STA674: Regression Analysis and Design of Experiments
Assignment #1

Submission:

You must format your assignments as a pdf. Handwritten assignments will not be accepted.

When are ready to submit your assignment, copy your R (or RStudio work) or SAS code and paste it at the end of your document. *Don't forget to add comments to help the grader follow your work.* Collaboration during the process of solving the problems is not only allowed but encouraged; that said, the submissions are each expected to be an individual effort reflecting the individual's work. Identical submissions or even submissions found to be **"too close to be coincidental" will be flagged and given no credit** (1 warning, then enforcement.)

Each problem is worth 4 points, for 20 points total. Longer problem parts are worth more. Please submit it to your instructor **by the due date on Canvas** via electronic submission to Canvas.

Questions

1. In many cases, statistical methods are used to estimate the relationship between overhead (y) and the level of production (x) using historical data. As a simple example, consider the historical data for a certain plant:

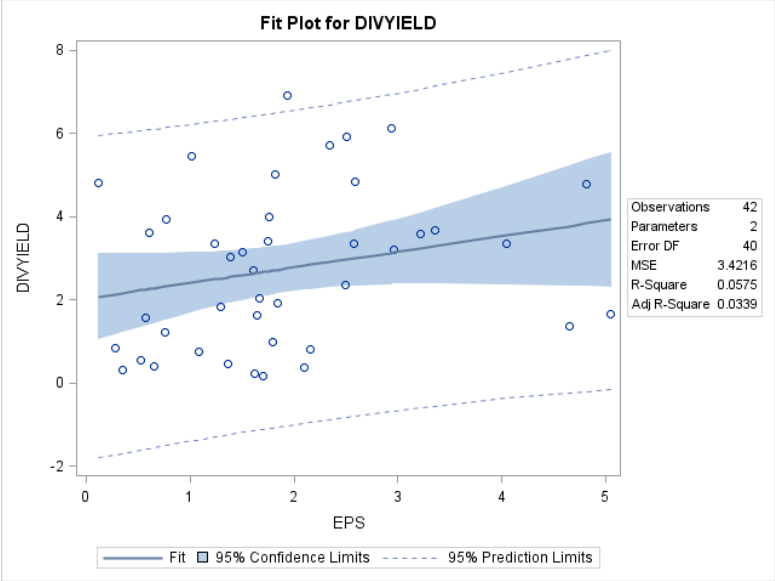
Production (in 10,000) units:	5	6	7	8	9	10	11
Overhead costs (in $1000):	12	11.5	14	15	15.4	15.3	17.5

   a. Construct a scatterplot of y versus x and graph the regression line on the scatterplot.

   b. Find the least-squares regression line relating overhead costs to production and provide an interpretation of the fitted line and b1.

   c. Compute the standard errors of b_0 and b_1.

2. Dividends A random sample of 42 firms was chosen from the S&P 500 firms listed in the Spring 2003 Special Issue of Business Week (The Business Week Fifty Best Performers). The dividend yield (DIVYIELD) and the 2002 earnings per share (EPS) were recorded for these 42 firms These data are in a file named DIV3. Using dividend yield as the dependent variable and EPS as the independent variable, a regression was run. Use the results to answer the questions. The scatterplot of DIVYIELD and EPS and regression results are shown below.



Number of Observations Used	42

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8.34483	8.34483	2.44	0.1262
Error	40	136.86356	3.42159		
Corrected Total	41	145.20839			

Root MSE	1.84975	R-Square	0.0575
Dependent Mean	2.75048	Adj R-Sq	0.0339
Coeff Var	67.25213		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	2.03364	0.54052	3.76	0.0005	0.94121	3.12607
EPS	1	0.37396	0.23946	1.56	0.1262	-0.11000	0.85791

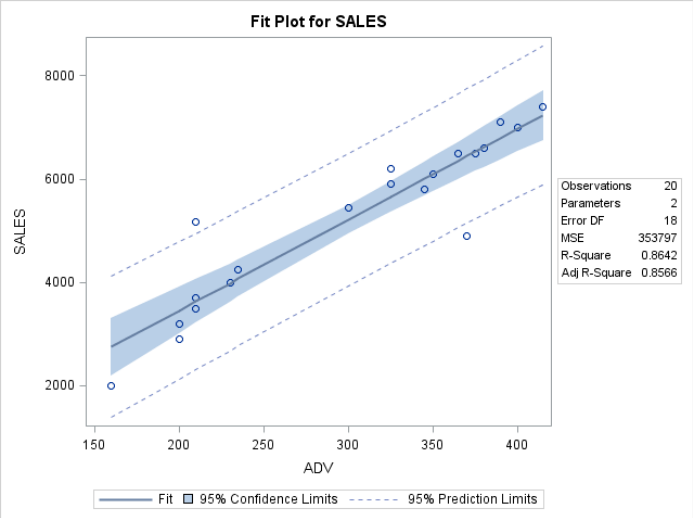
a. What is the sample regression equation relating dividends to EPS?

b. Is there a linear relationship between dividend yield and EPS? Use $\alpha = 0.05$. State the hypotheses to be tested, the decision rule, the test statistic, and your decision. What conclusion can be drawn from the test result?

c. Construct 95% confidence interval estimates of β_0 and β_1.

3. The vice-president of marketing for a large firm is concerned about the effect of advertising on sales of the firm's major product. To investigate the relationship between advertising and sales, data on the two variables were gathered from a random sample of 20 sales districts. These data

are available in a file named SALESAD3. (Sales and advertising are both expressed in hundreds of dollars. For example, 4250 represents $425,000). The scatterplot and results for the regression of sales (SALES) on advertising (ADV) are below.



Number of Observations Used	20

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	40523671	40523671	114.54	<.0001
Error	18	6368342	353797		
Corrected Total	19	46892014			

Root MSE	594.80820	R-Square	0.8642
Dependent Mean	5209.25000	Adj R-Sq	0.8566
Coeff Var	11.41831		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-57.28092	509.75030	-0.11	0.9118
ADV	1	17.56974	1.64168	10.70	<.0001

Using the information given, answer the following questions:

a. Is there a linear relationship between sales and advertising? Use $\alpha = 0.05$. State the hypotheses to be tested, the decision rule, the test statistic, and your decision. What implications does the test result have for the firm?

b. What is the sample regression equation relating sales to advertising? Provide an interpretation of the fitted line.

c. If ADV increases by $1000, what would be the resulting change in our prediction of SALES?

On 4 & 5: remember from the lectures—"point estimate" just means give a single number that is your best guess. Other terms that mean the same thing: "point prediction", and "plug it into the regression equation" ☺. The important notion in these exercises is not the value from the regression equation (you could do that in elementary school)—the important idea is finding/understanding the accompanying statistical variability (variability we attribute to having measured a sample rather than the entire population.)

4. Refer to exercise 1.

   a. Find a point estimate as well as a 95% confidence interval estimate of the overhead costs, on average, for production runs of 80,000 units.

   b. Find a point prediction of the overhead costs as well as a 95% prediction interval for a single production run of 80,000 units.

   c. Comment on the portions of the answers to parts a. and b. that remained the same and that changed and why.

5. Refer to exercise 2.

   a. Find a point prediction of the dividend yield as well as a 95% prediction interval for a single sampled firm with $2 earnings per share.

   b. Find a point prediction of the dividend yield as well as a 95% prediction interval for a single sampled firm with $5 earnings per share.

   c. Comment on the portions of the answers to parts a. and b. that remained the same and that changed (accuracy *and* precision) and why.