Visegrad Fund

Visegrad Grant *No. 21730020*
http://vinmes.eu/

Visegrad Network

V4 Seminars for Young Scientists on Publishing Techniques in the Field of Engineering Science

Statistical analysis of the results in brief

Agata Skwarek, PhD
Institute of Electron Technology, Poland

Visegrad Fund

Statistics

Visegrad Network

- Statistics - it is the method of getting the knowledge about tested parameters of the whole population (people, animals, plants, group of materials) on the basis of properly selected (REPRESENTATIVE!!!) sample.
- In practice: with statistics we are searching the answer to the question:" Do the samples differ from each other?"
- Finally… there is no 100% sure answer. Statistics analysis forcasts the results with some level of uncertainty (usually 5%).

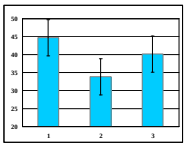


Fig. 1. Measured values of different samples

Visegrad Fund

Statistics

Visegrad Network

There are three levels of lies:

- forecasting
- diplomatic message
- statistics

Visegrad Fund

Hipothesis testing

The general idea of hypothesis testing involves:

- Making an initial assumption (H_0 – null hypothesis, H_1 – alternative hypothesis, determining the significance level (0.05)).
- (Significance level – the probability of H_0 rejection)
- Collecting evidence (data, variables).
- Based on the available evidence (data), deciding whether to reject or not reject (H_0) the initial assumption.

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

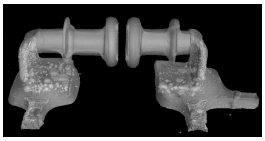
Hipothesis testing - examples

Proper formulating of the problem/research hypothesis.

Examples:

- H_0 – there is no difference in the voids content between the joints fabricated with convection reflow or VPS technology.
- H_1- there is the difference in the voids content between the joints fabricated with convection reflow or VPS technology.

But: The void total volume in the joint is important but also the size of the single void.



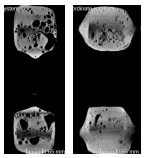


Fig. 2. Void fromation in the solder joints

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Hipothesis testing - examples

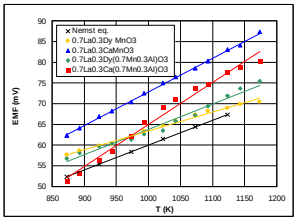


Fig. 3. EMF values versus temperature

- H_0 – the plots of the electromotive force changes for the SOFC samples with different electrode materials are the same as theoretical curve obtained from Nernst equation.
- H_1 – the plots of the electromotive force changes for the SOFC changes differ form theoretical curve obtained from Nernst equation.

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Hipothesis testing

Visegrad Network

- We always assume the null hypothesis is true and that is no difference between the samples.

But:

- If we reject the null hypothesis, we do not prove that the alternative hypothesis is true.
- If we do not reject the null hypothesis, we do not prove that the null hypothesis is true.

„*Presumption of innocence*" - one is considered innocent unless proven guilty.

- Two types of errors :"Type I error" and "Type II error
  - Type I error: The null hypothesis is rejected when it is true.
  - Type II error: The null hypothesis is not rejected when it is false.

A. Skwarek – Statystical analysis of the results in brief

And so one and so one……..

Visegrad Fund

Hipothesis testing

Visegrad Network

In statistics, there are two ways to determine whether the evidence is likely or unlikely given the initial assumption:

- "critical value approach" (favored in many of the older textbooks) → Excel, manually calculated statistics.
- "P-value approach" (what is used most often in research, journal articles, and statistical software) →advanced statistics sofwares.

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Variables (data)

Visegrad Network

- A **variable** is an object, event, idea, feeling, time period, or any other type of category you are trying to measure. There are two types of variables-independent and dependent.
- **Independent variables** are variables that are manipulated or are changed by researchers and whose effects are measured and compared. The other name for independent variables is Predictor(s).
- The **independent variables** are called as such because independent variables predict or forecast the values of the dependent variable in the model.
- The other variable(s) are also considered the **dependent variable**(s). The dependent variables refer to that type of variable that measures the affect of the independent variable(s) on the test units.

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Variables scales

- **Nominal Scales** - are used for labeling variables, without any quantitative value (gender, colours, place of living)
- **Ordinal Scale** - the order of the values is what's important and significant, but the differences between each one is not really known. Ordinal scales are typically measures of non-numeric concepts like satisfaction, happiness, discomfort, etc.
- **Interval Scales** - are numeric scales in which we know not only the order, but also the exact differences between the values (Celsius temperature - the difference between each value is the same.)
- **Ratio Scales** - tell about the order - exact value between units, AND they also have an absolute zero–which allows for a wide range of both descriptive and inferential statistics to be applied.

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Scales - comparision

Provides:	Nominal	Ordinal	Interval	Ratio
The "order" of values is known		✔	✔	✔
"Counts," aka "Frequency of Distribution"	✔	✔	✔	✔
Mode	✔	✔	✔	✔
Median		✔	✔	✔
Mean			✔	✔
Can quantify the difference between each value			✔	✔
Can add or subtract values			✔	✔
Can multiple and divide values				✔
Has "true zero"				✔

http://www.mymarketresearchmethods.com/types-of-data-nominal-ordinal-interval-ratio/

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Normal distribution (Gaussian)

The Normal Distribution has:

- mean = median = mode (central tendency)
- symmetry about the center → Parametrical tests
- 50% of values less than the mean and 50% greater than the mean

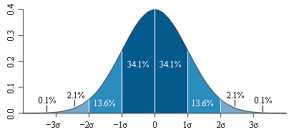


Fig. 3. Normal distribution
https://pl.wikipedia.org/wiki/Rozk%C5%82ad_normalny

σ - Standard deviation (s)

68% of values are within 1 standard deviation of the mean

95% of values are within 2 standard deviations of the mean

99,7% of values are within 3 standard deviations of the mean

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Normal distribution (Gaussian)

Visegrad Network

- How to check if your data the "normally distributed"?
  - Kolmogorov-Smirnov test for normality
  - Shapiro–Wilk test
- The other way is to calculate the average, SD, median, min and max and evaluate the results.
- If SD is high and median is sligthly different form the avaerage there is the risk that the distribution in not normal.
- If the range (min and max values) is wide it is recomended to use nonparateric tests.
- Pay attantion on the **outliers** (the observations that are distant from other observations). If the measured values are not included in the range of ±2SD from the average it should be rejceted or replaced!!

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Means

Visegrad Network

- Arithmetic mean - the sum of the numbers divided by how many numbers are being averaged. Arithemtic average should be calculated only when when adding up the values makes sense !!

$$\overline{X} = \frac{\sum X}{N}$$

- **Weighted mean** - if the numbers of next measumnets are different

eg. The researcher measured of the shear strength for the SAC solder joints. The results came from 3 substrates with mounted resistors (1 substrate - 10 resisotrs, 2 – substrate -15 resistors, 3 – substrate 5 resistors).

$$\overline{X}_w = \frac{10 \cdot \overline{X} + 15 \cdot \overline{X} + 5 \cdot \overline{X}}{10+15+5}$$

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Variance, standard deviation

Visegrad Network

- The **variance** (σ^2 – whole population, s^2 – sample population) is a measure of **how far each value in the data set is from the mean.**

$$s^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}{n-1}$$

- Standard deviation is calculated as the square root of the variance.

$$s = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}{n-1}}$$

x_i is the i^{th} observation from a sample of the population,

x-bar is the sample mean,

n (sample size) -1 is degrees of freedom,

Σ is the summation

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Median, Mode

- The **median** is the value separating the higher half of a data sample, a population, or a probability distribution, from the lower half. For a data set, it may be thought of as the "middle" value (the median is the most resistant statistics).
  Examples:
  1, 3, 3, 6, 7, 8, 9 — median = 6
  1, 2, 3, 4, 5, 6, 8, 9 — median = 4.5
- The **modal** it is the value that is most likely (the most frequent)
  Examples:
  1, 3, 6, 6, 6, 6, 7, 7, 12, 12, 17
  modal = 6
  (the most frequent size of the single void)

Visegrad Fund

Parametric tests

- **Parametric test** is one that makes assumptions about the parameters (defining properties) of the population distribution(s) from which one's data are drawn.
- A parametric test is more able to reject of H_0.
- One- and two-tailed tests
  - A two-tailed test is appropriate if the estimated value may be more than or less than the reference value.
  - A one-tailed test is appropriate if the estimated value may depart from the reference value in only one direction (just differ).

Visegrad Fund

Nonparametric tests

- If the distribution is not normal and the number of the sample is not high.
- **Nonparametric tests** do not rely on any distribution.
- They can thus be applied even if parametric conditions of validity are not met.

Visegrad Fund

Parametric and nonparametric tests

Visegrad Network

Parametric tests (normal distibution)	Nonparametric tests (any distribution including normal)
1. **Differences testing between independent groups**	
t-test ANOVA	U-test (Mann-Whitney) Kruskall-Wallis test by ranks
2. Differences testing between dependent groups	
t-test	Wilcoxon signed-rank test
3. Correlation between variables	
Pearson correlation coefficient	R Spearman test
Regression	

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Statistics software

Visegrad Network

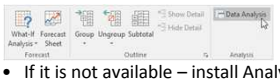
- Free statistical software is a practical alternative to commercial packages.
- These packages come from a variety of sources, including governments, nongovernmental organizations (NGOs) like UNESCO, and universities, and are also developed by individuals.
  - PSPP
  - EasyReg
  - Openstat
  - Online
    - http://www.socscistatistics.com/tests/Default.aspx
- Commercial
  - Excel
  - Statistica (Statsoft –Dell)

A. Skwarek – Statystical analysis of the results in brief

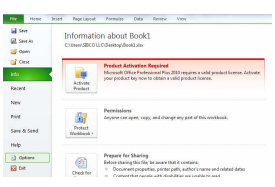
Visegrad Fund

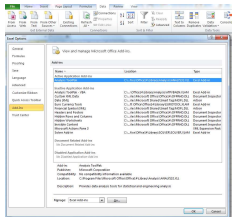
Excel- data analysis

Visegrad Network



- If it is not available – install Analysis ToolPak
- Click the File tab, click Options, and then click the Add-Ins category
- In the **Add-Ins** box, check the **Analysis ToolPak** check box, and then click **OK**





A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

t-test (parametric)

Visegrad Network

- The *t-test* is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis.
- A *t-test* is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known.
- The *t-test* can be used, for example, to determine if two sets of data are significantly different from each other.
- Several groups can not be compared with each other by performing *t-test* several times.
- The idea is to compare the mean and standard deviation of one group of subjects with the predetermined value.

$$t = \frac{\overline{X_1} + \overline{X_2}}{\sqrt{\left(\frac{(N_1-1)s_1^2 + (N_2-1)s_2^2}{N_1+N_2-2}\right)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}}$$

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

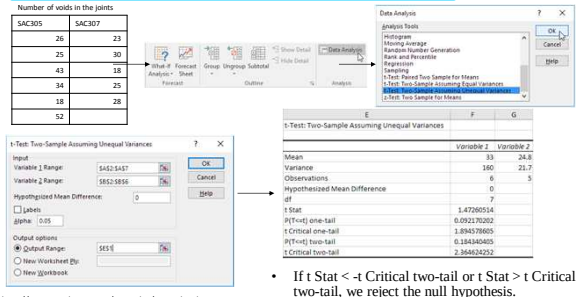
t-test example (Excel)

Visegrad Network

Comparision of the number of voids in SAC alloys – H_0 says that SAC305 and SAC307 has the same number of the voids.

Number of voids in the joints

SAC305	SAC307
26	23
25	30
43	18
34	25
18	28
52	



t-Test: Two-Sample Assuming Unequal Variances

	Variable 2	Variable 2
Mean	33	24.8
Variance	160	21.7
Observations	6	5
Hypothesized Mean Difference	0	
df	7	
t Stat	1.47260514	
P(T<=t) one-tail	0.092170202	
t Critical one-tail	1.894578605	
P(T<=t) two-tail	0.184340405	
t Critical two-tail	2.364624252	

http://www.excel-easy.com/examples/anova.html

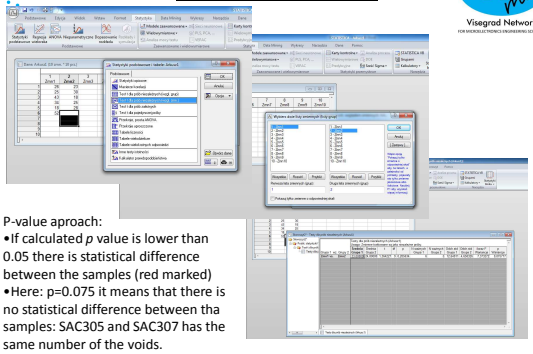
- If t Stat < -t Critical two-tail or t Stat > t Critical two-tail, we reject the null hypothesis.
- This is not the case, -2.365 < 1.473 < 2.365. Therefore, we do not reject the null hypothesis.

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

t-test example (Satistica)

Visegrad Network



P-value aproach:
- If calculated *p* value is lower than 0.05 there is statistical difference between the samples (red marked)
- Here: p=0.075 it means that there is no statistical difference between tha samples: SAC305 and SAC307 has the same number of the voids.

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Anova (parametric)

Visegrad Network

- Is one of the most popular and most commonly used statistical analyzes.
- More precisely - ANOVA it is the group of analyzes used to examine the influence of factors (independent variables) on the dependent variable.
- It is more advanced tool than *t-test.*
- The idea of variance analysis is to check whether certain independent variables (factors) influence the level of the dependent variable (measured values). Depending on the type of factors, we use different types of variance analysis.
- Analysis of variance is the ratio of the variance that we calculated between the studied groups and the average variance that we have we observed inside the groups.

Visegrad Fund

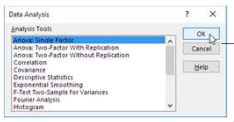
Anova example (Excel)

Visegrad Network

Comparision of the shear strength values of the solder joints - H_0 says that that shear strength of the different solder joints doesn't differ

Shear stremth values

SAC305	SAC307	mSAC
42	69	35
53	54	40
49	58	53
53	64	42
43	64	50
44	55	39
45	56	55
52		39
54		40



Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	9	435	48.33333	23.5
Column 2	7	420	60	32.33333
Column 3	9	393	43.66667	50.5

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1085.84	2	542.92	15.19623	7.16E-05	3.443357
Within Groups	786	22	35.72727			
Total	1871.84	24				

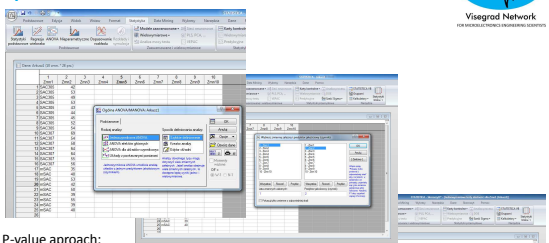
- Conclusion: if F > F crit, we reject the null hypothesis.
- This is the case, 15.196 > 3.443. Therefore, we reject the null hypothesis. The means of the three samples are not all equal.

http://www.excel-easy.com/examples/t-test.html

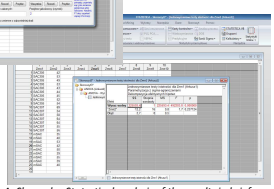
Visegrad Fund

Anova example (Statistica)

Visegrad Network



P-value aproach:
•If calculated *p* value is lower than 0.05 there is statistical difference between the samples (red marked)
•Here: p=0.000000... it means that there is statistical difference between tha samples: shear strength of the different solder joints (SAC305, SAC307, mSAC) differs.



Visegrad Fund

Ranking

Visegrad Network

- is a relationship between a set of items such that, for any two items, the first is either 'ranked higher than', 'ranked lower than' or 'ranked equal to' the second.

Value
57
74
43
78
61
64
58

Arrange the data in ascending order →

Value
43
57
58
61
64
74
78

Give ranks →

Rank
1
2
3
4
5
6
7

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

U-test (Mann-Whitney) (nonparametric)

Visegrad Network

- The *U-test* is one of the most popular alternatives for the *t-test* for independent trials.
- The dependent variable must be measured on an ordinal scale (it may also be measured on a quantitative scale).
- The use of the *U-test* does not require the parallelity of groups, normal distribution or homogeneous variances. This makes it widely applicable.
- The *U-test* is about ranking.
- The null hypothesis asserts that the medians of the two samples are identical.

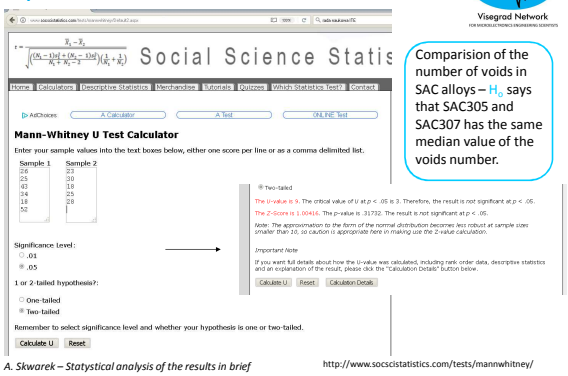
$$U = NM + \frac{N(N+1)}{2} - \sum_{x_i} Rank(x_i)$$

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

U-test example (Socscistatistics)

Visegrad Network



Comparision of the number of voids in SAC alloys – H_0 says that SAC305 and SAC307 has the same median value of the voids number.

A. Skwarek – Statystical analysis of the results in brief

http://www.socscistatistics.com/tests/mannwhitney/

Visegrad Fund

Kruskall-Wallis test by ranks (Socscistatistics)

Visegrad Network

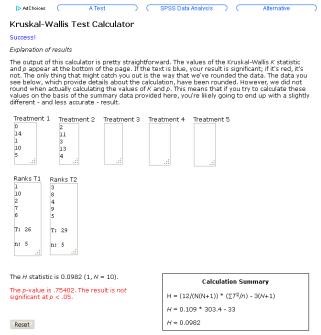
- The Kruskal-Wallis test is a non-parametric test, which means that it does not assume that the data come from a distribution that can be completely described by two parameters, mean and standard deviation (the way a normal distribution can).
- Alternative to the one-factor ANOVA test for **independent measures**.
- Like most non-parametric tests, you perform it on ranked data, so you convert the measurement observations to their ranks in the overall data set: the smallest value gets a rank of 1, the next smallest gets a rank of 2, and so on.
- You lose information when you substitute ranks for the original values, which can make this a somewhat less powerful test than a one-way Anova.

$$H = \frac{12}{N(N+1)} \cdot \frac{\sum T^2}{n} - 3(N+1)$$

Visegrad Fund

Kruskall-Wallis test by ranks example (Socscistatistics)

Visegrad Network


Kruskal-Wallis Test Calculator

Success!

Explanation of results

The output of this calculator is pretty straightforward. The values of the Kruskal-Wallis *H* statistic and *p* appear at the bottom of the page. If the text is blue, your result is significant; if it's red, it's not. The only thing that might catch you out is the way that we've rounded the data. The data you see below, which provide details about the calculation, have been rounded. However, we did not round when actually calculating the values of *H* and *p*. This means that if you try to calculate these values on the basis of the summary data provided here, you're likely going to end up with a slightly different - and less accurate - result.

Treatment 1: 0, 14, 1, 10, 5

Treatment 2: 2, 11, 3, 13, 4

Treatment 3

Treatment 4

Treatment 5

Ranks T1: 1, 10, 2, 7, 6 — T1: 26 — n1: 5

Ranks T2: 3, 9, 4, 8, 5 — T2: 29 — n2: 5

The *H* statistic is 0.0982 (1, *N* = 10).

The p-value is .75402. The result is not significant at $p < .05$.

Reset

Calculation Summary

H = (12/(N(N+1)) * (ΣT²/n) - 3(N+1)

H = 0.109 * 303.4 - 33

H = 0.0982

Comparision of the number of voids in SAC alloys – H_0 says that SAC305 and SAC307 has the same median value of the voids number.

http://www.socscistatistics.com/tests/kruskal/Default.aspx

Visegrad Fund

Wilcoxon signed-rank test

Visegrad Network

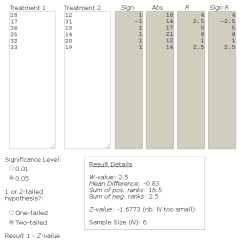
- The Wilcoxon test is a nonparametric test designed to evaluate the difference between two treatments or conditions where the samples are correlated (**dependant measures**).
- In particular, it is suitable for evaluating the data from a repeated-measures design in a situation where the prerequisites for a dependent samples t-test are not met.
- So, for example, it might be used to evaluate the data from an experiment that looks at the reading ability of children before and after they undergo a period of intensive training.

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Visegrad Fund

Wilcoxon signed-rank test example (Socscistatistics)

Visegrad Network



Comparision of the number of Sn whisker before and after thermal shocks – H_0 says that there is no difference in Sn whisker number before and after thermal shocks.

- Wilcoxon test requires more than 20 measured values.
- Depends on the population size eithe z value, eithe W value is compared

A. Skwarek – Statystical analysis of the results in brief

http://www.socscistatistics.com/tests/signedranks/Default.aspx

Visegrad Fund

Regression

Visegrad Network

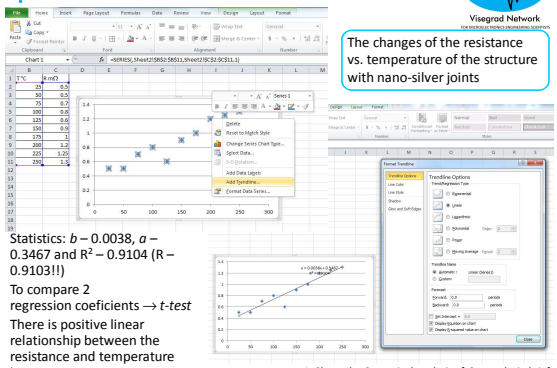
- The main idea of regression is prediction, forecasting data for a certain variable based on other variables. In other words, what value a given variable will take when we know the value of another variable.
- In regression it is assumed that the increase of one variable (predictor, predictors) is accompanied by an increase or decrease on the second variable.
- If the regression function as a linear function, i.e. y = bx + a – it is linear regression (y- dependent variable, x - independent variable, b - regression coefficient, a –absolute term).
- Determination coefficient R^2 - is a statistic that will give some information about the goodness of fit of a model.
- Correlation coefficent R - square root of the determination coefficient!

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Regression example (Excel)

Visegrad Network



The changes of the resistance vs. temperature of the structure with nano-silver joints

- Statistics: *b* – 0.0038, *a* – 0.3467 and R^2 – 0.9104 (R – 0.9103!!)
- To compare 2 regression coeficients → *t-test*
- There is positive linear relationship between the resistance and temperature increase.

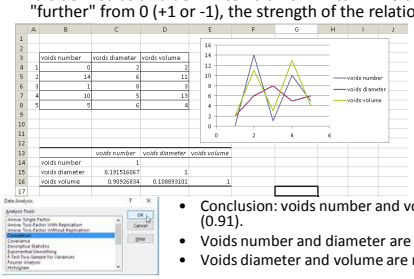
A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

Correlation – Pearson coefficient

Visegrad Network

- Correlation means the relationship between variables.
- The Pearson correlation coefficient is used to measure the strength of a linear association between two variables.
- It is defined as a value in intervals from -1 to 1. The coefficient is "further" from 0 (+1 or -1), the strength of the relationship is greater.



Voids number in the solder joint is correlated with their diameter and their volume.

- Conclusion: voids number and volume are correlated (0.91).
- Voids number and diameter are not correlated (0.19).
- Voids diameter and volume are not correlated (0.11) .

A. Skwarek – Statystical analysis of the results in brief

http://www.excel-easy.com

Visegrad Fund

R Spearman test

Visegrad Network

- Spearman's Rho measures the strength and direction of the relationship between two variables.
- Requirements
  - Scale of measurement must be ordinal (or interval, ratio)
  - Data must be in the form of matched pairs
  - The association must be monotonic (i.e., variables increase in value together, or one increases while the other decreases).

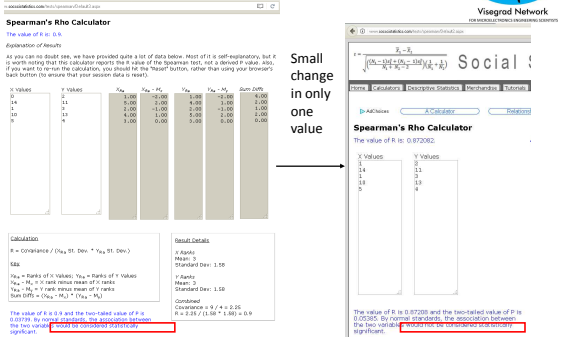
$$r_s = 1 - \frac{6\sum D^2}{N^3 - N}$$

A. Skwarek – Statystical analysis of the results in brief

Visegrad Fund

R Spearman test example (Socscistatistics)

Visegrad Network



Small change in only one value

A. Skwarek – Statystical analysis of the results in brief

http://www.socscistatistics.com/tests/spearman/Default.aspx

Visegrad Fund

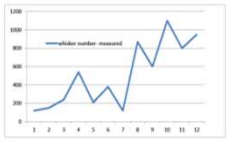
Expotential smoothing (Excel)

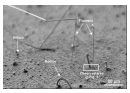
Visegrad Network

Exponential smoothing is used to smooth out irregularities (peaks and valleys) to easily recognize trends.

Number of whiskers

150
240
540
210
380
120
870
400
1100
700
950





Click in the Damping factor box and type 0.9.

Smaller damping factors also mean that your smoothed values are closer to the actual data points than larger damping factors.

A. Skwarek – Statystical analysis of the results in brief

http://www.excel-easy.com

Visegrad Fund

Social Science Statistics

Visegrad Network

The web site offers free resources for students and researchers working with statistics

http://www.socscistatistics.com/Default.aspx