Engineering Statistics II, HW 6

Due Start of class, Monday, Oct. 20

Instructions: This homework is due at the beginning of class Monday, Oct 20. A subset of these four problems will be graded.

Additional Instructions: Be sure to include copies of all relevant Minitab output in your hw. For additional instructions see those for hw3 which is posted on the course website.

1: A company operates two production lines for making soap bars. For each line, the relation between the speed of the line and the amount of scrap for the day was studied. The relationship between speed and amount of scrap appears to be approximately linear for both machines. A formal test is desired to determine whether or not the relationship between scrap amount and line speed is the same for both machines.

i. Download the data set for this problem from the following URL:

`www.rose-hulman.edu/~inlow/soap_line.MTW`

ii. Devise a full or complete model with sufficient flexibility to allow a different linear relationship (both slope and intercept) between line speed and scrap for each production line. Your model should have four coefficients.

iii. Use a **single nested model F test** to determine if any differences exist between the relationship of scrap and line speed for the two production lines. Use $\alpha = 0.05$. Comment on all assumptions that are necessary for your test to be valid and check all the assumptions you can. Show your work for credit.

iv. If you reject in part iii, conduct further analyses to characterize the nature of the differences. For example, determine if interaction is present.

2: In class we discussed the fact the one-way ANOVA F-test for comparing means across k populations/treatments is simply the nested model F-test for testing H_0: $\beta_1 = \cdots = \beta_{k-1} = 0$ vs. H_1: *at least one β is nonzero*. In this problem you will compare the mean vibration level for three different electric motor bearings using Minitab's regression and ANOVA procedures as follows:

i. Download the bearing data from the following URL:

`http://www.rose-hulman.edu/~inlow/bearing.MTW`

ii. See next page...

ii. Construct a regression model using indicator variables to represent the different levels of the categorical variable *Brand*. Let *Brand1* be the reference category.

iii. Assuming the full model meets all regression assumptions (do not check them), test the hypotheses mentioned above using a nested model F-test. At $\alpha = 0.05$ what do you conclude? Show your work for credit.

iv. Assuming all ANOVA assumptions are met (do not check them), test the ANOVA hypotheses H_o: $\mu_1 = \mu_2 = \mu_3$ vs. H_a: *at least two differ* using Minitab's `Stat -> ANOVA -> Onewa`. You F-test value, degrees of freedom, and p-value should agree with your result in part iii.

v. One advantage of analyzing the data using regression instead of a special-purpose ANOVA procedure is that you can determine which means differ by testing the regression coefficients. Using t-tests of the coefficients in your regression model in part ii, which of the means actually differ at $\alpha = 0.05$?

3: (Example 13-2.1 in your textbook) In this example you will analyze data from an experiment to investigate the relationship between the strength of paper grocery bags and the hardwood concentration in the pulp. The goal was to determine how the strength of the bags increased as the hardwood concentration was increased. Was there, for example, a point of diminishing return? The engineering team decided to investigate four levels of hardwood concentration: 5%, 10%, 15%, and 20%. Six specimens from each concentration level were tested. The resulting specimens were tested on a lab tensile strength tester in random order. The data is provided by the following URL

```
www.rose-hulman.edu/~inlow/hardwood.MTW
```

Analyze this data as follows:

i. Note that although hardwood concentration is interpretable as a continuous variable, since it has only 4 levels we can also treat it as a categorical variable. Do so by constructing indicator variables for 3 of the 4 concentrations and stacking the corresponding strength measurements to create the response variable. Fit the corresponding complete regression model containing all three indicator variables and check the regression assumptions. Do you feel they are met?

ii. Test the ANOVA hypotheses H_0: $\mu_5 = \mu_{10} = \mu_{15} = \mu_{20}$ vs. H_1: *at least two means differ* by testing H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ vs. H_1: *at least one of the β's is nonzero* using a nested model F test. What is the resulting F statistic and p-value? At $\alpha = 0.05$ what do you conclude? Show you work for credit.

iii. Verify your answer in part ii by using Minitab's ANOVA procedure `Stat -> ANOVA -> One-way`.

4: See next page...

4: In problems 2 and 3 you used nested model F tests to determine the significance of a categorical predictor variable. Nested model F tests have other uses. For example they are used to test if any of the second order terms in a second order polynomial model are signficant. In other words, suppose we fit the full or complete second order model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \epsilon$$

and want to know if any of the second order terms are needed or if a linear model will suffice. We can answer this question by testing the significance of the second order tems using a nested model F test of

H_0: $\beta_3 = \beta_4 = \beta_5 = 0$ vs.
H_1: *at least one* β *nonzero*

Test these hypotheses using a nested model F test as follows:

i. Open the data set `resp_surf.MTW -- response surface example data set` at the bottom of the course website. We analyzed this data previously.

ii. Recall that before doing a nested model F test you need to verify that the complete model meets the regression assumptions. Since we previously analyzed this data and verified the complete model meets the regression assumptions, do not check them.

iii. Since the complete model meets the regression assumptions, use a nested model F test to test the hypotheses above at $\alpha = 0.05$. What do you conclude? Show your work for full credit. Helpful hint: You don't need to standardize your predictors in order to do your F test in this problem.