**DATA 557**

**Homework Assignment 6**

Submissions are due by **5:00pm, Wednesday February 20.**

This homework is an individual (not group) assignment. You may work together to help each other solve problems, but you should do all the work, create your own solutions, and hand in your own work without copying others’ work.

It is not necessary to include R code with your solutions for this assignment.

Data set: ‘cells.csv’

Summary: a randomized clinical trial of immune cell stimulation in 40 patients

Variables:

id: subject id #

dose: drug dose (0, 10, or 100mg)

sex: sex (0=female, 1=male)

age: age (yrs)

count0: pre-treatment cell count

count1: post-treatment cell count (the response variable)

1. Use ANOVA to test for a difference between mean post-treatment cell count between dose groups. Is there evidence for an effect of dose on post-treatment cell count?

2. Use linear regression to assess the effect of dose on post-treatment cell count? Is there evidence for an effect using regression? Give an interpretation in words of the estimated coefficient for dose. Compare the results using ANOVA and linear regression.

3. Add the variable sex to the ANOVA and linear regression models. Describe how the results change when the variable sex is added to the model.

4. Using ANOVA and linear regression, test for interaction between sex and dose. State the interpretations of the coefficients in the linear regression model with interaction. Give a graphical display of this linear regression model that shows the relationship between dose and response for males and females.

**Data set for Questions 5-7**: ‘Sales.csv’

Variables:

LAST\_SALE\_PRICE: the sales price of the home

SQFT: area of the house (sq. ft.)

LOT\_SIZE: area of the lot (sq. fit.)

BEDS: number of bedrooms

BATHS: number of bathrooms

5. Use scatterplots to display the associations between sales price and each of the following predictors: SQFT, LOT\_SIZE, BEDS and BATHS. Calculate Pearson correlation coefficients for each plot. Describe the associations in terms of linearity, strength of association and whether the association is positive or negative.

6. Use separate linear regression models to assess the association between sales price and each of the four predictor variables (a separate model for each predictor). Interpret the estimated regression coefficients for each model.

7. Fit a linear regression model with all 4 predictor variables included. Describe how the estimated coefficients for the predictors change compared to their values in the separate models. Compare the R-squared values for all of the models.