D	C
Pořadí	14	15	16	17	18	19	20	21	22	23	24	25	26
Výrobce	D	C	B	C	A	C	D	D	C	C	A	C	A

Na základě těchto údajů posudte, zda původ výrobků má vliv na jeho jakost³.

6. Na základě dat o bytech z předminulého cvičení otestujte na hladině významnosti $\alpha = 0,05$, zda cena bytu v Českých Budějovicích závisí na typu vlastnictví.
7. Pět ras psů se účastnilo armádního výcviku v Grabštejně. Po ukončení výcviku byly provedeny zkoušky u vybraných psů z jednotlivých ras. Jednalo se o zkoušky: chůze u nohy, hledání stopy, cviky sedni–lehni–vstaň, aportování, zadržení pachatele a překonání překážky. Jednotliví psi byli v každé disciplíně bodováni. Ukazatelem úrovně výcviku byl součet získaných bodů. Rozhodněte, zda lze předpokládat, že kvalita výcviku závisí nejenom na tom, kdo výcvik vede, ale i na tom, jaká rasa se cvičí, máte-li k dispozici výsledky zkoušek (viz tabulka 5).

Tabulka 5: Výsledky hodnocení psích dovedností

Rasa	Pes					
	1	2	3	4	5	6
Německý ovčák	98	95	89	90	97	95
Dobrman	80	85	76	56	80	70
Rotvailer	95	90	93	89	87	75
Velký knírač	69	81	78	85	89	90
Briard	78	69	65	80	78	78

8. U 15 osob byly zkoušeny 3 vyučovací metody: samostudium, studium prostřednictvím počítače a studium s pomocí učitele. Pokus probíhal tak, že každý člověk nastudoval prostřednictvím každé metody určitou oblast a poté se podrobil testu sestávajícího ze 20 otázek. Počet správných odpovědí je uveden v tabulce (viz tabulka 6).

Lze předpokládat, že je efektivnost všech tří vyučovaných metod stejná (volte $\alpha = 0,05$)?

³Pořadí, které je známo ze zadání příkladu, jistě nesplňuje předpoklad normality. Máte však k dispozici test, který normalitu dat nevyžaduje, dokonce je přímo na pořadí založen. Použijte ten.

Tabulka 6: Ověřování efektivnosti vyučovacích metod

	osoby														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. metoda	13	15	12	14	12	13	11	13	11	12	9	11	10	7	10
2. metoda	15	13	14	15	14	13	14	12	12	14	10	14	9	9	6
3. metoda	16	17	15	16	15	17	9	13	15	13	13	14	9	14	8

9. Na 10% hladině významnosti rozhodněte zda ovlivňuje počet vytvořených čárek („čím více tím lépe“ způsob jejich zápisu za předem určený časový limit 10 s. Nezávislý pokus (tj. každá osoba byla zařazena právě do jedné skupiny) byl proveden na přednášce z předmětu Statistika (ZS 2007/2008) na EF JU v Českých Budějovicích. V tabulce 7 indikuje **R** výsledky při použití „šikovnější“ ruky, **L** při použití „méně šikovné“ ruky k psaní čárek a **B** indikuje dosažené počty čárek vytvořených křídou na tabuli.

Tabulka 7: Počty vytvořených čárek za časový limit 10 s v závislosti na provedení

Počet	25	32	42	45	54	54	55	62	64	66	42	43	46	48	53
Provedení	L	L	L	L	L	P	P	P	P	P	T	T	T	T	T

10. Na ekonomické fakultě cvičí statistiku 4 pedagogové: Č, B, K a R. Statistická témata jsou rozdělena do 4 částí: *Pravděpodobnost, deskriptivní metody, induktivní metody a regresní analýzy*. Jednotliví pedagogové cvičí jednotlivá témata. Mají přitom individuální pedagogické metody. Na konci každého tématu se píše test, který má 20 otázek. Správně zodpovězená otázka je za 1 bod. Aby se eliminoval nezodpovědný přístup studentů k přípravě na test, je na testy vázán zápočet: Zápočet ze statistiky dostane student, u kterého počet bodů ani u jednoho z testů neklesne pod 10 bodů. Opravné testy se nepřipouštějí.

Otestujte, zda účinnost různých pedagogických metod je stejná či nikoli (účinnost metod je měřena počtem bodů z testu – viz tabulka 8). Ze všech studentů bylo náhodně vybráno 10 studentů a u nich zjišťován počet bodů.

Lze předpokládat, že efektivnost metod všech vyučujících je stejná (volte $\alpha = 0,01$)?

11. [aneb Dcv. 1.] Úlohu 4 a 5 naleznete vzorově vyřešenou v dokumentu [nonparametrics.pdf](#), jež připravila RNDr. Marie Kletečková. Zkonfrontujte její a Vaše řešení.

Tabulka 8: Ověřování efektivnosti vyučovacích metod jednotlivých pedagogů

Student č.	Vyučující			
	C	B	K	R
1	13	15	16	12
2	15	15	15	12
3	12	15	12	12
4	17	15	16	14
5	12	12	12	12
6	19	19	19	12
7	17	13	17	13
8	11	14	10	10
9	18	18	18	18
10	17	17	17	17