

# Introduction to Applied Statistical Methods

## Practical Session 5 - Solutions

Please download and open the SPSS Data file “Practical 5 data.sav” from KEATS

### *Background:*

The data set consists of a sample of  $n=154$  students attending three postgraduate programmes (affective disorders AF, clinical neuropsychiatry CN, and mental health studies MH), at the IoPPN in 2017. In this dataset you will find the following variables:

- **programme**: the programme the students were attending (1: AF, 2:CN, 3:MH)
- **group**: the teaching group (1: AF and CN, 2: MH)
- **anx**: the scores on the ‘anxiety related to statistics’ scale, which the students completed during the first week of Term 1.
- **catgrade**: the ability category (with respect to statistics) at which each student belonged at the beginning of the term, based on the Prior Knowledge Quiz (1: Low, 2: Sufficient, 3: Good, 4: High).
- **quiz1**: the grades on the practical quiz 1, which the students completed on KEATS
- **quiz2**: the grades on the practical quiz 2, which the students completed on KEATS

### Task 1

First, identify the type of each variable in the dataset.

<b>programme</b> is	<b>a categorical nominal</b>	variable
<b>group</b> is	<b>a categorical nominal (binary)</b>	variable
<b>anx</b> is	<b>a numerical continuous</b>	variable
<b>catgrade</b> is	<b>a categorical ordinal (4 points)</b>	variable
<b>quiz1</b> is	<b>a numerical continuous</b>	variable

quiz2 is	a numerical continuous	variable
----------	------------------------	----------

## Task 2

Fill in the blanks to appropriately describe the data, giving your answers to one decimal place.

In the sample, there are 19 (12.3 %) students from the affective disorders programme, 28 ( 18.2 %) students from the clinical neuropsychiatry, and 107 ( 69.5 %) students from the mental health programme. With respect to their prior knowledge in statistics, 13 ( 8.4 %) students started their training with *low* prior knowledge, 56 ( 36.4 %) students started with *sufficient* prior knowledge, 53 ( 34.4 %) students started with good prior knowledge, 15 (9.7 %) students started with *high* prior knowledge, and 17 ( 11.0 %) students did not take the test.

Regarding the '*anxiety related to statistics*' scale, the students' scores were normally distributed with mean anxiety = 35.2 (sd= 16.4, min= 0, max= 70). The scores for both Progress Quizzes were negatively skewed, with most students scoring towards the upper end of the distribution. Specifically, the median for first Quiz was 75.0 (min= 15.9 - max= 95.5) and the median for second Quiz was 82.5 (min= 12.5 - max= 100.0).

## Task 3

Based on the summary, for which of the variables you would use parametric and for which non-parametric tests?

parametric tests: <b>anxiety scores</b>
non-parametric tests: <b>quiz 1 and quiz 2 scores</b>

## Task 4

Use the appropriate test to see if there are statistically significant differences between the two teaching groups (*AD&CN* and *MH*) in the percentages of students belonging to each of the 'prior knowledge' categories (*low*, *sufficient*, *good*, *high*). Report on the results.

The screenshot displays the SPSS software interface. The 'Analyze' menu is open, and 'Crosstabs...' is selected. The 'Crosstabs' dialog box is open, showing 'Prior Knowledge Quiz results [catgrade]' in the 'Row(s):' field and 'Group [group]' in the 'Column(s):' field. The 'Exact Tests' sub-dialog box is also open, showing the 'Exact' method selected with a confidence level of 99% and a time limit of 5 minutes. The 'Crosstabs: Statistics' sub-dialog box is open, showing the 'Chi-square' test selected, along with other options like 'Nominal', 'Ordinal', and 'Nominal by Interval'.

**Analyze** | Graphs | Utilities | Extensions | Window | Help

Power Analysis  
Reports  
Descriptive Statistics  
Bayesian Statistics  
Tables  
Compare Means  
General Linear Model  
Generalized Linear Models  
Mixed Models  
Correlate  
Regression  
Loglinear  
Neural Networks  
Classify  
Dimension Reduction  
Scale  
Nonparametric Tests  
Forecasting  
Survival  
Multiple Response  
Missing Value Analysis...  
Multiple Imputation  
Complex Samples  
Simulation...  
Quality Control  
Spatial and Temporal Modeling...  
Direct Marketing

**Crosstabs**

Row(s):  
Prior Knowledge Quiz results [catgrade]

Column(s):  
Group [group]

Exact...  
Statistics...  
Cells...  
Format...  
Style...

**Crosstabs: Statistics**

☒ Chi-square  
☐ Correlations

**Nominal**  
☐ Contingency coefficient  
☐ Phi and Cramer's V  
☐ Lambda  
☐ Uncertainty coefficient

**Ordinal**  
☐ Gamma  
☐ Somers' d  
☐ Kendall's tau-b  
☐ Kendall's tau-c

**Nominal by Interval**  
☐ Eta

☐ Kappa  
☐ Risk  
☐ McNemar

☐ Cochran's and Mantel-Haenszel statistics  
Test common odds ratio equals: 1

**Exact Tests**

☐ Asymptotic only  
☐ Monte Carlo  
Confidence level: 99 %  
Number of samples: 10000  
☒ Exact  
☒ Time limit per test: 5 minutes

Exact method will be used instead of Monte Carlo when computational limits allow.

For nonasymptotic methods, cell counts are always rounded or truncated in computing the test statistics.

? Cancel Continue ? Cancel Continue

Output:

Counts

☒ Observed

☐ Expected

☐ Hide small counts

Less than

Percentages

☐ Row

☒ Column

☐ Total

☐ Create APA style table

Noninteger Weights

☒ Round cell counts

☐ Truncate cell counts

☐ No adjustments

☐ Round case weights

☐ Truncate case weights

z-test

☐ Compare column proportions

☐ Adjust p-values (Bonferroni method)

Residuals

☐ Unstandardized

☐ Standardized

☐ Adjusted standardized

?

Cancel

Continue

Prior Knowledge Quiz results * Group Crosstabulation					
			Group		Total
			AD and CN	MH	
Prior Knowledge Quiz	Low	Count	4	9	13

results		% within Group	10.80%	9.00%	9.50%
	Sufficient	Count	11	45	56
		% within Group	29.70%	45.00%	40.90%
	Good	Count	18	35	53
		% within Group	48.60%	35.00%	38.70%
	High	Count	4	11	15
		% within Group	10.80%	11.00%	10.90%
	Total	Count	37	100	137
		% within Group	100.00 %	100.00 %	100.00 %

Chi-Square Tests						
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	2.935 <sup>a</sup>	3	0.402	0.412		
Likelihood Ratio	2.978	3	0.395	0.421		
<b>Fisher-Freeman-Halton Exact Test</b>	<b>3.069</b>			<b>0.393</b>		
Linear-by-Linear Association	.535 <sup>b</sup>	1	0.465	0.481	0.27	0.072
N of Valid Cases	137					
<b>a 2 cells (25.0%) have expected count less than 5. The minimum expected count is 3.51.</b>						
b The standardized statistic is -.731.						

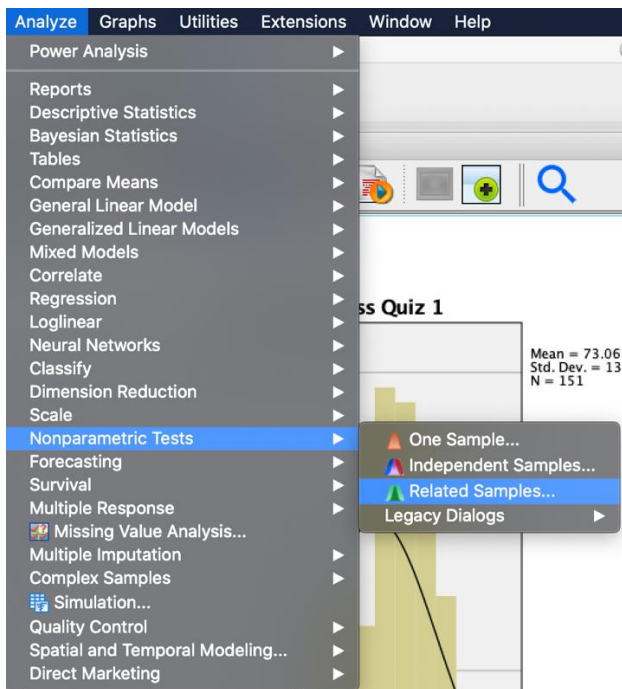
**For two independent groups, the appropriate test is Pearson's chi square, provided the assumptions of the test hold. However, the assumptions did not hold, so we report instead on the Fisher's exact test.**

**The was no association between the group membership and the level of prior knowledge in statistics (Fisher's exact test p=0.393).**

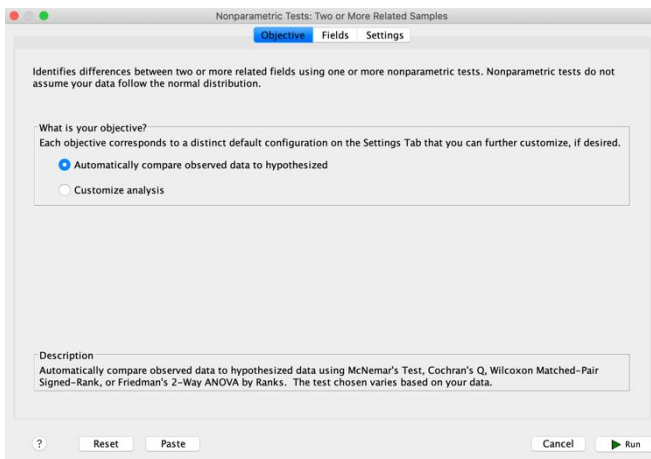
## Task 5

Use the appropriate test to see if there are statistically significant differences between the student's results on Progress Quiz 1 and Progress Quiz 2. Was there any improvement?

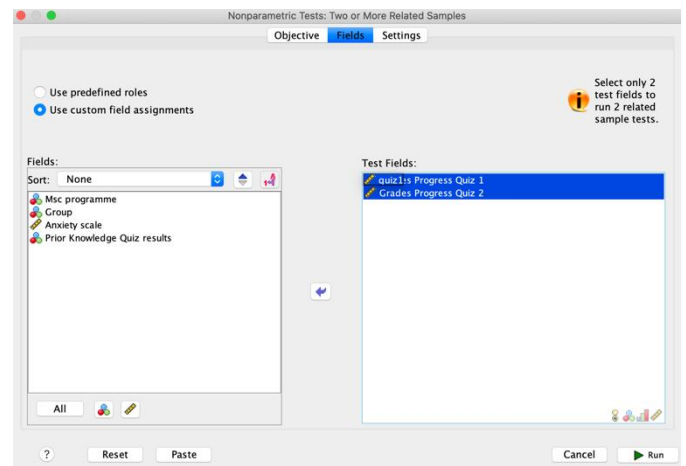
We can see from the histogram (not shown) of the difference between the two variables of interest (Progress Quiz 1 and Progress Quiz 2) that the distribution of the difference has a skewed distribution. Therefore, we use the Wilcoxon signed rank test instead of a repeated measures t-test.



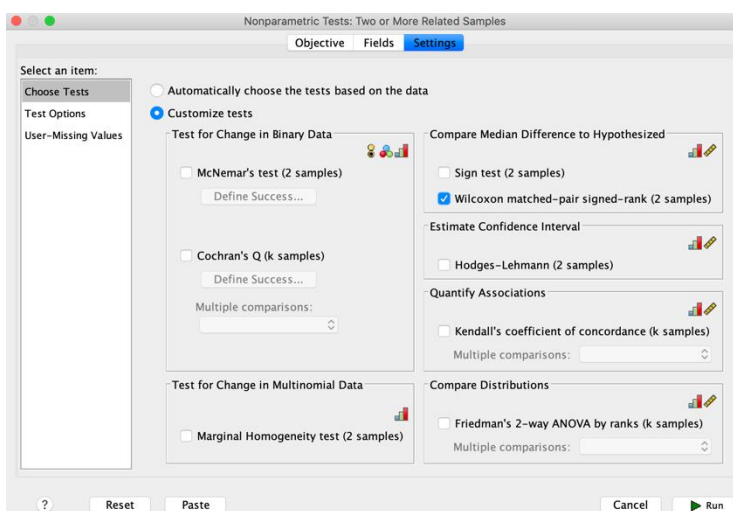
1.



2.



3.



Output:

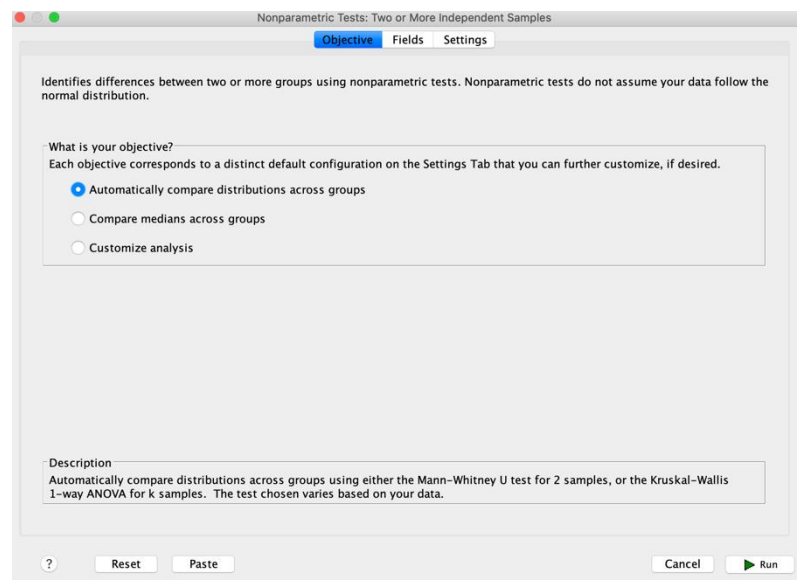
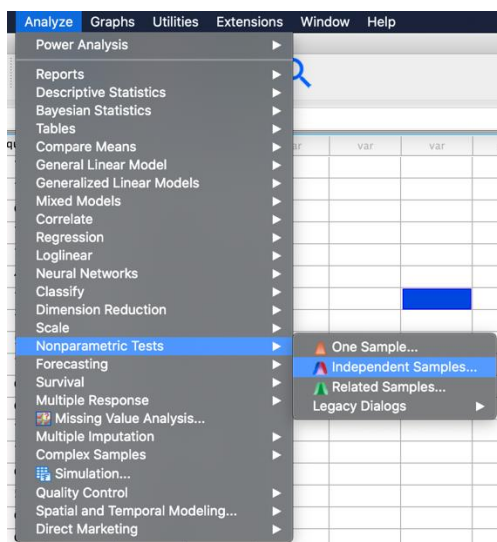
Hypothesis Test Summary				
	Null Hypothesis	Test	Sig. <sup>a,b</sup>	Decision
1	The median of differences between Grades Progress Quiz 1 and Grades Progress Quiz 2 equals 0.	Related-Samples Wilcoxon Signed Rank Test	0	Reject the null hypothesis.
a The significance level is .050.				
b Asymptotic significance is displayed.				

Related-Samples Wilcoxon Signed Rank Test Summary	
Total N	122
Test Statistic	5700
Standard Error	391.37
Standardized Test Statistic	4.979
Asymptotic Sig.(2-sided test)	0

**The students scored higher in Quiz 2 than Quiz 1 (Quiz 1 median= 75.0; Quiz 2 median=82.5). This was a significant difference, according to the Wilcoxon signed rank test (Z=4.979, p<0.001).**

## Task 6

Use the appropriate test to see there were statistically significant differences between the two teaching groups with respect to their performance in the progress quizzes.





Nonparametric Tests: Two or More Independent Samples

Objective Fields Settings

☐ Use predefined roles  
☒ Use custom field assignments

Fields:

Sort: None

- Msc programme
- Anxiety scale
- Prior Knowledge Quiz results

Test Fields:

- Grades Progress Quiz 1
- Grades Progress Quiz 2

Groups:

- Group

All

? Reset Paste Cancel Run

Nonparametric Tests: Two or More Independent Samples

Objective Fields Settings

Select an item:

- Choose Tests
- Test Options
- User-Missing Values

☐ Automatically choose the tests based on the data  
☒ Customize tests

Compare Distributions across Groups

☒ Mann-Whitney U (2 samples)  
☐ Kolmogorov-Smirnov (2 samples)  
☐ Test sequence for randomness (Wald-Wolfowitz for 2 samples)  
☐ Kruskal-Wallis 1-way ANOVA (k samples)  
 Multiple comparisons:   
☐ Test for ordered alternatives (Jonckheere-Terpstra for k samples)  
 Hypothesis order:   
 Multiple comparisons:

Compare Ranges across Groups

☐ Moses extreme reaction (2 samples)  
☒ Compute outliers from sample  
☐ Custom number of outliers  
 Outliers: 1

Compare Medians across Groups

☐ Median test (k samples)  
☒ Pooled sample median  
☐ Custom  
 Median: 0  
 Multiple comparisons:

Estimate Confidence Interval across Groups

☐ Hodges-Lehmann estimate (2 samples)

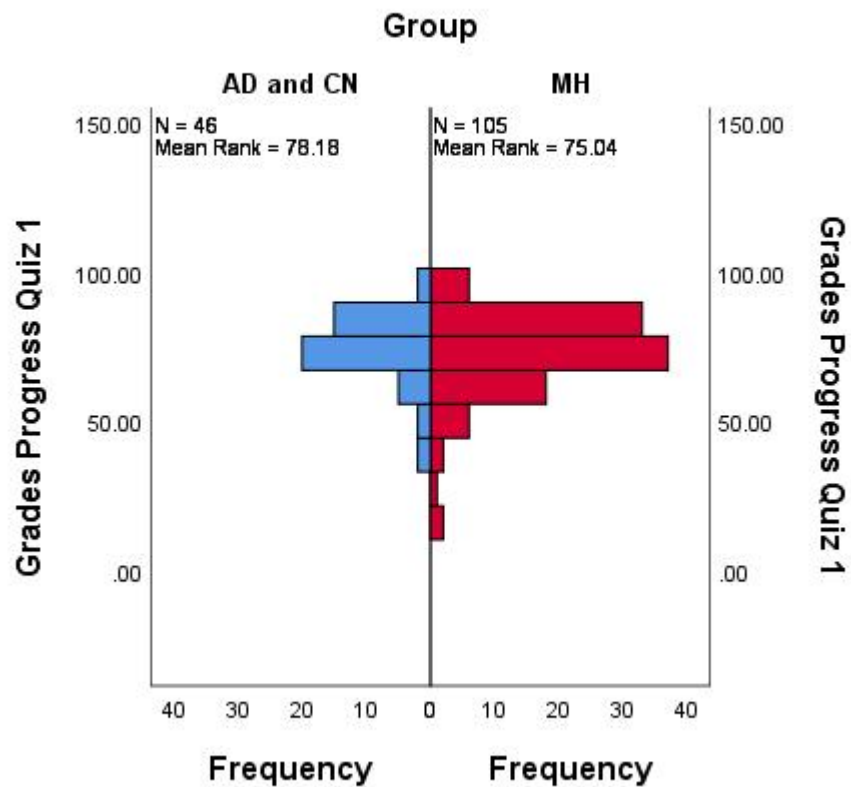
? Reset Paste Cancel Run

Output:

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig. <sup>a,b</sup>	Decision
1	The distribution of Grades Progress Quiz 1 is the same across categories of Group.	Independent-Samples Mann-Whitney U Test	0.684	Retain the null hypothesis.
2	The distribution of Grades Progress Quiz 2 is the same across categories of Group.	Independent-Samples Mann-Whitney U Test	0.989	Retain the null hypothesis.
a The significance level is .050.				
b Asymptotic significance is displayed.				

Grades Progress Quiz 1 across Group

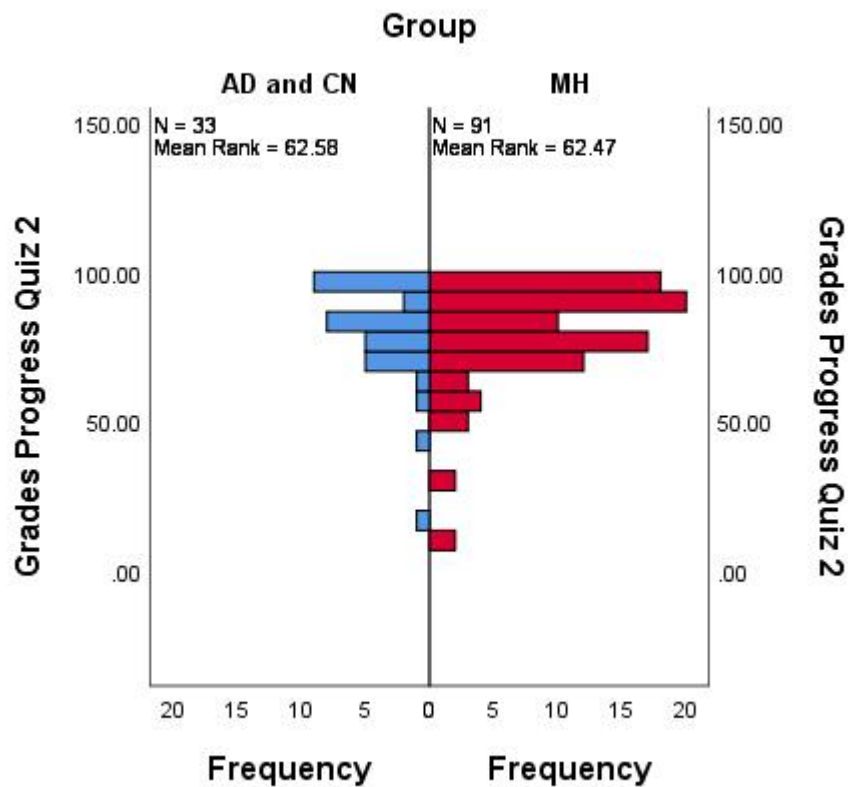
## Independent-Samples Mann-Whitney U Test



Independent-Samples Mann-Whitney U Test Summary	
Total N	151
Mann-Whitney U	2314.5
Wilcoxon W	7879.5
Test Statistic	2314.5
Standard Error	246.733
Standardized Test Statistic	-0.407
Asymptotic Sig.(2-sided test)	0.684

Grades Progress Quiz 2 across Group

## Independent-Samples Mann-Whitney U Test



Independent-Samples Mann-Whitney U Test Summary	
Total N	124
Mann-Whitney U	1499
Wilcoxon W	5685
Test Statistic	1499
Standard Error	176.441
Standardized Test Statistic	-0.014
Asymptotic Sig.(2-sided test)	0.989

**There were no statistically significant differences in progress quiz scores between the two groups, according to the Mann-Whitney U test (Quiz 1:  $U=2314.5$ ,  $p=0.684$ ; Quiz 2:  $U=1499.0$ ,  $p=0.989$ ).**