THE KRUSKAL–WALLLIS TEST

TEODORA H. MEHOTCHEVA

Wednesday, 23rd April 08

Seminar in Methodology & Statistics

THE KRUSKAL-WALLIS TEST:

The non-parametric alternative to ANOVA:

testing for difference between several independent groups

NON PARAMETRIC TESTS: CHARACTERISTICS

Distribution-free tests?

⇒ Not exactly, they just less restrictive than parametric tests

⇒ Based on ranked data

- ⇒ By ranking the data we lose some information about the magnitude of difference between scores
  - ⇒ the non-parametric tests are less powerful than their parametric counterparts, i.e. a parametric test is more likely to detect a genuine effect in the data , if there is one, than a non-parametric test.

WHEN TO USE KRUSKAL-WALLIS

We want to compare several independent groups but we don’t meet some of the assumptions made in ANOVA:

$\Rightarrow$ Data should come from a normal distribution

$\Rightarrow$ Variances should be fairly similar (Levene’s test)

EXAMPLE:
EFFECT OF WEED ON CROP

4 groups: 0 weeds/meter
1 weed/meter
3 weeds/meter
9 weeds/meter

4 samples x group (*N*16)

Weeds	Corn	Weeds	Corn	Weeds	Corn	Weeds	Corn
0	166.7	1	166.2	3	158.6	9	162.8
0	172.2	1	157.3	3	176.4	9	142.4
0	165.0	1	166.7	3	153.1	9	162.7
0	176.9	1	161.1	3	165.0	9	162.4

Corn crop by weeds (Ex. 15.13, Moore & McCabe, 2005)

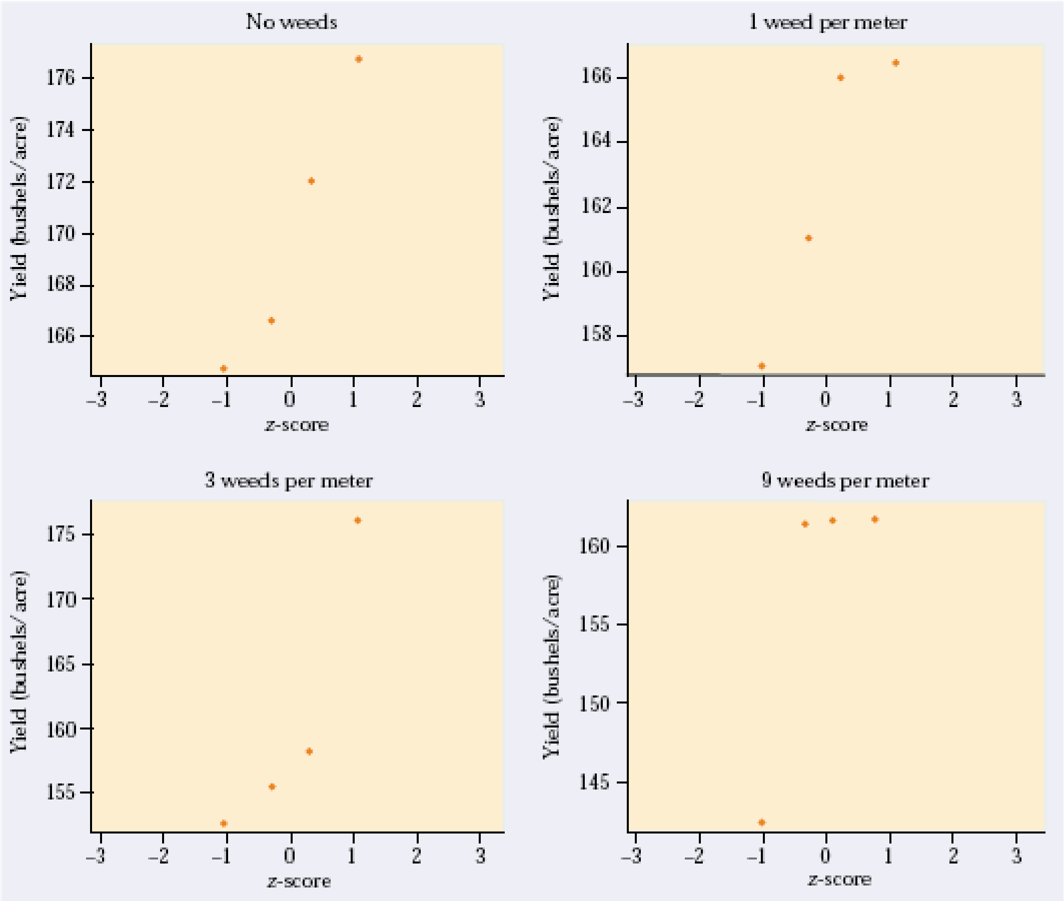
EFFECT OF WEED ON CROP: EXPLORING THE DATA

Weeds	n	Mean	Std.dev.
0	4	170.200	5.422
1	4	162.825	4.469
3	4	161.025	10.493
9	4	157.575	10.118

Summary statistics for Effect of Weed on Crop

! For ANOVA: the largest standard deviation should NOT exceed twice the smallest.

EFFECT OF WEED ON CROP: EXPLORING THE DATA: Q-Q PLOTS



Ex. 15.9, Moore & McCabe, 2005

KRUSKAL-WALLIS: HYPOTHESISING

H_0: *All four populations have the same* ***median*** *yield.*

H_a: *Not all four* ***median*** *yields are equal.*

! ANOVA *F:*

H_0: $\mu 0 = \mu 1 = \mu 3 = \mu 9$

H_a: *not all four means are equal.*

KRUSKAL-WALLIS: HYPOTHESISING

⇒ Non-parametric tests hypothesize about the ***median*** instead of the ***mean*** (as parametric tests do).

mean – a hypothetical value not necessarily present in the data ($\mu = \Sigma x_i / n$)

median – the middle score of a set of ordered observations. In the case of even number of observations, the median is the average of the two scores on each side of what should be in the middle

The ***mean*** is more sensitive to outliers than the ***median.***

Ex: 1, 5, 2, 8, 38

μ = 10,7 [(1+5+2+8+38)/5]

median = 5 (1, 2, 5, 8, 38)

! Median when the observations are even:

Ex: 5, 2, 8, 38 = 6,5

(2, 5, 8, 38)

↳ (5+8)/2 = 6,5

THE KRUSKAL-WALLIS TEST: THE THEORY

⇒ We take the responses from all groups and rank them; then we sum up the ranks for each group and we apply one way ANOVA to the ranks, not to the original observations.

⇒ We order the scores that we have from lowest to highest, ignoring the group that the scores come from, and then we assign the lowest score a rank of 1, the next highest a rank of 2 and so on.

EFFECTS OF WEED ON CROP: KRUSKAL-WALLIS TEST: RANKING THE DATA

Score	142,4	157,3	158,6	161,1	…	166,2	166,7	166,7	172,2	…
Rank	1	2	3	4	…	11	12	13	14	…
Act. Rank	1	2	3	4	…	11	12,5	12,5	14	…
Group	9	3	3	1	…	1	0	1	0	…

! Repeated values (tied ranks) are ranked as the average of the potential ranks for those scores, i.e.

(12+13)/2=12,5

EFFECTS OF WEED ON CROP: KRUSKAL-WALLIS TEST: RANKS

When the data are ranked we collect the scores back in their groups and add up the ranks for each group = R_i (i determines the particular group)

Weeds	Ranks				Sum of ranks
0	10	12,5	14	16	52,5
1	4	6	11	12,5	33,5
3	2	3	5	15	25,0
9	1	7	8	9	25,0

Ex. 15.14, Moore & McCabe, 2005

THE KRUSKAL-WALLIS TEST: THE THEORY

- In ANOVA , we calculate the total variation (total sum of squares, SST) by adding up the variation among the groups (sum of squares for groups, SSG) with the variation within group (sum of squares for error, SSE):

$$SST=SSG+SSE$$

In Kruskal-Wallis: one way ANOVA to the ***ranks***, not the original scores.

If there are *N* observations in all, the ranks are always the whole numbers from 1 to *N*. The total sum of squares for the ranks is therefore a fixed number no matter what the data are $\Rightarrow$ no need to look at both SSG and SSE $\Rightarrow$

Kruskal-Wallis = SSG for the ranks

KRUSKAL-WALLIS TEST: *H* STATISTIC

The test statistic *H* is calculated:

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

$\Rightarrow$ The Kruskal-Wallis test rejects the H_o when H is large.

EFFECTS OF WEED ON CROP: KRUSKAL-WALLIS TEST: *H* STATISTIC

Our Example: $I = 4$, $N = 16$, $n_i=4$, $R = 52.5, 33.5, 25.0, 25.0$

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

$$= \frac{12}{(16)(17)}\left(\frac{52.5^2}{4} + \frac{33.5^2}{4} + \frac{25^2}{4} + \frac{25^2}{4}\right) - 3(17)$$

$$= \frac{12}{272}\,(689.0625 + 280.5625 + 156.25 + 156.25) - 51$$

$$= 0.0441(1282{,}125) - 51$$

$$= 56.5643 - 51$$

$$= 5.56$$

EFFECTS OF WEED ON CROP: KRUSKAL-WALLIS TEST: *P VALUE*

- ⇒ H has approximately the chi-square distribution with $k - 1$ degrees of freedom
  - ⇒ df = 3 (4-1) & $H = 5.56$
  - ⇒ $0.10 < P > 0.15$

THE STUDY: THE EFFECT OF SOYA ON CONCENTRATION

- ⇒ 4 groups:
  - O Soya Meals per week
  - 1 Soya Meal per week
  - 4 Soya Meals per week
  - 9 Soya Meals per week
- ⇒ 20 participants per group ⇒ N=80
- ⇒ Tested after one year: RT when naming words

THE EFFECT OF SOYA ON CONCENTRATION: EXPLORATORY STATISTICS

Soya	n	Mean	Std.dev.
0	20	4.9868	5.08437
1	20	4.6052	4.67263
4	20	4.1101	4.40991
9	20	1.6530	1.10865

Summary Statistics for Soya on Concentration

! Violation of the rule of thumb for using ANOVA:
the largest standard deviation should NOT exceed twice the smallest.

THE EFFECT OF SOYA ON CONCENTRATION: TEST OF NORMALITY

Tests of Normality

	Number of Soya Meals Per Week	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
RT (Ms)	No Soya Meals	,181	20	,085	,805	20	,001
	1 Soya Meal Per Week	,207	20	,024	,826	20	,002
	4 Soyal Meals Per Week	,267	20	,001	,743	20	,000
	7 Soya Meals Per Week	,204	20	,028	,912	20	,071

a. Lilliefors Significance Correction

Significance of data ⇒ the distribution is significantly different from a normal distribution, i.e. it is non-normal.

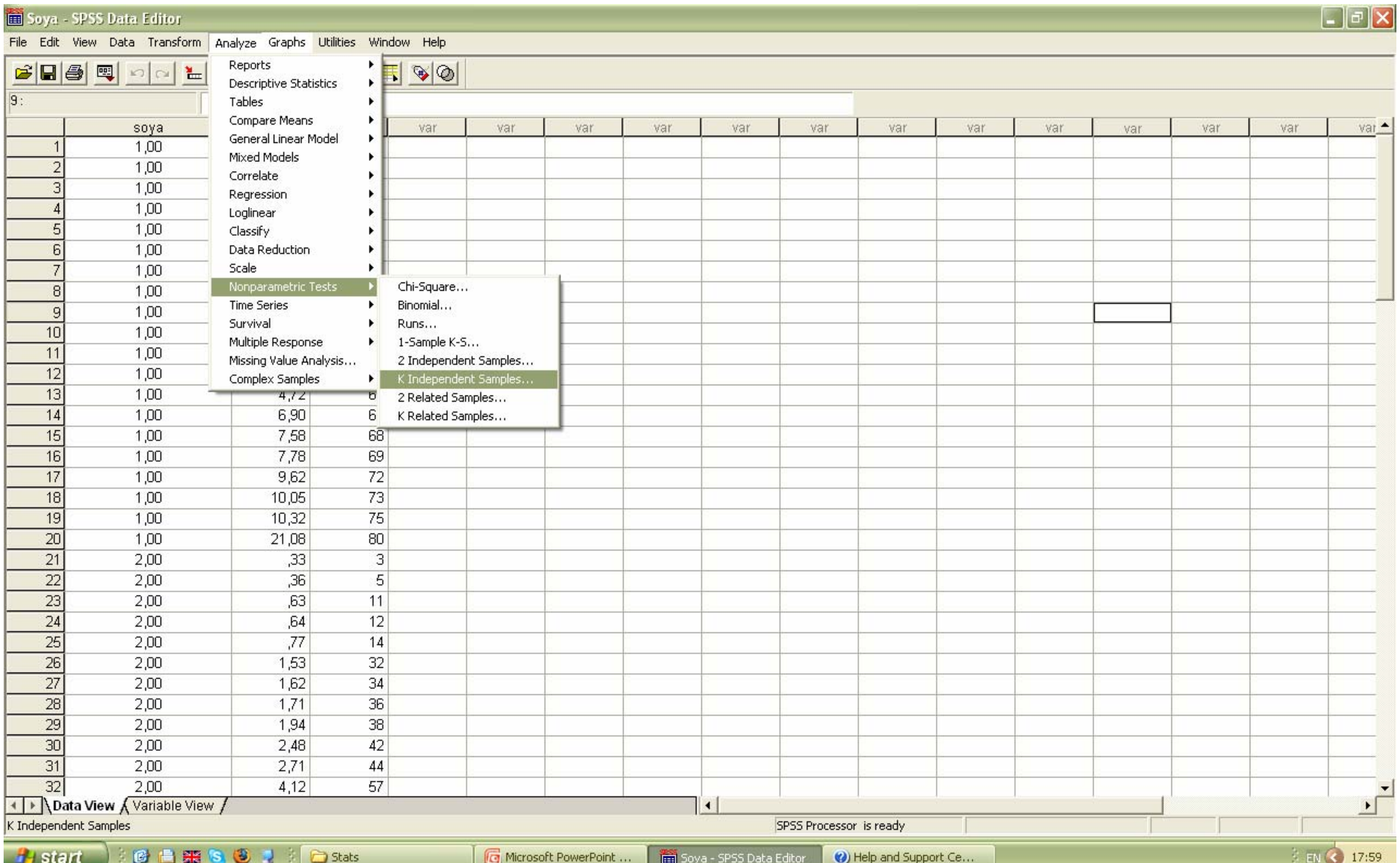
THE EFFECT OF SOYA ON CONCENTRATION: HOMOGENEITY OF VARIANCE

Test of Homogeneity of Variance

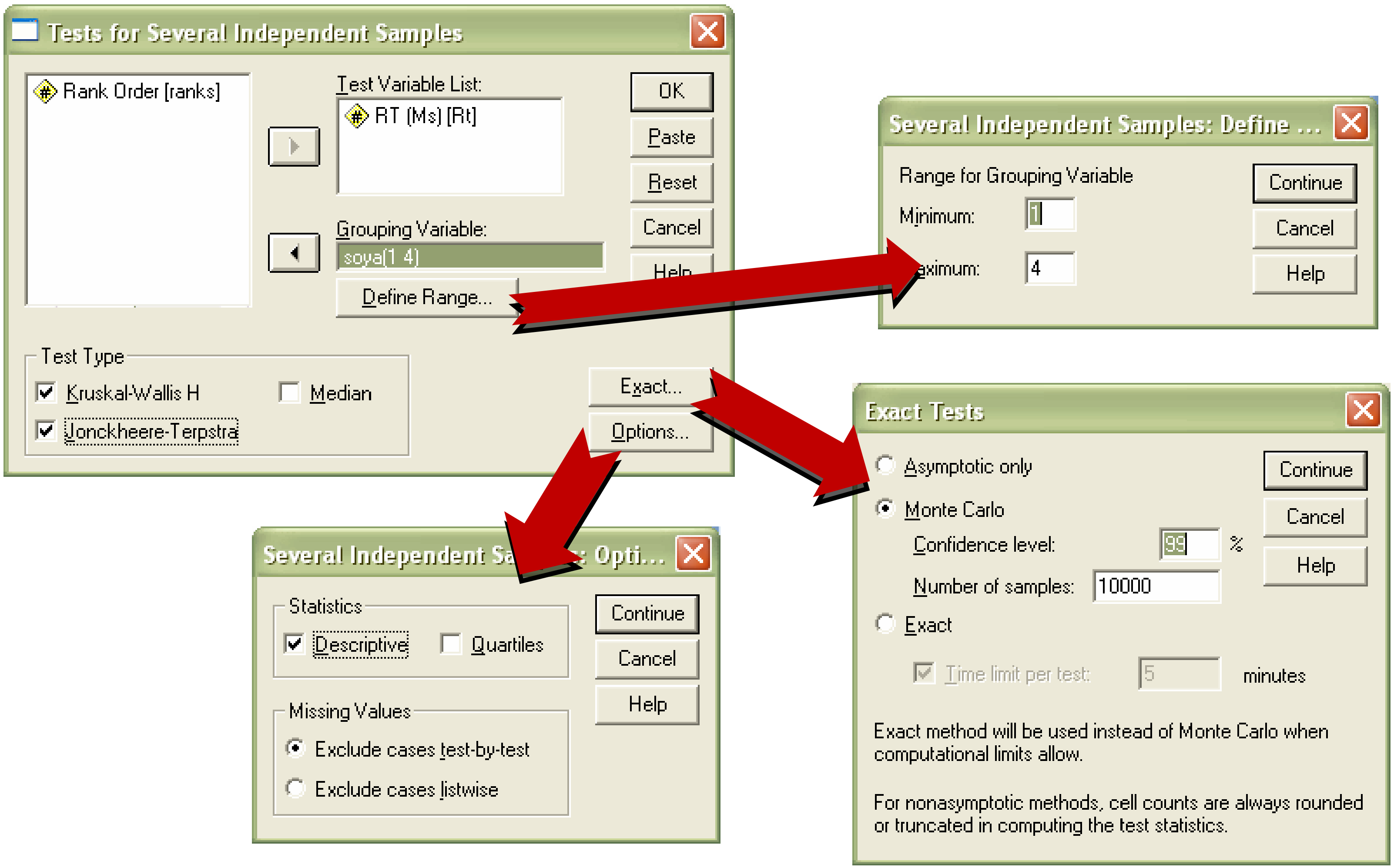
		Levene Statistic	df1	df2	Sig.
RT (Ms)	Based on Mean	5,117	3	76	,003
	Based on Median	2,860	3	76	,042
	Based on Median and with adjusted df	2,860	3	58,107	,045
	Based on trimmed mean	4,070	3	76	,010

Significance of data ⇒ the variances in different groups are significantly different ⇒ data are not homogenous

KRUSKAL-WALLIS: SPSS



KRUSKAL-WALLIS: SPSS



THE EFFECT OF SOYA ON CONCENTRATION: SPSS: RANKS

Ranks

	Number of Soya Meals	N	Mean Rank
RT (Ms)	No Soya Meals	20	46,35
	1 Soya Meal Per Week	20	44,15
	4 Soyal Meals Per Week	20	44,15
	7 Soya Meals Per Week	20	27,35
	Total	80	

THE EFFECT OF SOYA ON CONCENTRATION: SPSS: TEST STATISTICS

Test Statistics [b,c]

			RT (Ms)
Chi-Square			8,659
df			3
Asymp. Sig.			,034
Monte Carlo Sig.	Sig.		,031[a]
	99% Confidence Interval	Lower Bound	,027
		Upper Bound	,036

a. Based on 10000 sampled tables with starting seed 2000000.

b. Kruskal Wallis Test

c. Grouping Variable: Number of Soya Meals Per Week

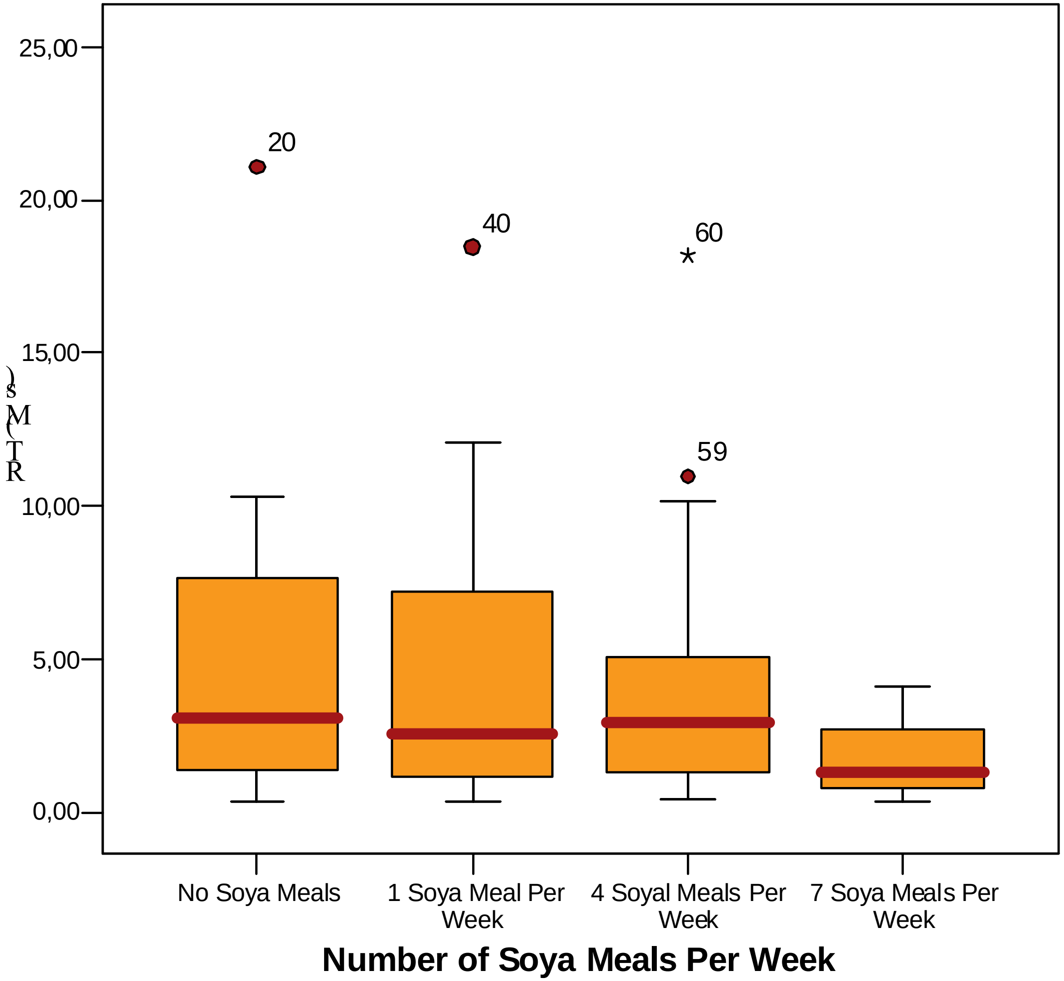
$\Rightarrow$ Test significance p <.034

$\Rightarrow$ Confidence Interval .028-.037 – does not cross the boundary of .05 $\Rightarrow$ a lot of confidence that the significant effect is genuine

THE EFFECT OF SOYA ON CONCENTRATION: THE CONCLUSION

We know that there is difference but we don’t know exactly where!

KRUSKAL-WALLIS: FINDING THE DIFFERENCE



KRUSKAL-WALLIS:
POST HOC TESTS: MANN-WHITNEY TEST

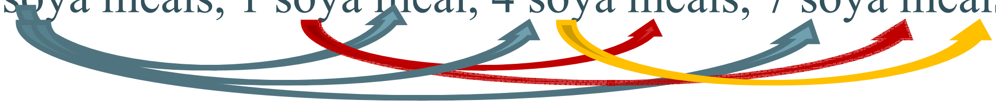
Mann-Whitney tests = Wilcoxon Rank Sum Test

a non-parametric test for comparing two independent groups based on ranking

! Lots of Wilcoxon Rank Sum Tests ⇒ inflation of the Type I error (the probability of falsely rejecting the H_0)

- ⇒ Bonferroni correction - .05/ the number of test to be conducted
  - ⇒ The value of significance becomes too small, i.e.:

0 soya meals, 1 soya meal, 4 soya meals, 7 soya meals = 6 tests



⇒ .05/6= .0083

KRUSKAL-WALLIS: *POST HOC TESTS*: MANN-WHITNEY TEST

Select a number of comparisons to make, i.e.:

Test 1: 1 soya meal per week compared to 0 soya meals
Test 2: 4 soya meals per week compared to 0 soya meals
Test 3: 7 soya meals per week compared to 0 soya meals

$\Rightarrow \alpha$ level = .05/3= .0167

KRUSKAL-WALLIS: *POST HOC TESTS:* MANN-WHITNEY TEST

1. 0 soya vs. 1 meal per week

Test Statistics[b]

	RT (Ms)
Mann-Whitney U	191,000
Wilcoxon W	401,000
Z	-,243
Asymp. Sig. (2-tailed)	,808
Exact Sig. [2*(1-tailed Sig.)]	,820[a]

a. Not corrected for ties.

b. Grouping Variable: Number of Soya Meals Per Week

2. 0 soya vs. 4 meals per week

Test Statistics[b]

	RT (Ms)
Mann-Whitney U	188,000
Wilcoxon W	398,000
Z	-,325
Asymp. Sig. (2-tailed)	,745
Exact Sig. [2*(1-tailed Sig.)]	,758[a]

a. Not corrected for ties.

b. Grouping Variable: Number of Soya Meals Per Week

3. 0 soya vs. 7 meals per week

Test Statistics[b]

	RT (Ms)
Mann-Whitney U	104,000
Wilcoxon W	314,000
Z	-2,597
Asymp. Sig. (2-tailed)	,009
Exact Sig. [2*(1-tailed Sig.)]	,009[a]

a. Not corrected for ties.

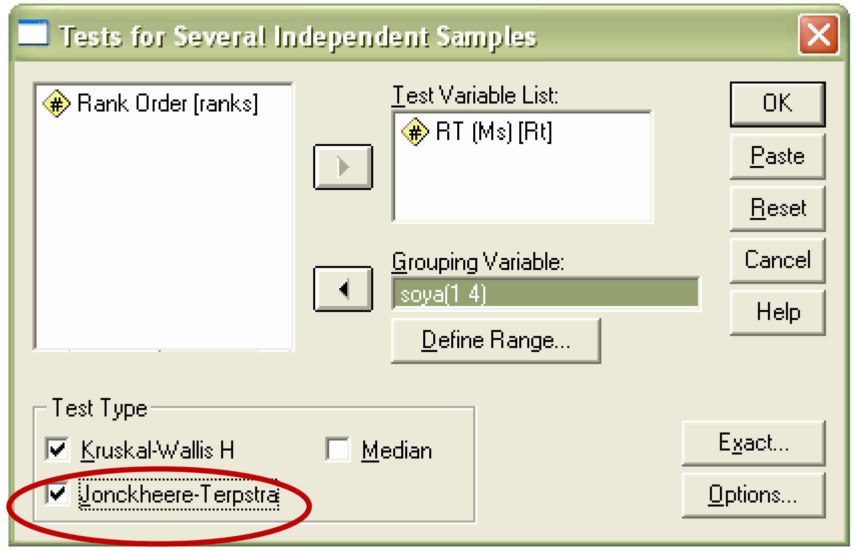
b. Grouping Variable: Number of Soya Meals Per Week

! α level = .0167 not .05

KRUSKAL-WALLIS: TESTING FOR TRENDS: JONCKHEERE-TERPSTRA TEST

If we expect that the groups we compare are ordered in a certain way.

I.e. the more soya a person eats the more concentrated and faster they become (shorter RTs)



KRUSKAL-WALLIS: TESTING FOR TRENDS: JONCKHEERE-TERPSTRA TEST

Jonckheere-Terpstra Test[b]

			RT (Ms)
Number of Levels in Number of Soya Meals Per Week			4
N			80
Observed J-T Statistic			912,000
Mean J-T Statistic			1200,000
Std. Deviation of J-T Statistic			116,333
Std. J-T Statistic			-2,476
Asymp. Sig. (2-tailed)			,013
Monte Carlo Sig. (2-tailed)	Sig.		,012[a]
	99% Confidence Interval	Lower Bound	,009
		Upper Bound	,015
Monte Carlo Sig. (1-tailed)	Sig.		,006[a]
	99% Confidence Interval	Lower Bound	,004
		Upper Bound	,008

a. Based on 10000 sampled tables with starting seed 2000000.

b. Grouping Variable: Number of Soya Meals Per Week

normal distribution

z score calculated: -2,476

If $> 1.65 \Rightarrow$ significant result .

"-" descending medians $\Rightarrow$ scores get smaller

"+" ascending medians $\Rightarrow$ scores get bigger

Medians get smaller the more soya meals we eat :

$\Rightarrow$ RTs become faster

$\Rightarrow$ more soya better concentration and more speed!

CALCULATING AN EFFECT SIZE

A standardized measure of the magnitude of the observed effect

$\Rightarrow$ the measured effect is meaningful or important

Cohen's *d* or Pearson's correlation coefficient *r:*

$$1 > r < 0$$

$r = .10$ small effect 1% of the total variance

$r = .30$ medium effect 9 %of total variance

$r = .50$ large effect 25 %of total varince

Converting *z* score into the effect size estimate

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

CALCULATING AN EFFECT SIZE

⇒ Difficult to convert χ^2 statistic with df >1 to an effect size r

⇒ Instead of the Kruskal-Wallis we can do it for the Wilcoxon's (Mann-Whitney) tests

$$r_{\text{NoSoya–1 meal}} = \frac{-0.243}{\sqrt{40}} = -.04$$ very small effect (close to 0)

$$r_{\text{NoSoya- 4 meals}} = \frac{-0.325}{\sqrt{40}} = -.05$$ very small effect (close to 0)

$$r_{\text{NoSoya-7meals}} = \frac{-2.597}{\sqrt{40}} = -.41$$ medium effect

$$r_{\text{Jonckheere}} = \frac{-2.476}{\sqrt{80}} = -.28$$