

## Kontingenční tabulka

Menu:	QCExpert	Testování	Kontingenční tabulka
-------	----------	-----------	----------------------

Tento modul testuje nezávislost dvou diskrétních číselných nebo nečíselných proměnných  $A$ ,  $B$  na základě experimentálně zjištěných výskytů jednotlivých kombinací úrovní těchto proměnných. Počet úrovní proměnné  $A$  je označen  $r$ , počet úrovní proměnné  $B$  je označen  $c$ . Základem je tabulka četností, zadat je třeba pouze tučně ohraničené jednotlivé četnosti, nikoli celkové součty.

	$B$			
$A$	$B_1$	...	$B_c$	Celkem
$A_1$	$n_{11}$	...	$n_{1c}$	$n_{1.}$
...	...	...	...	...
$A_r$	$n_{r1}$		$n_{rc}$	$n_{r.}$
Celkem	$n_{.1}$	...	$n_{.c}$	$n$

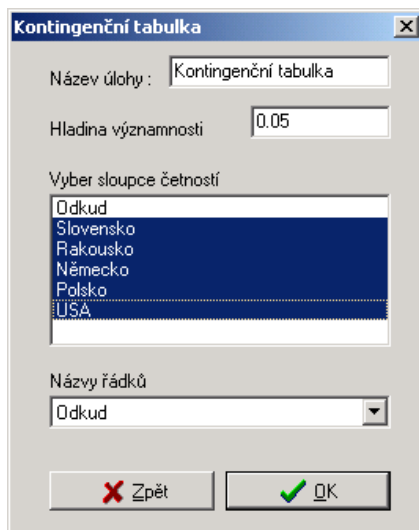
Z této tabulky se vypočítá testovací statistika  $C$ , která má  $\chi^2$  rozdělení:

$$C = n \sum_{i=1}^r \sum_{j=1}^c \frac{n_{ij}^2}{n_{i.} n_{.j}} - n$$

Testovací statistika  $C$  se porovná s kritickou hodnotou  $\chi^2$  rozdělení  $\chi^2_{(r-1),(c-1)}(1-\alpha)$ . Je-li  $C > \chi^2$ , hypotéza  $H_0$  o nezávislosti  $A$  a  $B$  se zamítá. Závislost  $A$  a  $B$  znamená, že pravděpodobnosti a tím i počty výskytů ve sloupcích pro aspoň jednu úroveň (hodnotu) proměnné  $A$  jsou ovlivněny úrovněmi proměnné  $B$ . Jsou-li  $A$  a  $B$  nezávislé, pak je pravděpodobnost  $p_{ij}$  jevu  $A_j \wedge B_i$  rovna součinu marginálních pravděpodobností  $p_i, p_j$ . Tyto pravděpodobnosti se dají vypočítat ze součtů četností v řádcích a sloupcích,  $p_i = n_{i.}/n, p_j = n_{.j}/n$ . Platí, že  $\sum p_i = \sum p_j = 1$ .

### Data a parametry

Data ve formě četností z tučně ohraničené části výše uvedené tabulky musí být v datové tabulce. Názvy sloupců jsou v záhlaví tabulky, názvy řádků mohou být v kterémkoliv (obvykle prvním) sloupci datové tabulky. V dialogovém okně se vyberou sloupce četností a sloupce s názvy řádků. V poli *Hladina významnosti* se zadá požadovaná hladina významnosti testu  $\alpha$ . V poli *Vyber sloupce* četností se vyberou sloupce, v nichž jsou četnosti  $n_{ij}$  jednotlivých kombinací úrovní proměnných  $A$  a  $B$ . Doporučená minimální četnost je  $n_{ij} = 5$ , aby byla zajištěna přesnost testu. Kombinací myši a klávesy *Ctrl* lze vybrat libovolné sloupce. Názvy sloupců, tedy úrovní proměnné  $A$  se očekávají v záhlaví sloupců. Jsou-li ve sloupci datové tabulce názvy řádků (tedy úrovní proměnné  $B$ ), vybere se tento sloupec v poli *Názvy řádků*. Stiskem tlačítka OK se provede test a výsledky se zapíše do protokolu. Po výpočtu se dialogové okno zavře.



Obrázek 1 Dialogový panel pro kontingenční tabulku

### Příklad

Na základě regionálních ekonomických dat o exportu do jednotlivých zemí máme rozhodnout, zda poloha regionu (proměnná  $B$ ) má vliv na objem exportu do dané země (proměnná  $A$ ) za určité období. Tedy například zda Ostravský kraj vyváží větší procento exportu do Polska, než třeba Plzeňský, atd. Objemy exportu jednotlivých kombinací regionu a země lze považovat za četnosti. Tyto objemy spolu s názvy regionů v prvním sloupci a názvy zemí v záhlaví se zapíší do datové tabulky následujícím způsobem.

<i>Odkud</i>	<i>Slovensko</i>	<i>Rakousko</i>	<i>Německo</i>	<i>Polsko</i>	<i>USA</i>
<i>Ostrava</i>	350	216	189	626	46
<i>Brno</i>	387	489	274	126	115
<i>Plzeň</i>	52	83	264	132	51
<i>Praha</i>	484	594	737	447	141

V dialogovém okně (Obrázek 1) se zadá hladina významnosti 0.05, vyberou se sloupce četností *Slovensko*, *Rakousko*, *Německo*, *Polsko*, *USA* a názvy řádků ve sloupci *Odkud*. Stiskem tlačítka OK provedeme test, který prokáže na hladině významnosti, že poloha regionů má vliv na objem exportu do sledovaných zemí. Chí-kvadrát statistika je 884, kritická hodnota je 21, nezávislost proměnných se zamítá. To znamená, že struktura cílových zemí exportu je ovlivněna lokalitou exportéra.

### Protokol

Analyza kontingenční tabulky	Název modulu.
Název úlohy	Název úlohy z dialogového okna.
Tabulka počtů	
<i>Název řádku</i>	Zadané četnosti z datové tabulky, v posledním sloupci jsou řádkové marginální četnosti, tedy součty řádků.
Teoretické:	V závorkách jsou teoretické četnosti, které by vyplývaly z nezávislosti proměnných.
Celkem	Tento řádek obsahuje součty četností ve sloupcích, tedy sloupcové marginální četnosti.
Tabulka poměrů a	

pravděpodobností	
<i>Název řádku</i>	Empirické pravděpodobnosti vypočítané ze zadaných četností z datové tabulky jako podíl $n_{ij}/n$ , v posledním sloupci jsou řádkové marginální pravděpodobnosti, tedy součty řádků.
Teoretické:	Teoretické pravděpodobnosti vypočítané za předpokladu nezávislosti <i>proměnných</i> jako podíl $n_{i.} \cdot n_{.j}/n^2$ , v posledním sloupci jsou opět řádkové marginální pravděpodobnosti.
Celkem	Tento řádek obsahuje součty empirických pravděpodobností ve sloupcích, tedy sloupcové marginální pravděpodobnosti.
Závěr	Slovní formulace závěru testu.
Hladina významnosti	Zadaná hladina významnosti, obvykle se pracuje na hladině 0.05.
Stupně volnosti	Počet stupňů volnosti $(r - 1)(c - 1)$ .
Chi2 statistika	Testovací statistika vypočítaná z dat.
Kritická hodnota	Kritická hodnota (kvantil) $\chi^2$ rozdělení $\chi^2_{(r-1),(c-1)}(1-\alpha)$ .
p-hodnota	Vypočítaná <i>p</i> -hodnota testu.