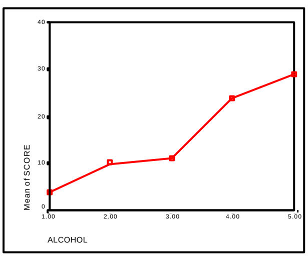
Multiple comparisons procedures

Experiment

- Effects of alcohol on driving ability
- 5 conditions: alcohol 1 (no alcohol), alcohol 2, alcohol 3, alcohol 4, alcohol 5 (high alcohol)
- 8 participants in each condition
- Driving in simulator
- Error rate

Experiment

- Results: One-way ANOVA: $F(4,39) = 27.33$; $p < 0.00$



Questions

- Significant difference between each conditions?
- Predictions from other study:
  - Sig. diff. between alcohol 3 and alcohol 4
  - Sig. diff. between alcohol 1 and alcohol 2

Learning objectives

- Why multiple comparison procedures?
- Characteristics of procedures
- Most popular procedures

ANOVA

- Important:
  Rejection of null hypothesis means at least one of the means is different from at least one other mean, but we do **not** know exactly which means are different from which other means!

Multiple comparisons

	Alco. 1	Alco. 2	Alco. 3	Alco. 4	Alco. 5
Alco. 1					
Alco. 2					
Alco. 3					
Alco. 4					

Multiple comparison procedures

- **Planned (a priori) comparisons:**
  Small number of comparisons which were chosen before the experiment, e.g. following up predictions from theories

- **Post hoc comparisons:**
  all possible comparisons

Error in multiple testing

- **For one test the probability of basing conclusions on an error is the α-level (Type I error)**
- **For multiple testing conclusions are based on a family of tests.**
- **The error for multiple testing is given by the familywise error rate (FW)**
- **FW is the probability of making least one Type I error in a family of tests.**

Problem with multiple testing

- **FW increases with number of tests**
- **FW is an accumulation of α-levels**

- **Illustration:**
  The more often a dice is rolled the more likely it is that there is a one amongst the throws

Multiple comparison procedures

- **Objective:**
  Keep the familywise error at bay, but maintaining power of test
- **more "conservative" test: emphasis on keeping FW under control at the cost of losing power**
- **more "liberal" test: lose control of FW, but higher power of test**

Warning

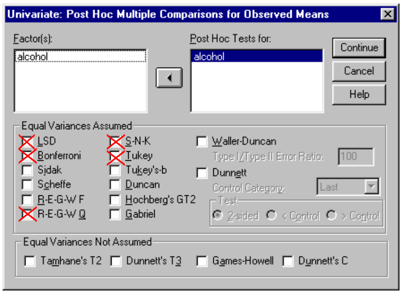
- **Look at the data and then test the two means with the largest contrast**

- **Important: Implicitly made comparisons between all other means. Consequently, the Type I error from all the other implicit tests add to the familywise error.**

Controversy

- Past: Multiple comparisons only with significant overall ANOVA
- However: Most post hoc tests don't require significant overall ANOVA
- Outside this course: Follow tradition and make the ANOVA before

SPSS: Post hoc's test



Fisher's Least Significance Difference Test (LSD)

- = Fisher's protected t
- Modified t-test
- Post hoc comparison
- Requires significant overall F
- For three means the familywise error stays at α

Multiple Comparisons

Dependent Variable: SCORE
LSD

(I) ALCOHOL	(J) ALCOHOL	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1.00	2.00	-6.0000*	2.82843	.041	-11.7420	-.2580
	3.00	-7.0000*	2.82843	.018	-12.7420	-1.2580
	4.00	-20.0000*	2.82843	.000	-25.7420	-14.2580
	5.00	-25.0000*	2.82843	.000	-30.7420	-19.2580
2.00	1.00	6.0000*	2.82843	.041	.2580	11.7420
	3.00	-1.0000	2.82843	.726	-6.7420	4.7420
	4.00	-14.0000*	2.82843	.000	-19.7420	-8.2580
	5.00	-19.0000*	2.82843	.000	-24.7420	-13.2580
3.00	1.00	7.0000*	2.82843	.018	1.2580	12.7420
	2.00	1.0000	2.82843	.726	-4.7420	6.7420
	4.00	-13.0000*	2.82843	.000	-18.7420	-7.2580
	5.00	-18.0000*	2.82843	.000	-23.7420	-12.2580
4.00	1.00	20.0000*	2.82843	.000	14.2580	25.7420
	2.00	14.0000*	2.82843	.000	8.2580	19.7420
	3.00	13.0000*	2.82843	.000	7.2580	18.7420
	5.00	-5.0000	2.82843	.086	-10.7420	.7420
5.00	1.00	25.0000*	2.82843	.000	19.2580	30.7420
	2.00	19.0000*	2.82843	.000	13.2580	24.7420
	3.00	18.0000*	2.82843	.000	12.2580	23.7420
	4.00	5.0000	2.82843	.086	-.7420	10.7420

Based on observed means.

*. The mean difference is significant at the .05 level.

LSD-Test: Results

	Alco. 1	Alco. 2	Alco. 3	Alco. 4	Alco. 5
Alco. 1		p=0.041	p=0.018	p<0.001	p<0.001
Alco. 2			p=0.726	p<0.001	p<0.001
Alco. 3				p<0.001	p<0.001
Alco. 4					p=0.086

Characteristics of LSD-Test

- Powerful for 3 means/groups
- But loses control of FW for more than 3 means

Bonferroni Procedure

- A priori and post hoc comparison
- Modified t-tests with new a-level
- new a-level is given by dividing the original a-level by number of tests
- keeps the familywise error at bay

Bonferroni Procedure

- Testing two predictions from example (a priori):
  - new a-level = 0.05 / 2 = 0.025
  - Without correction: FW " 0.05 + 0.05 = 0.1
- Post hoc comparison in example:
  - 10 comparisons (4 + 3 + 2 + 1)
  - new a-level = 0.05 / 10 = 0.005
  - Without correction: FW " 10 * 0.05 = 0.5

Multiple Comparisons

Dependent Variable: SCORE
Bonferroni

(I) ALCOHOL	(J) ALCOHOL	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1.00	2.00	-6.0000	2.82843	.411	-14.4741	2.4741
	3.00	-7.0000	2.82843	.183	-15.4741	1.4741
	4.00	-20.0000*	2.82843	.000	-28.4741	-11.5259
	5.00	-25.0000*	2.82843	.000	-33.4741	-16.5259
2.00	1.00	6.0000	2.82843	.411	-2.4741	14.4741
	3.00	-1.0000	2.82843	1.000	-9.4741	7.4741
	4.00	-14.0000*	2.82843	.000	-22.4741	-5.5259
	5.00	-19.0000*	2.82843	.000	-27.4741	-10.5259
3.00	1.00	7.0000	2.82843	.183	-1.4741	15.4741
	2.00	1.0000	2.82843	1.000	-7.4741	9.4741
	4.00	-13.0000*	2.82843	.001	-21.4741	-4.5259
	5.00	-18.0000*	2.82843	.000	-26.4741	-9.5259
4.00	1.00	20.0000*	2.82843	.000	11.5259	28.4741
	2.00	14.0000*	2.82843	.000	5.5259	22.4741
	3.00	13.0000*	2.82843	.001	4.5259	21.4741
	5.00	-5.0000	2.82843	.858	-13.4741	3.4741
5.00	1.00	25.0000*	2.82843	.000	16.5259	33.4741
	2.00	19.0000*	2.82843	.000	10.5259	27.4741
	3.00	18.0000*	2.82843	.000	9.5259	26.4741
	4.00	5.0000	2.82843	.858	-3.4741	13.4741

Based on observed means.
*. The mean difference is significant at the .05 level.

Bonferroni Procedure: Results

- Post hoc comparison in example:

	Alco. 1	Alco. 2	Alco. 3	Alco. 4	Alco. 5
Alco. 1		p=0.441	p=0.183	P<0.001	P<0.001
Alco. 2			p=1.000	P<0.001	P<0.001
Alco. 3				P=0.001	P<0.001
Alco. 4					p=0.858

were significant in LSD

Bonferroni Procedure as a priori test: Results

- Testing two predictions from example (a priori):
  alcohol 1 vs. alcohol 2 not confirmed (p = 0.041)
  alcohol 3 vs. alcohol 4 confirmed (p < 0.001)
- Technical problem with SPSS

Used for a priori test

Dependent Variable: SCORE
LSD

(I) ALCOHOL	(J) ALCOHOL	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1.00	2.00	-6.0000*	2.82843	.041	-11.7420	-.2580
	3.00	-7.0000*	2.82843	.018	-12.7420	-1.2580
	4.00	-20.0000*	2.82843	.000	-25.7420	-14.2580
	5.00	-25.0000*	2.82843	.000	-30.7420	-19.2580
2.00	1.00	6.0000*	2.82843	.041	.2580	11.7420
	3.00	-1.0000	2.82843	.726	-6.7420	4.7420
	4.00	-14.0000*	2.82843	.000	-19.7420	-8.2580
	5.00	-19.0000*	2.82843	.000	-24.7420	-13.2580
3.00	1.00	7.0000*	2.82843	.018	1.2580	12.7420
	2.00	1.0000	2.82843	.726	-4.7420	6.7420
	4.00	-13.0000*	2.82843	.000	-18.7420	-7.2580
	5.00	-18.0000*	2.82843	.000	-23.7420	-12.2580
4.00	1.00	20.0000*	2.82843	.000	14.2580	25.7420
	2.00	14.0000*	2.82843	.000	8.2580	19.7420
	3.00	13.0000*	2.82843	.000	7.2580	18.7420
	5.00	-5.0000	2.82843	.086	-10.7420	.7420
5.00	1.00	25.0000*	2.82843	.000	19.2580	30.7420
	2.00	19.0000*	2.82843	.000	13.2580	24.7420
	3.00	18.0000*	2.82843	.000	12.2580	23.7420
	4.00	5.0000	2.82843	.086	-.7420	10.7420

Based on observed means.
*. The mean difference is significant at the .05 level.

Characteristics of Bonferroni Procedure

- Firm control of FW, but little power
- Very conservative for post hoc comparison
- Best used for a priori testing

"Sorting"-procedures

- (Student-)Newman-Keuls Test (S-N-K)
- Tukey's Test
- Ryan Procedure (REGWQ)

"Sorting"-procedure

- Sort the means in ascending order
- Finds homogenous groups
- Pairwise comparison of means through a modified t-test utilizing order of means
- Different tests: Different criteria for forming groups (similar to Bonferroni)

Characteristics of "Sorting"-procedures

- (Student-)Newman-Keuls Test (SNK): Powerful, but tend to lose control of FW
- Tukey's Test: firm control of FW, but less powerful
- Ryan Procedure (REGWQ): firm control of FW and more powerful than Tukey's Test

SNK-Test: Results

Homogeneous Subsets

SCORE

Student-Newman-Keuls[a,b]

ALCOHOL	N	Subset 1	Subset 2	Subset 3
1.00	8	4.0000		
2.00	8		10.0000	
3.00	8		11.0000	
4.00	8			24.0000
5.00	8			29.0000
Sig.		1.000	.726	.086

Means for groups in homogeneous subsets are displayed.
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 32.000.

a. Uses Harmonic Mean Sample Size = [illegible]000.

b. Alpha = .05.

Minimal p-value within the group

SNK: Results

	Alco. 1	Alco. 2	Alco. 3	Alco. 4	Alco. 5
Alco. 1		p<0.05	p<0.05	p<0.05	p<0.05
Alco. 2			p>0.726	p<0.05	p<0.05
Alco. 3				p<0.05	p<0.05
Alco. 4					p>0.086

Tukey's HSD test

- HSD = honestly significant difference test
- Post hoc comparison
- Very popular

Multiple Comparisons

Dependent Variable: SCORE

	(I) ALCOHOL	(J) ALCOHOL	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Tukey HSD	1.00	2.00	-6.0000	2.82843	.234	-14.1319	2.1319
		3.00	-7.0000	2.82843	.120	-15.1319	1.1319
		4.00	-20.0000*	2.82843	.000	-28.1319	-11.8681
		5.00	-25.0000*	2.82843	.000	-33.1319	-16.8681
	2.00	1.00	6.0000	2.82843	.234	-2.1319	14.1319
		3.00	-1.0000	2.82843	.996	-9.1319	7.1319
		4.00	-14.0000*	2.82843	.000	-22.1319	-5.8681
		5.00	-19.0000*	2.82843	.000	-27.1319	-10.8681
	3.00	1.00	7.0000	2.82843	.120	-1.1319	15.1319
		2.00	1.0000	2.82843	.996	-7.1319	9.1319
		4.00	-13.0000*	2.82843	.000	-21.1319	-4.8681
		5.00	-18.0000*	2.82843	.000	-26.1319	-9.8681
	4.00	1.00	20.0000*	2.82843	.000	11.8681	28.1319
		2.00	14.0000*	2.82843	.000	5.8681	22.1319
		3.00	13.0000*	2.82843	.000	4.8681	21.1319
		5.00	-5.0000	2.82843	.408	-13.1319	3.1319
	5.00	1.00	25.0000*	2.82843	.000	16.8681	33.1319
		2.00	19.0000*	2.82843	.000	10.8681	27.1319
		3.00	18.0000*	2.82843	.000	9.8681	26.1319
		4.00	5.0000	2.82843	.408	-3.1319	13.1319

Based on observed means.

*. The mean difference is significant at the .05 level.

Tukey: Results

Homogeneous Subsets

SCORE

Tukey HSD[a]

ALCOHOL	N	Subset for alpha = .05 1	2
1.00	8	4.0000	
2.00	8	10.0000	
3.00	8	11.0000	
4.00	8		24.0000
5.00	8		29.0000
Sig.		.120	.408

Means for groups in homogeneous subsets are display

a. Uses Harmonic Mean Sample Size = 8.000.

Tukey: Results

	Alco. 1	Alco. 2	Alco. 3	Alco. 4	Alco. 5
Alco. 1		p>0.120	p>0.120	p<0.05	p<0.05
Alco. 2			p>0.120	p<0.05	p<0.05
Alco. 3				p<0.05	p<0.05
Alco. 4					p>0.408

Ryan Procedure: Results

Homogeneous Subsets

SCORE

Ryan-Einot-Gabriel-Welsch Range[a,b]

ALCOHOL	N	Subset 1	2
1.00	8	4.0000	
2.00	8	10.0000	
3.00	8	11.0000	
4.00	8		24.0000
5.00	8		29.0000
Sig.		.077	.201

Means for groups in homogeneous subsets are display
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 32.000.

a. Critical values are not monotonic for these data. Substitutions have been made to ensure monotonicity. Type I error is therefore smaller.

b. Alpha = .05.

Ryan Procedure: Results

	Placebo	Alco. 1	Alco. 2	Alco. 3	Alco. 4
Placebo		p>0.077	p>0.077	p<0.05	p<0.05
Alco. 1			p>0.077	p<0.05	p<0.05
Alco. 2				p<0.05	p<0.05
Alco. 3					p>0.201

2. Example

SCORE

	FACTOR1	N	Subset		
			1	2	3
Tukey HSD [a,b]	5	10	2.6416		
	3	10	4.2628	4.2628	
	4	10	4.5643	4.5643	
	2	10		6.7183	6.7183
	1	10			7.2063
	Sig.		.198	.054	.980
Ryan-Einot-Gabriel-Welsch Range [c,b]	5	10	2.6416		
	3	10	4.2628		
	4	10	4.5643		
	2	10		6.7183	
	1	10		7.2063	
	Sig.		.132	.885	

Means for groups in homogeneous subsets are displayed.
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 3.811.

a. Uses Harmonic Mean Sample Size = 10.000.

b. Alpha = .05.

c. Critical values are not monotonic for these data. Substitutions have been made to ensure monotonicity. Type I error is therefore smaller.

Tukey: 2. Example

	1	2	3	4	5
1		p>0.980	<u>p<0.05</u>	<u>p<0.05</u>	<u>p<0.05</u>
2			p>0.054	p>0.054	<u>p<0.05</u>
3				p > 0.198	p > 0.198
4					p>0.198

Recommendation

- A few a priori tests:
  Bonferroni
- Post hoc comparisons:
  Tukey HSD procedure or REGWQ-procedure
  For 3 groups: LSD

Multiple comparison procedure for Two-way ANOVA

- Comparisons between all cells
- Analysis of simple effects
- Analysis of means within the simple effects
- For analysis of simple effects the problem with the familywise error applies as well

Multiple comparisons for within-participants design

- None available in SPSS!
- Recommendation:
  paired t-test together with Bonferroni-correction