

# Srovnání průměrů

POLb1139 Statistické myšlení

# Souvislost

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

# T-test - využití

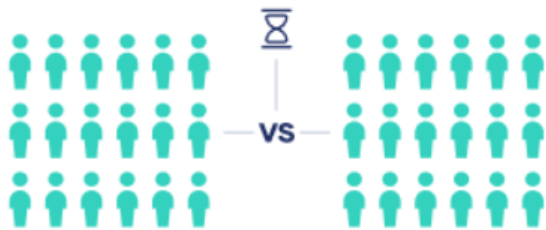
- Jak se liší průměry jedné proměnné, podle kategorií druhé proměnné
  - Pohlaví x příjem
  - Věkové kategorie x levice-pravice
  - Seminární skupiny x hodnocení práce
- Experimenty (2 skupiny)
- Základ pro analýzu ANOVA (3+ skupiny)

# 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

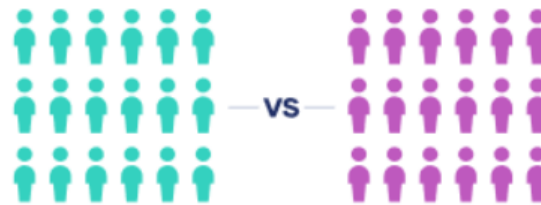
# Druhy T-testů

## Paired-samples t test



Investigate whether there's a difference within a group between two points in time (within-subjects).

## Independent-samples t test



Investigate whether there's a difference between two groups (between-subjects).

## One-sample t test



Investigate whether there's a difference between a group and a standard value or whether a subgroup belongs to a population.

(<https://www.scribbr.com/statistics/t-test/>)

# One-sample T-test

- T-test pro jediný výběr
- 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

- Ideální věk odchodu do důchodu vs. skutečný odchod do důchodu
  - Důchodový věk je podle právních předpisů 65 let
  - Myslí si lidé, že je to ideální věk pro odchod do důchodu?
  - Porovnááme průměr ideálního důchodového věku (co si lidé myslí) se skutečným důchodovým věkem (65)
- SPSS:
  - Analyze -> Compare means -> One sample t-test
  - Test variable: retired permanently, ideal age [iagrtr]
  - 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

	t	df	Significance		Mean Difference	95% Confidence Interval of the Difference	
			One-Sided p	Two-Sided p		Lower	Upper
Retire permanently, ideal age. SPLIT BALLOT	-53,998	1930	<,001	<,001	-4,577	-4,74	-4,41

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



# Independent samples T-test

- T-Test pro nezávislé výběry
- Nezávislé – srovnání 2 skupin
  - Rozdíl průměrného příjmu mezi muži a ženami
  - Rozdíl mezi průměrným podílem hlasů Demokratů a Republikánů

# Příklad

- Liší se příjem mezi muži a ženami?
- SPSS:
  - Analyze -> Compare means -> Independent samples t-test
  - Test variable: your usual net income [*netinum*]
  - Grouping variable: Gender
  - Define groups: 0 1

### Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Your usual net income	Man	779	22383,93	30184,366	1081,467
	Woman	1001	18995,37	23824,091	753,007

### Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
Your usual net income	Equal variances assumed	,704	,401	2,647	1778	,004	,008	3388,565	1280,133	877,841	5899,289
	Equal variances not assumed			2,571	1450,065	,005	,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 statisticky signifikantní?
- Je rozdíl věcně významný?

# Homogenita rozptylu

- Levenův test
- Testuje nulovou hypotézu, že rozptyly v různých skupinách jsou stejné
  - Pokud test vyjde jako statisticky **signifikantní**, je předpoklad homogenity rozptylů narušený [Equal variances not assumed]
  - Pokud test vyjde jako statisticky **nesignifikantní**, je předpoklad homogenity rozptylů splněn [Equal variances assumed]
- Při velkém počtu hodnot mohou i malé odlišnosti mezi rozptyly vést k signifikantním výstupům

# Levenův test

- SPSS
  - Analyze → Descriptive Statistics → Explore
  - Příslušné proměnné vložit do „Dependent list“ a „Factor list“
  - V „Plots“ si zvolit jednu z možností v „Spread vs Level with Levene Test“ (untransformed)

### Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Your usual net income	Man	779	22383,93	30184,366	1081,467
	Woman	1001	18995,37	23824,091	753,007

### Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
Your usual net income	Equal variances assumed	,704	,401	2,647	1778	,004	,008	3388,565	1280,133	877,841	5899,289
	Equal variances not assumed			2,571	1450,065	,005	,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 statisticky signifikantní?
- Je rozdíl věcně významný?

# Paired samples T-test

- 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?
- SPSS:
  - Analyze -> Compare means -> Paired samples t-test
  - Variable 1: Year first started in paid employment [pdempyr]
  - Variable 2: Year first left parents [lvpntyr]



### Paired Samples Statistics

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

### Paired Samples Test

		Paired Differences					t	df	Significance	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				One-Sided p	Two-Sided p
					Lower	Upper				
Pair 1	Year first left parents for living separately for 2 months or more - Year first started in paid employment or apprenticeship	1,995	5,471	,131	1,737	2,252	15,196	1736	<,001	<,001

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

# Určení síly

- Cohenovo D
- Čím vyšší, tím silnější souvislost
  - 0,2 = malý efekt
  - 0,5 = střední efekt
  - 0,8 = velký efekt
- Pouze orientační hodnoty
- 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 -> Nonparametric tests -> Legacy dialogs
- Když si nejste jistí, který použít, použijte oba