

TEST χ^2 DOBRÉ SHODY

Sledujeme kategoriální veličinu X (např.):

- pohlaví (zastoupení samců a samic);
- kvalita výrobku (I.jakost, II.jakost, zmetek);

Chceme prokázat:

- jsou obě pohlaví zastoupena rovnoměrně, tedy v poměru 1:1 (50:50 %)?
- jsou výrobky dle jakosti zastoupeny v poměru 3:1:1 (60:20:20 %)?

TEST χ^2 DOBRÉ SHODY

Testovaná dvojice hypotéz

- obecně pro veličinu X s r -kategoriemi:

■ $H_0: P(x_1) = \pi_1 ; P(x_2) = \pi_2 ; \dots ; P(x_r) = \pi_r$

■ $H_1: \text{non } H_0$

kde $\pi_1, \dots, \pi_r$ jsou konkr. čísla: $\pi_1 + \dots + \pi_r = 1$

TEST χ^2 DOBRÉ SHODY

Testovaná dvojice hypotéz konkr.:

- u příkladu samci vs samice:

■ $H_0: P(x_1) = 0,5 ; P(x_2) = 0,5$

- u příkladu s jakostí:

■ $H_0: P(x_1) = 0,6 ; P(x_2) = 0,2 ; P(x_3) = 0,2$

TEST χ^2 DOBRÉ SHODY

Z dat určíme absolutní, tzv.

pozorované četnosti $n_1; n_2; \dots; n_r$

přičemž $n_1 + \dots + n_r = n$

Pro jednotlivé kategorie spočteme tzv.

očekávané četnosti $o_1; o_2; \dots; o_r$

a to podle vzorce:

$$o_i = n \cdot \pi_i \quad (i=1, \dots, r)$$

TEST χ^2 DOBRÉ SHODY

***Např. necht' při kontrole jakosti bylo
110 výrobků I. jakosti ($n_1 = 110$),
56 výrobků II. jakosti ($n_2 = 56$)
34 zmetků ($n_3 = 34$), tj. $n = 200$;***

***při testu, zda $\pi_1 = 0,6$; $\pi_2 = 0,2$; $\pi_3 = 0,2$
dostaneme jaké očekávané četnosti?***

$$o_1 = n \cdot \pi_1 = 200 \cdot 0,6 = 120$$

$$o_2 = n \cdot \pi_2 = 200 \cdot 0,2 = 40$$

$$o_3 = n \cdot \pi_3 = 200 \cdot 0,2 = 40$$

TEST χ^2 DOBRÉ SHODY

Podstatou testové statistiky je porovnání četností pozorovaných s očekávanými:

$$T = \sum (n_i - o_i)^2 / o_i \quad (i=1 \dots r)$$

Př. (pokrač.):

$$\begin{aligned} T &= (110-120)^2/120 + \\ &+ (56-40)^2/40 + \\ &+ (34-40)^2/40 = \\ &= 0,83 + 6,40 + 0,90 = 8,13 \end{aligned}$$

TEST χ^2 DOBRÉ SHODY

Řešení pomocí Excelu:

	A	B	C	D
1		π_i	n_i	o_i
2		0,6	110	120
3		0,2	56	40
4		0,2	34	40
5	suma	1	200	200
6				
7		=chitest(C2:C4;D2:D4)		
8		CHITEST(aktuální; očekávané)		

p=0,017

...tj.?

TEST χ^2 DOBRÉ SHODY

Poznámka:

V principu lze takto testovat i normalitu;

Sledovanou číselnou veličinu „rozsekáme“ intervalovým rozdělením do kategorií a porovnáme pozorované četnosti v těchto kategoriích s očekávanými, které by odpovídaly pravděpodobnostem dle „gaussovského“ modelu

TEST χ^2 NEZÁVISLOSTI

Sledujeme dvojici kategoriálních veličin X, Y

- např. u každého respondenta jeho pohlaví (M-Ž) a krevní skupinu (A-B-AB-0);
- nebo u každého výrobku jeho kvalitu (I.jakost, II.jakost, zmetek) a to, během jaké směny vznikl (dopolední – odpolední - noční směna);

TEST χ^2 NEZÁVISLOSTI

Chceme prokázat:

- **závisí nebo nezávisí krevní skupina na pohlaví?**

(ve smyslu, zda jsou nebo nejsou mezi muži a ženami významné rozdíly v zastoupení jednotlivých krevních skupin)

TEST χ^2 NEZÁVISLOSTI

Resp. (příklad s výrobky) chceme prokázat:

- **závisí nebo nezávisí kvalita výrobku na tom, během jaké směny vznikl?**

(ve smyslu, zda jsou nebo nejsou mezi jednotlivými směnami významné rozdíly v zastoupení jednotlivých kvalitativních kategorií)

TEST χ^2 NEZÁVISLOSTI

Testovaná dvojice hypotéz:

H_0 : nezávislost (mezi X a Y)

H_1 : non H_0 (tj. závislost mezi X a Y)

TEST χ^2 NEZÁVISLOSTI

Data:

	A	B	C
1	výrobek č.	jakost	směna
2	1	I	d
3	2	I	n
4	3	z	d
5	4	II	o
6	5	I	o
7		...	
81	80	II	n

TEST χ^2 NEZÁVISLOSTI

***Data přehledně – kontingenční tabulka
pozorovaných absolutních četností:***

Pozorované četnosti	$Y = y_1$	$Y = y_2$	...	$Y = y_s$	suma
$X = x_1$	n_{11}	n_{12}	...	n_{1s}	$n_{1\bullet}$
$X = x_2$	n_{21}	n_{22}	...	n_{2s}	$n_{2\bullet}$
...	...	...	...	...	...
$X = x_r$	n_{r1}	n_{r2}	...	n_{rs}	$n_{r\bullet}$
suma	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet s}$	n

r = počet „řádkových“ kategorií

s = počet „sloupcových“ kategorií

VSUVKA: KONTINGENČNÍ TABULKA (vytvoření)

2.

3.

Kontingenční tabulka

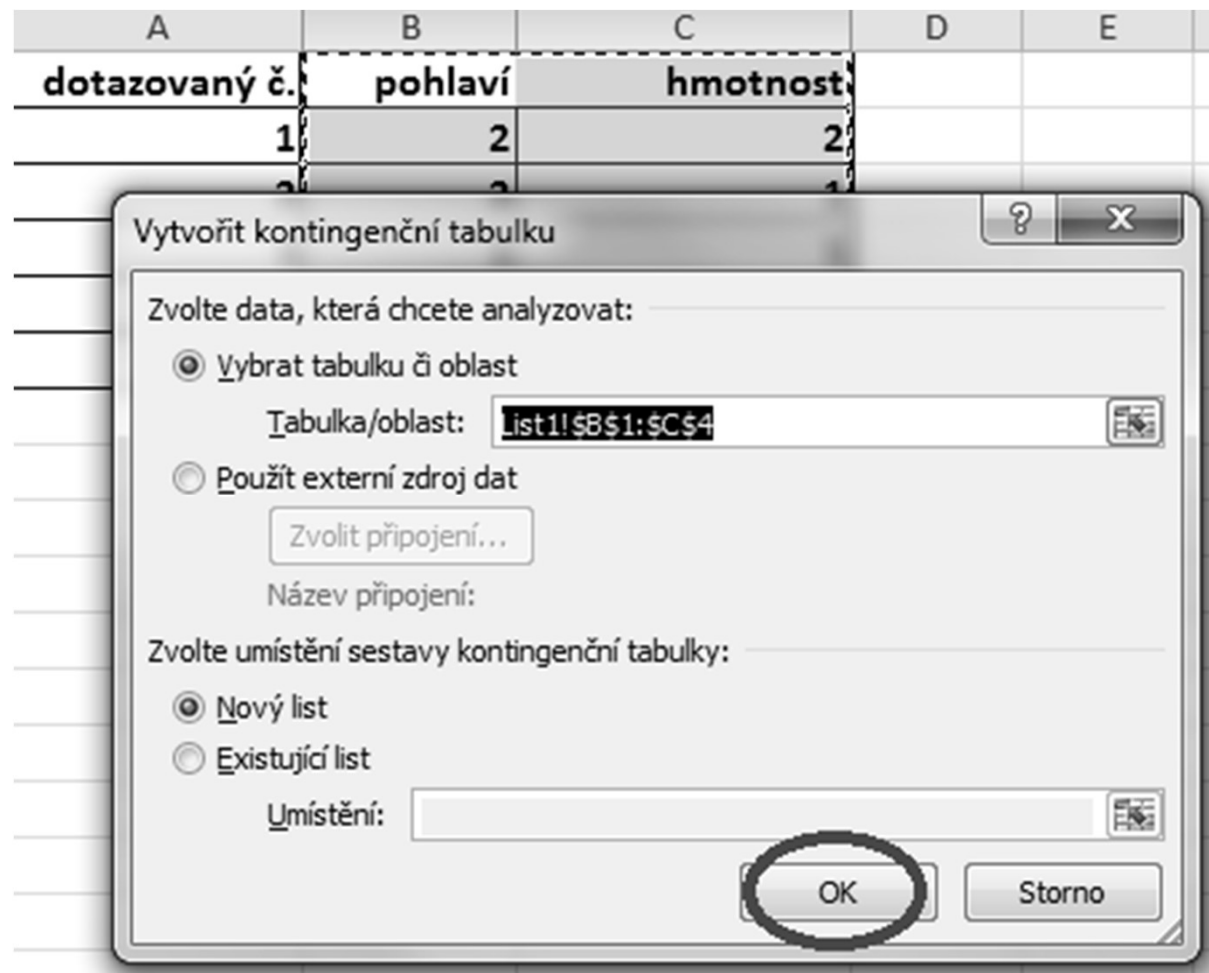
1.

dotazovaný č.	pohlaví	hmotnost
1	2	2
2	2	1
3	1	1
...	...	...
200	1	2

1=chlapec
2=dívka

1=v normě
2=nadváha

VSUVKA: KONTINGENČNÍ TABULKA (vytvoření)



Charakteristiky kategoriálních veličin

Seznam polí kontingenční tabulky

Zvolte pole, které chcete přidat do sestavy:

- ☒ pohlaví
- ☒ hmotnost

tažením myší

Přetáhněte pole do jedné z následujících oblastí:

1. **Filtr sestavy**

2. **Popisky sloupců**

3. **Popisky řádků**

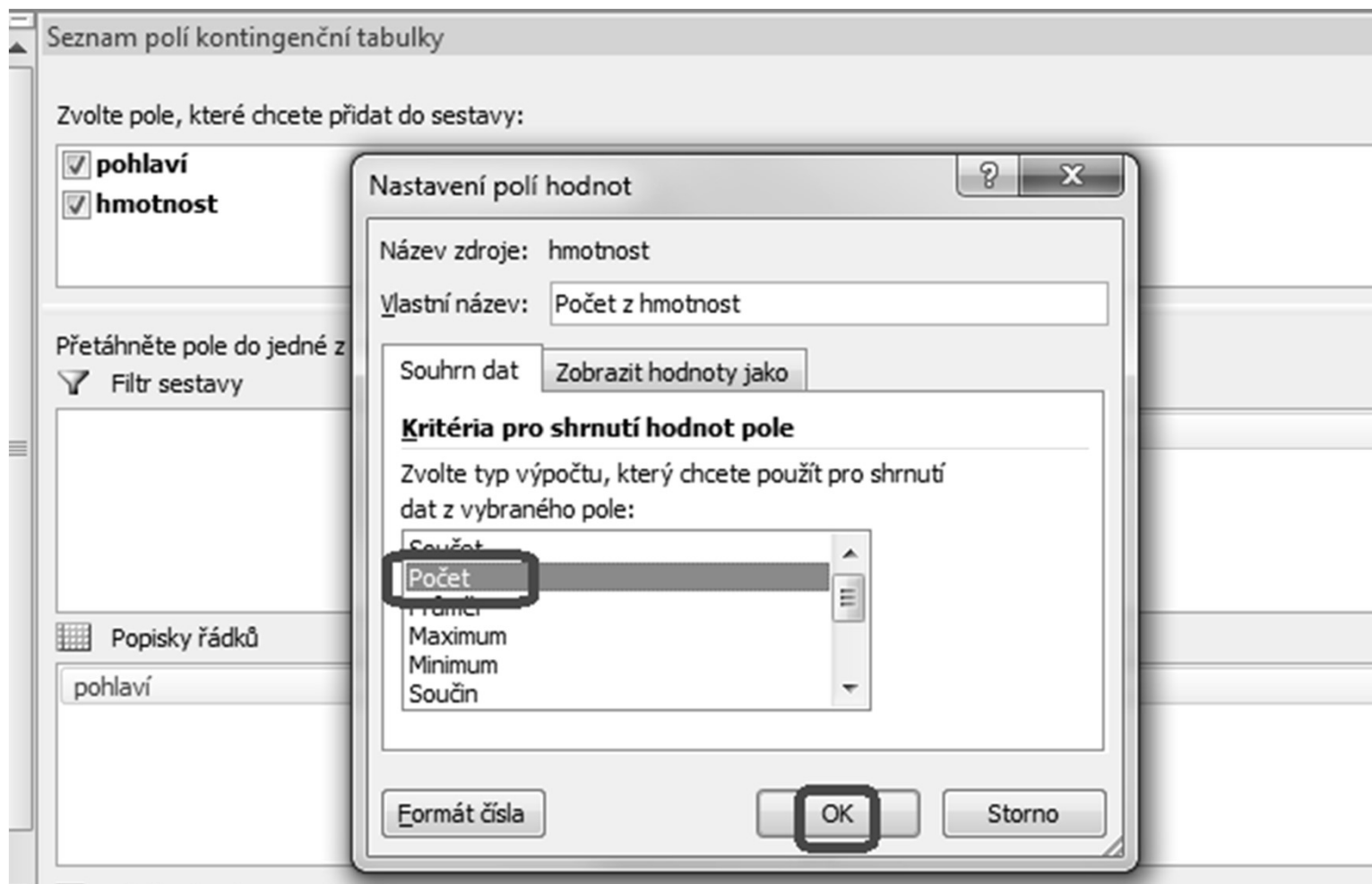
Hodnoty

Součet z hmotnost

- Přesunout nahoru
- Přesunout dolů
- Přesunout na začátek
- Přesunout na konec
- Přejít k filtru sestavy
- Přejít k popiskům řádků
- Přejít k popiskům sloupců
- Σ Přejít k hodnotám
- ✕ Odstranit pole
- Nastavení polí hodnot...**

Aktualizovat

VSUVKA: KONTINGENČNÍ TABULKA (vytvoření)



TEST χ^2 NEZÁVISLOSTI

Kontingenční tabulka - příklad:

	dopol	odpol	noc	suma
I. jakost	12	15	11	38
II. jakost	7	9	11	27
zmetky	4	5	6	15
suma	23	29	28	80

např. $n_{12} = 15$ $n_{21} = 7$
 $n_{1\cdot} = 38$ $n_{\cdot 1} = 23$

TEST χ^2 NEZÁVISLOSTI

Připravíme tabulku očekávaných četností:

Očekávané četnosti	$Y = y_1$	$Y = y_2$	...	$Y = y_s$	suma
$X = x_1$	o_{11}	o_{12}	...	o_{1s}	$n_{1\bullet}$
$X = x_2$	o_{21}	o_{22}	...	o_{2s}	$n_{2\bullet}$
...	...	...	...	...	...
$X = x_r$	o_{r1}	o_{r2}	...	o_{rs}	$n_{r\bullet}$
suma	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet s}$	n

$$o_{ij} = n_{i\bullet} \cdot n_{\bullet j} / n$$

$$\text{např. } o_{12} = n_{1\bullet} \cdot n_{\bullet 2} / n$$

TEST χ^2 NEZÁVISLOSTI

Očekávané četnosti – příklad (pokrač.):

	dopol	odpol	noc	suma
I. jakost	10,9	13,8	13,3	38
II. jakost	7,8	9,8	9,5	27
zmetky	4,3	5,4	5,3	15
suma	23	29	28	80

např. $o_{12} = n_{1.} \cdot n_{.2} / n = 38 \cdot 29 / 80 = 13,8$

! součty stejné jako původně (až na zaokr.)!

TEST χ^2 NEZÁVISLOSTI

Řešení pomocí Excelu:

	A	B	C	D	E
1		dopol	odpol	noc	suma
2	I. jakost	12	15	11	38
3	II. jakost	7	9	11	27
4	zmetky	4	5	6	15
5	suma	23	29	28	80
6					
7		dopol	odpol	noc	suma
8	I. jakost	10,9250	13,7750	13,3000	38
9	II. jakost	7,7625	9,7875	9,4500	27
10	zmetky	4,3125	5,4375	5,2500	15
11	suma	23	29	28	80
12					
13		<u>=chitest(B2:D4;B8:D10)</u>			

p=0,883...tj.?

TEST χ^2 NEZÁVISLOSTI

Podmínka použití:

- *pozor – u obou typů testu (dobré shody i nezávislosti) musí být všechny kategorie dostatečně zastoupeny, aneb všechny očekávané četnosti mají být aspoň 5;*
- *není-li splněno, doporučuje se sloučit některé (obvykle sousední) kategorie*