WEEK 11

COMPARING SEVERAL MEANS: ONE WAY ANOVA

COMPARING SEVERAL MEANS

Suppose we have 4 groups (A, B, C, D)

A-B	**B-C**
A-C	**B-D**
A-D	**C-D**

Can we make multiple "t tests" to compare means?

Remember that with every single t test:

The probability of incorrectly rejecting the H_0 (Tip 1 error rate) = 5%
Therefore, probability of no Type 1 error =95%

If we assume that each each test is independent then the overall probability of not doing a type 1 error is:
.95 * .95. * 95 * .95 * .95. * 95 = .736 (so making a Tip 1 error rate= 1-0.736 = 0.264 => 26.4%)

T test

H_0 : two samples have the same mean.

F test → ANOVA

H_0: whether three or more means are the same

COMPARING MORE THAN 2 GROUPS...

Data Collection

Different groups of people take part in each experimental condition

Between group, independent design

Same participants take part in each experimental condition

Within-subjects design, repeated measures

Parametric test assumptions met: → One way ANOVA | Repated measurement ANOVA

Parametric test assumptions violated: → Kruskal Wallis test | Freidman test

- ANOVA
  - H0= The mean (average value of the dependent variable) is the same for all groups..

At the end of the data analysis..

What if H_0 is Rejected? (there is a difference; $p<0.05$).
"It does not provide specific information about which groups were different!?"

We need post-hoc comparison tests

Equal variances assumed

- LSD
- Bonferroni
- Sidak
- Scheffe
- R-E-G-W-f
- R-E-G-W-Q
- S-N-K
- Tukey
- Duncan
- Hochberg's GT2
- Gabriel
- Dunnet

Equal variances not assumed

- Tamhane's T2
- Dunnett's T3
- Games-Howell
- Dunnett's C

SELECTING THE PROPER POST HOC TEST?

Gruplarınızın denek sayısı eşitse ve varyasyon homojense;	• REGWQ veya Tukey testi
Tip 1 hata üzerinde kontrollü olunmak isteniyorsa;	• Bonferroni
Grupların denek sayısında az çok bir farklılık varsa;	• Gabriel
Grupların deney sayıları çok farklıysa;	• Hochberg's GT2
Varyasyonlar heterojense;	• Games-Howell

ASSUMPTIONS OF ANOVA

- Variances in each experimental condition need to be fairly similar
- Observations should be independent
- Dependent variable should be measured at least on interval scale

TEST STATISTIC

$$\text{test statistics } (F \text{ test}) = \frac{\text{amount of variance explained by the model}}{\text{amount of variance not explained by the model}} = \frac{\text{effect}}{\text{error}} = \frac{MS_M}{MS_R}$$

- The total amount of variation within our data is called Total sum of squares (SS_T)

$$SS_T = \sum (x_i - \bar{x}_{grand})^2$$

- How much of the total amount of variation is explained by the model ➔ (Model Sum of Squares - SS_M)

$$SS_M = \sum n_k(\bar{x}_k - \bar{x}_{grand})^2$$

- How much of the total amount of variation can not be explained by the model ➔ (Residual Sum of Squares- SS_R)

$$SS_R = \sum (x_{ik} - \bar{x}_k)^2$$

TEST STATISTIC

- As, both SS_M and SS_R are summed values they will be influenced by the number of scores that were summed. To eliminate this bias we calculate average sum of squares ==> MS

$$MS_M = \frac{SS_M}{df_M} \qquad MS_R = \frac{SS_R}{df_R}$$

$$test\ statistics\ (F\ test) = \frac{amount\ of\ variance\ explained\ by\ the\ model}{amount\ of\ variance\ not\ explained\ by\ the\ model} = \frac{effect}{error} = \frac{MS_M}{MS_R}$$

	age_group	bodylength
1	2 years old	50,00
2	2 years old	45,00
3	2 years old	48,00
4	2 years old	47,00
5	2 years old	45,00
6	2 years old	49,00
7	2 years old	50,00
8	2 years old	54,00
9	2 years old	57,00
10	2 years old	55,00
11	3 years old	63,00
12	3 years old	55,00
13	3 years old	54,00
14	3 years old	49,00
15	3 years old	65,00
16	3 years old	46,00
17	3 years old	53,00
18	3 years old	67,00
19	3 years old	58,00
20	3 years old	50,00
21	4 years old	71,00
22	4 years old	67,00

Example

Suppose that a researcher wants to examine the effect of age on the body length measurements (*eg. body length*) in the awasi sheep at the end of shearing season.

Dependent variable:
Body length

Independent variable:
Age group
- 2 years old
- 3 years old
- 4 years old

Hypothesis ?

Data set > Awasi_age.sav

a)Normality assumption:

H_0= The data follow a normal distribution
H_1= The data do not follow a normal distribution

Step 1: Testing the assumptions

Tests of Normality

Age_Group		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Body_Length	2 years old	,200	10	,200*	,929	10	,437
	3 years old	,156	10	,200*	,949	10	,652
	4 years old	,145	10	,200*	,956	10	,741

P>0.05
H_0 is accepted

b) Homogeneity of variances assumption:

H_0= The population variances are equal
H_1= The population variances are not equal

Test of Homogeneity of Variance

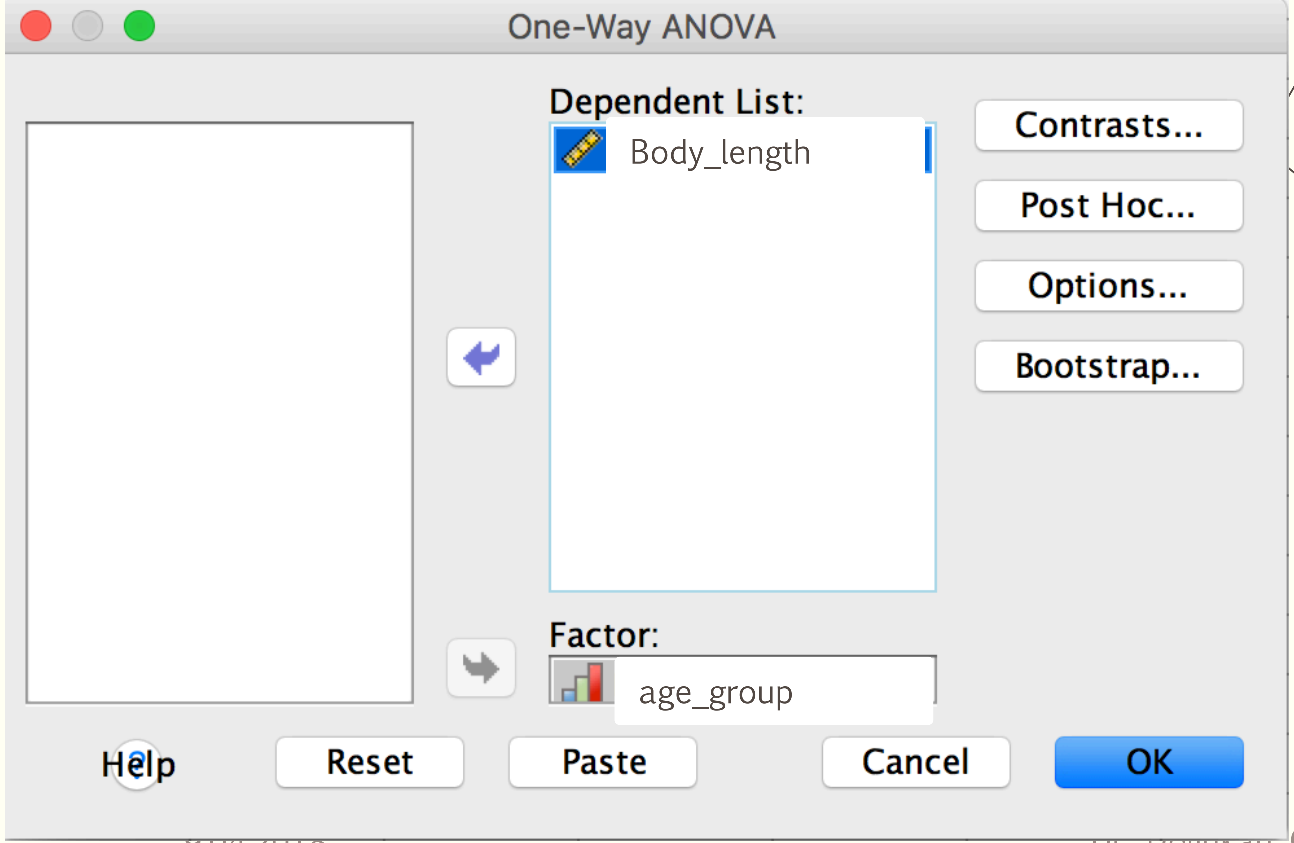
		Levene Statistic	df1	df2	Sig.
body_length	Based on Mean	2,569	2	27	,095
	Based on Median	1,734	2	27	,196
	Based on Median and with adjusted df	1,734	2	19,176	,203
	Based on trimmed mean	2,527	2	27	,099

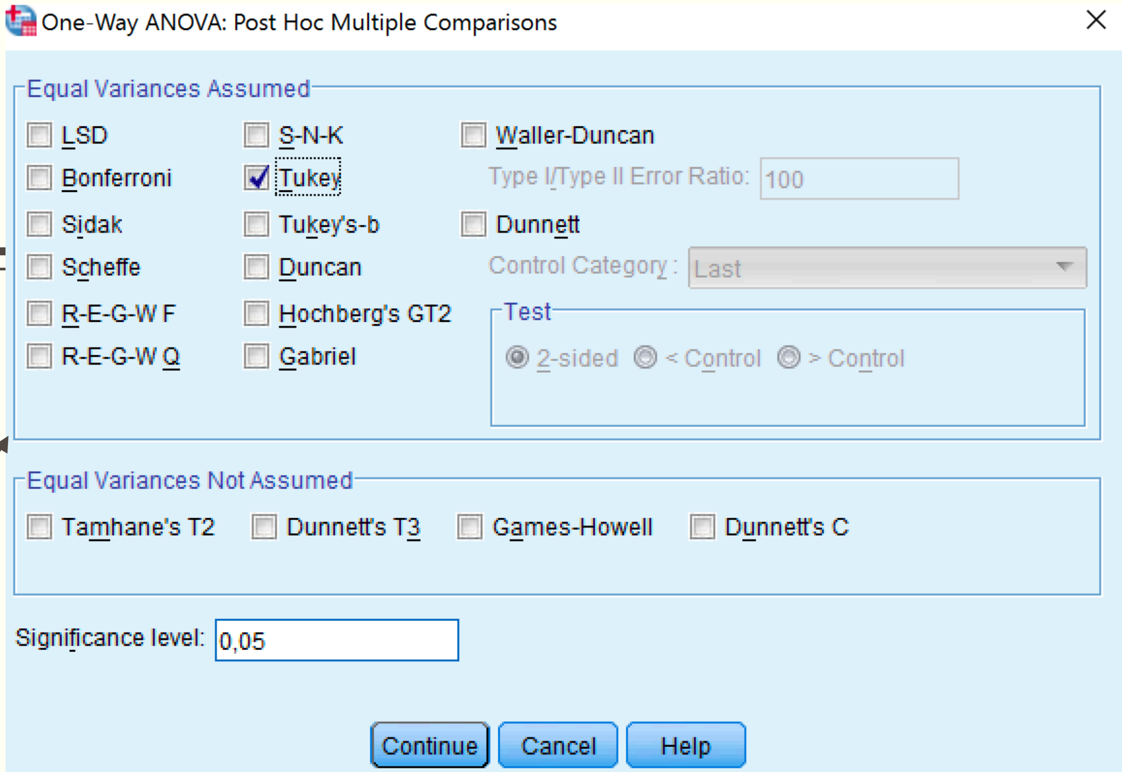
P>0.05
H_0 is accepted

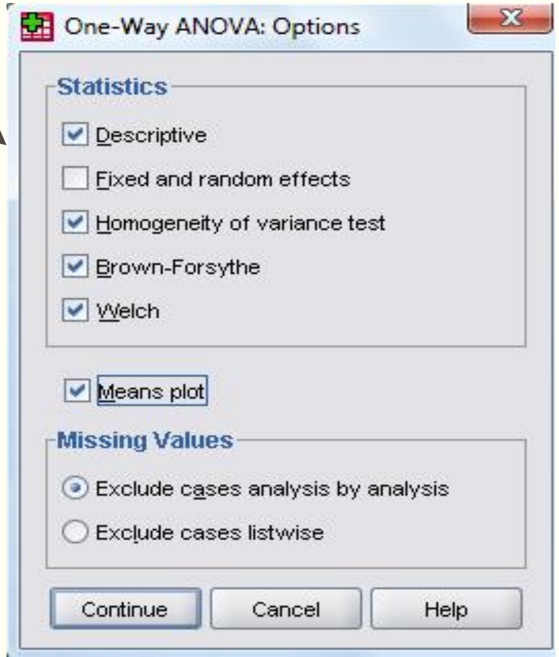
Step 2: Data analysis: One way ANOVA

Analyze > Compare Means > One Way ANOVA or

(Analyze > General Linear Model > Univariate)







Output

Descriptives

Body_length

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
2 years old	10	50,0000	4,13656	1,30809	47,0409	52,9591	45,00	57,00
3 years old	10	56,0000	7,10243	2,24598	50,9192	61,0808	46,00	67,00
4 years old	10	65,4000	4,29987	1,35974	62,3241	68,4759	58,00	71,00
	30	57,1333	8,26181	1,50839	54,0483	60,2183	45,00	71,00

Test of Homogeneity of Variances

body_length

Levene Statistic	df1	df2	Sig.
2,569	2	27	,095

Robust Tests of Equality of Means

Body_length

	Statistic[a]	df1	df2	Sig.
Welch	32,235	2	17,336	,000
Brown-Forsythe	21,008	2	20,959	,000

a. Asymptotically F distributed.

ANOVA

body_length

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1205,067	2	602,533	21,008	,000
Within Groups	774,400	27	28,681		
Total	1979,467	29			

P<0.05 => H0 is rejected => There is a statistically significant difference in body length measurements of sheeps among the age groups.

But, which groups are different exactly?

Output

Multiple Comparisons

Dependent Variable: body_length
Tukey HSD

(I) Age_group	(J) Age_group	Mean Difference (I–J)	Std. Error	Sig.
2 years old	3 years old	-6,00000*	2,39506	,047
	4 years old	-15,40000*	2,39506	,000
3 years old	2 years old	6,00000*	2,39506	,047
	4 years old	-9,40000*	2,39506	,002
4 years old	2 years old	15,40000*	2,39506	,000
	3 years old	9,40000*	2,39506	,002

*. The mean difference is significant at the 0.05 level.

body_length

Tukey HSD[a]

Age_group	N	Subset for alpha = 0.05		
		1	2	3
2 years old	10	50,0000		
3 years old	10		56,0000	
4 years old	10			65,4000
Sig.		1,000	1,000	1,000

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 10,000.

Reporting the results ==>

Variable	n	Mean ± Std. Error	P
2 years old	10	50 ± 1,31 [c]	<0,001
3 years old	10	56 ± 2,25 [b]	
4 years old	10	65,4 ± 1,36 [a]	

Interpretation?

a, b: Different letters in the same column indicate statistical significance (p<0.05)

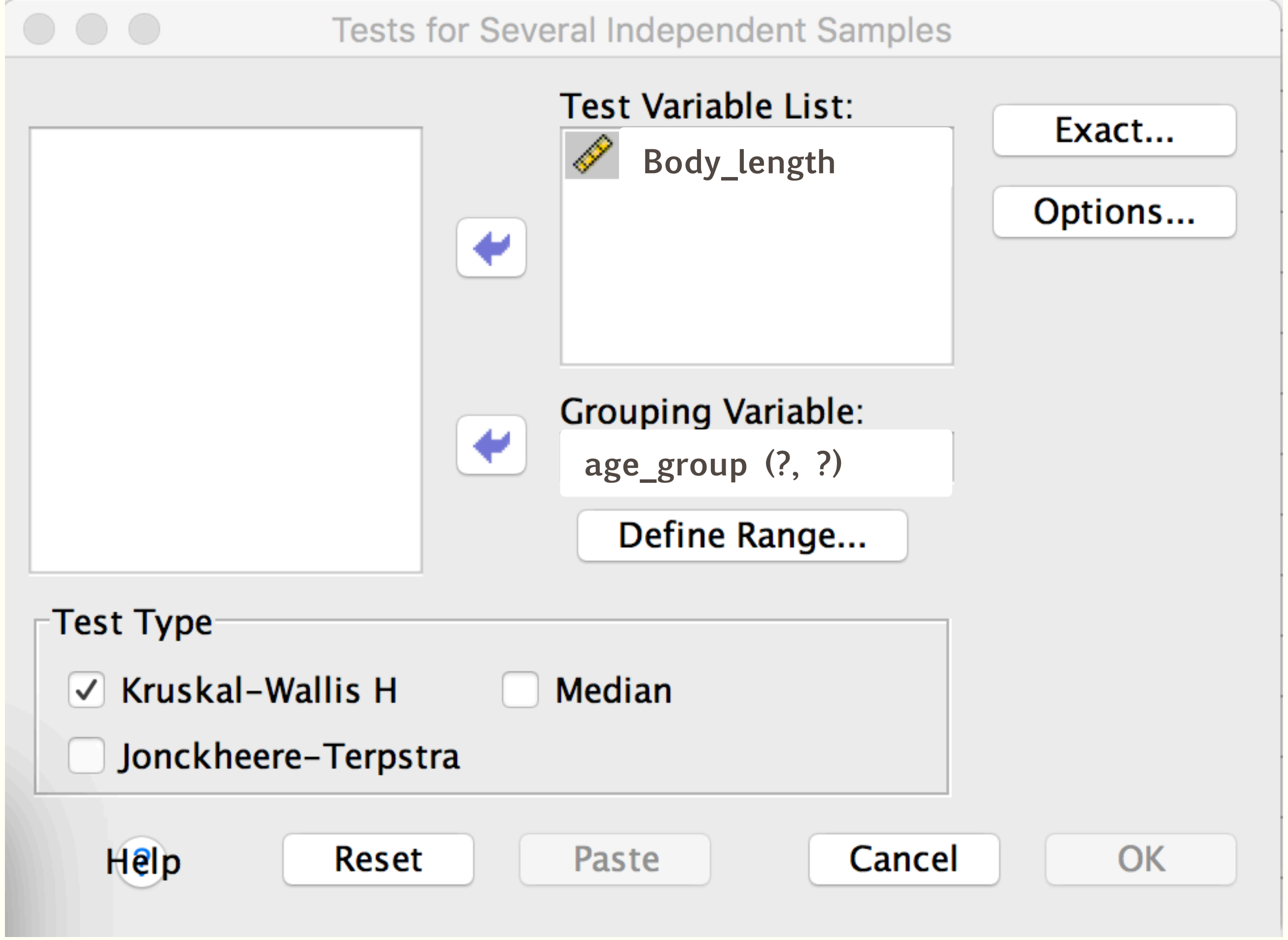
What if the parametric test assumptions are violated?

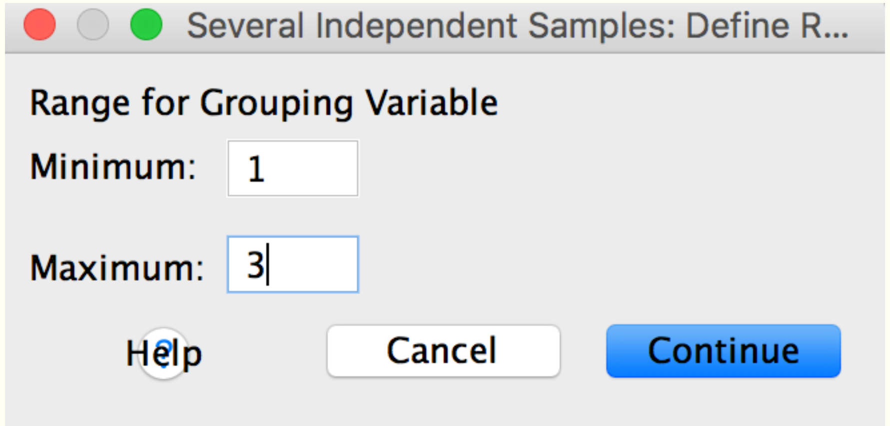
Data Collection

	Different groups of people take part in each experimental condition (Between group, independent design)	Same participants take part in each experimental condition (Within-subjects design, repeated measures)
Parametric test assumptions met:	One way ANOVA	Repated measurement ANOVA
Parametric test assumptions violated:	Kruskal Wallis test	Freidman test

Let’s use the same dataset and assume that the assumptions are violated

- Analyze > Non-Parametric Tests > Legacy Dialogs > K Independent Samples





Ranks

	Age_group	N	Mean Rank
Body_length	2 years old	10	7,95
	3 years old	10	14,45
	4 years old	10	24,10
	TOTAL	30	

Test Statistics[a,b]

	Body_length
Chi-Square	17,086
df	2
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable: Keratoplasti_Yöntemi

But, which groups are different from each other?

Post hoc testing procedures for Kruskal Wallis test:

(1) Dunn's Test

(2) Using multiple Mann Whitney U tests → Bonferroni correction

↓

P= 0.05 / number of comparision → P=0.0167

2 y- 3 y

Test Statistics[a]

	Body_length
Mann-Whitney U	24,500
Wilcoxon W	79,500
Z	-1,933
Asymp. Sig. (2-tailed)	,053
Exact Sig. [2*(1-tailed Sig.)]	,052[b]

2 y- 4 y

Test Statistics[a]

	Body_length
Mann-Whitney U	,000
Wilcoxon W	55,000
Z	-3,782
Asymp. Sig. (2-tailed)	,000
Exact Sig. [2*(1-tailed Sig.)]	,000[b]

3 y – 4 y

Test Statistics[a]

	Body_length
Mann-Whitney U	14,000
Wilcoxon W	69,000
Z	-2,725
Asymp. Sig. (2-tailed)	,006
Exact Sig. [2*(1-tailed Sig.)]	,005[b]