Statistical Analysis

1. What is an economic model and, ideally, what characteristics would we like our economic models to satisfy?

2. Suppose that we have a simple economic model that shipper demands for transportation services depend upon the transportation rate charged.

    a. Identify the cause and effect relationship in this model.
    b. Why is this model a qualitative model of shipper demands?
    c. How would you transform this economic model into an econometric model of shipper demand?
    d. In your econometric model, how would you test the hypothesis that demand depends upon transportation rate charged?

3. You are concerned about the damage that automobile pollution is causing our nation's environment and are interested in the effect that gasoline prices have upon gasoline consumption,

    a. Specify your economic model as well as your econometric model of gasoline consumption. In your econometric model, what are the unknown parameters and what is the interpretation of each parameter?
    b. What hypothesis will you test in your econometric model of gasoline consumption? Would this be a one-tail or two-tail test and why?

4. What do we mean when we say that we want our economic models to be representative? The table below identifies two groups of individuals. The "population" is the group on which we would like to get information, and the "sample" is the group from whom we collect information and make inferences about the population. For each pair of groups, identify whether the sample would be representative of the population.

Population	Sample
Vacation travellers	Households with incomes > $50,000
Airline operations	Airlines currently offering service
Public transit firms	Bus companies in Indiana
Work trips	Automobile owners
Freight shippers	Agricultural producers
Air travellers	Travellers at Chicago's O'Hare Airport

5. Consider the following econometric model of T individual bus demands:

$$(\text{Bus Trips})_t = \alpha + \beta\,(\text{Fare})_t + \gamma\,(\text{Income})_t + \varepsilon_t$$

where $(\text{Fare})_t$ is the fare (in dollars) per bus trip paid by individual t and $(\text{Income})_t$ (in dollars) is individual t's income.

a. What sign would you expect β and γ, respectively, to have, and why?
b. From the econometric model, what impact will a 50-cent increase in bus fare have upon ridership? What impact will a $1 increase in household income? What about a $1000 increase in household income?
c. The method of least squares estimates α, β and γ by minimizing the sum of errors, $\Sigma_t \varepsilon_t^2$. Why is this method preferred to an alternative procedure that would estimate α, β and γ by minimizing the sum of errors, $\Sigma_t \varepsilon_t$?
d. From a statistical viewpoint, what desirable properties do linear regression estimates have?

6. For a policy-maker, one objective of econometric models is to forecast the values of economic variables. Suppose, for example, that the head of the Department of Transportation wants to predict the rail petroleum consumption next year. To do so, she looks at petroleum consumption over the past 20 years, takes the average, and uses the average as an estimate of next year's consumption.

a. Why is this estimate likely to be an incorrect estimate of next year's petroleum consumption?
b. Show that estimating next year's consumption by the average consumption over the past 20 years is comparable to the following regression model:

$$\text{Rail Petroleum Consumption} = \alpha + \varepsilon_t, \qquad t = 1, \ldots, 20,$$

where α is estimated by the method of least squares. According to this model, what impact will a $1 increase in the price of oil have upon rail petroleum consumption?

c. On the advice of her economic staff, the policy-maker revises her model to include the price of oil:

$$\text{Rail Petroleum Consumption} = \alpha + \beta \,(\text{Price of Oil}) + \varepsilon_t, \qquad t = 1, \ldots, 20,$$

In this case, what would be the policy head's prediction of next year's consumption of oil by railroads? And what effect will a $1 increase in the price of oil have upon rail petroleum consumption?

7. You work for an airline and are trying to determine the effect that fuel prices and distance travelled have upon airline fares. You estimate the following model using regression analysis:

$$\log(\text{fare}) = -4{,}5 + 0{,}36 \log(\text{fuel price}) + 0{,}68 \log(\text{distance}) \qquad R^2 = 0{,}90,$$
$$ {\scriptstyle (1,3)} \quad {\scriptstyle (0,15)} \qquad\qquad\qquad {\scriptstyle (0,03)}$$

where the numbers in parenthesis are the standard errors. The model was estimated on a cross-section of 100 routes flown by a major airline.

  a. What is the economic model underlying the relationship between airline fares, fuel prices, and distance travelled? What are the qualitative hypotheses associated with this model?
  b. Are the econometric results presented above consistent with the qualitative economic hypotheses?
  c. Derive a 95% confidence interval for the true values of the coefficients of the two explanatory variables. What does this interval tell you?
  d. According to the estimated equation, what effect will a 1% increase in fuel prices have upon airline fares? In the early 1980s, fuel prices increased by 60 %. Based upon the above equation, what effect would this have upon airline fares?
  e. Suppose that the average length of trip on a major airline is 1,000 miles, whereas the average trip length on a smaller regional airline is 400 miles. According to the above equation, what effect would this difference in trip length have upon airline fares?

8. You are interested in the effect that increased speed limits have had upon intercity freight carriage by truck. To that end, you estimate the following econometric model on time series data from 1976 to 1991. From 1976 to 1986, the speed limit on rural interstates was 55 mph. In 1987, the federal government passed legislation that allowed individual states to raise the speed limits to 65 mph:

$$(\text{Truck Carriage})_t = \alpha + \beta_1 (\text{GDP})_t + \beta_2 (\text{Speed Limit})_t + \varepsilon_t,$$
$$t = 1976, \ldots, 1991,$$

where Truck Rate is the truck shipping cost per ton-mile, Rail Rate is the railroad shipping price per ton-mile, and GDP is gross domestic product. Speed Limit is a dummy variable that equals 0 for the years 1976–86 and equals 1 for the years 1987–1991.

  a. What is the expected sign on β_1? Would you use a one-tail or two-tail test to test this hypothesis?
  b. From the above equation, identify the percentage effect that a 5% increase in GDP would have upon truck shipments. That is, what is the elasticity of Truck Freight Carriage with respect to GDP? Alternatively, what effect would a 5% increase in GDP have upon the absolute amount of ton-miles shipped?
  c. What do you think the effect of higher speed limits on truck shipments would be? That is, do you expect β_2 to be positive, negative, or uncertain? Graphically depict the regression line during the period 1976–86 and then show how the increase speed limit would affect the regression line.

9. For small-package transportation firms, there is a seasonal increase in the demand for their services during the holiday season. Consider the following quarterly time series model of firm revenues for small-package transportation firms:

$$\log (\text{Revenues})_t = \alpha + \beta_1 (\text{Time})_t + \beta_2 (\text{Fourth Quarter})_t + \varepsilon_t, \qquad t = 1, \ldots, T,$$

where the variable Time takes on the value 1 for $t = 1$, 2 for $t = 2$, and so on. Fourth Quarter is a dummy variable that equals one for the fourth quarter of the year (October–December) and is included to capture the holiday seasonal effect. During the first three quarters, Fourth Quarter equals zero.

    a. What is the interpretation of β_1?

    b. What is the interpretation of β_2, and would you expect β_2 to be greater than or less than zero? Assuming that $\beta_1 > 0$, graph the relationship between the dependent variable and Time for the months of January through September. What does this relationship look like during the fourth quarter of the year?

10. You operate an inland water transport service and hypothesize that your average costs per cargo ton-mile of operation ($AVCOST$) depend upon the average length of haul (AVH) and route density (DEN). You estimate the following equation using a cross-section of 38 observations of water transport firms:

$$AVCOST = 0{,}04 - 0{,}15\ AVH - 0{,}80\ DEN, \qquad\qquad R^2 = 0{,}85,$$
$$\underset{(0,01)}{}\ \underset{(0,042)}{}\ \ \ \underset{(0,235)}{}$$

where $AVCOST$ = average costs = total costs divided by cargo ton-miles; AVH = average length of haul = cargo ton-miles divided by tons carried; and DEN = route density = cargo ton-miles divided by route mileage. Numbers in parenthesis are standard errors.

    a. How well do the two explanatory variables explain average costs?

    b. Interpret coefficients.

    c. Calculate a 99% confidence region for the true marginal effect of density on average cost.

    d. Assuming that the average cost per ton-mile is 18 cents and average length of haul is 300 miles, what is the elasticity of average cost with respect to average length of haul?