Comparing groups using continuous not-Normal data (Some non-parametric tests)

UNIVERSITY OF LIVERPOOL

Gabriela Czanner PhD CStat
Department of Biostatistics
Department of Eye and Vision Science

ST PAULS EYE UNIT

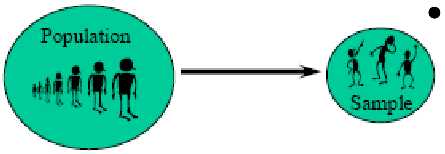
29 May 2013

MERSEY POSTGRADUATE TRAINING PROGRAMME

Workshop Series: Basic Statistics for Eye Researchers and Clinicians

How to compare several groups?

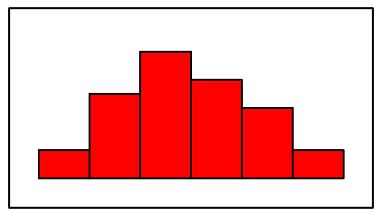
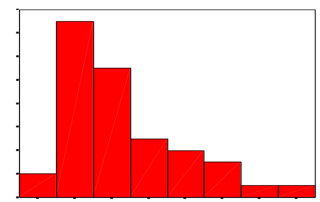
- Research question
  - Is Visual Acuity same across three patients groups e.g. diabetic retinopathy, diabetic maculopathy and healthy?
  - Is the degree of staining in conjuctivial tissues same across 5 different storage methods?
  - Is contrast sensitivity different at two different time points e.g. before and after a treatment?



- A strategy
  - To collect data on samples in each group
  - Then use the data to compare the groups via an appropriate data analysis method

Comparisons of two groups

Type of Data

Categorical	Continuous Normal	Continuous Skewed
e.g. Color blindness yes/no Or 2 severity grades of diabetic maculopathy		

Tests for differences, independent data:

Categorical	Continuous Normal	Continuous Skewed
Chi-square χ^2 test Fisher's exact test	t-test	Mann-Whitney U test

Tests for differences, non-independent (paired) data:

Categorical	Continuous Normal	Continuous Skewed
McNemar's test	Paired (single sample) t-test	Wilcoxon signed rank test

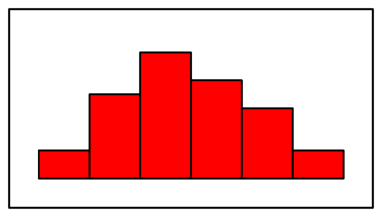
Comparisons of <u>three (or more)</u> groups

Type of Data

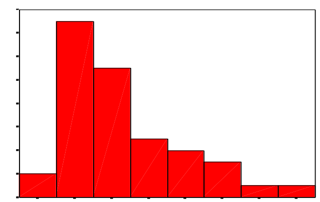
Categorical

e.g. 3 severity grades of diabetic maculopathy and 3 categories of age

Continuous Normal



Continuous Skewed



Tests for differences, independent data:

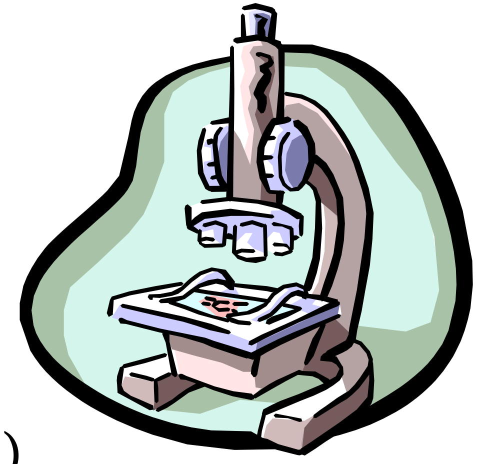
Categorical: Chi-square χ^2 test; Fisher's exact test

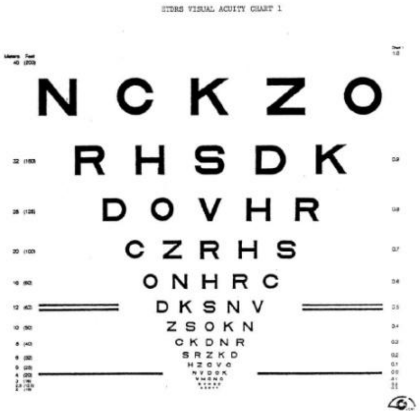
Continuous Normal: ANOVA & post-hoc analyses via two-sample t-test

Continuous Skewed: Kruskal-Wallis & post-hoc analyses with Mann-Whitney U test

Outline

- Comparison for 2 groups using data that are continuous, paired (non-independent) and skewed (not-Normal)
  - Wilcoxon test
- Comparisons for 2 groups using data that are continuous, independent and skewed (not-Normal)
  - Mann-Whitney's test





Nonparametric methods of groups comparison for continuous data

- Parametric methods (such as t-test, ANOVA) for continuous (interval) data assume response measurement is approximately Normally distributed. Assumption may not be justifiable, or difficult to check because of small amount of data.
- **Nonparametric** (or **distribution-free**) techniques do not require the Normality assumption where numerical data are replaced by rank values.

For paired continuous data: **Wilcoxon signed rank**

For independent continuous data: **Mann-Whitney U test**

Wilcoxon signed rank test

- The non-parametric equivalent of the paired (one sample) *t-test is the **Wilcoxon signed rank test.***
- For n pairs of observations the procedure is:
  - Calculate the difference between the pairs of observations.
  - Rank the differences in order of magnitude, ignoring the signs.
  - Sum the ranks of the negative differences to give T^- *and those of the positive differences to give* T^+
  - Set T *to be the smaller of* T^- *and* T^+
- The distribution of T *is tabulated for* $n < 50$*. A normal approximation can be used for large n.*

Example: Visual acuity and Wilcoxon signed rank test

- 13 patients of early diabetic retinopathy are seen twice: at a baseline and at six moths.
- Has their visual acuity (BCVA in letters) changed across the two time point?

BCVAbaseline	BCVAat6months
94	93
90	95
90	84
81	85
70	80
85	83
87	79
82	75
86	95
73	84

Null hypothesis ???
H0:

Alternative hypothesis???
H1:

Example: Visual acuity and Wilcoxon signed rank test

- 13 patients of early diabetic retinopathy are seen twice: at a baseline and at six moths.
- Has their visual acuity (BCVA in letters) changed across the two time point?

BCVAbaseline	BCVAat6months
94	93
90	95
90	84
81	85
70	80
85	83
87	79
82	75
86	95
73	84

Null hypothesis
H0: The distribution of BCVA is same at the two time points.

Alternative hypothesis
H1: the distribution is different

Data are too short to check Normality, so we will use Wilcoxon test.

Example: Visual acuity and Wilcoxon signed rank test

BCVAbaseline	BCVAat6months	Difference	Abs Diff	Rank	Rank +	Rank -
94	93	1				
90	95	-5				
90	84	6				
81	85	-4				
70	80	-10				
85	83	2				
87	79	8				
82	75	7				
86	95	-9				
73	84	-11				

1 step Calculate the difference between the pairs of observations.

Example: Visual acuity and Wilcoxon signed rank test

BCVAbaseline	BCVAat6months	Difference	Abs Diff	Rank	Rank +	Rank -
94	93	1	1	1		
90	95	-5	5	4		
90	84	6	6	5		
81	85	-4	4	3		
70	80	-10	10	9		
85	83	2	2	2		
87	79	8	8	7		
82	75	7	7	6		
86	95	-9	9	8		
73	84	-11	11	10		

2 step Rank the differences in order of magnitude, ignoring the signs.

Example: Visual acuity and Wilcoxon signed rank test … manually

BCVAbaseline	BCVAat6months	Difference	Abs Diff	Rank	Rank +	Rank -
94	93	1	1	1	1	
90	95	-5	5	4		4
90	84	6	6	5	5	
81	85	-4	4	3		3
70	80	-10	10	9		9
85	83	2	2	2	2	
87	79	8	8	7	7	
82	75	7	7	6	6	
86	95	-9	9	8		8
73	84	-11	11	10		10
				SUM	21	34
Here T- equals	34					
Here T+ equals	21					

Step 3 Find ranks in each group: Rank + and Rank-; and sum them up.

Example: Visual acuity and Wilcoxon signed rank test … manually

BCVAbaseline	BCVAat6months	Difference	Abs Diff	Rank	Rank +	Rank -
94	93	1	1	1	1	
90	95	-5	5	4		4
90	84	6	6	5	5	
81	85	-4	4	3		3
70	80	-10	10	9		9
85	83	2	2	2	2	
87	79	8	8	7	7	
82	75	7	7	6	6	
86	95	-9	9	8		8
73	84	-11	11	10		10
				SUM	21	34

Step 4 Test statistic.

Here T- equals	34
Here T+ equals	21

- Hence test statistic T= min(T- and T+) = **21**
- Under hypothesis of equality of distributions at 0 and 6 months, what value of T we expect ?
  - (a) Small and close to 0 ?
  - (b) Large and close to max rank which is 55 ?
  - (b) Or the average value of 1, 2, 3,….10 which is 55/2=27.5 ?

Example: Visual acuity and Wilcoxon signed rank test

BCVAbaseline	BCVAat6months	Difference	Abs Diff	Rank	Rank +	Rank -
94	93	1	1	1	1	
90	95	-5	5	4		4
90	84	6	6	5	5	
81	85	-4	4	3		3
70	80	-10	10	9		9
85	83	2	2	2	2	
87	79	8	8	7	7	
82	75	7	7	6	6	
86	95	-9	9	8		8
73	84	-11	11	10		10
				SUM	21	34

Here T- equals 34
Here T+ equals 21

- Hence test statistic T= min(T- and T+) = **21**
- Under hypothesis of equality of distributions at 0 and 6 months, what value of T we expect?
  - (a) Small and close to 0 ? **NO**
  - (b) Large and close to max rank which is 55 ? **NO**
  - (b) Or the average value of 1, 2, 3,….10 which is 55/2=27.5 ? **YES**

Example: Visual acuity and Wilcoxon signed rank test

BCVAbaseline	BCVAat6months	Difference	Abs Diff	Rank	Rank +	Rank -
94	93	1	1	1	1	
90	95	-5	5	4		4
90	84	6	6	5	5	
81	85	-4	4	3		3
70	80	-10	10	9		9
85	83	2	2	2	2	
87	79	8	8	7	7	
82	75	7	7	6	6	
86	95	-9	9	8		8
73	84	-11	11	10		10
				SUM	21	34

Here T- equals	34
Here T+ equals	21

- Hence test statistic T= min(T- and T+) =21
- Is value 21 small enough to reject H0?
- The 5% quantile of Wilcoxon test is **8** (from stats tables) i.e. we would reject null hypothesis of equality **if T<=8**.
- **Conclusion? Do we reject null hypothesis of equality of BCVA distribution across the two time points?**

Example: Visual acuity and Wilcoxon signed rank test

BCVAbaseline	BCVAat6months	Difference	Abs Diff	Rank	Rank +	Rank
94	93	1	1	1	1	
90	95	-5	5	4		4
90	84	6	6	5	5	
81	85	-4	4	3		3
70	80	-10	10	9		9
85	83	2	2	2	2	
87	79	8	8	7	7	
82	75	7	7	6	6	
86	95	-9	9	8		8
73	84	-11	11	10		10
				SUM	21	34

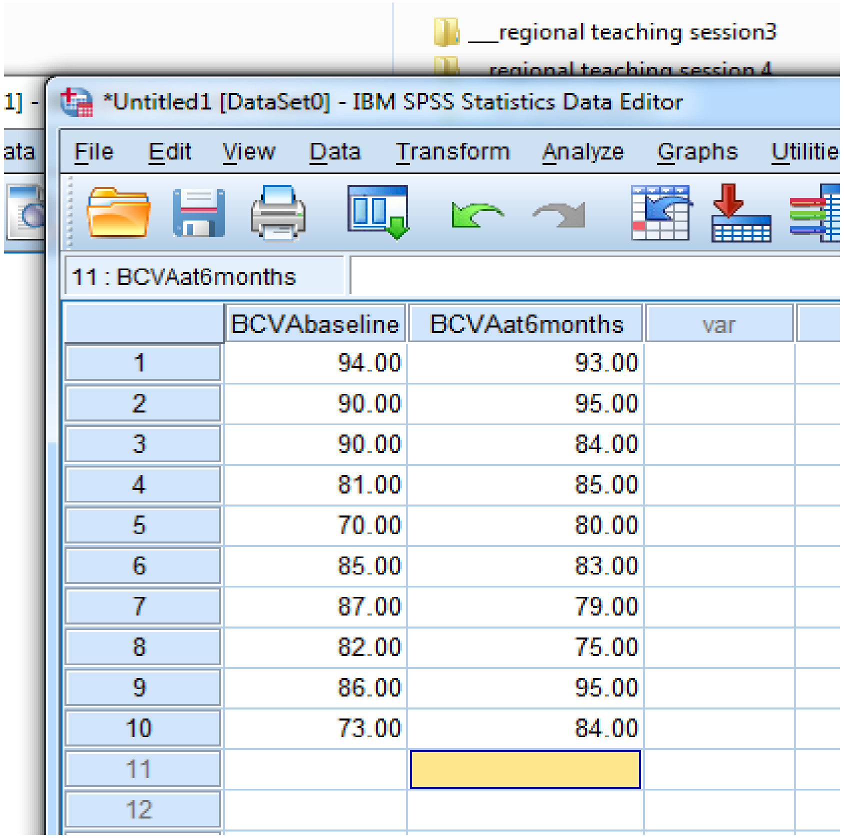
Here T- equals	34
Here T+ equals	21

- Hence test statistic T= min(T- and T+) =21
- Is value 21 small enough to reject H0?
- The 5% quantile of Wilcoxon test is **8** (from stats tables) i.e. we would reject null hypothesis of equality **if T<=8**.
- **Conclusion? Do we reject null hypothesis of equality of distribution of BCVA across the two time points? The answer is "NO"**

Example: Visual acuity and Wilcoxon test manually

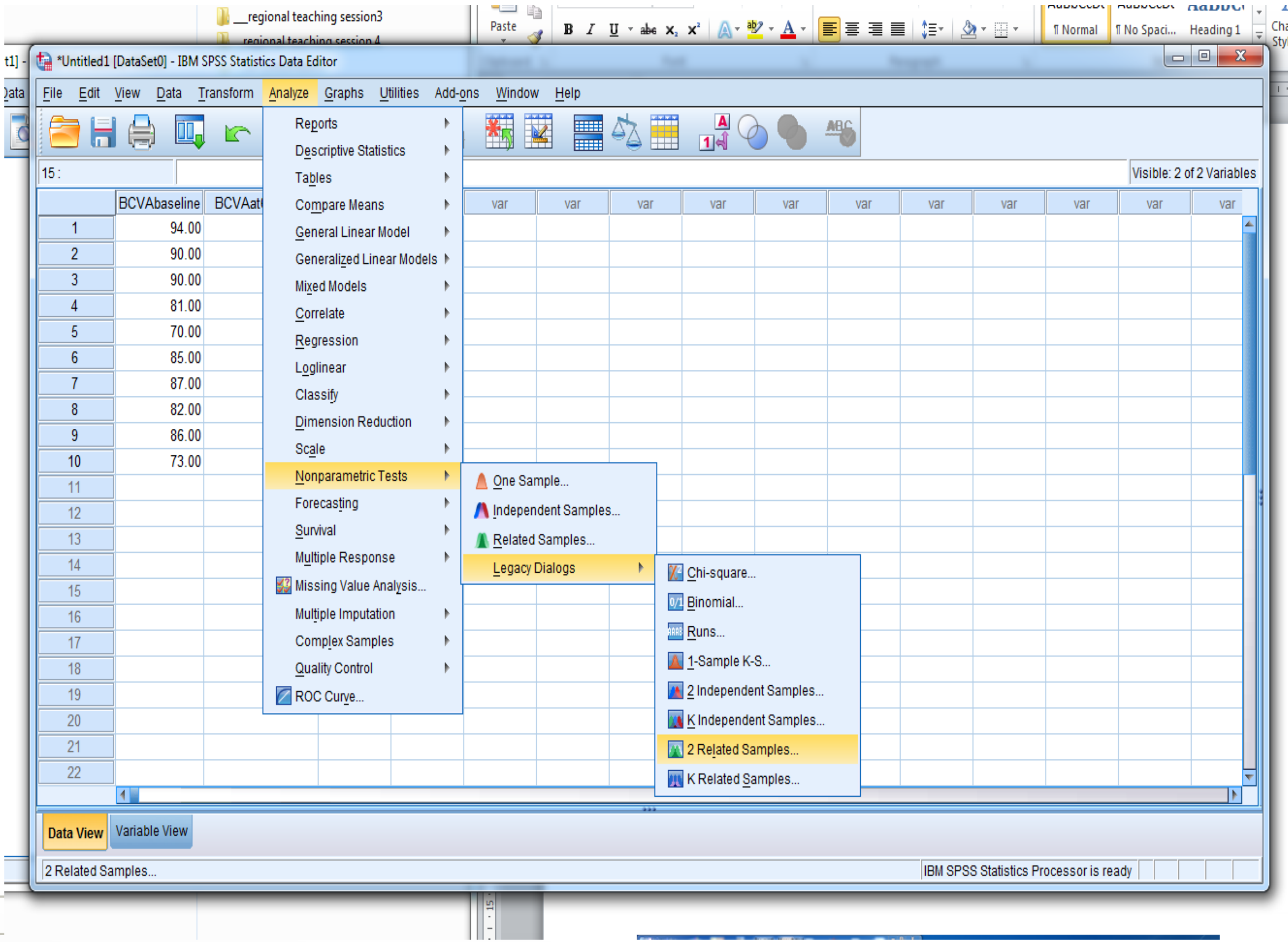
- From statistics tables the 5% quantile (critical value) is 8, hence T>8 i.e. 21>8, hence the result is NOT significant at the 5% significance level.
- The conclusion is that we do not have strong evidence of change in BCVA across the 6 months.
- Note that:
  - Zero differences are discounted and *n is decreased accordingly.*
  - Tied values are assigned the average of the ranks covered.

Wilcoxon test in SPSS software: data supplied in two columns

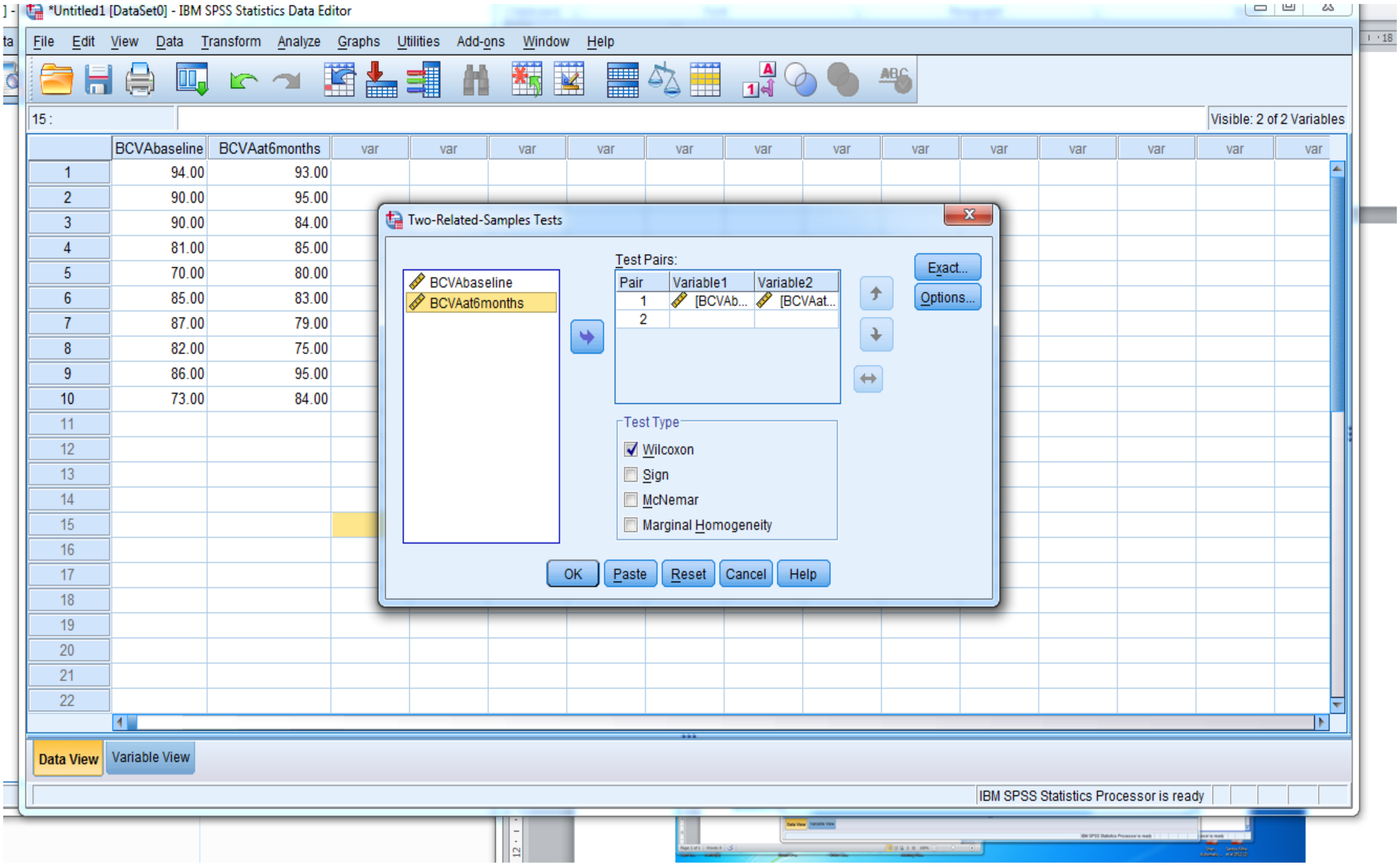


	BCVAbaseline	BCVAat6months	var
1	94.00	93.00	
2	90.00	95.00	
3	90.00	84.00	
4	81.00	85.00	
5	70.00	80.00	
6	85.00	83.00	
7	87.00	79.00	
8	82.00	75.00	
9	86.00	95.00	
10	73.00	84.00	
11			
12			

Wilcoxon test in SPSS software: where to find the test



Wilcoxon test in SPSS software: defining variables



Wilcoxon test in SPSS software: results

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
BCVAat6months - BCVAbaseline	Negative Ranks	5[a]	4.20	21.00
	Positive Ranks	5[b]	6.80	34.00
	Ties	0[c]		
	Total	10		

a. BCVAat6months < BCVAbaseline

b. BCVAat6months > BCVAbaseline

c. BCVAat6months = BCVAbaseline

Test Statistics[a]

	BCVAat6months - BCVAbaseline
Z	-.663[b]
Asymp. Sig. (2-tailed)	.508

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

From SPSS: p-value 0.508.
Do we reject null hypothesis at 5% significance level?

The conclusion is that we do not have strong evidence of change in BCVA across the 6 months.

Non-parametric group-comparison test for independent continuous non-Normal data

- Suppose we have two groups of individuals, with n_1 *in the first group and* n_2 *in the second.*

- We observe a response measurement for each individual in each group and wish to compare the measurements between the two groups.

- If response cannot be assumed to be normally distributed: means (or, the difference in means) are not an adequate summary of the data.

Example: Contrast sensitivity

- Contrast sensitivity (CS, in letters) was measured on 6 normal healthy subjects and 7 subjects with diabetic maculopathy ($n_1 = 6$, $n_2 = 7$).
- The distributions in each group are skewed, so a non-parametric test is preferable to the two-sample *t-test:*

Normals	DM
CS	CS
38	24
39	25
40	25
42	27
42	35
42	37
	41

Mann-Whitney U test

For two distinct groups, we use the **Mann-Whitney U test and proceed as follows:**

- Rank **all the measurements (from low to high)**
- To those ranks that are tied assign the average of the tied ranks.
- Sum the ranks of group 1, giving T_1
- Calculate $U_1 = n_1 n_2 + \frac{1}{2} n_1(n_1 + 1) - T_1$ and $U_2 = n_1 n_2 - U_1$
- Set *U to be the smaller of* U_1 *and* U_2
- The distribution of *U is tabulated for* n_1 *and* n_2 *less than* ***20***. *For larger values of n, a normal approximation can be used.*

Example: Contrast Sensitivity

- We are interested if there is a difference between normal healthy subjects and subjects with diabetic maculopahty in terms of contrast sensitivity.

Normals	DM
CS	CS
38	24
39	25
40	25
42	27
42	35
42	37
	41

H0: healthy and diabetic maculopathy subjects have <u>same distribution</u> of contrast sensitivity

H1: the distributions are <u>different</u>

We shall use Mann-Whitney U test because of independent continuos data, and because data are skewed.

Example: Contrast Sensitivity and Mann-Whitney U test

Normals	DM	Normals	DM
CS	CS	Ranks	Ranks
38	24	7	1
39	25	8	2.5
40	25	9	2.5
42	27	12	4
42	35	12	5
42	37	12	6
	41		10

Total T1=	60
Total T2=	31
n1=	6
n2=	7
U1=	3
U2=	39

Test statistic U = smaller of U1 and U2 = 3

From stats tables the quantile (critical value) is = 6

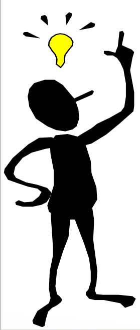
Example: Contrast Sensitivity and Mann-Whitney U test

- Remember that tied values are assigned the average of the ranks covered.
- Thus here:

$$U_1 = n_1 n_2 + ½\, n_1(n_1 + 1) - T_1 = 6 \times 7 + ½ \times 6 \times 7 - 60 = 3$$

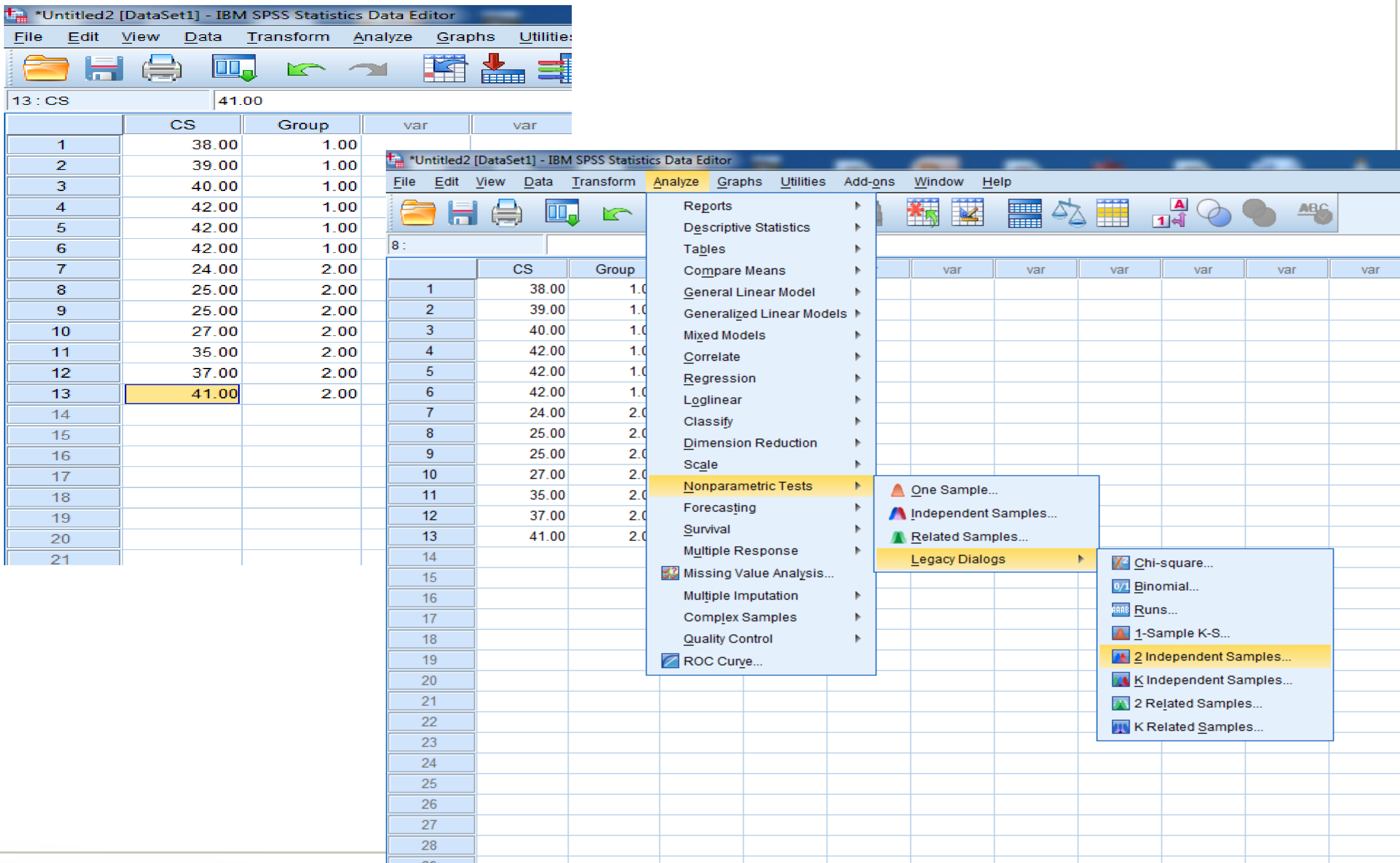
$$U_2 = n_1 n_2 - U_1 = 6 \times 7 - 3 = 39$$

- Test statistic $U = \min(U_1, U_2) = 3$
- From tables, we have quantile (critical value of 6), hence it is seen that the result is significant at the 5% significance level, because $U < 6$.

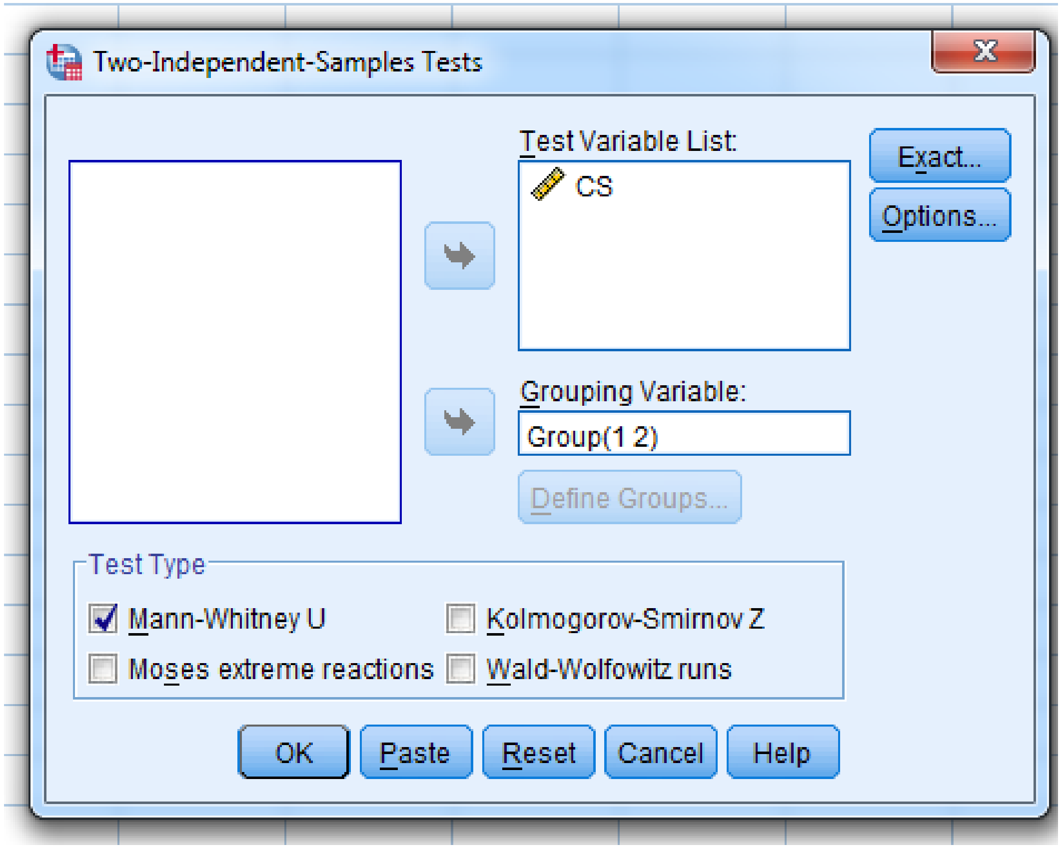


There is strong evidence to suggest that contrast sensitivity differs between the two groups.

Mann-Whitney U test in SPSS software: contrast sensitivity data supplied in one column, another column for groups

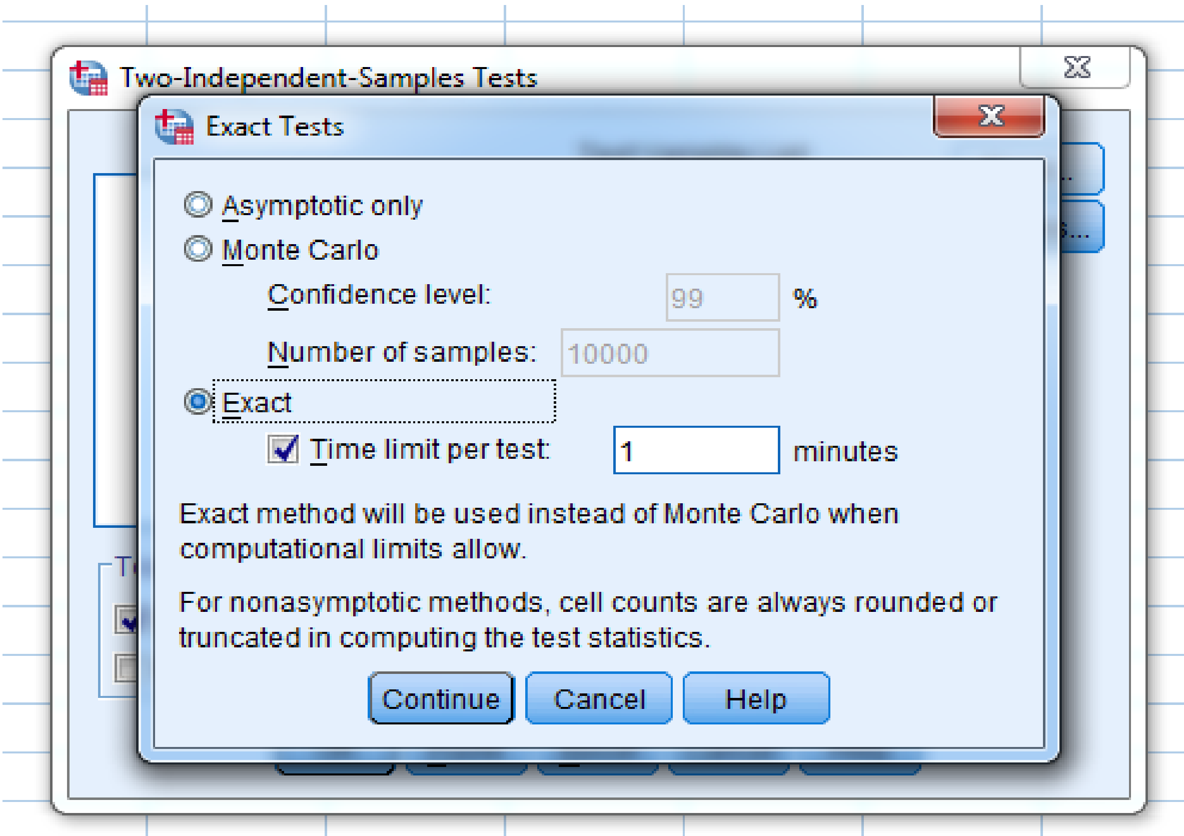


Mann-Whitney U test in SPSS software: specification of variables for analysis



Note: Do not forget to define the group values!

Mann-Whitney test in SPSS software: asking for exact p-values when samples are small (i.e. both n1 and n2 are less than 20)



Note: the Asymptotic option is the default in SPSS, the Asymptotic option uses a simplified formula for calculation of p-value, and this formula works well if large enough data (>20).

Mann-Whitney U test in SPSS software: results

Mann-Whitney Test

Ranks

	Group	N	Mean Rank	Sum of Ranks
CS	Normal	6	10.00	60.00
	DM	7	4.43	31.00
	Total	13		

Test Statistics[a]

	CS
Mann-Whitney U	3.000
Wilcoxon W	31.000
Z	-2.589
Asymp. Sig. (2-tailed)	.010
Exact Sig. [2*(1-tailed Sig.)]	.008[b]
Exact Sig. (2-tailed)	.008
Exact Sig. (1-tailed)	.004
Point Probability	.002

a. Grouping Variable: Group

b. Not corrected for ties.

Conclusion?
Do we reject the null hypothesis of equality of CS distributions?

Mann-Whitney U test in SPSS software: results

Mann-Whitney Test

Ranks

	Group	N	Mean Rank	Sum of Ranks
CS	Normal	6	10.00	60.00
	DM	7	4.43	31.00
	Total	13		

Test Statistics[a]

	CS
Mann-Whitney U	3.000
Wilcoxon W	31.000
Z	-2.589
Asymp. Sig. (2-tailed)	.010
Exact Sig. [2*(1-tailed Sig.)]	.008[b]
Exact Sig. (2-tailed)	.008
Exact Sig. (1-tailed)	.004
Point Probability	.002

a. Grouping Variable: Group

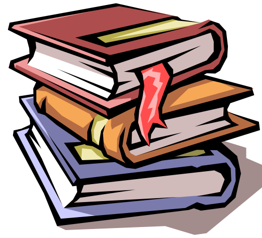
b. Not corrected for ties.

Conclusion?
Do we reject the null hypothesis of equality of CS distributions?
Answer is Yes, b/c pvalue=0.008<0.05..

Summary

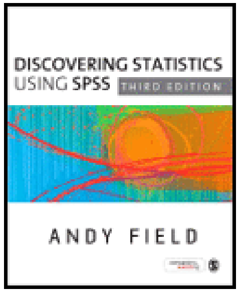
- When to use non-parametric tests for group comparisons?
  - When data are not Normal
  - Or when data too small checked normality
- If we want to compare more than 3 using a continuous independent data then we use Kruskal-Wallis test
  - If the test finds significant difference than we do post-hoc analyses using Wilcoxon test
  - Need to adjust for multiple comparisons.

Resources



Books

- Practical statistics for medical research by Douglas G. Altman
- Medical Statistics from Scratch by David Bowers



Journals' with series on how to do statistics in clinical research

- American Journal of Ophthalmology has **Series on Statistics**
- British Medical Journal has series **Statistics Notes**

Manual for SPSS statistical software - with lots of worked-out examples

- Andy Field, Discovering statistics using SPSS

Workshops organized by Biostatistics Department, U of Liverpool

http://www.liv.ac.uk/translational-medicine/research/departmentsandgroups/biostatistics/coursesandworkshops/

- 3 June 2013 Statistical Process Control
- 4 June 2013 Validity and reliability of diagnostic tests and other methods of measurement
- (Fall) Unbiased design, error prediction and statistical analysis when quantifying biological structures

Thank you for your attention

These slides and worksheet can be found on: http://pcwww.liv.ac.uk/~czanner/

Planned future workshops:

- How to diagnose if a patient is having a disease or how to determine if he is at high risk? Classification methods. (Spt/Oct)
- How to make sense of many measured characteristics? Multivariate stats methods.
- Odd-ratios: What are the odds of developing diseases in smokers as opposed to non-smokers?
- Ideas are welcome!

Statistical Clinics for ophthalmic clinicians and researchers !
Run by appointment.
Email: czanner@liv.ac.uk
Phone: +44-151-706-4019
Further information: http://pcwww.liv.ac.uk/~czanner/