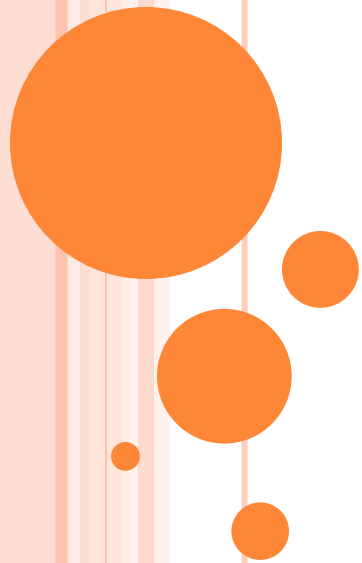


STATISTICAL TESTS



POPULATION - SAMPLE

- When a statistic investigation is carried out on a sample, any result obtained has a relative value.
- We can not even say with certainty what is the difference between the observed data in the sample and the data in population, since the state of the population is usually unknown.



STATISTICAL HYPOTHESIS

- Therefore, referring to the populations studied starting from the samples, the researchers can not state judgments with certainty, but can express some assumptions which are called **statistical hypotheses**.
- **Statistical tests** will evaluate the statistical hypotheses. After the test, the statistical hypotheses will be accepted or rejected.



THE MODEL FOR THE STATISTICAL TESTS

- Statistical hypothesis testing is done by comparing two competing hypotheses :
 - Null hypothesis - the model that the researcher wants to replace
 - Alternative hypothesis - the new model to replace the null hypothesis



THE GOAL OF THE STATISTICAL TEST

- The goal of the statistical test is to reject the null hypothesis
- The result of the test:
 - Reject the null hypothesis and accept the alternative one
 - Can't reject the null hypothesis, can't demonstrate the alternative one



FORMULATION OF PROBLEMS IN TERMS OF STATISTICAL HYPOTHESIS

- Define the null hypothesis - the test is conducted under the assumption that the null hypothesis is true
- Define the alternative hypothesis - also called a working hypothesis – contradicts the null hypothesis



THE LEVEL OF SIGNIFICANCE

- The level of significance is the size of the risk for errors which the researcher is willing to accept and to assume
- Usually the level of significance is between 1 și 5%
- The confidence level = $1 - \text{level of significance}$



STATISTICAL SIGNIFICANCE

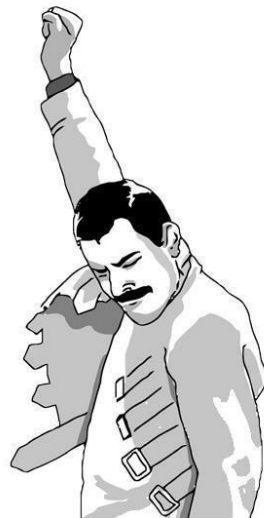
- P – Probability
- Theoretical value for the minimal level of significance for which the null hypotheses can be rejected
- It is specific to statistical software



THE RESULT OF THE TEST USING THE STATISTICAL SIGNIFICANCE

- The null hypothesis can be rejected if the results are at least significant
- If the result of the test are not significant the null hypothesis can not be rejected

$p < 0.05$



ERRORS

		Truth	
		H_0 True	H_0 False
Result	H_0 rejected	Type I error (α)	Correct
	H_0 not rejected	Correct	Type II error (β)



TYPE I ERROR

- = H_0 is rejected but is true
- We concluded that there are real differences although they are due to chance
- We conclude that a treatment is effective based on a misinterpretation



TYPE II ERROR

- = H_0 is not rejected even that it is false
- We concluded that the observed differences are due to chance even when they arise due to differences between samples
- We could abandon a new treatment or a new research direction
- “Power of the test” = $1 - \beta$



DESCRIPTIVE STATISTICS PARAMETERS

- Mean
- Median
- Variance, etc.

- Frequency

- Difference
- Relationship
- Independence
- Normality
- Correlation, etc.



QUANTITATIVE PARAMETERS

- We compare the height of people who do performance sports with height of people who do not practice performance sports
- Null hypotheses: there is no difference between the height of people who do performance sports and height of people who do not practice performance sports



ALTERNATIVE HYPOTHESES

- The height of people who practice performance gymnastics is lower than the height of people who do not practice performance sports
- The height of people who practice performance basketball is higher than height of people who do not practice performance sports
- The height of people who practice performance skiing is different than height of people who do not practice performance sports



ONE TAIL TEST VS, TWO TAIL TEST

- When we are interested in a specific direction of the comparison (lower, higher) – one tail test
- When we just want to see difference – no matter in which direction – two tail test



EXAMPLE – TWO TAIL TEST

- We are interested to assess the performance of an automatic blood pressure monitor, so we'll compare the results of that with the results given by a manual blood pressure monitor.
- The automatic blood pressure monitor is **NOT CONFORM** if its results vary more than 10% of the results given by a manual blood pressure monitor.



EXAMPLE – ONE TAIL TEST

- We want to assess the efficacy of a new toothpaste comparing with an older product.
- The new toothpaste will be better if the number of dental caries for people which use it is lower than the number of dental caries for people which use the other product.



SAMPLE TYPE

- Independent sample – there is no connection between the objects in the samples
- Dependent (pair) sample – for each object in first sample we have the corresponding one in the second sample



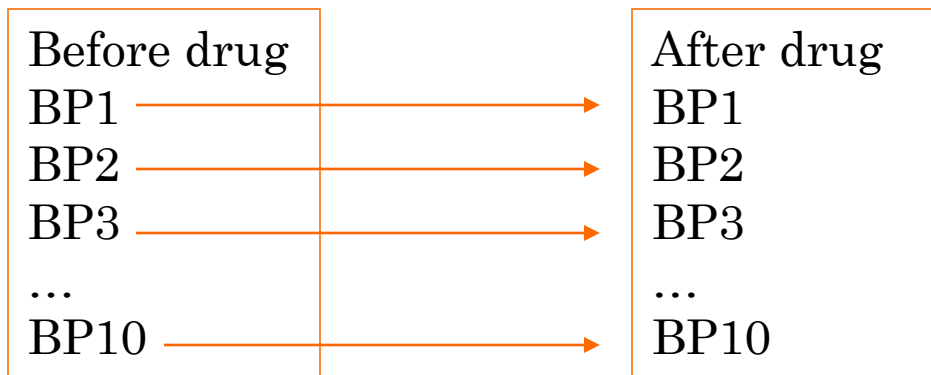
EXAMPLE OF INDEPENDENT SAMPLES

- We want to assess which sport develops better long-distance running ability. For this, we compare the 5 km running time of a group of football players with the 5 km running time of a group of handball players



EXAMPLE OF DEPENDENT SAMPLE

- In order to assess the efficacy of a drug in reducing the blood pressure we measure for 10 persons the blood pressure before and after taking that drug



CONCLUSION

- To apply a statistical test, we must know:
- What I want to test
- If it is a quantitative variable -do I use one tail or two tail test
- What kind of samples are involved (independent, pairs)



Quantitative variables

Data follow normal distribution

- **Z**
- **T**
- **ANOVA**
- ...

Parametrical tests

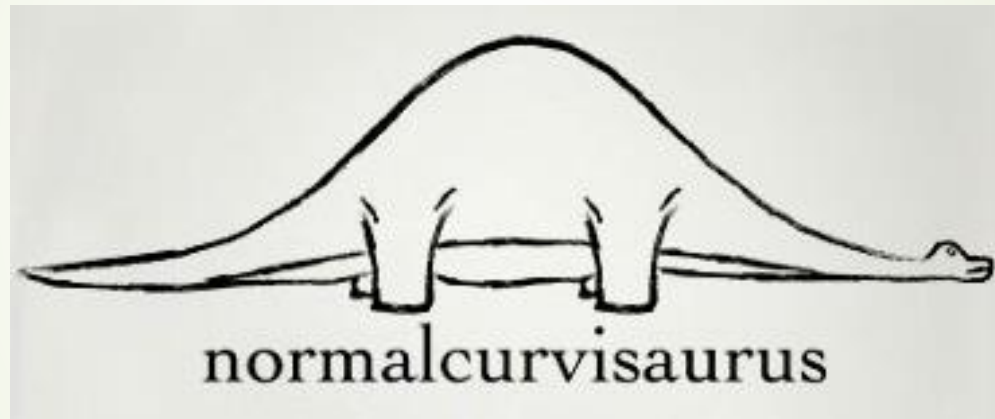
Data not follow normal distribution

- **Kruskal-Wallis**
- **Mann-Whitney U**
- **Mann-Whitney-Wilcoxon**
- **Wilcoxon signed-rank**
- ...

Non - Parametrical tests

Quantitative variables

A **normality test** is used to determine whether sample data has been drawn from a normally distributed population (within some tolerance).



Quantitative variables

- ▶ Example
- ▶ We are interested to assess the performance of an automatic blood pressure monitor, so we'll compare the results of that with the results given by a manual blood pressure monitor.
- ▶ The automatic blood pressure monitor is NOT CONFORM if its results vary more than 10% of the results given by a manual blood pressure monitor.
- ▶ In order to do that we'll measure the blood pressure in a sample of 98 persons with both devices - variable: SBP/DBP – continuous quantitative variable
- ▶ Should we use a parametrical test, or should we use a non-parametrical test?



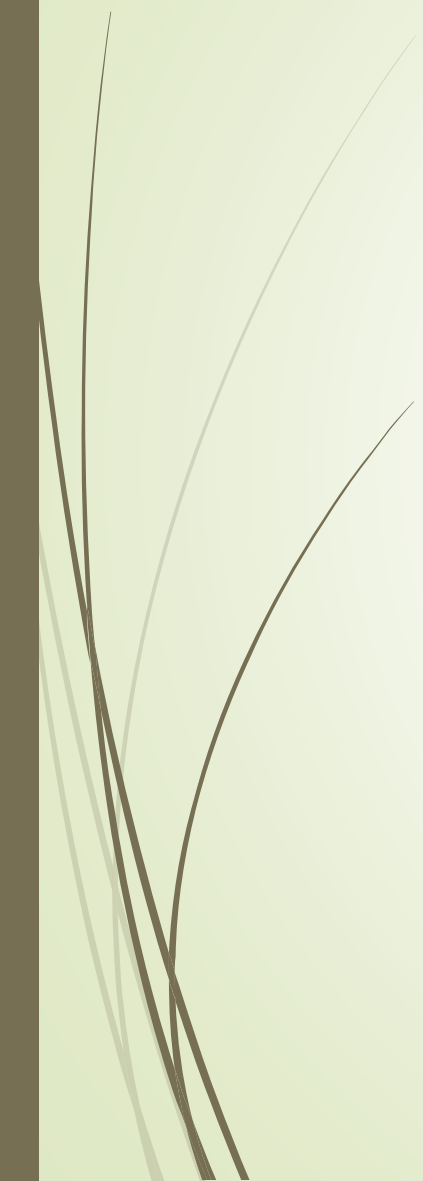


Normality tests

- **Kolmogorov-Smirnov (if sample size ≥ 50)**
- **Shapiro-Wilk (if sample size < 50)**
- Test the internal quality of the sample



Normality tests - Hypotheses

- ▶ H_0 : There is no difference between the sample distribution and normal distribution
 - ▶ H_1 : There is difference between the sample distribution and normal distribution
- 



Normality tests - decision

The result of the test: p

If $p \leq 0,05$ H_0 can be rejected, H_1 could be accepted with 95% of confidence – the sample distribution is different than normal distribution

If $p > 0,05$ H_0 cannot be rejected, we cannot demonstrate that the sample distribution is different than normal distribution

Example - decision

- H_0 : There is no difference between the distribution of the data in the sample and the normal distribution
- H_1 : There is difference between the distribution of the data in the sample and the normal distribution
- We apply Kolmogorov-Smirnov (sample size=98 \geq 50)
- $p=0,03$
- Because $p < 0,05$ we can reject H_0 and accept H_1 with 95% of confidence, so with 95% of confidence, there is difference between the distribution of the data in the sample and the normal distribution, so we must use a non-parametrical statistical test





- We decide to repeat the experiment, this time we have a sample size 32
- We apply Shapiro-Wilk test (sample size=32<50)
- $p=0,09$
- Because $p>0,05$ we cannot reject H_0 , we cannot demonstrate that there is difference between the distribution of the data in the sample and the normal distribution, so we must use a parametrical statistical test

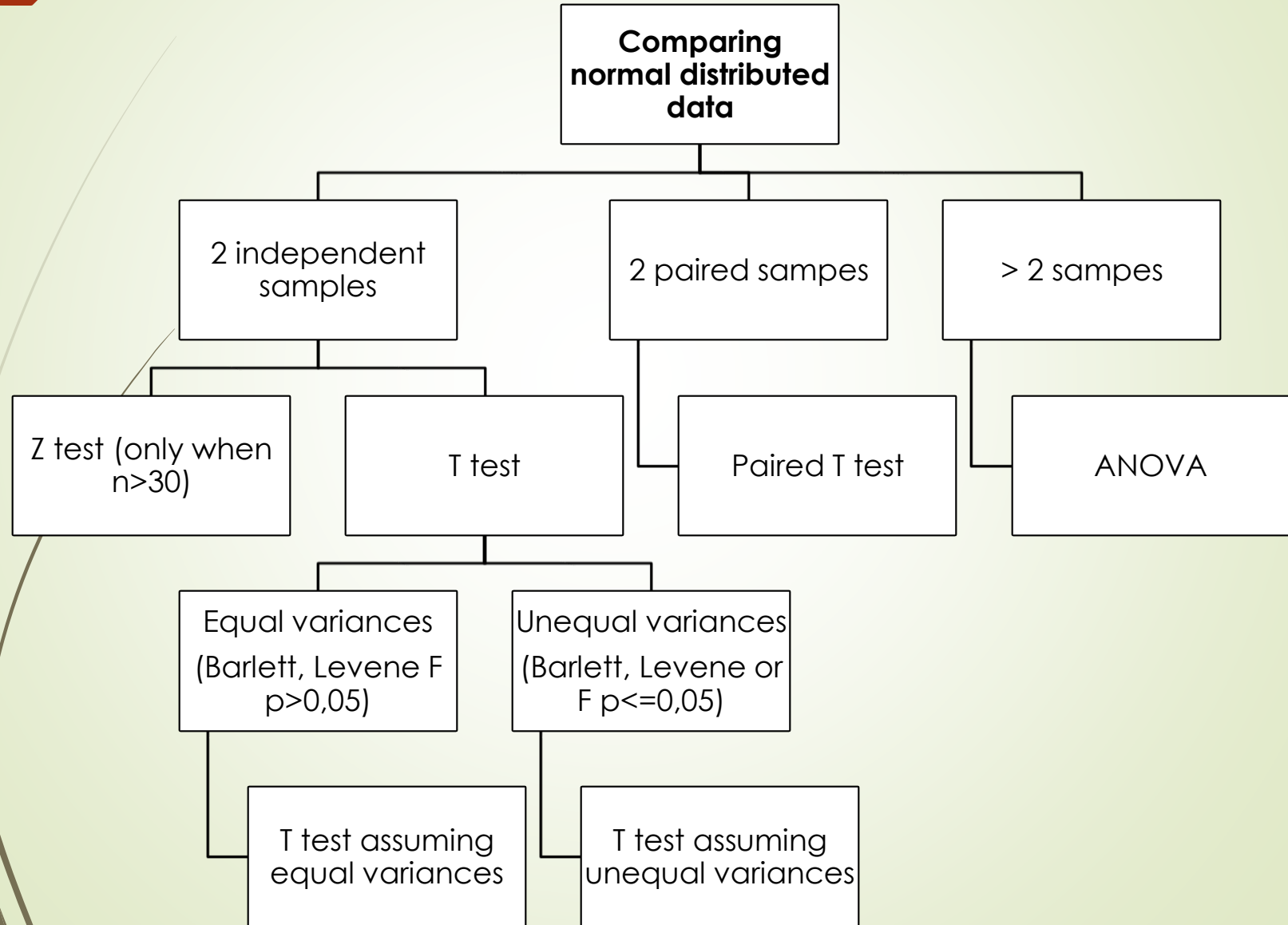




Conclusion

- ▶ When we have to analyze a sample according to a quantitative variable we must decide if we'll use parametrical or non-parametrical tests. For that we must check if the data distribution is similar with normal distribution using normality tests (Kolmogorov-Smirnov or Shapiro-Wilk test).
- ▶ If $p > 0,05$, parametrical tests should be used
- ▶ If $p \leq 0,05$ non-parametrical tests should be used

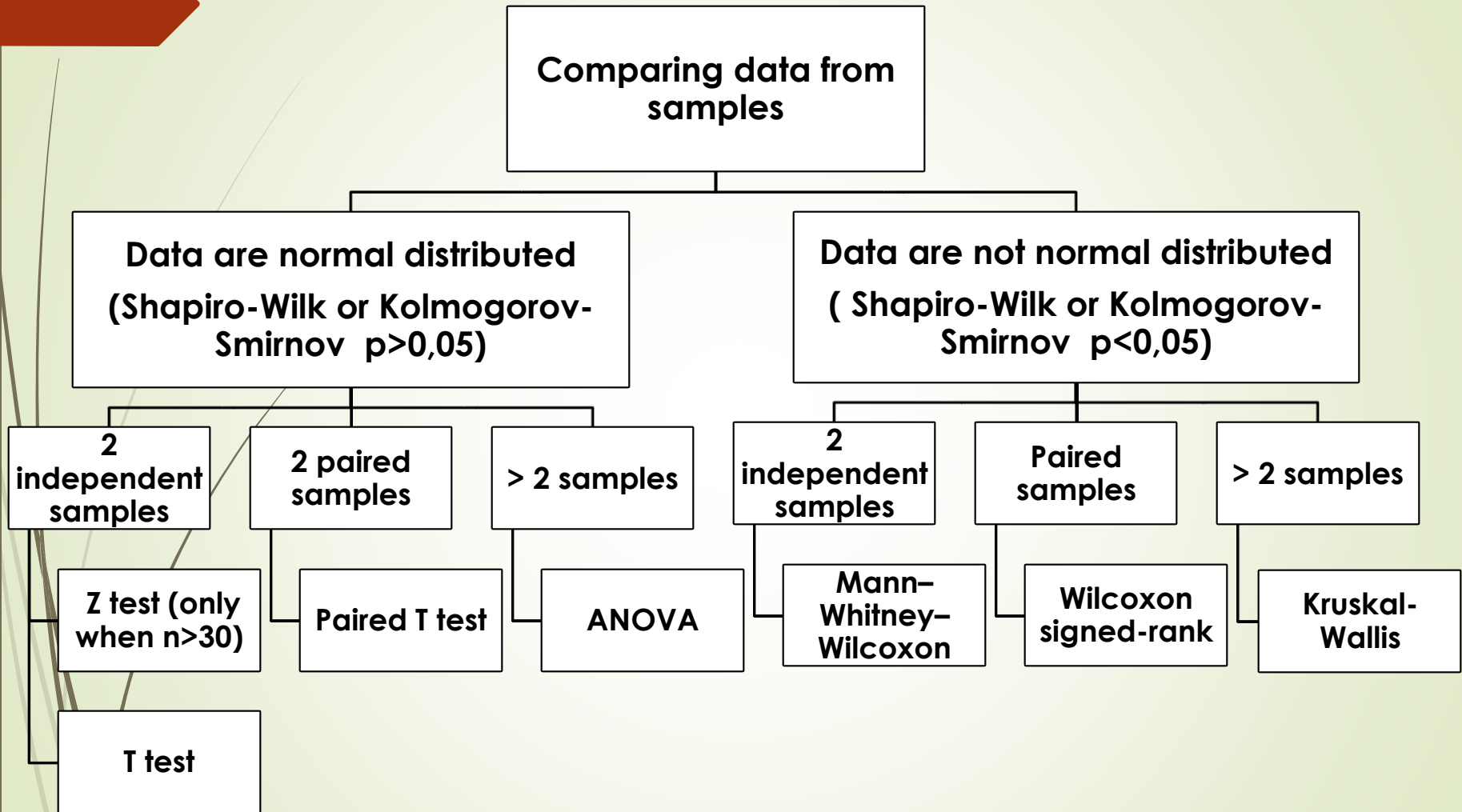
Choosing the test– quantitative normal distributed data



Parametrical and non-parametrical tests

Parametrical test	Non-parametrical equivalent test
Z, ANOVA	Kruskal-Wallis
Student (independent samples) First apply test for variances F, Levine, Barlet, etc.	Mann-Whitney U Mann-Whitney-Wilcoxon Wilcoxon-Mann-Whitney Wilcoxon rank-sum
Student (paired samples)	Wilcoxon signed-rank

Choosing the test – quantitative variables



CHI SQUARE TEST

- IS USED TO TEST THE INDEPENDENCE (ASSOCIATION) BETWEEN QUALITATIVE VARIABLES
- THE NULL HYPOTHESIS: THERE IS NO ASSOCIATION BETWEEN VARIABLES
- THE ALTERNATIVE HYPOTHESIS: THERE IS ASSOCIATION BETWEEN VARIABLES


CHI SQUARE BY EXAMPLE – DICHOTOMIAL VARIABLES

- WE ARE SEARCHING FOR HYPOTHETICAL ASSOCIATION BETWEEN SMOKING AND LUNG CANCER. FOR THAT, FROM ONCOLOGY DEPARTMENT RECORDS, WE SELECT A GROUP OF 160 PEOPLE WHO WERE DIAGNOSED WITH LUNG CANCER. WE ALSO SELECT A CONTROL GROUP OF 240 PERSONS WHICH WERE NEVER DIAGNOSED WITH LUNG CANCER. THE STUDY OF THIS SAMPLE LEADS US TO THE FOLLOWING CONTINGENCY TABLE:

	Cancer present	Cancer absent	Total
Smoking yes	80	50	130
Smoking No	80	190	270
Total	160	240	400



THE HYPOTHESES, LEVEL OF SIGNIFICANCE, CRITICAL REGION

- H_0 : THERE IS NO ASSOCIATION BETWEEN SMOKING AND LUNG CANCER
 - H_1 : THERE IS ASSOCIATION BETWEEN SMOKING AND LUNG CANCER
 - WE'LL USE A LEVEL OF SIGNIFICANCE OF 5%
- 

HTTPS://STATPAGES.INFO/CTAB2X2.HTML

2-way Contingency Table Analysis

This page computes various statistics from a 2-by-2 table. It will calculate the Yates-corrected chi-square, the Mantel-Haenszel chi-square, the Fisher Exact Test, and other indices relevant to various special kinds of 2-by-2 tables:

1. analysis of risk factors for unfavorable outcomes (odds ratio, relative risk, difference in proportions, absolute and relative reduction in risk, number needed to treat)
2. analysis of the effectiveness of a diagnostic criterion for some condition (sensitivity, specificity, prevalence, pos & neg predictive values, adjusted predictive values, pos & neg likelihood ratios, diagnostic and error odds ratios)
3. measures of inter-rater reliability (% correct or consistent, mis-classification rate, kappa, Forbes' NMI)
4. other measures of association (contingency coefficient, Cramer's phi coefficient, Yule's Q)

Many of these concepts are explained in detail in an online [Evidence-based Medicine Glossary](#) or [Center for Evidence-based Medicine](#). For more information about a particular index, click on the ["more info"](#) link for that index.

Confidence intervals for the estimated parameters are computed by a general method (based on "constant chi-square boundaries") given in: *Statistical Methods for Rates and Proportions* (2nd Ed.) Section 5.6, by Joseph L. Fleiss (Pub: John Wiley & Sons, New York, 1981). This method is also described in *Numerical Recipes in C* (2nd Ed.) Section 15.6, by William H. Press et al. (Pub: Cambridge University Press, Cambridge UK, 1992). Wilson's method is used to find CI for adjusted predictive values. Ref: *Statistics with Confidence* (2nd Ed.) by DG Altman et al. (Pub: British Medical Journal Books, UK 2000) p. 46-7. The reference used for CI calculation for RIOCI is: Relative Improvement Over Chance (RIOCI) and Phi as Measures of Predictive Efficiency and Strength of Association in 2 × 2 Tables David P. Farrington and Rolf Loeber. *Journal of Quantitative Criminology* vol. 5, No. 3 (September 1989), p. 201-213.

Enter the observed numbers into the four cells below. Move between cells using tab-key. Make sure that the row and column totals add up correctly. Then click the Compute button.

Warning: Do not enter cell counts with a leading zero! That is, if a cell count is 34, enter it as 34, not as 034. Some browsers will mis-interpret some numbers entered with leading zeros, and will produce wrong results (with no warning message).

Observed Contingency Table

		Condition /disease		Totals
		Present	Absent	
T e s t	Positive	<input type="text"/> = a (TP)	<input type="text"/> = b (FP)	
	Negative	<input type="text"/> = c (FN)	<input type="text"/> = d (TN)	
Totals				

Confidence Level: 95 %

Select from list or enter a more appropriate conf. level

Compute

RESULTS

Observed Contingency Table

	Condition /disease		Totals
	Present	Absent	
Positive	80 = a (TP)	50 = b (FP)	130 = r1
Negative	80 = c (FN)	190 = d (TN)	270 = r2
Totals	160 = c1	240 = c2	400 = t

Confidence Level: 95 %

Select from list or enter a more appropriate conf. level

Compute | Download result as CSV

Diagnostic Test

Specify disease prevalence, test sensitivity, test specificity and sample size

Prevalence (eg. 0.1)	1	* Enter your population prevalence estimate here
Sensitivity (eg. 0.8)	.8	A) if you want to run a diagnostic test
Specificity (eg. 0.8)	.8	B) if it turns out that the sample and actual population prevalence differs substantially.
Total sample size	1000	Consider reporting the adjusted PPV/NPV's.
Run diagnostic test with above parameters		<more info>

Expected Frequencies			Cell Proportions			Row Proportions			Column Proportions		
52.000	78.000	130.000	0.200	0.125	0.325	0.615	0.385	0.325	0.500	0.208	
108.000	162.000	270.000	0.200	0.475	0.675	0.296	0.704	0.675	0.500	0.792	
160.000	240.000		0.400	0.600		0.400	0.600		0.400	0.600	

Chi-Square Tests

Type of Test	Chi Square	d.f.	p-value
Pearson Uncorrected	37.227	1	0.000
Yates Corrected	35.909	1	0.000
Mantel-Haenszel	37.134	1	0.000

Fisher Exact Test

Type of comparison (Alternate Hypothesis)	p-value
Two-tailed (to test if the Odds Ratio is <i>significantly different</i> from 1): If you don't know which Fisher Exact p-value to use, use this one . This is the p-value produced by SAS, SPSS, R, and other software.	0.000
Left-tailed (to test if the Odds Ratio is <i>significantly less</i> than 1):	1.000
Right-tailed (to test if the Odds Ratio is <i>significantly greater</i> than 1):	0.000
Two-tailed p-value calculated as described in Rosner's book: (2 times whichever is smallest: left-tail, right-tail, or 0.5) It tends to agree closely with Yates Chi-Square p-value.	0.000
Probability of getting <i>exactly</i> the observed table: (This is not really a p-value; don't use this as a significance test.)	0.000
Verification of computational accuracy: (This number should be very close to 1.0; the closer, the better.)	1.0000000000

Quantities derived from a 2-by-2 table

Quantities Derived from the 2-by-2 Contingency Table	Value	95% CI	
o (OR) = (a/b)/(c/d);	3.800	2.449	5.897
risk (RR) = (a/r1)/(c/r2);	2.077	1.637	2.596
	0.301	0.199	0.397

CONCLUSION

- BECAUSE
 - $P < 0,05$ THERE IS ASSOCIATION BETWEEN SMOKING AND LUNG CANCER (95% CONFIDENCE)
 - $RR = 2,077$ 95%CI = [1,637 , 2,596] – SMOKING IS A RISK FACTOR FOR LUNG CANCER



WHEN WE APPLY CHI SQUARE?

- **TWO GROUPS**
- **A QUALITATIVE VARIABLE OF INTEREST**
- **THE ASSOCIATION BETWEEN TWO QUALITATIVE VARIABLES**
- **COMPARING THE OBSERVED DISTRIBUTION WITH THEORETICAL DISTRIBUTION**


COCHRAN RULE

	G1	G2	...	Total
E1	a	b	...	a + c + ...
E2	c	d	...	c + d + ...
...
Total	a + c + ...	b + d +	a + b + c + d + ...

$$Chi^2 = \sum \frac{(O - T)^2}{T}$$

- **CHI SQUARE TEST RESULT IS VALID ONLY WHEN 80% OF THEORETICAL FREQUENCIES ARE MORE THAN 5 AND ALL THEORETICAL FREQUENCIES ARE MORE THAN 1**

THEORETICAL CONTINGENCY TABLE – COCHRANE RULE



	Group 1	Group 2	Total
Var 1	a^t	b^t	a^t+b^t
Var 2	c^t	d^t	c^t+d^t
Total	a^t+c^t	b^t+d^t	$a^t+b^t+c^t+d^t$

Theoretical table

**All Values >5 – Chi square
test**

**More values between 0 and 5–
Fisher Exact test**

**One value between 2 and 5–
Yates Corrected Chi Square**