Assignment 4, due March 31, 2020 at beginning of class

(Out of 40 points)

1. (10 points) Data from a sample of 10 pharmacies are used to examine the relationship between prescription sales volume and the percentage of prescription ingredients purchased directly from the supplier. The sample data are shown below:

Pharmacy	Sales Volume, y (in $10,000)	% Ingredients Purchased directly, x
1	25	10
2	55	18
3	50	25
4	75	40
5	110	50
6	138	63
7	90	42
8	60	30
9	10	5
10	100	55

The data summaries are $\bar{y} = 71.3$, $\bar{x} = 33.8$, $SXY = 6714.6$, $SXX = 3407.6$, $SYY = 13882.1$.

(2)   (a) Calculate Pearson's correlation coefficient r.

(1)   (b) What proportion of the variation in Sales is explained by % Ingredients purchased directly? (Hint: it is r^2).

(c) Use the method of least squares to find:

(3)   i. the slope estimate $\hat{\beta}$.

(3)   ii. the intercept estimate $\hat{\alpha}$.

(1)   iii. write the equation of the regression line

(Hint: please refer to page 12 of the course note (lecture 8) to get the $\hat{\beta}$ and $\hat{\alpha}$. Please also see the example on pages 17 to 19.)

2. (15 points) A study involving 42 subjects found that bone mineral density (BMD) in g/cm^2, measured at the left femural neck, was related to Weight (in kg) according to the least squares equation

$$BMD = 0.47 + .0049 Weight$$

(1)   (a) What is the predicted BMD for a subject with weight 80kg?

(2)   (b) What increase in BMD is expected with an increase of weight of 5kg?

(2)   (c) What weight is predicted with a BMD of .8 g/cm^2?

(d) Assess the hypothesis that there is no association between Weight and BMD. Use the fact that the standard error of the slope estimate is 0.002.
(Hint: please refer to page 14 and 15 of lecture 8 to do the hypothesis test and calculate the confidence interval; for calculating 95% confidence interval, the $t_{\alpha/2, n-2}$ term in the formula is the t-value from the t−table corresponding to .95 confidence interval and $n - 2$ degrees of freedom).

(2)   i. State the hypotheses.

(2)   ii. Calculate the test statistic.

(1)   iii. What is the degrees of freedom?

(1)   iv. Determine the P value as accurately as possible using the tables on the class web site.

(4)   (e) Calculate a 95% confidence interval for the slope coefficient.

3. (15 points) Investigators wish to assess whether there is a difference in the efficacy of salbutamol and ipratropium bromide in the treatment of asthma. They will measure forced expiratory volume in 1 second (FEV_1) after two weeks of treatment. They wish to detect a difference of 0.2 liters with a two-sided alternative using $\alpha = .05$. Assume that the SD is 1.0 liter in both groups. (Note: for a two-sided alternative test with $\alpha = .05$, use $z_{\alpha/2} = 1.96$; for having a 80% power (i.e., $\beta = .2$), use $z_\beta = 0.84$; and for having a 90% power (i.e., $\beta = .1$), use $z_\beta = 1.28$ in calculation).

(5)   (a) What would be the power of the test if they included 100 subjects in each group?

(5)   (b) How large a sample in each group should they take to obtain 80% power?

(5)   (c) How large a sample in each group should they take to obtain 90% power?