A utility company is interested in investigating how Y, the electricity consumption by each household, is related to monthly income (X_1), number of persons (X_2) in the household, and the living area (X_3) of the house (or apartment). A simple random sample of 34 households served by this utility company is surveyed, and the following information is obtained for each household:

Y = (total) electric bill (in dollars) for the past year
X_1 = monthly income for the household (in dollars)
X_2 = number of persons in the household
X_3 = living area (in square feet) of the house or apartment

The investigator assumes that the regression function of Y on X_1, X_2, and X_3 is of the form (at least approximately)

$$\mu_Y(x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

for $2{,}000 \le x_1 \le 6{,}000$, $x_2 = 1,2,3,4,5,6$, and $500 < x_3 < 4{,}000$.

Suppose that the assumptions for multiple linear regression are met.

The utility company is interested in predicting the electricity usage patterns by various households for next year so the target population is a *future* population. However, the study population is assumed to be very similar to the target population, and so the sample data from the study population may be used to make inferences about the target population.

Questions / tasks [#10 and #11 are advanced]

1 Find the values of $\hat{\beta}_0$ to $\hat{\beta}_3$, SE[$\hat{\beta}_0$] to SE[$\hat{\beta}_3$], $\hat{\mu}_Y(x_1, x_2, x_3)$

2 How well can Y, the annual electric bill for a randomly chosen household, be predicted using X_1, X_2 and X_3 ?

3 How well can Y, the annual electric bill for a randomly chosen household, be predicted if no predictor factors are used?

use $$, not $^{2} or R^2, to answer 2 and 3 !!

4-6

What is the difference between the average annual electric bill of households ... *if they have the same*

consisting of 5 individuals **vs.** 4 individuals	• monthly income • living area
with monthly income of $4,000.00 **vs.** $3,000.00	• # individuals in house • living area
with a living area of 2,000 sq. ft **vs.** l,500 sq. ft	• monthly income • # individuals in house

Obtain a point estimate and a 95% upper confidence bound for each of these 3 quantities.

7 The monthly income of a *particular* household is $3,200. There are 6 individuals in the household, and the living area equals 2,800 square feet. Predict the annual monthly electric bill for this household.

8 In Problem 7, compute a 90% upper confidence bound for the annual electric bill for this household.

9 Repeat Problems 7 and 8 for the *average* annual electric bill of all households with this monthly income, number of individuals in the household, and living area.

10 What is the difference between the average annual electric bill for households consisting of 7 individuals with a living area of 3,500 square feet and a monthly income of $5,400.00 and that for households consisting of 4 individuals with a living area of 2,400 square feet and a monthly income of $5,000.00? Estimate this difference and also compute a 95% two-sided confidence interval for this quantity.

11 The model above would be considered adequate for predicting Y if the standard deviation of the prediction errors, i.e., σ, is less than $50.00. If this is not so, then the investigator will look for additional explanatory variables to include in the regression model. To assist the investigator in making a decision in this connection, compute a two-sided 95% confidence interval for σ. Do you believe that the investigator needs to look for additional explanatory variables? Or do you believe that the model is adequate? Explain