

# Stat 511, HW 3 Answers

Parts are worth one point, except where marked otherwise.

## 1. Orthogonal polynomials

(a) testing the hypothesis that all treatments have the same mean. There are 4 treatment, so we need  $3=4-1$  orthogonal polynomials besides the intercept column.

(b) Here is the orthogonal polynomials

c0	c1	c2	c3
1.00	-17.00	11.54	-13.93
1.00	-9.00	-5.07	29.01
1.00	3.00	-14.98	-18.57
1.00	23.00	8.50	3.48
1.00	-17.00	11.54	-13.93
1.00	-9.00	-5.07	29.01
1.00	3.00	-14.98	-18.57
1.00	23.00	8.50	3.48
1.00	-17.00	11.54	-13.93
1.00	-9.00	-5.07	29.01
1.00	3.00	-14.98	-18.57
1.00	23.00	8.50	3.48

(c) SS of orthogonal polynomials

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
c1	1	72.076	72.076	6.9906	0.02953 *
c2	1	0.005	0.005	0.0005	0.98305
c3	1	0.034	0.034	0.0033	0.95540
Residuals	8	82.483	10.310		

(d) Yes, they are the same

(e) The SSE is different. The sse for linear regression includes ss for intercept and the other two orthogonal polynomials.

(f) From the plot, the responds are linearly correlated with treatment. However, there is no enough indication of differences between four treatments.

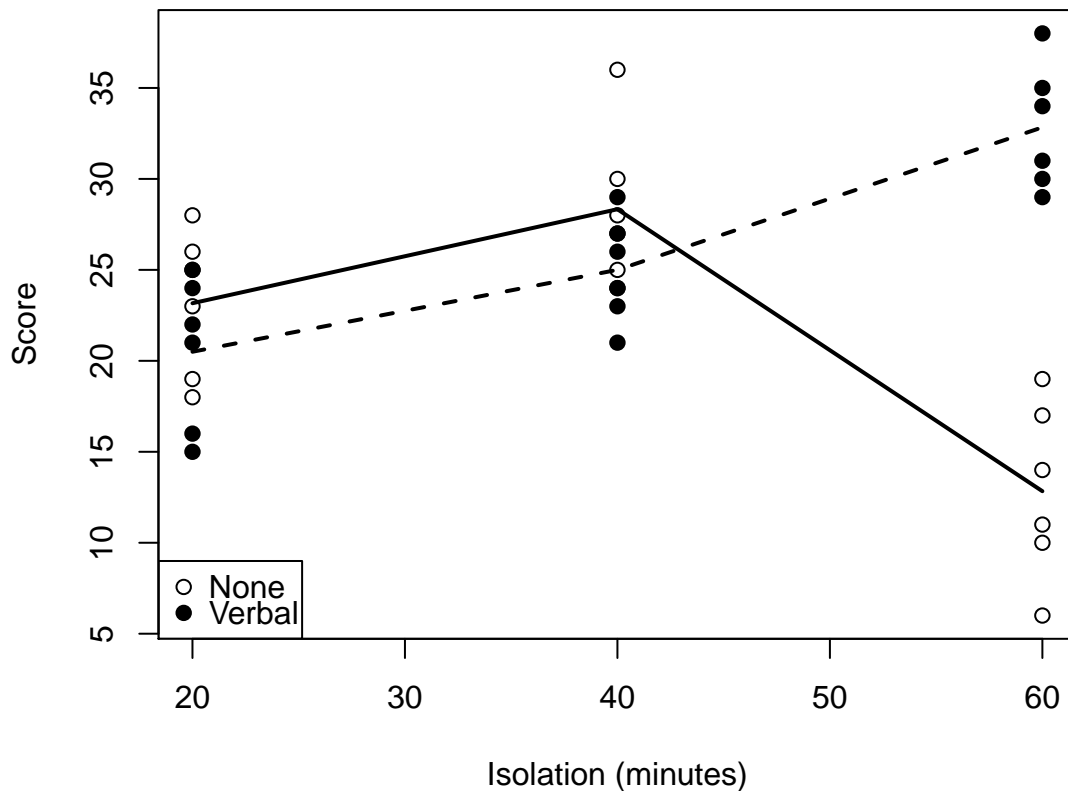
(g) The p-value from linear regression. Because the data shows strong linear relation between treatments and responds.

## 2. Children's memory

Note: sample sizes are the same for all groups.

- (a) The plot of means shows that the lines connecting the verbal reinforcement means and the non-verbal reinforcement means cross. This suggests that there is an interaction. There might be a main effect of isolation because the values at 40 minutes are larger than those at 20; it is hard to see what is happening on average at 60 min without computing the marginal means. The main effect of reinforcement is hard to see (slightly + at two isolation times, strongly - at the last).

Here's our graph, but one was not expected in your answer.



- (b) The p-values are: reinforce: 0.0014, isolation: 0.0139, and the interaction:  $< 0.0001$ .

Note: Here is the full ANOVA table.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
reinf	1	196.000000	196.000000	12.42	0.0014
isol	2	156.222222	78.111111	4.95	0.0139
reinf*isol	2	1058.666667	529.333333	33.55	<.0001

- (c) My contrasts and SS are:

	None			Verbal			SS
	20	40	60	20	40	60	
Reinforcement	1	1	1	-1	-1	-1	196.00
Isolation time	1	-1	0	1	-1	0	140.17
(2 df)	1	1	-2	1	1	-2	16.06
Interaction	1	-1	0	-1	1	0	0.67
(2 df)	1	1	-2	-1	-1	2	1058.00

If you don't round before adding, you get Reinforce: 196.00, Isolation: 156.22, Interaction: 1058.67

Note: The appropriate contrast coefficients depend on how you label the 6 groups. I numbered groups the reinforcement= 'none' groups as 1 through 3 and the reinforcement='verbal' as 4-6. The isolation = 20 groups were 1 and 4, =40 were 2 and 5 and =60 were 3 and 6.

- (d) A plot of  $x$ =predicted values vs  $y$  = residuals appears consistent with constant variance. There is no sign of any outliers. A normal quantile-quantile plot of the residuals exhibits a fairly straight line indicating the normal error assumption is met, and no outliers are present. The assumption of independence is reasonable since the students were randomly assigned to the treatments.
- (e) Using the pooled error and using contrasts for each pairwise difference gives you: at 20 min:  $p = 0.25$ , at 40 min:  $p = 0.16$ , at 60 min:  $p < 0.0001$

Here is the full table:

Time	SS	F	P
20 min	21.3	1.35	0.25
40 min	33.3	2.11	0.16
60 min	1200.0	76.06	0.0001

Note: it is hard to decide whether to use a multiple comparisons adjustment here. The three comparisons are orthogonal and we aren't doing too many, so I (and I believe) most folks would not adjust. We also accepted Bonferroni. Note: Tukey HSD is not appropriate because this is not the set of all pairs.

- (f) There are lots of possible answers here. Here is my summary. Some features to notice: I emphasize the magnitudes of effects, not the p-values. My choice of features to include in the summary depended on which effects were significant in the ANOVA table. I did not describe all pairs of means, but instead used the treatment structure to focus my summary. Other folks might emphasize different points; traditions vary across scientific disciplines.

On average, verbal reinforcement increases memory scores by 4.7 units, but this difference is not consistent across the isolation times. Considered separately at each isolation time, there is no evidence of an effect of reinforcement at 20 or 40 minutes, but verbal reinforcement increases retention by 20 units ( $p < 0.0001$ ) at 60 minutes. With verbal reinforcement, longer isolation time tends to enhance children's memory. Without verbal reinforcement, moderate isolation time (40 min.) leads to improved performance, but too much isolation time (60 min.) reduces children's performance.

The population is fourth graders from this specific city, but it is very common for the folks who collected these data to extrapolate (on non-statistical grounds) and make general conclusions about children.

### 3. Alpine plant growth

(a) 2 pts.

1) Averaged over the species, there is no evidence of a difference in yield between sites ( $F = 0.67$ ,  $p = 0.52$ ).

2) The effect of site is very different in the two species ( $F = 16.6$ ,  $p < 0.0001$ ).

Note: The first question is addressed by the Type III SS and F test for sites. The second is addressed by the interaction SS and F test. Remember interactions concern equality of simple effects.

(b) 1 pt. Background: A quick look at the plot should indicate why there is an interaction. As you move from dry to wet sites, the mean log biomass of *P. gracilis* increases while the mean log biomass of *P. nivea* decreases. The increase in one species cancels the decrease in the other, so the average of the two species stays approximately constant.

One possible answer is: “The two species differ in their responses to site (Interaction  $F = 16.6$ ,  $p < 0.0001$ ). As you move from dry to wet sites, the mean log biomass of *P. gracilis* increases while the mean log biomass of *P. nivea* decreases.

Note: This is an interpretation question, so just about anything reasonable is acceptable. In journals that expect a p-value for every claim, then you would add p-values from the analysis in part d. In my opinion, it is appropriate to omit some p-values. Just make sure your interpretation is supported by the data.

(c) Contrasts and splitting the data test the same hypothesis. They use different error variances. The MSE for contrasts is pooled over both species; the MSE for splitting is the species-specific MSE.

(d) If two parts of the data (e.g. two species) have very different error variances, then two separate analyses are better. You get information about the MSE for each species from the ‘by spp’ analysis. They are  $MSE=0.205$  for *gracilis* and  $MSE=0.308$  for *nivea*. These are similar. I would use the contrast analysis. If the variances differ by more than 10 fold, then I am worried about the effects of unequal variances.

Note: You could also look at equality of variances by fitting the 2 way ANOVA and constructing box plots of residuals in the two species. The usual plot of residual vs predicted value isn’t helpful here because it look at a different cause of unequal variances.

(e) 2 pts. Either by contrasts or splitting, you see strong evidence of site effects on each species individually. For *gracilis*,  $F=5.75$   $p=0.0060$  by contrasts or  $F=7.23$ ,  $p=0.0039$  by considering one species at a time. For *nivea*,  $F=13.53$ ,  $p<0.0001$  by contrasts and  $F=11.32$ ,  $p=0.0004$  by considering one species at a time.

Notes: In this case, the conclusions (from tests or from estimates and means) are similar. This may not be so. If the two error variances were very different, there is a cost to

pooling. If the error d.f. are small, there is a cost to separating the species (larger critical values).

Questions like these are intended to mimic what you will have to do when you analyze your own data and write up the results. Often, you will have to choose among various ways to analyze data. The important things are: 1) you can't show all possibilities, 2) that you can justify your choice. Those thoughts motivated our grading. Did you choose a method and justify that choice? Lost points if you didn't choose (i.e. gave both methods) or weren't consistent with your choice in 2c. Note: I try to make it clear when I want you to choose like this. The question explicitly said to choose and report one method, not both.

- (f) 2 pts . In this case, the conclusions (from tests or descriptions of pattern) are similar. Basically, the choice of missing value or 2 doesn't make much difference for these data.

Notes: Sometimes it does matter which approach you choose for missing observations. It really matters when the number of missing values responds to the treatment. I have seen studies where the mean size is larger in a supposedly stressful environment. Most plants died in that stressful environment and the lucky survivors got very large. There, the treatment of the dead plants really changed the conclusions.