VI. Základy testování hypotéz

Princip statistického testování hypotéz
Pojmy statistických testů
Normalita dat a její význam pro testování

Vytvořil Institut biostatistiky a analýz, Masarykova univerzita
J. Jarkovský, L. Dušek

Anotace

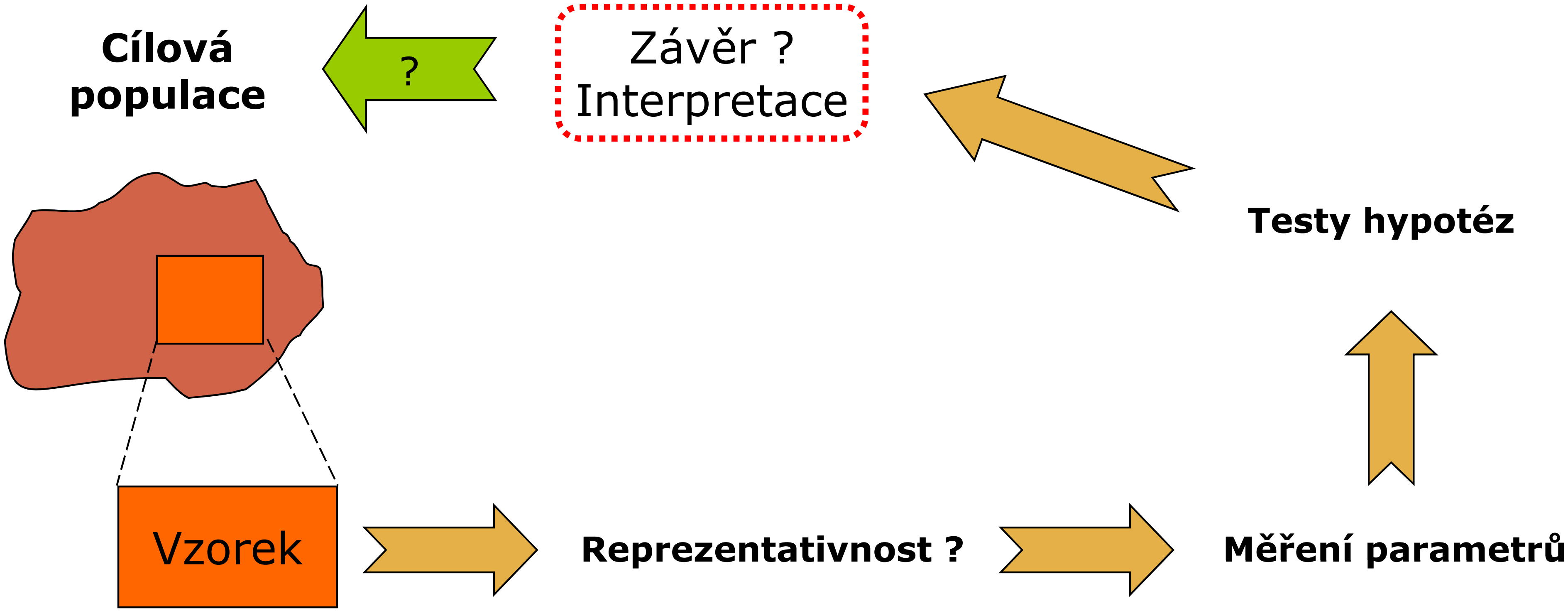
- Testování hypotéz je po popisné statistice druhým hlavním směrem statistických analýz. Při testování pokládáme hypotézy, které se snažíme s určitou pravděpodobností potvrdit nebo vyvrátit.
- Tzv. nulovou hypotézu lze nejlépe popsat jako situaci, kdy předpokládáme vliv náhody (rozdíl mezi skupinami je pouhá náhoda, vztah dvou proměnných je pouhá náhoda apod.), alternativní hypotéza předpokládá vliv nenáhodného faktoru.
- Výsledkem statistického testu je v zásadě pravděpodobnost nakolik je hodnocený jev náhodný nebo ne, při překročení určité hranice (nejčastěji méně než 5% pravděpodobnost, že jev je pouhá náhoda) deklarujeme, že pravděpodobnost náhody je pro nás dostatečně nízká abychom jev prohlásili za nenáhodný
- Statistická významnost je ovlivnitelná velikostí vzorku a tak je pouze indicií k prohlášení např. rozdílu dvou skupin pacientů za skutečně významný. V ideální situaci je nezbytné aby rozdíl byl významný nejenom statisticky (=nenáhodný), ale i prakticky (=nejde pouze o artefakt velikosti vzorku).

Hypotézy

- H_0 - tvrdenie o parametroch alebo type rozloženia, z ktorého pochádza náhodný výber. Vyjadruje nejaký teoretický predpoklad, často skeptického rázu a užívateľ ho musí stanoviť vopred, bez prihliadnutia k dátovému súboru.
- H_1 – alternatívna hypotéza, ktorá hovorí, čo platí ak neplatí nulová hypotéza.
- Hypotézy musia byť stanovené tak, aby nemohli platiť zároveň.
- Testovanie hypotéz – postup, ktorý je založený na danom náhodnom výbere a s pomocou ktorého rozhodneme o zamietnutí alebo nezamietnutí nulovej hypotézy.

Princip testování hypotéz

- **Formulace hypotézy**
- **Výběr cílové populace a z ní reprezentativního vzorku**
- **Měření sledovaných parametrů**
- **Použití odpovídajícího testu** → **závěr testu**
- **Interpretace výsledků**

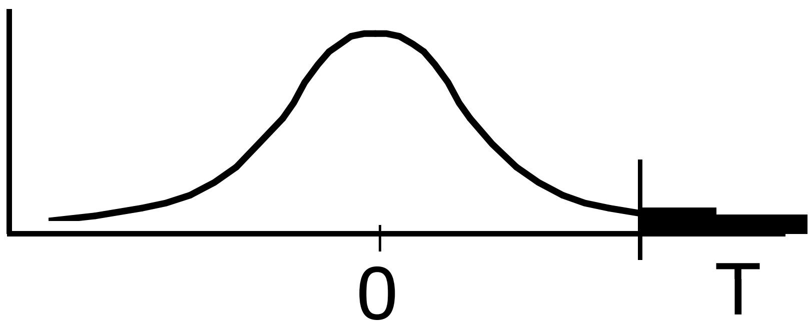


Statistické testování – základní pojmy

- **Nulová hypotéza H_O** — H_O: sledovaný efekt je nulový
- **Alternativní hypotéza H_A** — H_A: sledovaný efekt je různý mezi skupinami
- **Testová statistika**

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

- **Kritický obor testové statistiky**



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

Možné chyby při testování hypotéz

- I přes dostatečnou velikost vzorku a kvalitní design experimentu se můžeme při rozhodnutí o zamítnutí/nezamítnutí nulové hypotézy dopustit chyby.

Správné rozhodnutí

Chyba I. druhu

		Závěr testu	
		Hypotézu nezamítáme	Hypotézu zamítáme
Skutečnost	H_0 Platí	1- α	α
	H_0 Neplatí	β	1- β

Správné rozhodnutí

Chyba II. druhu

Význam chyb při testování hypotéz

Pravděpodobnost chyby 1. druhu

α → **Pravděpodobnost nesprávného zamítnutí nulové hypotézy**

Pravděpodobnost chyby 2. druhu

β → **Pravděpodobnost nerozpoznání neplatné nulové hypotézy**

Síla testu

$1-\beta$ → **Pravděpodobnostně vyjádřená schopnost rozpoznat neplatnost hypotézy**

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 α (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 α a přijímáme H_A.
- Je-li p-hodnota $> \alpha$, pak H_0 nezamítáme na hladině významnosti α.

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

Parametrické vs. neparametrické testy

Parametrické testy

- Mají předpoklady o rozložení vstupujících dat (např. normální rozložení)
- Při stejném N a dodržení předpokladů mají vyšší sílu testu než testy neparametrické
- Pokud nejsou dodrženy předpoklady parametrických testů, potom jejich síla testu prudce klesá a výsledek testu může být zcela chybný a nesmyslný

Neparametrické testy

- Nemají předpoklady o rozložení vstupujících dat, lze je tedy použít i při asymetrickém rozložení, odlehlých hodnotách, či nedetekovatelném rozložení
- Snížená síla těchto testů je způsobena redukcí informační hodnoty původních dat, kdy neparametrické testy nevyužívají původní hodnoty, ale nejčastěji pouze jejich pořadí

One-sample vs. two sample testy

Jedno-výběrové testy (one-sample)

- Srovnávají jeden vzorek (one sample, jednovýběrové testy) s referenční hodnotou (popřípadě se statistickým parametrem cílové populace)
- V testu je tedy srovnáváno rozložení hodnot (vzorek) s jediným číslem (referenční hodnota, hodnota cílové populace)
- Otázka položená v testu může být vztažena k průměru, rozptylu, podílu hodnot i dalším statistickým parametrům popisujícím vzorek

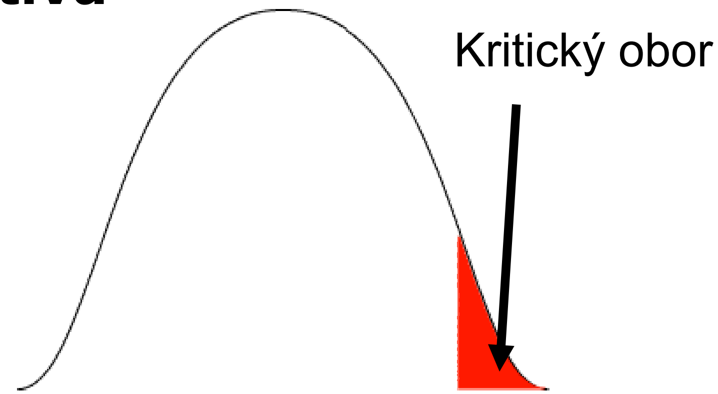
Dvou-výběrové testy (two-sample)

- Srovnávají navzájem dva vzorky (two sample, dvouvýběrové testy)
- V testu jsou srovnávány dvě rozložení hodnot
- Otázka položená v testu může být opět vztažena k průměru, rozptylu, podílu hodnot i dalším statistickým parametrům popisujícím vzorek
- Kromě testů pro dvě skupiny hodnot existují samozřejmě i testy pro více skupin dat

One-tailed vs. Two-tailed testy

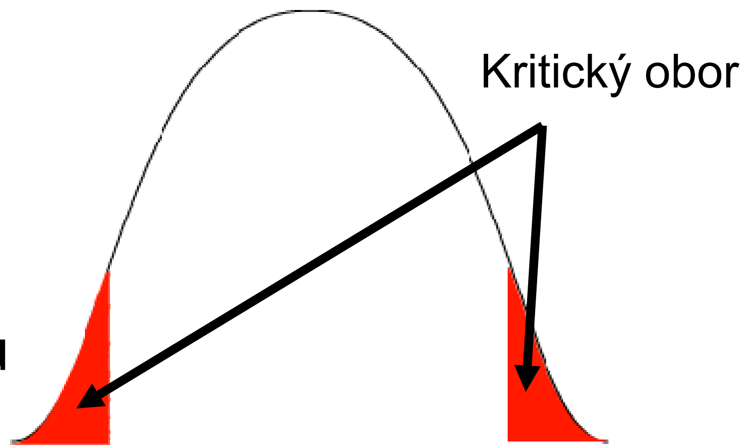
One – tailed testy= jednostranná alternatíva

- Hypotéza testu je postavena asymetricky, tedy ptáme se na **větší než/ menší než**
- Test může mít pouze dvojí výstup – jedna z hodnot je větší (menší) než druhá a všechny ostatní případy



Two – tailed testy=obojstranná alternatíva

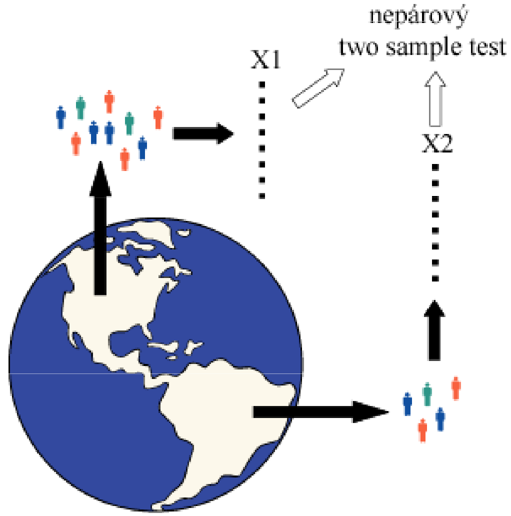
- Hypotéza testu se ptá na otázku **rovná se/nerovná se**
- Test může mít trojí výstup – **menší - rovná se – větší než**
- Situace **nerovná se** je tedy souhrnem dvou možných výstupů testu (**menší+větší**)



Nepárový vs. párový design

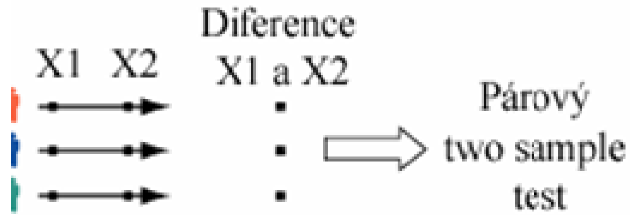
Nepárový design

- Skupiny srovnávaných dat jsou na sobě zcela nezávislé (též nezávislý, independent design), např. lidé z různých zemí, nezávislé skupiny pacientů s odlišnou léčbou atd.
- Při výpočtu je nezbytné brát v úvahu charakteristiky obou skupin dat



Párový design

- Mezi objekty v srovnávaných skupinách existuje vazba, daná např. člověkem před a po operaci, reakce stejného kmene krys atd.
- Vazba může být buď přímo dána nebo pouze předpokládána (v tom případě je nutné ji ověřit)
- Test je v podstatě prováděn na diferencích skupin, nikoliv na jejich původních datech



Statistické testy a normalita dat

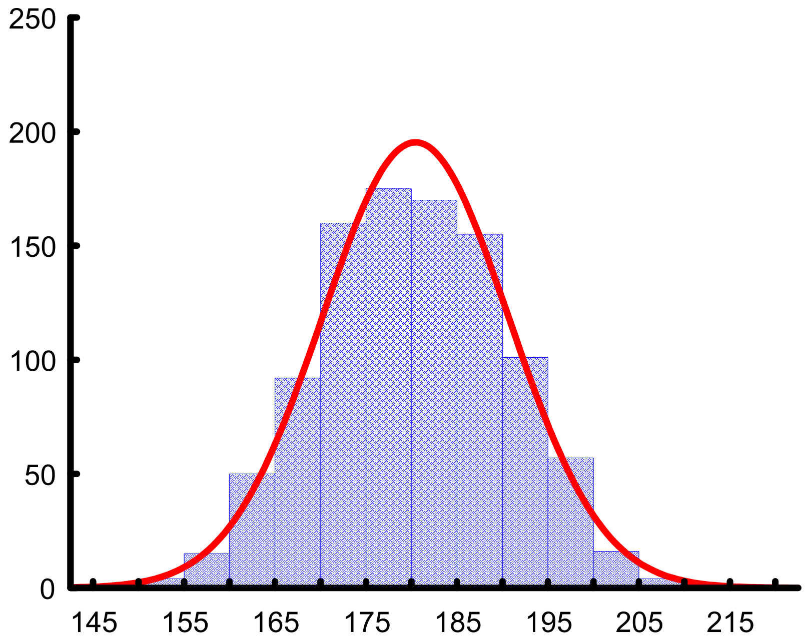
- **Normalita dat je jedním z předpokladů tzv. parametrických testů (testů založených na předpokladu nějakého rozložení) – např.** *t*-testy
- **Pokud data nejsou normální, neodpovídají ani modelovému rozložení, které je použito pro výpočet (*t*-rozložení) a test tak může lhát**

- **Řešením je tedy:**
  - Transformace dat **za účelem dosažení normality jejich rozložení**
  - Neparametrické testy **– tyto testy nemají žádné předpoklady o rozložení dat**

Typ srovnání	Parametrický test	Neparametrický test
2 skupiny dat nepárově:	Nepárový t-test	Mann Whitney test
2 skupiny dat párově:	Párový t-test	Wilcoxon test, znaménkový test
Více skupin nepárově:	ANOVA	Kruskal- Wallis test
Korelace:	Pearsonův koeficient	Spearmanův koeficient

Testy normality

- Testy normality pracují s nulovou hypotézou, že není rozdíl mezi zpracovávaným rozložením a normálním rozložením. Vždy je ovšem dobré prohlédnout si i histogram, protože některé odchylky od normality, např. bimodalitu některé testy neodhalí.



•Test dobré shody

V testu dobré shody jsou data rozdělena do kategorií (obdobně jako při tvorbě histogramu), tyto intervaly jsou normalizovány (převedeny na normální rozložení) a podle obecných vzorců normálního rozložení jsou k nim dopočítány očekávané hodnoty v intervalech, pokud by rozložení bylo normální. Pozorované normalizované četnosti jsou poté srovnány s očekávanými četnostmi pomocí $\chi 2$ testu dobré shody. Test dává dobré výsledky, ale je náročný na n, tedy množství dat, aby bylo možné vytvořit dostatečný počet tříd hodnot.

•Kolgomorov –Smirnov test

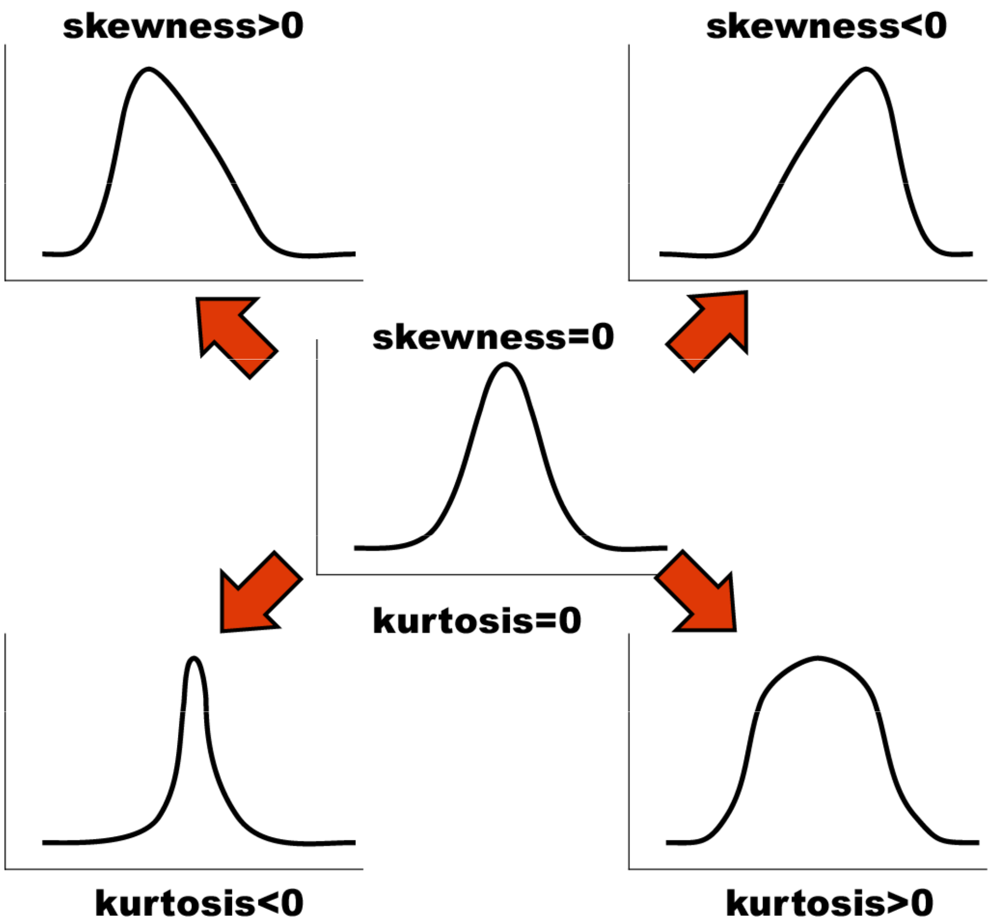
Tento test je často používán, dokáže dobře najít odlehlé hodnoty, ale počítá spíše se symetrií hodnot než přímo s normalitou. Jde o neparametrický test pro srovnání rozdílu dvou rozložení. Je založen na zjištění rozdílu mezi reálným kumulativním rozložením (vzorek) a teoretickým kumulativním rozložením. Pro testování normálního rozložení, ak odhadujeme parametry z výběru, se používá modifikace – Lilieforsův test.

•Shapiro-Wilk`s test

Jde o neparametrický test použitelný i při velmi malých n (10) s dobrou sílou testu, zvláště ve srovnání s alternativními typy testů, je zaměřen na testování symetrie.

Šikmost a špičatost jako testy normality

- Parametry normálního rozložení, skewness a kurtosis mohou být využity pro testování normality, ale pouze pro velké vzorky (šikmost – 100, špičatost – 500).



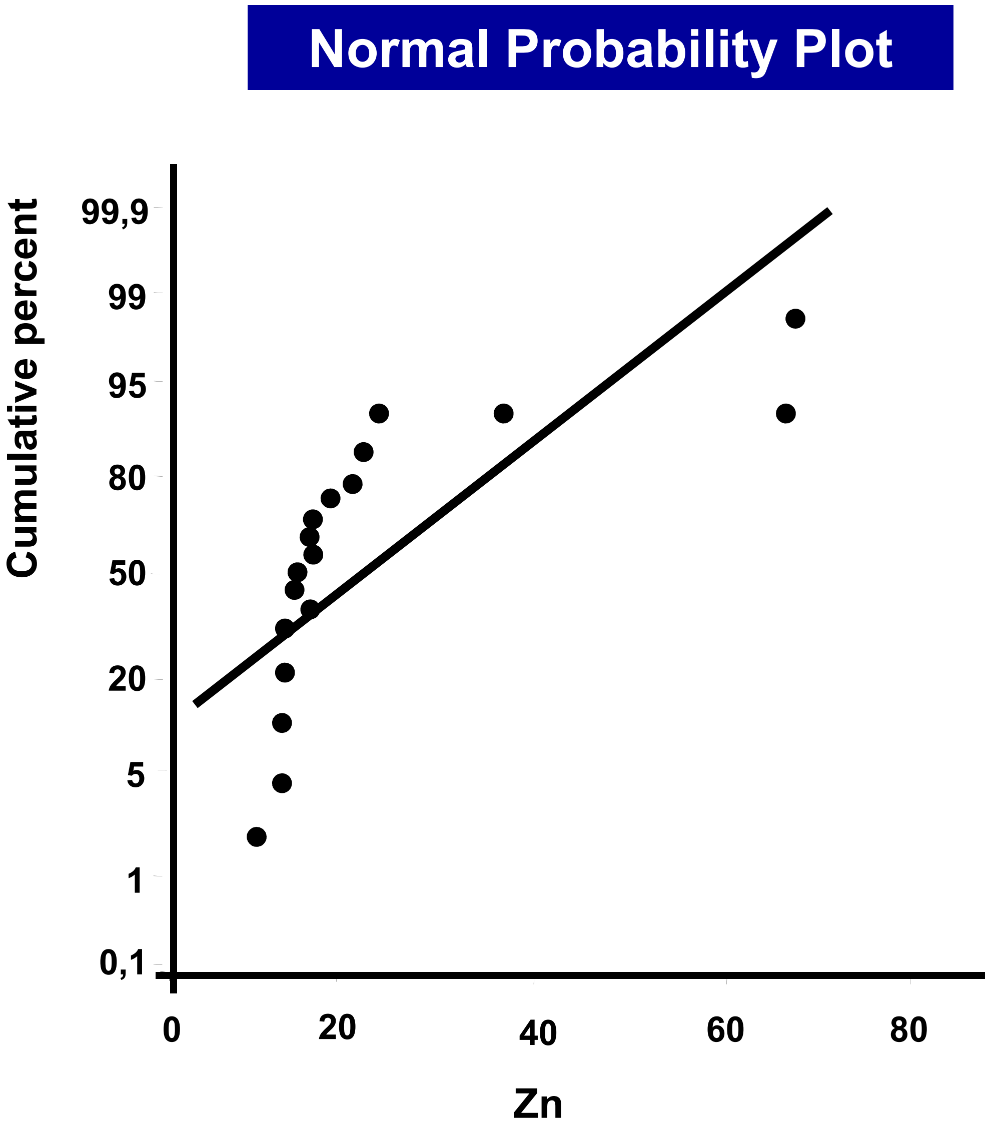
Shapiro-Wilksov test (S-W test)

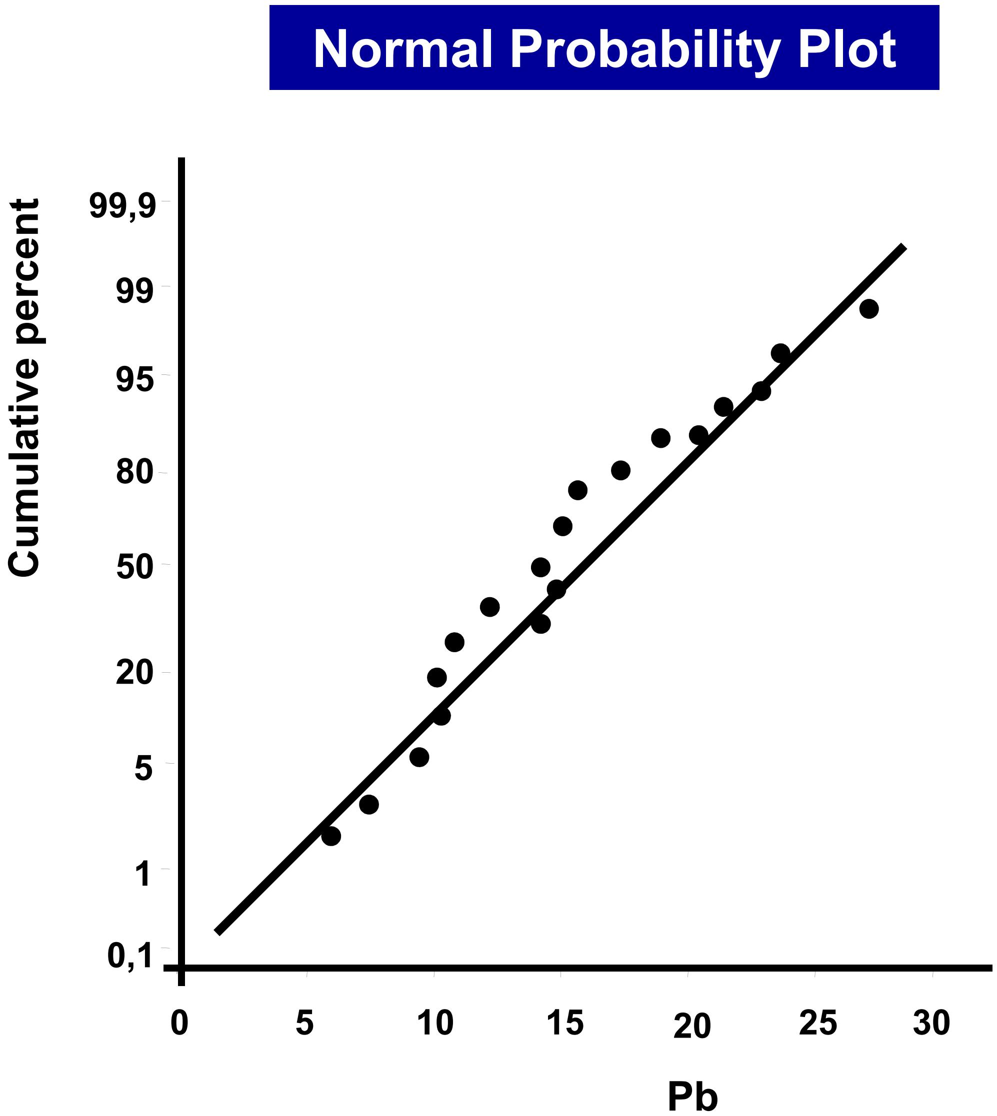
- H0: náhodný výber pochádza z normálneho rozloženia
- Testová štatistika

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n} x_i - \bar{x}}.$$

- a_i – váhy odvodené od stredných hodnôt a rozptylov náhodného výberu z rozloženia N(0,1)
- Hodnota W určuje podobnosť s normálnym rozložením
- Čím je W bližšie k 1, tým je zhoda väčšia
- Rozhodujeme sa na základe p-hodnoty

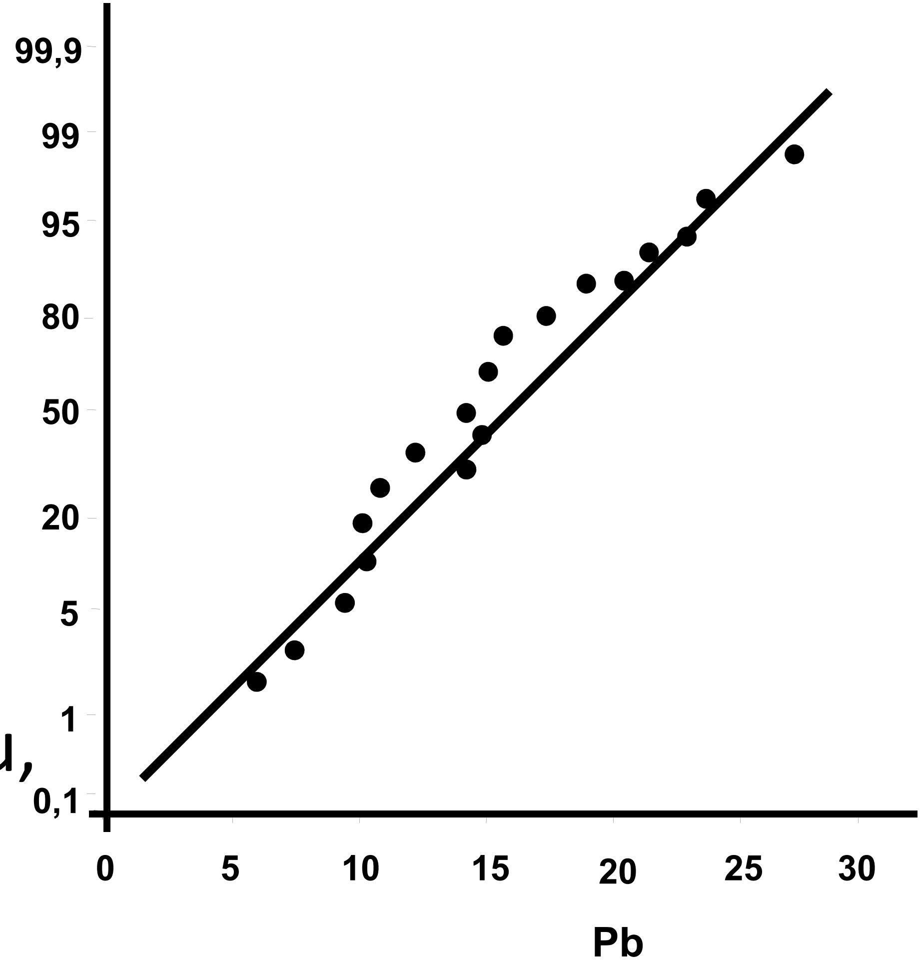
Grafická diagnostika normality



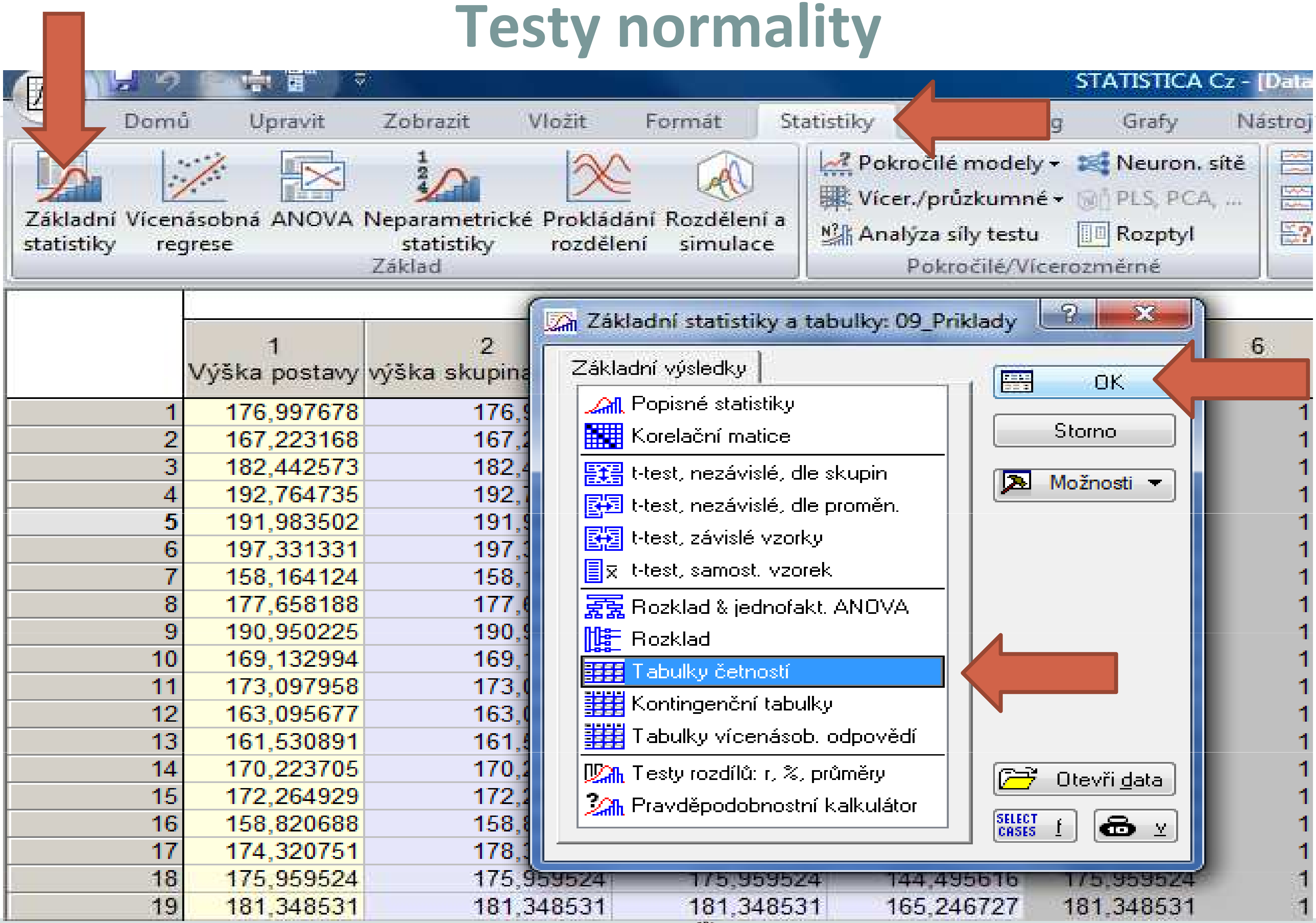


N-P plot (Normal-probability plot)

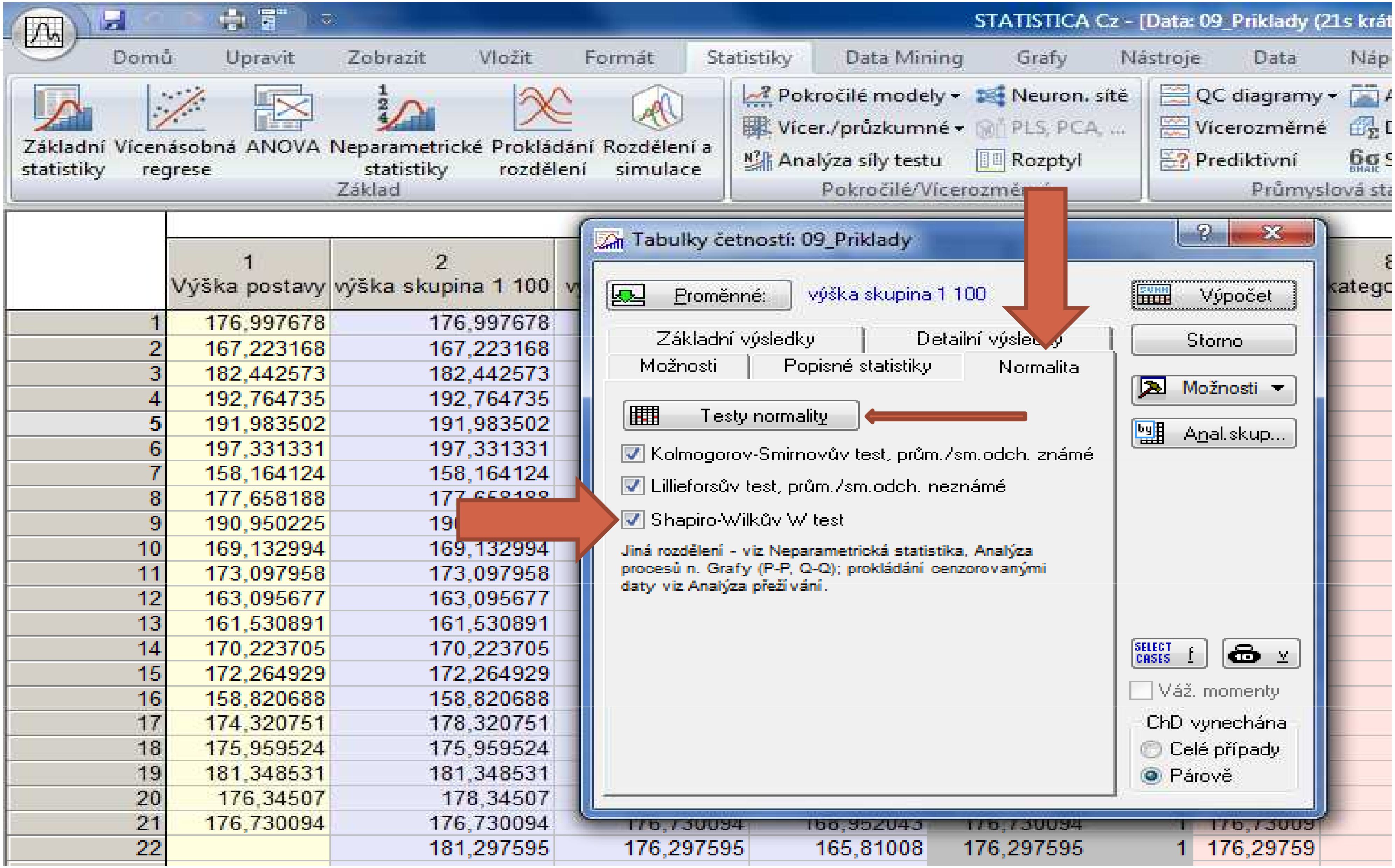
- Máme náhodný výber
- Odhadneme strednú hodnotu a rozptyl
- Na osu x vynášame namerané hodnoty (z našich dát)
- Na osu y teoretické očakávané hodnoty
- Týmito bodmi preložíme krivku, ak je priamka, výber pochádza z normálneho rozloženia



Testy normality

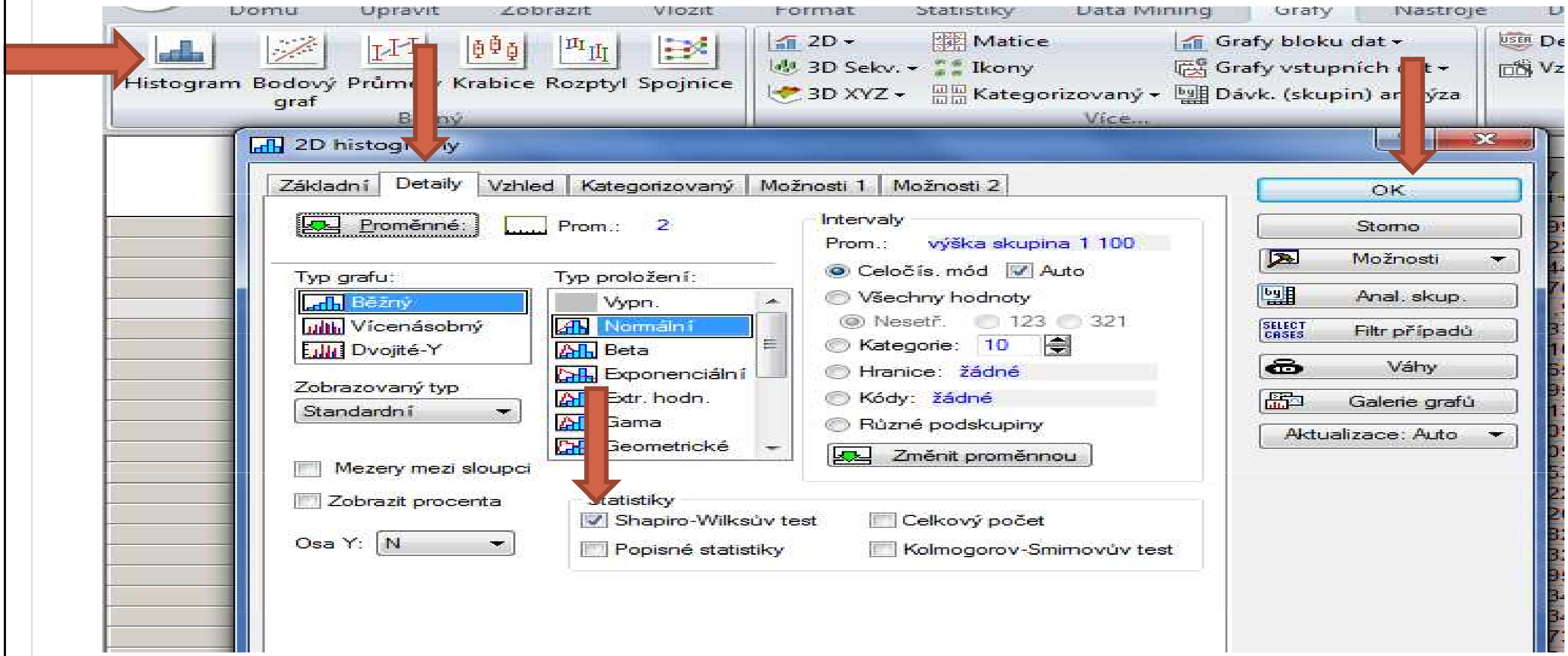


Vyberieme premenné a zaklikneme S-W test



Histogram

- V záložke *Grafy* vyberieme *Histogram*, zvolíme premennú a v záložke *Detaily* v časti *Statistiky* zvolíme *Shapiro-Wilksov test*



N-P plot

- Záložka: *Statistiky* vyberieme *Základní statistiky ,* zvolíme *Popisné statistiky*
- V záložke *Pravd. & bodové grafy* vyberieme *Normálni pravděpod. graf*
- nezabudneme vybrať premennú

Príklad

- Otvorte si súbor 06_Priklady.sta
- Testujte normalitu premenných
  - Výška – prem. 19
  - Váha – prem. 20
  - Parametr – prem. 15

- testujte pomocou histogramu s krivkou normálneho rozloženia, N-P testu a Shapiro-Wilksovho testu

Statistické testy o parametrech jednoho výběrů

Jednovýběrový t-test
Jednovýběrový test rozptylu

Anotace

- Jednovýběrové statistické testy srovnávají některou popisnou statistiku vzorku (průměr, směrodatnou odchylku) s jediným číslem, jehož význam je ze statistické hlediska hodnota cílové populace
- Z hlediska statistické teorie jde o ověření, zda daný vzorek pochází z testované cílové populace.

“One sample“ testy I

V případě one sample testů jde o srovnání výběru dat (tedy one sample) s cílovou populací. Pro parametrické testy musí mít datový soubor normální rozložení.

Průměr – cílová vs. výběrová populace

$$t = \frac{\bar{x} - \mu}{s} \sqrt{n}$$

H_0	H_A	Testová statistika	Interval spolehlivosti
$\bar{x} \leq \mu$	$\bar{x} > \mu$	**t**	$t > t_{1-\alpha}^{(n-1)}$
$\bar{x} \geq \mu$	$\bar{x} < \mu$	**t**	$t < t_{\alpha}^{(n-1)}$
$\bar{x} = \mu$	$\bar{x} \neq \mu$	**t**	$\lvert t \rvert > t_{1-\alpha/2}^{(n-1)}$

Softvér nám vypíše p-hodnotu, ktorú porovnáme s hladinou významnosti, inak by sme museli hľadať hodnotu Intervalu spoľahlivosti v Štatistických tabuľkách

“One sample“ testy II

V případě one sample testů jde o srovnání výběru dat (tedy one sample) s cílovou populací. Pro parametrické testy musí mít datový soubor normální rozložení.

Rozptyl – cílová vs. výběrová

$$\chi^2 = \frac{(n-1).s^2}{\sigma^2}$$

H_0	H_A	Testová statistika	Interval spolehlivosti
$s^2 \leq \sigma^2$	$s^2 > \sigma^2$	χ^2	$\chi^2 > \chi^2_{1-\alpha}$ ***(n-1)***
$s^2 \geq \sigma^2$	$s^2 < \sigma^2$	χ^2	$\chi^2 < \chi^2_{\alpha}$ ***(n-1)***
$s^2 = \sigma^2$	$s^2 \neq \sigma^2$	χ^2	$\chi^2 > \chi^2_{1-\alpha/2}$ ***nebo*** $\chi^2 < \chi^2_{\alpha/2}$

Srovnání odhadu průměru s předpokládanou hodnotou I

10krát nezávisle bola odmeraná konštanta μ. Výsledky merania sú 2, 1.8, 2.1, 2.4, 1.9, 2.1, 2, 1.8, 2.3, 2.2.

Smerodajná odchýlka bola určená ako 0.2. Nejaká teória tvrdí, že μ=0.95.

- **H0:**
- **H1:**
- **Testová štatistika:**

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n}$$

- **riadi sa normálnym rozložením N(0,1)**

Srovnání odhadu průměru s předpokládanou hodnotou I

- Pomocou kritického oboru:
  - $W=(-\infty, u_{\alpha/2}>U<u_{1-\alpha/2},\infty)$
  - H0 zamietame, ak t leží v W
- Pomocou intervalu spoľahlivosti:
  - $(d,h)=(\bar{x} - \sigma/sqrt(n)*u_{1-\alpha/2} , \bar{x}+\sigma/sqrt(n)*u_{1-\alpha/2})$
  - Zamietame ak μ neleží v tomto intervale
- Pomocou p-hodnoty:
  - $p=2 \min \{P(T\leq t),P(T\geq t)\}$
  - Zamietame ak $p<\alpha$

Srovnání odhadu průměru s předpokládanou hodnotou II

Aktivita enzymu v buňkách

Při zjišťování aktivity enzymu v buňkách na vzorku 25 měření byl zjištěn průměr 3,5 jednotek a směrodatná odchylka 1.

1. otázka zní, zda se naměřené hodnoty našeho vzorku liší od výsledků dřívější rozsáhlé studie zaměřené na celou cílovou populaci, kde byla zjištěna průměrná aktivita 2,5 jednotky?

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{3{,}5 - 2{,}5}{1}\sqrt{25} = 5$$

H0: x=μ tedy two tailed test

$$t_{0{,}975}^{24} = 2{,}064 \Rightarrow t > t_{1-\alpha/2}^{24} \Rightarrow$$ ***H0 zamítnuta při $\alpha \leq 0{,}05$***

od jiné hodnoty bychom zachytili při daných hodnotách?

2. otázka – jakou minimální odchylku X od jiné hodnoty bychom zachytili při daných hodnotách?

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{d}{s}\sqrt{n} \rightarrow d = \frac{t_{1-\alpha/2}^{\nu}}{\sqrt{n}} s \rightarrow d = \frac{2{,}064}{5}1$$

3. za předpokladu, že z praktického hlediska je významná odchylka již 0,2 jednotky, jaký minimální počet měření musíme provést, abychom ji byli schopni prokázat ?

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{d}{s}\sqrt{n} \rightarrow n = \left(\frac{t_{1-\alpha/2}^{\nu}}{d} s\right)^2$$

Jednovýberový t-test

- V záložke *Statistiky* vyberieme *Základní statistiky*
- V okienku vyberieme *t-test samost. Vzorky- OK*
- V záložke *Detailní výsledky* zvolíme referenčnú hodnotu a zvolíme *výpočet*

Príklad

- Otestujte, či priemerná výška v premennej Výška skupina 1100 je 175cm
- Otestujte, či priemerná výška v rovnakej premennej je 179cm

Samostatný príklad

- Testujte, či priemerný počet leukocytov je 7
- Testujte, či priemerný počet leukocytov je 9
- Testujte, či je priemerná váha 84kg
- Nezabudnite najskôr otestovať normalitu☺