ASTAc/01 Biostatistika
4. cvičení

Opakování – základy testování hypotéz
Shrnutí statistických testů
Jednovýběrové parametrické testy
Dvouvýběrové parametrické testy

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

Opakování

1. Co je **nulová** a **alternativní hypotéza**?
2. Co je **testová statistika**?
3. Jakými **způsoby** můžeme **testovat hypotézy**?
4. Co je **chyba I. a II. druhu**?
5. Co vyjadřuje **p-hodnota**?
6. Jaký je rozdíl mezi **parametrickými** a **neparametrickými** testy?
7. Jaký je rozdíl mezi **jednovýběrovými** a **dvouvýběrovými** testy?
8. Jaký je rozdíl mezi **jednostrannými** a **oboustrannými** testy?
9. Jaký je rozdíl mezi **párovými** a **nepárovými** testy?

Shrnutí statistických testů

Shrnutí statistických testů

Typ srovnání	Nulová hypotéza	Parametrický test	Neparametrický test
1 výběr dat vs. referenční hodnota	Střední hodnota je rovna zvolené referenční hodnotě.	jednovýběrový t-test / z-test	Wilcoxonův test; znaménkový test
2 nezávislé skupiny dat (test shody středních hodnot)	Střední hodnoty/rozdělení se mezi skupinami neliší.	nepárový t-test	Mannův-Whitneyův test; Mediánový test
2 nezávislých skupin dat (test shody rozptylů = homoskedasticity)	Rozptyl obou skupin je shodný.	F-test	Levenův test
2 párově závislé výběry dat	Rozdíl (diference) párových hodnot je nulový.	párový t-test	Wilcoxonův test; znaménkový test
Shoda rozdělení výběru s teoretickým rozdělením	Rozdělení dat odpovídá teoretickému (vybranému) rozdělení.	test dobré shody (χ^2 test)	Shapirův-Wilkův test; Kolmogorovův-Smirnovův test; Lilieforsův test
Více než 2 skupin nepárově (test shody středních hodnot)	Střední hodnoty/rozdělení se mezi skupinami neliší.	ANOVA	Kruskalův-Wallisův test; Mediánový test
Korelace	Neexistuje vztah mezi hodnotami dvou výběrů.	Pearsonův korelační koeficient	Spearmanův korelační koeficient

Základní rozhodování o výběru statistických testů

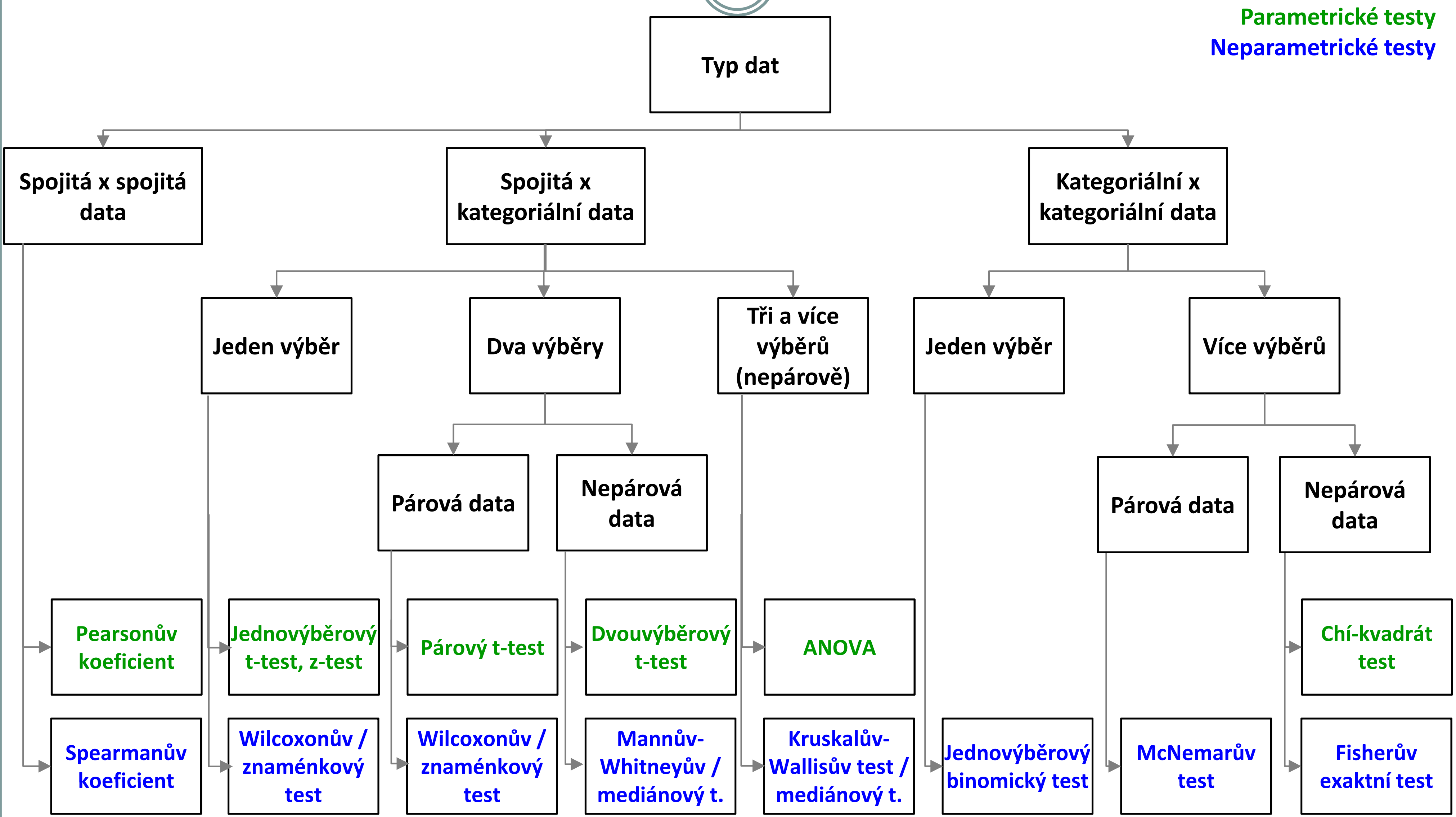
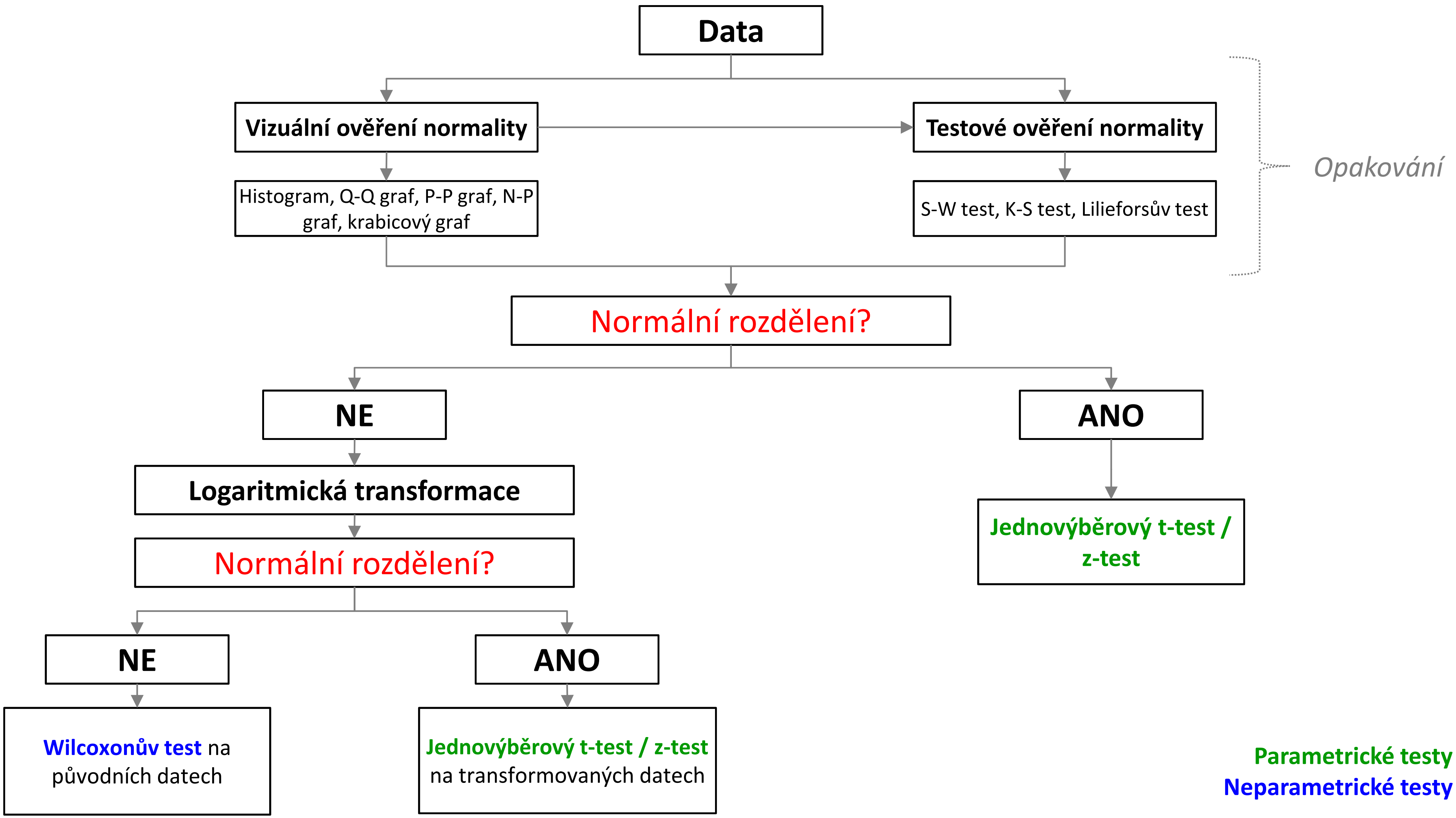


Schéma při testování pomocí jednovýběrových testů



Vytvořil Institut biostatistiky a analýz, Masarykova univerzita
E. Janoušová, L. Dušek

Schéma při testování pomocí párových testů

Data

Normální rozdělení?
(normální rozdělení diferencí!)

NE

Párový Wilcoxonův test / znaménkový test

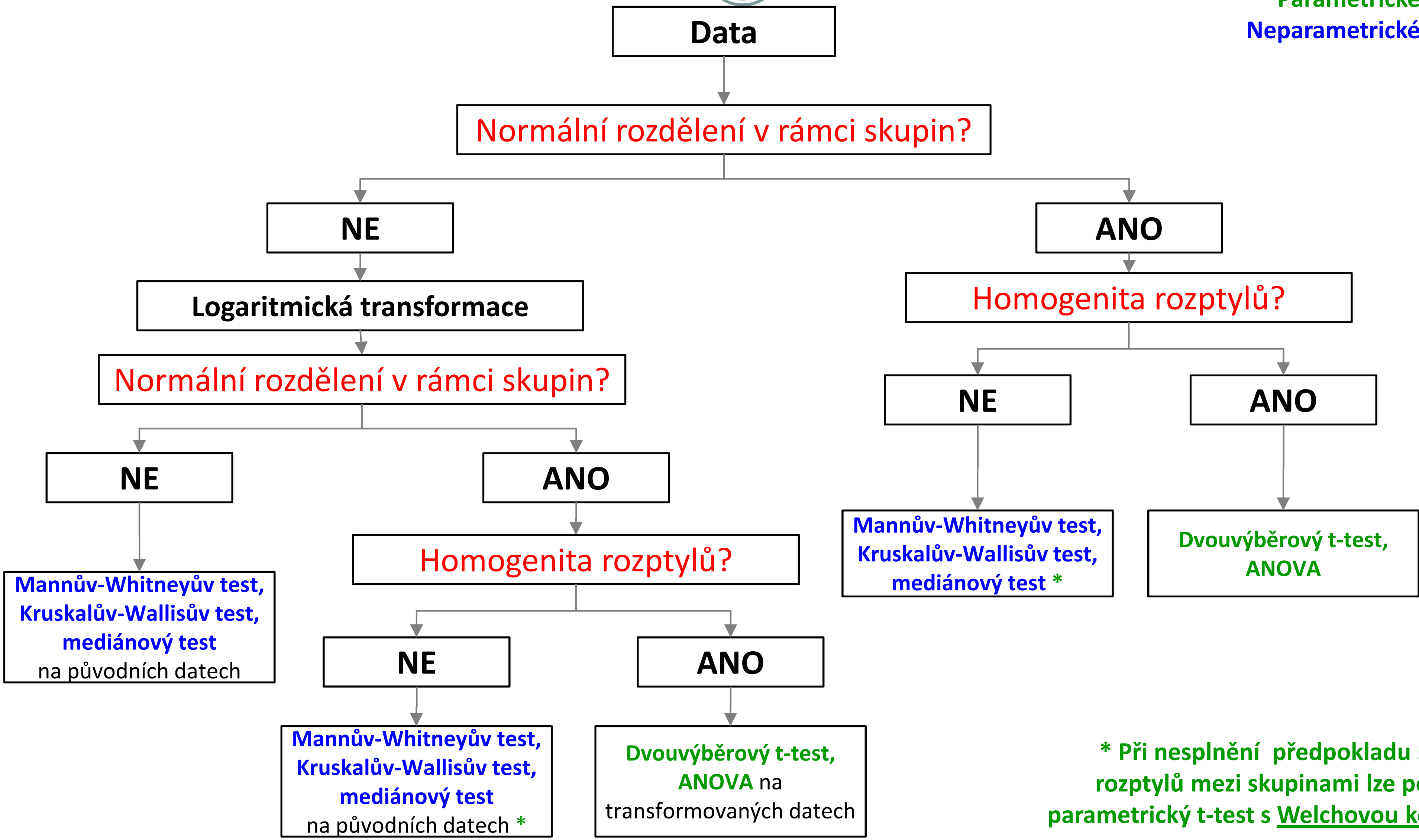
ANO

Párový t-test

Parametrické testy
Neparametrické testy

Schéma při testování 2 a více skupin

Parametrické testy
Neparametrické testy



* Při nesplnění předpokladu shody rozptylů mezi skupinami lze použít i parametrický t-test s Welchovou korekcí

Parametrické testy

Parametrické testy

- Předpoklad: <u>normalita dat</u>
- ***<u>Jednovýběrový z-test</u>*** (porovnání základního a výběrového souboru, známe střední hodnotu <u>a rozptyl</u> základního souboru)
- ***<u>Studentův t-test</u>*** (testování rozdílů dvou středních hodnot)
  1. ***Jednovýběrový t-test*** (porovnání základního a výběrového souboru, známe střední hodnotu ale <u>neznáme rozptyl</u> základního souboru; nahrazujeme jej výběrovým rozptylem našich dat)
  2. ***Dvouvýběrový t-test*** (porovnání dvou výběrových souborů, neznáme střední hodnotu základního souboru):

     - ***párový*** (závislé výběry)
     - ***nepárový*** (nezávislé výběry)
- ***<u>F-test</u>*** (testování rozdílů dvou rozptylů)

1. Statistické testy o parametrech jednoho výběru

Jednovýběrový t-test

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.

Jednovýběrové testy I

V případě „one sample“ testů jde o srovnání jednoho výběru dat (proto „one sample“) s cílovou populací. Pro parametrické testy musí mít výběr normální rozložení.

Průměr – cílová vs. výběrová populace

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

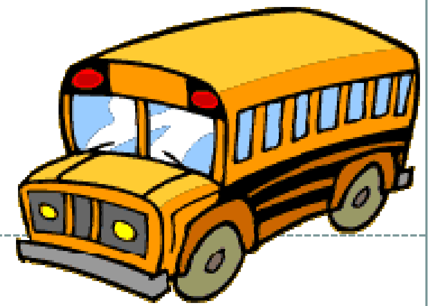
H_0	H_A	Testová statistika	Interval spolehlivosti
$\bar{x} \leq \mu$	$\bar{x} > \mu$	t	$t > t_{1-\alpha}^{(n-1)}$
$\bar{x} \geq \mu$	$\bar{x} < \mu$	t	$t < t_{\alpha}^{(n-1)}$
$\bar{x} = \mu$	$\bar{x} \neq \mu$	t	$\|t\| > t_{1-\alpha/2}^{(n-1)}$

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

$t_{1-\alpha/2}$ (n-1) = kvantil Studentova *t*-rozdělení pro dané stupně volnosti (n-1) a zvolené α

* Pokud známe rozptyl základního souboru, tento nahradí výběrový odhad rozptylu z dat a výsledná statistika Z∾ N(0,1)

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.
- **<u>Postup:</u>**

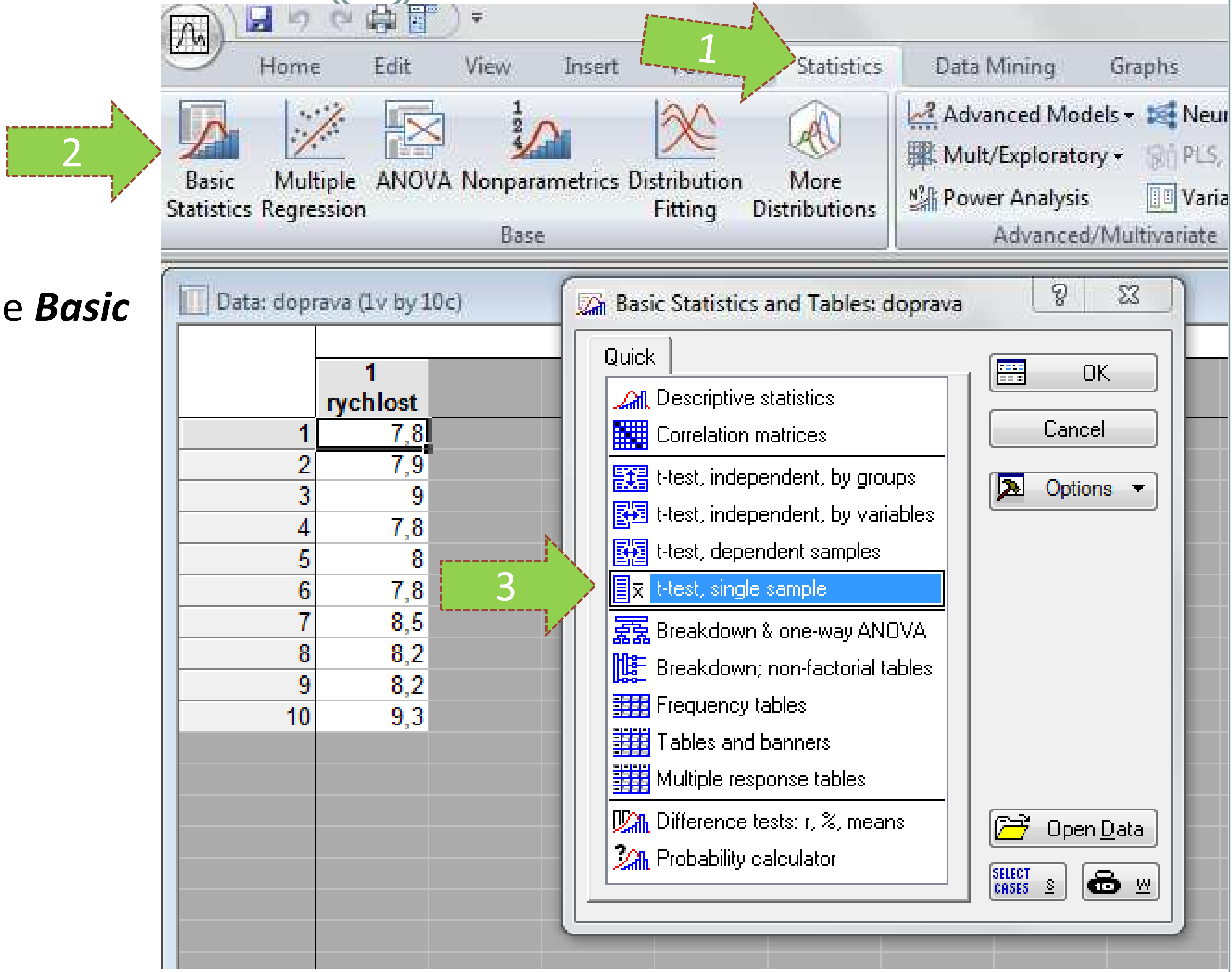
1. Na hladině významnosti 0,05 testujeme **hypotézu H_0:** $\mu = 8$ **, proti H_A :** $\mu \neq 8$
2. Vypočteme **aritmetický průměr** a **rozptyl výběrového souboru.**
3. Vypočteme **testovou statistiku t:**

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{8{,}25 - 8}{0{,}530}\sqrt{10} = 1{,}492$$

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***



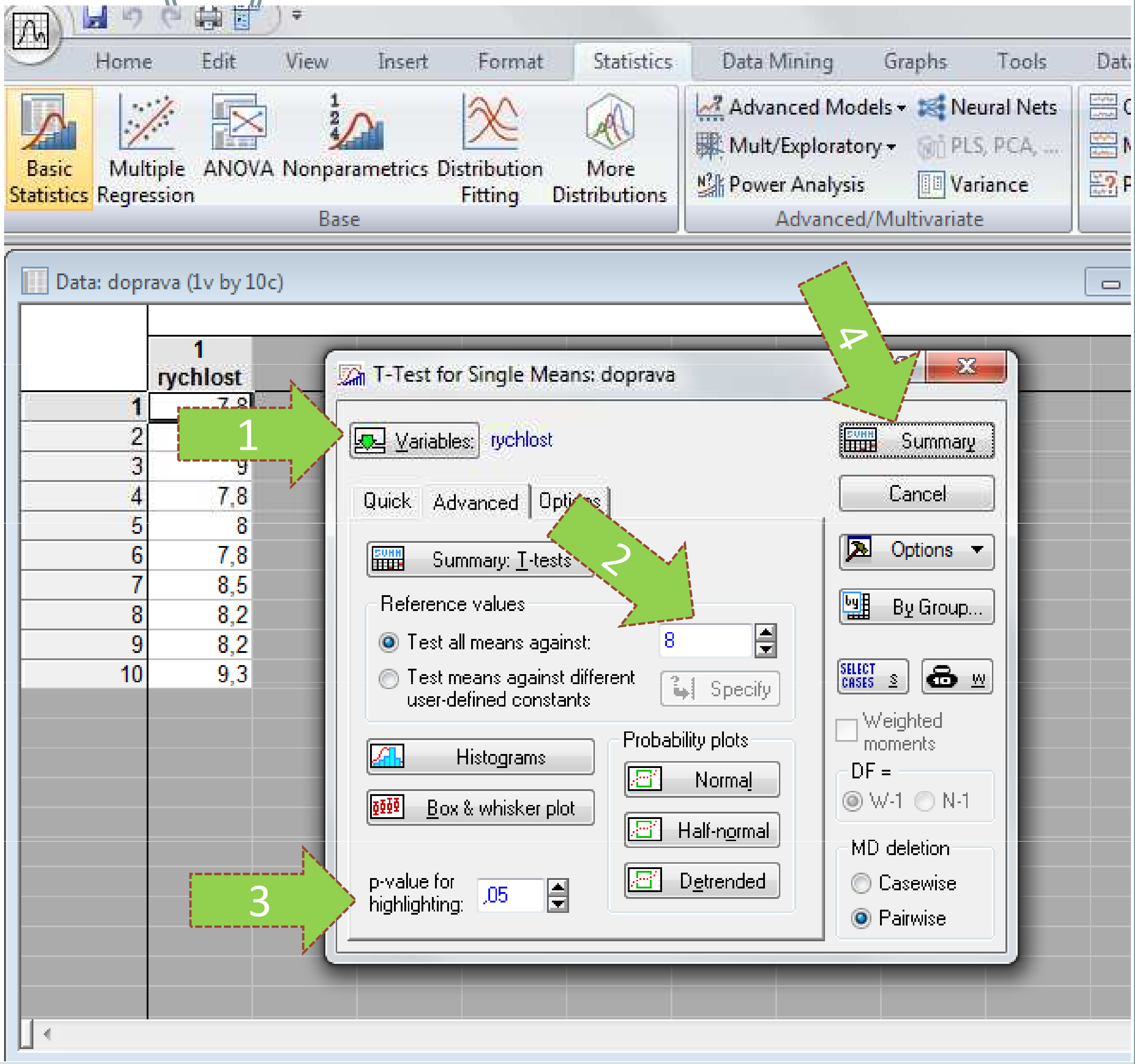
Řešení v softwaru Statistica II

- 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-hodnoty lze změnit

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



Řešení v softwaru Statistica III

Výběrový průměr (průměr pozorovaných dat)

Rozsah výběru

Standardní chyba

Hodnota testovacího kritéria

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 (pozorovaných dat)

POZOR: Platí pro oboustranný test!!!

Referenční konstanta-předpokládaná velikost střední hodnoty

2. Statistické testy o parametrech dvou výběrů

Dvouvýběrový párový a nepárový t-test

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

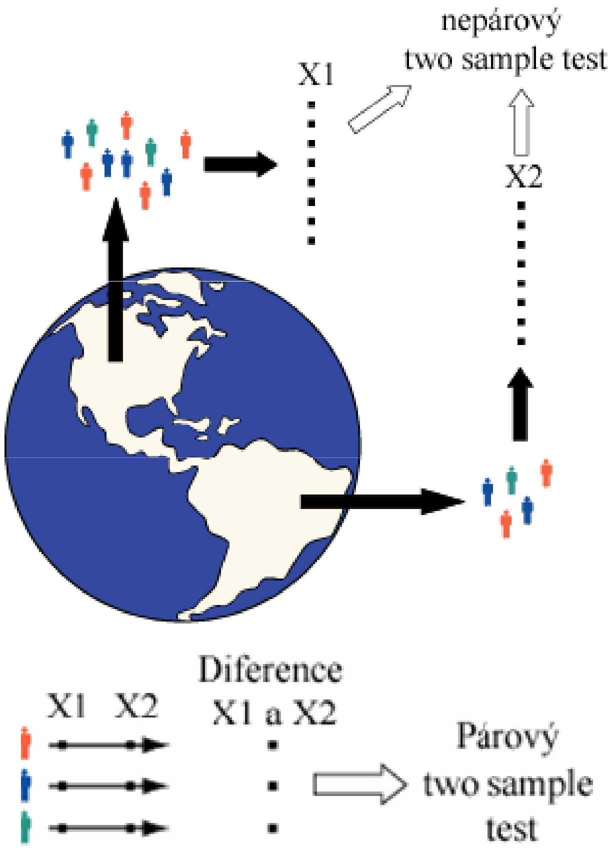
Anotace

- 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 dvouvýběrových testů srovnáváme spolu dvě rozložení. Jejich základním dělením je podle designu experimentu na testy párové a nepárové.



- Základním testem pro srovnání dvou <u>nezávislých</u> rozložení spojitých čísel je **nepárový dvouvýběrový t-test**

- Základním testem pro srovnání dvou <u>závislých</u> rozložení spojitých čísel je **párový dvouvýběrový t-test**

Dvouvýběrové testy: párové a nepárové II

Data

X_1 X_2

Nezávislé uspořádání

X_1 X_2

$$H_0 : \mu_1 = \mu_2$$

<u>Charakteristiky výběrů hodnotíme nezávisle na sobě</u>:

Výběr x_1: n_1 (počet vzorků), $\overline{x}_1$ (výběr. průměr), s_1^2 (výběr. rozptyl)

Výběr x_2: n_2 (počet vzorků), $\overline{x}_2$ (výběr. průměr), s_2^2 (výběr. rozptyl)

Pozor: u nepárového designu se počet vzorků výběrů může lišit!

Párové uspořádání

$X_1 - X_2 = D$

$$H_0 : \overline{D} = 0$$

<u>Charakteristiky počítáme z diferencí (D) párových pozorování</u>:

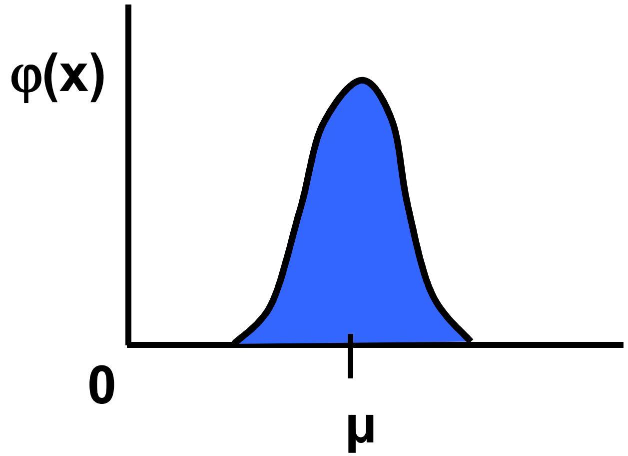
Výběr D: n_D (počet párů vzorků), $\overline{x}_D$ (výběr. průměr), s_D^2 (výběr. rozptyl)

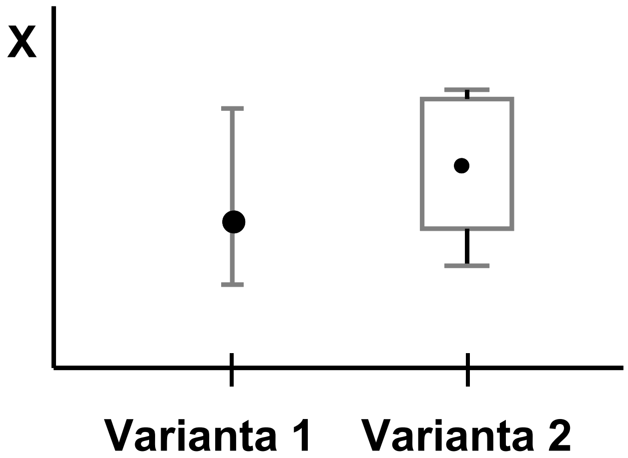
Pozor: u párového designu se počet vzorků výběrů (n_1 a n_2) nesmí lišit!

Pozn.: při párovém uspořádání převádíme na design jednovýběrových testů

Předpoklady nepárového dvouvýběrového t-testu

- Náhodný výběr subjektů jednotlivých skupin z jejich cílových populací
- Nezávislost obou srovnávaných vzorků
- Přibližně **normální rozložení proměnné v rámci skupin** (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** („homoskedasticita rozptylu“). 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 (oboustranný – „ 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
$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F = \frac{s_1^2}{s_2^2}$
$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F = \frac{s_2^2}{s_1^2}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$F = \frac{\max(s_1^2; s_2^2)}{\min(s_1^2; s_2^2)}$

Oboustranný F-test

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** (oboustranná); v případě shodných rozptylů je vše v pořádku a je možné pokračovat ve výpočtu t-testu, v opačném případě není vhodné test počítat nebo zvolit Welchovu korekci.

Nepárový dvouvýběrový t-test – výpočet II

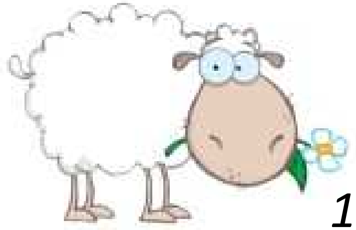
3. Výpočet **testové statistiky t-testu** (stupně volnosti jsou $\upsilon = n_1 + n_2 - 2$):

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

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$ 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 skupiny 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 dvouvýběrového nepárového t-testu, můžeme spočítat testovou statistiku, 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.

$$t = \frac{Rozdil.prumeru}{SE(rozdil.prumeru)} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad \upsilon = n_1 + n_2 - 2$$

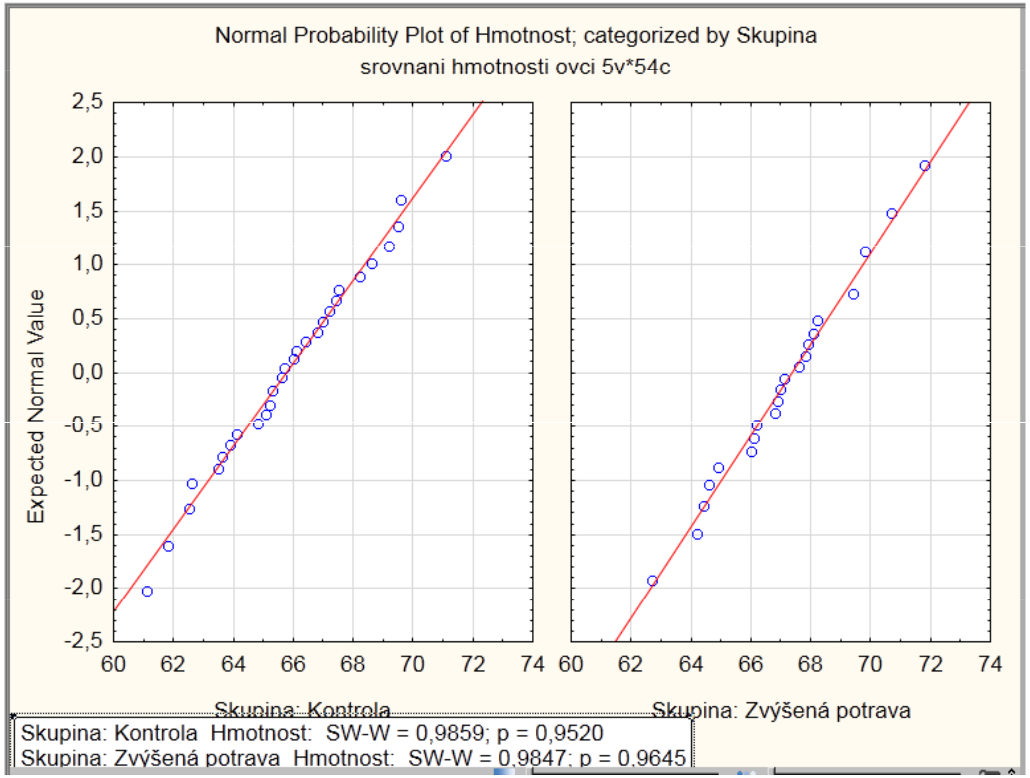
- Pro rozdíl mezi oběma soubory jsou spočítány 95% intervaly spolehlivosti jako 1,59±2.01*(0,655) kg, což odpovídá rozsahu 0,28 až 2,91 kg. <u>To, že interval spolehlivosti nezahrnuje 0 je dalším potvrzením, že mezi skupinami je významný rozdíl</u> – jde o další způsob testování významnosti rozdílů mezi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

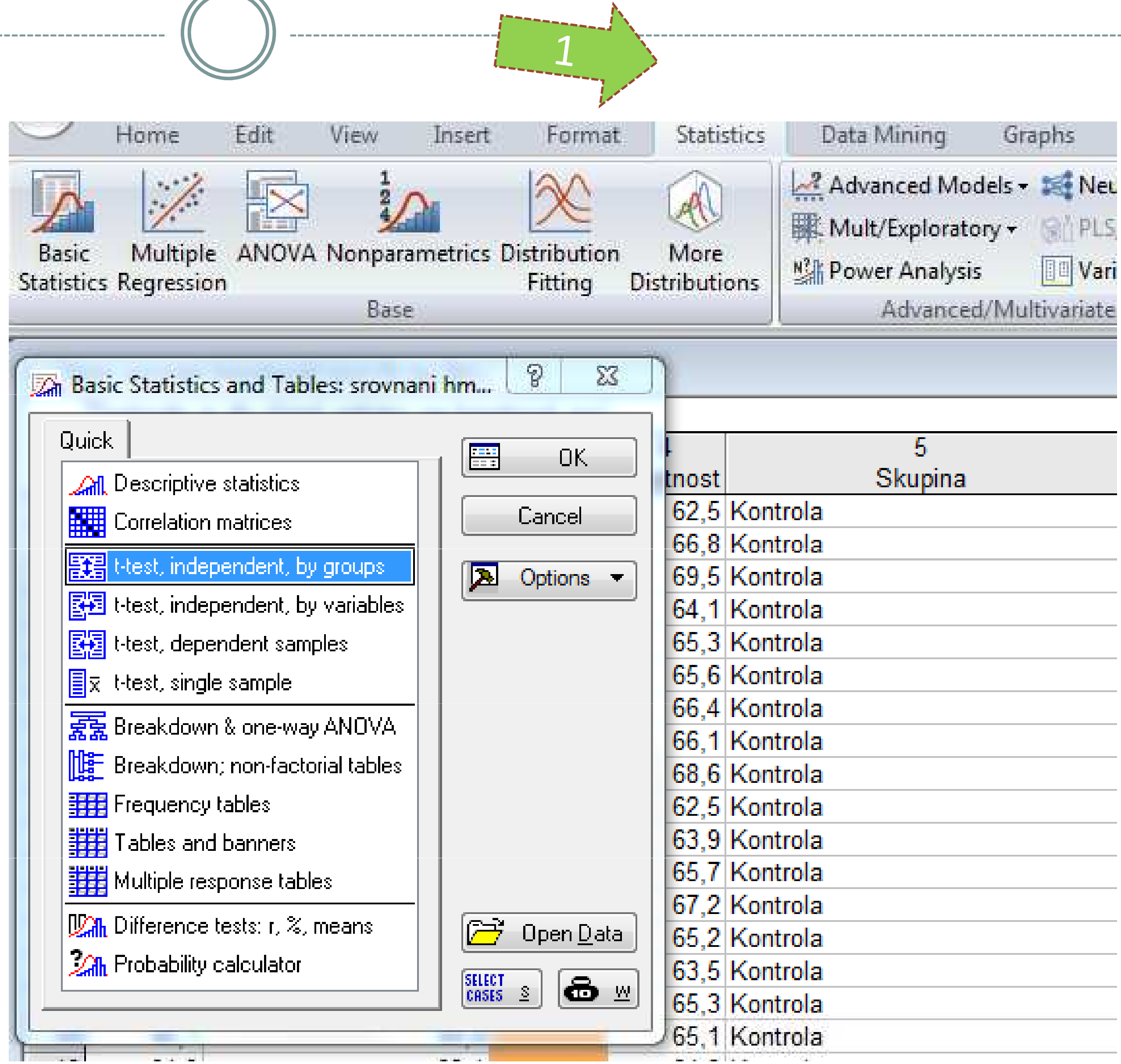
• Nejprve ověřte normalitu hmotnosti jednak ve skupině kontroly a ve skupině se zvýšenou potravou



•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.

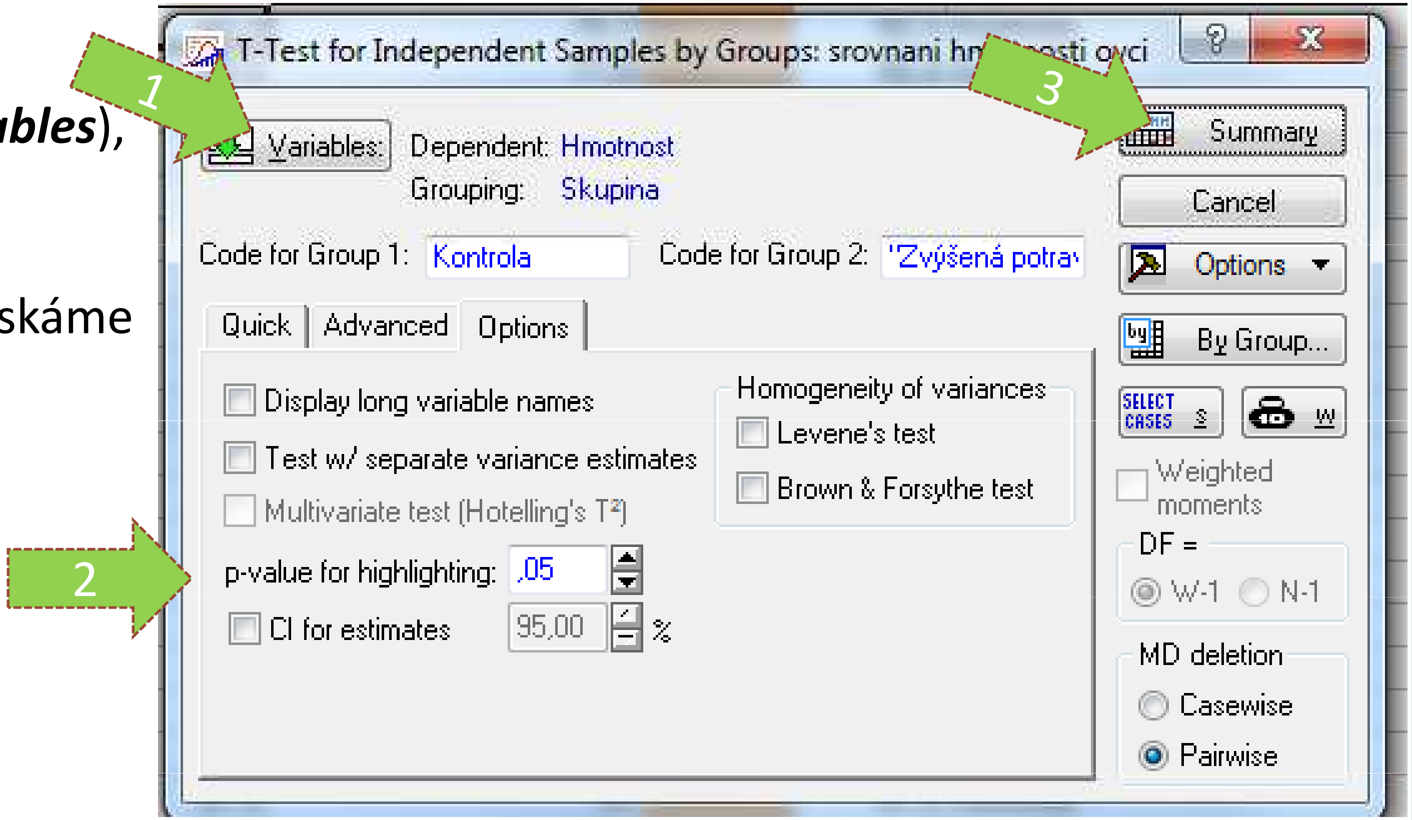
Příklad 2: Řešení v softwaru Statistica I

- 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 testové statistiky
(pro test shody středních hodnot)

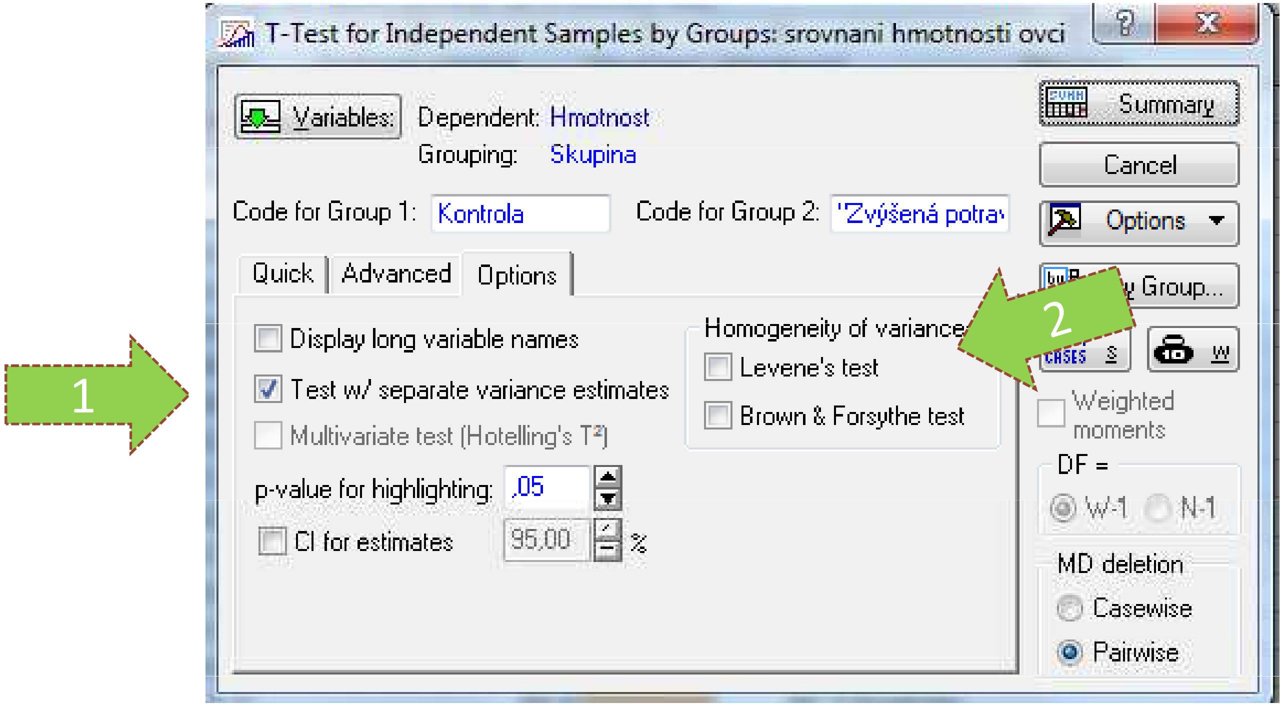
Počet stupňů volnosti

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

- Pokud F-test prokázal odlišnost rozptylů, 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***

Příklad 3: Párový dvouvýběrový t-test

Byl prováděn pokus s dietou u 18 diabetických krys, každá krysa byla vystavena dvěma dietám (jedné nové speciální a jedné kontrolní dietě). 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 (zda se liší hmotnost krys po nové speciální a po kontrolní dietě).

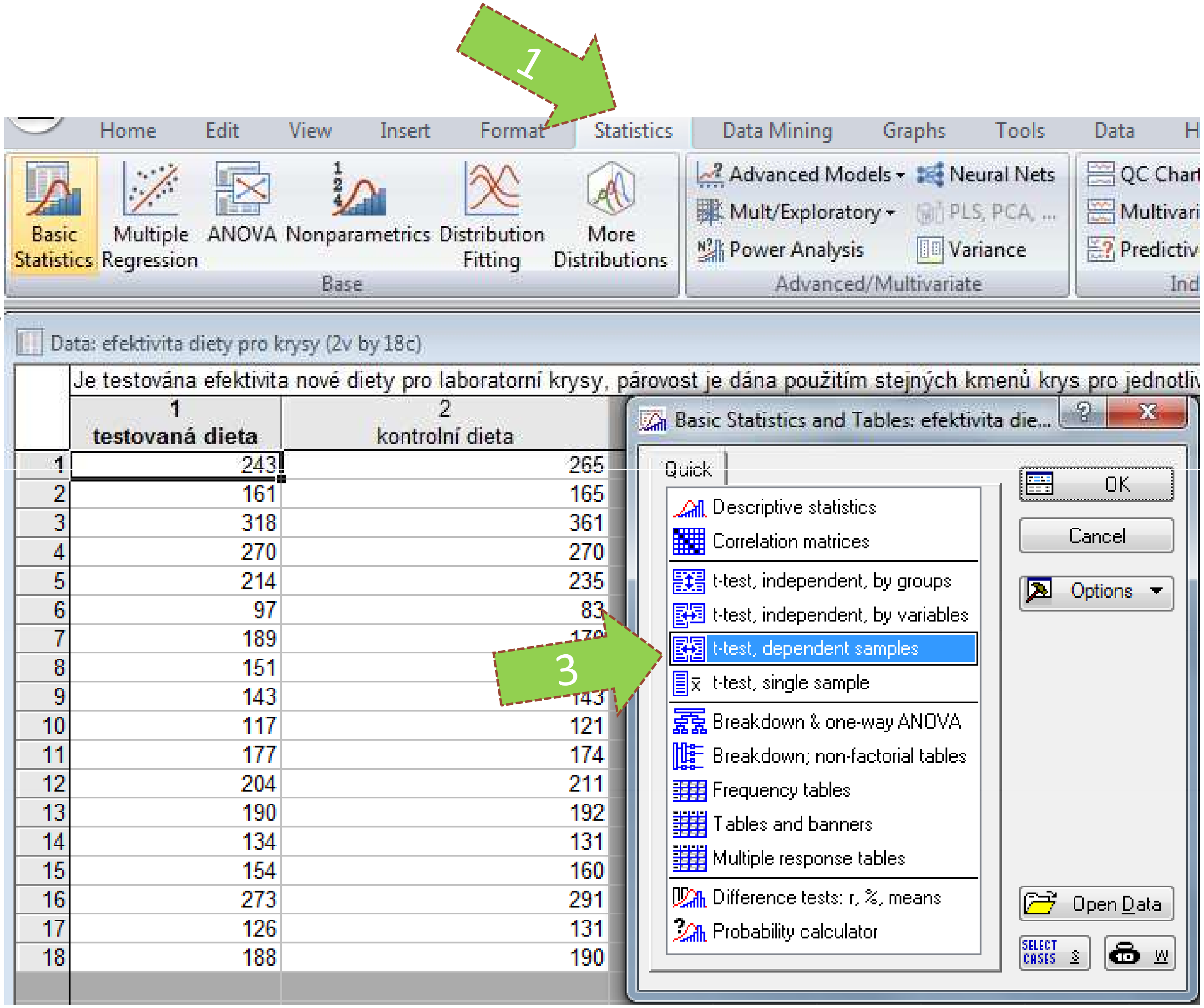
1. **Nulová hypotéza** zní, že skutečný průměrný rozdíl v hmotnosti krys po speciální a kontrolní dietě je nulový (speciální dieta nevedla ke změně hmotnosti ve srovnání s kontrolní dietou), **alternativní hypotéza** zní, že rozdíl hmotností je odlišný od nuly (speciální dieta vedla ke změně hmotnosti ve srovnání s kontrolní dietou).
2. Pro každou krysu je spočítán rozdíl hmotností naměřených po obou dietách a měly by být ověřeny předpoklady pro jednovýběrový t-test – alespoň přibližně **normální rozložení diferencí**.
3. Je spočítána **testová statistika**, výpočet vlastně probíhá jako jednovýběrový t-test, kde je zjišťována významnost průměru diferencí obou souborů jako rozdíl mezi touto hodnotou a nulou (0 je hodnota, kterou by průměrná diference měla nabývat, pokud platí nulová hypotéza). T=-1,72 s 17 stupni volnosti, skutečná p-hodnota=0,102 a tedy na hladině významnosti α=0,05 nemůžeme nulovou hypotézu zamítnout.

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

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

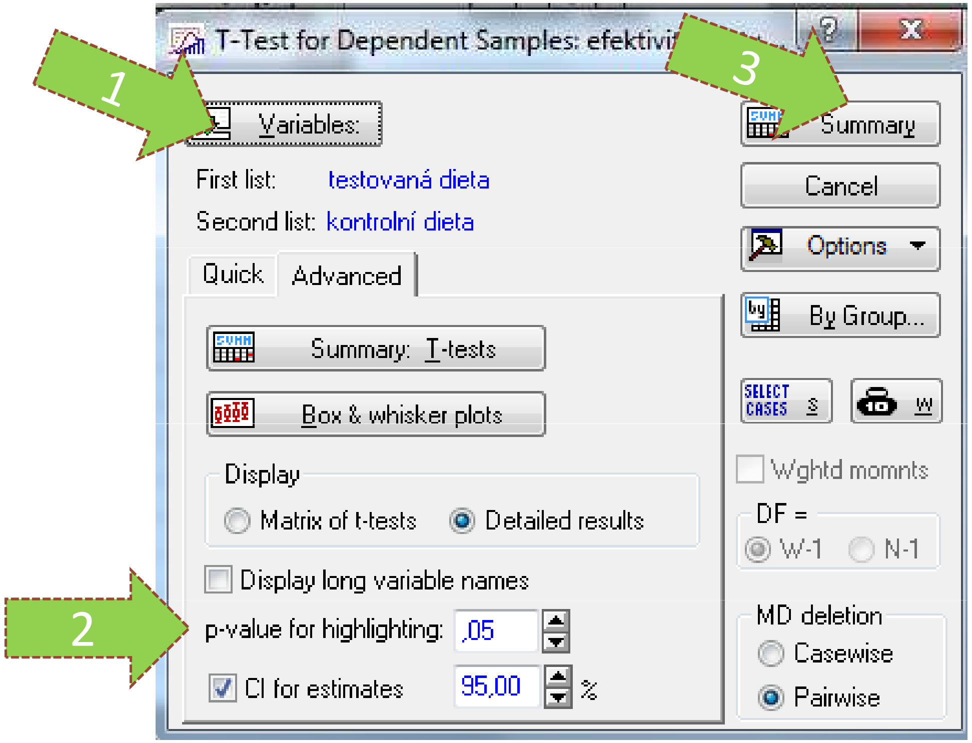
Příklad 3: Řešení v softwaru Statistica I

• 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

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