

Cvičení 12. Vícevýběrové testy

Typ proměnné	Požadovaný typ analýzy	Předpoklady		Testy
Alespoň tři nezávislé spojité proměnné	Ověření shody rozptylů (homoskedasticity)	Normalita	Vyvážené třídění	Cochranův test
			Nevyvážené třídění	Hartleyův test
		—		Bartlettův test
	Ověření shody měr polohy (středních hodnot, resp. mediánů)	—		Leveneův test
		Normalita	Shoda rozptylů (homoskedasticita)	ANOVA (Analýza rozptylu = test shody stř. hodnot (Poznámka: V případě zamítnutí H_0 je vhodné provést post hoc analýzu, např. Schéffeho metodou.)
		Symetrické rozdělení		Kruskalův-Wallisův test (test shody mediánů, resp. shody distribucí; použití i pro pořadová data. Poznámka: V případě zamítnutí H_0 je vhodné provést post hoc analýzu, např. Dunnové metodou.)

Tabulka ANOVA					
Zdroj variability	Součet čtverců	Počet stupňů volnosti	Rozptyl (prům. součet čtverců)	F -poměr	p -hodnota
Skupinový faktor	$SS_B = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2$	$df_B = \underline{k - 1}$	$\underline{MS_B} = \frac{SS_B}{df_B}$	$\frac{MS_B}{MS_e}$	$1 - F_0(x_{OBS})$
Reziduální faktor	$SS_e = \sum_{i=1}^k (n_i - 1) s_i^2$	$df_e = \underline{n - k}$	$\underline{MS_e} = \frac{SS_e}{df_e}$	—	—
Celkový faktor	$SS_T = \sum_{i=1}^k (X_{ij} - \bar{X})^2$	$df_T = n - k1$	—	—	—

$$\hat{s}_i^2 = \frac{\sum (x_{ij} - \bar{x}_i)^2}{n_i - 1}$$

$$SS_T = SS_B + SS_e$$

Příklady

Příklad 1.

Testujeme nulovou hypotézu $\mu_1 = \mu_2 = \mu_3$. Bylo zjištěno, že data, která máme k dispozici jsou výběry z normálního rozdělení splňující předpoklad homoskedasticity (shody rozptylů). Na základě údajů získaných explorační analýzou doplňte tabulku ANOVA a vyplývající závěry.

Faktor	Rozsah výběru	Průměr	Výběrová směrodatná odchylka
Skupina 1	40	300	33
Skupina 2	40	290	34
Skupina 3	42	310	31
Celkem	122		—

Zdroj variability	Součet čtverců	Počet stupňů volnosti	Rozptyl (prům. součet čtverců)	<i>F-poměr</i>	<i>p-hodnota</i>
Skupinový	8146,7	2	4098,4	3,84	0,024
Reziduální	126956,0	119	1066,4		
Celkový					

$$\sum_{i=1}^n n_i \cdot \bar{X}_i = \bar{X} = 300,2$$

ANOVA

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$$H_A: \neg H_0$$

$$p\text{-hodnota} = 0,024$$

$$\alpha = 0,05$$

\Rightarrow Zamišľan H_0

\Rightarrow Existuje stat. významný rozdiel
medzi strednými hodnotami μ_1, μ_2, μ_3

Příklad 2.

122 pacientů, kteří podstoupili operaci srdce, bylo náhodně rozděleno do tří skupin.

Skupina 1: Pacienti dostali 50 % oxidu dusného a 50 % kyslíkové směsi nepřetržitě po dobu 24 hodin.

Skupina 2: Pacienti dostali 50 % oxidu dusného a 50 % kyslíkové směsi pouze během operace.

Skupina 3: Pacienti nedostali žádný oxid dusný, ale dostali 35-50 % kyslíku po dobu 24 hodin.

Data v souboru kyselina listova.xls odpovídají koncentracím soli kyseliny listové v červených krvinkách ve všech třech skupinách po uplynutí 24 hodin ventilace. Ověřte, zda pozorované rozdíly mezi koncentracemi soli kyseliny listové jsou statisticky významné, tj. zda existuje vliv složení směsi na sledovaný parametr.

OP? - nejsou

$\mu / \chi_{0,5}$ ve skupinách

ANOVA

/ KX test

NORM. ✓

SYMMETRIE

HODNOSK. ✓

SW-TEST NORMALITY

SK1: p-hod = 0,477

SK2: p-hod = 0,771

SK3: p-hod = 0,525

} NEZAMÍTAJÍ
NORMALITU

Bartlett t.

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

$$H_A: \neg H_0$$

$$p\text{-hod} = 0,645 \Rightarrow \text{NEZAMÍTÁM} \\ \text{HOMOSKEDASTICITU}$$

ANALÝZA

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$$H_A: \neg H_0$$

$$p\text{-hod} < 0,001$$

$$\Rightarrow \text{ZAMÍTÁM } H_0$$

Existujú stat. významné rozdiely
mezi stí, bod.

⇒ POST-HOC

TUKEY HSD

$\alpha = 0,05$

	Sk1	Sk3	Sk2
Sk1	X	0,002 X	0,001 X
Sk3		X	0,003 X
Sk2			X

EFEKT		HM1	HM2	HM3
26,7	Sk1	a		
-0,2	Sk3		b	
-26,5	Sk2			c

Příklad 3.

Na farmě jsou chována tři plemena králíků. Byl proveden pokus kralici.xls, jehož cílem bylo zjistit, zda i když chováme a vykrmujeme všechny králíky po stejnou dobu a za stejných podmínek, existuje statisticky významný (průkazný) rozdíl mezi plemeny v hmotnostech králíků. Ověřte.

Q P ? - 1 OP \rightarrow odstraněn
(38,1)

$\mu / x_{0,5}$ vah králíků

ANOVA

1 k u test

NORM. ✓

SYM.

HOMOS. ✓

NORM - (S u t) $\alpha = 0,05$

VIDEL : p-hod = 0,825

CESTK : p-hod = 0,278

KALIF : p-hod = 0,169

\Rightarrow NEZAMÍTAJN N NORMALITV

$$H_0: \sigma_v^2 = \sigma_c^2 = \sigma_k^2 \quad (\text{Bart. test})$$

$$H_A: \neg H_0$$

$$p\text{-hod} = 0,217$$

\Rightarrow NEZAMÍTAJN $H_0 \rightarrow$ NEZ. HOMOS.

ANOVA

$$H_0: \mu_v = \mu_c = \mu_k$$

$$H_A: \neg H_0$$

$$p\text{-hod} < 0,001$$

\Rightarrow POST-HOC

(TUKEY)

	KALIF	VID	CEK
KAL	X	0,007 X	X 0,007
VID		X	X 0,007
CEK			X
$\alpha \approx 0,25$			

EFFET			
0,88	KAL	a	
-0,08	VID		b
-0,64	CEK		c

Příklad 4.

Soutěž o nejlepší jakost výrobků obeslali čtyři výrobci A, B, C, D celkem 66 výrobky. Porota sestavila pořadí (uvedeno pouze pořadí výrobku od nejlepšího k nejhoršímu), jež je uvedené v souboru jakost.xls. Na základě uvedených údajů posuďte, zda původ výrobků má vliv na jeho jakost.

→ Spojená?

NE!

→ NEzávislá?

NE!

→ NEJROBUSTNĚJŠÍ TEST

KW-TEST

(π NEST.
ZAV)

SYMETRIE

A tibble: 4 × 3

skupina	sikmost	test.pval
<chr>	<dbl>	<dbl>
1 A	-0.211238463	0.8295537
2 B	0.459593402	0.5207930
3 C	-0.147647782	0.5681381
4 D	0.009223978	0.8996529

⇒ NEZAMÍTÁM SYMETRII

KU - t.

$$H_0: X_{0,5;A} = X_{0,5;B} = X_{0,5;C} = X_{0,5;D}$$

$$H_A: \neg H_0$$

$$p\text{-hod} = 0,1295$$

\Rightarrow NE zamietam H_0

\rightarrow neexistuje stat. významný rozdiel.

(niektoré predpok. nebyli
splnené \Rightarrow test nemusí
byť spoľahlivý)

Příklad 5.

Byl sledován vliv tří preparátů na srážlivost krve. Kromě jiných ukazatelů byl zjišťován tzv. trombinový čas. Údaje o 42 sledovaných osobách jsou zaznamenány v souboru trombin.xls. Záviseí velikost trombinového času na tom, jaký byl použit preparát?

OP 2. neyman

$\mu / \sigma_{0,5}$ ve stupních

ANOVA

1

K-U. test

Norm. ~~X~~

Symet. ✓

Homosk.

Norm (S-U)

p-hodnoty

	skupina	norm.pval	
	<chr>	<dbl>	
1	A	0.03179805	X
2	B	0.94597139	
3	C	0.27138568	

$$\alpha = 0,05$$

Zamínám normalizaci pro A

SYMMETRIE? ✓ (Symmetry test)

skupina	sikmost	test.pval
<chr>	<dbl>	<dbl>
1 A	0.5400617	1.0000000 ✓
2 B	0.2886751	0.7777511 ✓
3 C	0.2975920	0.8792017 ✓

$$\alpha = 0,05$$

⇒ NEZAMÍTÁM SYMMETRII

K-W test

$$H_0: X_{0,r;A} = X_{0,r;B} = X_{0,r;C}$$

$$H_A: \neg H_0$$

$$p\text{-hod} < 0,001$$

$$\alpha = 0,05$$

⇒ ZAMÍTÁM H_0

→ Existují stat. výs. rozdíly mezi

mediány Árendiných úsů

→ POST-HOC (Dun-test)
Bonferroni

$$\alpha = 0,05$$

	C	B	A
C	X	0,022 _X	0,001 _X
B		X	0,004 _X
A			X

EFET			
8,00	C	a	
0,00	B		b
-6,00	A		c

PŘÍKLAD 6.

OP? NESOL

ANOVA

NORM ✓

HOMOGEN. ✓

1

kw

SYMETRIE

S-W test NORMALITY

$$\alpha = 0,05$$

	typ	p.hodnota
<chr>		<dbl>
1.	Dřímál	0.8295308
2	Kejchal	0.8162545
3	Profa	0.8265843
4	Rejpal	0.6555333
5	Stydlín	0.9177484
6	Šmudla	0.1177217
7	Štístko	0.1866139

⇒ NEZAMÍTÁME NORMALITU

$$H_0: \sigma_1^2 = \dots = \sigma_7^2$$

(Bartlett)

$$H_A: \neg H_0$$

$$p\text{-hodnota} = 0,522$$

\Rightarrow NEZAMÍTAME HPMOSK.

ANOVA

$$H_0: \mu_1 = \dots = \mu_7$$

$$H_A: \neg H_0$$

$$p\text{-hodnota} \ll 0,001$$

\Rightarrow ZAMÍTAME H_0

\rightarrow existujú stat. významné rozdiely
medzi sled. hodnotami.

\Rightarrow POST-HOC (TUKEY HSD)

	$\check{S}M$	KE	RE	$\check{S}T$	ST	$D\check{R}$	PR
$\check{S}M$	X	0,001	0,001	0,001	0,001	0,001	0,001
KE		X	0,998	0,773	0,001	0,001	0,001
RE			X	0,944	0,006	0,001	0,001
$\check{S}T$				X	0,125	0,013	0,001
ST					X	0,985	0,001
$D\check{R}$						X	0,005
PR							X

0,84	\tilde{SM}	a							
0,27	KE		b						
0,20	RE		b	c					
0,08	\tilde{ST}		b	c	d				
-0,24	ST				d	e			
-0,34	\tilde{PR}					e	b		
-0,80	PR							g	

