

# 7. Kontingenční tabulky a $\chi^2$ test



Znaménkový test a  $\chi^2$  test  
Shrnutí statistických testů  
Kontingenční tabulky

# Statistické testování – základní pojmy



➤ **Nulová hypotéza  $H_0$**

$H_0$ : sledovaný efekt je nulový

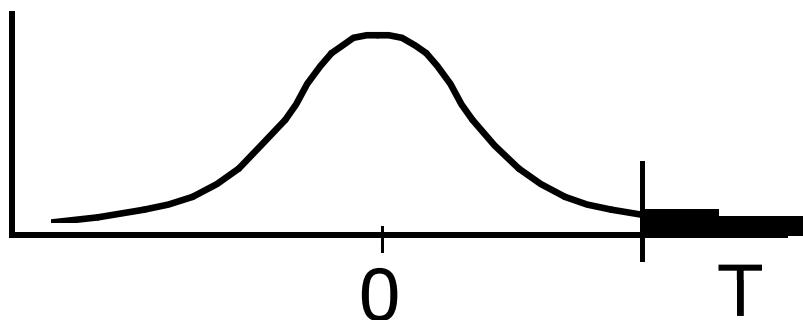
➤ **Alternativní hypotéza  $H_A$**

$H_A$ : sledovaný efekt je různý mezi skupinami

➤ **Testová statistika**

$$\text{Testová statistika} = \frac{\text{Pozorovaná hodnota} - \text{Očekávaná hodnota}}{\text{Variabilita dat}} * \sqrt{\text{Velikost vzorku}}$$

➤ **Kritický obor testové statistiky**



**Statistické testování odpovídá na otázku zda je pozorovaný rozdíl náhodný či nikoliv. K odpovědi na otázku je využít statistický model – testová statistika.**

# P-hodnota



Významnost hypotézy hodnotíme dle získané tzv. p-hodnoty, která vyjadřuje pravděpodobnost, s jakou číselné realizace výběru podporují  $H_0$ , je-li pravdivá.

P-hodnotu porovnáme s  $\alpha$  (hladina významnosti, stanovujeme ji na **0,05**, tzn., že připouštíme 5% chybu testu, tedy, že zamítneme  $H_0$ , ačkoliv ve skutečnosti platí).

P-hodnotu získáme při testování hypotéz ve statistickém softwaru.

- Je-li **p-hodnota  $\leq \alpha$** , pak  **$H_0$  zamítáme** na hladině významnosti  $\alpha$  a **přijímáme  $H_A$**
- Je-li **p-hodnota  $> \alpha$** , pak  **$H_0$  nezamítáme** na hladině významnosti  $\alpha$

P-hodnota vyjadřuje pravděpodobnost za platnosti  $H_0$ , s níž bychom získali stejnou nebo extrémnější hodnotu testové statistiky.

# Test dobré shody - základní teorie

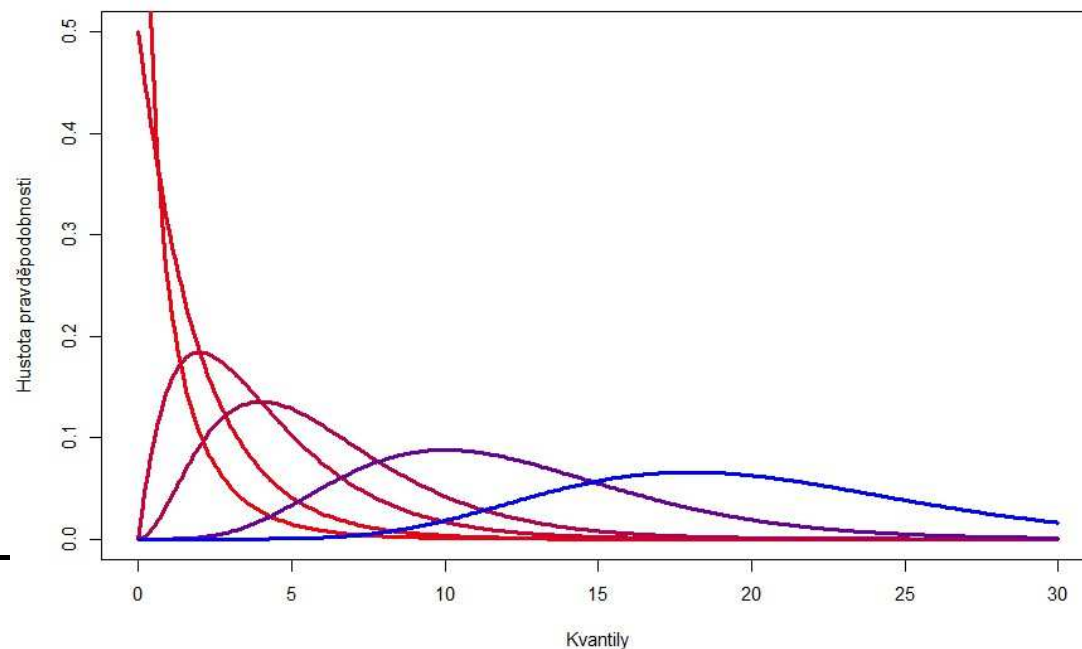


Testuje shodu reálné distribuce hodnot do  $n$  skupin s teoretickou distribucí. Předpokladem je, že velikost rozdílu mezi očekávaným a skutečným počtem hodnot v každé skupině je náhodně rozdělená → multinomické rozdělení.

Součet druhých mocnin relativních rozdílů očekávaného a skutečného počtu hodnot má přibližně  $\chi^2$  rozdělení.

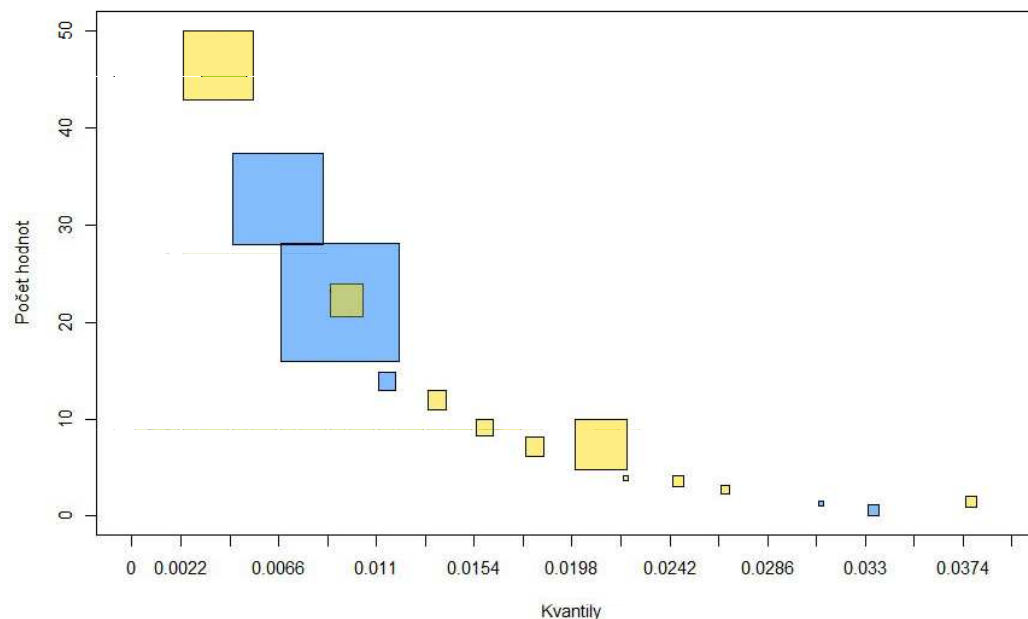
$\chi^2$  rozdělení pro kladné hodnoty (suma čtverců) se liší podle počtu stupňů volnosti  $k$  (počtu skupin) - se zvyšujícím se  $k$  přechází v normální rozdělení.

$$\chi^2_{(s.v.)} = \sum \frac{\left[ \begin{array}{c} \text{pozorovaná} \\ \text{četnost} \end{array} - \begin{array}{c} \text{očekávaná} \\ \text{četnost} \end{array} \right]^2}{\text{očekávaná četnost}}$$



# Test dobré shody - základní teorie

$$\chi^2_{(s.v.)} = \frac{\left[ \begin{array}{c} \text{pozorovaná} \\ \text{četnost} \end{array} - \begin{array}{c} \text{očekávaná} \\ \text{četnost} \end{array} \right]^2}{\underbrace{\text{očekávaná četnost}}_{\text{1. jev}}} + \frac{\left[ \begin{array}{c} \text{pozorovaná} \\ \text{četnost} \end{array} - \begin{array}{c} \text{očekávaná} \\ \text{četnost} \end{array} \right]^2}{\underbrace{\text{očekávaná četnost}}_{\text{2. jev}}} + \dots$$



# Znaménkový test



Zjednodušení neparametrického párového Wilcoxonova testu.

Namísto velikosti rozdílů se počítá pouze jejich orientace (signum).

Případy, kde  $sgn(d) = 0$  se z analýzy vylučují.

Sečou se kladné a záporné rozdíly a menší ze součtů je hledaná statistika  $m$ .

Statistika  $m$  se porovná s tabulkovou hodnotou pro danou hladinu pravděpodobnosti:

Počet párů $n$	Hladina významnosti ( $\alpha$ )		
	0,01	0,05	0,10
5	-	-	0
6	-	0	0
7	-	0	0
8	0	0	1
9	0	1	1
10	0	1	1
11	0	1	2
12	1	2	2
13	1	2	3
14	1	2	3
15	2	3	3
16	2	3	4
17	2	4	4
18	3	4	5
19	3	4	5
20	3	5	5

# Shrnutí statistických testů



Typ srovnání	Nulová hypotéza	Parametrický test	Neparametrický test
1 skupina dat vs. etalon	Střední hodnota je rovna hodnotě etalonu.	jednovýběrový t-test	Mann-Whitneyův test
2 skupiny dat nepárově	Obě skupiny hodnot pochází ze stejného rozdělení.	nepárový t-test	Mann-Whitneyův test
2 skupiny dat párově	Zkoumaný efekt mezi páry hodnot je nulový.	Párový t-test	Wilcoxonův test; znaménkový test
shoda rozdělení	rozdělení dat ve skupině odpovídá teoretickému (vybranému) rozdělení.	Shapiro-Wilksův test; Kolmogorovův-Smirnovův test; test dobré shody	test dobré shody
homoskedasticita (shoda rozptylů)	rozptyl obou (všech) skupin je shodný.	Levenův test	-
více skupin nepárově	Zkoumaný efekt mezi skupinami hodnot je nulový.	ANOVA	Kruskal- Wallisův test
korelace	Neexistuje (příčinná, důsledková) vazba mezi skupinami hodnot.	Pearsonův koeficient	Spearmanův koeficient; Kendallův koeficient

# Kontingenční tabulky

## H<sub>0</sub> :Nezávislost dvou jevů A a B

**Kontingenční  
tabulka  
2 x 2**

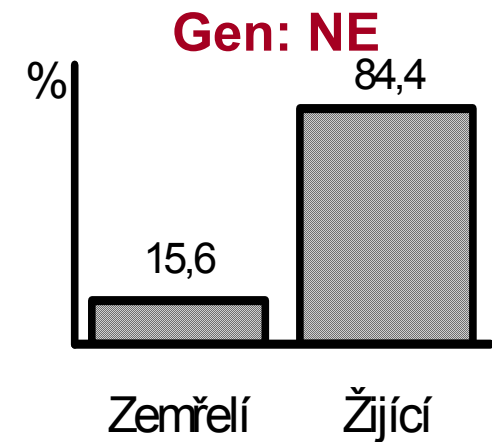
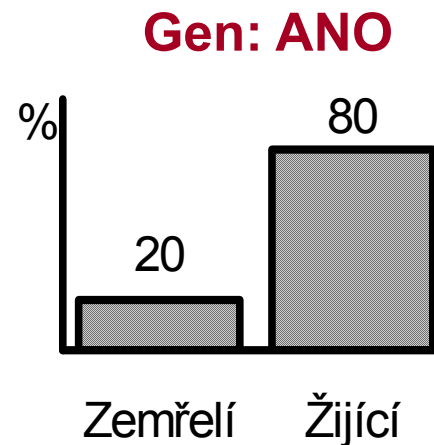
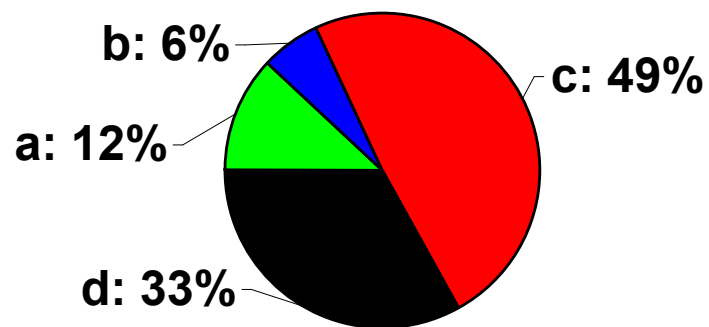
$\downarrow$ B $\rightarrow$ A	+	-	$\Sigma$
+	a	b	
-	c	d	
$\Sigma$			suma sum



# Kontingenční tabulky: příklad

gen \ †	Ano	Ne	$\Sigma$
Ano	20	82	102
Ne	10	54	64
$\Sigma$	30	136	166

## Kontingenční tabulka v obrázku



# Příklad – závislost pohlaví na onemocnění



	Zdraví	Nemocní	Celkem
Muži	50	50	100
Ženy	50	50	100
Celkem	100	100	200

	Zdraví	Nemocní	Celkem
Muži	45	55	100
Ženy	55	45	100
Celkem	100	100	200

	Zdraví	Nemocní	Celkem
Muži	35	65	100
Ženy	65	35	100
Celkem	100	100	200

	Zdraví	Nemocní	Celkem
Muži	5	95	100
Ženy	95	5	100
Celkem	100	100	200

# Příklad – závislost pohlaví na onemocnění

## Pozorované hodnoty

	Zdraví	Nemocní	Celkem
Muži	50	50	100
Ženy	50	50	100
Celkem	100	100	200

	Zdraví	Nemocní	Celkem
Muži	45	55	100
Ženy	55	45	100
Celkem	100	100	200

	Zdraví	Nemocní	Celkem
Muži	35	65	100
Ženy	65	35	100
Celkem	100	100	200

	Zdraví	Nemocní	Celkem
Muži	5	95	100
Ženy	95	5	100
Celkem	100	100	200



## Očekávané hodnoty pro všechny tabulky vlevo

	Zdraví	Nemocní	Celkem
Muži	50	50	100
Ženy	50	50	100
Celkem	100	100	200

$$\chi^2_{(s.v.)} = \sum \frac{\left[ \begin{array}{cc} \text{pozorovaná} & - & \text{očekávaná} \\ \text{četnost} & & \text{četnost} \end{array} \right]^2}{\text{očekávaná četnost}}$$

# Příklad – závislost pohlaví na onemocnění



	Zdraví	Nemocní	Celkem
Muži	50	50	100
Ženy	50	50	100
Celkem	100	100	200

$$X^2 = 0,0$$

$$p = 1,000$$

	Zdraví	Nemocní	Celkem
Muži	45	55	100
Ženy	55	45	100
Celkem	100	100	200

$$X^2 = 2,0$$

$$p = 0,157$$

	Zdraví	Nemocní	Celkem
Muži	35	65	100
Ženy	65	35	100
Celkem	100	100	200

$$X^2 = 18,0$$

$$p < 0,0001$$

	Zdraví	Nemocní	Celkem
Muži	5	95	100
Ženy	95	5	100
Celkem	100	100	200

$$X^2 = 162,0$$

$$p < 0,0001$$