

# **Non-parametric Procedures**

# What are Non-parametric Statistics?

---

Methods of analyzing data that examine the relative position or rank of the data rather than the actual values.

Non-parametric statistics do not:

- assume that the data come from a normal distribution.
- create any parameter estimates (e.g., means; standard deviations) to assess whether one set of numbers is statistically different from another set of numbers.
- You can use median scores and ranges for descriptives

# Which test to use?

How many sets of scores?			
Two		More than two	
Within or between subjects?		Within or between subjects?	
Within	Between	Within	Between
<b>Wilcoxon Signed- Rank</b>	<b>Mann- Whitney U</b>	<b>Friedman ANOVA</b>	<b>Kruskal- Wallis H</b>

# Definitions

- Non-parametric Statistics
  - An inferential statistic which requires no assumptions about the shape of the population distribution. These methods of analyzing data examine the relative position or rank of the data rather than the actual values.
- Mann-Whitney U Test
  - Non-parametric equivalent of the independent groups  $t$  test when group sizes are too small or unequal to insure robustness against violation of the parametric assumptions required by the  $t$  test or when the study involves a discrete ordinal variable. Under certain conditions, it will fail to detect the presence of a relationship that the parametric alternative can detect.
- The Wilcoxon Signed-Rank test
  - Non-parametric equivalent of the dependent groups  $t$  test when the group sizes are too small or unequal to insure robustness against violation of the parametric assumptions required by the  $t$  test or when the study involves a discrete ordinal variable. Under certain conditions, it will fail to detect the presence of a relationship that the parametric alternative can detect.

- Kruskal-Wallis  $H$  Test
  - Non-parametric equivalent of one-way independent groups ANOVA when the group sizes are too small or unequal to insure robustness against violation of the parametric assumptions required by the  $F$  test, when populations are believed to be severely non-normal, or when the study involves a discrete ordinal variable. Under certain conditions, it will fail to detect the presence of a relationship that the parametric alternative can detect.
- Friedman's ANOVA
  - Non-parametric equivalent of the one-way dependent groups ANOVA test when group sizes are too small or unequal to insure robustness against violation of the parametric assumptions required by the  $F$  test, when populations are believed to be severely non-normal, or when the study involves a discrete ordinal variable. Under certain conditions, it will fail to detect the presence of a relationship that the parametric alternative can detect.

# The Mann-Whitney $U$ Test:

Non-parametric equivalent of the independent  $t$  test

Tests whether two independent samples are from the same population. It uses rank ordering of data.

Assumptions: Random and independent sampling

The P value answers this question:

If the groups are sampled from populations with identical distributions, what is the chance that random sampling would result in **the mean ranks** being as far apart (or more so) as observed in this experiment?

# Steps in the Analysis

1. Combine the data from the two groups.
2. Rank order the data from lowest to highest.
3. The lowest score is replaced with a rank of 1, the next lowest score with a 2, and so on. Replace tied scores with the mean of their positions in the list.
4. Compute the  $U$  value for each group, where:

$$U = n_1n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

R: sum of ranks

## Example: Stereotyping (stereotyping.sav)

Children with working/non-working mothers were asked to interpret stories and were given a gender stereotype score based on their answers. 100 = extreme stereotyping, 0 = no stereotyping

Non-parametric test was used because scoring was subjective

Mother has full-time job	Mother has no job outside home
17	19
32	63
39	78
27	29
58	39
25	59
31	77
	81
	68



# Sums of ranks

Mother has full-time job	Ranks	Mother has no job outside home	Ranks
17	1	19	2
32	7	63	12
39	8.5	78	15
27	4	29	5
58	10	39	8.5
25	3	59	11
31	6	77	14
		81	16
		68	13
	39.5		96.5

Shortcut: take each score and count the number of scores that are higher in the other group. Give 0.5 points for the same score in the other group. Add up the points to get U.

	Mother has full-time job		Mother has no job outside home	
	Score	Points	Score	Points
	17	9	19	6
	32	7	63	0
	39	6.5	78	0
	27	8	30	4
	58	6	39	1.5
	25	8	59	0
	30	7	77	0
			81	0
			68	
<b>Total</b>		<b>51.5</b>		<b>11.5</b>

# The p value for Mann Whitney U

- When the smaller sample has 100 or fewer values, software computes the exact P value, even with ties.
  - It tabulates every possible way to shuffle the data into two groups of the sample size actually used, and computes the fraction of those shuffled data sets where the difference between mean ranks was as large or larger than actually observed.
- When the samples are large (the smaller group has more than 100 values), software uses the approximate method
  - converts U or sum-of-ranks to a Z value, and then looks up that value on a Gaussian distribution to get a P value.

- Note: the two medians may be the same
- but there still may be a significant difference between the groups if the probability of their **mean ranks** being what they are is very low if they come from the same population

## Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
<b>1</b>	The distribution of Stereotyping is the same across categories of Mother.	Independent-Samples Mann-Whitney U Test	.031 <sup>1</sup>	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

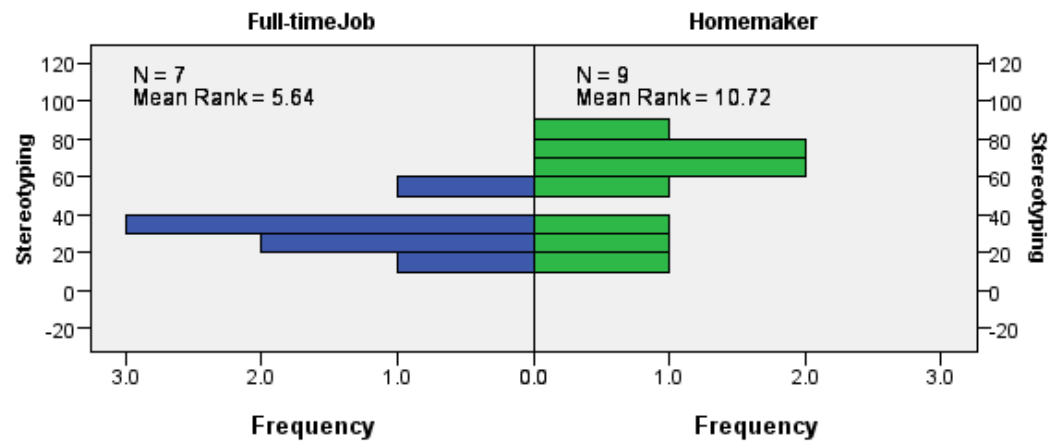
<sup>1</sup>Exact significance is displayed for this test.

Output via Nonparametric Tests > Independent Samples

Double click to get details

## Independent-Samples Mann-Whitney U Test

Mother



Total N	16
Mann-Whitney U	51.500
Wilcoxon W	96.500
Test Statistic	51.500
Standard Error	9.440
Standardized Test Statistic	2.119
Asymptotic Sig. (2-sided test)	.034
Exact Sig. (2-sided test)	.031

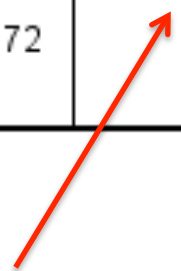
This is U →

This is z →

## Output via Legacy dialogs

### Ranks

	Mother	N	Mean Rank	Sum of Ranks
Stereotyping	Full-timeJob	7	5.64	39.50
	Homemaker	9	10.72	96.50
	Total	16		



### Test Statistics<sup>a</sup>

	Stereotyping
Mann-Whitney U	11.500
Wilcoxon W	39.500
Z	-2.119
Asymp. Sig. (2-tailed)	.034
Exact Sig. [2*(1-tailed Sig.)]	.031 <sup>b</sup>

a. Grouping Variable: Mother

b. Not corrected for ties.

To obtain Ranks:

1. Arrange all data in increasing order
2. Assign 1 to the lowest value, 2 to the second lowest value, etc.
3. For the Sum of Ranks, add up the rank separately for the two groups

# Computing Effect Size

$$r = \frac{Z}{\sqrt{N}}$$

Note:  $Z$  = z-score from SPSS output.

$$r = \frac{2.119}{\sqrt{16}} = \frac{2.119}{4} = .53$$



# Reporting the Results

The results showed that children whose mothers work full-time (N = 7) are less likely to show stereotypical behaviour, with the mean rank of this group being 5.57, while the children whose mothers do not work outside the home (N = 9) had a mean rank of 10.78. A Mann-Whitney test revealed that this difference was statistically significant,  $U = 11.5$ ,  $p = .034$ ,  $r = .53$ .

# The Kruskal-Wallis $H$ Test

This is a nonparametric equivalent to one-way independent-groups ANOVA. Tests whether several independent samples are from the same population.

Assumptions: Random and independent sampling

# Steps in the Analysis

1. Combine the data from the groups.
2. Rank order the data from lowest to highest.
3. The lowest score is replaced with a rank of 1, the next lowest score with a 2, and so on.
4. Replace tied scores with the mean of their positions in the list.
5. Compute the  $H$  value for the sample.

Degrees of freedom:  $k - 1$

$$H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1)$$

# Follow-up analysis

- Pairwise comparisons
  - Multiple Mann-Whitney tests with correction for multiple comparisons
- Planned contrasts
  - Stepwise comparisons starting with group with lowest sum of ranks
    - Compares 1 to 2
    - If significant, separates 1 and compares 2 to 3, etc.
    - If not significant, keep them together and adds next condition
  - Trend analysis (Jonckheere-Terpstra J)
    - Compares each condition to the next

# Effect size

- Run pairwise comparisons for z scores
- Calculate r for each pair

$$r = \frac{z}{\sqrt{N}}$$

N: the number of people in the two groups compared

# An example: Coffee and Driving (coffeeDriving.sav)

- Does coffee improve driving performance?
- Three groups of participants:
  - Coffee
  - Decaff
  - Water
- Simulated driving performance (higher score = better driving)
- Non-parametric test was used because of subjective scoring (considered ordinal rather than scale)

# SPSS output via Nonparametric tests > Independent Samples

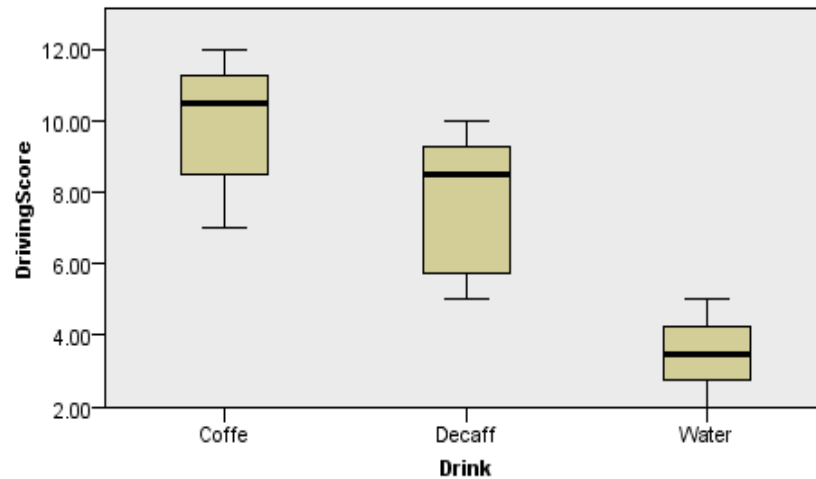
## Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
<b>1</b>	The distribution of DrivingScore is the same across categories of Drink.	Independent-Samples Kruskal-Wallis Test	.002	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Double click for details

### Independent-Samples Kruskal-Wallis Test



<b>Total N</b>	18
<b>Test Statistic</b>	12.788
<b>Degrees of Freedom</b>	2
<b>Asymptotic Sig. (2-sided test)</b>	.002

1. The test statistic is adjusted for ties.

Independent samples test view

### Homogeneous Subsets based on DrivingScore

		Subset	
		1	2
Sample <sup>1</sup>	Water	3.583	
	Decaff		10.500
	Coffe		14.417
<b>Test Statistic</b>		. <sup>2</sup>	3.464
<b>Sig. (2-sided test)</b>			.063
<b>Adjusted Sig. (2-sided test)</b>			.063

Homogeneous subsets are based on asymptotic significances. The significance level is .05.

<sup>1</sup>Each cell shows the sample average rank of DrivingScore.

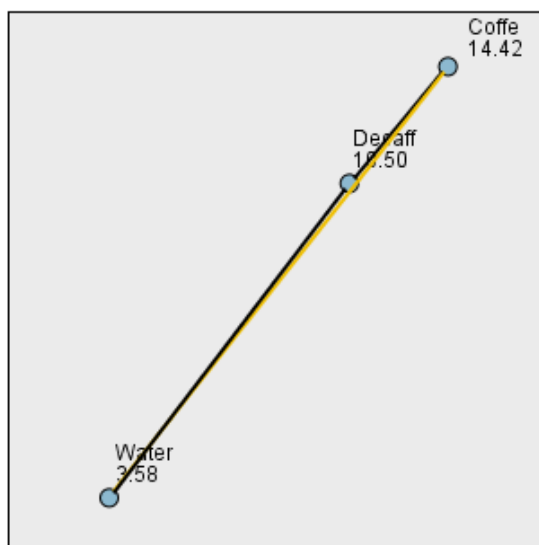
<sup>2</sup>Unable to compute because the subset contains only one sample.

Homogeneous subsets view  
The result of stepwise comparisons



# Output of pairwise comparisons

Pairwise Comparisons of Drink



Each node shows the sample average rank of Drink.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Water-Decaff	6.917	3.068	2.255	.024	.072
Water-Coffe	10.833	3.068	3.531	.000	.001
Decaff-Coffe	3.917	3.068	1.277	.202	.605

Z scores

p value adjusted for multiple comparisons

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

# The Kruskal-Wallis Test

Ranks			
	Drink	N	Mean Rank
DrivingScore	Coffe	6	14.42
	Decaff	6	10.50
	Water	6	3.58
	Total	18	

Test Statistics <sup>a,b</sup>	
	DrivingScore
Chi-Square	12.788
df	2
Asymp. Sig.	.002
a. Kruskal Wallis Test	
b. Grouping Variable: Drink	

Output from Legacy Dialogs

H = Chi square

# Reporting the Results

The results showed that the subjects' driving performance was significantly affected by the type of drink given to them,  $H(2) = 12.79$ ,  $p = .002$ . Step-down follow-up analysis showed that people who received either coffee (M Rank = 14.42) or a decaffeinated drink (M Rank = 10.5) drove significantly better than the group who drank water (M Rank = 3.58). There was no significant difference between the coffee and the decaff conditions.

## **The Wilcoxon signed rank $T$ Test: (sometimes $W$ is used for $T$ )**

Non-parametric equivalent of the dependent  $t$  test

Tests whether two dependent samples are from the same population. It uses rank ordering of data that are at least ordinal level.

Assumptions: Random and independent sampling

Note: Use same formula as for the Mann-Whitney Test to compute effect size.

Example:  
discomfortLight.sav

People's  
discomfort in  
a brightly lit  
room and in a  
dark room  
was estimated  
on a scale of 1  
to 50.

Bright room	Dark room
23	33
14	22
35	38
26	30
28	31
19	17
42	42
30	25
26	34
31	24
18	21
25	46
23	29
31	40
30	41

# Steps in the Analysis

1. Compute the difference scores for each associated pair of values.
2. Rank order the absolute value of these differences scores from lowest to highest ignoring 0 differences.  
(Equal differences get the average of their ranks)
3. Sum the ranks associated with the positive difference scores, then sum the ranks associated with the negative difference scores:

The computed  $T$  value is the smaller sum.

## Wilcoxon *signed-rank* Test: Computing by hand

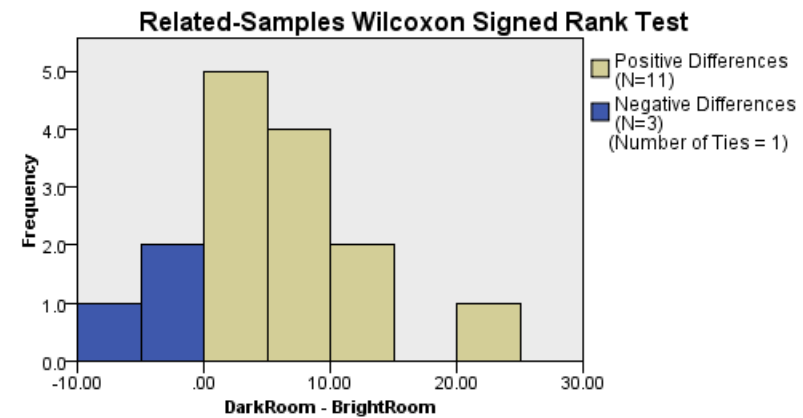
Bright	Dark	Difference	Rank of difference
23	33	10	12
14	22	8	9.5
35	38	3	3
26	30	4	5
28	31	3	3
19	17	-2	1
42	42	0	
30	25	-5	6
26	34	8	9.5
31	24	-7	8
18	21	3	3
25	46	21	14
23	29	6	7
31	40	9	11
30	41	11	13

$$T = 1 + 6 + 8 = 15$$

## Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between BrightRoom and DarkRoom equals 0.	Related-Samples Wilcoxon Signed Rank Test	.018	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.



This is T →

This is z →

Total N	15
Test Statistic	90.000
Standard Error	15.910
Standardized Test Statistic	2.357
Asymptotic Sig. (2-sided test)	.018



### Ranks

	N	Mean Rank	Sum of Ranks
DarkRoom - BrightRoom			
Negative Ranks	3 <sup>a</sup>	5.00	15.00
Positive Ranks	11 <sup>b</sup>	8.18	90.00
Ties	1 <sup>c</sup>		
Total	15		

a. DarkRoom < BrightRoom

b. DarkRoom > BrightRoom

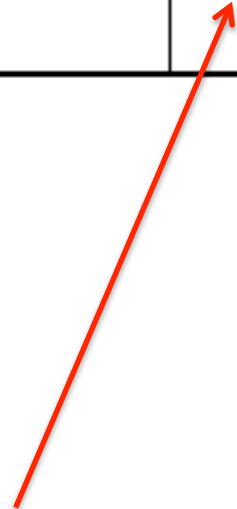
c. DarkRoom = BrightRoom

### Test Statistics<sup>a</sup>

	DarkRoom - BrightRoom
Z	-2.357 <sup>b</sup>
Asymp. Sig. (2-tailed)	.018

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.



The sum of positive Ranks is T

# Computing Effect Size

$$r = \frac{Z}{\sqrt{N}}$$

Note:  $Z$  = z-score from SPSS output.

$$r = \frac{2.357}{\sqrt{15}} = \frac{2.357}{3.87} = .61$$

# Reporting the Results

A Wilcoxon Signed Ranks test revealed that people experienced significantly more discomfort in a dark room (Mdn = 31) than in a brightly lit room (Mdn = 26),  $T = 90$ ,  $p = .018$ ,  $r = .61$ .

# Friedman's ANOVA ( $\chi_F^2$ )

The non-parametric equivalent of one-way repeated measures ANOVA and so is used for testing differences between experimental conditions when there are more than two conditions and the same participants have been used in all conditions.

Follow-up analysis

# Calculating

- Take one person at a time
- Rank that person's scores in the different conditions
- Take the next person and rank his or her scores, etc.
- Add up the ranks for each condition
- Calculate  $X_F^2$

$$X_F^2 = \left[ \frac{12}{Nk(k-1)} \sum R_i^2 \right] - 3N(k+1)$$

# **Example: Creativity and reward**

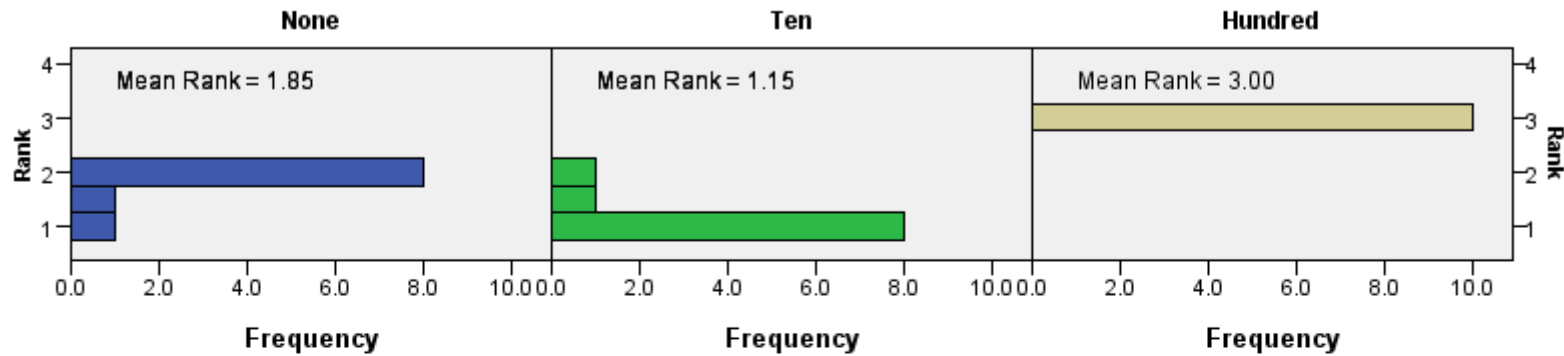
**(creativityReward.sav)**

A researcher conducts a study to test the effects of a reward on creativity. Participants are offered nothing, 10 dollars and 100 dollars as a reward if they complete a task. The task is creating a collage from a set of shapes. The creativity of the result is judged by artists on a scale of 1 to 15.

### Hypothesis Test Summary

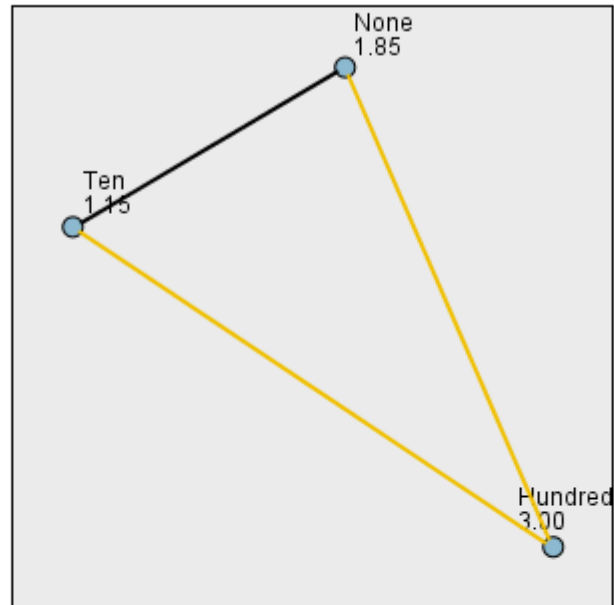
	Null Hypothesis	Test	Sig.	Decision
1	The distributions of None, Ten and Hundred are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	.000	Reject the null hypothesis.

### Related-Samples Friedman's Two-Way Analysis of Variance by Ranks



Total N	10
Test Statistic	17.897
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.000

### Pairwise Comparisons



Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Ten-None	.700	.447	1.565	.118	.353
Ten-Hundred	-1.850	.447	-4.137	.000	.000
None-Hundred	-1.150	.447	-2.571	.010	.030

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05.



# Friedman Test

Ranks	
	Mean Rank
None	1.85
Ten	1.15
Hundred	3.00

Test Statistics <sup>a</sup>	
N	10
Chi-Square	17.897
df	2
Asymp. Sig.	.000
a. Friedman Test	

# Reporting the Results

The results of a Friedman's ANOVA showed that rewards significantly affected creativity,  $X_F^2(2) = 17.89, p < .001$ . Post-hoc pairwise comparisons revealed that people who were given a large reward (M Rank = 3) were significantly more creative than those who were either given no reward (M Rank = 1.85) ( $p = .03$ ) or a small reward (M Rank = 1.15) ( $p < .001$ ).

# Homework

- DownloadFestival.sav
  - Do hygiene standards decline at the Sziget festival?
  - Participants at a three-day music festival
  - Hygiene standards measured: higher score = higher standardDon't forget to set missing values to listwise!
- Coulrophobia.sav
  - Are advertisements such as McDonald's clown advert harmful?
  - Four groups of children:
    - Watched a scary advert with a clown
    - Listened to a story about a nice clown
    - Met a nice clown in person
    - No exposure to clowns
  - The children's fear of clowns was taken as dependent variable (higher score = more fear)