4.5 Vzťahy medzi premennými

4.5.1 Analýza kontingencie

Tieto testy zisťujú vzťahy medzi dvoma nominálnymi premennými, t.j. používajú sa k analýze závislostí nominálnych premenných. Patrí sem skupina neparametrických testov, ktoré vychádzajú z kontingenčnej tabuľky (Tabuľka 24). Tieto testy overujú nulovú hypotézu, ktorá tvrdí, že premenné sú nezávislé. Jedna skupina testov je určená iba pre štvorpoľné kontingenčné tabuľky (2x2), v ktorých vystupujú dve dichotomické premenné.

Tabuľka 24 Kontingenčná tabuľka pozorovaných početností RxS

X\Y	y_1	y_2	...	y_S	**Σ**
x_1	a_{11}	a_{12}	...	a_{1S}	r_1
x_2	a_{21}	a_{22}	...	a_{2S}	r_2
...	...	...	...	...	...
x_R	a_{R1}	a_{R2}	...	a_{RS}	r_R
Σ	s_1	s_2	...	s_S	n

Chí-kvadrát test nezávislosti môžeme použiť k analýze viacpoľných tabuliek. Chí-kvadrát test nezávislosti predstavuje rozšírenie chí-kvadrát testu dobrej zhody a vychádza z kontingenčnej tabuľky (Tabuľka 24) pozorovaných početností, kde pozorovaná početnosť a_{ij} je početnosť (frekvencia) kombinácie $x_i \wedge y_j$.

Očakávané početnosti e_{ij} sú také, ktoré zodpovedajú nulovej hypotéze o nezávislosti dvoch premenných. Očakávaná početnosť príslušnej bunky sa rovná podielu súčinu príslušnej pozorovanej početnosti riadku a stĺpca a celkovému počtu pozorovaní.

$$e_{ij} = \frac{r_i s_j}{n}$$

Chí-kvadrát test overuje, či môžu byť rozdiely skutočných a očakávaných početností iba náhodné (premenné sú nezávislé) alebo štatisticky významné (premenné sú závislé) (Tabuľka 25).

Chí-kvadrát test môžeme použiť iba v prípade, že očakávané početnosti sú dostatočne veľké.

$$e_{ij} = \frac{r_i s_j}{n} \geq 5$$

Taktiež výsledky z kontingenčných tabuliek, ktoré obsahujú nulové početnosti, treba brať s rezervou.

V prípade, že očakávané početnosti nebudú dostatočne veľké, môžeme použiť **Fisherov test**, ktorý je však použiteľný iba pre štvorpoľné tabuľky (Tabuľka 25).

Tabuľka 25 Procedúra testovania - chí-kvadrát test nezávislosti/Fisherov test

	Chí-kvadrát test nezávislosti	Fisherov test
1.	H0: $P(X \wedge Y) = P(X)P(Y)$ proti alternatíve *X, Y* sú závislé	
2.	$\alpha = 0{,}05 \vee \alpha = 0{,}01 \vee \ldots$	
3.	$\chi^2 = \sum_{i=1}^{R}\sum_{j=1}^{S} \frac{(a_{ij} - e_{ij})^2}{e_{ij}}$	$P = \sum_{i=0}^{\min(r_1, s_1) - a_{11}} \frac{r_1! r_2! s_1! s_2!}{n!(a_{11}+i)!(a_{12}-i)!(a_{21}-i)!(a_{22}+i)!}$
4.	$\chi^2 \geq \chi^2_{(R-1)(S-1)}(\alpha) \vee p \leq \alpha$	$P \leq \alpha$

Kontingenčné koeficienty (Pearsonov, Cramerov V) predstavujú mieru vzťahu medzi dvoma nominálnymi premennými. Nadobúdajú hodnoty z intervalu 0 (žiadny vzťah) až 1 (dokonalý vzťah). Chí-kvadrát testom nezávislosti môžeme testovať **významnosť kontingenčných koeficientov**.

Ďalej ako mieru vzťahu dvoch nominálnych premenných môžeme použiť symetrické korelačné koeficienty **Spearmanov, Kendallov, koeficient gama** a asymetrický koeficient **Sommerovo D**, ktorý je rozšírením gama koeficientu. Asymetrický koeficient rozlišuje závislú a nezávislú premennú na rozdiel od predchádzajúcich koeficientov. Štatistický softvér väčšinou ponúkne k výpočtu obidve možnosti: *D(X|Y), D(Y|X)*.

Fí-kvadrát môžeme použiť ako mieru vzťahu iba v štvorpoľných kontingenčných tabuľkách (2x2), t.j. medzi dvoma dichotomickými premennými.

Všetky miery môžeme interpretovať rovnako ako korelačné koeficienty (Tabuľka 26).

4.5.2 Korelačná analýza

Korelačná analýza skúma tesnosť štatistickej závislosti medzi kvantitatívnymi premennými. Korelačná analýza na rozdiel od regresie nevyjadruje príčinno-následný vzťah. Premenná *Y* nezávisí na premennej *X*, ale dve náhodné premenné *X* a *Y* sa spoločne menia. Regresná analýza predpokladá, že premenná *Y* je náhodná a premenná *X* fixná. Korelačný koeficient je mierou lineárnej závislosti dvoch premenných.