University of Ljubljana
SEB School of Economics and Business

RMT

LECTURE 5
HYPOTHESIS TESTING

2023/2024

Denis Marinšek

EQUIS ACCREDITED
AACSB ACCREDITED
Association of MBAs ACCREDITED

HYPOTHESIS TESTING

HYPOTHESIS TESTING

1. Research Question: Is there a difference in the share of online banking users between younger and older generations?
2. Research Hypothesis: A higher percentage of younger generations use online banking compared to older generations.
3. Collect the data, find the appropriate statistical test, and see if the data supports your hypothesis.

- Hypothesis can be tested with parametric or with nonparametric statistical tests. What is the difference?

- When to use non-parametrical tests:
  - If assumptions of parametrical tests are violated (for example assumption about normal distribution) or if there are serious problems of outliers.
  - If variables are measured on ordinal measurement scale.

University *of Ljubljana*

SEB SCHOOL OF ECONOMICS AND BUSINESS

HYPOTHESIS TESTING

Parametrical test		Alternative non-parametrical test
Hypothesis about the population arithmetic mean	*t*-test Effect size: *CohenD – Sawilowsky, 2009*	Wilcoxon Signed Rank Test (at least interval variables) – Hypothesis about the population median Effect size: *Bisserial Correlation – Funder, 2019*
Hypothesis about the difference between two population arithmetic means (dependent or paired samples)	Paired *t*-test Effect size: *CohenD – Sawilowsky, 2009*	Sign Test (ordinal variables) – Special version of Binomial test Wilcoxon Signed Rank Test (at least interval variables) Effect size: *Bisserial Correlation – Funder, 2019*
Hypothesis about the difference between two population arithmetic means (independent samples)	Independent *t*-test with Welch correction Effect size: *CohenD – Sawilowsky, 2009*	Wilcoxon Rank Sum Test, sometimes called Mann-Whitney U-test (at least ordinal variables) Effect size: *Bisserial Correlation – Funder, 2019*
Hypothesis about the equality of three or more population arithmetic means for dependent samples	One-way repeated measures ANOVA, rANOVA Effect size: η^2 - *Cohen, 1992*	Friedman ANOVA (at least ordinal variables) Effect size: *Kendall W - Landis, 1977*

HYPOTHESIS TESTING

Parametrical test		Alternative non-parametrical test
Hypothesis about the equality of three or more population arithmetic means for independent samples	*One-way ANOVA* Effect size: η^2 - *Cohen, 1992* *Welch Heteroscedastic F-test* – robustness check	*Kruskal-Wallis Rank Sum Test* (at least ordinal varibles) Effect size: η^2 - *Field, 2013*
Hypothesis about the population proportion	χ^2-test (Test of proportion)	Binomial test
Hypothesis about association between two categorical variables	χ^2-test Effect size: CramerV statistics (*Funder, 2019*) or Odds ratio (*Chen et al., 2010*)	Fisher`s Exact Probability Test
Hypothesis about population distribution for given probabilities	χ^2-test	
Hypothesis about the equality in two population proportions	χ^2-test (Test of proportion)	

HYPOTHESIS TESTING

Sawilowsky (2009) ("`sawilowsky2009`")

- **d < 0.1** - Tiny
- **0.1 <= d < 0.2** - Very small
- **0.2 <= d < 0.5** - Small
- **0.5 <= d < 0.8** - Medium
- **0.8 <= d < 1.2** - Large
- **1.2 <= d < 2** - Very large

Funder & Ozer (2019) ("`funder2019`"; default)

- **r < 0.05** - Tiny
- **0.05 <= r < 0.1** - Very small
- **0.1 <= r < 0.2** - Small
- **0.2 <= r < 0.3** - Medium
- **0.3 <= r < 0.4** - Large
- **r >= 0.4** - Very large

Landis & Koch (1977) ("`landis1977`"; default)

- **0.00 <= w < 0.20** - Slight agreement
- **0.20 <= w < 0.40** - Fair agreement
- **0.40 <= w < 0.60** - Moderate agreement
- **0.60 <= w < 0.80** - Substantial agreement
- **w >= 0.80** - Almost perfect agreement

Cohen (1992) ("`cohen1992`") applicable to one-way anova, or to *partial* eta / omega / epsilon squared in multi-way anova.

- **ES < 0.02** - Very small
- **0.02 <= ES < 0.13** - Small
- **0.13 <= ES < 0.26** - Medium
- **ES >= 0.26** - Large

Chen et al. (2010) ("`chen2010`"; default)

- **OR < 1.68** - Very small
- **1.68 <= OR < 3.47** - Small
- **3.47 <= OR < 6.71** - Medium
- **OR >= 6.71 ** - Large

HYPOTHESIS ABOUT THE POPULATION ARITHMETIC MEAN

HYPOTHESIS ABOUT THE POPULATION MEDIAN

HYPOTHESIS ABOUT THE ARITHMETIC MEAN

Testing the population arithmetic mean with the chosen parameter.

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

$$t = \frac{\bar{y} - \mu_0}{se(\bar{y})}$$

$$se(\bar{y}) = \frac{s_y}{\sqrt{n}} \qquad df = n - 1$$

Conditions and assumptions:

- Variable is numeric.
- Normality - variable on the population is normally distributed.
- No outliers.

HYPOTHESIS ABOUT THE ARITHMETIC MEAN

Example: Body mass

A researcher examined the effects of online schooling on the body weight of ninth graders. Based on data from the 2018/2019 school year, the average weight was 59.5 kg and the median was 58.3 kg. At the beginning of the 2021/2022 school year, the researchers randomly selected 50 ninth graders and measured their weight. What conclusion can the researcher draw?

```
mydata <- read.table("./Body mass.csv", header=TRUE, sep=";", dec=",")
head(mydata)
```

```
##   ID Mass
## 1  1 62.1
## 2  2 64.5
## 3  3 56.5
## 4  4 53.4
## 5  5 61.3
## 6  6 62.2
```

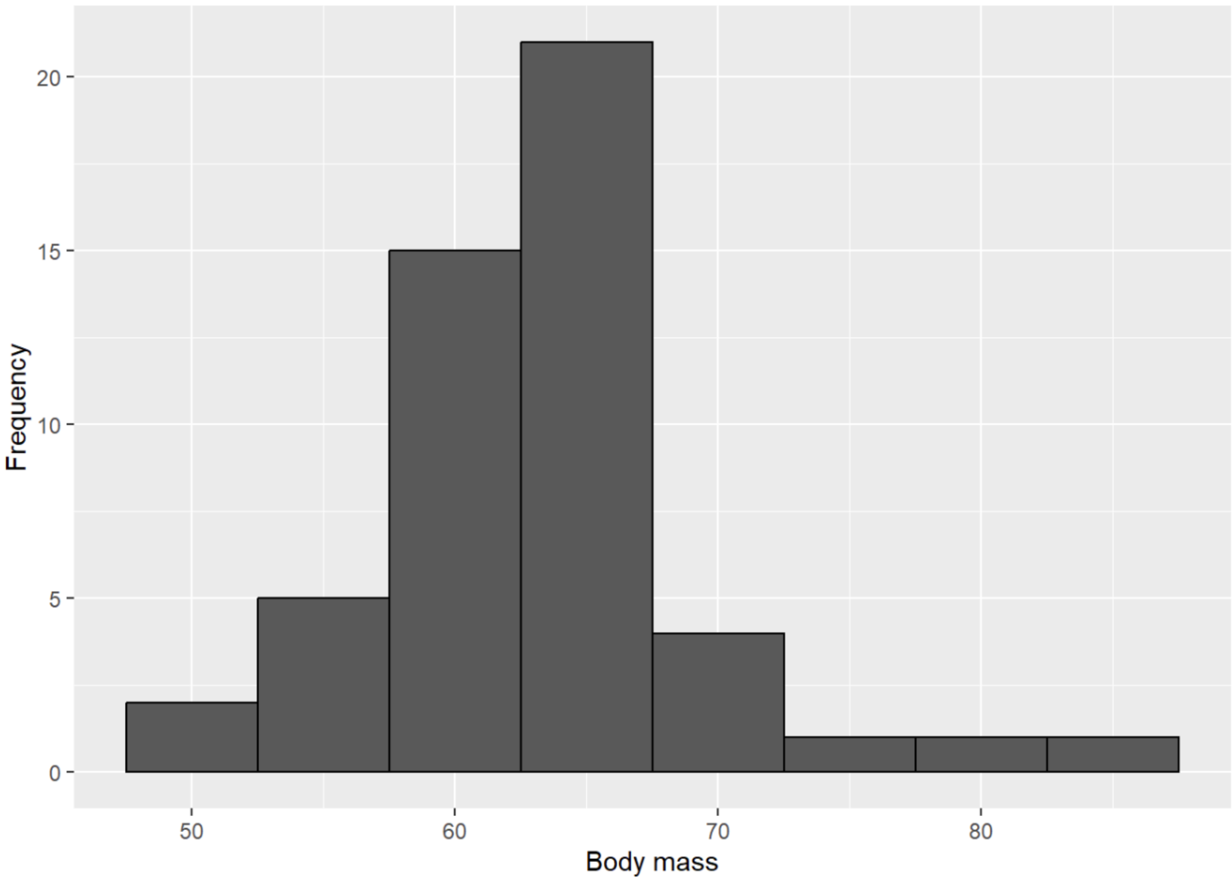
Description of data:

- ID: Identificator.
- Mass: Body mass in kg.

HYPOTHESIS ABOUT THE ARITHMETIC MEAN

Example: Body mass

```
library(ggplot2)
ggplot(mydata, aes(x = Mass)) +
  geom_histogram(binwidth = 5, colour = "black") +
  ylab("Frequency") +
  xlab("Body mass")
```



HYPOTHESIS ABOUT THE ARITHMETIC MEAN

Example: Body mass

```
mean(mydata$Mass)
```

```
## [1] 62.876
```

```
sd(mydata$Mass)
```

```
## [1] 6.011403
```

```
nrow(mydata)
```

```
## [1] 50
```

HYPOTHESIS ABOUT THE ARITHMETIC MEAN

Example: Body mass

```
ybar = mean(mydata$Mass); sd = sd(mydata$Mass); n = nrow(mydata)

se = sd/sqrt(n)

ybar_lower_t = ybar + qt(0.025, df=n-1)*se
ybar_upper_t = ybar + qt(0.975, df=n-1)*se
```

95-% confidence interval: $61.168 < \mu < 64.584$

```
t.test(mydata$Mass,
       mu = 59.5,
       alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  mydata$Mass
## t = 3.9711, df = 49, p-value = 0.000234
## alternative hypothesis: true mean is not equal to 59.5
## 95 percent confidence interval:
##  61.16758 64.58442
## sample estimates:
## mean of x
##    62.876
```

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

HYPOTHESIS ABOUT THE ARITHMETIC MEAN

Example: Body mass

```
#install.packages("effectsize")
library(effectsize)
```

```
effectsize::cohens_d(mydata$Mass, mu = 59.5)
```

```
## Cohen's d |       95% CI
## ------------------------
## 0.56      | [0.26, 0.86]
##
## - Deviation from a difference of 59.5.
```

```
interpret_cohens_d(0.56, rules = "sawilowsky2009")
```

```
## [1] "medium"
## (Rules: sawilowsky2009)
```

Sawilowsky (2009) ("sawilowsky2009")

- **d < 0.1** - Tiny
- **0.1 <= d < 0.2** - Very small
- **0.2 <= d < 0.5** - Small
- **0.5 <= d < 0.8** - Medium
- **0.8 <= d < 1.2** - Large
- **1.2 <= d < 2** - Very large

$$d = \frac{\bar{y} - \mu_0}{s_y} = \frac{62.876 - 59.5}{6.011} = 0.56$$

Conclusions:

Based on the sample data, we found that the average weight of ninth graders at the beginning od theta 2021/2022 school year was 62.88 kg and has increased compared to the 2018/2019 school year ($p < 0.001$, $d = 0.56$ – medium sized effect).

HYPOTHESIS ABOUT THE MEDIAN

Testing the population median with the chosen parameter.

H_0: $Me = Me_0$
H_1: $Me \neq Me_0$

$$z = \frac{T^+ - \bar{T}}{SE_T}$$

$$\bar{T} = \frac{n(n+1)}{4} \qquad SE_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Requirement:

- Variable must be numeric.

HYPOTHESIS ABOUT THE MEDIAN

Example: Body mass

```
median(mydata$Mass)
```

```
## [1] 62.8
```

```
wilcox.test(mydata$Mass,
            mu = 58.3,
            correct = FALSE)
```

```
##
##  Wilcoxon signed rank test
##
## data:  mydata$Mass
## V = 1116, p-value = 3.847e-06
## alternative hypothesis: true location is not equal to 58.3
```

HYPOTHESIS ABOUT THE MEDIAN
Example: Body mass

```
effectsize(wilcox.test(mydata$Mass,
           mu = 58.3,
           correct = FALSE))
```

```
## r (rank biserial) |       95% CI
## ---------------------------------
## 0.75              | [0.58, 0.86]
##
## - Deviation from a difference of 58.3.
```

```
interpret_rank_biserial(0.75, rules = "funder2019")
```

```
## [1] "very large"
## (Rules: funder2019)
```

Funder & Ozer (2019) ("`funder2019`"; default)

- **r < 0.05** - Tiny
- **0.05 <= r < 0.1** - Very small
- **0.1 <= r < 0.2** - Small
- **0.2 <= r < 0.3** - Medium
- **0.3 <= r < 0.4** - Large
- **r >= 0.4** - Very large

Conclusions:

Based on the sample data, we found that the median weight of ninth graders at the beginning of the 2021/2022 school year was 62.8 kg and has increased compared to the 2018/2019 school year ($p < 0.001$, $r = 0.75$ – very large effect).

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO POPULATION ARITHMETIC MEANS

SIGN TEST

WILCOXON SIGNED RANK TEST

WILCOXON RANK SUM TEST

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

t-test

t-test for repeated measures (Paired samples *t*-test):

Pairs of data belong to the same unit, measured at different conditions.

t-test for independent measures (Independent samples *t*-test):

Data belong to two different groups of units (from two different populations). Each unit is measured once.

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Assumptions

Independent Samples *t*-test and Paired Samples *t*-test are parametric tests:

- Variable is numeric.
- Normality:
  - Paired Samples *t*-test: Differences on the population are normally distributed.
  - Independent Samples *t*-test: The distribution of the variable is normal in both populations.
- For independent Samples *t*-test:
  - The data must come from two independent populations,
  - Variable has the same variance in both populations – since this assumption is often violated, we apply Welch correction.

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Paired samples *t*-test

H_0: $\mu_1 = \mu_2$ or $\mu_d = 0$

H_1: $\mu_1 \neq \mu_2$ or $\mu_d \neq 0$

$$t = \frac{\bar{d} - \mu_d}{se(\bar{d})}$$

$$d_i = y_{i1} - y_{i2} \qquad \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i\,; \quad se(\bar{d}) = \frac{s_d}{\sqrt{n}};$$

$$s_d = \sqrt{s_d^2}$$

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

$$df = n - 1$$

University of Ljubljana
SEB SCHOOL OF ECONOMICS AND BUSINESS

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Independent samples *t*-test ($\sigma_1^2 \neq \sigma_2^2$)

H_0: $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 \neq \mu_2$ or $\mu_1 - \mu_2 \neq 0$

$$t = \frac{\bar{y}_1 - \bar{y}_2}{se(\bar{y}_1 - \bar{y}_2)}$$

$$se(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^{\mathbf{2}}}{\left(\frac{s_1^2}{n_1}\right)^{\mathbf{2}} \frac{1}{n_1 - 1} + \left(\frac{s_2^2}{n_2}\right)^{\mathbf{2}} \frac{1}{n_2 - 1}}$$

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Paired samples *t*-test – Example 1: Literacy

In elementary school, it was decided to test students' literacy skills. 92 sixth grade students were randomly selected, and each of them took a simple test on reading comprehension (maximum 100 points) and writing knowledge (maximum 100 points). Can we say that the average scores obtained in reading comprehension and written knowledge are different? If so, where did the students perform better?

```
mydata <- read.table("./Literacy.csv", header=TRUE, sep=";", dec=",")
head(mydata)
```

```
##   ID Reading Writing
## 1  1   81.50   81.44
## 2  2   85.25   73.27
## 3  3   86.88   84.24
## 4  4   88.68   73.16
## 5  5   77.30   80.45
## 6  6   87.53   70.79
```

Description:

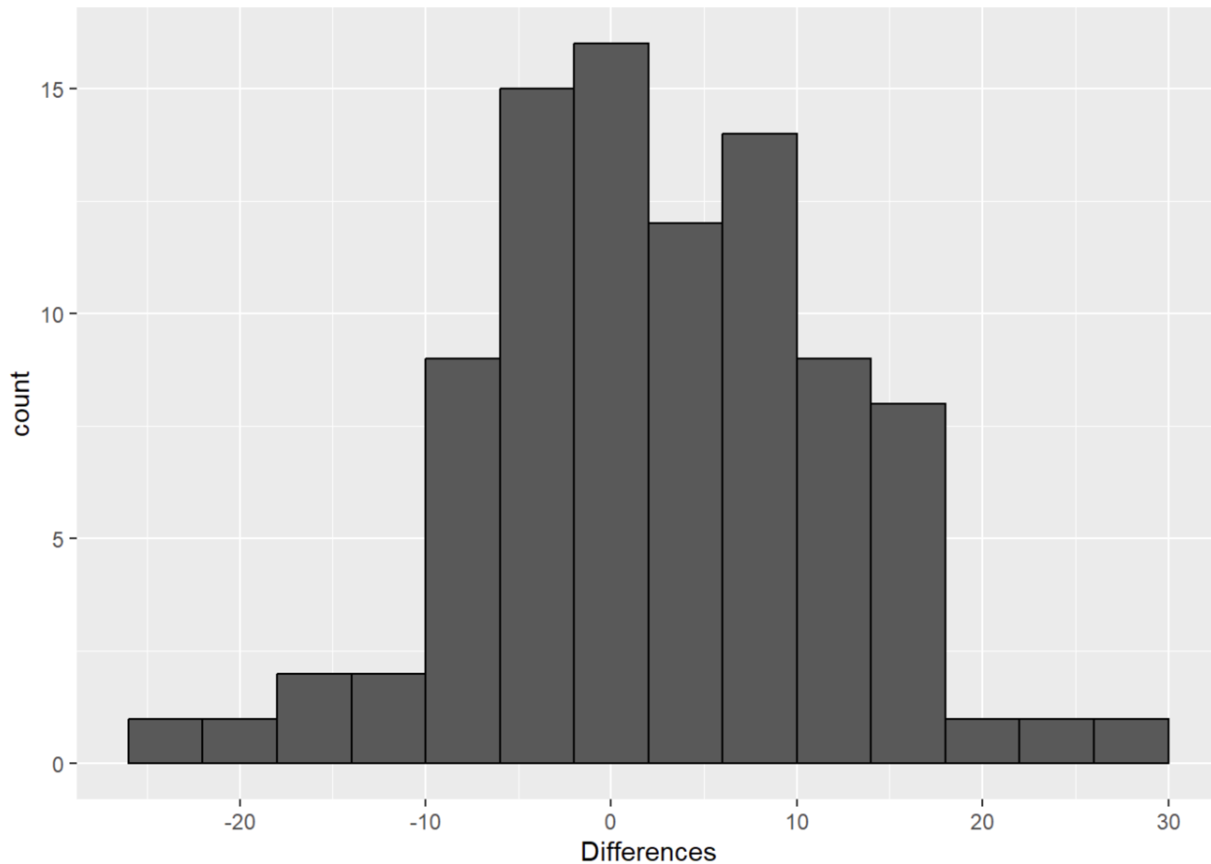
- ID: Student ID
- Reading: Result at reading
- Writing: Result at writing

University of Ljubljana
SEB SCHOOL OF ECONOMICS AND BUSINESS

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Paired samples *t*-test – Example: Literacy

```
mydata$Difference <- mydata$Reading - mydata$Writing

library(ggplot2)
ggplot(mydata, aes(x = Difference)) +
  geom_histogram(binwidth = 4, color = "black") +
  xlab("Differences")
```



HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Paired samples *t*-test – Example: Literacy

```
shapiro.test(mydata$Difference)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mydata$Difference
## W = 0.99245, p-value = 0.8842
```

```
library(psych)
```

```
describe(mydata[ , -1])
```

```
##            vars  n  mean   sd median trimmed  mad    min    max range  skew kurtosis   se
## Reading       1 92 80.99 7.94  81.34   81.20 7.67  55.11 100.00 44.89 -0.43     0.90 0.83
## Writing       2 92 78.47 5.63  78.75   78.43 6.31  64.06  92.30 28.24 -0.03    -0.48 0.59
## Difference    3 92  2.51 9.37   2.05    2.50 9.10 -25.51  28.33 53.84 -0.03     0.26 0.98
```

University *of Ljubljana*

SEB SCHOOL OF ECONOMICS AND BUSINESS

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Paired samples *t*-test – Example: Literacy

```
t.test(mydata$Reading, mydata$Writing,
       paired = TRUE,
       alternative = "two.sided")
```

```
##
##  Paired t-test
##
## data:  mydata$Reading and mydata$Writing
## t = 2.5711, df = 91, p-value = 0.01176
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.5714427 4.4539920
## sample estimates:
## mean difference
##        2.512717
```

```
library(effectsize)
```

```
cohens_d(mydata$Difference)
```

```
## Cohen's d |       95% CI
## ------------------------
## 0.27      | [0.06, 0.48]
```

```
interpret_cohens_d(0.27, rules = "sawilowsky2009")
```

```
## [1] "small"
## (Rules: sawilowsky2009)
```

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Paired samples *t*-test – Example: Literacy

Conclusions:

Based on sample data we found the difference between average points received at reading ($\bar{y} = 80.99$) and average points received at writing ($\bar{y} = 78.47$). The difference is statistically significant at $p = 0.012$ (effect size is small, $d = 0.27$).

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 2: Grades

We have received data from 8 elementary school students about the result of the test in mathematics and English they wrote last week. Can we claim that there is a difference in the knowledge of elementary school students in the two subjects mentioned?

ID	Math		English		Sign	Difference	Abs(Difference)	Rank
1	Excl	90	Good	70	Plus	20	20	5
2	Good	72	Very good	83	Minus	-11	11	4
3	Very good	85	Excellent	95	Minus	-10	10	3
4	Fair	60	Very good	82	Minus	-22	22	6
5	Very good	70	Good	68	Plus	2	2	1,5
6	Very good	74	Very good	72	/	2	2	1,5
7	Fair	62	Excelent	95	Minus	-33	33	8
8	Fair	57	Very good	88	Minus	-31	31	7

Positive signs	2
Negative signs	5
Sum of positive signs	8
Sum of negative signs	28

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 2: Grades

```
mydata <- read.table("./Grades.csv", header=TRUE, sep=";", dec=",")
print(mydata)
```

```
##   ID     Math Math_points  English English_points
## 1  1     Excl          90     Good             70
## 2  2     Good          72 VeryGood             83
## 3  3 VeryGood          85     Excl             95
## 4  4     Fair          60 VeryGood             82
## 5  5 VeryGood          70     Good             68
## 6  6 VeryGood          74 VeryGood             72
## 7  7     Fair          62     Excl             95
## 8  8     Fair          57 VeryGood             88
```

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 2: Grades

```
library(dplyr)
```

```
mydata <- mydata %>%
  mutate(Math_code = dplyr::recode(mydata$Math,
    "Excl" = 5, "VeryGood" = 4, "Good" = 3, "Fair" = 2))

mydata <- mydata %>%
  mutate(English_code = dplyr::recode(mydata$English,
    "Excl" = 5, "VeryGood" = 4, "Good" = 3, "Fair" = 2))

print(mydata)
```

```
##   lD     Math Math_points  English English_points Math_code English_code
## 1  1     Excl          90     Good             70         5            3
## 2  2     Good          72 VeryGood             83         3            4
## 3  3 VeryGood          85     Excl             95         4            5
## 4  4     Fair          60 VeryGood             82         2            4
## 5  5 VeryGood          70     Good             68         4            3
## 6  6 VeryGood          74 VeryGood             72         4            4
## 7  7     Fair          62     Excl             95         2            5
## 8  8     Fair          57 VeryGood             88         2            4
```

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 2: Grades

Sign Test

```
mydata$Difference <- mydata$Math_code - mydata$English_code

table(mydata$Difference)
```

```
##
## -3 -2 -1  0  1  2
##  1  2  2  1  1  1
```

```
binom.test(x = 5,
           n = 7,
           p = 0.5,
           alternative = "two.sided")
```

```
##
##  Exact binomial test
##
## data:  5 and 7
## number of successes = 5, number of trials = 7, p-value = 0.4531
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2904209 0.9633074
## sample estimates:
## probability of success
##              0.7142857
```

$$\binom{n}{k} p^k (1-p)^{n-k}$$

$$\binom{7}{5} 0{,}5^5 (1-0{,}5)^{7-5} = 21 \cdot \frac{1}{128} = 0{,}1641$$
$$\binom{7}{6} 0{,}5^6 (1-0{,}5)^{7-6} = 7 \cdot \frac{1}{128} = 0{,}0547$$
$$\binom{7}{7} 0{,}5^7 (1-0{,}5)^{7-7} = 1 \cdot \frac{1}{128} = 0{,}0078$$

$$\binom{7}{2} 0{,}5^3 (1-0{,}5)^{7-2} = 21 \cdot \frac{1}{128} = 0{,}1641$$
$$\binom{7}{1} 0{,}5^2 (1-0{,}5)^{7-1} = 7 \cdot \frac{1}{128} = 0{,}0547$$
$$\binom{7}{0} 0{,}5^1 (1-0{,}5)^{7-0} = 1 \cdot \frac{1}{128} = 0{,}0078$$

$$\sum = 0{,}4531$$

University of Ljubljana

SEB School of Economics and Business

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 2: Grades

Wilcoxon Signed Rank Test

```
wilcox.test(mydata$Math_points,  mydata$English_points,
            paired = TRUE,
            correct = FALSE,
            exact = FALSE,
            alternative = "two.sided")
```

```
##
##  Wilcoxon signed rank test
##
## data:  mydata$Math_points and mydata$English_points
## V = 8, p-value = 0.1609
## alternative hypothesis: true location shift is not equal to 0
```

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 3: Party

10 participants attend the party two weekends in a row. On the first weekend they consume alcohol and cannabis, on the second weekend they consume only alcohol. The day after the party, each of them rates how severe the headache is on a scale of 1 to 7 (higher score - higher intensity). Check if there are differences in the intensity of the headache after the first and second weekend.

```
mydata <- read.table("./Party.csv", header=TRUE, sep=";", dec=",")
print(mydata)
```

```
##    ID Cannabis Alcohol
## 1   1        3       3
## 2   2        7       3
## 3   3        3       4
## 4   4        4       3
## 5   5        5       3
## 6   6        4       2
## 7   7        5       3
## 8   8        3       2
## 9   9        2       1
## 10 10        6       2
```

Description:

- ID: Person ID
- Cannabis: The headaches (1 - weak to 7 - strong)
- Alcohol: The headaches (1 - weak to 7 - strong)

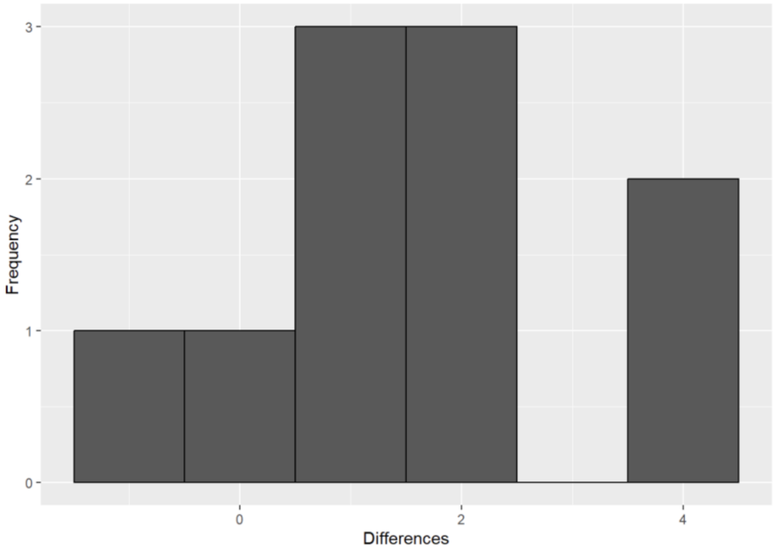
HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 3: Party

```
mydata$Difference <- mydata$Cannabis - mydata$Alcohol

describe(mydata[, -1])
```

```
##            vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## Cannabis      1 10  4.2 1.55    4.0    4.12 1.48   2   7     5  0.35    -1.22 0.49
## Alcohol       2 10  2.6 0.84    3.0    2.62 0.74   1   4     3 -0.28    -0.84 0.27
## Difference    3 10  1.6 1.58    1.5    1.62 0.74  -1   4     5  0.14    -1.08 0.50
```

```
library(ggplot2)
ggplot(mydata, aes(x=Difference)) +
  geom_histogram(binwidth = 1, color = "black") +
  xlab("Differences") +
  ylab("Frequency")
```



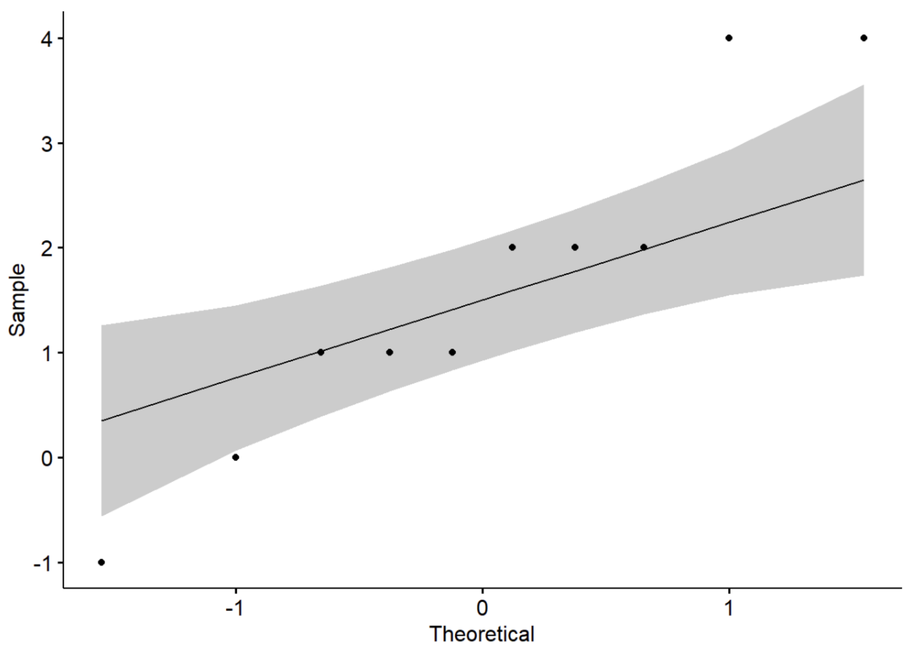
HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 3: Party

```
shapiro.test(mydata$Difference)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mydata$Difference
## W = 0.92897, p-value = 0.4378
```

```
library(ggpubr)
ggqqplot(mydata$Difference)
```



HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 3: Party

```
wilcox.test(mydata$Cannabis, mydata$Alcohol,
            paired = TRUE,
            correct = FALSE,
            exact = FALSE,
            alternative = "two.sided")
```

```
##
##  Wilcoxon signed rank test
##
## data:  mydata$Cannabis and mydata$Alcohol
## V = 42.5, p-value = 0.01634
## alternative hypothesis: true location shift is not equal to 0
```

```
library(effectsize)
```

```
effectsize(wilcox.test(mydata$Cannabis, mydata$Alcohol,
                       paired = TRUE,
                       correct = FALSE,
                       exact = FALSE,
                       alternative = "two.sided"))
```

```
## r (rank biserial) |       95% CI
## ------------------------------
## 0.89              | [0.62, 0.97]
```

Funder & Ozer (2019) ("`funder2019`"; default)

- **r < 0.05** - Tiny
- **0.05 <= r < 0.1** - Very small
- **0.1 <= r < 0.2** - Small
- **0.2 <= r < 0.3** - Medium
- **0.3 <= r < 0.4** - Large
- **r >= 0.4** - Very large

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 3: Party

Conclusions:

Using the sample data, we find that the intensity of the headache with cannabis and alcohol differs from the intensity of the headache with alcohol alone (p = 0.017). More severe headaches are expected after the consumption of alcohol and cannabis together (effect size r = 0.89).

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Independent samples *t*-test – Example 4: Math

The math professor was interested in whether there is a difference in exam scores between students who exercise at least three times a week and those who do not. Over the past few years, she has collected data on first exam scores, with students also indicating whether they are actively exercising.

```
mydata <- read.table("./Math.csv", header=TRUE, sep=";", dec=",")
head(mydata)
```

```
##   ID Active  Math
## 1  1      1 93.78
## 2  2      0 86.66
## 3  3      1 85.69
## 4  4      0 85.57
## 5  5      0 85.32
## 6  6      0 85.19
```

Description:

- ID: ID of a student
- Active: Is student regularly exercising?
- Math: Result of math exam

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Independent samples *t*-test – Example 4: Math

```
mydata$ActiveF <- factor(mydata$Active,
                         levels = c(0, 1),
                         labels = c("No", "Yes"))

library(psych)
```

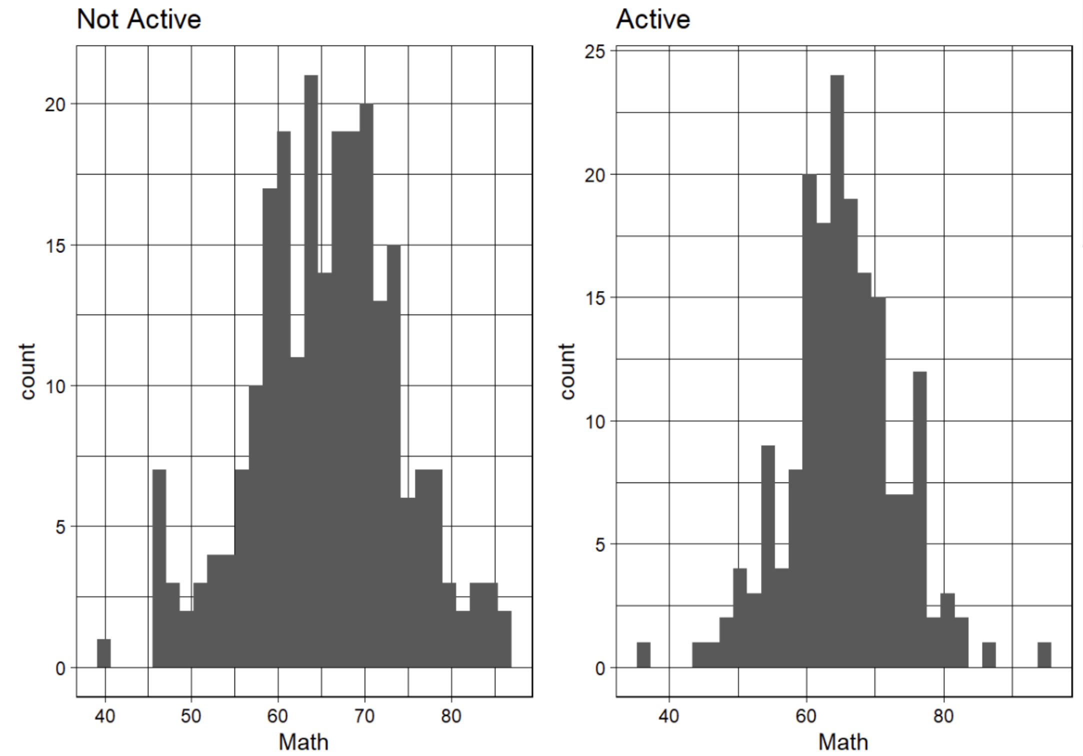
```
describeBy(mydata$Math, mydata$ActiveF)
```

```
##
##  Descriptive statistics by group
## group: No
##    vars   n  mean   sd median trimmed  mad   min   max range  skew kurtosis   se
## X1    1 242 65.58 8.58  65.78   65.72 8.04 40.38 86.66 46.28 -0.14     0.06 0.55
## ------------------------------------------------------------------------------------------
## group: Yes
##    vars   n  mean   sd median trimmed  mad   min   max range  skew kurtosis   se
## X1    1 180 65.32 8.12  65.12   65.37 6.99 35.32 93.78 58.46 -0.06     1.07 0.61
```

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Independent samples *t*-test – Example 4: Math

```
library(ggplot2)

Active_no <- ggplot(mydata[mydata$ActiveF == "No",  ], aes(x = Math)) +
  theme_linedraw() +
  geom_histogram() +
  ggtitle("Not Active")

Active_yes <- ggplot(mydata[mydata$ActiveF == "Yes",  ], aes(x = Math)) +
  theme_linedraw() +
  geom_histogram() +
  ggtitle("Active")

library(ggpubr)
ggarrange(Active_no, Active_yes,
          ncol = 2, nrow = 1)
```



HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Independent samples *t*-test – Example 4: Math

```
t.test(mydata$Math ~ mydata$ActiveF,
       paired = FALSE,
       var.equal = FALSE,
       alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  mydata$Math by mydata$ActiveF
## t = 0.31548, df = 396.45, p-value = 0.7526
## alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
## 95 percent confidence interval:
##  -1.351309  1.867895
## sample estimates:
##  mean in group No mean in group Yes
##          65.57818          65.31989
```

HYPOTHESIS ABOUT THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS

Independent samples *t*-test – Example 4: Math

```
library(effectsize)
```

```
effectsize::cohens_d(mydata$Math ~ mydata$ActiveF,
                     pooled_sd = FALSE)
```

```
## Cohen's d |         95% CI
## -------------------------
## 0.03      | [-0.16, 0.22]
##
## - Estimated using un-pooled SD.
```

```
interpret_cohens_d(0.03, rules = "sawilowsky2009")
```

```
## [1] "tiny"
## (Rules: sawilowsky2009)
```

Conclusions:

Using the sample data, the professor was unable to demonstrate that the average score in mathematics differed between athletically active and non-athletically active students (the effect size is tiny, $d = 0.03$).

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 5: Shopping

You wonder if there are differences between men and women in the amount of time they spend at the shopping mall. You measured the time a person spends in a single visit to a large shopping mall in Vienna for 17 men and 20 women. What is your conclusion?

```
mydata <- read.table("./Shopping.csv", header=TRUE, sep=";", dec=",")
head(mydata)
```

```
##   ID Gender Time
## 1  1      1   22
## 2  2      1  167
## 3  3      1  160
## 4  4      1  205
## 5  5      1  279
## 6  6      1   15
```

Description:

- ID: Identifactor.
- Gender: 0 = Male, 1 = Female,
- Time: Time shopping in minutes

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS
Example 5: Shopping

```
mydata$GenderF <- factor(mydata$Gender,
                         levels = c(0, 1),
                         labels = c("Male", "Female"))
```

```
library(psych)
describeBy(mydata$Time, g = mydata$GenderF)
```

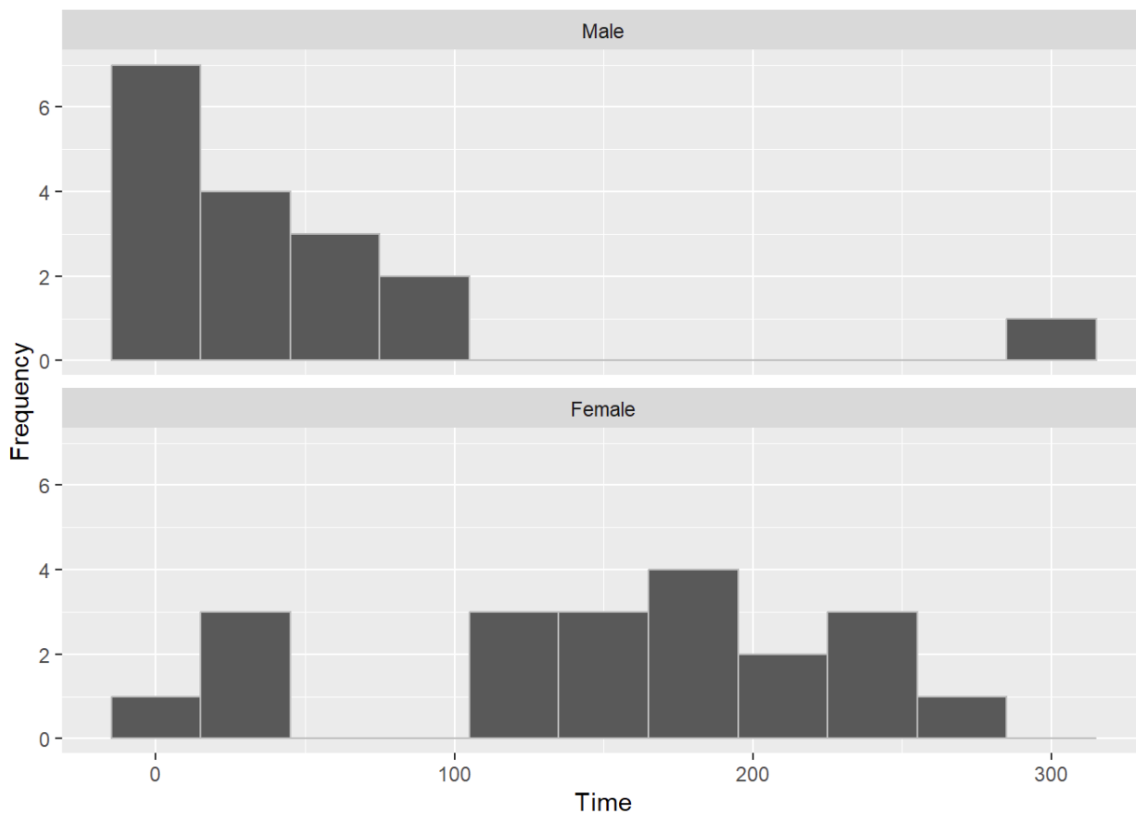
```
##
##  Descriptive statistics by group
## group: Male
##    vars  n  mean    sd median trimmed   mad min max range skew kurtosis    se
## X1    1 17 52.82 70.85     22      39 20.76   8 305   297 2.59      6.5 17.18
## ---------------------------------------------------------------------------
## group: Female
##    vars  n   mean    sd median trimmed   mad min max range  skew kurtosis    se
## X1    1 20 151.35 80.67  163.5  153.81 66.72  15 279   264 -0.39       -1 18.04
```

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 5: Shopping

```
library(ggplot2)
```

```
ggplot(mydata, aes(x = Time)) +
  geom_histogram(binwidth = 30, colour="gray") +
  facet_wrap(~GenderF, ncol = 1) +
  ylab("Frequency")
```



HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 5: Shopping

```
library(rstatix)
```

```
mydata %>%
  group_by(GenderF) %>%
  shapiro_test(Time)
```

```
## # A tibble: 2 × 4
##   GenderF variable statistic         p
##   <fct>   <chr>        <dbl>     <dbl>
## 1 Male    Time         0.602 0.0000109
## 2 Female  Time         0.927 0.138
```

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 5: Shopping

```
wilcox.test(mydata$Time ~ mydata$GenderF,
            paired = FALSE,
            correct = FALSE,
            exact = FALSE,
            alternative = "two.sided")
```

```
##
##  Wilcoxon rank sum test
##
## data:  mydata$Time by mydata$GenderF
## W = 52.5, p-value = 0.0003355
## alternative hypothesis: true location shift is not equal to 0
```

HYPOTHESIS ABOUT THE DIFFERENCE IN TWO DISTRIBUTION LOCATIONS

Example 5: Shopping

```
library(effectsize)
```

```
effectsize(wilcox.test(mydata$Time ~ mydata$GenderF,
                       paired = FALSE,
                       correct = FALSE,
                       exact = FALSE,
                       alternative = "two.sided"))
```

```
## r (rank biserial) |         95% CI
## ---------------------------------
## -0.69             | [-0.84, -0.44]
```

```
interpret_rank_biserial(0.69)
```

```
## [1] "very large"
## (Rules: funder2019)
```

Funder & Ozer (2019) ("`funder2019`"; default)

- **r < 0.05** - Tiny
- **0.05 <= r < 0.1** - Very small
- **0.1 <= r < 0.2** - Small
- **0.2 <= r < 0.3** - Medium
- **0.3 <= r < 0.4** - Large
- **r >= 0.4** - Very large

Conclusions:

Based on the sample data, we find that men and women differ in the amount of time spent at the mall ($p < 0.001$) - women spend more time, the difference in distribution is very large ($r = 0.69$)

REPEATED MEASURES ANALYSIS OF VARIANCE, rANOVA

FRIEDMAN ANALYSIS OF VARIANCE

REPEATED MEASURES ANOVA - rANOVA

Comparison of three or more arithmetic means for dependent samples (extension of paired samples *t*-test), rANOVA.

H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$
H_1: At least one μ_j is different

- The same unit (e.g., person) is measured over time (at least 3 times) and the differences in the averages are compared.
- Using three or more comparable variables (e.g., agreement measured on a Likert scale of 1-5), belonging to the same unit (e.g., person), and checking to see if there are differences in averages.

REPEATED MEASURES ANOVA - rANOVA

Conditions and assumptions:

- Analyzed variables are numeric.
- Normality: The variables in the population are normally distributed (this can be checked with the Shapiro-Wilk test).
- Sphericity: the variances of the differences between the examined variables are equal (tested with the Mauchly test for sphericity).
- No significant outliers.

FRIEDMAN ANOVA

Comparison of three or more locations for dependent samples (extension of Wilcoxon Signed Rank Test).

H_0: All distribution locations of variables are the same

H_1: At least one distribution location of variable is different

Comparison of three or more distribution locations of variables belonging to the same units:

- The same unit is measured over time (at least 3 times) and differences in the location of variable distributions are compared using rank analysis.
- Using three or more comparable variables (e.g., agreement measured on a Likert scale of 1-5) belonging to the same unit and checking by rank analysis whether there are differences in the location of the variable distributions.

REPEATED MEASURES ANOVA – rANOVA
Example: Trust

In his master's thesis, the student investigated the level of trust in different social systems in Slovenia. He conducted an online survey using the 1KA program and received 1,318 responses. He identified the following as possible social systems, on which respondents indicated how much they trust them on a scale between 0 and 10 (the higher the value, the greater the trust):

- Trust in parliament (trstprl)
- Trust in the rule of law (trstlgl)
- Trust in the police (trstplc)
- Trust in politics (trstplt)

Are there differences in trust in these social systems?

REPEATED MEASURES ANOVA – rANOVA
Example: Trust

```
mydata <- read.table("./Trust.csv", header=TRUE, sep=";", dec=",")
head(mydata)
```

```
##   ID trstprl trstlgl trstplc trstplt
## 1  1       3       0       7       0
## 2  2       2       3       3       2
## 3  3       0       4      10       0
## 4  4       0       0       0       0
## 5  5       1       6       7       3
## 6  6       4      88       7       4
```

Description:

- ID: Person ID
- trstprl: Trust in the parliament (0-10)
- trstlgl: Trust in the legal system (0-10)
- trstplc: Trust in the police (0-10)
- trstplt: Trust in the politics (0-10)

77, 88, 99 missing values

```
set.seed(1) #Setting initial point of sampling.
mydata <- mydata[sample(nrow(mydata), 50), ] #Random sample of 50 units.
```

REPEATED MEASURES ANOVA – rANOVA
Example: Trust

```
library(dplyr)
library(naniar)

#Missing values are changed to NA.
mydata <- mydata %>%
             replace_with_na(replace = list(trstprl = c(77, 88, 99),
                                            trstlgl = c(77, 88, 99),
                                            trstplc = c(77, 88, 99),
                                            trstplt = c(77, 88, 99)))

tail(mydata, 10) #Showing last 10 units.
```

```
##        ID trstprl trstlgl trstplc trstplt
## 810   810       4       0       4       0
## 526   526      NA       0       9       0
## 642   642       2       4       6       3
## 1069 1069       3       2       7       6
## 22     22       4       3       7       3
## 193   193       6       8       9       3
## 499   499       5       4       7       5
## 1128 1128       5       3       7       5
## 983   983       3       2       4       1
## 843   843       5       4       6       5
```

REPEATED MEASURES ANOVA – rANOVA
Example: Trust

```
library(tidyr)
```

```
mydata <- drop_na(mydata) #Removing units with missing values.
```

```
tail(mydata, 10)
```

```
##       ID trstprl trstlgl trstplc trstplt
## 40   435       0       0       0       0
## 41   810       4       0       4       0
## 42   642       2       4       6       3
## 43 1069       3       2       7       6
## 44    22       4       3       7       3
## 45   193       6       8       9       3
## 46   499       5       4       7       5
## 47 1128       5       3       7       5
## 48   983       3       2       4       1
## 49   843       5       4       6       5
```

REPEATED MEASURES ANOVA – rANOVA
Example: Trust

```
library(rstatix)
```

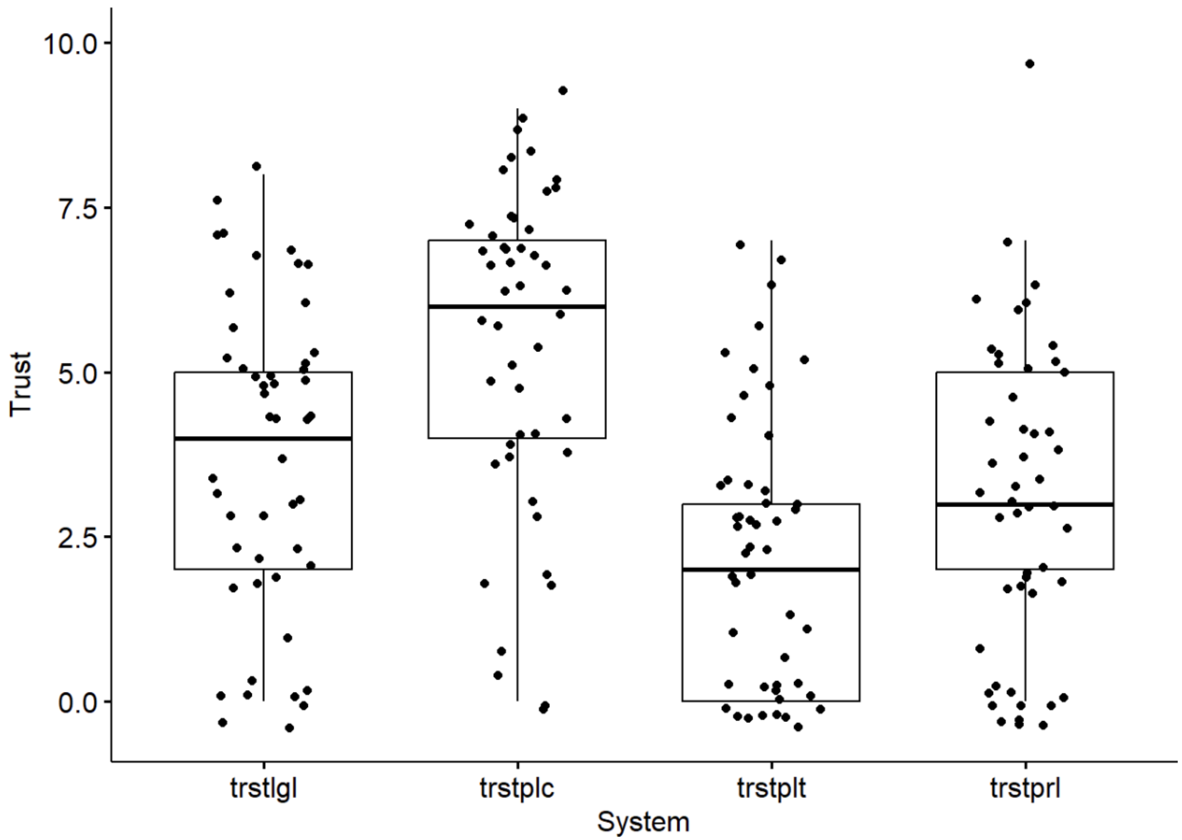
```
library(tidyverse)
```

```
#Wide format of data is changed to long format.
mydata_long <- mydata %>%
                 pivot_longer(
                   cols = c("trstprl", "trstlgl", "trstplc", "trstplt"),
                   names_to = "System",
                   values_to = "Trust") %>%
                 convert_as_factor(System)


mydata_long <- as.data.frame(mydata_long)

tail(mydata_long, 10)
```

```
##       ID  System Trust
## 187 1128 trstplc     7
## 188 1128 trstplt     5
## 189  983 trstprl     3
## 190  983 trstlgl     2
## 191  983 trstplc     4
## 192  983 trstplt     1
## 193  843 trstprl     5
## 194  843 trstlgl     4
## 195  843 trstplc     6
## 196  843 trstplt     5
```

REPEATED MEASURES ANOVA – rANOVA
Example: Trust

```
library(ggpubr)

#Boxplot for each variable.
ggboxplot(mydata_long,
          x = "System",
          y = "Trust",
          add = "jitter")
```



REPEATED MEASURES ANOVA – rANOVA
Example: Trust

```
library(tidyverse)
library(ggpubr)
library(rstatix)

#Finding outliers.
mydata_long %>%
  group_by(System) %>%
  identify_outliers(Trust)
```

```
## # A tibble: 1 × 5
##   System     ID Trust is.outlier is.extreme
##   <fct>   <int> <int> <lgl>      <lgl>
## 1 trstprl   248    10 TRUE       FALSE
```

```
#Removing outliers.
mydata_long <- mydata_long %>%
  filter(!ID == 248)
```

```
mydata_long %>%
  group_by(System) %>%
  identify_outliers(Trust)
```

```
## [1] System     ID         Trust      is.outlier is.extreme
## <0 rows> (or 0-length row.names)
```

REPEATED MEASURES ANOVA – rANOVA
Example: Trust

```
library(rstatix)

#Checking normality with Shapiro-Wilk test
mydata_long %>%
  group_by(System) %>%
  shapiro_test(Trust)
```

```
## # A tibble: 4 × 4
##   System  variable statistic        p
##   <fct>   <chr>        <dbl>    <dbl>
## 1 trstlgl Trust        0.935 0.0104
## 2 trstplc Trust        0.914 0.00187
## 3 trstplt Trust        0.891 0.000317
## 4 trstprl Trust        0.919 0.00269
```

REPEATED MEASURES ANOVA – rANOVA

Example: Trust

```
library(rstatix)

#Descriptive statistics for each variable.
mydata_long %>%
  group_by(System) %>%
  get_summary_stats(Trust, type = "common")
```

```
## # A tibble: 4 × 11
##   System  variable     n   min   max median   iqr  mean    sd    se    ci
##   <fct>   <fct>    <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 trstlgl Trust       48     0     8      4  3     3.71  2.40 0.347 0.698
## 2 trstplc Trust       48     0     9      6  3     5.46  2.43 0.351 0.706
## 3 trstplt Trust       48     0     7      2  3     2.23  1.98 0.286 0.575
## 4 trstprl Trust       48     0     7      3  3.25  2.94  2.07 0.298 0.6
```

REPEATED MEASURES ANOVA – rANOVA
Example: Trust

```
library(rstatix)

#rANOVA
ANOVA_results <- anova_test(dv = Trust, #Dependent variable
                            wid = ID, #Subject identifier
                            within = System, #Within-subject factor variable
                            data = mydata_long)

ANOVA_results #Summary of results.
```

```
## ANOVA Table (type III tests)
##
## $ANOVA
##   Effect DFn DFd      F        p p<.05   ges
## 1 System   3 141 39.534 1.33e-18     * 0.229
##
## $`Mauchly's Test for Sphericity`
##   Effect     W     p p<.05
## 1 System 0.747 0.021     *
##
## $`Sphericity Corrections`
##   Effect   GGe      DF[GG]    p[GG] p[GG]<.05   HFe       DF[HF]    p[HF] p[HF]<.05
## 1 System 0.874 2.62, 123.3 1.45e-16         * 0.931 2.79, 131.25 1.76e-17         *
```

```
get_anova_table(ANOVA_results, correction = "auto")
```

```
## ANOVA Table (type III tests)
##
##   Effect  DFn   DFd      F        p p<.05   ges
## 1 System 2.62 123.3 39.534 1.45e-16     * 0.229
```

REPEATED MEASURES ANOVA – rANOVA
Example: Trust

```r
library(effectsize)
```

```r
interpret_eta_squared(0.229, rules = "cohen1992")
```

```
## [1] "medium"
## (Rules: cohen1992)
```

Cohen (1992) (`"cohen1992"`) applicable to one-way anova, or to *partial* eta / omega / epsilon squared in multi-way anova.

- **ES < 0.02** - Very small
- **0.02 <= ES < 0.13** - Small
- **0.13 <= ES < 0.26** - Medium
- **ES >= 0.26** - Large

REPEATED MEASURES ANOVA – rANOVA

Post-Hoc tests

Post-Hoc tests compare all combinations of two variables.

Because the probability of a Type I error increases when the statistical tests are repeated on the same sample of data, the criterion for rejecting null hypothesis is more stringent. The most commonly used method is the Bonferroni correction, which increases the reported *p*-value according to the number of tests performed.

REPEATED MEASURES ANOVA – rANOVA

Example: Trust

```
library(rstatix)

#Comparing all paires of variables
pwc <- mydata_long %>%
  pairwise_t_test(Trust ~ System,
                  paired = TRUE,
                  p.adjust.method = "bonferroni")

pwc
```

```
## # A tibble: 6 × 10
##    .y.   group1  group2     n1    n2 statistic    df        p    p.adj p.adj.signif
## * <chr> <chr>   <chr>   <int> <int>     <dbl> <dbl>    <dbl>    <dbl> <chr>
## 1 Trust trstlgl trstplc    48    48     -5.09    47 6.27e- 6 3.76e- 5 ****
## 2 Trust trstlgl trstplt    48    48      4.44    47 5.36e- 5 3.22e- 4 ***
## 3 Trust trstlgl trstprl    48    48      2.50    47 1.6 e- 2 9.6 e- 2 ns
## 4 Trust trstplc trstplt    48    48     10.1     47 2   e-13 1.2 e-12 ****
## 5 Trust trstplc trstprl    48    48      7.54    47 1.24e- 9 7.44e- 9 ****
## 6 Trust trstplt trstprl    48    48     -3.24    47 2   e- 3 1.3 e- 2 *
```

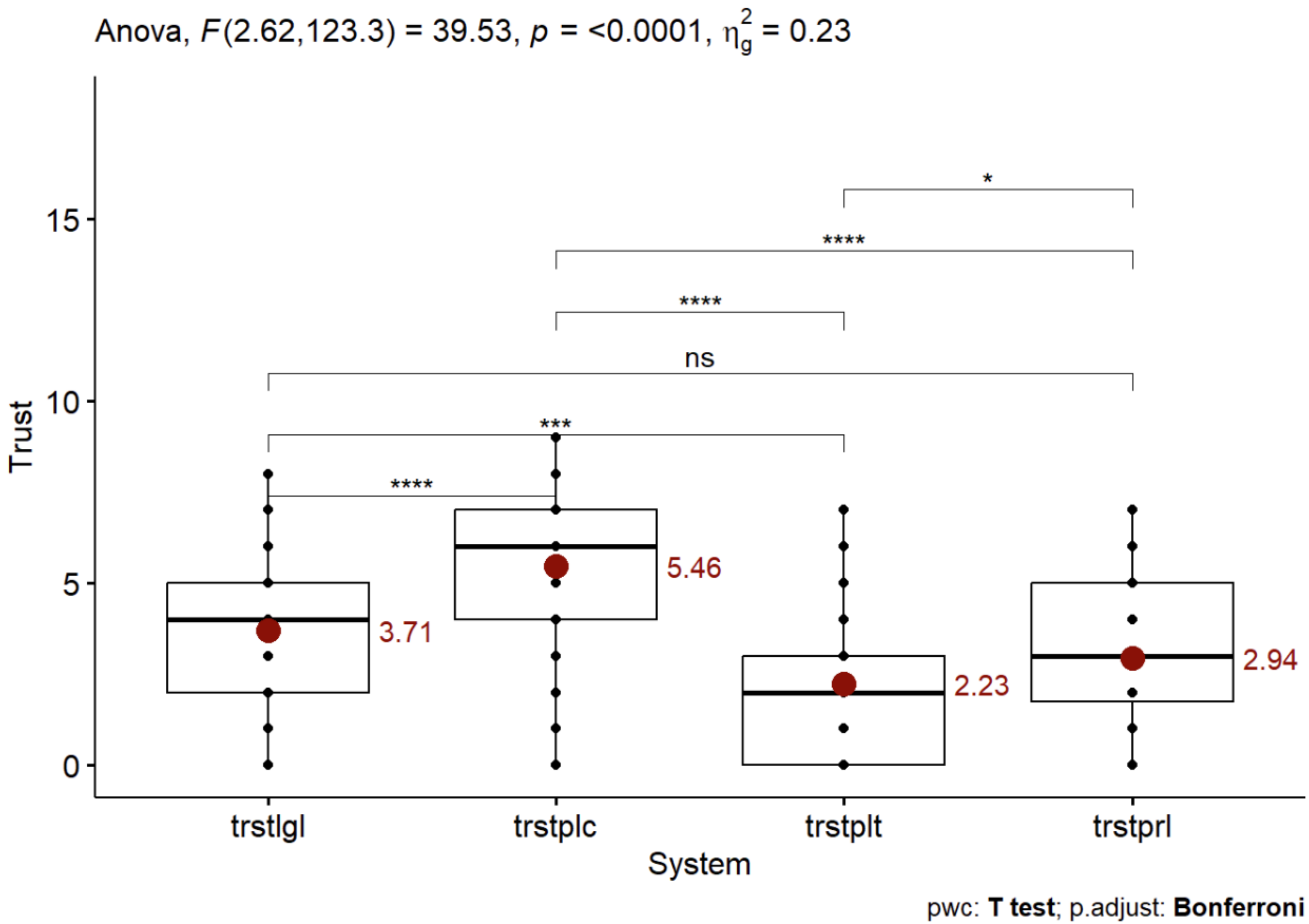
REPEATED MEASURES ANOVA – rANOVA

Example: Trust

```r
library(rstatix)
pwc <- pwc %>%
       add_y_position(fun = "median", step.increase = 0.35)

library(ggpubr)
ggboxplot(mydata_long, x = "System", y = "Trust", add = "point", ylim=c(0, 18)) +
  stat_pvalue_manual(pwc, hide.ns = FALSE) +
  stat_summary(fun = mean, geom = "point", shape = 16, size = 4,
               aes(group = System), color = "darkred",
               position = position_dodge(width = 0.8)) +
  stat_summary(fun = mean, colour = "darkred",
               position = position_dodge(width = 0.8),
               geom = "text", vjust = 0.5, hjust = -2,
               aes(label = round(after_stat(y), digits = 2), group = System)) +
  labs(subtitle = get_test_label(ANOVA_results,  detailed = TRUE),
       caption = get_pwc_label(pwc))
```

REPEATED MEASURES ANOVA – rANOVA

Example: Trust



FRIEDMAN ANOVA
Example: Trust

```
library(rstatix)

#Friedman ANOVA.
FriedmanANOVA <- friedman_test(Trust ~ System | ID,
                               data = mydata_long)

FriedmanANOVA #Summary of results.
```

```
## # A tibble: 1 × 6
##   .y.       n statistic    df        p method
## * <chr> <int>     <dbl> <dbl>    <dbl> <chr>
## 1 Trust    48      68.2     3 1.02e-14 Friedman test
```

```
library(effectsize)
effectsize::kendalls_w(Trust ~ System | ID,
                       data = mydata_long)
```

```
## Warning: 36 block(s) contain ties, some containing only 1 unique ranking.
```

```
## Kendall's W |       95% CI
## --------------------------
## 0.47        | [0.37, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
interpret_kendalls_w(0.47)
```

```
## [1] "moderate agreement"
## (Rules: landis1977)
```

Landis & Koch (1977) ("`landis1977`"; default)

- **0.00 <= w < 0.20** - Slight agreement
- **0.20 <= w < 0.40** - Fair agreement
- **0.40 <= w < 0.60** - Moderate agreement
- **0.60 <= w < 0.80** - Substantial agreement
- **w >= 0.80** - Almost perfect agreement

FRIEDMAN ANOVA
Example: Trust

```r
library(rstatix)

#Wilcoxon signed rank tests - comparing all possible paires.
paires_nonpar <- wilcox_test(Trust ~ System,
                             paired = TRUE,
                             p.adjust.method = "bonferroni",
                             data = mydata_long)


paires_nonpar
```

```
## # A tibble: 6 × 9
##   .y.   group1  group2     n1    n2 statistic            p        p.adj p.adj.signif
## * <chr> <chr>   <chr>   <int> <int>     <dbl>        <dbl>        <dbl> <chr>
## 1 Trust trstlgl trstplc    48    48     109    0.0000279    0.000167     ***
## 2 Trust trstlgl trstplt    48    48     548.   0.00013      0.00078      ***
## 3 Trust trstlgl trstprl    48    48     464.   0.014        0.082        ns
## 4 Trust trstplc trstplt    48    48     942.   0.0000000133 0.0000000798 ****
## 5 Trust trstplc trstprl    48    48     908.   0.000000138  0.000000828  ****
## 6 Trust trstplt trstprl    48    48      56.5  0.004        0.023        *
```

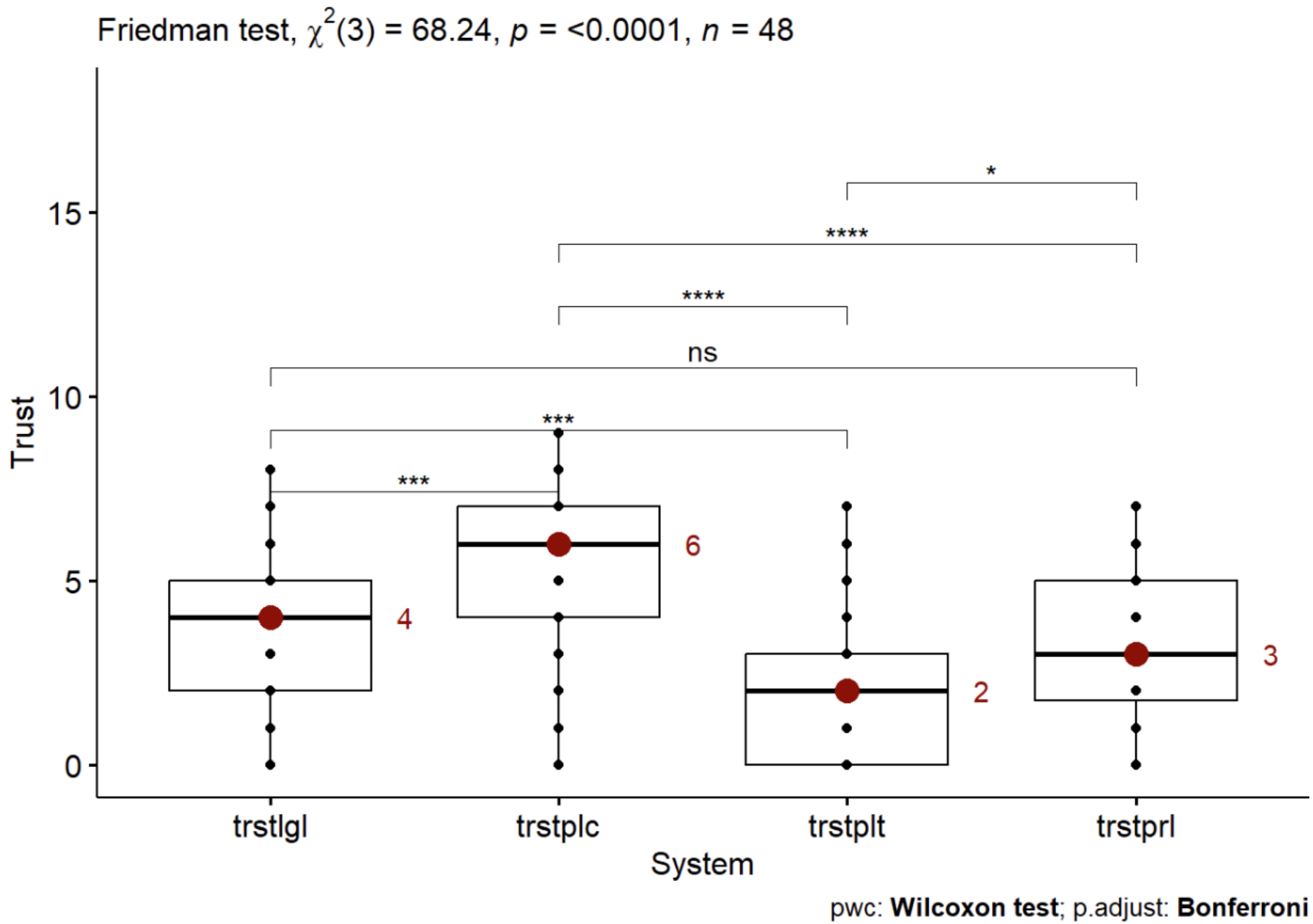
University *of Ljubljana*

SEB SCHOOL OF ECONOMICS AND BUSINESS

FRIEDMAN ANOVA
Example: Trust

```
library(rstatix)
comparisons <- paires_nonpar %>%
                 add_y_position(fun = "median", step.increase = 0.35)

library(ggpubr)
ggboxplot(mydata_long, x = "System", y = "Trust", add = "point", ylim=c(0, 18)) +
  stat_pvalue_manual(comparisons, hide.ns = FALSE) +
  stat_summary(fun = median, geom = "point", shape = 16, size = 4,
               aes(group = System), color = "darkred",
               position = position_dodge(width = 0.8)) +
  stat_summary(fun = median, colour = "darkred",
               position = position_dodge(width = 0.8),
               geom = "text", vjust = 0.5, hjust = -8,
               aes(label = round(after_stat(y), digits = 2), group = System)) +
  labs(subtitle = get_test_label(FriedmanANOVA,  detailed = TRUE),
       caption = get_pwc_label(comparisons))
```

FRIEDMAN ANOVA
Example: Trust



ONE-WAY ANALYSIS OF VARIANCE (ANOVA)

KRUSKAL-WALLIS RANK SUM TEST

ONE-WAY ANALYSIS OF VARIANCE - ANOVA

Comparison of three or more arithmetic means for independent samples (extension of independent samples *t*-test).

H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$
H_1: At least one μ_j is different from others

Different groups of units are measured (at the same time) and the differences in the group averages are compared.

ONE-WAY ANALYSIS OF VARIANCE - ANOVA

Conditions and assumptions:

- Analyzed variable is numeric.
- Variable in the population is normally distributed within each group.
- Homoscedasticity: the variance of analyzed variable is the same within all groups.
- No significant outliers.

KRUSKAL-WALLIS RANK SUM TEST

Comparison of three or more distribution locations of variables for independent samples (extension of the Wilcoxon Rank Sum Test).

H_0: All distribution locations of variables are the same

H_1: At least one distribution location of variable is different

Comparison of three or more distribution locations of the same variable, belonging to different groups of units.

ONE-WAY ANALYSIS OF VARIANCE – ANOVA
Example: Puppies

Puppy therapy.

90 people is randomly assigned to one of three groups:
- Control group – no puppy (28 people)
- 15 minutes of puppy therapy (32 people)
- 30 minutes of puppy therapy (30 people)

The analyzed variable is a measure of happiness between 1 (as unhappy as I can imagine) and 10 (as happy as I can imagine). Are puppy therapy and happiness related?

ONE-WAY ANALYSIS OF VARIANCE – ANOVA
Example: Puppies

```
mydata <- read.table("./Puppies.csv", header=TRUE, sep=";", dec=".")
head(mydata)
```

```
##   ID Dose Happiness
## 1  1    1         3
## 2  2    1         3
## 3  3    1         4
## 4  4    1         1
## 5  5    1         3
## 6  6    1         6
```

Description:

- Person: ID
- Dose: 1: Control group, 2: 15 minutes exposed to Puppy, 3: 30 minutes exposed to Puppy
- Happiness: How happy is a person between 1 and 10.

```
mydata$DoseFactor <- factor(mydata$Dose,
                            levels = c(1, 2, 3),
                            labels = c("Control", "15 minutes", "30 minutes"))
```

```
library(psych)
```

```
describe(mydata$Happiness)
```

```
##    vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 90 5.33 2.12      5    5.33 2.97   1  10     9 0.05    -0.87 0.22
```

ONE-WAY ANALYSIS OF VARIANCE – ANOVA
Example: Puppies

```
describeBy(x = mydata$Happiness, group = mydata$DoseFactor)
```

```
##
##  Descriptive statistics by group
## group: Control
##    vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 28  3.5 1.32    3.5     3.5 0.74   1   6     5    0    -0.67 0.25
## ------------------------------------------------------------------------------------------
## group: 15 minutes
##    vars  n mean   sd median trimmed  mad min max range skew kurtosis  se
## X1    1 32 5.06 1.72      5       5 1.48   2   9     7 0.24    -0.54 0.3
## ------------------------------------------------------------------------------------------
## group: 30 minutes
##    vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 30 7.33 1.24      7    7.33 1.48   5  10     5    0    -0.73 0.23
```

ONE-WAY ANALYSIS OF VARIANCE – ANOVA

Example: Puppies

```
library(car)
leveneTest(mydata$Happiness, group = mydata$DoseFactor)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  1.0404 0.3577
##       87
```

```
library(dplyr)
```

```
library(rstatix)
```

```
mydata %>%
  group_by(DoseFactor) %>%
  shapiro_test(Happiness)
```

```
## # A tibble: 3 × 4
##   DoseFactor variable  statistic     p
##   <fct>      <chr>         <dbl> <dbl>
## 1 Control    Happiness     0.946 0.161
## 2 15 minutes Happiness     0.964 0.356
## 3 30 minutes Happiness     0.943 0.108
```

ONE-WAY ANALYSIS OF VARIANCE – ANOVA

Example: Puppies

```
ANOVA_results <- aov(Happiness ~ DoseFactor,
                     data = mydata)

summary(ANOVA_results)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## DoseFactor   2  216.5  108.23    51.3 1.92e-15 ***
## Residuals   87  183.5    2.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(effectsize)
```

```
eta_squared(ANOVA_results)
```

```
## # Effect Size for ANOVA
##
## Parameter  | Eta2 |       95% CI
## -------------------------------
## DoseFactor | 0.54 | [0.42, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

$$\eta^2 = \frac{k_k}{k_y} = \frac{216.5}{216.5 + 183.5} = 0.54$$

```
interpret_eta_squared(0.54, rules = "cohen1992")
```

```
## [1] "large"
## (Rules: cohen1992)
```

Cohen (1992) (`"cohen1992"`) applicable to one-way anova, or to *partial* eta / omega / epsilon squared in multi-way anova.

- **ES < 0.02** - Very small
- **0.02 <= ES < 0.13** - Small
- **0.13 <= ES < 0.26** - Medium
- **ES >= 0.26** - Large

ONE-WAY ANALYSIS OF VARIANCE – ANOVA
Example: Puppies

If the assumption of homogeneity of variances is violated, we use Welch's F-test, which is less affected by the conditions of heteroskedasticity.

```
library(onewaytests)
welch.test(Happiness ~ DoseFactor,
           data = mydata)
```

```
##
##   Welch's Heteroscedastic F Test (alpha = 0.05)
## -------------------------------------------------------------
##   data : Happiness and DoseFactor
##
##   statistic  : 65.16615
##   num df     : 2
##   denom df   : 57.50688
##   p.value    : 1.655095e-15
##
##   Result     : Difference is statistically significant.
## -------------------------------------------------------------
```

ONE-WAY ANALYSIS OF VARIANCE – ANOVA
Example: Puppies

```
pairwise.t.test(x = mydata$Happiness, g = mydata$DoseFactor,
                p.adj = "none")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  mydata$Happiness and mydata$DoseFactor
##
##            Control 15 minutes
## 15 minutes 7.5e-05 -
## 30 minutes 3.3e-16 2.3e-08
##
## P value adjustment method: none
```

```
pairwise.t.test(x = mydata$Happiness, g = mydata$DoseFactor,
                p.adj = "bonf")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  mydata$Happiness and mydata$DoseFactor
##
##            Control 15 minutes
## 15 minutes 0.00023 -
## 30 minutes 9.9e-16 6.8e-08
##
## P value adjustment method: bonferroni
```

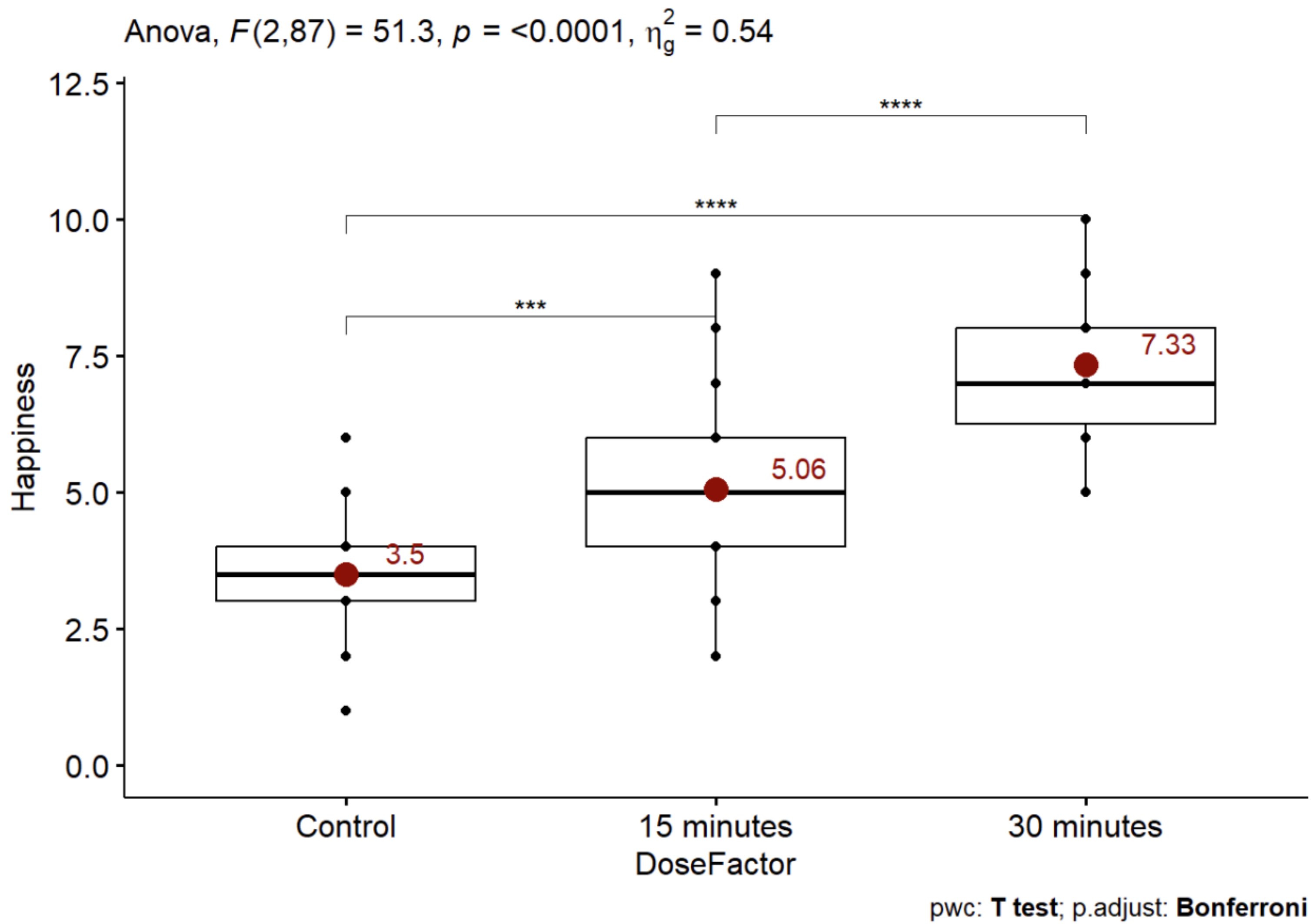
ONE-WAY ANALYSIS OF VARIANCE – ANOVA

Example: Puppies

```
pwc <- mydata %>%
  pairwise_t_test(Happiness ~ DoseFactor,
                  paired = FALSE,
                  p.adjust.method = "bonferroni")

ANOVA_results <- anova_test(Happiness ~ DoseFactor,
                            data = mydata)

library(rstatix)
pwc <- pwc %>%
       add_y_position(fun = "median", step.increase = 0.35)

library(ggpubr)
ggboxplot(mydata, x = "DoseFactor", y = "Happiness", add = "point", ylim=c(0, 12)) +
  stat_pvalue_manual(pwc, hide.ns = FALSE) +
  stat_summary(fun = mean, geom = "point", shape = 16, size = 4,
               aes(group = DoseFactor), color = "darkred",
               position = position_dodge(width = 0.8)) +
  stat_summary(fun = mean, colour = "darkred",
               position = position_dodge(width = 0.8),
               geom = "text", vjust = -0.5, hjust = -1,
               aes(label = round(after_stat(y), digits = 2), group = DoseFactor)) +
  labs(subtitle = get_test_label(ANOVA_results,  detailed = TRUE),
       caption = get_pwc_label(pwc))
```

ONE-WAY ANALYSIS OF VARIANCE – ANOVA

Example: Puppies



ONE-WAY ANALYSIS OF VARIANCE – ANOVA
Example: Puppies

Conclusions:

We found a statistically significant relationship between puppy therapy and average happiness level ($F = 51.3$, $p < 0.001$), the effect size is large ($\eta^2 = 0.54$). Post hoc tests revealed differences for each pair of groups ($p < 0.001$).

KRUSKAL-WALLIS RANK SUM TEST

Example: Puppies

```
kruskal.test(Happiness ~ DoseFactor,
             data = mydata)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Happiness by DoseFactor
## Kruskal-Wallis chi-squared = 49.612, df = 2, p-value = 1.686e-11
```

```
kruskal_effsize(Happiness ~ DoseFactor,
                data = mydata)
```

```
## # A tibble: 1 × 5
##   .y.           n effsize method  magnitude
## * <chr>     <int>   <dbl> <chr>   <ord>
## 1 Happiness    90   0.547 eta2[H] large
```

KRUSKAL-WALLIS RANK SUM TEST

Example: Puppies

```
library(rstatix)

groups_nonpar <- wilcox_test(Happiness ~ DoseFactor,
                             paired = FALSE,
                             p.adjust.method = "bonferroni",
                             data = mydata)

groups_nonpar
```

```
## # A tibble: 3 × 9
##   .y.       group1     group2        n1    n2 statistic        p    p.adj p.adj.signif
## * <chr>     <chr>      <chr>      <int> <int>     <dbl>    <dbl>    <dbl> <chr>
## 1 Happiness Control    15 minutes    28    32       218 5.48e- 4 2    e- 3 **
## 2 Happiness Control    30 minutes    28    30        14 2.02e-10 6.06e-10 ****
## 3 Happiness 15 minutes 30 minutes    32    30       142 1.49e- 6 4.47e- 6 ****
```

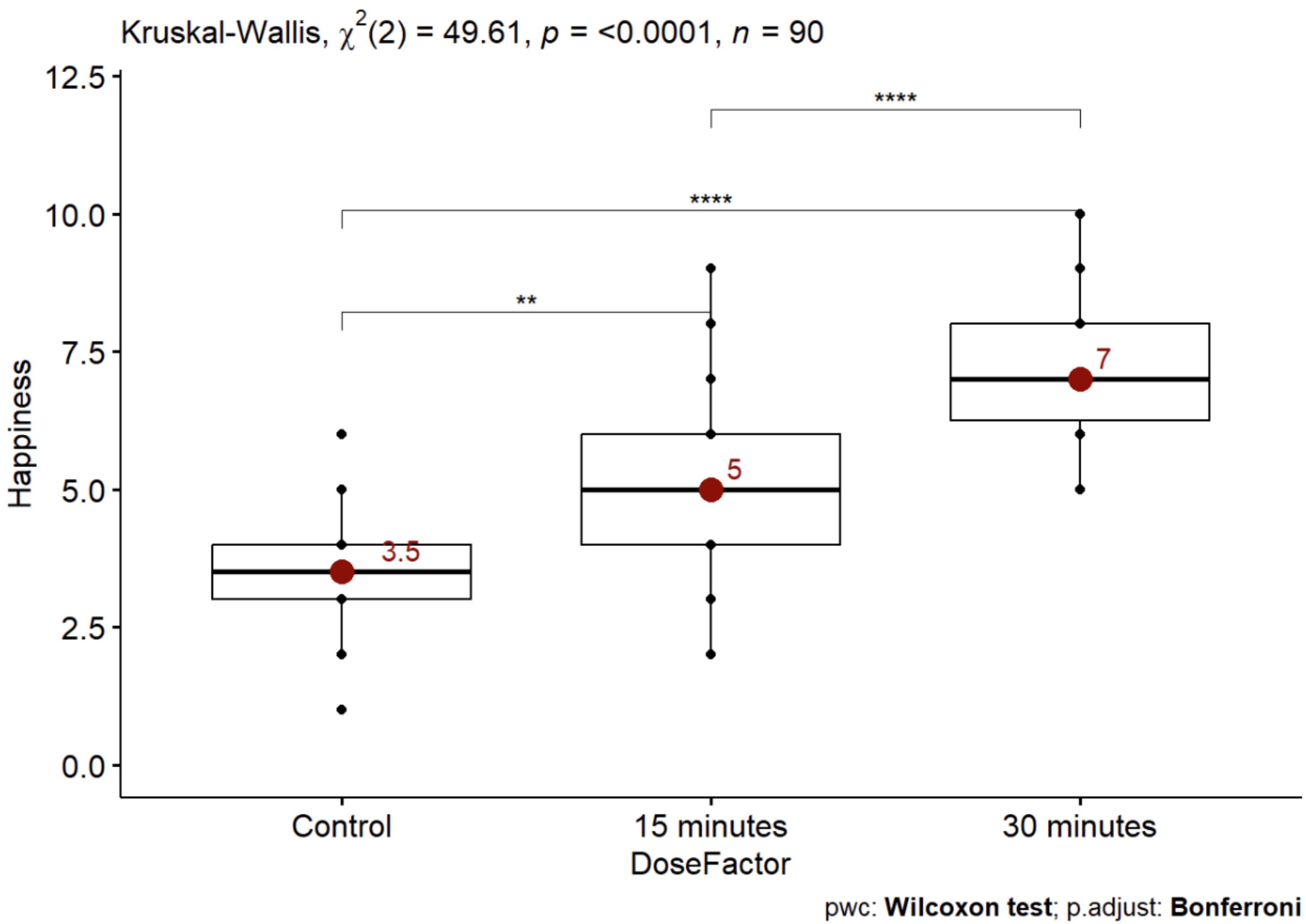
ONE-WAY ANALYSIS OF VARIANCE – ANOVA

Example: Puppies

```
pwc <- mydata %>%
  wilcox_test(Happiness ~ DoseFactor,
              paired = FALSE,
              p.adjust.method = "bonferroni")

Kruskal_results <- kruskal_test(Happiness ~ DoseFactor,
                                data = mydata)

library(rstatix)
pwc <- pwc %>%
       add_y_position(fun = "median", step.increase = 0.35)

library(ggpubr)
ggboxplot(mydata, x = "DoseFactor", y = "Happiness", add = "point", ylim=c(0, 12)) +
  stat_pvalue_manual(pwc, hide.ns = FALSE) +
  stat_summary(fun = median, geom = "point", shape = 16, size = 4,
               aes(group = DoseFactor), color = "darkred",
               position = position_dodge(width = 0.8)) +
  stat_summary(fun = median, colour = "darkred",
               position = position_dodge(width = 0.8),
               geom = "text", vjust = -0.5, hjust = -1,
               aes(label = round(after_stat(y), digits = 2), group = DoseFactor)) +
  labs(subtitle = get_test_label(Kruskal_results,  detailed = TRUE),
       caption = get_pwc_label(pwc))
```

ONE-WAY ANALYSIS OF VARIANCE – ANOVA

Example: Puppies



ONE-WAY ANALYSIS OF VARIANCE – ANOVA
Example: Puppies

Conclusions:

We found that the distribution of happiness levels differs for at least one of the puppy doses ($\chi^2 = 49.6,\ p < 0.001$), the effect size was high ($\eta^2 = 0.55$). Post-Hoc tests revealed differences for each pair of groups ($p < 0.01$).

TEST OF POPULATION PROPORTION

BINOMIAL TEST

TEST OF EQUALITY IN TWO POPULATION PROPORTIONS

TEST OF POPULATION PROPORTION

Testing a population proportion with the chosen parameter.

$H_0: \pi = \pi_0$

$H_1: \pi \neq \pi_0$

$$\chi^2 = \left(\frac{p - \pi_0}{SE(p)}\right)^2 \qquad p = \frac{n_a}{n} \qquad SE(p) = \sqrt{\frac{\pi_0(1 - \pi_0)}{n}} \qquad m = 1$$

Assumption: $n\pi_0 > 5$ and $n(1 - \pi_0) > 5$

TEST OF POPULATION PROPORTION
Example 1

The sample of 1,870 children born at the maternity hospital over the last year included 965 boys and 905 girls. Can you conclude that there are more boys than girls in the newborn population?

```
prop.test(x = 965,
          n = 1870,
          p = 0.5,
          correct = FALSE,
          alternative = "greater")
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  965 out of 1870, null probability 0.5
## X-squared = 1.9251, df = 1, p-value = 0.08265
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.4970246 1.0000000
## sample estimates:
##         p
## 0.5160428
```

BINOMIAL TEST
Example 2

The inspector visited 33 businesses in a particular food industry and found serious violations in 9 of them. Can she conclude from this that the proportion of undertakings in the industry in question is more than 0.10, which is the proportion of infringements at the last inspection five years ago?

```
binom.test(x = 9,
           n = 33,
           p = 0.10,
           alternative = "greater")
```

```
## 
##  Exact binomial test
## 
## data:  9 and 33
## number of successes = 9, number of trials = 33, p-value = 0.004134
## alternative hypothesis: true probability of success is greater than 0.1
## 95 percent confidence interval:
##  0.1502883 1.0000000
## sample estimates:
## probability of success 
##              0.2727273
```

TEST OF EQUALITY IN TWO POPULATION PROPORTIONS

Testing the equality in two population proportions.

H_0: $\pi_1 = \pi_2$

H_1: $\pi_1 \neq \pi_2$

University of Ljubljana
SEB School of Economics and Business

TEST OF EQUALITY IN TWO POPULATION PROPORTIONS
Example 3

Researcher surveyed ($n = 323$) the general population about their use of online banking. Of 234 people under 40, 195 use online banking, while of the 89 people over 40, 45 use online banking. Is there a difference in the proportion of online banking users between the younger and older populations?

```r
prop.test(c(195, 45), c(234, 89),
          correct = TRUE,
          alternative = "two.sided")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(195, 45) out of c(234, 89)
## X-squared = 34.571, df = 1, p-value = 4.11e-09
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.2056395 0.4497912
## sample estimates:
##    prop 1    prop 2
## 0.8333333 0.5056180
```