PARAMETRICKÉ TESTY

Testujeme rovnost průměru - **předpokladem normální rozdělení**

I) Jednovýběrový t-test

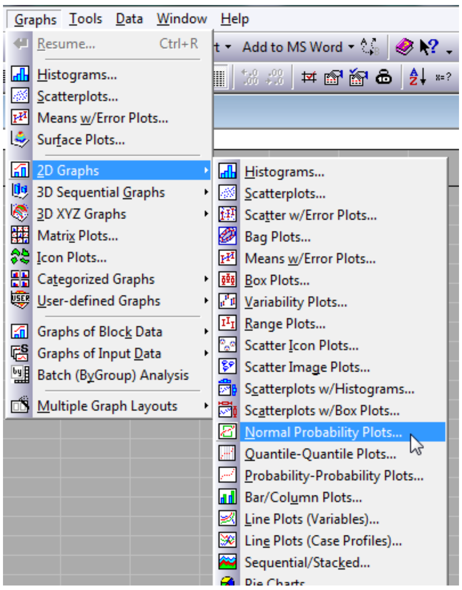
1) Měření Etalonu. Dataset - *mereni_etalonu.sta* - 9 měření etalonu srovnáváme s PŘEDPOKLÁDANOU HODNOTOU *10.*

H0: *Není statisticky významný rozdíl mezi naměřenými hodnotami a očekávanou hodnotou 10. (μ=10)*
H1: *Naměřené hodnoty se statisticky významně liší od očekávané hodnoty 10. (μ≠10)*

Krok A) Předpoklad testu -> pochází měření z normálního rozdělení? - Pokud by nesplňoval museli bychom využít neparametrickou obdobu tesu - Wilcoxonův jednovýběrový test.

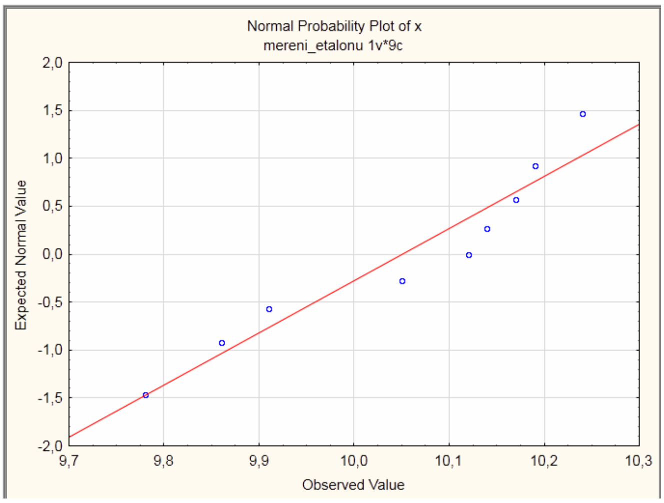
- Vybrat jednu ze 3 testovacích možností (ideální statisticky+histogram)- N-P plot, HISTOGTRAM, statisticky

- N-P plot - *Grafy -> 2D grafy -> Normální pravděpodobnostní grafy*

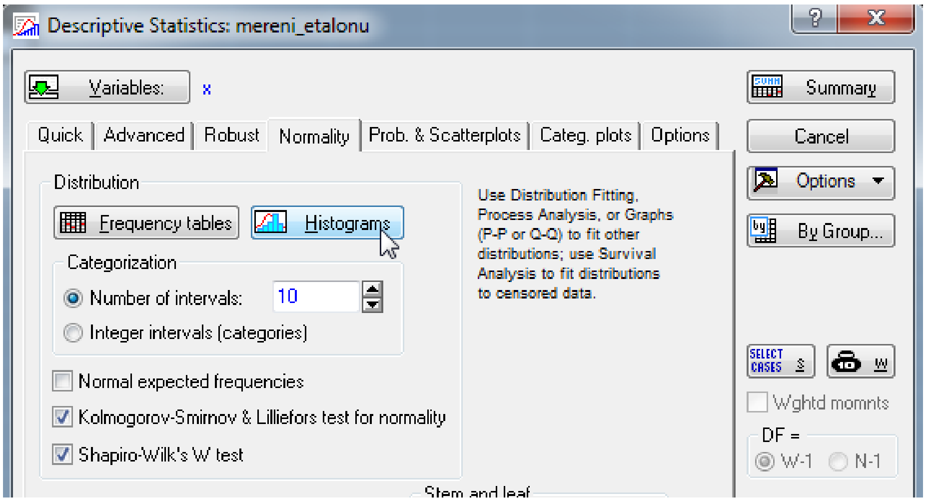


Jako proměnnou dáme proměnnou, kterou chceme zkoumat - *x - OK -> OK*

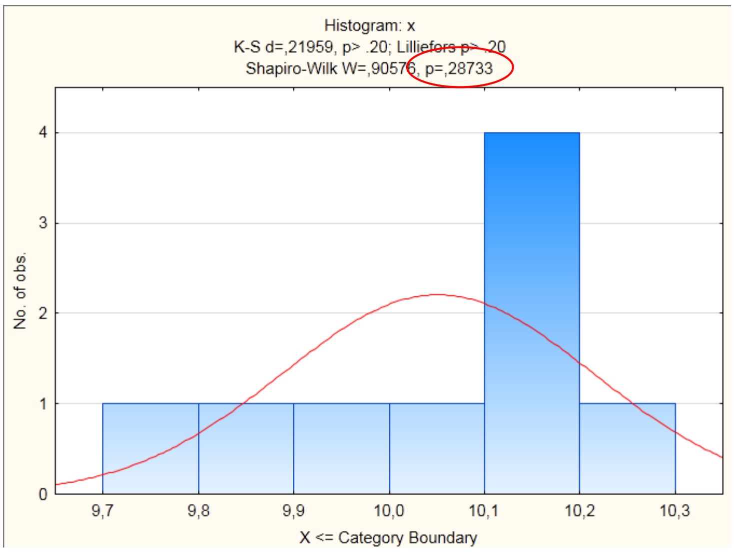
-data celkem kopírují přímku dalo by se považovat za normální



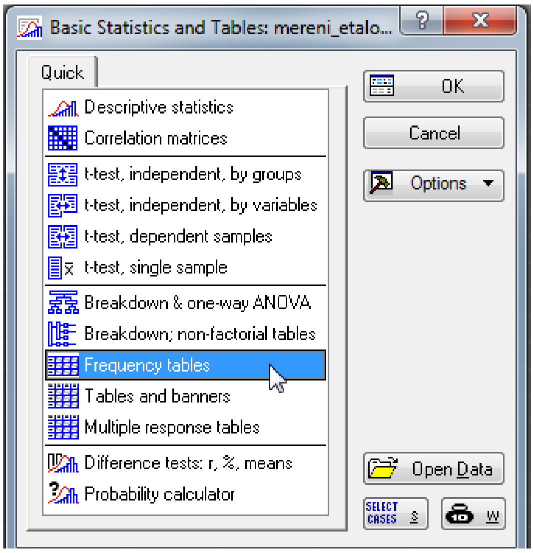
- Histogram + testování (používat Lillieforsův a Shapiro-Wilkův test, u menších vzorků do 30 spíše Shapiro-Wilkův)
  Statistiky -> *Základní statistiky* -> *Popisné statistiky* -> Jako proměnné dát zkoumané proměnné (*x*) -> Záložka *Normalita* -> zaškrtnout Lillieforsův a Shapiro-Wilkův test -> *Histogramy*



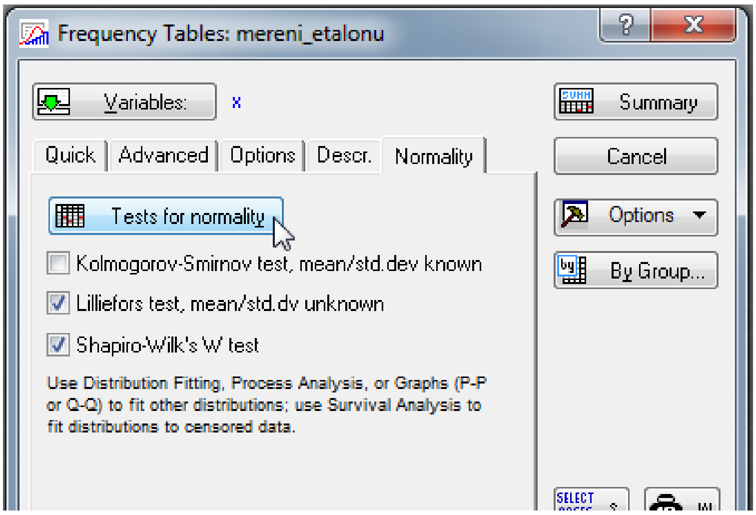
- díky malému počtu pozorování nám toho histogram moc neřekne, ale dle S-W testu (0,287) nezamítáme, že by data pocházela z normálního rozdělení...



- Pouze testové statistiky
  Statistiky ->Základní statistiky -> tabulky četností -> OK



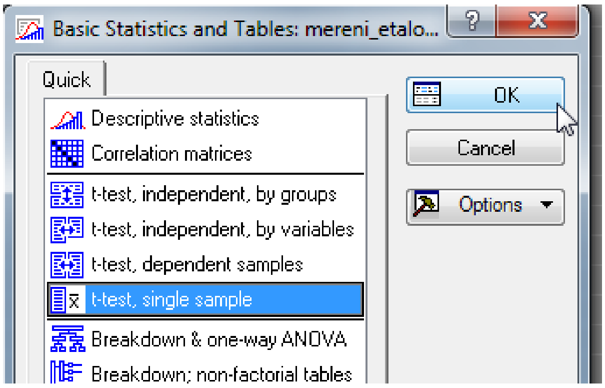
Jako proměnnou dát testované proměnné (x) -> Záložka Normalita -> Zaškrtnout Lillieforse a Shapiro-Wilka -> Testování normality



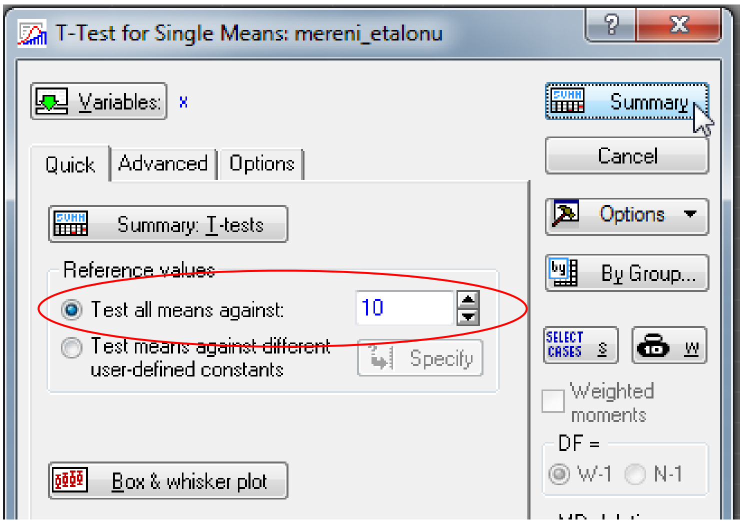
	Tests of Normality (mereni_etalonu)				
Variable	N	max D	Lilliefors p	W	p
x	9	0,219587	p > .20	0,905761	0,287325

<u>Krok B)</u> Výpočet testovací statistiky

Statistiky -> Základní statistiky -> t-test, samost. vzorek -> OK



Proměnná ->naměřenné hodnoty (x) -> referenční hodnota 10 -> Výpočet



Variable	Test of means against reference constant (value) (mereni_etalonu) Mean	Std.Dv.	N	Std.Err.	Reference Constant	t-value	df	p
x	10,05111	0,162669	9	0,054223	10,00000	0,942611	8	0,373470

Dle p-hodnoty 0,373 > 0,05 nezamítáme nulovou hypotézu a můžeme říct, že odchylky měření od očekávané hodnoty byly na 5% hladině významnosti způsobeny jen náhodou.

2) Táborníkům byl zadán úkol, aby odhadli trvání 1 minuty. Testujte na hladině významnosti 0,05 že se jejich odhad neodlišoval od skutečné doby trvání 1 minuty.

Dataset - odhad_minuty.sta

[H0: odhad táborníků se neliší od skutečné doby 1 minuty
$p < 0{,}001$, zamítáme nulovou hypotézu]

3) Při nanášení tenkých kovových vrstev stříbra se vyžaduje, aby tloušťka vrstvy byla 0,020 µm. Zjistěte, zda se statisticky významně odlišují naměřené hodnoty dle spektroskopie od této požadované tloušťky.

Dataset - vrstva_stribra.sta

[H0: naměřené hodnoty se neliší od předpokládané tloušťky vrstvy stříbra
$p = 0{,}026$, zamítáme nulovou hypotézu]

II) Dvouvýběrový nepárový t-test

Máme k dispozici dva výběry a srovnáváme jejich průměry... '

Předpoklady *Normální rozdělení obou výběrů

***Homogenita rozptylu (zda mají oba výběry stejnou sm. odchylku)**

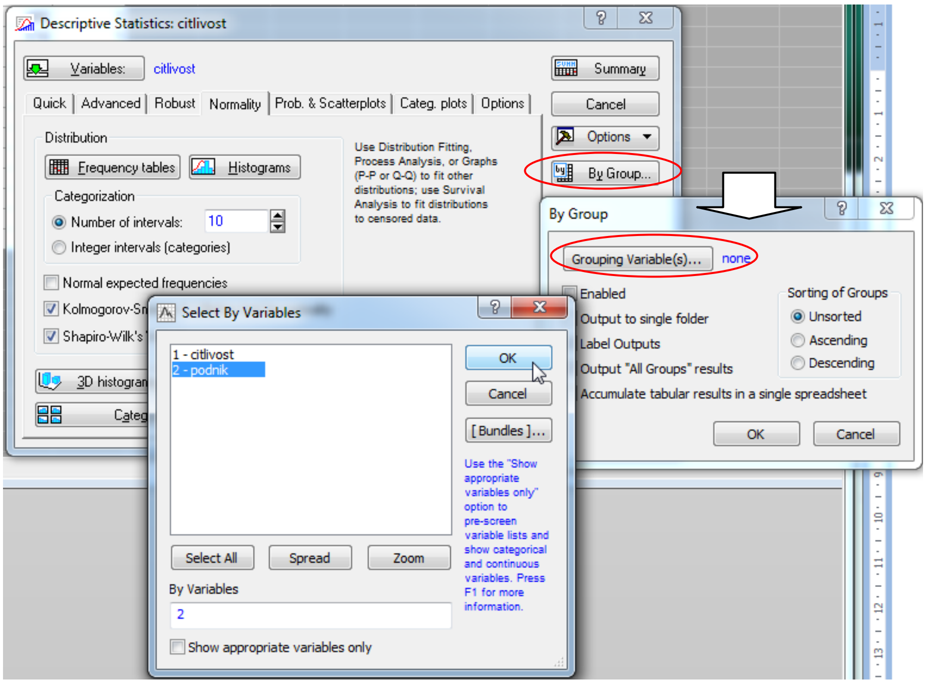
1) Měříme citlivost zařízení ve 2 podnicích. Testujte, zda se citlivost zařízení v jednotlivých podnicích liší. (citlivost.sta**)**

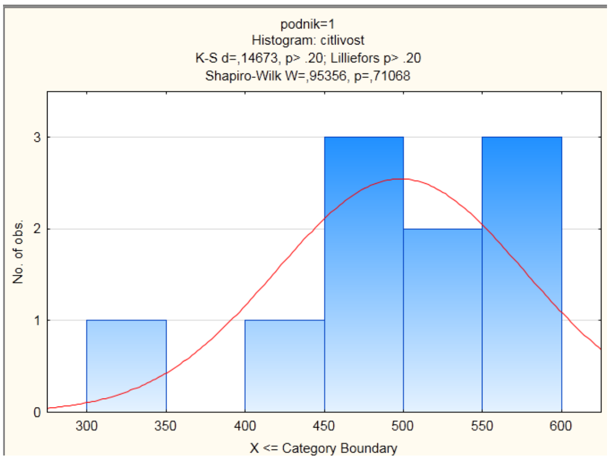
H0: *Není statisticky významný rozdíl mezi citlivostí zařízení v prvním a v druhém podniku. ($\mu_1 = \mu_2$)*

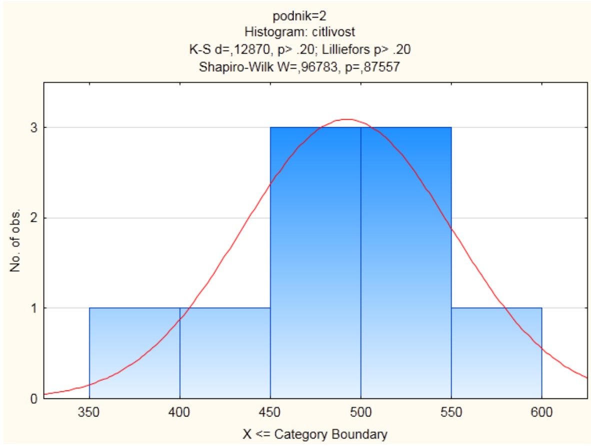
H1: *Citlivost zařízení je statisticky významně odlišná v jednotlivých podnicích($\mu_1 \neq \mu_2$)*

Krok A) Postupujeme stejně jako při jednovýběrovém testu - pomocí testovacích statistik (Liliefors a Shapiro-Wilk) zkoumeme, zda oba výběry pochází z normálního rozdělení - **POKUD NE PAK PŘISTUPUJEME K NEPARAMETRICKÝM TESTŮM**

-využijeme skupinové proměnné (podnik)

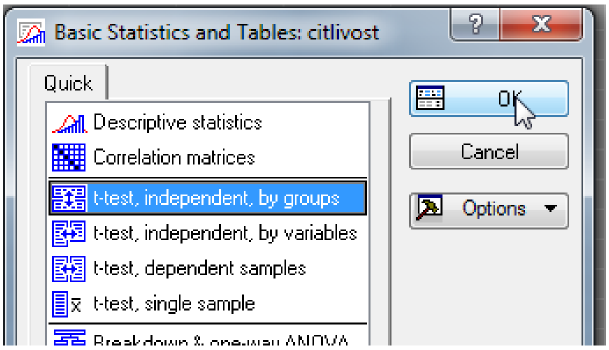




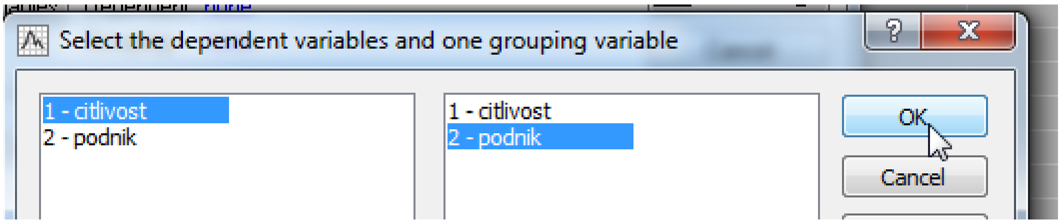


Krok B) Dle p-hodnoty v obou podnicích vidíme, že normalitu dat v jednotlivých výběrech nezamítáme. Nyní můžeme přistoupit k testování.

Statistiky -> Základní statistiky -> t-test, nezávislý, dle skupin -> OK



Jako závisle proměnnou dáme *citlivost* jako skupinovou proměnnou dáme *podnik…*



-> *Výpočet*

T-tests; Grouping: podnik (citlivost)
Group 1: 1
Group 2: 2

Variable	Mean 1	Mean 2	t-value	df	p	Valid N 1	Valid N 2	Std.Dev. 1	Std.Dev. 2	F-ratio Variances	p Variances
citlivost	498,0000	492,2222	0,180933	17	0,858560	10	9	78,28722	58,04692	1,818961	0,411634

Tabulku čteme odzadu!!! - NEJPRVE koukáme na test o shodě rozptylů, pokud je splněn (p>0,05) pak interpretujeme hodnotu testu - p>0,05 - zamítáme nulovou hypotézu o shodě průměrů citlivosti zařízení v jednotlivých podnicích.

POKUD není splněn předpoklad o homogenitě rozptylů, pak přistoupíme k NEPARAMETRICKÉ VARIANTĚ

2) Byly použity dva typy hnojení. Testujte zda výnos pro první typ hnojení se neliší od výnosu při druhém typu hnojení.

Dataset: hnojeni.sta
-**nápověda -** jedná se o stejný typ testu, i když jsou data zadána trochu jinak než tomu bylo v ukázkovém příkladě. Jediné co se však změní je, že místo skupinové proměnné při testu normality a při testování používáme dvě proměnné (tedy skupinovou proměnnou zadávat nemusíme). Další změna je že místo t-test nezávislý dle skupin vybíráme t-test nezávislý dle proměnných.

[H0: Výnos při prvním hnojení je stejný jako výnos při druhém typu hnojení
p =0,269, nezamítáme nulovou hypotézu]

3) Byla naměřena výška studentek z informatiky a studentek z BT-BIO. Porovnejte, zda jsou studentky z obou oborů stejně vysoké.

Dataset - studentky.sta

[H0: naměřené hodnoty se neliší od předpokládané tloušťky vrstvy stříbra
p = 0,088, nezamítáme nulovou hypotézu]

III) Párový t-test

Posuzujeme, zda se významně změnili hodnoty před nějakou událostí a po ní/2 měření stejného parametru jiným způsobem aj., máme vždy dva údaje k jednomu případu (pacientu) – před a po

- Pokud si nejsem jistý, zda se jedná o párovost – vyzkouším korelaci obou parametrů- pokud významná, tak se zřejmě bude jednat o párovou variantu

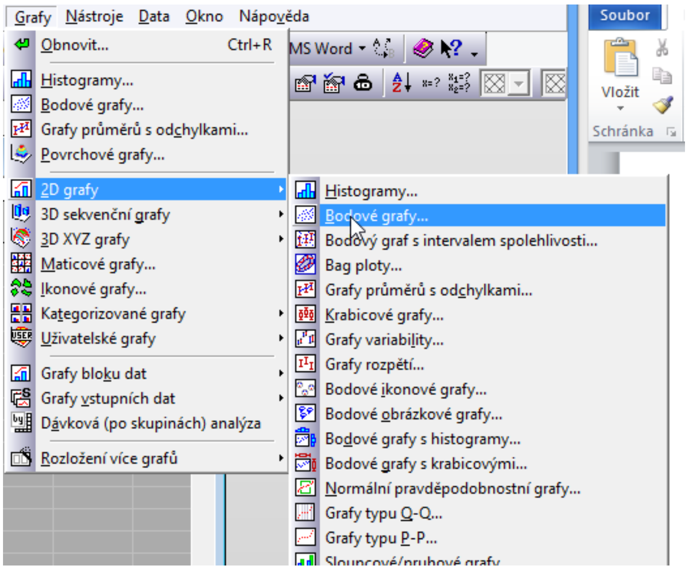
Předpoklady – normální rozdělení rozdílu hodnot před a po!

1) **Máme 2 metody (*A* a *B*) pro odhadnutí nějakého parametru. Testujte na 5% hladině významnosti, že se tyto dvě metody neliší.**

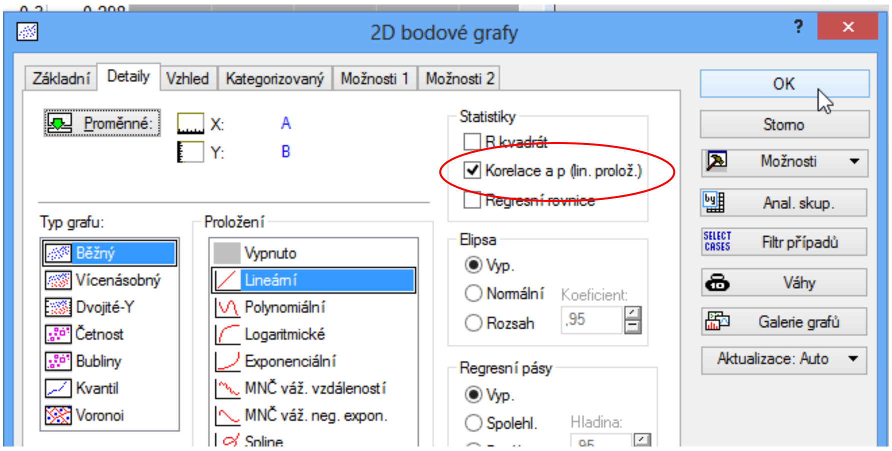
H0: Není rozdíl mezi danými metodami (μ_1- μ_2=0)
H1: Mezi metodami je statisticky významný rozdíl (μ_1- μ_2≠0)

Nepovinný krok) – Zjistím, zda se jedná opravdu o párový test, vypočtu si korelaci těchto dvou měření.

Grafy -> 2D grafy -> bodový graf



Jako proměnné vybereme naše dvě metody *A* a *B* a na záložce *Detaily* zaškrtneme *korelace a p(lin.prolož.) ->OK*

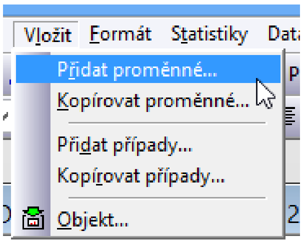


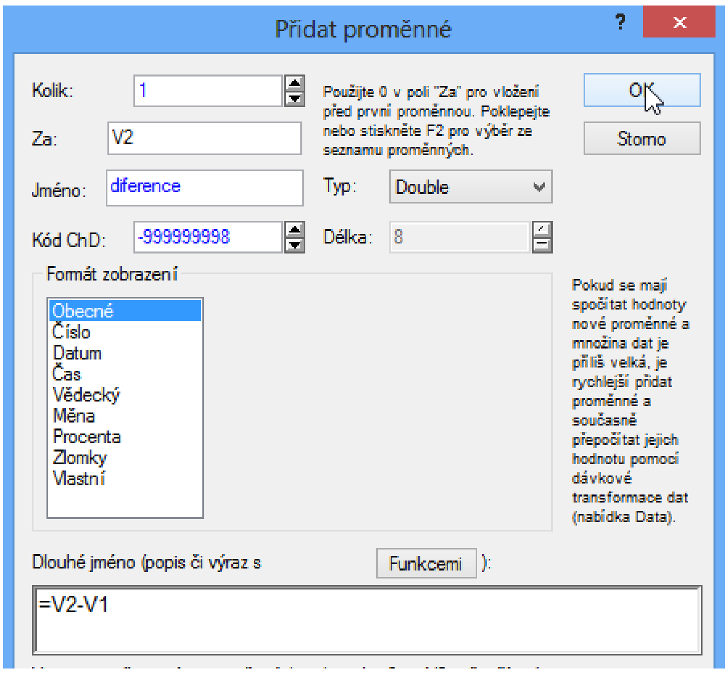
Z grafu a vysoké korelace můžeme usoudit, že se jedná o párový test…

Krok A) Ověříme předpoklady daného testu

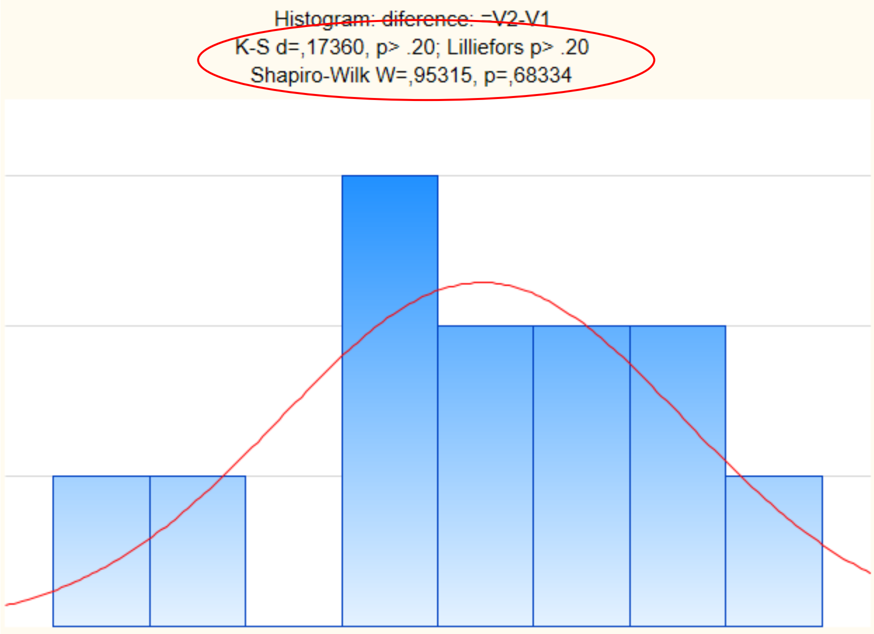
- Vypočteme novou proměnnou, která bude rozdíl mezi oběma metodami *A* a *B…*

Vytvoříme novou proměnnou : Záložka *Vložit -> Přidat proměnné->* Dáme za druhou proměnnou, nazveme *diference a* do *Dlouhého jména* napíšeme vzorec pro výpočet proměnné *=V2-V1 -> OK*



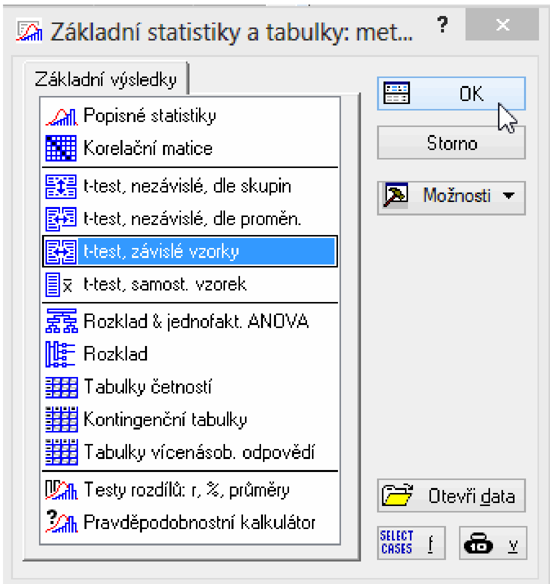


Tuto novou diferenci otestujeme na normalitu (viz výše)…

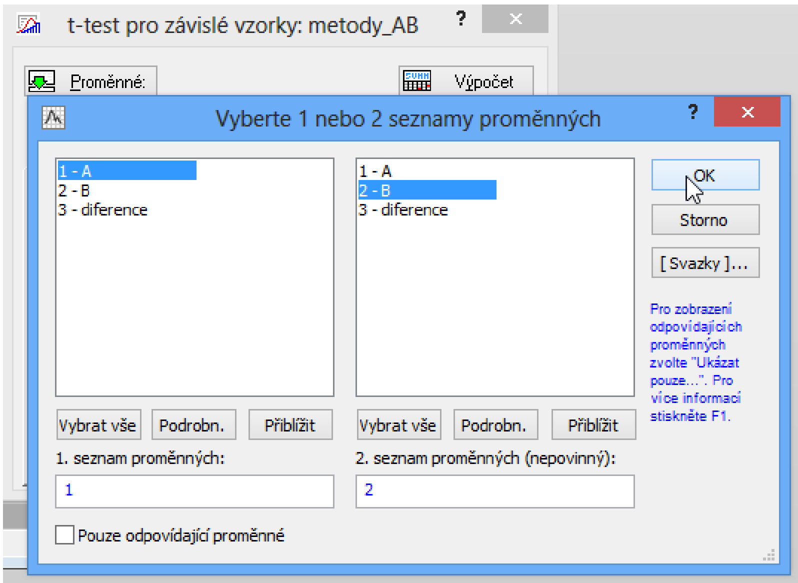


Krok B) Z výsledku je patrné, že normalita diference je splněna (**pokud by nebyla – neparametrický párový Wilcoxonův test**). Nyní můžeme přistoupit k samostatnému testování:

Statistiky -> Základní statistiky -> t-test, závislé vzorky -> OK …



Jako výběr proměnných zvolíme jednotlivé metody -> *OK*-> *Výpočet …*



Z výsledné tabulky vidíme, že nulovou hypotézu zamítáme a že metody *A* a *B* nejsou srovnatelné.

t-test pro závislé vzorky (metody_AB)
Označ. rozdíly jsou významné na hlad. p < ,05000

Proměnná	**Průměr**	Sm.odch.	N	Rozdíl	Sm.odch. rozdílu	t	sv	p	Int. spolehl. -95,000%	Int. spolehl. +95,000%
A	0,305500	0,030658								
B	0,308417	0,030146	12	-0,002917	0,004188	-2,41262	11	0,034454	-0,005577	-0,000256

2) Je testována hloubka dezénu pneumatik před projetím jistého úseku a po něm. Na hladině významnosti 0,05 testujte, že se hloubka dezénu nezměnila.

Dataset: pneumatiky.sta

[H0: Hloubka dezénu zůstala stejná
p =0,341, nezamítáme nulovou hypotézu]

3) je naměřen tlak před podáním léku a po něm. Testujte, zda má daný lék vliv na krevní tlak.

Dataset - tlak.sta

[H0: tlak před podáním léku a po jeho podání je stejný
p = 0,039, zamítáme nulovou hypotézu]