Math 356H Assignment #4 Solutions

1. Chapter 13, #53

   (a) We transform the model as

   $$\ln(Q) = Y = \ln(\alpha) + \beta \ln(a) + \gamma \ln(b) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon'$$

   with the appropriate definitions.

   So taking the logarithm of each variable and performing the multiple regression we get $\hat{\beta}_0 = 1.5652$, $\hat{\beta}_1 = .9450$, $\hat{\beta}_2 = .1815$. For $a = 10$ and $b = .01$, $x_1 = \ln(10) = 2.3026$ and $x_2 = \ln(.01) = -4.6052$, from which $\hat{y} = 2.9053$ and $\hat{Q} = e^{2.9053} = 18.27$.

   (b) Again taking the natural log, $Y = \ln(Q) = \ln(\alpha) + \beta a + \gamma b + \ln(\varepsilon)$, so to fit this model it is necessary to take the natural log of each Q value (and not transform a or b) before using multiple regression analysis.

   (c) We simply take the exponential at each endpoint: $(e^{.217}, e^{1.755}) = (1.24, 5.78)$.

2. Chapter 13, #58

   In the first step of the backward elimination method, the variable `sumrfib` was eliminated because it had the smallest t-ratio that was in absolute value less than 2. In step 2, variable `spltabs` was eliminated because its t-ratio was the smallest, and still smaller than 2 in absolute value. In step 3, the variable `sprngfib` was eliminated because its t-ratio was the smallest, and still smaller than 2 in absolute value. After that, no variable was eliminated because all t-ratios are larger than 2 in absolute value. The variables kept in the final model are `%sprwood` and `sumltabs`.

   In the forward selection, variable `%sprwood` is added in the first step, with the highest t-ratio in absolute value greater than 2. Next, `sumltabs` is added with the highest t-ratio greater than 2. No more variables are added because there is no t-ratio larger than 2 in absolute value. As before, the variables kept in the final model are `%sprwood` and `sumltabs`, so that both procedures agree.

3. Refer to the data in Chapter 10, #6

   (a) What is the response variable? What is the factor?
   Response: total Fe; factor: type of iron formation.

   (b) How many levels of the factor are being studied?
   Four

   (c) Check and comment on the ANOVA assumptions for this problem:

   i. Is there any reason to believe that errors are not independent?
   Since we do not know the order in which observations were taken, we have no way of checking independence. We must believe that it is valid.

   ii. Does total FE look normally distributed for each of the factors?
   The P-values for normality tests for all four factor levels are:

Factor level	1	2	3	4
P-value	.113	.210	.075	.293

   At $\alpha = .05$, there is no evidence that the data are not normally distributed. We can also see that the boxplots look reasonably symmetric, with at most one possible outlier. Silicate has 2 possible outliers, so it would be the most questionable case; we saw that the formal test fails to reject, so it can be considered normal as well.

   iii. Calculate the sizes of the sample variances and do a visual check of the data by looking at boxplots of the data to see whether the spread in each sample looks about the same.

Factor level	1	2	3	4
s^2	11.49	19.62	8.15	23.33

   The assumption of constant variance does not seem to be satisfied.

(d) Perform the ANOVA test.

Analysis of Variance

Source	DF	SS	MS	F	P
Treatment	3	509.1	169.7	10.85	0.000
Error	36	563.1	15.6		
Total	39	1072.3			

With P-value of .000, we reject H_0 of equality of means in all treatments.

(e) If applicable, use Tukey's procedure to identify differences.

We use

$$w = Q_{\alpha,I,IJ-I}\sqrt{MSE/J} \approx 3.79 \cdot \sqrt{15.6/10} = 4.73$$

Then Tukey's procedure would result in:

Treatment	silicate	carbonate	magnetite	hematite
Mean	24.69	26.08	29.95	33.34
	A	A		
		B	B	
			C	C

4. The following partial ANOVA table is taken from the article "Perception of Spatial Incongruity" in which the abilities of three different groups to identify a perceptual incongruity were assessed and compared. All individuals in the experiment had been hospitalized to undergo psychiatric treatment. There were 21 individuals in the depressive group, 32 individuals in the functional "other" group, and 21 individuals in the brain-damaged group. Complete the ANOVA table and carry out the F test at level $\alpha = 0.01$.

Source	df	sum of squares	Mean square	F	P
Groups	**2**	**152.18**	76.09	**5.564**	.0057
Error	**71**	**970.96**	**13.675**		
Total	**73**	1123.14			

We reject H_0 of equality of means.

5. A chemical engineer is studying a newly developed polymer to be used in removing toxic wastes from water. Experiments are conducted at five different temperatures. The response noted is the percentage of impurities removed by the treatment:

Temperature				
I	II	III	IV	V
40	36	49	47	55
35	42	51	49	60
42	38	53	51	62
48	39	53	52	63
50	37	52	50	59
51	40	50	51	61

(a) Test the hypothesis of equal treatment means.

Analysis of Variance

Source	DF	SS	MS	F	P
Treatment	4	1535.5	383.9	32.64	0.000
Error	25	294.0	11.8		
Total	29	1829.5			

We reject the hypothesis of equal treatment means.

(b) Use Tukey's procedure to compare all possible pairs of means.

We use use

$$w = Q_{\alpha,I,IJ-I}\sqrt{MSE/J} \approx 4.17 \cdot \sqrt{11.8/6} = 5.65$$

Then Tukey's procedure would result in:

Treatment	II	I	IV	III	V
Mean	38.67	44.33	50	51.33	60
	A	A			
		B	B		
			C	C	
					D

6. Chapter 10, #17. Keep in mind that if all sample sizes are equal, MSE is just the average of the variances (see p.407).

$\theta = \sum_i c_i\mu_i$, with $c_1 = c_2 = .5$, $c_3 = -1$, so that $\hat{\theta} = .5\bar{x}_{1\cdot} + .5\bar{x}_{2\cdot} - \bar{x}_{3\cdot} = .5(1.63) + .5(1.56)$ - $(1.42) = .175$. We need MSE to evaluate the estimated standard deviation of the contrast: $MSE = (.27^2 + .24^2 + .26^2)/3 = .0660$, so that $\sqrt{MSE(.5^2 + .5^2 + 1)/10} = .3147$. Since $t_{.025,27} = 2.053$, a 95% CI for the contrast is $.175 \pm 2.053(.3147) = (-.029, .379)$. Since the interval includes 0, we cannot reject the null hypothesis that the contrast is zero.