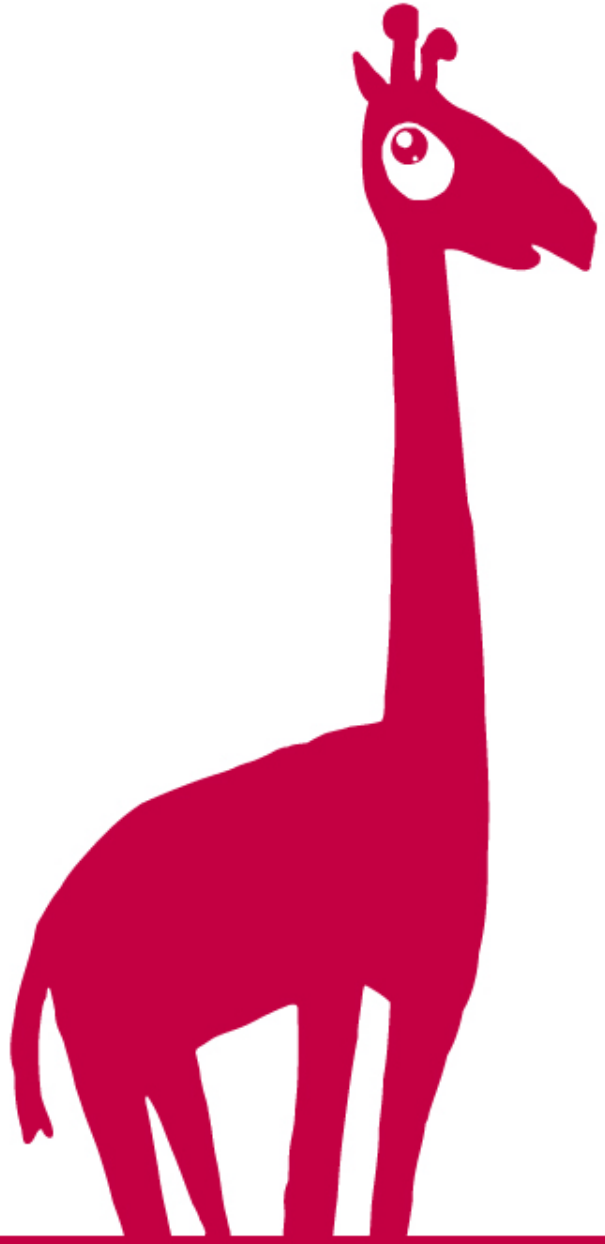
J.J. Strossmayer University of Osijek
Faculty of Economics Osijek
ICES
International Centre for Entrepreneurial Studies

Methodology of Entrepreneurial Research – Selected statistical tests

Nataša Šarlija

Osijek, January, 2021.



Be very careful about:

- Ethics of the research
- Assumptions and correct usage of each and every statistical procedure
- Data sets

Example 1

- Import dataset ‘5_GEM 2010 APS Croatia.sta’ in Statistica and calculate:
- Frequency table of variable number 155: gender. Interpret it. Does it seem right?
- Descriptive statistics of variable number 156: age. Interpret it.
- Note: description of the variables you can find in 4_variables description_GEM_2010.xlsx (sheet: APS Croatia):

Example 2

- Import dataset ‘5_GEM 2010 APS Croatia.sta’ in Statistica and calculate:
- Frequency table of variable number 188: frfail10. Interpret it.
- Descriptive statistics of variable number 193: tea10. Interpret it.
- Note: description of the variables you can find in 4_variables description_GEM_2010.xlsx (sheet: APS Croatia):
- Note: do not forget to turn-on weight_l and wghtd momnts

Example 3.

- Import dataset ‘6_GEM 2010 APS Master.sta’ in Statistica and calculate:
- Descriptive statistics for TEA index, TEA male, TEA female for all countries in the data set (var.number 37, 38, 39)

- Note: description of the variables you can find in 4_variables description_GEM_2010.xlsx (sheet: master APS)
- Note: no need to use weighting

Two types of statistical tests

Parametric tests

- Compare parameters of distributions

↓

μ, p ...

- e.g. Is there a difference in mean values of age between ent.active women and men

Nonparametric tests

- Compare the shape of distributions

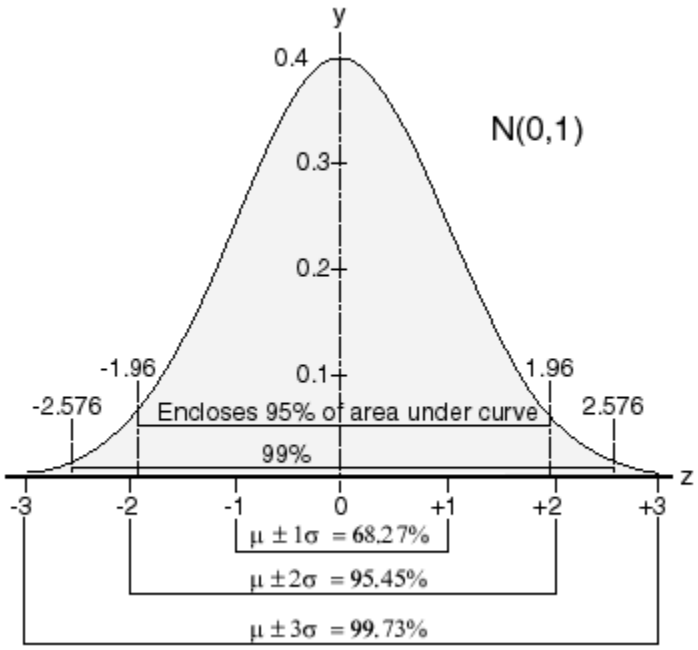
↓

- e.g. Is there a difference in distributions of age between ent.active women and men

Selected tests

- Among parametic tests, we will use those that are based on normal distribution
  - Those tests are used if we know that the variable is normally distributed
  - We don’t know if the variable is normally distributed but we have a sample larger than 30

Normal distribution



- Continuous probability distribution
- Many phenomenon are following normal distribution
- Central limit theorem
- 99,73% of all data lies between $\mu \pm 3\sigma$

Rejecting H_0

- In these tests conclusion about rejecting H_0 is based on a comparison between α and p-value
- α
  - set in advance
  - max. probability that we will make a mistake in rejecting H_0 when it is true
- p-value
  - test statistic, calculated from the data
- if $p < \alpha$, reject H_0

One variable (1/2)

- We have one continuous variable
- We want to test whether mean (μ) of a variable is equal to some specific value

t-test [Sheskin, pp.121,135]

- Assumptions:
  - variable is normally distributed
  - variance is known (z-test)
  - sample can be small
  - we don't know if variable is n.d.
  - we have large sample (n>30)

One variable (2/2)

- Hypotheses for t-test (one sample)
- $H_0 : \mu = \mu_0$
  $H_1 : \mu \neq \mu_0$
- Only for large samples:
- $H_0 : p = p_0$
  $H_1 : p \neq p_0$
- If $p < \alpha \longrightarrow$ reject H_0

Example 4

- We have read somewhere that the average age of entrepreneurially active person is 43 years. We want to check if this is true for Croatia.

  (5_GEM 2010 APS Croatia.sta)

Example 5

- In the dataset ‘6_GEM 2010 APS Master’ in the variable OPPORT10 (13) there are percentages of people for 60 countries who think that there will be a good business opportunities in the next 6 months.
- Make desctiptive statistics of the variable.
- Check if mean value is statistically different than 50%. What does it mean?

Example 6

- I believe that in Croatia there is at least 50% of the population who agree with the statement that you will often see stories in the public media about successful new businesses.
- Am I right? Check the data and let me know.

  (5_GEM 2010 APS Croatia.sta)

Note: NBMEDI10 – variable number 192

Two variables (1/5)

We have 2 continuous variables

- Parametric test

T-test

- difference in μ
- Variables are n.d.

  Samples are small

  Variances are known (z-test)
- We don't know if variable is n.d.

  We have n>30

[Sheskin, pp.375]

- Non-parametric test

Mann-Whitney

- difference in distributions
- Variables are not normally distributed or we don't know distributions

  We have n<30

[Sheskin, pp.423]

Two variables (2/5)

We have 2 continuous variables

- Parametric test

T-test

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

- Only for large samples:

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 \neq p_2$$

- If $p < \alpha$, reject H_0

- Non-parametric test

Mann-Whitney

H_0: distributions are equal

H_1: distributions are not equal

[Sheskin, pp.423]

- If $p < \alpha$, reject H_0

Example 7

- We want to test whether there is a difference in percentage of fear of failure (188) between TEA active and TEA non-active (193).

  (5_GEM 2010 APS Croatia.sta)

Example 8

- We would like to see whether there is a difference in total money required to start a business (SUMONTUS: 275) between women and men (gender: 155) who are entrepreneurial active (TEA10: 193).

  Do women start less expensive businesses?

  We know nothing about distribution of total

  money required.

  (5_GEM 2010 APS Croatia.sta)

Example 9

- We would like to see whether there is a difference in entrepreneurial activity (TEA10: 37) between efficiency driven and innovation driven countries (CAT_GCR2: 4).

  Is it the same for opportunity (TEA10OPP: 42) and necessity (TEA10NEC: 43)?

  What do you expect?

  (6_GEM 2010 APS Master.sta)

Two variables (3/5)

We have 2 numerical variables

We want to test correlation

Parametric test

Pearson corr.coef.

- at least one variable is n.d.
- linear relationship

[Sheskin, pp.945]

Non-parametric test

Spearman corr.coef.

- variables not n.d.
- at least one is ordinal
- not linear relationship
- small sample

[Sheskin, pp.1061]

Two variables (4/5)

We have 2 numerical variables

We want to test correlation

Parametric test

Pearson corr.coef.

$$H_0 : r = 0$$
$$H_1 : r \neq 0$$

- If $p < \alpha$, reject H_0

Non-parametric test

Spearman corr.coef.

H_0: no correlation

H_1: correlation exists

- If $p < \alpha$, reject H_0

Example 10

- Is there a correlation between FRFAIL (15) and SUSKILL (14)?

  Whether higher percentage of those who have knowledge means lower percentage of fear of failure.

  Test it for GEM countries.

  (6_GEM 2010 APS Master.sta)

Example 11

- Is there a relationship between Knoent10 (12) and TEA10 (37).
- Test it separatey for factor driven countries and efficiency driven countries (CAT_GCR2).

  (6_GEM 2010 APS Master.sta)

Two variables (5/5)

- We have two categorical variables
- We want to test for independence

χ^2 test [Sheskin, pp.493]

- Assumptions:
  - no more than 20% of expected values may be less than 5

H_0: there is no dependence between two variables

H_1: there is dependence between two variables

  - If $p < \alpha$, reject H_0

Example 12

- We would like to examine whether knowledge and skills to start a business (SUSKIL10: 187) depend on gender (gender: 155), in other words, is there a difference in perception of having knowledge and skills for business between women and men.

(5_GEM 2010 APS Croatia.sta)

Example 13

- Examine whether there is a dependence between TEA10 (193) and OPPORT10 (186).

  (5_GEM 2010 APS Croatia.sta)

Example 14

- Examine whether oppinion about equal standard of living depends on gender.
- GENDER (155)
- EQUALI10 (189): most people would prefer that everyone had a similar standard of living
- (5_GEM 2010 APS Croatia.sta)

Three and more variables (1/2)

We have 3 or more continuous variables

We can use:

Parametric test

ANOVA

- difference in μ
- Variables are n.d.

  Samples are small

  Variances are known
- We don't know if variables are n.d.

  We have n>30

[Sheskin, pp.667]

Non-parametric test

Kruskal-Wallis ANOVA

- difference in distributions
- Variables are not normally distributed

  We have n<30

[Sheskin, pp.757]

Three and more variables (2/2)

We have 3 or more continuous variables

We can use:

Parametric test

ANOVA

$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots$

$H_1 :$ at least one μ is different

- If $p < \alpha$, reject H_0

Non-parametric test

Kruskal-Wallis ANOVA

H_0: distributions are equal

H_1: distributions are not equal

- If $p < \alpha$, reject H_0

Example 15

We would like to examine whether there is a difference in TEA (193) between regions in Croatia (hrregion: 163).

(5_GEM 2010 APS Croatia.sta)

Example 16

- Explore the relationship between Kontinent - Continent (1) and TEA10 (37) for all GEM countires.

(6_GEM 2010 APS Master.sta)

Nataša Šarlija

University J.J. Strossmayer in Osijek
Gajev trg 7
HR-31000 Osijek
Phone: +385 31 224 452
Fax: +385 31 224 438
E-mail: natasa@efos.hr

www.icesefos.hr