Parametrické testy

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

Parametrické testy

- Předpoklad: **normalita dat**
- *Studentův t-test* (testování rozdílů dvou středních hodnot)

  varianty t-testu:

  1. ***Jednovýběrový t-test*** (porovnání základního a výběrového souboru, známe střední hodnotu základního souboru)
  2. ***Dvouvýběrový t-test*** (porovnání dvou výběrových souborů, neznáme střední hodnotu základního souboru):

     - ***párový*** (závislé výběry)
     - ***nepárový*** (nezávislé výběry)
- *F-test* (testování rozdílů dvou rozptylů)

1. Statistické testy o parametrech jednoho výběru

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 statistického 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

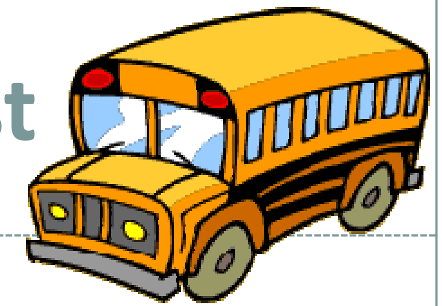
$$t = \frac{\bar{x} - \mu}{s} \sqrt{n}$$

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**	$\lvert t \rvert > t_{1-\alpha/2}^{(n-1)}$

μ - střední hodnota základního souboru
$\bar{x}$ - průměr výběrového souboru
s^2 - rozptyl výběrového souboru
n - počet členů výběrového souboru

1-α/2 kvantil Studentova *t*- rozdělení pro dané stupně volnosti (n-1) a zvolené α

Příklad 1: Jednovýběrový t-test

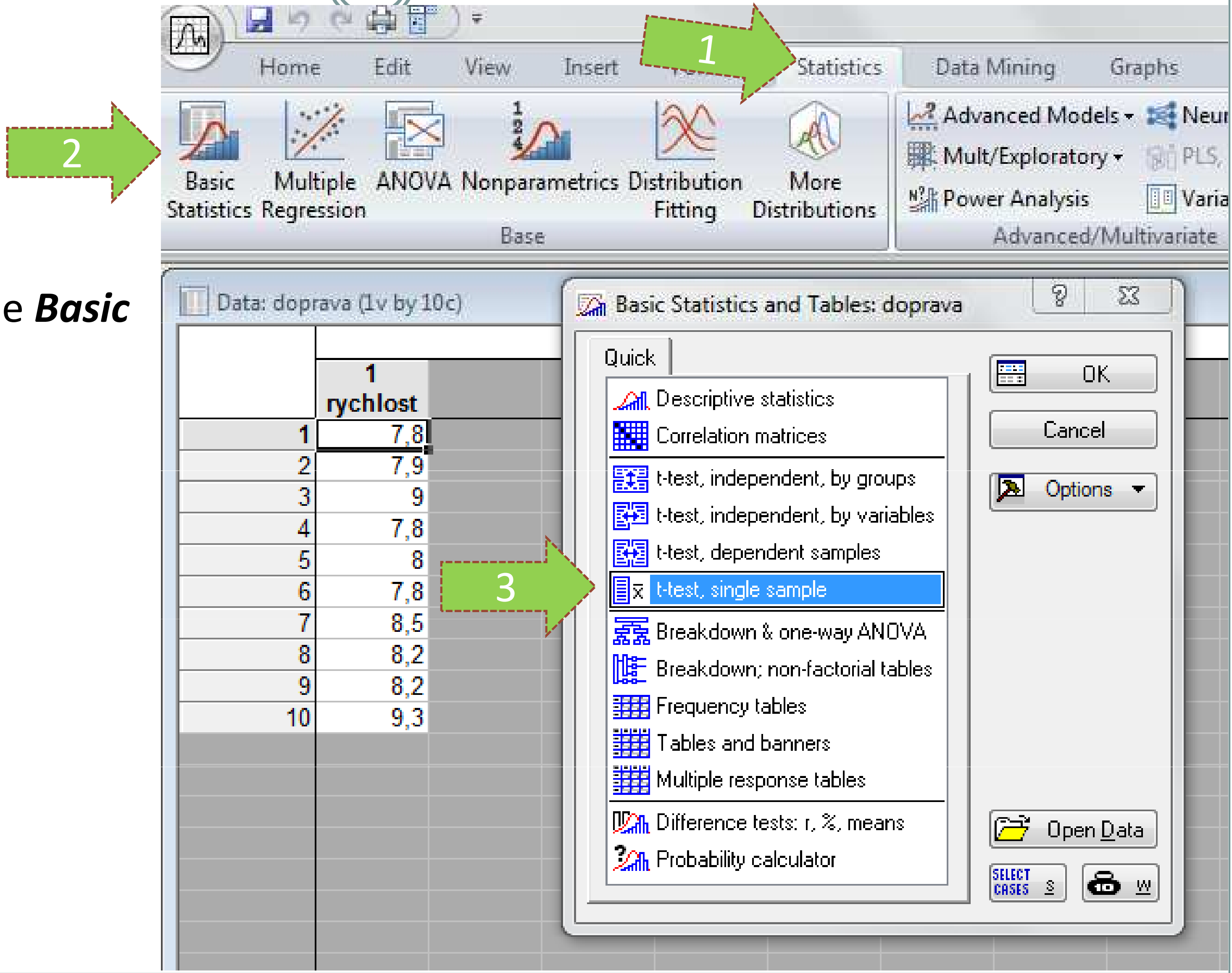


- Určitá linka autobusové městské dopravy má v době dopravní špičky průměrnou rychlost 8 km/hod. Uvažovalo se o tom, zda změna trasy by vedla ke změně průměrné rychlosti. Nová trasa byla proto projeta v deseti náhodně vybraných dnech a byly zjištěny tyto průměrné rychlosti: 7,8 7,9 9,0 7,8 8,0 7,8 8,5 8,2 8,2 9,3. Rozhodněte, zda změna trasy vede ke změně průměrné rychlosti. Předpokládáme normální rozdělení a α=0,05.
- **Postup:**

1. **Na hladině významnosti 0,05 testujeme hypotézu H_0: $\mu = 8$, proti H_A : $\mu \neq 8$**
2. **Vypočteme aritmetický průměr a rozptyl výběrového souboru.**
3. **Vypočteme testové kritérium t:**
$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{8{,}25 - 8}{0{,}530}\sqrt{10} = 1{,}492$$
4. **Vypočtené t porovnáme s kritickou hodnotou $t_{1-\alpha/2(n-1)}$:** $t_{0{,}975}(9) = 2{,}262$
5. **Je-li $t \leq t_{1-\alpha/2(n-1)}$ ➡ statisticky nevýznamný rozdíl testovaných parametrů při zvolené** α; nulovou hypotézu nezamítáme, na hladině významnosti α=0,05 se nepodařilo prokázat, že by změna trasy měla za následek změnu průměrné rychlosti.

Příklad 1: Řešení v softwaru Statistica I

• V menu ***Statistics*** zvolíme ***Basic statistics*** ,vybereme ***t-test,single sample***



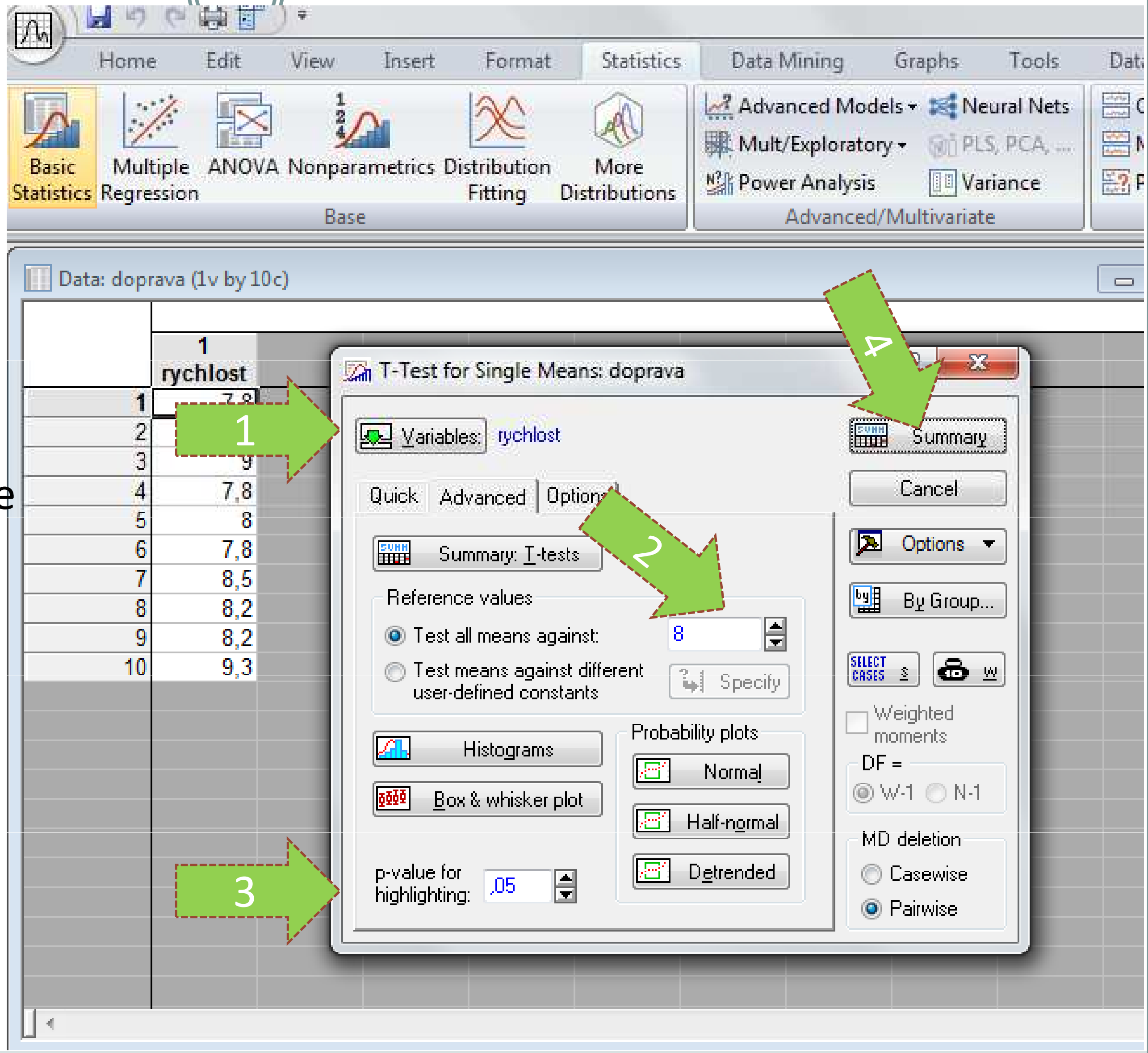
Řešení v softwaru Statistica II

• Vybereme proměnnou, kterou chceme testovat

• Na kartě ***Advanced*** napíšeme do okénka ***Test all means against*** velikost střední hodnoty populace (lze také na kartě ***Quick, Options***)

• ***p-value for highlighting-*** Úroveň p lze změnit

•Kliknutím na ***Summary t-test*** nebo na ***Summary*** získáme výstupy



Řešení v softwaru Statistica III

Rozsah výběru

Hodnota testovacího kritéria

Standardní chyba

Výběrový průměr stat. znaku

Stupeň volnosti

Variable	Test of means against reference constant (value) (doprava)							
	Mean	Std.Dv.	N	Std.Err.	Reference Constant	t-value	df	p
rychlost	8,250000	0,529675	10	0,167498	8,000000	1,492556	9	0,169753

Výběrová směrodatná odchylka stat.znaku

POZOR: Platí pro oboustranný test!!!

Referenční konstanta-předpokládaná velikost střední hodnoty

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

"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 populace

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

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** $\chi^2 < \chi^2_{\alpha/2}$

2. Statistické testy o parametrech dvou výběrů

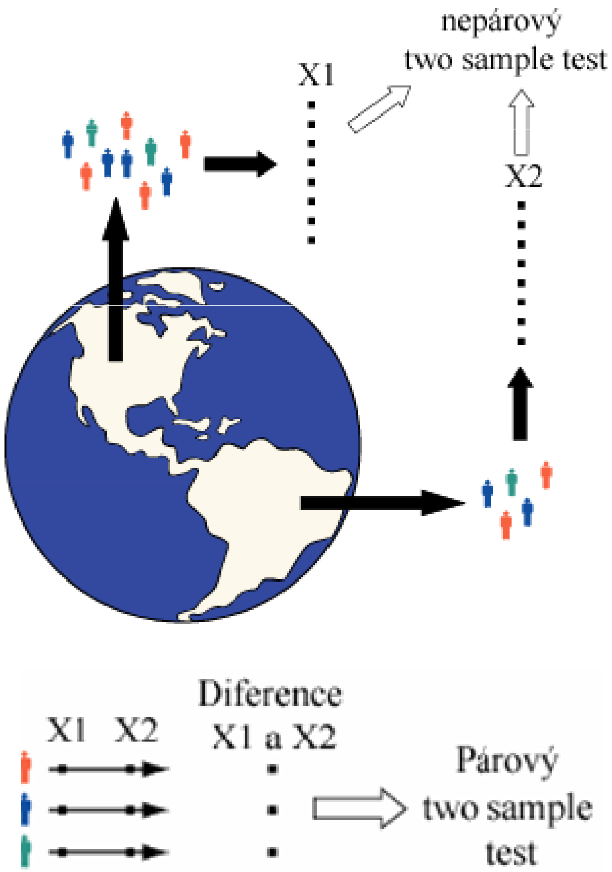
Dvouvýběrový párový a nepárový t-test

Anotace

- Jedním z nejčastějších úkolů statistické analýzy dat je srovnání spojitých dat ve dvou skupinách pacientů. Na výběr je celá škála testů, výběr konkrétního testu se pak odvíjí od toho, zda je o srovnání párové nebo nepárové a zda je vhodné použít test parametrický (má předpoklady o rozložení dat) nebo neparametrický (nemá předpoklady o rozložení dat, nicméně má nižší vypovídací sílu).
- Nejznámějšími testy z této skupiny jsou tzv. t-testy používané pro srovnání průměrů dvou skupin hodnot

Dvouvýběrové testy: párové a nepárové I

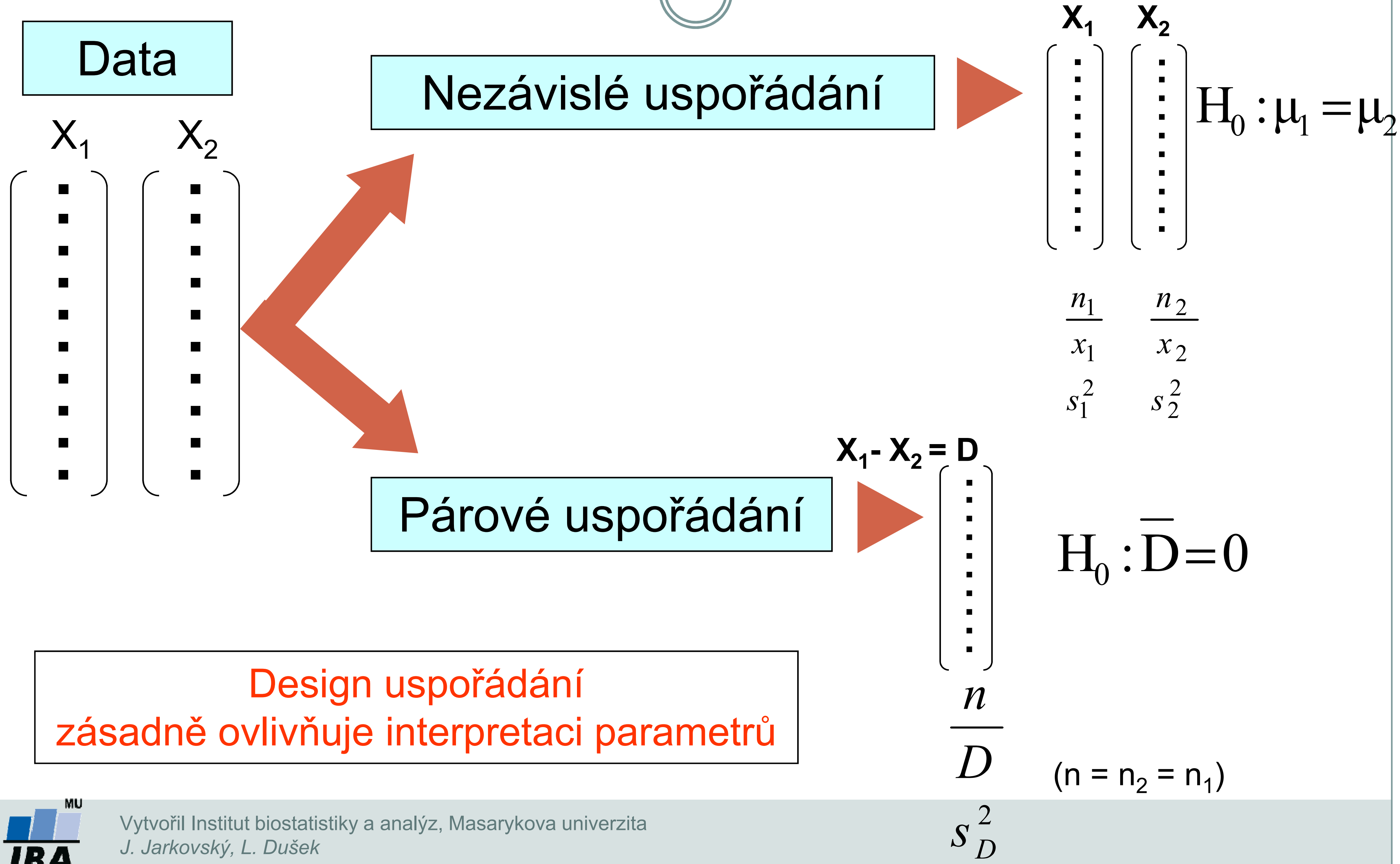
- Při použití two sample testů srovnáváme spolu dvě rozložení. Jejich základním dělením je podle designu experimentu na testy párové a nepárové.



- Základním testem pro srovnání dvou nezávislých rozložení spojitých čísel je **nepárový two-sample t-test**

- Základním testem pro srovnání dvou závislých rozložení spojitých čísel je **párový two-sample t-test**

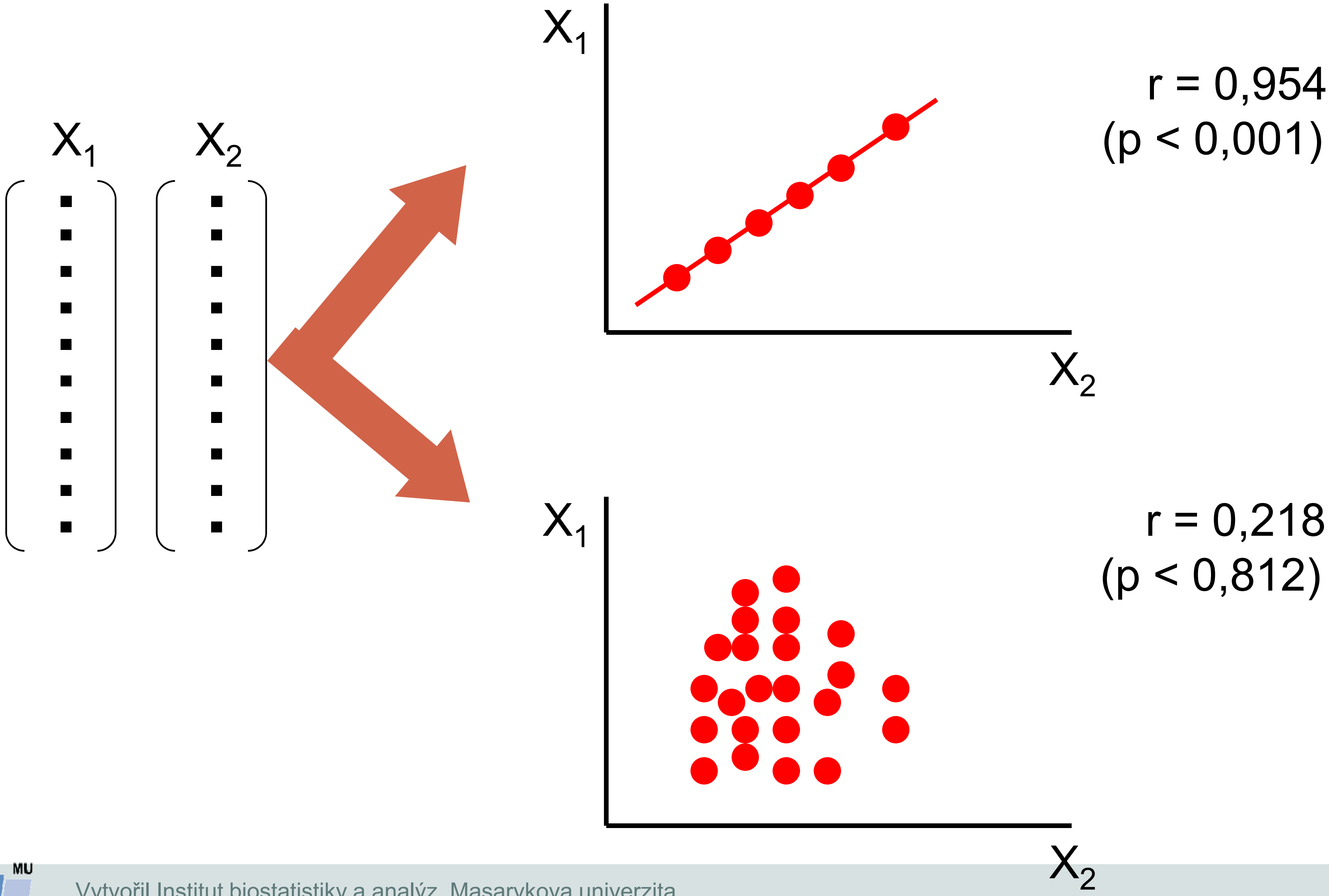
Dvouvýběrové testy: párové a nepárové II



Design uspořádání
zásadně ovlivňuje interpretaci parametrů

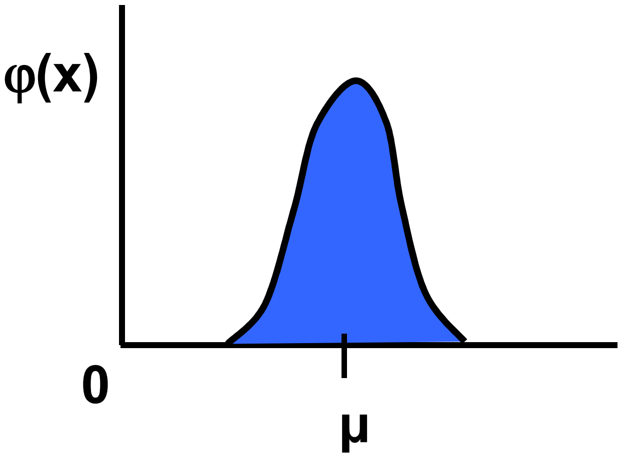
Dvouvýběrové testy: párové a nepárové III

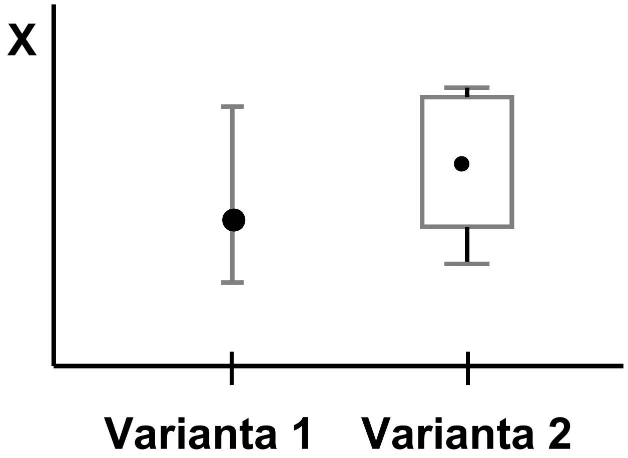
Identifikace párovitosti (Korelace, Kovariance)



Předpoklady nepárového dvouvýběrového t-testu

- Náhodný výběr subjektů jednotlivých skupin z jejich cílových populací
- Nezávislost obou srovnávaných vzorků
- Přibližně **normální rozložení proměnné ve vzorcích**, drobné odchylky od normality ovšem nejsou kritické, test je robustní proti drobným odchylkám od tohoto předpokladu, normalita může být testována testy normality
- **Rozptyl v obou vzorcích by měl být přibližně shodný** (homoscedastic). Tento předpoklad je testován několika možnými testy – *Levenův test* nebo *F-test*.
- Vždy je vhodné prohlédnout histogramy proměnné v jednotlivých vzorcích pro okometrické srovnání a ověření předpokladů normality a homogenity rozptylu – nenahradí statistické testy, ale poskytne prvotní představu.





Nepárový dvouvýběrový t-test – výpočet I

1. Nulová hypotéza: průměry obou skupin jsou shodné, alternativní hypotéza je, že nejsou shodné, two tailed test
2. Prohlédnout průběh dat, průměr, medián apod. pro zjištění odchylek od normality a nehomogenitu rozptylu, provést F –test

H_0	H_A	Testová statistika
$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F = \frac{s_1^2}{s_2^2}$
$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F = \frac{s_2^2}{s_1^2}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$F = \frac{\max\left(s_1^2; s_2^2\right)}{\min\left(s_1^2; s_2^2\right)}$

F-test pro srovnání dvou výběrových rozptylů

- Používá se pro srovnání rozptylu dvou skupin hodnot, často za účelem ověření homogenity rozptylu těchto skupin dat.

- V případě ověření homogenity je testována hypotéza shody rozptylů (two tailed); v případě shodných rozptylů je vše v pořádku a je možné pokračovat ve výpočtu t-testu, v opačném případě není vhodné test počítat.

Nepárový dvouvýběrový t-test – výpočet II

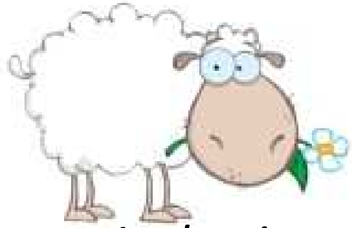
3. Výpočet testové statistiky (stupně volnosti jsou $\upsilon = n_1 + n_2 - 2$):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$ vážený odhad rozptylu

4. výsledné *t* srovnáme s tabulární hodnotou t pro dané stupně volnosti a α (obvykle α=0,05)
5. Lze spočítat interval spolehlivosti pro rozdíl průměrů (např. 95%), počet stupňů volnosti a s^2 odpovídají předchozím vzorcům

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0,975} SE(\bar{x}_1 - \bar{x}_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{0,975}\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$



Příklad 2: Nepárový dvouvýběrový t-test

1. skupina, N=30

Průměrná hmotnost ovcí v čase páření byla srovnávána pro kontrolní skupinu a skupinu krmenou zvýšenou dávkou potravy. Kontrolní skupina obsahuje 30 ovcí, skupina se zvýšeným příjmem potravy pak 24 ovcí.

- Vlastní experiment byl prováděn tak, že na začátku máme 54 ovcí (ideálně stejného plemene, stejně staré atd.), které náhodně rozdělíme do dvou skupin (náhodné rozdělování objektů do pokusných skupin je objektem celého specializovaného odvětví statistiky nazývaného randomizace). Poté co experiment proběhne, musíme nejprve ověřit teoretický předpoklad pro využití nepárového t-testu. Pro obě proměnné jsou vykresleny grafy (můžeme též spočítat základní popisnou statistiku), na kterých můžeme posoudit normalitu a homogenitu rozptylu, kromě okometrického pohledu můžeme pro <u>ověření normality použít testy normality, pro ověření homogenity rozptylu pak F-test</u>
- Pokud platí všechny předpoklady Two sample nepárového t-testu, můžeme spočítat testovou charakteristiku, výsledné t je 2,43 s 52 stupni volnosti, podle tabulek je a $t_{0,975\,(52)}$= 2,01, tedy t> $t_{0,975\,(52)}$= a nulovou hypotézu můžeme zamítnout, skutečná pravděpodobnost je pak 0,018. Rozdíl mezi skupinami je 1,59 kg ve prospěch skupiny se zvýšeným příjmem.

$$t = \frac{Rozdíl\_průrůmě}{SE(rozdílprůo\ ěrů)} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \qquad s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \qquad \upsilon = n_1 + n_2 - 2$$

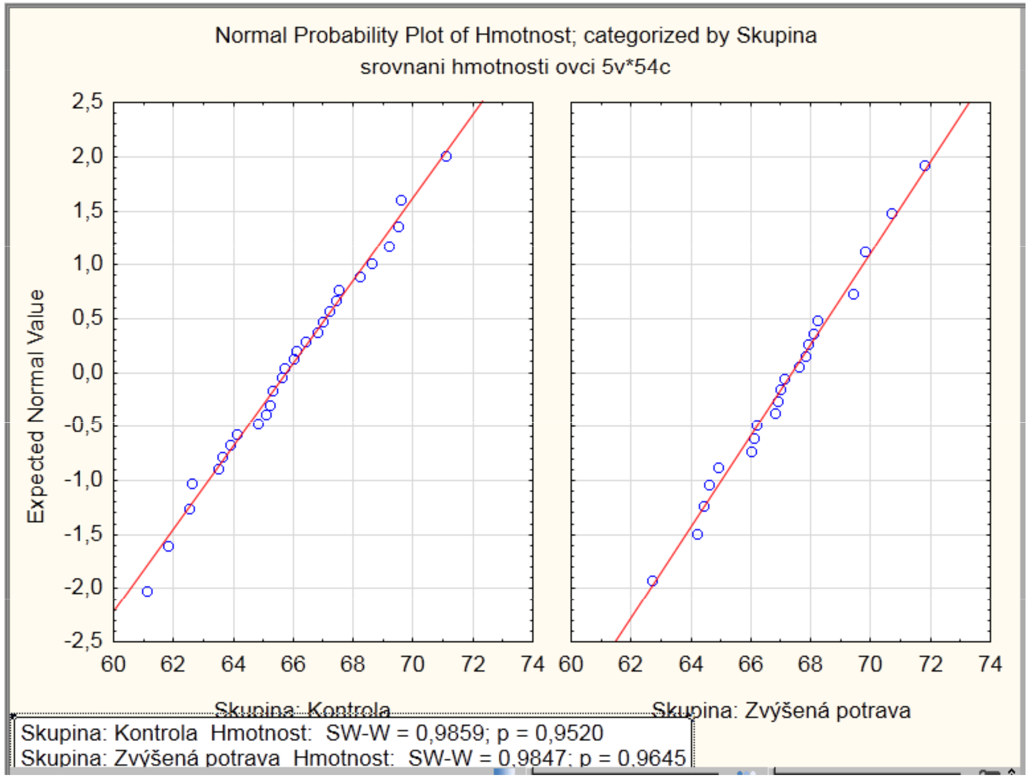
- Pro rozdíl mezi oběma soubory jsou spočítány 95% konfidenční intervaly jako 1,59±2.01*(0,655) kg, což odpovídá rozsahu 0,28 až 2,91 kg. To, že konfidenční interval nezahrnuje 0 je dalším potvrzením, že mezi skupinami je významný rozdíl – jde o další způsob testování významnosti rozdílů mezi skupinami dat – nulovou hypotézu o tom, že rozdíl průměrů dvou skupin dat je roven nějaké hodnotě zamítáme v případě, kdy 95% konfidenční interval rozdílu nezahrnuje tuto hodnotu (v tomto případě 0).

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0,975} SE(\bar{x}_1 - \bar{x}_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{0,975}\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

2. skupina, N=24

Příklad 2: Řešení v softwaru Statistica

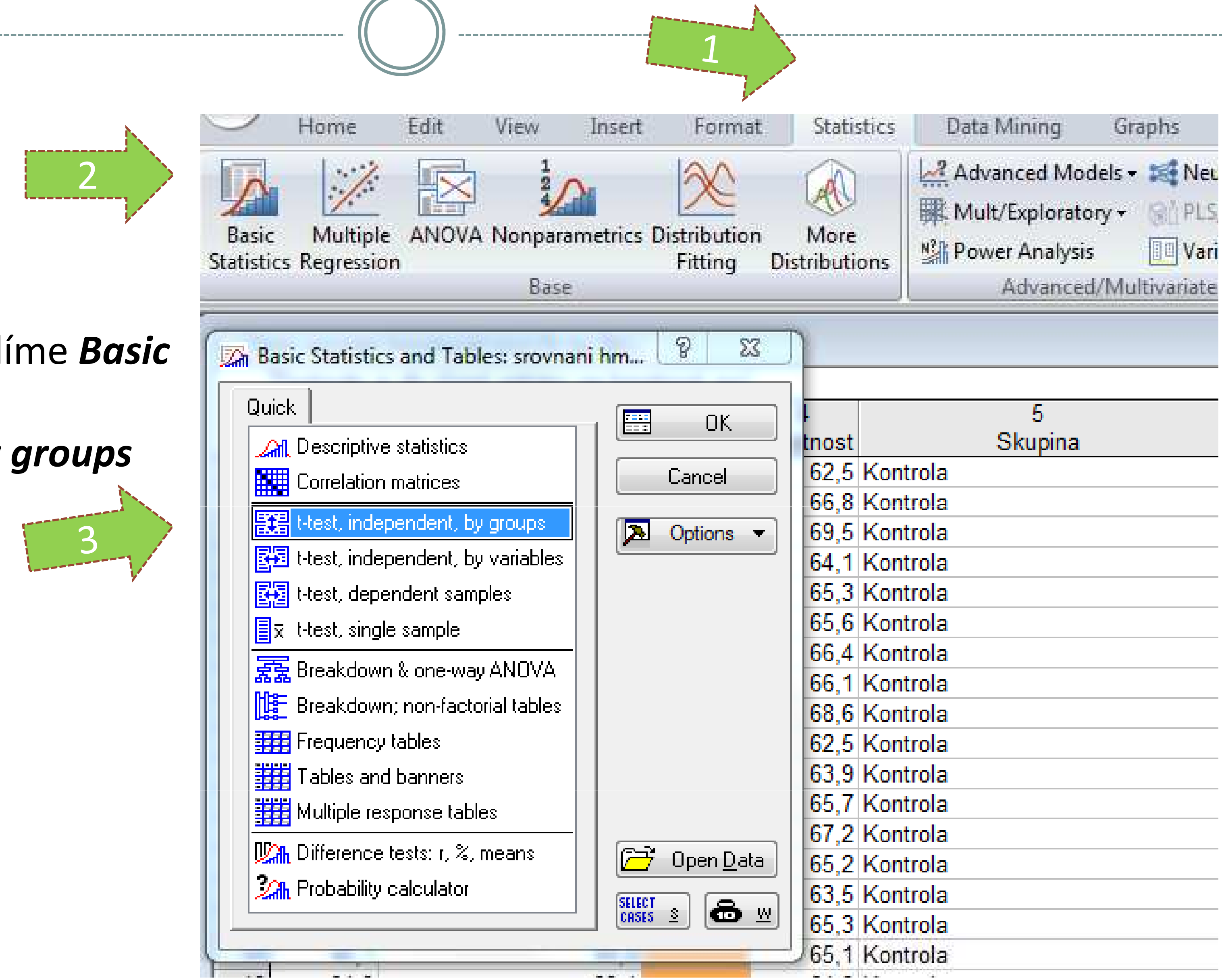
• Nejprve ověřte normalitu hmotnosti jednak ve skupině kontroly a ve skupině se zvýšenou potravou



•V obou případech se tečky odchylují od přímky jenom málo a p-hodnoty S-W testu převyšují 0,05. Předpoklad o normálním rozložení dat v obou skupinách je oprávněný.

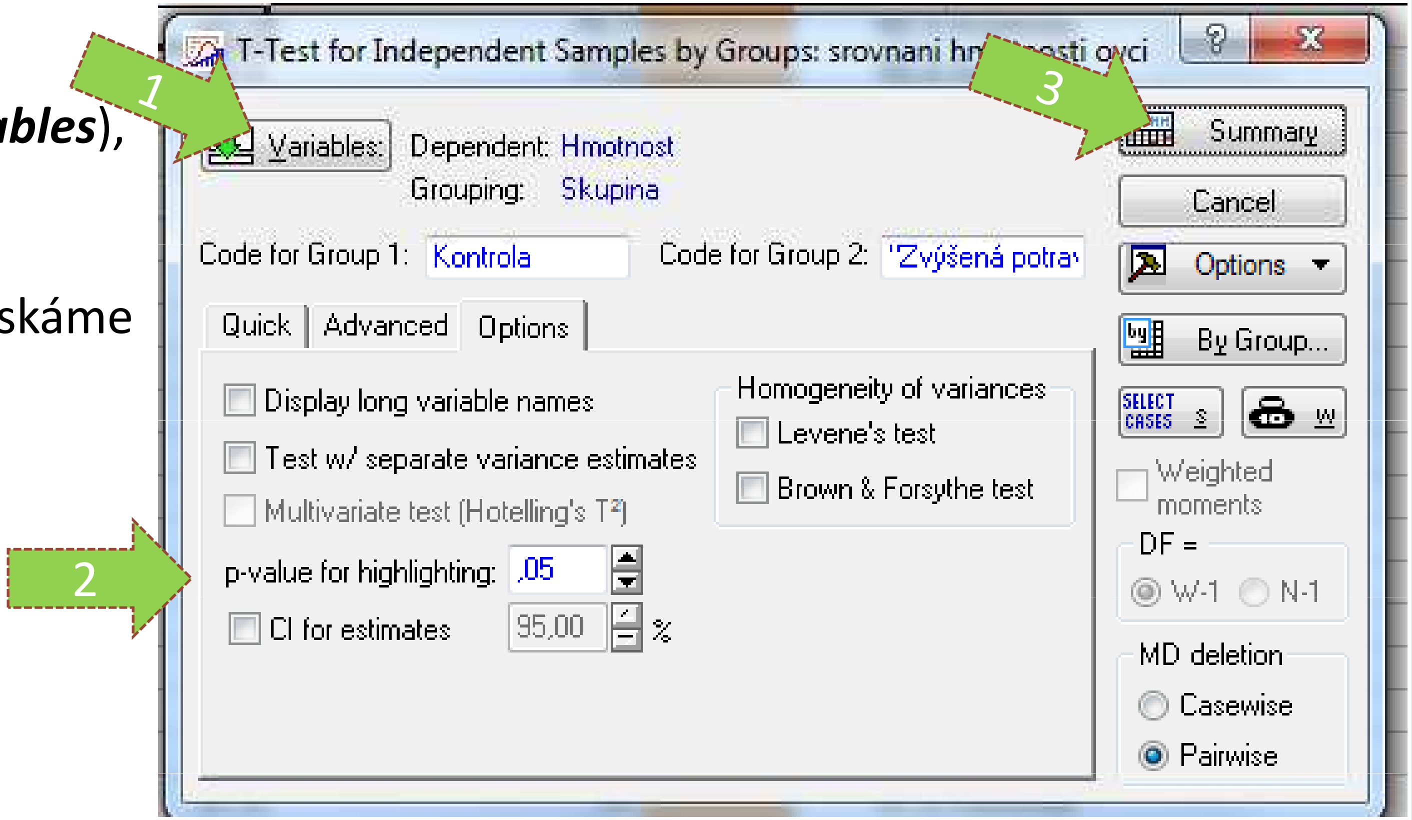
Příklad 2: Řešení v softwaru Statistica I

• V menu ***Statistics*** zvolíme ***Basic statistics*** ,vybereme ***t-test, independent, by groups***



Příklad 2: Řešení v softwaru Statistica II

- Zvolíme proměnné (***Variables***),
- Kliknutím na ***Summary*** získáme výstupy



Příklad 2: Řešení v softwaru Statistica III

•***POZOR: Výstupní tabulku vyhodnocujeme zezadu!!!***

Výběrový průměr u 1. skupiny

Výběrový průměr u 2. skupiny

Výběrová směrodatná odchylka u 2. skupiny

Rozsah výběru 1. skupiny

Rozsah výběru 2. skupiny

T-tests; Grouping: Skupina (srovnani hmotnosti ovci)
Group 1: Kontrola
Group 2: Zvýšená potrava

Variable	Mean Kontrola	Mean Zvýšená potrava	t-value	df	p	Valid N Kontrola	Valid N Zvýšená potrava	Std.Dev. Kontrola	Std.Dev. Zvýšená potrava	F-ratio Variances	p Variances
Hmotnost	65,77333	67,36667	-2,43226	52	0,018483	30	24	2,497162	2,252470	1,229066	0,617383

Hodnota testovacího kritéria
(pro test shody středních hodnot)

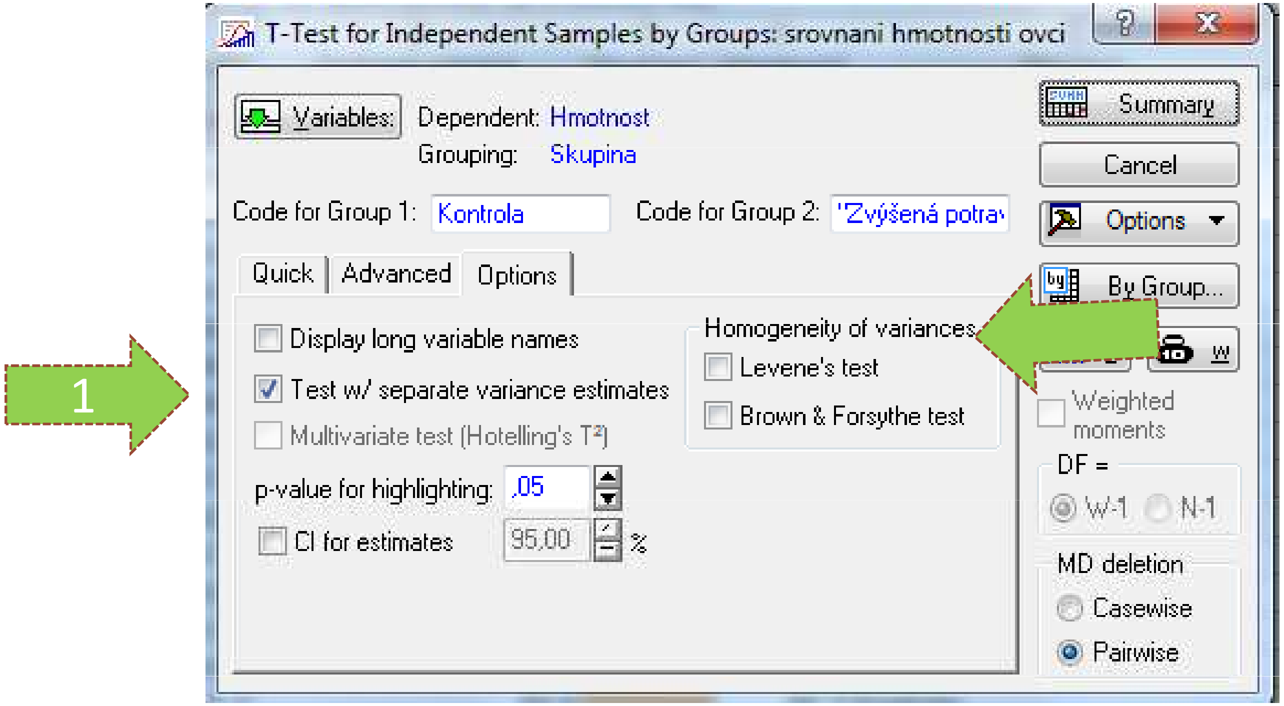
Počet stupňů volnosti

Testová statistika pro test shody rozptylů
(F-test)

Tyto sloupce lze interpretovat pouze pokud rozdíl mezi rozptyly byl neprůkazný !!!

Příklad 2: Řešení v softwaru Statistica IV, F-test

- <u>Pokud F-test prokázal odlišnost rozptylů</u>, je nutné na záložce ***Options*** odškrtnout ***Test w/separate variance estimates (t-test se samost. odhady rozptylů)***



- Chceme-li homogenitu rozptylů testovat ještě jiným testem, než F-testem, vybereme test z nabídky ***Homogeneity of variances***

Párové dvouvýběrové testy – předpoklady

- Skupiny dat jsou spojeny přes objekt měření, příkladem může být měření parametrů pacienta před léčbou a po léčbě (nemusí jít přímo o stejný objekt, dalším příkladem mohou být např. krysy ze stejné linie).
- Oba soubory musí mít shodný počet hodnot, protože všechna měření v jednom souboru musí být spárována s měřením v druhém souboru. Při vlastním výpočtu se potom počítá se změnou hodnot (diferencí) subjektů v obou souborech.
- Před párovým testem je vhodné ověřit si zda existuje vazba mezi oběma skupinami – vynesení do grafu, korelace.

Existuje několik možných designů experimentu, stručně lze sumarizovat:

1. pokus je párový a jako párový se projeví
2. párové provedení pokusu – párově se neprojeví
   - možná párovost není
   - špatně provedený pokus – malé n, velká variabilita, špatný výběr jedinců
3. čekali jsme nezávislé a jsou
4. čekali jsem nezávislé a nejsou
   - vazba
   - náhoda

Párový dvouvýběrový t-test

- Tento test nemá žádné předpoklady o rozložení vstupních dat, protože je počítán až na základě jejich diferencí.
- Tyto diference by měly být normálně rozloženy a otázkou v párovém t-testu je, zda se průměrná hodnota diferencí rovná nějakému číslu, typicky jde o srovnání s nulou jako důkaz neexistence změny mezi oběma spárovanými skupinami.
- V podstatě jde o one sample t-test, kde místo rozdílu průměru vzorku a cílové populace je uveden průměr diferencí a srovnávané číslo (0 v případě otázky, zda není rozdíl mezi vzorky).
- Pro srovnání s 0 (testovou statistikou je t rozložení): $t = \frac{\bar{D}}{s}\sqrt{n} \qquad \upsilon = n - 1$
- Někdy je obtížné rozhodnout, zda jde nebo nejde o párové uspořádání, párový test by měl být použit pouze v případě, že můžeme potvrdit vazbu (korelace, vynesení do grafu), jedním z důvodů proč toto ověřovat je fakt, že v případě párového t-testu není nutné brát ohled na variabilitu původních dvou souborů, tento předpoklad však platí pouze v případě vazby mezi proměnnými. Výpočet obou typů testů se vlastně liší v použité s, jednou jde o s diferencí, v druhém případě o složený odhad rozptylu obou souborů.
- Zda je párové uspořádání efektivnější lze určit na základě:
    - Síly vazby
    - Je-li s_D výrazně menší než s_{x1-x2}
- Závislost je možné rozepsat pomocí vzorce: $s_D^2 \cong \sigma_{x_1}^2 + \sigma_{x_2}^2 - 2Cov(x_1; x_2)$
- v případě Cov=0, tedy v případě neexistence vazby pak s_D^2 odpovídá součtu původních rozptylů, tedy přibližně S_{x1-x2}.

Příklad 3: Párový dvouvýběrový t-test

Byl prováděn pokus s dietou u 18 diabetických krys, každá krysa byla vystavena dvěma dietám. Protože každá krysa absolvovala obě diety, jde o párové uspořádání, kdy hodnoty v obou pokusech jsou spojeny přes pokusné zvíře. Zjistěte, zda testovaná dieta způsobí změnu hmotnosti u krys.

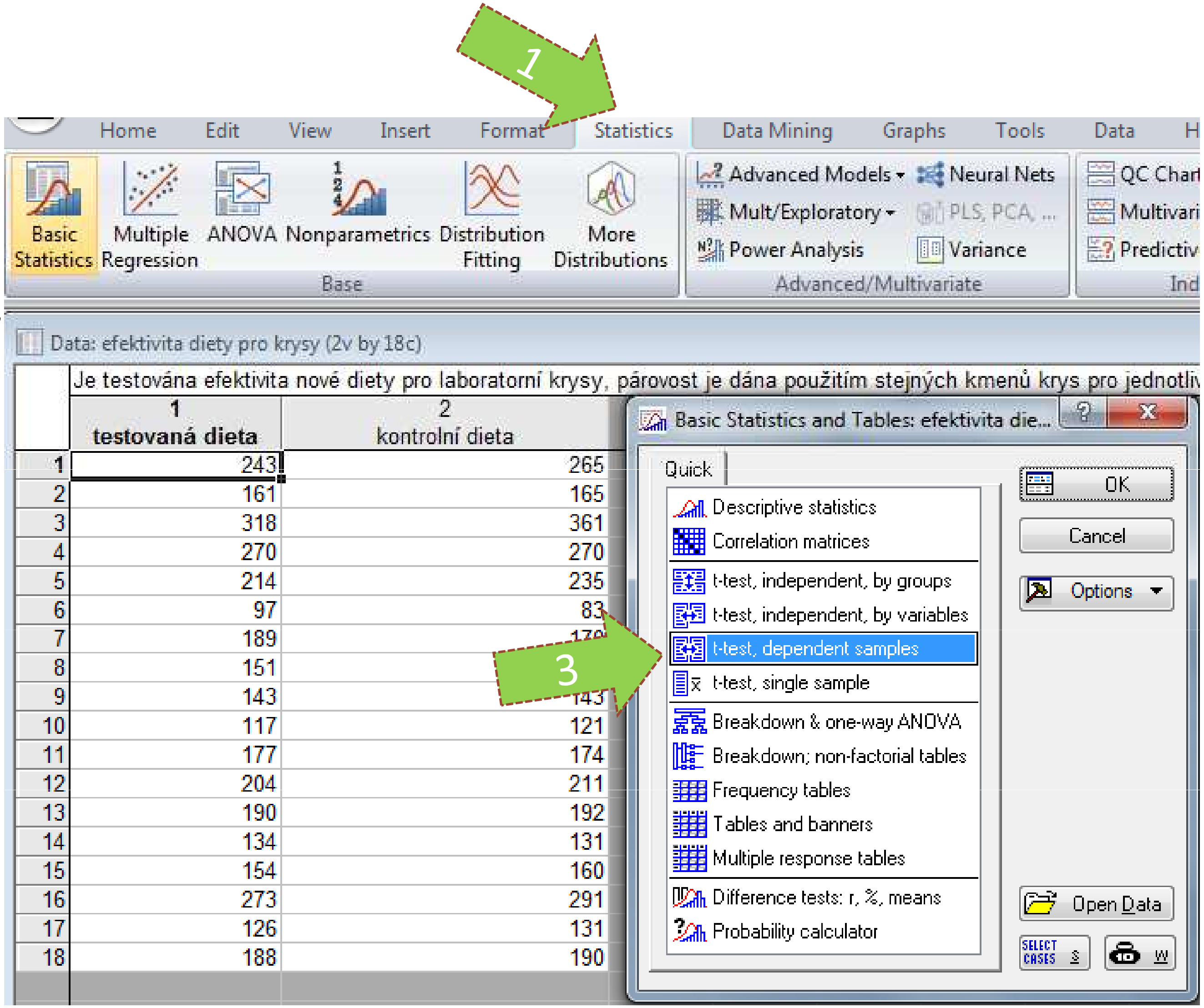
1. Nulová hypotéza zní, že skutečný průměrný rozdíl mezi oběma dietami je 0, alternativní hypotéza zní, že to není 0.
2. Pro každou krysu je spočítán rozdíl mezi hmotnosti při obou dietách a měly by být ověřeny předpoklady pro one sample t-test – tedy alespoň přibližně normální rozložení.
3. Je spočítána testová charakteristika, výpočet vlastně probíhá jako one-sample t-test, kde je zjišťována významnost průměru diferencí obou souborů jako rozdíl mezi touto hodnotou a nulou (nula je hodnota, kterou by průměrná diference měla nabývat, pokud platí nulová hypotéza). T=-1,72 s 17 stupni volnosti, skutečná hodnota p=0,102 a tedy na hladině p=0,05 nemůžeme nulovou hypotézu zamítnou

$$t = \frac{rozdíl\_průměru\_vzorku\_a\_populace}{SE(průměru)} = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{\overline{x} - \mu}{s}\sqrt{n}$$

4. Závěrem můžeme říci, že nulová hypotéza neexistence rozdílu mezi oběma dietami nebyla zamítnuta, což znamená, že testovaná dieta nemá významný vliv na snížení hmotnosti.

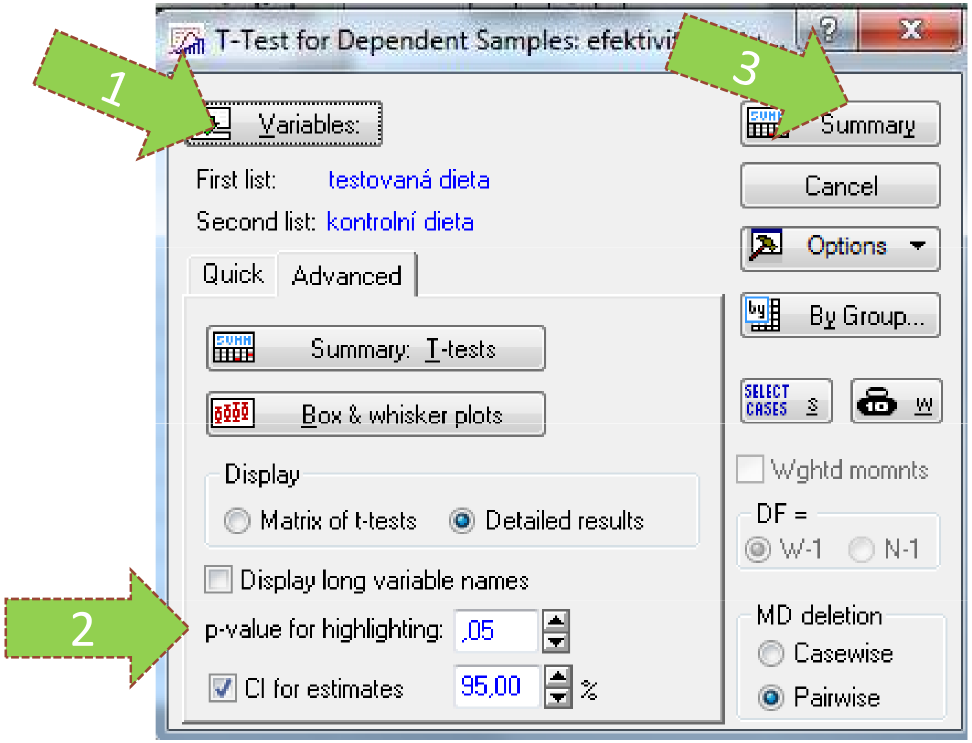
Příklad 3: Řešení v softwaru Statistica I

- V menu ***Statistics*** zvolíme ***Basic statistics*** ,vybereme ***t-test, dependent samples***



Příklad 3: Řešení v softwaru Statistica II

- Zvolíme proměnné (***Variables***),
- Kliknutím na ***Summary*** získáme výstupy



Příklad 3: Řešení v softwaru Statistica III

Výběrový průměr

Výběrová směrodatná odchylka

Počet pozorování

T-test for Dependent Samples (efektivita diety pro krysy)
Marked differences are significant at p < ,05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
testovaná dieta	186,0556	59,52011						
kontrolní dieta	191,7222	69,65022	18	-5,66667	13,91994	-1,72714	17	0,102266

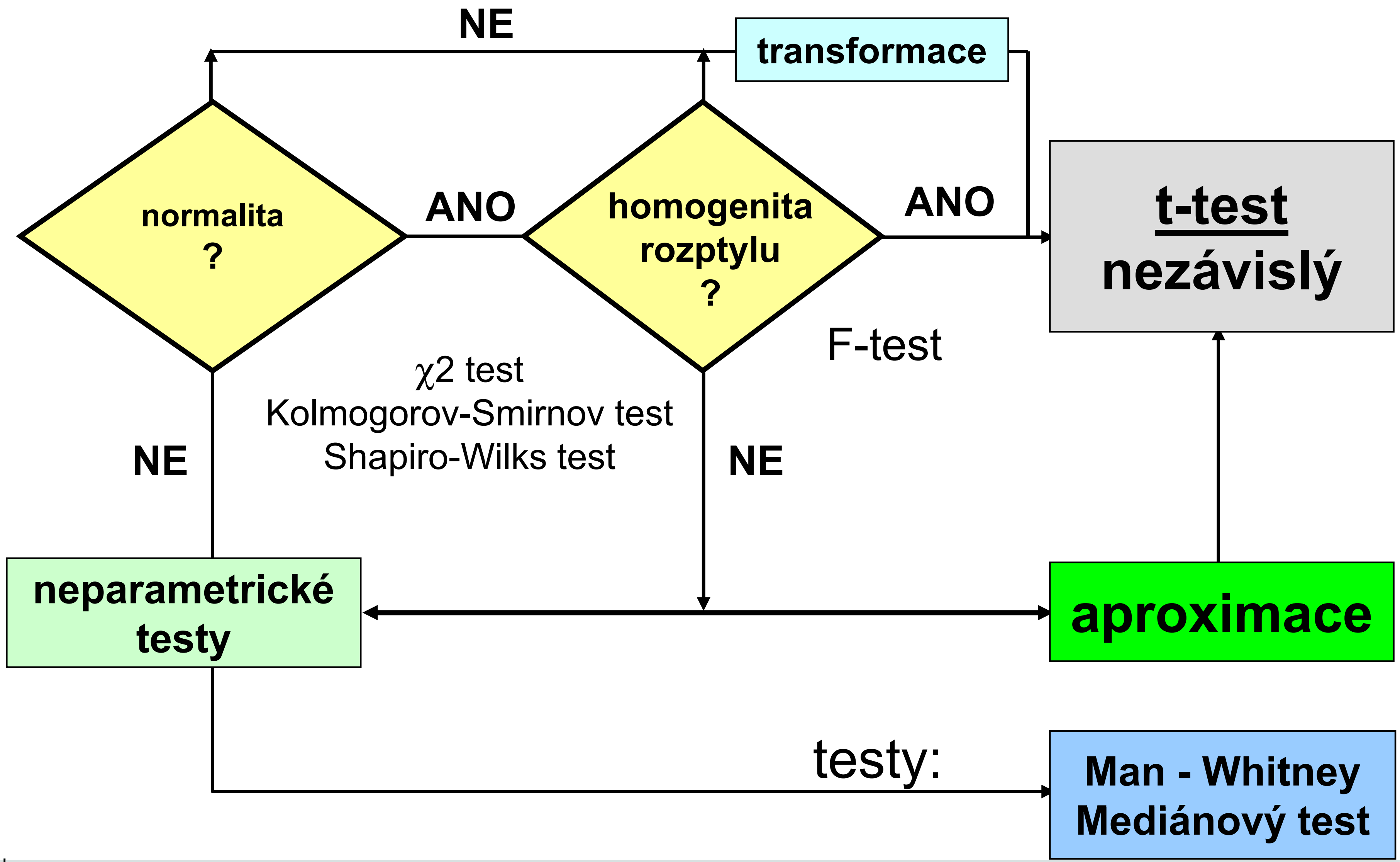
Hodnota testovacího kritéria

Průměrná hodnota diferencí

Výběrová směrodatná odchylka diferencí

Dvouvýběrové testy: schéma analýzy

Nezávislé uspořádání



Dvouvýběrové testy: schéma analýzy

Párové uspořádání