

# ANOVA a její neparametrická alternativa

# ANOVA

- Testuje rozdíl v průměrech pro teoreticky neomezený počet skupin, tzn. zvládne více než dvě skupiny
- Příklad: Existuje rozdíl v průměrném příjmu mezi lidmi s rozdílnou úrovní vzdělání?
- V podstatě lepší alternativa t testu
- Možné chápat jako typ lineární regrese (později)
- Základní předpoklady:
  - Nezávislost pozorování
  - Kardinální data (v praxi i ordinální)
  - Normalita rozdělení
  - Stejnost rozptylů

# ANOVA

- Založená na porovnávání meziskupinového a vnitroskupinového rozptylu
- $H_0$  = Rozdíly v průměrech všech skupin jsou nulové

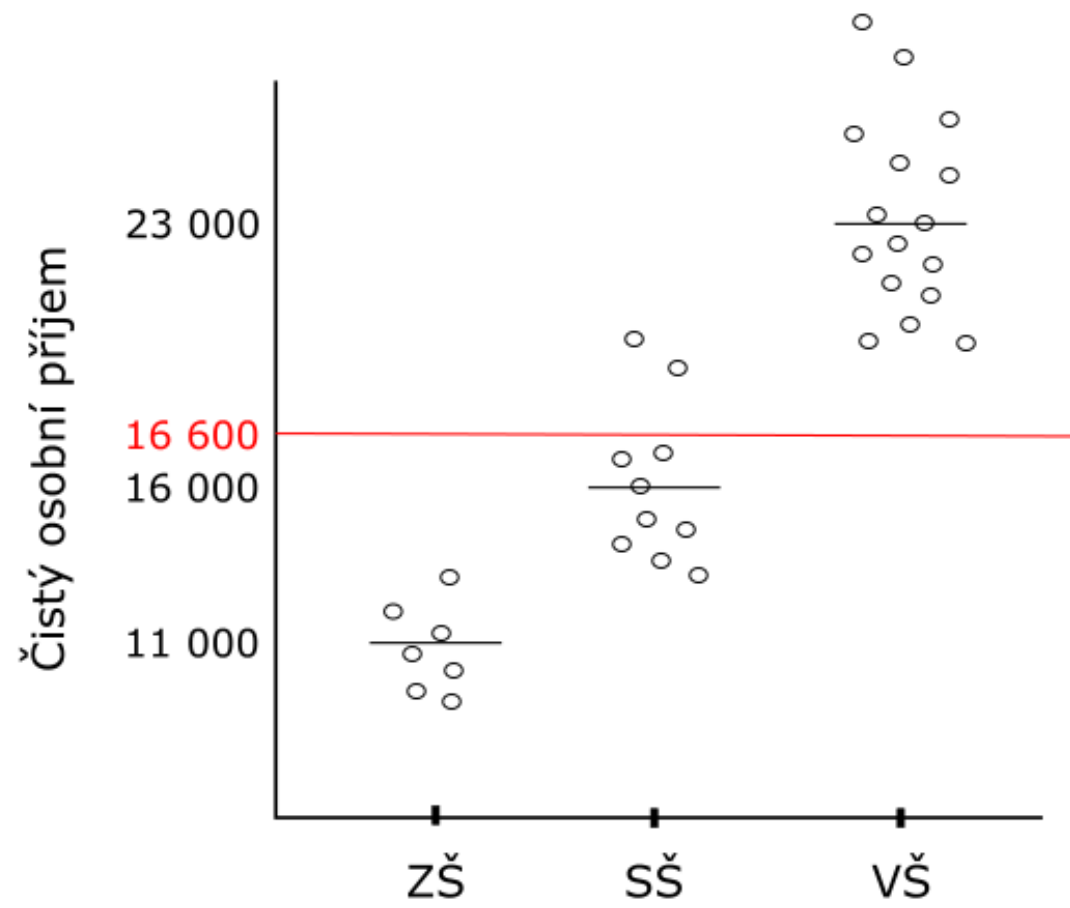
$$F = \frac{\text{Rozptyl mezi skupinami}}{\text{Rozptyl uvnitř skupin}} = \frac{\frac{SS_{\text{mezi skupinami}}}{DF_{\text{meziskupinami}}}}{\frac{SS_{\text{uvnitř skupin}}}{DF_{\text{uvnitř skupin}}}}$$

SS = suma čtverců  
DF = stupně volnosti

Pro připomenutí:

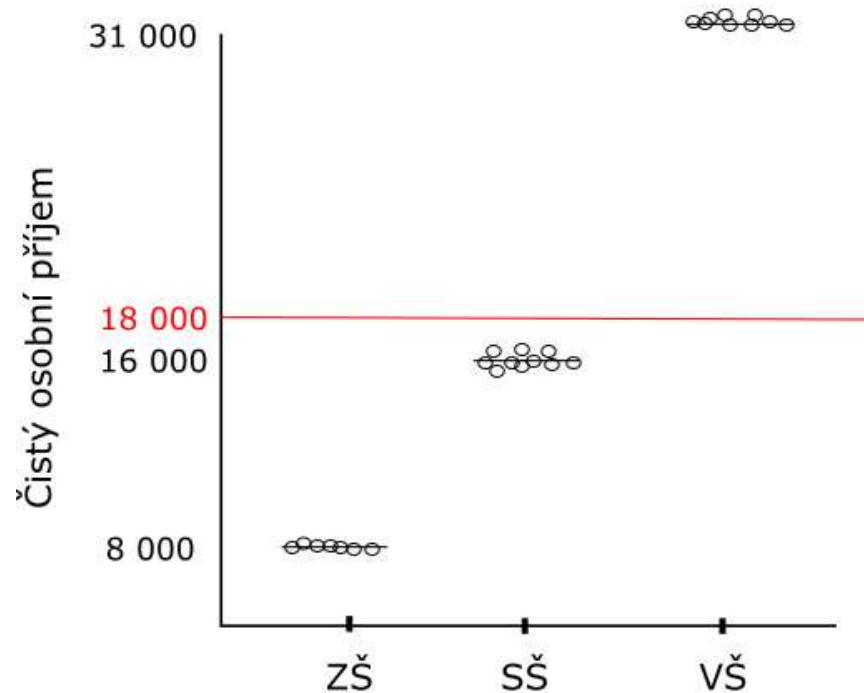
$$\text{Rozptyl} = \frac{SS}{DF} = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

# ANOVA

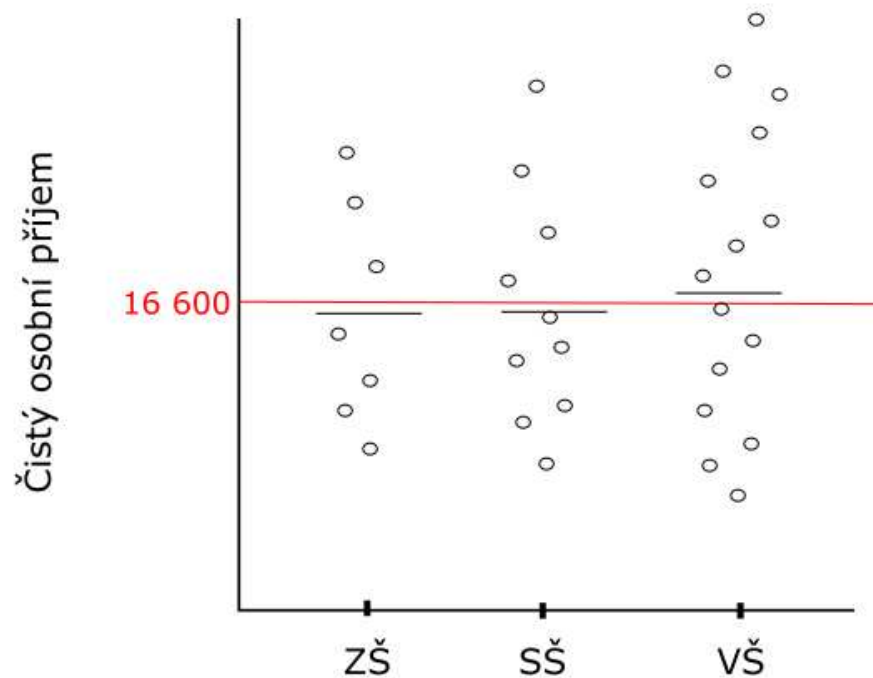


# ANOVA – Ideální typy

Zámítáme  $H_0$



Nezamítáme  $H_0$



# ANOVA

- ANOVA samotná není schopná identifikovat, mezi kterými skupinami je rozdíl -> Post hoc testy pro párové srovnání všech skupin
- ALE nelze použít klasické testy kvůli zvyšování/„nafukování“ pravděpodobnosti chyby 1. druhu (hladiny statistické významnosti)
- Např.:
  - Klasická hladina spolehlivosti je 5%
  - Máme 4 kategorie, srovnáváme všechny páry - > 6 párů
  - $0,05 * 6 = 0,3$  -> Pravděpodobnost chyby 1. druhu se zvedla na 30%

# ANOVA – Post hoc testy

- Post hoc testy vyrovnávají růst chyby 1. druhu
- Nejznámější: Tukeyho test (případně Games-Howell, pokud není dosažena shoda rozptylů)
- Post hoc testy jsou schopné najít rozdíly mezi jednotlivými kategoriemi

# ANOVA -

- ANOVA dokáže testovat vliv více než jedné kategoriální proměnné (tzv. multiway ANOVA).
  - Např. Závisí příjem na úrovni vzdělání a pohlaví?
- Je možné i zjistit, zda existuje vztah mezi kategoriálními proměnnými (tzv. Multiway ANOVA s interakcemi)
  - Např. Závisí příjem na úrovni vzdělání a pohlaví A existuje vztah mezi pohlavím a vzděláním
- ANOVA dokáže pracovat i s párovými daty (tzv. ANOVA s náhodnými bloky, v podstatě lepší verze párového t testu)
  - Např. Měnila se příjem domácností v posledních 5 letech?



# Kruskall-Wallisův test

- Neparametrická verze ANOVY
- Nevyžaduje shodu rozptylu ani normalitu rozdělení
- Výpočet velmi podobný klasické ANOVĚ (naměřené hodnoty nahrazeny pořadím)
  
- Nelze použít klasické post hoc testy
- Dvě alternativy:
  - Ruční korekce (např. Bonferroniho korekce)
  - Neparametrický post hoc test (např. Dunnův test)

# Kruskall a Wallis – post hoc testy

- Bonferroniho korekce:

- Hladina spolehlivosti se vydělí počtem testů
- Např.:
  - Hladina významnosti = 0,05
  - 4 kategorie -> 6 testů

$$\text{Upravená hladina významnosti} = \frac{0,05}{4} = 0,0125$$

- Nová hladina významnosti, se kterou porovnááme p hodnotu je 0,0125

- Dunnův test:

- Adjustuje p hodnoty, aby bylo možné zachovat standardní hladinu významnosti

# Neparametrické alternativy ANOVY

- Existuje neparametrická verze ANOVY s náhodnými bloky, např. Friedmanův test
- Neparametrická verze multiway ANOVY problematická (teoreticky existuje, v praxi málokdo zná, SPSS nenabízí vůbec)

# (Jednoduchá) ANOVA v SPSS

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the 'Compare Means' option is selected. A sub-menu is displayed, showing 'One-Way ANOVA...' as the chosen option. The background shows a data table with 18 rows and 3 columns: Name, Type, and Values.

|    | Name    | Type    | Values |
|----|---------|---------|--------|
| 1  | CD      | Numeric |        |
| 2  | NUTS3   | Numeric |        |
| 3  | VSO     | Numeric |        |
| 4  | OCT     | Numeric |        |
| 5  | KRAJ    | Numeric |        |
| 6  | IDE_9   | Numeric |        |
| 7  | IDE_8   | Numeric |        |
| 8  | IDE_2   | Numeric |        |
| 9  | t_VEK_5 | Numeric |        |
| 10 | t_VZD   | Numeric |        |
| 11 | IDE_6B  | Numeric |        |
| 12 | OV_1    | Numeric |        |
| 13 | PS_21A  | String  |        |
| 14 | PS_21B  | String  |        |
| 15 | IDE_1   | Numeric |        |
| 16 | EV_10   | Numeric |        |
| 17 | PS_1    | Numeric |        |
| 18 | PI_1A   | Numeric |        |

Analyze -> Compare Means  
-> One-Way ANOVA

# SPSS – One-way ANOVA

## Test of Homogeneity of Variances

IDE.10a Osobní čistý měsíční příjem

| Levene Statistic | df1 | df2 | Sig. |
|------------------|-----|-----|------|
| 11,403           | 3   | 596 | ,000 |

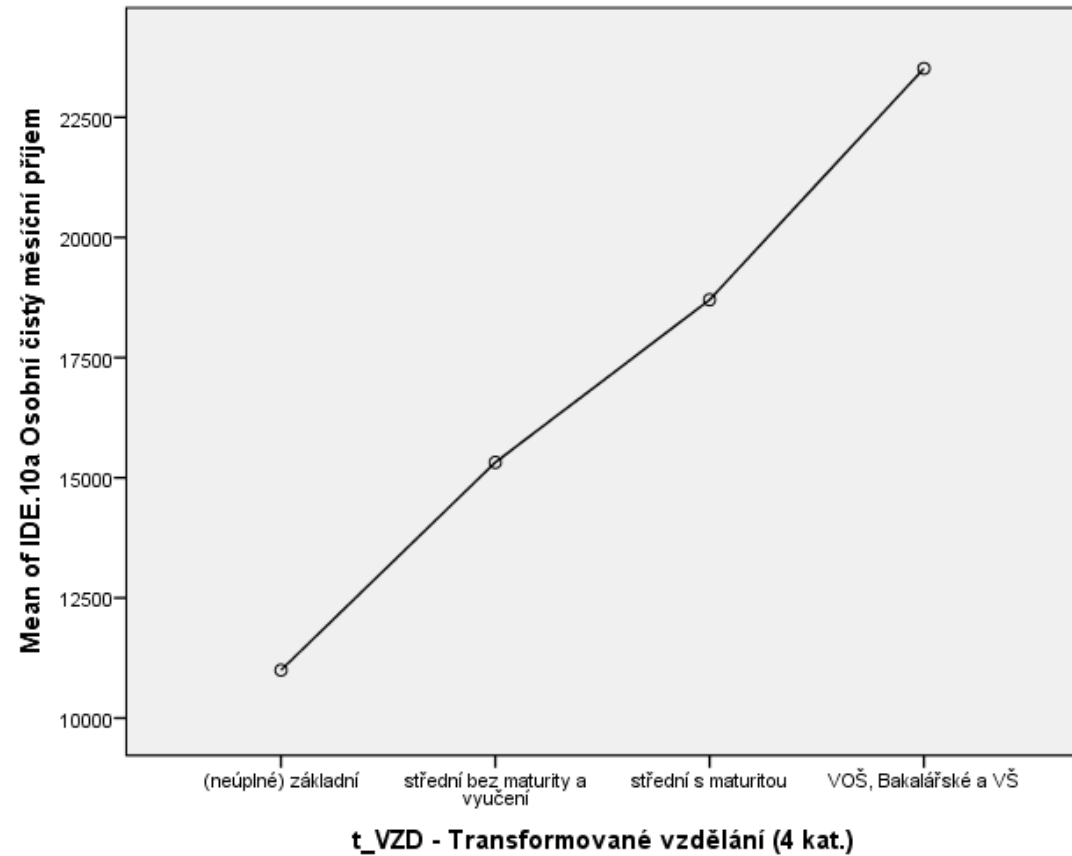
## ANOVA

IDE.10a Osobní čistý měsíční příjem

|                | Sum of Squares | df  | Mean Square | F      | Sig. |
|----------------|----------------|-----|-------------|--------|------|
| Between Groups | 8274481635     | 3   | 2758160545  | 40,237 | ,000 |
| Within Groups  | 4,085E+10      | 596 | 68548402,48 |        |      |
| Total          | 4,913E+10      | 599 |             |        |      |

**Nezapomenout na testování  
Předpokladů!**

## Means Plots



# SPSS – ANOVA post hoc testy

## Multiple Comparisons

Dependent Variable: IDE.10a Osobní čistý měsíční příjem

|                      | (I) t_VZD -<br>Transformované vzdělání<br>(4 kat.) | (J) t_VZD -<br>Transformované vzdělání<br>(4 kat.) | Mean<br>Difference (I-<br>J) | Std. Error | Sig.    | 95% Confidence Interval |             |
|----------------------|--|--|------------------------------|------------|---------|-------------------------|-------------|
|                      |  |  |                              |            |         | Lower Bound             | Upper Bound |
| Tukey HSD            | (neúplné) základní                                 | střední bez maturity a vyučení                     | -4323,202 <sup>*</sup>       | 1096,034   | ,001    | -7146,91                | -1499,50    |
|                      |  | střední s maturitou                                | -7707,599 <sup>*</sup>       | 1122,041   | ,000    | -10598,31               | -4816,89    |
|                      |  | VOŠ, Bakalářské a VŠ                               | -12518,878 <sup>*</sup>      | 1251,987   | ,000    | -15744,36               | -9293,39    |
|                      | střední bez maturity a vyučení                     | (neúplné) základní                                 | 4323,202 <sup>*</sup>        | 1096,034   | ,001    | 1499,50                 | 7146,91     |
|                      |  | střední s maturitou                                | -3384,397 <sup>*</sup>       | 810,161    | ,000    | -5471,61                | -1297,18    |
|                      |  | VOŠ, Bakalářské a VŠ                               | -8195,676 <sup>*</sup>       | 982,271    | ,000    | -10726,29               | -5665,06    |
|                      | střední s maturitou                                | (neúplné) základní                                 | 7707,599 <sup>*</sup>        | 1122,041   | ,000    | 4816,89                 | 10598,31    |
|                      |  | střední bez maturity a vyučení                     | 3384,397 <sup>*</sup>        | 810,161    | ,000    | 1297,18                 | 5471,61     |
|                      |  | VOŠ, Bakalářské a VŠ                               | -4811,279 <sup>*</sup>       | 1011,208   | ,000    | -7416,45                | -2206,11    |
| VOŠ, Bakalářské a VŠ | (neúplné) základní                                 | 12518,878 <sup>*</sup>                             | 1251,987                     | ,000       | 9293,39 | 15744,36                |             |
|                      | střední bez maturity a vyučení                     | 8195,676 <sup>*</sup>                              | 982,271                      | ,000       | 5665,06 | 10726,29                |             |
|                      | střední s maturitou                                | 4811,279 <sup>*</sup>                              | 1011,208                     | ,000       | 2206,11 | 7416,45                 |             |

# SPSS – Kruskal Wallis

Analyze -> Nonparametric tests -> Legacy Dialogs

ALE pokud chcete i post hoc testy, tak

Analyze -> Nonparametric tests -> Independent samples

Nonparametric Tests: Two or More Independent Samples

Objective Fields Settings

Use predefined roles  
 Use custom field assignments

Fields:

Sort: None

- CD číslo dotazníku
- NUTS 3
- Velikostní skupina obce
- Osobní číslo tazatele
- kraje ČR
- IDE 9 Okres
- IDE 8 Pohlaví
- IDE 2 Věk
- t\_VEK\_5 - transformovaný věk -
- IDE 6b Vzdělání
- OV.1 Spokojenost s životem
- PS.21a Nejdůležitější celospole
- PS.21b Nejdůležitější celospole
- IDE 1 Životní úroveň domácnost
- EV.10 Hodnocení současné ek
- PS.1 Spokojenost se současn
- PI.1a Důvěra - prezident
- PI.1b Důvěra - vláda ČR
- PI.1c Důvěra - Poslanecká sně

Test Fields:

- IDE.10a Osobní čistý měsíční

Groups:

- t\_VZD - Transformované vzd...

Run Paste Reset Cancel Help

Nonparametric Tests: Two or More Independent Samples

Objective Fields Settings

Select an item:

- Choose Tests
- Test Options
- User-Missing Values

Automatically choose the tests based on the data  
 Customize tests

Compare Distributions across Groups

- Mann-Whitney U (2 samples)
- Kruskal-Wallis 1-way ANOVA (k samples)  
Multiple comparisons: All pairwise
- Kolmogorov-Smirnov (2 samples)
- Test for ordered alternatives (Jonckheere-Terpstra for k samples)  
Hypothesis order: Smallest to largest
- Test sequence for randomness (Wald-Wolfowitz for 2 samples)  
Multiple comparisons: All pairwise

Compare Ranges across Groups

- Moses extreme reaction (2 samples)
  - Compute outliers from sample
  - Custom number of outliers  
Outliers: 1

Compare Medians across Groups

- Median test (k samples)
  - Pooled sample median
  - Custom  
Median: 0

Multiple comparisons: All pairwise

Estimate Confidence Interval across Groups

- Hodges-Lehmann estimate (2 samples)

Run Paste Reset Cancel Help

# SPSS – Kruskal Wallis + post hoc

## Independent-Samples Kruskal-Wallis Test Summary

|                               |                      |
|-------------------------------|----------------------|
| Total N                       | 600                  |
| Test Statistic                | 128,009 <sup>a</sup> |
| Degree Of Freedom             | 3                    |
| Asymptotic Sig.(2-sided test) | ,000                 |

a. The test statistic is adjusted for ties.

## Pairwise Comparisons of t\_VZD - Transformované vzdělání (4 kat.)

| Sample 1-Sample 2   | Test Statistic | Std. Error | Std. Test Statistic | Sig. | Adj. Sig. |
|---|----------------|------------|---------------------|------|-----------|
| (neúplné) základní-<br>střední bez maturity a<br>vyučení  | -110,058       | 22,931     | -4,799              | ,000 | ,000      |
| (neúplné) základní-<br>střední s maturitou                | -191,097       | 23,476     | -8,140              | ,000 | ,000      |
| (neúplné) základní-VOŠ,<br>Bakalářské a VŠ                | -268,314       | 26,194     | -10,243             | ,000 | ,000      |
| střední bez maturity a<br>vyučení-střední s<br>maturitou  | -81,040        | 16,950     | -4,781              | ,000 | ,000      |
| střední bez maturity a<br>vyučení-VOŠ, Bakalářské<br>a VŠ | -158,256       | 20,551     | -7,701              | ,000 | ,000      |
| střední s maturitou-VOŠ,<br>Bakalářské a VŠ               | -77,217        | 21,157     | -3,650              | ,000 | ,002      |

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.