1. Short Answer

Note: Each part is unrelated.

a. List the three major assumptions of a simple linear regression model. For each assumption, give one way that you might check whether that assumption is satisfied and identify one remedy that might be used to adjust for the problem if it exists.

Assumption #1:

Diagnostic:

Remedy:

Assumption #2:

Diagnostic:

Remedy:

Assumption #3:

Diagnostic:

Remedy:

b. You are designing an experiment in which you want to relate corn yield (Y, in bushels per acre) to the amount of fertilizer used on the plot (X, in pounds per acre). Based on previous experiments, you believe that the ideal fertilizer amount is between 10 and 12 pounds per acre. You would like to obtain prediction intervals at $X = 10.5$, $X = 11$, and $X = 11.5$. (i) In terms of experimental design, explain what you could do to try to minimize the widths of these intervals. (ii) Explain what you would do to obtain these three intervals with a *joint confidence region* of *at least* 95%.

2. In a simple linear regression problem with n = 62, the following estimates were obtained: $b_0 = 50$, $b_1 = 6$, $s\{b_1\} = 5$, MSE = 16, $\bar{X} = 3$.

a. Write the simple linear regression model. Include the distributional assumptions.

b. Write the estimated regression line.

c. Predict the value of Y when X = 3.

d. Calculate the estimated standard deviation of the error term in the model.

e. Give a 95% confidence interval for the slope of the regression line.

f. Give a 95% prediction interval for a new observation Y when X = 3.

3. Refer to the SAS output marked OUTPUT FOR PROBLEM 3. The data are from a study of company executives. The response variable is annual `salary` (dollars), and the two explanatory variables used are `gender` (0 = female, 1 = male) and `exper` (experience in years). The variable `expgen` is the product of `gender` and `exper`.

a. Write down the linear model used in this analysis, including the distributional assumption.

b. Write the estimated regression equation. Then, write two separate fitted lines predicting `salary` from `exper`: one for females and one for males.

c. Estimate the mean salary for women with 5 years of experience.

d. How do you justify that either gender or experience or their product, or any combination of those three variables is useful to predict salary?

e. Does residual show anything strange? Explain why or why not.

OUTPUT FOR PROBLEM 3

The REG Procedure

Dependent Variable: salary

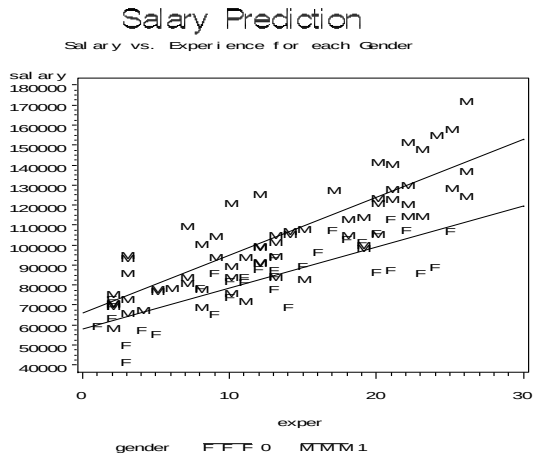
Analysis of Variance

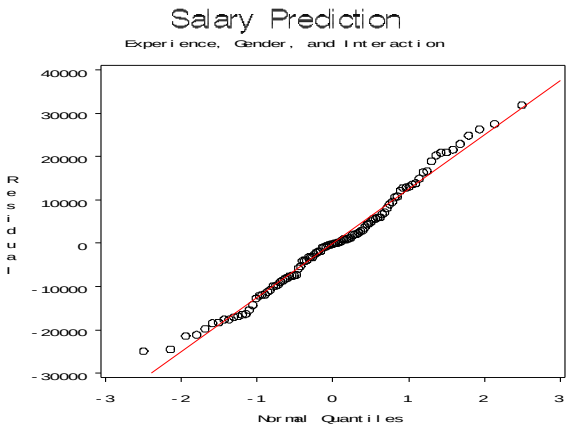
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	47463712734	15821237578	98.09	<.0001
Error	96	15484034766	161292029		
Corrected Total	99	62947747500			

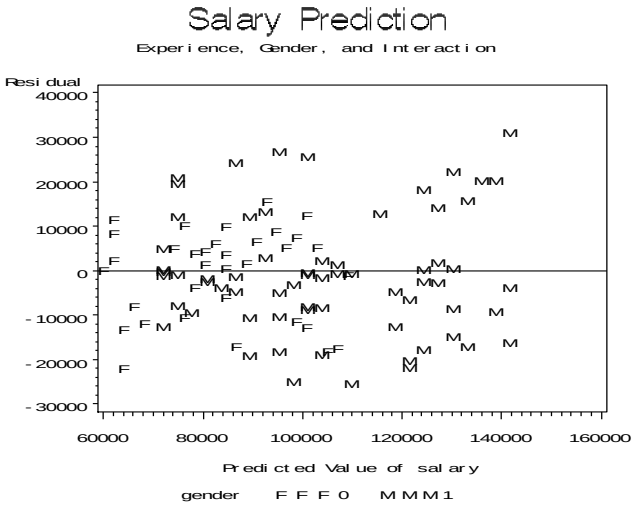
Root MSE	12700	R-Square	0.7540
Dependent Mean	97535	Adj R-Sq	0.7463
Coeff Var	13.02105		

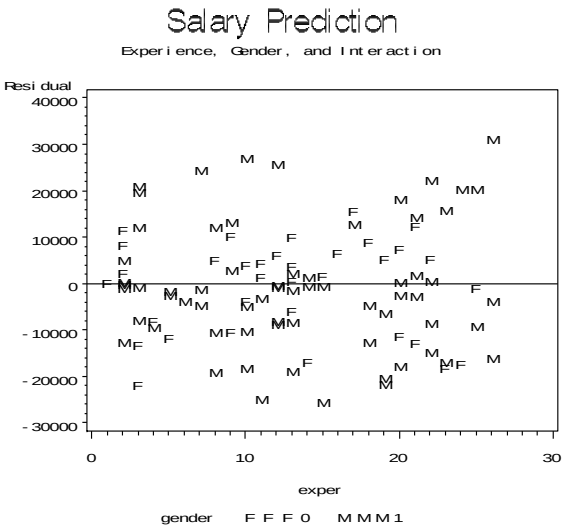
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS	Variance Inflation
Intercept	1	58050	4461.17930	13.01	<.0001	9.513076E11	27309548988	0
gender	1	7799	5497.47020	1.42	0.1593	9651865067	324571036	4.20471
exper	1	2045	308.56451	6.63	<.0001	36949216648	7081290125	3.15069
expgen	1	864	373.65327	2.31	0.0229	862631019	862631019	6.58957









4. Suppose we have an incomplete ANOVA table in studying the linear model of predicting Salary by quality, experience and publication.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	[ ]	[ ]	[ ]	[ ]
Error	[ ]	61.44300	[ ]		
Corrected Total	23	689.26000			

a. Complete the ANOVA table.

b. Write down the linear model and the hypothesis for the *F* test. What is your conclusion from this test?

Suppose that we miss the last three columns, `t Value`, `Pr > |t|`, `95% Confidence Limits`, in the following table.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits
Intercept	1	17.84693	2.00188			
quality	1	1.10313	0.32957			
experience	1	0.32152	0.03711	[ ]	[ ]	[,]
publications	1	1.28894	0.29848			

c. Fill in the blanks for experience based on the information from the ANOVA table and the above table.

d. Write down the hypothesis for the *t* test in c. What is your conclusion from this test?