χ2 – test o nezávislosti v kontingenční tabulce

= test nezávislosti kategoriálních znaků.

- **Kontingenční tabulka** – dvourozměrná tabulka, do které jsou uspořádány údaje o dvou slovních proměnných (příp. jedné proměnné slovní a druhé číselné).

Kontingenční tabulka:

a_i \ b_j	b_1	b_2	....	b_s	Součty četností $n_{i\bullet}$
a_1	n_{11}	n_{12}	....	n_{1s}	$n_{1\bullet}$
a_2	n_{21}	n_{22}	....	n_{2s}	$n_{2\bullet}$
⋮	⋮	⋮	⋮	⋮	⋮
a_r	n_{r1}	n_{r2}	....	n_{rs}	$n_{r\bullet}$
Součty četností $n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	....	$n_{\bullet s}$	n

- Četnosti n_{ij} uvnitř tabulky se nazývají *sdružené (simultánní) četnosti*.
- Na okrajích tabulky najdeme *okrajové (marginální) četnosti* $n_{i\bullet}$ a $n_{\bullet j}$, kde tečkou je naznačeno sčítání.
- Platí, že:

$$n = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} \qquad \text{a} \qquad n = \sum_{i=1}^{r} n_{i\bullet} = \sum_{j=1}^{s} n_{\bullet j}\ .$$

- Podstatou χ^2 – testu nezávislosti je porovnání empirických četností s teoretickými četnostmi, které bychom očekávali v případě nezávislosti.
- Teoretické četnosti značíme n'_{ij} a vypočítáme je podle :

$$n'_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}\ .$$

- **Předpoklad testu:** χ^2 – test nezávislosti lze použít tehdy, jsou-li všechna políčka kontingenční tabulky dostatečně obsazena, tj. platí-li $n'_{ij} \geq 5$. Pokud tato podmínka splněna není, musíme některé třídy sloučit nebo zvětšit rozsah výběru.

- **Test:**

1) H_0 : znaky *a* a *b* jsou nezávislé
   H_1 : znaky *a* a *b* jsou závislé (non H_0)

2) **Testové kritérium:**

$$G=\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(n_{ij}-n'_{ij}\right)^2}{n'_{ij}}\ ,\qquad G\in\left\langle 0,n\cdot h\right\rangle,\ \text{h ... min (r – 1), (s – 1).}$$

G má při platnosti H_0 rozdělení χ^2 s $\nu=(r-1)\cdot(s-1)$ stupni volnosti.

3) **Kritický obor:**

$$W\equiv\left\{G,G>\chi^2_{1-\alpha}\left[(r-1)(s-1)\right]\right\}$$

- Pokud jsme testem prokázali závislost, má smysl posuzovat její sílu – to lze pomocí kontingenčních koeficientů:

  Cramérův koeficient kontingence

  $$C_C=\sqrt{\frac{G}{n\cdot h}}\qquad C_C\in\left\langle 0,1\right\rangle$$

  Pearsonův koeficient kontingence

  $$C_P=\sqrt{\frac{G}{G+n}}\qquad C_P\in\left\langle 0,1\right)$$

- **Speciálním případem** kontingenční tabulky je tabulka ***ASOCIAČNÍ*** – čtyřpolní dvourozměrná tabulka, která obsahuje 2 slovní proměnné, které nabývají pouze 2 hodnot (alternativní znaky).

A_i \ B_j	B_1	B_2	$n_{i\bullet}$
A_1	n_{11}	n_{12}	$n_{1\bullet}$
A_2	n_{21}	n_{22}	$n_{2\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	n

1) H_0 : znaky *A* a *B* jsou nezávislé
   H_1 : znaky *A* a *B* jsou závislé (non H_0)

2) $G=n\cdot\frac{\left(n_{11}n_{22}-n_{12}n_{21}\right)^2}{n_{\bullet 1}n_{\bullet 2}n_{1\bullet}n_{2\bullet}}$

   G má při platnosti H_0 rozdělení $\chi^2(1)$.

3) $W\equiv\left\{G,G>\chi^2_{1-\alpha}(1)\right\}$

- K měření těsnosti závislosti jevů *A* a *B* se používá **koeficient asociace $\mathbf{r_{AB}}$** :

  $$r_{AB}=\frac{n_{11}n_{22}-n_{12}n_{21}}{\sqrt{n_{1\bullet}n_{2\bullet}n_{\bullet 1}n_{\bullet 2}}}\ ,\qquad r_{AB}\in\left\langle -1,1\right\rangle$$