

# Metodologický seminář – II. T-test pro pokročilé v JASP

Pavλίna Janošová

[janosova@htf.cuni.cz](mailto:janosova@htf.cuni.cz)

Konzultace **po dohodě**:

v MS Teams



## Pojmosloví - míry středních hodnot

- **Průměr** (symbol „ $x$ “ nebo „ $m$ “ - v angličtině „**mean**“). Ne vždy se však hodí pro porovnání – pokud se nejedná o tzv. normální rozložení určité vlastnosti, není vypovídající – **viz křivka platů v ČR**. Pak uvádíme medián, případně modus:
- **Medián** (**Median**, **Mdn**) – prostřední hodnota všech respondentů.

### **Příklad:**

1; 5; 5; **7**; 12; 14; 20      --- Mdn = **7**

1; 5; 5; **7; 12**; 14; 20; 22      --- Mdn = průměr dvou prostředních hodnot  
(7 a 12), tj. **9,5** (průměr je něco přes **11**)

- **Modus** (**Mode**) – nejčastěji se vyskytující hodnota v celém souboru respondentů. V obou skupinách bude modus = **5**

## POPIS PROMĚNNÝCH

míry střední hodnoty, grafy, normalita rozložení hodnot

V JASP Descriptives => plots

**U spojitých/kardinálních proměnných:** průměry a směrodatné odchylky i různé varianty grafů.

**U pořadových/ordinálních proměnných:** medián i modus, případně i percentily, koláčový graf ([Pie chart](#)), nebo histogram ([Distribution plots](#)).

**U nominálních/kategoriálních proměnných:** četnost kategorií v procentech, koláčový graf.

- Podrobněji viz **video P. Soukup: Seminář č. 2 pro JASP**

<https://www.youtube.com/watch?v=XhSPTQHdCLQ>

DESCRIPTIVES: **5.-7. minuta**, dále **33.-44. minuta**

**= 20 minut i s technickou rezervou.**

## t-test pro dva nezávislé výběry v JASP – podmínky pro použití testu

Pro použití **t-testu** je zapotřebí splnit určité podmínky:

- **Normální rozložení hodnot u kardinální/spojité závisle proměnné** (nad 100 respondentů však začíná platit „centrální limitní věta“).
- **Podobný rozptyl u této proměnné v obou porovnávaných podsouborech.**

Pokud nejsou splněny, nemůžeme použít parametrický t-test.

- Není-li rozložení hodnot normální, použijeme místo t-testu **Mannův Whitneyův U-test** (viz nabídka pod „[Independent samples t-test](#)“).
- Nejsou-li v obou podsouborech podobné rozptyly, použijeme místo t-testu **Welchův test** (též v nabídce pod „[Independent samples t-test](#)“).

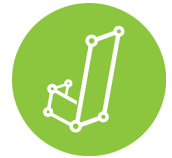
## t-test pro dva nezávislé výběry podmínky pro použití testu



- (1) Kontrola normálního rozložení proměnné** („okometricky“) - V JASP nahoře na liště: „**Descriptives**“:
- Do pravého horního okénka („**variables**“) přesuneme kontrolovanou proměnnou (získáme její průměr a směrodatnou odchylku)
  - Níže nalevo v nabídce otevřeme „**plots**“ a pod tím napravo označíme „**Distribution plots**“. Odpovědi by měly rozhazovat do většiny variant, neměly by být jen v jedné části grafu.
  - (jinou možností je výpočet z šikmosti (**skewness**) a špičatosti (**kurtosis**) nebo **Shapiro-Wilkův test**, viz papírové podklady).

Zpravidla však pracujeme s položkami z celého dotazníku, které sčítáme (spíš zprůměrujeme) jako celek a výsledný graf se pak podobá Gaussově křivce mnohem víc.

## t-test pro dva nezávislé výběry podmínky pro použití testu a výpočet t-testu



- (2) **Kontrolu podobnosti rozptylů v obou podsouborech** provádíme v JASP už rovnou při výpočtu t-testu. Zvolíme nahoře „t-tests“ => „Independent samples t-test“:
- Do pravého horního okénka („Variables“) přesuneme **závislou proměnnou**.
  - Do okénka pod tím („Grouping Variable“) přesuneme proměnnou, která definuje porovnávané podsoubory (při zkoumání rozdílů mezi muži a ženami to bude „**pohlaví**“).
  - Níže vpravo zaškrtneme „Equality of variances“ – ***pokud vyjde Leveneho F nevýznamné („p“ je vyšší než 0,05), můžeme použít výpočet t-testu v horní tabulce.*** Pokud by byl výsledek statisticky významný, museli bychom místo něj použít neparametrický „**Welchův test**“ nalevo v nabídce. Výpočet se objeví v tabulce s t-testem ve druhém řádku. U Welchova testu a Mann Whitneyeho testu (který tu je taky) se místo hodnoty „t“ do diplomky uvádí jejich hodnoty.
  - PRO INTERPRETACI VÝSLEDKŮ: V nabídce nalevo „Descriptives“ získáme popisná data pro oba z porovnávaných podsouborů (průměry/**means** a směrodatné odchylky/**SD**).

## t-test pro dva nezávislé výběry podmínky pro použití testu a výpočet t-testu

**Pokud nevíme**, jestli použít parametrický t-test či neparametrický Mann Whitneyův U-test, **provedeme oba**.

- Když vyjde významnost jen v t-testu, ale nevyjde v Mann Whitneyově testu, pak se řídíme tímto neparametrickým testem (a přijímáme závěr, že se oba podsoubory statisticky významně neliší).
- Jsou-li výsledky shodně významné (či nevýznamné) v obou testech, lze se řídit kterýmkoli z testů.

Independent Samples T-Test

	Test	Statistic	df	p
3_Humor	Student	1.785	115.000	0.077
	Welch	1.819	99.150	0.072
	Mann-Whitney	1975.500		0.086

## Párový t-test v JASP

Používá se při **porovnávání jedné skupiny respondentů ve stejném dotazníku naměřeném ve dvou časových obdobích:**

- např. *výsledky ve stejném znalostním testu u školní třídy před výukovým experimentem a po něm,*
- nebo *klima třídy (či postoje k šikaně) před stmelovacím kurzem a po něm,*
- nebo *dotazník životní spokojenosti při příchodu do domova pro seniory a po nějaké době adaptace* apod.
- Měří se takto také třeba *účinky terapie* před jejím zahájením a po nějaké době trvání apod.

V tomto případě **do Excelu** přepisujeme **číselné údaje od daného respondenta z obou dotazníků/testů na jeden řádek** – Excelový list s daty pak bude mít dvojnásobný počet sloupců a bude širší.



# Výpočet párového t-testu v JASP



Zvolíme na horní liště „t-tests“ => „Paired Samples t-test“:

- Do pravého horního okénka („Variables“) přesuneme kontrolovanou proměnnou v prvním čase měření i ve druhém čase měření.
- V nabídce nalevo využijeme možnost „Descriptives“ (v nabídce vlevo uprostřed), díky níž získáme popisná data pro oba z porovnávaných podsouborů (průměry/**means** a směrodatné odchylky/**SD**)

Nevýznamná hodnota „p“ (více než 0,05) znamená, že se hodnoty neliší a že v čase nedošlo k posunu ve sledované proměnné. Je-li podle „p“ rozdíl mezi prvním a druhým měřením významný, znamená to, že se sledovaná proměnná změnila. Získáním hodnot průměrů v možnosti „Descriptives“ zjistíme, zda sledovaná proměnná narostla, nebo se zmenšila.

# ANOVA

(analýza rozptylu)  
a WELCHŮV či BROWN-FORSYTŮV TEST



**Porovnááme-li více než dva podsoubory**, nemůžeme provést jednotlivé t-testy. Riskovali bychom, že neoprávněně přijmeme alternativní hypotézu.

Zde také máme **závisle proměnnou**: kardinální/spojitou, nebo ordinální/pořadovou (kolonka „**dependent variable**“), dále proměnnou třídící, resp. **nezávisle proměnnou**: buď nominální/kategoriální, nebo ordinální/pořadovou, složenou z několika málo (více než dvou) kategorií/podskupin (kolonka „**fixed factors**“).

**(PODMÍNKA 1)**: Pokud by **závisle proměnná byla příliš vychýlená od normálního rozdělení**, volíme výpočet **Kruskall-Walis**, který najdeme v ANOVA pod nabídkou „**Nonparametrics**“.

**(PODMÍNKA 2)**: Kontrolujeme **shodu rozptylů ve všech podskupinách shodné** - zda můžeme použít JASPEM vypočítaný parametrický Fischerův výpočet (F):

- pod „**Asumption checks**“ volíme „**Homogeneity test**“. Je-li výsledek Leveneho testu významný, rozptyly se liší. Místo parametrického Fischerova testu bychom pak museli použít buď robustnější **Brown-Forsythe test**, nebo **Welchův test**.

# ANOVA

## (analýza rozptylu) - INTERPRETACE VÝSLEDKŮ:



Abychom mohli výsledky INTERPRETOVAT, *musíme si říct o popisnou statistiku* („[Descriptive statistics](#)“) a podívat se na průměry.

***Které z podskupin jsou vzájemně statisticky významně odlišné?***

-Zvolíme nabídku „[post hoc tests](#)“. Do prázdného okénka přeneseme zkoumanou proměnnou. Při shodě rozptylů používáme již nastavený „[test](#)“. Máme-li neshodné rozptyly, používáme test „[Games-Howell](#)“.

... **ANOVA umožňuje do výpočtu zahrnout i více než jednu nezávisle proměnnou.** *Mohli bychom např. zkoumat, zda „spravedlivost učitele“ vyžadují ve stejné míře žáci různých tříd a zároveň i všechny dívky a chlapci.*

-Do kolonky „[Fixed factors](#)“ bychom zadali obě dvě proměnné (pohlaví i příslušnost ke školní třídě). Ve výsledcích dostaneme výpočty pro obě proměnné i pro jejich **interakci** (z ní se můžeme dozvědět za určitých okolností např. to, že spravedlivost učitele je extrémně důležitá pro chlapce z některých konkrétních tříd).

-Tento výpočet interakce (jde o proměnnou znázorněnou jako: „**pohlaví \* třída**“) můžeme zrušit v nabídce „[Model](#)“ tím, že ji zde odstraníme z pravého okénka.

# ANCOVA

(analýza rozptylu s kovariátem):



Jde o analýzu variance/rozptylu (ANOVA), kdy kombinujeme nominální/kategoriální nezávisle proměnné s nezávisle proměnnou spojitou/kardinální (té se říká „kovariát“). ANCOVA je tedy kombinací ANOVY s lineární regresí.

-mezi nezávisle proměnné se zařazuje proměnná nominální/kategoriální (kolonka „**fixed factors**“), ale i proměnná spojitá/kardinální (kolonka „**covariates**“). Výsledek pak zachycuje, jakou hrají tyto nezávisle proměnné roli u proměnné závislé, když je sledujeme obě najednou. Je jasné, že zahrnování více proměnných do výpočtu přináší přesnější výsledky, protože lépe odpovídají naší realitě. V běžném životě na nás také působí mnohem víc vlivů než jen jeden.

-ANCOVA umožňuje i výpočet s více nezávisle proměnnými, nominálními, i spojitými.

ANCOVA - Humor

Cases	Sum of Squares	df	Mean Square	F	p
pohlaví	0.003	1	0.003	0.003	0.955
Sympatie	19.893	1	19.893	22.685	< .001
Residuals	99.967	114	0.877		



## KORELACE

(souvislosti mezi spojitými nebo pořadovými proměnnými)

Korelacemi zjišťujeme, zda mezi dvěma (a více) proměnnými existuje významný lineární vztah. Tento vztah se může podobat přímé úměrnosti (a pak očekáváme pozitivní korelační koeficient), nebo nepřímé úměrnosti (pak by šlo o zápornou hodnotu).

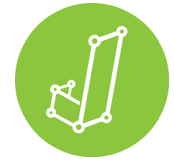
Korelační koeficient má hodnotu v rozmezí -1 až +1. Čím vyšší absolutní hodnotu má směrem k jedničce, tím je vztah mezi proměnnými těsnější, silnější. Čím víc se blíží nule, tím méně spolu obě proměnné souvisí.

Klikneme na: „Regression“ => „Correlation“. Do horního okénka přesuneme obě proměnné, jejichž vzájemný vztah chceme zkoumat. Získáváme **výpočet parametrického Pearsonova korelačního koeficientu** ( $r$ ) a pod ním hodnotu významnosti  $p$ .

V případě pořadových/ordinálních proměnných a také u kardinálních/spojitých proměnných s výrazně nesymetrickým (nenormálním) rozložením hodnot volíme **výpočty pořadových koeficientů**: **Spearmanův** ( $\rho$ ), nebo **Kendallův** ( $\tau$ ). Za těmito účely i zde testujeme normalitu rozložení těchto dat (pomocí „Assumption checks“ => „Shapiro“ pro dvě proměnné – „Pairwise“).

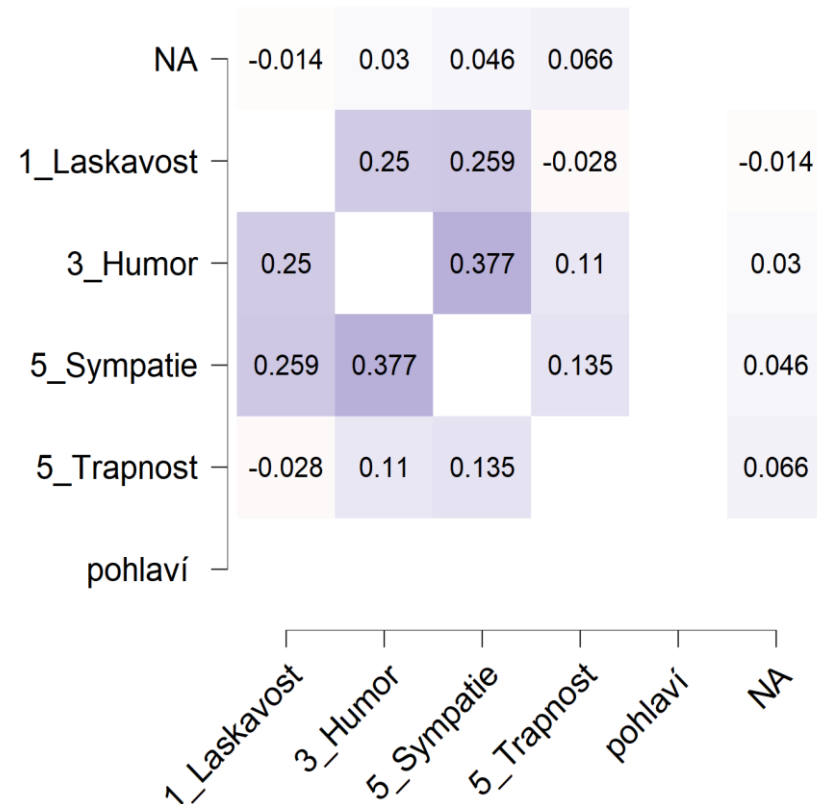
# KORELACE

## - grafické znázornění výsledků



Korelační vztahy mezi spojitými/kardinálními proměnnými lze rozkreslit v grafu: „Plots“ => „Scatter plot“. Uvidíme jednotlivé „respondenty“ v bodech, kde se v prostoru nacházejí na osách **x** a **y** (=nezávisle proměnná a závisle proměnná).

Pokud korelujeme víc proměnných, můžeme si říct o schéma „Heatmap“, které ukáže Pearsonovy hodnoty mezi každými z nich. Podle intenzity barvy daného políčka znázorní i těsnost korelačního vztahu mezi jednotlivými dvojicemi proměnných, přičemž modrá značí kladný korelační vztah a červená vztah záporný.





## VIDEO

(P. Soukup: Seminář 3 - JASP )

**Podíváme se na VIDEO P. Soukupa:**

<https://www.youtube.com/watch?v=RGHJxiUpPBI>

-Úsek od začátku do 17.30 min.

-Úsek od 25.30 min. – do konce.

-Celkem cca 40 minut.

**-Za 40-45 minut se sejdeme v MS Teams**

# Děkuji Vám za pozornost

- přejdeme k Vaším dotazům, poznámkám a námětům: