

MIDTERM EXAMINATION #2 VERSION A
“Two-Variable Regression”
February 23, 2006

INSTRUCTIONS: This exam is closed-book, closed-notes. You may use a calculator on this exam, but not a graphing calculator or a calculator with alphabetical keys. Point values for each question are noted in brackets. A table of the t-distribution is attached.

NOTATION: In this exam, $\hat{\beta}_1$ and $\hat{\beta}_2$ denote the least-squares estimators of the intercept and slope of the line $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, \hat{y}_i denotes a least-squares fitted value, $\hat{\varepsilon}_i$ denotes a least-squares residual, and the sample size is denoted n . The true or population value of the variance of the unobserved error term ε_i is denoted σ^2 . The (unbiased) least-squares estimate of σ^2 is denoted $\hat{\sigma}^2$. The sample means of x and y are denoted \bar{x} and \bar{y} respectively.

I. MULTIPLE CHOICE: Circle the one best answer to each question. Feel free to use margins for scratch work [3 pts each—48 pts total]

(1) Which equation holds necessarily, regardless of the data?

- a. $\sum x_i y_i = 0$.
- b. $\sum x_i \hat{y}_i = 0$.
- c. $\sum \hat{y}_i \hat{\varepsilon}_i = 0$.
- d. $\sum (x_i - \bar{x})(y_i - \bar{y}) = 0$.
- e. $\sum (\hat{y}_i - \bar{y})^2 = \sum (y_i - \bar{y})^2 + \sum \hat{\varepsilon}_i^2$.

(2) In the model $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, assuming $E(\varepsilon_i|x_i)=0$, the conditional mean of y (that is, $E(y_i|x_i)$) is

- a. zero.
- b. β_1 .
- c. β_2 .
- d. $\beta_2 x_i$.
- e. $\beta_1 + \beta_2 x_i$.

(3) Which assumption is required for the sum of the least-squares residuals to equal zero?

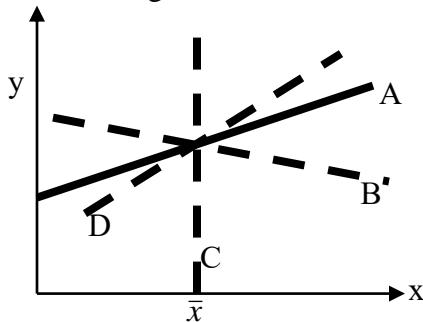
- a. $E(\varepsilon_i|x_i) = 0$.
- b. Homoskedasticity: $\text{Var}(\varepsilon_i) = \sigma^2$, a constant.
- c. No autocorrelation: $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.
- d. All of the above.
- e. None of the above.

(4) Which assumption is required for the least-squares estimators to be unbiased?

- a. $E(\varepsilon_i|x_i) = 0$.
- b. Homoskedasticity: $\text{Var}(\varepsilon_i) = \sigma^2$, a constant.
- c. No autocorrelation: $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.
- d. All of the above.
- e. None of the above.

(5) In the graph below, the solid line denoted "A" is the true population regression line. If the error term has mean zero but is *positively correlated* with x , then the least-squares estimated line will tend to resemble

- line A.
- line B.
- line C.
- line D.
- cannot be determined from the information given.



(6) Which assumption is required for the least-squares estimators to be Best Linear Unbiased Estimators?

- $E(\varepsilon_i|x_i) = 0$.
- Homoskedasticity: $\text{Var}(\varepsilon_i) = \sigma^2$, a constant.
- No autocorrelation: $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.
- All of the above.
- None of the above.

(7) Which assumption is required for the least-squares estimators to be method-of-moments estimators?

- $E(\varepsilon_i|x_i) = 0$.
- Homoskedasticity: $\text{Var}(\varepsilon_i) = \sigma^2$, a constant.
- No autocorrelation: $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.
- All of the above.
- None of the above.

(8) If the data observations fit exactly on a straight line, then the least-squares estimate of the variance of the error term would be

- $\hat{\sigma}^2 = n$.
- $\hat{\sigma}^2 = -1$.
- $\hat{\sigma}^2 = 0$.
- $\hat{\sigma}^2 = 1$.
- None of the above.

(9) The variance of the least-squares slope estimator $\hat{\beta}_2$ is smaller, and thus the true value of β_2 is estimated more precisely,

- the smaller the sample size.
- the larger the variance of the error term σ^2 .
- the larger the variation of the x values around the sample mean \bar{x} .
- All of the above.
- None of the above.

(10) The assumption that the error term is normally distributed:

$$\varepsilon_i \sim N(0, \sigma^2)$$

is required in order for the least-squares estimators to be

- consistent.
- best linear unbiased estimators.
- best unbiased estimators.
- asymptotically normally-distributed (that is, approximately normally-distributed in large samples).
- All of the above.

(11) Suppose we use the least-squares predictor ($\hat{y}_{n+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{n+1}$) to predict y_{n+1} . The variance of the prediction error ($y_{n+1} - \hat{y}_{n+1}$) is larger, and thus prediction is less precise,

- the smaller the variance of the error term σ^2 .
- the farther x_{n+1} is from \bar{x} .
- the larger the variation of the x -values in our sample around \bar{x} .
- All of the above.
- None of the above.

(12) The variance of the prediction error tends to decrease as the sample size used for estimation increases, finally approaching

- a. $\text{Var}(\hat{\beta}_1)$.
- b. $\text{Var}(\hat{\beta}_2)$.
- c. σ^2 .
- d. zero.
- e. one.

(13) In time-series data, any two variables that each have trends must be

- a. causally related.
- b. correlated in any finite sample.
- c. orthogonal.
- d. All of the above.
- e. None of the above.

(14) According to which model is the elasticity of y with respect to x equal to 0.5?

- a. $y = 4.6 + 0.5 x$.
- b. $y = 4.6 + 0.5 (1/x)$.
- c. $y = 4.6 + 0.5 \ln(x)$.
- d. $\ln(y) = 4.6 + 0.5 x$.
- e. $\ln(y) = 4.6 + 0.5 \ln(x)$.

(15) According to which model does a one-unit change in x cause approximately a three percent increase in y ?

- a. $y = 2.1 + 0.03 x$.
- b. $y = 2.1 + 0.03 (1/x)$.
- c. $y = 2.1 + 0.03 \ln(x)$.
- d. $\ln(y) = 2.1 + 0.03 x$.
- e. $\ln(y) = 2.1 + 0.03 \ln(x)$.

(16) Suppose a production function is estimated of the form $y_i = \beta_1 + \beta_2 x_i$, where y_i denotes output in gallons and x_i denotes labor input. Now suppose the output data are converted to liters (there are about 3.8 liters in a gallon) and the equation is re-estimated. Which of the following are true?

- a. $\hat{\beta}_1$ will increase by a factor of 3.8.
- b. $\hat{\beta}_2$ will increase by a factor of 3.8.
- c. The sum of squared residuals will increase by a factor of $(3.8)^2$.
- d. The r^2 value will be unaffected.
- e. All of the above.

II. PROBLEMS: Please write your answers in the boxes on this question sheet. Show your work and circle your final answers.

(1) [LS confidence intervals, tests, prediction: 24 pts] The relationship between per-capita income and population growth is estimated using a sample of $n=20$ countries. For each country i , let y_i denote its fertility rate and let x_i denote its per capita income in thousands of U.S. dollars. For example, if country #1 has a fertility rate of 2.5 and per-capita income of \$2,700 then $y_1 = 2.5$ and $x_1 = 2.7$.

The model $y_i = \beta_1 + \beta_2 x_i$ is estimated with the following results. (The numbers on top are the least-squares estimates of the intercept and slope, and the numbers at the bottom in parentheses are standard errors.)

Fertility rate	=	3.05	-	0.024	Per-capita income
		(0.06)		(0.008)	

a. [3 pts] Suppose a country's per-capita income rose from \$2 thousand dollars to \$5 thousand dollars. By how much would its fertility rate decrease? That is, predict Δy .

b. [3 pts] Suppose a country had a per-capita income of \$1 thousand dollars. According to these results, what would be its predicted fertility rate?

For the rest of this problem, assume the error term is normally-distributed.

c. [3 pts] What are the "degrees of freedom" for these estimates? Give an integer answer.

d. [6 pts] Compute a **95%** confidence interval for the intercept.

e. [9 pts] Test the null hypothesis that per-capita income has no effect on fertility, against the alternative hypothesis that per-capita income has a negative effect (a **one-tailed test**) at **5%** significance. Give the value of the test statistic, the critical point from a table, and your conclusion (whether you can reject null hypothesis).

Value of test statistic = _____. Critical point(s) = _____.

Reject null hypothesis? _____.

(2) [LS confidence intervals, tests, elasticity: 24 pts] The relationship between electricity rates and electricity usage is estimated using a sample of $n=1000$ households in a variety of electric utility districts. For each household i , let y_i denote the average monthly consumption of electricity (in kilowatt-hours) and x_i denote the rate or price per kilowatt-hour. For example, if household #1 faces a rate of \$0.07 per kilowatt-hour and consumes two thousand kilowatt-hours, then $x_1 = 0.07$ and $y_1 = 2000$.

The model $y_i = \beta_1 + \beta_2 x_i$ is estimated with the following results. The numbers on top are the least-squares estimates of the intercept and slope, and the numbers at the bottom in parentheses are standard errors.

Electricity usage per month in kilowatt-hours	=	2880 (465)	-	10800 (3375)	Rate per kilowatt- hour
---	---	---------------	---	-----------------	----------------------------

- a. [3 pts] Suppose electricity were free. According to these results, how much electricity is a typical household predicted to consume per month?

kilowatt-hours

- b. [3 pts] Suppose the electricity rate increased from \$0.06 per kilowatt-hour to \$0.10 per kilowatt-hour. By how many kilowatt-hours would electricity consumption decrease? That is, predict Δy .

kilowatt-hours

- c. [3 pts] Suppose the mean electricity usage in the sample is 1800 kilowatt-hours and the mean electricity rate is \$0.10 per kilowatt-hour. (That is, $\bar{y} = 1800$ and $\bar{x} = 0.10$.) Compute the estimated elasticity of demand for electricity at the sample means.

- d. [6 pts] Compute a **95%** confidence interval for the intercept.

- e. [9 pts] Test the null hypothesis that electricity rates have no effect on electricity usage, against the alternative hypothesis that rates have a negative effect (a **one-tailed test**) at **5%** significance. Give the value of the test statistic, the critical point from a table, and your conclusion (whether you can reject null hypothesis).

Value of test statistic = _____. Critical point(s) = _____.
 Reject null hypothesis? _____.

III. CRITICAL THINKING [4 pts] A researcher wants to estimate the effect of alcohol consumption on health status. Using a sample of 100 Drake employees, the researcher estimates the equation $y_i = \beta_1 + \beta_2 x_i$, where y_i denotes an index of the health status of employee i , and x_i denotes the number of alcoholic drinks consumed per month by the same employee. The researcher is shocked to find that the least squares estimator of β_2 is estimated as positive! Assuming that alcohol consumption, *ceteris paribus*, should actually *decrease* health status, why might least-squares give the "wrong" answer here? Which of the following assumptions would you suspect is violated in these data? Why? [Hint: The researcher has just discovered that a number of Drake employees suffer from serious medical conditions such as diabetes, cirrhosis, etc. and consequently have been forbidden by their doctors from consuming alcohol.]

- $E(\varepsilon_i|x_i) = 0$.
- Homoskedasticity: $\text{Var}(\varepsilon_i) = \sigma^2$, a constant.
- No autocorrelation: $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.

[end of exam]