

# Pravděpodobnost a zobecňující statistika

Inference (možnost zobecnění) jako  
předpoklad statistické práce.

Statistická signifikace. (Různé typy testů: Chi<sup>2</sup>,  
T-test, F-test. Neparametrické testy.)

# NEJPRVE STANDARDIZACE:

- Proč standardizace?
- Jak se provádí.
- Co to je Z-SCORE.

# Co je to nulová hypotéza?

Obvykle se testuje NULOVÁ HYPOTÉZA ( $H_0$ ) jako specifický model statistické hypotézy. NULOVÁ HYPOTÉZA PŘEDPOKLÁDÁ STAV „NEEXISTENCE“ (ROZDÍLU) ČI STAV SHODY:

- Rozložení hodnot znaku se neliší od nějakého teoretického rozložení (například normálního nebo náhodného).
- Rozložení četností hodnot proměnné (vlastností jednotky), např. příjmu, věku, míry anomie, spokojenosti v životě, ..., ve výběrovém souboru odpovídá rozložení proměnné v populaci (neliší se od něho).
- Mezi dvěma parametry, např. mezi průměrným příjmem mužů a žen, není v základním souboru rozdíl (usuzujeme na to ze zjištěných statistik ve výběrovém souboru).
- Mezi empirickým a náhodným rozložením hodnot v kontingenční tabulce není rozdíl (empirické rozložení je náhodné, neexistuje vztah nejen mezi 2 proměnnými, které tabulku tvoří, ale ani mezi jejich variantami).

# Nulová hypotéza

Hypotézy lze zásadně prohlásit za falešné (tedy zamítnout jejich platnost), nikoliv však dokazovat jejich platnost. Hypotéza nemůže být přímo dokázána, nýbrž může být jen zamítnuta jí odporující (nulová) hypotéza.

# DVA VÝSLEDKY TESTOVÁNÍ $H_0$

NEMÁME DŮVOD ZAMÍTNOUT MODEL  
NULOVÉ HYPOTÉZY A PROTO JÍ  
PŘIJÍMÁME.

DATA NEODPOVÍDAJÍ  $H_0$  (jejich  
existence je při platnosti  $H_0$  vysoce  
nepravděpodobná) A PROTO JÍ  
ZAMÍTÁME. JEJÍ ZAMÍTNUTÍ VŠAK  
VĚTŠINOU NESTAČÍ PRO PŘIJETÍ  
ALTERNATIVNÍ HYPOTÉZY.

# HLADINA VÝZNAMNOSTI (significance level)

Nazývá se tak pravděpodobnost, že náhodná odchylka (daná výběrovou chybou) překročí určitou danou hodnotu, nazývanou hranice významnosti či KRITICKÁ HODNOTA. Představuje velikost rizika chyby, jež připustíme.

Zjištěné (empirické) odchylky, vyskytující se s pravděpodobností MENŠÍ NEŽ JE ZVOLENÁ HLADINA VÝZNAMNOSTI (HV), se nazývají STATISTICKY VÝZNAMNÉ (signifikantní) na této zvolené hladině.

# TESTOVACÍ KRITÉRIUM

- Každému testovacímu kritériu PŘÍSLUŠÍ TEORETICKÉ ROZDĚLENÍ (normální rozložení, t neboli Studentovo rozložení, F rozložení, ...).
- Tabelovány bývají jeho KRITICKÉ HODNOTY.

# POSTUP TESTOVÁNÍ

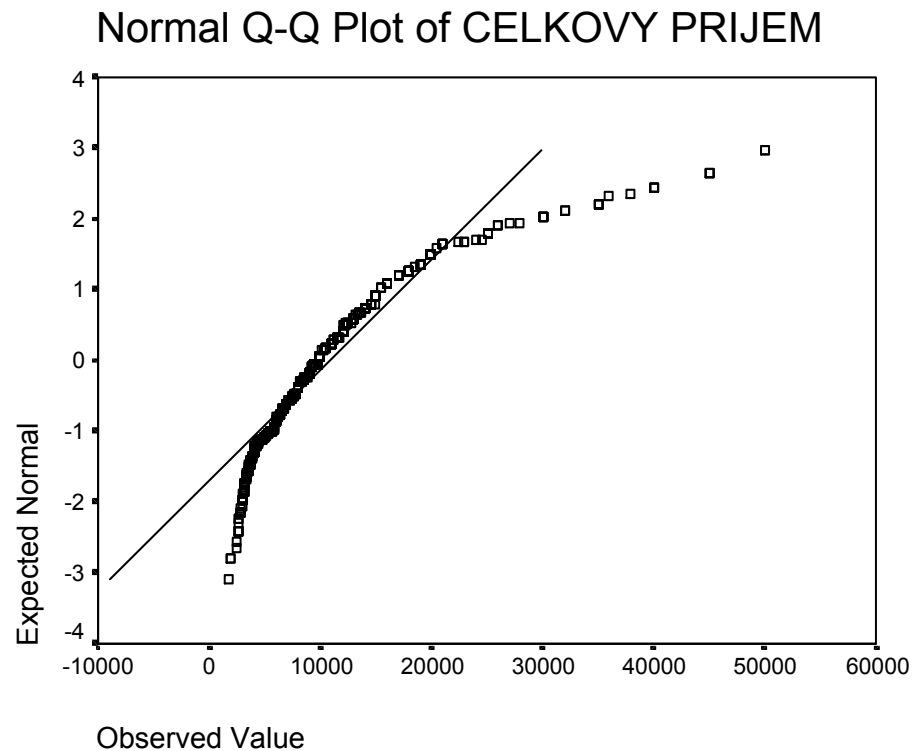
- Zvolíme vhodné TESTOVACÍ KRITÉRIUM, jehož teoretické rozložení (standardizované normální rozložení, Studentovo rozložení, rozložení chí-kvadrátu...).
- Vypočítáme z dat výběrového souboru jeho empirickou hodnotu (z-skóre jemuž odpovídá standardizované normální rozložení, t hodnotu jíž odpovídá Studentovo rozložení, chí-kvadrát jemuž odpovídá rozložení chí-kvadrát ...).
- Porovnáme vypočítanou statistiku s jejím teoretickým rozložením - s její KRITICKOU HODNOTOU ( $T^*$ ).



# 1. TEST NORMALITY ROZLOŽENÍ

## Procedura EXPLORE: Plots

### - NORMAL Q-Q PLOT



#### Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>		
	Statistic	df	Sig.
A85 CELKOVY PRI	,125	988	,000

a. Lilliefors Significance Correction

Intervaly spolehlivosti  
FREQUENCY

$$\sqrt{\frac{p(100 - p)}{N}}$$

# Interval spolehlivosti pro dichotomické proměnné

Trik spočívá v tom, že hodnoty dichotomie (ať byly kódovány jako 0 a 1, nebo jako 1 a 2) převedeme (rekódujeme procedurou *Recode*) na hodnoty 0 a 100.

Descriptives. – Zjistíme průměr, což je pravděpodobnost. A pomocí předcházejícího vzorce spočítáme STANDARDNÍ CHYBU.

Tj., například

=ODMOCNINA((55,9\*(100-55,9)/1803))

Krát 2 (HLADINA VÝZNAMNOSTI 95%)

# Interval spolehlivosti pro polytomické proměnné

- FREQUENCIES...
- PRAVDĚPODOBNOST je validní procento.

# Jak postupujeme?

Microsoft Excel - int-spol.xls

Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda

Times New Roman CE 11 B I U % , +.00 -.00

D14 = =ODMOCNINA(((C14\*(100-C14))/B18)\*2

	A	B	C	D	E	F	G	H	I
1	Q46_3 Většina žen touží po domově a dětech								
2			Frequency	Percent	Valid Percent	Cumulative Percent			
3	Valid	1 rozhodně souhlasí	213	11,2	12,0	12,0			
4		2 souhlasí	1070	56,1	60,1	72,1			
5		3 nesouhlasí	474	24,9	26,6	98,7			
6		4 rozhodně nesouhlasí	22	1,2	1,3	100,0			
7		Total	1780	93,3	100,0				
8	Missing	-2 neodpověděl/a	8	0,4					
9		-1 neví	120	6,3					
10		Total	128	6,7					
11	Total		1908	100,0					
12				<b>std. Error</b>	<b>CI</b>	<b>CI</b>			
13		<b>Freq.</b>	<b>Valid %</b>	<b>95%</b>	<b>dolní</b>	<b>dolní</b>			
14	1 rozhodně souhlasí	213	12,0	1,54	10,4	13,5			
15	2 souhlasí	1070	60,1	2,32	57,8	62,4			
16	3 nesouhlasí	474	26,6	2,10	24,6	28,7			
17	4 rozhodně nesouhlasí	22	1,3	0,53	0,7	1,8			
18	Total	1780	100,0						
19									
20									

$$\sqrt{\frac{p(100-p)}{N}}$$

List1 / List2 / List3 / List4 / List5 / List6 /

Připraven 123

Tedy Standardizovaná CHYBA je:

$$= \text{ODMOCNINA}((12*(100-12)/1780))$$

krát 2 (interval spolehlivosti 95%).