B. Best Linear Unbiased Estimation

B1. You are given the following regression model:

$$Y = \alpha + \beta X + \varepsilon,$$

where ε is a random error term with mean 0 and variance σ^2. Which of the following are true?

1. The least-squares estimators, α and β, are linear combinations of the values Y_i, $i = 1, 2, \ldots, n$.

2. $s_{\hat{\alpha}} = \sigma \sqrt{\dfrac{\sum X_i^2}{N \sum x_i^2}}$ 3. $s_{\hat{\beta}} = \sigma \sqrt{\dfrac{1}{\sum x_i^2}}$

A. 1,2 B. 1,3 C. 2,3 D. 1,2,3 E. None of these answers are correct. (84F–3–8)

B2. In a simple regression model, the variance of $\hat{\alpha}$ is a function of the number of observations. (86S–3–5–MC).

B3. In a simple regression model, the variance of $\hat{\beta}$ can be reduced by using a wider range of the independent variables. (86S–3–5–MC).

B4. You are given two simple linear regression models:

$Y_i = \alpha_1 + \beta_1 X_i + \varepsilon_{1i}$ $\quad Y_i = \beta_2 X_i + \varepsilon_{2i}$, where

If ε_{1i} and ε_{2i} are random error terms with mean 0 and variance σ^2, $\hat{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$.

(91S–120–10–MC)

B5. You fit a least-squares regression line to five pairs of observations (X_t, Y_t) using the model: $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$.

You determine:

$$\sum X_t = 10 \qquad \sum X_t^2 = 30 \qquad \sum (Y_i - \hat{Y}_i)^2 = 15$$

Determine $s^2_{\hat{\beta}1}$.

A. .3 B. .5 C. 1 D. 2 E. 3 (Sample–120–1)

B6. Four pairs of values (X_i, Y_i) are represented by a simple linear regression model $Y_i = \alpha + \beta X_i + \varepsilon_i$, where the variance of the error term is 4. You are given:

$$X_1 = -2 \quad X_2 = 1 \quad X_3 = 2 \quad X_4 = 3$$

Determine the correlation coefficient between $\hat{\alpha}$ and $\hat{\beta}$.

A. $-2/9$ B. $-\sqrt{2}/6$ C. $-2/7$ D. $-\sqrt{2}/3$ E. $-1/2$ (98F–120–2)

B7. For a two-variable regression based on seven observations, you are given:

$$\sum (X_i - \bar{X})^2 = 2{,}000 \qquad \sum \hat{\varepsilon}_i^2 = 967$$

Calculate $s_{\hat{\beta}}$, the standard error of $\hat{\beta}$.

A. .26 B. .28 C. .31 D. .33 E. .35 (01S–4–40)

B8. You observe N independent observations from a process whose true model is:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

You are given:

i) $Z_i = X_i^2$, for $i = 1, 2, \ldots, N$

ii) $$b^* = \frac{\Sigma (Z_i - \bar{Z})(Y_i - \bar{Y})}{\Sigma (Z_i - \bar{Z})(X_i - \bar{X})}$$

Which of the following is true?

A. b^* is a nonlinear estimator of β.
B. b^* is a heteroscedasticity-consistent estimator (HCE) of β.
C. b^* is a linear biased estimator of β.
D. b^* is a linear unbiased estimator of β, but not the best linear unbiased estimator (BLUE) of β.
E. b^* is the best linear unbiased estimator (BLUE) of β. (01F–4–35)

B9. You fit the model $Y_i = \alpha + \beta X_i + \varepsilon_i$ to twenty observations. You are given:

$$ESS = 2{,}000 \qquad \Sigma X_i = -300 \qquad \Sigma X_i^2 = 6{,}000$$

Determine $\hat{Cov}(\hat{\alpha}, \hat{\beta})$.

A. .7 B. .8 C. .9 D. 1.0 E. 1.1 (05–ASM–13)

C. Tests of Regression Coefficients

C1. The output of a least-squares regression analysis provides the following equation and statistics which indicate the success achieved with the estimating equation:

Equation: Income = −5.0 + 15.0Y − 225.0X

Variable	Estimated Coefficient	Standard Error
Constant	−5.0	4.0
Y	15.0	3.0
X	−225.0	15.0

The R-bar squared equals .995.

a. What is the t statistic for each variable used in the regression?
b. Would you consider eliminating any variable based on the results? Explain your answer. (80–10–17a&b–1/1)

C2. Multiple regression analysis has been used to determine the equation:

$$Y = A_0 + A_1X_1 + A_2X_2 + A_3X_3 + \text{Residual}$$

The following statistical results have been calculated:

R^2	.85	F statistic	49.11
Durbin-Watson d statistic	1.98	Number of data points	30
Standard error of the regression	21.32	t statistic for A_1	7.37
t statistic for A_2	−.21	t statistic for A_3	−5.20

How would you determine which, if any, of the coefficients A_1, A_2, A_3 differ(s) significantly from 0? (81–10–9c–1)

C3. The t-test measures how many standard errors a given regression coefficient is from 0. (82–10–17–1)

C4. The sign of the t statistic for a coefficient in a linear regression <u>is the same as</u> the sign of the coefficient. (83–10–17–1)

C5. You are given ten pairs of values (X_i, Y_i) that will be represented by the following model:

$$Y = \alpha + \beta X + \varepsilon,$$

where ε is a random error term with mean 0 and variance σ^2. You have determined:

$$\sum_{i=1}^{10} (X_i - \bar{X})^2 = 400 \qquad \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = 425 \qquad \sum_{i=1}^{10} (\hat{Y}_i - \bar{Y})^2 = 225$$

Calculate the value of the t statistic used for testing the hypothesis H_0: $\beta = 0$.

A. 1.5 B. 2.0 C. 2.5 D. 3.0 E. 3.5 (84F–3–6)

C6. You are representing ten observations (X_i, Y_i) by the following model:

$$Y = \alpha + \beta X + \varepsilon,$$

where ε is a random error term with mean 0 and variance σ^2. You have determined:

$$\hat{\beta} = 1.3 \qquad \sum_{i=1}^{10} X_i = 50 \qquad \sum_{i=1}^{10} X_i^2 = 650 \qquad s^2 = 169$$

Calculate the t value for testing H_0: $\beta = 0$.

A. 2 B. 4 C. 7 D. 10 E. 13 (86S–3–8)

C7. The following model of claims inflation is used for a line of insurance:

$$a_k(j) = (I_j/I_0)^{(1-\alpha)}(I_{k+j}/I_k)^{\alpha} a_k(0)$$

j is the interval between time 0 and the date of accident and k is the interval between the date of accident and the date of payment. All claims for this line are paid 2 years after the accident date. A regression using all successive partial severities having the same duration is performed to estimate α. The results of that regression are: $\hat{\alpha} = .10$ and $SE(\hat{\alpha}) = .06$. Assuming 2 standard errors are required for significance, what is the expected increase in average severity from accident year 85 to accident year 86? Use the claim cost index provided below. Explain your answer.

Claim Cost Index

84	85	86	87	88
1.00	1.10	1.25	1.40	1.50

(86–10–56–2)

C8. You are given the following model: $Y = \alpha + \beta X + \varepsilon$, where ε is a random error term with mean 0 and variance σ^2. You have used this model to represent a set of 25 observations (X_i, Y_i). You are given:

$$\bar{X} = 0 \qquad s^2 = 100 \qquad \hat{\alpha} = 3$$

Determine the length of the shortest 98-percent confidence interval for α.

A. 9.9 B. 10.0 C. 10.1 D. 10.2 E. 10.3 (86F–120–5)

C9. You are given:

X	1	2	3	4	5
Y	–2	4	2	–1	0

Using linear least squares, determine the t value for testing the hypothesis that no linear relationship exists between y and x.

A. .01 B. .03 C. .09 D. .11 E. .13 (87S–120–5)

C10. You are using a linear regression model to represent the following data:

Independent variable (X)	1	2	3	4	5	6	7	8	9
Dependent variable (Y)	2	4	1	3	6	8	7	5	9

You have determined that $\sum_{i=1}^{9} (Y_i - \hat{Y}_i)^2 = 21.6$. Which of the following are true?

1. The t value used to test H_0: $\beta = .25$ is 2.4 to the nearest .1.
2. The hypothesis H_0: $\beta = .25$ cannot be rejected at the 2 percent level of significance.

A. 1 B. 2 C. 1,2 D. None of these statements are true. (87S–120–8)

C11. Your assistant fitted a set of observations to a simple linear regression model with independent variable $X_i = i$, for $i = 1$ to 5. When you returned, you found that coffee had obliterated most of the information. The only things that you could make out were that the shortest 95-percent confidence interval for β was $(-1.5, 2.5)$ and that $\hat{\alpha}$ was 30. Determine the upper bound of the shortest 95-percent confidence interval for α.

erreur : lire table Student.

A. 5.0 B. 6.6 C. 8.8 D. 9.6
E. The answer cannot be determined from the information provided. (89S–120–6)

C12. For a simple linear regression model, you are given:

i) The absolute value of the t statistic for testing the hypothesis H_0: $\beta = 0$ versus H_a: $\beta \neq 0$ is 2.

ii) $\sum (X_i - \bar{X})Y_i = 4s$, where s^2 is the residual mean square.

Calculate $\sum (X_i - \bar{X})^2$.

A. 1 B. 2 C. 4 D. 8 E. 16 (89F–120–2)

C13. You are studying the average return on sales as a function of the number of firms in an industry. You have collected the data below for 1969–88 and performed a regression of the form:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Observation	Year	Number of Firms (X_i)	Return on Sales (Y_i)
1	1969	138	.167
2	1970	130	.155
3	1971	134	.211
4	1972	152	.166
5	1973	180	.121
6	1974	216	.100
7	1975	256	.147
8	1976	294	.058
9	1977	328	.092
10	1978	354	.146
11	1979	367	.069
12	1980	368	.041
13	1981	356	.116
14	1982	332	.085
15	1983	299	.076
16	1984	260	.087
17	1985	221	.155
18	1986	184	.142
19	1987	155	.190
20	1988	136	.119

$\hat{\beta} \pm t_{18,\frac{.05}{2}} \hat{\sigma}_{\hat{\beta}}$

$t_{18,\frac{.05}{2}} = 2.100$

$\hat{\beta} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - 20\bar{x}^2}$

$\hat{\sigma}_{\hat{\beta}} = \sqrt{\frac{\sigma^2}{\sum (x_i - \bar{x})^2}} =$

$\sum_{i=1}^{20} X_i = 4{,}860$ $\quad \sum_{i=1}^{20} Y_i = 2.452$ $\quad \sum_{i=1}^{20} X_iY_i = 539.309$ $\quad \sum_{i=1}^{20} X_i^2 = 1{,}330{,}224$

$\sum_{i=1}^{20} Y_i^2 = .341804$ $\quad \sum_{i=1}^{20} \varepsilon_i^2 = .0198$

Determine the upper bound of the shortest 95-percent confidence interval for the regression coefficient β.

A. –.0010 B. –.0008 C. –.0006 D. –.0004 E. –.0002 (89F–120–4)

C14. You are given a simple regression of the form:

$$Y = \alpha + \beta X + \varepsilon$$

Based on eighteen observations of the variables X and Y, you have determined that the correlation coefficient of the two variables is .6 and that the estimated slope coefficient $\hat{\beta}$ is 3.0. Determine $\hat{\sigma}_{\hat{\beta}}$.

A. .1 B. .4 C. 1.0 D. 1.2 E. 1.5 (90F–120–12)

C15. The model $Y_i = \alpha + \beta X_i + \varepsilon_i$, where the ε_i's are uncorrelated normal random variables with mean 0 and variance $\sigma^2 = 1$, is to be fitted to observations at X = –2, –1, 0, 1, and 2. Using this model, determine q such that $\Pr[\hat{\beta} - \beta < q] = .95$.

σ

A. .2 B. .3 C. .5 D. .6 E. .7 (91S–120–8)

C16. The following table shows the average rating (X) given by customers to eight sales people and their average monthly salary (Y) in thousands:

Salesperson	1	2	3	4	5	6	7	8
Rating	3.9	1.5	2.4	1.7	3.0	2.5	1.9	4.0
Salary	2.8	1.7	2.0	1.6	2.3	2.5	1.2	1.8

Using a simple linear regression model to test the relationship of a salesperson's rating, you have determined:

$$\bar{X} = 2.612 \qquad \bar{Y} = 1.987 \qquad \sum X_i^2 = 6.36 \qquad \sum Y_i^2 = 1.91 \qquad \sum X_iY_i = 2.08$$

Determine the upper bound of the shortest symmetric 90% confidence interval for β_1.

A. .35 B. .39 C. .58 D. .63 E. .68 (Sample–120–9)

C17. You fit the model $Y_j = \alpha + \beta X_j + \varepsilon_j$ to 10 observed values (X_i, Y_i). You determine:

$$\sum(Y_i - \hat{Y}_i)^2 = 2.79 \qquad \sum(X_i - \bar{X}_i)^2 = 180 \qquad (Y_i - \bar{Y}_i)^2 = 152.40 \qquad \bar{X} = 6 \qquad \bar{Y} = 7.78$$

Determine the width of the shortest symmetric 95% confidence interval for α.

A. 1.1 B. 1.2 C. 1.3 D. 1.4 E. 1.5 (95F–120–3)

C18. You fit a regression model $Y_j = \alpha + \beta X_i + \varepsilon_i$ to 12 observations. You determine that the symmetric 95% confidence interval for β is (1.2, 3.8) and that

$$\sum (X_i - \bar{X})^2 = .826$$

Determine the residual variance (s^2).

A. .1 B. .2 C. .3 D. .4 E. .5 (98S–120–2)

C19. You fit 16 observations (X, Y) to a simple linear regression model. You determine that (–.8, 5.6) is the symmetric 95% confidence interval for the slope parameter (β). Determine the value of the t statistic for testing the hypothesis that $\beta = 0$.

A. .6 B. .8 C. 1.0 D. 1.3 E. 1.6 (98F–120–3)

C20. You are investigating the relationship between per-capita consumption of natural gas and the price of natural gas. You gathered data from 20 cities and constructed the following model:

$$Y = \alpha + \beta X + \varepsilon, \text{ where}$$

Y is per capita consumption, X is the price, and ε is a normal random error term. You have determined:

$$\hat{\alpha} = 138.561 \qquad \hat{\beta} = -1.104 \qquad \sum X_i^2 = 90{,}048 \qquad \sum Y_i^2 = 116{,}058$$

$$\sum x_i^2 = \sum(X_i - \bar{X})^2 = 10{,}668 \qquad \sum y_i^2 = \sum(Y_i - \bar{Y})^2 = 20{,}838$$

$$\sum \hat{\varepsilon}_i^2 = \sum(Y_i - \hat{Y}_i)^2 = 7{,}832$$

Determine the shortest 95% confidence interval for β.

A. (–2.1, –.1) B. (–1.9, –.3) C. (–1.7, –.5) D. (–1.5, –.7) E. (–1.3, –.9)
(00F–4–5)

C21. You fit the following model to eight observations:

$$Y = \alpha + \beta X + \varepsilon$$

You are given:

$$\hat{\beta} = -35.69 \qquad \sum (X_i - \bar{X})^2 = 1.62 \qquad \sum (Y_i - \hat{Y})^2 = 2{,}394$$

Determine the symmetric 90-percent confidence interval for β.

A. (–74.1, 2.7) B. (–66.2, –5.2) C. (–63.2, –8.2) D. (–61.5, –9.9) E. (–61.0, –10.4)
(01F–4–5)

C22. For the model $Y_i = \alpha + \beta X_i + \varepsilon_i$, where i = 1, 2, . . . ,10, you are given:

i) $X_i = \begin{cases} 1 & \text{if the ith individual belongs to a specified group} \\ 0 & \text{otherwise} \end{cases}$

ii) 40 percent of the individuals belong to the specified group.

iii) The least squares estimate of β is $\hat{\beta} = 4$.

iv) $\Sigma(Y_i - \hat{\alpha} - \hat{\beta}X_i)^2 = 92$

Calculate the t statistic for testing H_0: $\beta = 0$.

A. .9 B. 1.2 C. 1.5 D. 1.8 E. 2.1 (03F–4–5)

Q22) $X_i \begin{cases} 1 & \text{si groupe} \\ 0 & \text{sinon.} \end{cases}$ 40% → grp A.

X_i : $\underbrace{1 \ldots 1}_{4 \text{ fois}}$ $\underbrace{0 \ldots 0}_{6 \text{ fois}}$.