T- test

Souvislost

- 2 x kategorická proměnná -> crosstab (minule)
- Kategorická proměnná x kardinální -> srovnání průměrů (t-test)
- Kardinální x kardinální -> korelace (příště)

T-test - využití

- Jak se liší průměry jedné proměnné, podle kategorií druhé proměnné
  - Pohlaví x příjem
  - Věkové kategorie x levice-pravice
  - Seminární skupiny x hodnocení práce
- Jak se liší hodnoty proměnné mezi experimentálními skupinami

- Základ pro anovu

Princip

- Je rozdíl mezi průměry statisticky významný
- Překrývají se intervaly spolehlivosti
- Menší průměr +2*SE < větší průměr -2*SE
  - 2 pokud chceme spolehlivost 95%, pro vyšší (99%) spolehlivost 3

One-sample mean

- Srovnává hodnotu se známou populační hodnotou nebo kritickou hodnotou
  - Liší se průměrný věk ve vzorku od skutečného průměrného věku populace?
  - Jak moc se liší doba zkažení jogurtu od minimální doby spotřeby?

Příklad

- Liší se ideální věk odchodu do důchodu od skutečného odchodu do důchodu
- SPSS:
- Analyze -> Compare means -> One sample t-test
- Test variable: retired pernamently, ideal age
- Test value: 65

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Retire permanently, ideal age. SPLIT BALLOT	1931	60,42	3,725	,085

One-Sample Test

	Test Value = 65					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Retire permanently, ideal age. SPLIT BALLOT	-53,998	1930	,000	-4,577	-4,74	-4,41

- Průměr ideálního věku do důchodu: 60,42
- Rozdíl oproti skutečnému věku (65): 4,5
- Je rozdíl signifikantní?
- Je rozdíl věcně významný?

Ukázat jak dostat tabulky z SPSS do excelu!

T-Test pro nezávislé výběry

- Nezávislé – srovnání 2 skupin
  - Rozdíl příjmu mezi muži a ženami
  - Rozdíl mezi průměrem pro muže a pro ženy

  - Pro srovnání více skupin potřeba udělat výběry z dat
  - Nebo obejít pomocí funkce explore

Příklad

- Liší se příjem mezi muži a ženami?
- SPSS:
- Analyze -> Compare means -> Independent samples t-test
- Test variable: your usual [weekly, monthly, annualy] net [….
- Grouping variable: Zena
- Define groups: 0 1

Group Statistics

	zena	N	Mean	Std. Deviation	Std. Error Mean
Your usual [weekly/monthly/annual] net [pay/pensions/social benefits]	,00	779	22383,93	30184,366	1081,467
	1,00	1001	18995,37	23824,091	753,007

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Your usual [weekly/monthly/annual] net [pay/pensions/social benefits]	Equal variances assumed	,704	,401	2,647	1778	,008	3388,565	1280,133	877,841	5899,289
	Equal variances not assumed			2,571	1450,065	,010	3388,565	1317,798	803,571	5973,559

- Průměr pro muže: cca 22 000 kč
- Průměr pro ženy: cca 19 000 kč
- Rozdíl mezi průměry: 3388 kč
- Je rozdíl signifikantní?
- Je rozdíl věcně významný?

Jak obejít množství skupin

- Analysis -> explore
- Dependent list: your usual [weekly, monthly, annualy] net [….
- Factor list: vzdelani_4kat
- Display statistics

Your usual [weekly/monthly/annual] net [pay/pensions/social benefits]	ZŠ	Mean		13328,09	716,978
		95% Confidence Interval for Mean	Lower Bound	11903,69	
			Upper Bound	14752,49	
		5% Trimmed Mean		13120,95	
		Median		12000,00	
		Variance		46779165,41	
		Std. Deviation		6839,530	
		Minimum		0	
		Maximum		39000	
		Range		39000	
		Interquartile Range		9000	
		Skewness		,554	,253
		Kurtosis		2,213	,500
	SŠbezM	Mean		19664,48	1176,553
		95% Confidence Interval for Mean	Lower Bound	17354,08	
			Upper Bound	21974,89	
		5% Trimmed Mean		16693,06	
		Median		16500,00	
		Variance		880400416,7	
		Std. Deviation		29671,542	
		Minimum		0	
		Maximum		400000	
		Range		400000	
		Interquartile Range		8000	
		Skewness		9,780	,097
		Kurtosis		105,594	,194
	SŠsM	Mean		21339,38	1098,521
		95% Confidence Interval for Mean	Lower Bound	19182,81	
			Upper Bound	23495,96	
		5% Trimmed Mean		18477,41	
		Median		20000,00	

ZŠ	13328.09
SŠbezM	19664.48
SŠsM	21339.38
VŠ	22272.8

		průměr
ZŠ	Spodní hranice	11903.69
	Horní hranice	14752.49
SŠbezM	Spodní hranice	17354.08
	Horní hranice	21974.89
SŠsM	Spodní hranice	19182.81
	Horní hranice	23495.96
VŠ	Spodní hranice	20920.55
	Horní hranice	23625.06

Párový t-test

- srovnání 2 měření pro stejné případy
- Srovnání různých měření v jednom čase
  - Kdo má vyšší důvěru – prezident nebo premiér?
- Srovnání stejného měření v různém čase
  - Snižuje požití paralenu horečku
  - Měření před požitím – průměr
  - Měření po požití – průměr
  - Rozdíl průměrů před a po

Příklad

- Jak se liší čas odstěhování od rodičů od zisku ekonomické soběstačnosti?
- Analyze -> Compare means -> Paired samples t-test
- Variable 1: year when starts first job
- Variable 2: year when first left parents

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Year first started in paid employment or apprenticeship	1986,02	1737	16,340	,392
	Year first left parents for living separately for 2 months or more	1988,02	1737	16,110	,387

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Year first started in paid employment or apprenticeship - Year first left parents for living separately for 2 months or more	-1,995	5,471	,131	-2,252	-1,737	-15,196	1736	,000

- Průměr první práce: 1986
- Průměr odstěhování: 1988
- Je rozdíl statisticky významný
- Je rozdíl věcně významný

Určení síly

- Cohenovo D
- Pearsonovo R

- Čím vyšší, tím silnější souvislost
- Pearsonovo R je kvůli způsobu výpočtu menší než cohenovo D

- Omezená využitelnost

Neparametrické testy

- T-test by měl být správně použit na normálně rozdělená data
  - U velkých vzorků (nad 500 případů) je to jedno

- Mann-Whitney test
  - Rozložení může být libovolné
  - Nepočítá s hodnotami proměnné ale s pořadím případů
  - ANALYZE -> Nonparametri tests -> Legacy dialog

- Když si nejste jistí, který použít, použijte oba