Chapter 12 Multiple Comparisons Among Treatment Means

12.1 The effects of food and water deprivation on a learning task:

a. ANOVA with linear contrasts:

Groups:	ad lib (1)	2/day (2)	food (3)	water (4)	f & w (5)	
Means:	18	24	8	12	11	
a_j:	.5	.5	-.333	-.333	-.333	$0.8333 = \Sigma a_j^2$
b_j:	1	-1	0	0	0	$2 = \Sigma b_j^2$
c_j:	0	0	.5	.5	-1	$1.5 = \Sigma c_j^2$
d_j:	0	0	1	-1	0	$2 = \Sigma d_j^2$

$$\psi_1 = (.5)(18) + (.5)(24) + (-.333)(8) + (-.333)(12)(-.333)(11) = 10.667$$

$$\psi_2 = (1)(18) + (-1)(24) + (0)(8) + (0)(12) + (0)(11) = -6$$

$$\psi_3 = (0)(18) + (0)(24) + (.5)(8) + (.5)(12) + (-1)(11) = -1$$

$$\psi_4 = (0)(18) + (0)(24) + (1)(8) + (-1)(12) + 0(11) = -4$$

$$SS_{contrast_1} = \frac{n\psi_1}{\Sigma a_j^2} = \frac{5(10.667^2)}{0.8333} = 682.667$$

$$SS_{contrast_2} = \frac{n\psi_2}{\Sigma b_j^2} = \frac{5(-6^2)}{2} = 90$$

$$SS_{contrast_3} = \frac{n\psi_3}{\Sigma c_j^2} = \frac{5(-1^2)}{1.5} = 3.333$$

$$SS_{contrast_4} = \frac{n\psi_4}{\Sigma d_j^2} = \frac{5(-4^2)}{2} = 40.000$$

Source	*df*	SS	MS	*F*
Deprivation	4	816.000	204.000	36.429*
1&2 vs 3,4,5	1	682.667	682.667	121.905*
1 vs 2	1	90.000	90.000	16.071*
3&4 vs 5	1	3.333	3.333	<1
3 vs 4	1	40.000	40.000	7.143*
Error	20	112.000	5.600	
Total	24	928.000		

$* p < .05 \quad [F_{.05(4,20)} = 2.87;\ F_{.05(1,20)} = 4.35]$

b. Orthogonality of contrasts:

Cross-products of coefficients:

$$\Sigma a_j b_j = (.5)(1) + (.5)(-1) + (.333)(0) + (.333)(0) + (.333)(0) = 0$$
$$\Sigma a_j c_j = (.5)(0) + (.5)(0) + (.333)(.5) + (.333)(.5) + (.333)(-1) = 0$$
$$\Sigma a_j d_j = (.5)(0) + (.5)(0) + (.333)(1) + (.333)(-1) + (.333)(0) = 0$$
$$\Sigma b_j c_j = (1)(0) + (-1)(0) + (0)(.5) + (0)(.5) + (0)(-1) = 0$$
$$c_j d_j = (0)(0) + (0)(0) + (.5)(1) + (.5)(-1) + (1)(0) = 0$$

c.

$$SS_{treat} = \Sigma SS_{contrast}$$
$$816.000 = 682.667 + 90.000 + 3.333 + 40.000$$

12.3 For $\alpha = .05$:

Per comparison error rate = α = .05
Familywise error rate = $1 - (1 - \alpha)^2$ = .0975.

12.5 Studentized range statistic for data in Exercise 11.2:

$\overline{X}_1 = 19.3 \quad n_1 = 10$

$\overline{X}_2 = 12.0 \quad n_2 = 10$

$$q_2 = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{MS_{error}}{n}}} = \frac{19.3 - 12.0}{\sqrt{\frac{10.56}{10}}} = \frac{7.3}{1.028} = 7.101$$

12.7 The Bonferroni test on contrasts in Exercise 12.2 (data from Exercise 11.1):

From Exercise 12.2: $\psi_1 = -15.00$ $\psi_2 = -20.00$ $n = 6$

$$\Sigma a_j^2 = 1.5 \qquad \Sigma b_j^2 = 2$$

$$\text{MS}_{\text{error}} = 26.167$$

$$t = \frac{\psi}{\sqrt{\frac{\Sigma a_j^2 MS_{error}}{n}}}$$

$$t_1' = \frac{-15}{\sqrt{\frac{(1.5)(26.167)}{6}}} = -5.86 \qquad t_1' = \frac{-20}{\sqrt{\frac{(2)(26.167)}{6}}} = -6.77$$

$[t_{.05}\left(df_{error} = 15; 2\,\text{comparisons}) = 2.49\right)$ Reject H_0 in each case.

12.9 Holm's multistage test for data in Exercise 12.1.

Comparison	*F*	*t*	c	$t'_{.05(20,c)}$	Signif
1&2 vs 3,4,5	121.905	11.04	4	2.74	*
1 vs. 2	16.071	4.01	3	2.61	*
3 vs. 4	7.143	2.67	2	2.42	*
3&4 vs. 5	<1	<1	1	2.09	

Reject the first three null hypotheses but not the fourth. If this had been a standard Bonferroni test we would have rejected only the first two null hypotheses.

12.11 Tukey's test on example in Table 11.2 :

recall

Tukey HSD[a]

conditio	N	Subset for alpha = .05	
		1	2
2.00	10	6.9000	
1.00	10	7.0000	
3.00	10		11.0000
5.00	10		12.0000
4.00	10		13.4000
Sig.		1.000	.429

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 10.000.

The counting and imagery groups are homogeneous, but are different from the adjective, intentional, and rhyming conditions, which are also homogeneous. This is the same pattern of differences that we found with the REGWQ.

For the Tukey test SPSS also produces the following table:

Multiple Comparisons

Dependent Variable: RECALL

Tukey HSD

(I) CONDTION	(J) CONDTION	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Counting	Rhyming	1.00E-01	1.39	1.000	-3.85	4.05
	Adjective	-4.00*	1.39	.046	-7.95	-4.77E-02
	Imagery	-6.40*	1.39	.000	-10.35	-2.45
	Intentional	-5.00*	1.39	.007	-8.95	-1.05
Rhyming	Counting	-1.00E-01	1.39	1.000	-4.05	3.85
	Adjective	-4.10*	1.39	.039	-8.05	-.15
	Imagery	-6.50*	1.39	.000	-10.45	-2.55
	Intentional	-5.10*	1.39	.006	-9.05	-1.15
Adjective	Counting	4.00*	1.39	.046	4.77E-02	7.95
	Rhyming	4.10*	1.39	.039	.15	8.05
	Imagery	-2.40	1.39	.429	-6.35	1.55
	Intentional	-1.00	1.39	.951	-4.95	2.95
Imagery	Counting	6.40*	1.39	.000	2.45	10.35
	Rhyming	6.50*	1.39	.000	2.55	10.45
	Adjective	2.40	1.39	.429	-1.55	6.35
	Intentional	1.40	1.39	.851	-2.55	5.35
Intentional	Counting	5.00*	1.39	.007	1.05	8.95
	Rhyming	5.10*	1.39	.006	1.15	9.05
	Adjective	1.00	1.39	.951	-2.95	4.95
	Imagery	-1.40	1.39	.851	-5.35	2.55

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

12.13 Games and Howell approach

I have organized the solution into a set of tables to keep some sort of order to the procedure.

(1) Matrix of mean differences

Group:		1	2	3	4	5
Mean:		10	18	19	20	29
1	10	--	8*	9*	11*	19*
2	18		--	1	3	11*
3	19			--	2	10*
4	21				--	8*
5	29					--

(2) Matrix of *df'*:

Group:	1	2	3	4	5
1	--	8	13.92	12.78	15.00
2		--	8.52	8.14	8.55
3			--	12.96	14.89
4				--	13.72
5					--

where

$$df' = \frac{\left[\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}\right]}{\frac{\left[\frac{s_i^2}{n_i}\right]^2}{n_i - 1} + \frac{\left[\frac{s_j^2}{n_i}\right]^2}{n_j - 1}}$$

(3) Matrix of $q(r, df)$

Group:	1	2	3	4	5
1	--	3.26	3.70	4.15	4.37
2		--	3.20	4.04	4.42
3			--	3.06	3.67
4				--	3.03
5					--

(4) Matrix of

$$\sqrt{\frac{\frac{s_i^2}{n_i}+\frac{s_j^2}{n_j}}{2}}$$

Group:	1	2	3	4	5
1	--	1.16	1.00	0.99	0.99
2		--	1.19	1.18	1.19
3			--	1.03	1.03
4				--	1.02
5					--

(5) Matrix of W_r = product of corresponding elements of (3) & (4):

Group:	1	2	3	4	5
1	--	3.78	3.70	4.11	4.33
2		--	3.81	4.77	5.26
3			--	3.15	3.78
4				--	3.09
5					--

The asterisks in (1) indicate the differences which are significant. Groups 1 and 5 differ from each other and all other groups.

I imagine your glad that you don't have to do that one again—I know I am.

12.15 Tukey's HSD test applied to the THC data in Table 11.5

Group:	1	2	3	4	5
µg THC	0	0.1	0.5	1	2
n_j	10	10	9	8	10

$$n_h = \frac{k}{\Sigma(\frac{1}{n_j})} = \frac{5}{\frac{1}{10}+\frac{1}{10}+\frac{1}{9}+\frac{1}{8}+\frac{1}{10}} = 9.326$$

Group	1	5	4	2	3	r	q_{HSD}	W_r
Means	34.00	38.10	18.50	50.80	60.33			
1	--	4.1	14.5	16.8	26.33*	5	4.04	20.51
5		--	10.4	12.7	22.23*	4	4.04	20.51
4			--	2.3	11.83	3	4.04	20.51
2				--	9.53	2	4.04	20.51
3					--			

$$w_r = q_{.05}(r, df)\sqrt{\frac{MS_{error}}{\overline{N}_h}} = 4.04\sqrt{\frac{240.35}{9.326}} = 20.51$$

The 0.5µg group is different from the control group and the 2µg group. No other differences are significant. The maximum familywise error rate is .05.

12.17 If you are willing to sacrifice using a common error term, you simply run the relevant *t* tests but evaluate them at $\alpha' = \alpha/c$.

12.19 Linear and quadratic trend in Conti and Musty (1984).

The results given below assume that you have added the three observations mentioned in the exercise.

Group:	**Control**	**0.1**	**0.5**	**1**	**2**	
Means	1.981	2.318	2.557	2.353	2.124	
Linear	-0.72	-0.62	-0.22	0.28	1.28	$\Sigma a_j^2 = 2.668$
Quadratic	0.389	0.199	-0.362	-0.612	0.387	$\Sigma b_j^2 = 0.846$

$$\psi_{Linear} = \Sigma a_j \overline{X}_j = -0.72(1.981) - 0.62(2.318) - 0.22(2.557)$$
$$+0.28(2.353) + 1.28(2.124) = -0.04846$$

$$\psi_{Quad} = \Sigma b_j \overline{X}_j = 0.389(1.981) + 0.199(2.318) - 0.362(2.557)$$
$$-0.612(2.353) + 0.387(2.124) = -0.31179$$

$$SS_{Linear} = \frac{n\psi^2}{\Sigma a_j^2} = \frac{10(-0.04846)^2}{2.668} = 0.0088$$

$$SS_{Quad} = \frac{n\psi^2}{\Sigma b_j^2} = \frac{10(-0.31179)^2}{0.846} = 1.1487$$

Source	*df*	SS	MS	*F*
Treatments	4	1.857	0.491	8.103*
Linear	1	0.0088	0.0088	0.145
Quadratic	1	1.1490	1.1487	18.987*
Error	45	2.726	0.0605	
Total	49	4.689		
Total	49	14491.22		

There is a significant quadratic trend, but no significant linear trend. This quadratic trend is clearly visible in the means.

12.21 Computer example.

12.23 Trend analysis for Epineq.dat separately at each interval.

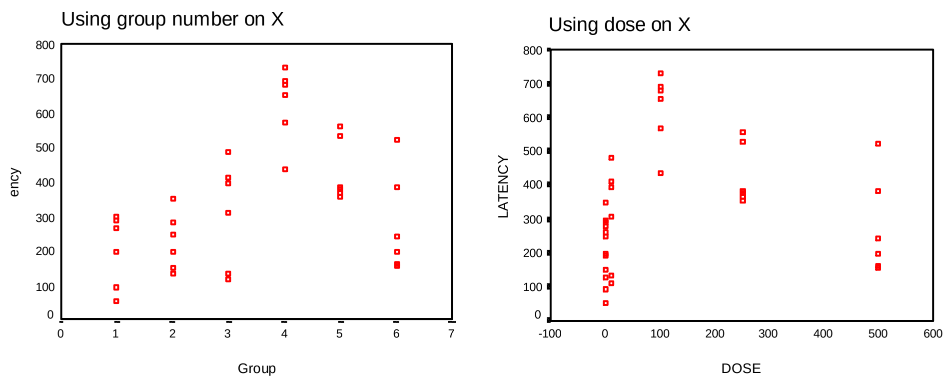
One Day: $F_{\text{Linear}} = 9.44$ (p = .0042); $F_{\text{Quad}} = 20.43$ (p = .0001)

One Week: $F_{\text{Linear}} = 4.33$ (p = .0453); $F_{\text{Quad}} = 13.23$ (p = .0009)

One Month: $F_{\text{Linear}} = 6.91$ (p = .0129); $F_{\text{Quad}} = 8.60$ (p = .0061)

12.25 Stone et al. (1992): Glucose and memory:

a.



b. Trend analysis using actual dose:

ANOVA

LATENCY

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		772106.472	5	154421.294	11.193	.000
	Linear Term	Contrast	7561.039	1	7561.039	.548	.465
		Deviation	764545.433	4	191136.358	13.855	.000
Within Groups			413869.833	30	13795.661		
Total			1185976.306	35			

c. Trend analysis using 1, 2, …6 as coding:

ANOVA

LATENCY

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		772106.472	5	154421.294	11.193	.000
	Linear Term	Contrast	152114.402	1	152114.402	11.026	.002
		Deviation	619992.070	4	154998.017	11.235	.000
Within Groups			413869.833	30	13795.661		
Total			1185976.306	35			

d. When we use the group number coding in our trend analysis we find a significant linear trend. As the dose of sucrose increases, memory increases accordingly.

e. The choice of coding system in not always obvious. Using 1, 2, … ,6 actually ranks the dose levels and ignores the fact that dose increases in an extreme way. (In other words, the difference between the first 2 doses is 1 mg/kg, whereas the difference between the last two doses is 250 mg/kg. Using 1, 2, …, 6 deliberately ignores this relationship. Apparently the human body responds in a nonlinear way to the increase in actual dose levels.

12.27 Effect sizes for contrasts in Exercise 12.1

$$d = \frac{\psi}{s_e} = \frac{\Sigma a_i \overline{X}_j}{s_e}$$

From the answers to Exercise 12.1:

$\psi_1 = 10.667$ $\psi_2 = -6$ $\psi_3 = -1$ $\psi_4 = -4$ $MS_{error} = 5.60$

$$d_1 = \frac{\psi}{s_e} = \frac{10.667}{\sqrt{5.60}} = \frac{10.667}{2.366} = 4.51$$

$$d_2 = \frac{\psi}{s_e} = \frac{-6}{\sqrt{5.60}} = \frac{-6}{2.366} = -2.54$$

$$d_3 = \frac{\psi}{s_e} = \frac{-1}{\sqrt{5.60}} = \frac{-1}{2.366} = -0.42$$

$$d_4 = \frac{\psi}{s_e} = \frac{-4}{\sqrt{5.60}} = \frac{-4}{2.366} = -1.69$$

All contrasts show large differences, with the smallest difference being about 4/10 of a standard deviation.

These are called standardized effect sizes because we divide the mean difference (the contrast) by the size of the standard deviation. This represents the difference in standard deviation units. An unstandardized effect would just be the difference between the means (or the means of sets of groups).

12.29 The study by Davey et al. (2003)

The group means are Negative mood =12.6, Positive mood = 7.0, No induction = 8.7

The SPSS ONEWAY solution with one contrast comparing the Negative and Positive mood groups is shown below.

ANOVA

Things listed to check

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	164.867	2	82.433	4.876	.016
Within Groups	456.500	27	16.907		
Total	621.367	29			

Contrast Coefficients

Contrast	Group		
	Negative	Positive	None
1	1	-1	0

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Things listed to check	Assume equal variances	1	5.6000	1.83888	3.045	27	.005
	Does not assume equal	1	5.6000	2.12498	2.635	13.162	.020

The contrast between the Positive and Negative mood conditions was significant ($t(27) = 3.045$, $p < .05$). This leads to an effect size of $d = \psi / \sqrt{MS_{error}} = 5.6 / \sqrt{16.907} = 5.6 / 4.11 = 1.36$. The two groups differ by over 1 1/3 standard deviations. It is evident that inducing a negative mood leads to more checking behavior than introducing a positive mood. (If we had compared the Positive and No mood conditions, the difference would not have been significant. However I had not planned to make that comparison.

12.31 **T**his requires students to make up their own example.