Výpočty hodnot + přidání jako novou proměnnou

Transform → Compute

Ve variable-view zkontrolovat druhy proměnné, případně donastavit LABEL.

Ukázka:

Vypočítejte index maskulinity, který vyjadřuje počet mužů připadající na 100 žen, a přidejte jej do souboru jako novou proměnnou.
Vypočítejte index stáří, který vyjadřuje, kolik je v populaci obyvatel ve věku 65 let a více na 100 dětí ve věku 0-14 let, a přidejte jej do souboru jako novou proměnnou.

K-S test

Analyze → Non-parametric statistics → Sample K-S

Zvolit proměnnou, zaškrtnout rozdělení a potvrdit

Je-li *p-hodnota (Asymp. Sig.) > 0,05*, tak rozdělení je NORMÁLNÍ.

Ukázka:

Zjistěte pomocí Kolmogorov-Smirnovova testu, zda hodnoty indexu maskulinity a indexu stáří mají normální rozdělení, nebo ne.

Oddělení souborů pro výpočty (PAK VRÁTIT ZPĚT)

Data → Split File

Vybrat proměnnou, zvolit Organize output by groups

Ukázka: (nejprve tento krok a pak krok následující)

Vypočítejte aritmetický průměr, směrodatnou odchylku, medián a horní a spodní kvartil pro hodnoty indexu stáří a maskulinity, rozdělené podle okresů (Liberecký a Jablonecký okres zvlášť).

Aritmetický průměr, směr. odchylku, medián, kvantily

Analyze → Descriptives Statistics → Explore

Do Dependent List vložit proměnné, ze kterých počítat – ve Statistics zvolit Descriptives a Percentiles

---

T-TEST jednovýběrový

Analyze → Compare Means → One Sample T-TEST

Známe hodnoty průměru/průměrů a testujeme, zda-li má výběrový soubor stejné parametry jako soubor základní. (Máme-li více proměnných a více průměrů, musíme to pro každý případ udělat zvlášť.)

Vybereme hodnotu proměnné, k níž se vztahuje průměr, který známe. Průměr, který známe zapíšeme jako Test Value a dáme OK.

Je-li *p-hodnota (Asymp. Sig.)* > 0,05 (mezi souborem a známou hodnotou průměru neexistuje statisticky významný rozdíl)

Ukázka:

Otestujte, zda výběrový soubor má stejné parametry jako základní soubor obyvatel ČR podle průměrného věku, když víte, že průměrný věk mužů v ČR je 40,6 let a žen 43,4 let.

T-TEST pro dva nezávislé výběry

Analyze → Compare Means → Independant Samples T-TEST

Zjisťujeme, zda-li jsou v jedné proměnné rozdíly např. ve dvou různých okresech.

Test Variable je název proměnné zastoupené pro dva okresy a Grouping Variable je třídící znak, JE NUTNO DODEFINOVAT případy v Define groups.

Sledujeme ve spodní velké tabulce hodnotu Sig. (2-tailed) → výsledkem jsou dvě hodnoty → výslednou hodnotu volíme dle Leveneho testu, viz obrázek:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Průměrný věk obyvatel k 1. 1. 2017	Equal variances assumed	2,37	,127	2,34	91,00	,022	1,15	,49
	Equal variances not assumed		→ 0,05	2,19	56,56	,032	1,15	,52

(lze Leveneho test dosáhnout přes ANOVu – Homogenity – pro kontrolu…)

Analýza rozptylu ANOVA

Analyze → Compare Means → One Way ANOVA

Pokud nechceme srovnávat pouze 2 nezávislé výběry, ale rovnou více, je výhodné využít analýzy rozptylu (Analysis of variance => ANOVA).

• Před samotnou ANOVA je třeba provést test shody rozptylů (Leveneho test, F-test), protože to je jeden z předpokladů pro ANOVA – výsledek testu by měl být > 0,05

• Pokud Leveneho test vyšel < 0,05, lze ANOVA použít, pokud poměr mezi největší a nejmenší směrodatnou odchylkou je maximálně 2

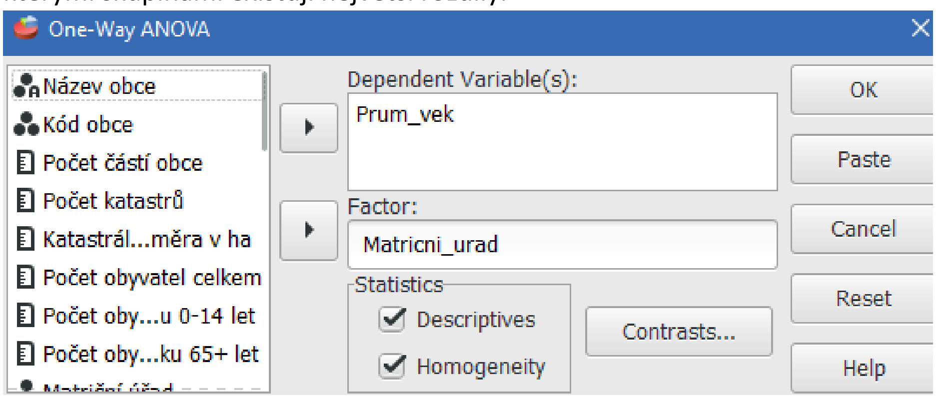
pro ANOVU cheme výsledek < 0,05 – zamítáme H0

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	205,73	23	8,94	2,10	,010
Within Groups	294,17	69	4,26		< 0,05
Total	499,90	92			

k obyvatel k 1. 1. 2017)

Ukázka:

Zjistěte pomocí analýzy rozptylu, zda existují statisticky signifikantní rozdíly v hodnotách proměnné Prum_vek mezi obcemi, které spadají pod jednotlivé matriční úřady. Pomocí Bonferroniho testu zjistěte, mezi kterými skupinami existují největší rozdíly.



→ zvolíme v menu ANOVu dle výše uvedeného, nedáváme ale OK ale PASTE, a přidáme syntax pro BONFERONIHO TEST

```
ONEWAY /VARIABLES= Prum_vek BY Matricni_urad
      /STATISTICS=DESCRIPTIVES HOMOGENEITY
      /POSTHOC BONFERRONI
      .
```

Rozdíl je ukázán v tabulce Mean Difference (I-J).

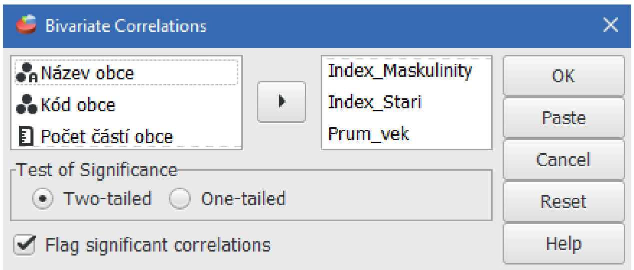
Korelace a regrese

Zadání:

Zjistěte, zda mezi proměnnými index maskulinity, index stáří a Prum_vek existuje statisticky významná korelace. Pro nejtěsnější vztah zjistěte rovnici regresní přímky.

1) Zjístíme, zda-li existuje významná korelace

Analyze → Bivariate correlation



0 – 0,2 = zanedbatelný vztah
0,2 – 0,4 = slabý vztah
0,4 – 0,7 = středně těsný vztah
0,7 – 1 = velmi těsný vztah

Nejtěsnější vztah mezi průměrným věkem obyvatel a indexem stáří. (Sig < 0,05 a koeficient je co největší do 1 – viz tab. výše)

< 0,05

Correlations

		Index_Maskulinity	*Index_Stari*	*Průměrný věk obyvatel k 1. 1. 2017*
Index_Maskulinity	*Pearson Correlation*	1,00	-,16	-,09
	Sig. (2-tailed)		~~,134~~	~~,399~~
	N	93	93	93
Index_Stari	*Pearson Correlation*	-,16	1,00	,91
	Sig. (2-tailed)	~~,134~~		,000
	N	93	93	93
Průměrný věk obyvatel k 1. 1. 2017	*Pearson Correlation*	-,09	,91	1,00
	Sig. (2-tailed)	~~,399~~	,000	
	N	93	93	93

2) Zobrazíme graf: **Graphs → Scatterplot** (vybereme tyto dvě hodnoty) – Y: Index, X: Věk (ř. h. vždy na X – depdendent – protože už je v dat. souboru uvedena)

3) Zjistíme parametry regresní přímky (pak dosadíme a máme rovnici): **Analyze → Regression → Linear** (dependent X - věk, independent Y – vypočtena až námi) + volba Statistics (odškrnout ANOVA)

	Unstandardized
	B
(Constant)	36,52
Index_Stari	,04

y = Bconstant + Bproměnné x → y = CONSTANT + Index_STARI . x