Part 1: 15 points

Refer to **Copier maintenance** Problem Assume that linear regression through the origin model is appropriate.

Obtain the estimated regression function.

Estimate β_1 with a 90 percent confidence interval. Interpret your interval estimate.

Predict the service time on a new call in which six copiers are to be serviced. Use a 90 percent prediction interval.

Plot the fitted regression line and the data. Does the linear regression function through the origin appear to be a good fit here?

Obtain the residuals e_i. Do they sum to zero? Plot the residuals against the fitted values $\hat{Y}_i$. What conclusions can be drawn from your plot?

Conduct a formal test for lack of fit of linear regression through the origin; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Part 2: 15 points

When joint confidence intervals for β_0 and β_1 are developed by the Bonferroni method with a family confidence coefficient of 90 percent, does this imply that 10 percent of the time the confidence interval for β_0 will be incorrect? That 5 percent of the time the confidence interval for β_0 will be incorrect and 5 percent of the time that for β_1 will be incorrect? Discuss.

Refer to **Copier maintenance** Problem

Will b_0 and b_1 tend to err in the same direction or in opposite directions here? Explain.

Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 95 percent family confidence coefficient.

A consultant has suggested that β_0 should be 0 and β_1 should equal 14.0. Do your joint confidence intervals in part (b) support this view?

Part 3: 15 points

Typographical errors. Shown below are the number of galleys for a manuscript (X) and the total dollar cost of correcting typographical errors (Y) in a random sample of recent orders handled by a firm specializing in technical manuscripts. Since Y involves variable costs only, an analyst wished to determine whether regression-through-the-origin model (4.10) is appropriate for studying the relation between the two variables.

i:	1	2	3	4	5	6	7	8	9	10	11	12
X_i:	7	12	10	10	14	25	30	25	18	10	4	6
Y_i:	128	213	191	178	250	446	540	457	324	177	75	107

Fit regression model (4.10) and state the estimated regression function.

Plot the estimated regression function and the data. Does a linear regression function through the origin appear to provide a good fit here? Comment.

In estimating costs of handling prospective orders, management has used a standard of \$17.50 per galley for the cost of correcting typographical errors. Test whether or not this standard should be revised; use $\alpha = .02$. State the alternatives, decision rule, and conclusion.

Obtain a prediction interval for the correction cost on a forthcoming job involving 10 galleys. Use a confidence coefficient of 98 percent.

Obtain the residuals e_i. Do they sum to zero? Plot the residuals against the fitted values $\hat{Y}_i$. What conclusions can be drawn from your plot?

Conduct a formal test for lack of fit of linear regression through the origin; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Part 4: 20 points

Sales growth. A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data are as follows, where X is the year (coded) and Y is sales in thousands of units:

i:	1	2	3	4	5	6	7	8	9	10
X_i:	0	1	2	3	4	5	6	7	8	9
Y_i:	98	135	162	178	221	232	283	300	374	395

a. Prepare a scatter plot of the data. Does a linear relation appear adequate here?

b. Use the Box-Cox procedure and standardization (3.36) to find an appropriate power transformation of Y. Evaluate SSE for $\lambda = .3, .4, .5, .6, .7$. What transformation of Y is suggested?

c. Use the transformation $Y' = \sqrt{Y}$ and obtain the estimated linear regression function for the transformed data.

d. Plot the estimated regression line and the transformed data. Does the regression line appear to be a good fit to the transformed data?

e. Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot. What do your plots show?

f. Express the estimated regression function in the original units.

Part 5: 15 points

Re **Sales growth** Problem

a. Divide the range of the predictor variable (coded years) into five bands of width 2.0, as follows: Band 1 ranges from $X = -.5$ to $X = 1.5$; band 2 ranges from $X = 1.5$ to $X = 3.5$; and so on. Determine the median value of X and the median value of Y in each band and develop the band smooth by connecting the five pairs of medians by straight lines on a scatter plot of the data. Does the band smooth suggest that the regression relation is linear? Discuss.

b. Create a series of seven overlapping neighborhoods of width 3.0 beginning at $X = -.5$. The first neighborhood will range from $X = -.5$ to $X = 2.5$; the second neighborhood will range from $X = .5$ to $X = 3.5$; and so on. For each of the seven overlapping neighborhoods, fit a linear regression function and obtain the fitted value $\hat{Y}_c$ at the center X_c of the neighborhood. Develop a simplified version of the lowess smooth by connecting the seven $(X_c, \hat{Y}_c)$ pairs by straight lines on a scatter plot of the data.

c. Obtain the 95 percent confidence band for the true regression line and plot it on the plot prepared in part (b). Does the simplified lowess smooth fall entirely within the confidence band for the regression line? What does this tell you about the appropriateness of the linear regression function?

Part 6: 20 points

When the predictor variable is so coded that $\bar{X} = 0$ and the normal error regression model (2.1) applies, are b_0 and b_1 independent? Are the joint confidence intervals for β_0 and β_1 then independent?

Derive an extension of the Bonferroni inequality (4.2a) for the case of three statements, each with statement confidence coefficient $1 - \alpha$.

Show that for the fitted least squares regression line through the origin $\sum X_i e_i = 0$.

Show that $\hat{Y}$ as defined for linear regression through the origin is an unbiased estimator of $E\{Y\}$.

Part 7: 20 (bonus) points

Refer to the **CDI** data set Consider the regression relation of number of active physicians to total population.

a. Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 95 percent family confidence coefficient.

b. An investigator has suggested that β_0 should be -100 and β_1 should be .0028. Do the joint confidence intervals in part (a) support this view? Discuss.

c. It is desired to estimate the expected number of active physicians for counties with total population of $X = 500, 1{,}000, 5{,}000$ thousands with family confidence coefficient .90. Which procedure, the Working-Hotelling or the Bonferroni, is more efficient here?

d. Obtain the family of interval estimates required in part (c), using the more efficient procedure. Interpret your confidence intervals.

Assignment 3

- E-mail your responses in a **single** pdf file by midnight, Thursday, October 26.
- Use the following file name:
  LASTNAME_FIRSTNAME_ASUID_ASSIGNMENTNUMBER
- Prepare your pdfs carefully; each week some of you will present their work.
- Include the `R` commands you used.