1. Explain the difference between **ˆ

1

and ** ; between the residual

*u*ˆ*i*

and the regression

error *ui* ; and between the OLS predicted value *Y*ˆ

*i*

and

*E*[*Y* | *X* ]

1. For each least squares assumption, provide an example in which the assumption is valid , and then provide an example in which the assumption fails.

*i*

*i*

1. Sketch a hypothetical scatterplot of data for an estimated regression with Sketch a hypothetical scatterplot of data for a regression with
2. Suppose that a researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression,
   1. A classroom has 22 students. What is the regression’s prediction for that classroom’s average test score?
   2. Last year a classroom had 19 students, and this year it has 23 students. What is the regression’s prediction for the change in the classroom average test score?
   3. The sample average class size across the 100 classrooms is 21.4. What is the sample average of the test scores across the 100 classrooms?
   4. What is the sample standard deviation of test scores across the 100 classrooms?
3. Suppose that a random sample of 200 20 year old men is selected from a population and that these men’s height and weight are recorded. A regression of weight on height yields

where Weight is measured in funds and Height is measured in inches.

1. What is the regression’s weight prediction for someone who is 70 inches tall? 65 inches tall? 74 inches tall?
2. A man has a late growth spurt and grows 1.5 inches over the course of a year. What is the regression’s prediction for the increase in this man’s weight?
3. Suppose that instead of measuring weight and height in funds and inches, these variables are measured in centimetres and kilograms. What are the regression estimates from this new centimetre-kilogram regression?
4. A regression of average weekly earnings (AWE, in dollars on age (in years) using a random sample of college-educated full-time workers aged 25-65 yields the following:
5. Explain what the coefficient values 696.7 and 9.6 mean.
6. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER (dollars? years? or is SER unit-free)?
7. The regression

*R*2 is 0.023. What are the units of measurement for the

*R*2 (dollars?

years? or is SER unit-free)?

1. What is the regression predicted earnings for a 25 year-old worker? A 45 year-old worker?
2. Will the regression give reliable predictions for a 99 year-old worker? What or why not?
3. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal?
4. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?
5. Show that the first least squares assumption, , implies that
6. Show that **ˆ0 is an unbiased estimator of **0 .
7. a. A linear regression yields Show that

b. A linear regression yields . Does this imply that

1. Consider the regression model Derive the formula for the least squares estimator for **1 .
2. Consider the regression model Derive the formula for the least squares estimator for **0 .
3. Consider the regression model Suppose you know that **0  0. Derive the formula

for the least squares estimator for **1 .

1. Consider the regression model Suppose you know that

**0  4. Derive the formula

for the least squares estimator for **1 .

1. Show that the regression

*R*2 in the regression of Y on X is the squared value of the

sample correlation between X and Y. That is, show that

*R*2  *r*2 .

1. Show that the

*XY*

*R*2 from the regression of Y on X is the same as the

*R*2 from the

regression of X on Y.

1. The output voltage for an electric circuit is specified to be 130. A sample of 40 independent readings on the voltage for this circuit gave a sample mean 128.6 and standard deviation 2.1. Test the hypothesis that the average output voltage is 130 against the alternative that it is less than 130. Use a test with level .05.

If the voltage falls as low as 128, serious consequences may result. For testing *H*0 : μ = 130 versus *Ha* : μ = 128, find the probability of a type II error, β, for the rejection region used in the above exercise.

1. The Rockwell hardness index for steel is determined by pressing a diamond point into the steel and measuring the depth of penetration. For 50 specimens of an alloy of steel, the Rock- well hardness index averaged 62 with standard deviation 8. The manufacturer claims that this alloy has an average hardness index of at least 64. Is there sufficient evidence to refute the manufacturer’s claim at the 1% significance level? The steel is sufficiently hard to meet usage requirements if the mean Rockwell hardness measure does not drop below 60. Using the rejection region found in the above exercise, find β for the specific alternative μ*a* = 60.
2. A study by Children’s Hospital in Boston indicates that about 67% of American adults and about 15% of children and adolescents are overweight.4 Thirteen children in a random sample of size 100 were found to be overweight. Is there sufficient evidence to indicate that the percentage reported by Children’s Hospital is too high? Test at the α = 0.05 level of significance. Find β for the specific alternative p*a* = 0.2.
3. A manufacturer claimed that at least 20% of the public preferred her product. A sample of 100 persons is taken to check her claim. With α = .05, how small would the sample percentage need to be before the claim could legitimately be refuted? (Notice that this would involve a one-tailed test of the hypothesis.)
4. In March 2001, a Gallup poll asked, “How would you rate the overall quality of the environment in this country today—as excellent, good, fair or poor?” Of 1060 adults nationwide, 46% gave a rating of excellent or good. Is this convincing evidence that a majority of the nation’s adults think the quality of the environment is fair or poor? Test using α = .05. Find β for the specific alternative p*a* = 0.2
5. A political researcher believes that the fraction *p*1 of Republicans strongly in favor of the death penalty is greater than the fraction *p*2 of Democrats strongly in favor of the death penalty. He acquired independent random samples of 200 Republicans and 200 Democrats and found 46 Republicans and 34 Democrats strongly favouring the death penalty. Does this evidence provide statistical support for the researcher’s belief? Use α

= .05. Find β for the specific alternative *pd* = 0.09 (d: stands for proportion difference)

1. A random sample of 500 measurements on the length of stay in hospitals had sample mean 5.4 days and sample standard deviation 3.1 days. A federal regulatory agency hypothesises that the average length of stay is in excess of 5 days. Do the data support this hypothesis? Use α = .05. Find β for the specific alternative μ*a* = 5.5
2. The hourly wages in a particular industry are normally distributed with mean $13.20 and standard deviation $2.50. A company in this industry employs 40 workers, paying them an average of $12.20 per hour. Can this company be accused of paying substandard wages? Use an α = .01 level test. Find β for the specific alternative μ*a* = 10.2