

PAST CAS AND SoA EXAMINATION QUESTIONSA. Multiple Regression Model

A1. Advertising expenditures and sales for the last five quarters have been as follows:

Quarter	1	2	3	4	5
Advertising	1	1	2	2	4
Sales	4	5	6	7	8

In quarter 3, a new product was introduced that would influence sales in quarters 3, 4, and 5. The following model is established:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \text{ where}$$

Y - sales

 X_1 - advertising expenditures X_2 - variable that is 1 when the new product is available and 0 otherwise ε - error component.Find the least-squares estimate of β_1 .

A. 25/14 B. 29/16 C. 33/16 D. 29/14 E. 33/14 (83S-3-10)

A2.

You are examining the relationship between the number of fatal car accidents on a tollway each month and three other variables: precipitation, traffic volume, and the occurrence of a holiday weekend during the month. You are using the following model:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon, \text{ where}$$

Y - number of fatal car accidents

 X_1 - precipitation in inches X_2 - traffic volume in thousands of cars X_3 - dummy variable that equals 1 if a holiday weekend occurs during the month and 0 otherwise ε - random error term with mean 0 and variance σ^2 .

The following data was collected for a 12-month period:

Month	Y	X_1	X_2	X_3
1	1	3	1	1
2	3	2	1	1
3	1	2	1	0
4	2	5	2	1
5	4	4	2	1
6	1	1	2	0
7	3	0	2	1
8	2	1	2	1
9	0	1	3	1
10	2	2	3	1
11	1	1	4	0
12	3	4	4	1

You have determined:

$$(X'X)^{-1} = \frac{1}{6,506} \begin{pmatrix} 257 & -82 & -446 \\ -82 & 254 & -364 \\ -446 & -364 & 2,622 \end{pmatrix} \quad s^2 = 1.45241$$

Determine $\hat{\beta}_1$. (Answer to the nearest .01.)

A. -.07 B. .15 C. .24 D. .70 E. 1.30 (85F-3-7)

- A3. You are using a multiple regression model to examine the relationship between average automobile insurance claim amounts and two other variables: the level of the Consumer Price Index and a dummy variable that is 0 when the speed limit is 70 miles per hour and 1 when the limit is 55 miles per hour. You have collected the following data:

Yr.	Ave. Auto Claim Amount (\$000s) - Y	Consumer Price Index - X ₁	Speed Limit - X ₂
1971	.753	1.500	0
1972	.634	1.612	0
1973	.779	1.803	0
1974	1.139	2.059	0
1975	.903	2.044	0
1976	1.008	2.222	0
1977	1.069	2.396	0
1978	.861	2.370	0
1979	1.021	2.608	1
1980	1.059	2.700	1
1981	1.159	2.830	1
1982	1.125	3.037	1
1983	1.189	3.131	1
1984	1.189	3.249	1
1985	1.491	3.408	1

You have determined:

$$\sum Y = 15.379 \quad \sum YX_1 = 39.478 \quad \sum YX_2 = 8.234 \quad \sum X_1X_2 = 20.963 \quad s^2 = .014$$

$$(X'X)^{-1} = \begin{pmatrix} 3.189 & -1.531 & 1.397 \\ -1.531 & .765 & -.761 \\ 1.397 & -.761 & 1.024 \end{pmatrix}$$

Calculate the predicted value of the average automobile insurance claim in 1986, using the estimated CPI value of 3.570 and assuming that the speed limit is 70 miles per hour.

- A. \$1,200 B. \$1,300 C. \$1,400 D. \$1,500 E. \$1,600 (86S-120-9)

- A4. You are given:

Y	1	2	4	4	5
X ₁	4	5	5	6	7
X ₂	1	1	2	3	3

You are to represent the data by the following model:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2 + \epsilon,$$

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

$$(X'X)^{-1} = \begin{pmatrix} 87.9 & -18.7 & -32.8 & 6.8 \\ -18.7 & 4.1 & 6.4 & -1.4 \\ -32.8 & 6.4 & 17.1 & -3.1 \\ 6.8 & -1.4 & -3.1 & .6 \end{pmatrix}$$

Calculate $\hat{\beta}_3$.

- A. -2.4 B. -1.8 C. -1.2 D. -.6 E. 0 (88S-120-10)

- A5. You are given the following weekly data (\$000s) on profits, food sales and nonfood sales for a supermarket:

Profit (Y)	2	3	4	6	10
Food sales (X_1)	1	1	2	3	3
Nonfood sales (X_2)	1	1	2	2	4

You are to represent the data by the following model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon,$$

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

$$(X'X)^{-1} = \begin{pmatrix} 1.20 & -.50 & .00 \\ -.50 & .75 & -.50 \\ .00 & -.50 & .50 \end{pmatrix}$$

Calculate $\hat{\beta}_1$.

- A. -2 B. -.75 C. -.5 D. .5 E. .75 (88F-120-8)
- A6. You have collected the following data on the response of a variable (Y) to two other variables:

Y	6.70	8.08	20.09	18.09	30.00	30.85
X_1	-5	-3	-1	1	3	5
X_2	-1	1	-1	1	-1	1

You have determined:

$$(X'X)^{-1} = \begin{pmatrix} .167 & .000 & .000 \\ .000 & .016 & -.016 \\ .000 & -.016 & .182 \end{pmatrix}$$

$\hat{\beta}_1$ is the estimate of the coefficient of X_1 in the multiple regression:

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

\hat{b}_1 is the estimate of the coefficient of X_1 in the multiple regression:

$$Y = b_0 + b_1 X_{1i} + \epsilon_i$$

Determine $\hat{\beta}_1 - \hat{b}_1$.

- A. -4 B. -2 C. 0 D. .1 E. .3 (89F-120-15)

A7. You are given:

Period	1	2	3	4	5
Y	1.3	1.5	1.8	1.6	1.7
X ₁	6	7	7	8	8
X ₂	4.5	4.6	4.5	4.7	4.6

You are to use the following regression model:

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \quad i = 1, 2, \dots, 5$$

You have determined:

$$(X'X)^{-1} = \begin{pmatrix} 1,522.73 & 26.87 & -374.67 \\ 26.87 & .93 & -7.33 \\ -374.67 & -7.33 & 93.33 \end{pmatrix}$$

Calculate $\hat{\varepsilon}_2$.

- A. -.2 B. -.1 C. 0 D. .1 E. .2 (89F-120-13)

A8. You are given the following model:

$$E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

You have collected the following data:

Y	6.3	7.5	8.1	10.8	4.7	6.8	8.3	7.0
X ₁	-3	-1	1	3	-3	-1	1	3
X ₂	-1	-1	-1	-1	1	1	1	1

Determine $\hat{\beta}_2$.

- A. -1.5 B. -.7 C. .6 D. 1.5 E. 2.1 (90S-120-1)

A9. You are performing a multiple regression of the form:

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

You have obtained the following data:

Y	1	2	3	4
X ₁	-1	1	-1	1
X ₂	-1	-1	1	1

Determine $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2$.

- A. 3.5 B. 4.0 C. 4.5 D. 5.0 E. 5.5 (90F-120-19)

A10. You are given:

Motorist	1	2	3	4	5	6
X_1	0	0	0	1	1	1
X_2	0	1	2	0	1	2
Y	1	0	2	1	3	5

You use the following model:

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \quad i = 1, 2, \dots, 5$$

where the ε_i 's are independent, normally distributed random variables with mean 0 and variance σ^2 . You have determined:

$$(X'X)^{-1} = \frac{1}{12} \begin{pmatrix} 7 & -4 & -3 \\ -4 & 8 & 0 \\ -3 & 0 & 3 \end{pmatrix}$$

Calculate $\hat{\beta}_2$.

- A. -.25 B. .25 C. 1.25 D. 2.00 E. 4.25 (93S-120-11)

A11. You fit the model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$ to the following data:

Y	1	2	4	3
X_2	-1	1	-1	1
X_3	-1	-1	1	1

Determine $\hat{\beta}_2$.

- A. 0 B. 1 C. 2 D. 3 E. 4 (96S-120-2)

A12. You fit the following model to four observations:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i, \quad i = 1, 2, 3, 4$$

You are given:

i	X_{2i}	X_{3i}
1	-3	-1
2	-1	3
3	1	-3
4	3	1

The least-squares estimator of β_3 is expressed as $\hat{\beta}_3 = \sum_{i=1}^4 w_i Y_i$. Determine (w_1, w_2, w_3, w_4) .

- A. (-.15, -.05, .05, .15) B. (-.05, .15, -.15, .05) C. (-.05, .05, -.15, .15)
 D. (-.3, -.1, .1, .3) E. (-.1, .3, -.3, .1) (01F-4-13)

B. Regression Statistics

B1. You are given the following model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \text{ where}$$

ε is a random error term with mean 0 and variance σ^2 . Using this model and a set of observations, you have determined that the estimated variance matrix is as follows:

$$C_S^2 = \begin{pmatrix} 900 & -144 & 225 \\ -144 & 1,600 & 400 \\ 225 & 400 & 625 \end{pmatrix}$$

You have also determined: $\hat{\beta}_2 = 40$. Calculate the t value used to test $H_0: \beta_2 = 0$.

A. .6 B. .8 C. 1 D. 1.3 E. 1.6 (86S-3-8)

B2. A sample of 25 observations has been represented by a model of the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon,$$

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

$$(X'X)^{-1} = \begin{pmatrix} 188.9832 & .8578 & -28.0275 \\ .8578 & .2500 & -.6000 \\ -28.0275 & -.6000 & 5.0625 \end{pmatrix}$$

$$s^2 = .0361 \quad \hat{\beta} = \begin{pmatrix} 87/72 \\ 207/72 \\ 189/72 \end{pmatrix}$$

Determine the shortest symmetric 95-percent confidence interval for β_1 .

A. (-.26, .54) B. (-.12, .40) C. (-.06, .34) D. (.06, .22) E. (.12, .16) (88S-120-8)

B3. Performing a regression of y_t on x_{t1} and x_{t2} , you determine that the regression equation is:

$$y_t = 1.2360 + .8683x_{t1} + .8517x_{t2}.$$

You are given:

Source	SS	df	MS	F
Regression	7.5753	2	3.7877	18.19
Error	1.8739	9	.2082	
Total	9.4492	11		

$$(X'X)^{-1} = \begin{pmatrix} 7.70997 & -.82314 & -1.05459 \\ -.82314 & .10044 & .10480 \\ -1.05459 & .10480 & .15284 \end{pmatrix}$$

Determine k such that a 95% confidence interval for β_2 is given by $\beta_2 \pm k$.

A. .16 B. .18 C. .33 D. .37 E. .40 (Sample-120-5)

B4. You fit the multiple regression model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$ to a set of data. You determine

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 6.1333 & -0.0733 & -0.1933 \\ -0.0733 & 0.0087 & -0.0020 \\ -0.1933 & -0.0020 & 0.0087 \end{pmatrix}$$

$$s^2 = 280.1167$$

Determine the estimated standard error of $(\hat{\beta}_2 - \hat{\beta}_3)$.

- A. 1.9 B. 2.2 C. 2.5 D. 2.8 E. 3.1 (95F-120-4)

B5. You perform a regression of Y on X_2 and X_3 . You determine: $Y_i = 20 - 1.5X_{2i} - 2X_{3i}$.

Source	Sum of Squares	Degrees of Freedom	SS/DF	F Ratio
Regression	42	2	21	5.25
Error	12	3	4	
Total	54	5		

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 4/3 & -1/4 & -1/3 \\ -1/4 & 1/16 & 0 \\ -1/3 & 0 & 2/3 \end{pmatrix}$$

Calculate the value of the t statistic for testing the null hypothesis $H_0: \beta_3 = 1$.

- A. -.9 B. -1.2 C. -1.8 D. -3.0 E. -5.0 (96S-120-5)

B6. You fit the multiple regression model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \epsilon_i$ to 30 observations. You are given:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 2.8195 & -.0286 & -.0755 & -.0263 \\ & .0027 & .0010 & -.0014 \\ & & .0035 & -.0010 \\ & & & .0032 \end{pmatrix}$$

$$\mathbf{Y}'\mathbf{Y} = 7,995 \quad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} 261.5 \\ 4,041.5 \\ 6,177.5 \\ 5,707.0 \end{pmatrix} \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 5.22 \\ 1.62 \\ .21 \\ -.45 \end{pmatrix}$$

Determine the length of the symmetric 95% confidence interval for $\hat{\beta}_3$.

- A. .3 B. .6 C. .7 D. 1.5 E. 1.8 (98S-120-6)

B7. You fit the following model to 30 observations:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

$$s^2 = 10 \quad r_{X_2 X_3} = .5 \quad \sum (X_2 - \bar{X}_2)^2 = 4 \quad \sum (X_3 - \bar{X}_3)^2 = 8$$

Determine the estimate of the standard deviation of the least-squares estimate of the difference between β_2 and β_3 .

- A. 1.7 B. 2.2 C. 2.7 D. 3.2 E. 3.7 (00S-4-35)

B8. For the model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$, you are given:

i) $N = 15$

ii) $(X'X)^{-1} = \begin{pmatrix} 13.66 & -.33 & 2.05 & -6.31 \\ -.33 & .03 & .11 & .00 \\ 2.05 & .11 & 2.14 & -2.52 \\ -6.31 & .00 & -2.52 & 4.32 \end{pmatrix}$

iii) $ESS = 282.82$

Calculate the standard error of $\hat{\beta}_3 - \hat{\beta}_2$.

- A. 6.4 B. 6.8 C. 7.1 D. 7.5 E. 7.8 (03F-4-36)

B9. You are given:

- i) Y is the annual number of discharges from a hospital.
 ii) X is the number of beds in the hospital.
 iii) Dummy variable D is 1 if the hospital is private and 0 if the hospital is public.
 iv) The classical three-variable linear regression model $Y = \beta_1 + \beta_2 X + \beta_3 D + \varepsilon$ is fitted to N cases using ordinary least squares.
 v) The matrix of estimated variances and covariances of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ is:

$$\begin{pmatrix} 1.89952 & -.00364 & -.82744 \\ -.00364 & .00001 & -.00041 \\ -.82744 & -.00041 & 2.79655 \end{pmatrix}$$

Determine the standard error of $\hat{\beta}_1 + 600\hat{\beta}_2$.

- A. 1.06 B. 1.13 C. 1.38 D. 1.90 E. 2.35 (05-ASM-2)

C. F Tests, R², and Corrected R²

- C1. What does the R-bar squared statistic measure? (80-10-17c-1)
 C2. You are given twenty pairs of values (X_i, Y_i) that are represented by the following model:

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

where ϵ is a random error term with mean 0 and variance σ^2 . You have determined that the F ratio for testing the hypothesis $H_0: \beta = 0$ is equal to 12. Calculate R².

- A. .3 B. .4 C. .5 D. .6 E. .7 (86S-3-10)

- C3. You have represented twenty observations by the following model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon, \text{ where}$$

ϵ is a random error term with mean 0 and variance σ^2 . You have determined that the coefficient of determination (R²) is .8. Calculate the F ratio used to test the significance of the regression.

- A. 4 B. 8 C. 12 D. 15 E. 18 (86F-120-8)

- C4. You are given ten pairs of values (X_i, Y_i), which are represented by the following model:

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

ϵ is a random error term with mean 0 and variance σ^2 . You have determined that the coefficient of determination for this regression is .90. Calculate the F ratio for testing the hypothesis $H_0: \beta = 0$.

- A. 24 B. 45 C. 60 D. 72 E. 81 (87S-120-10)

- C5. A straight line of the form $Y = \alpha + \beta X$ has been fitted by the method of least squares to six pairs of observations (X_i, Y_i), yielding the fitted values \hat{Y}_i . You are given:

X	7.0	14.0	15.5	25.1	21.4	32.8
Y	14.2	13.4	13.2	26.7	17.7	25.5
\hat{Y}	11.3	15.4	16.2	21.8	19.7	26.3

Determine the F ratio used to test the hypothesis that $\beta = 0$.

- A. 1 B. 3 C. 7 D. 11 E. 14 (87F-120-5)

- C6. According to Pindyck and Rubinfeld, what is the standard deviation of the errors in the following forecasts?

Period	1	2	3	4	5	6	7
Observed value	100	105	115	107	123	135	138
Forecast value	101	107	112	117	124	130	139

(89-10-24-1)

- C7. In their text, Pindyck and Rubinfeld propose using \bar{R}^2 rather than R² when evaluating regression models. Why do they favor using \bar{R}^2 over R²? (89-10-32a-1)

- C8. According to Pindyck and Rubinfeld, \bar{R}^2 is R² adjusted for degrees of freedom. (90-10-17-MC)

- C9. According to Pindyck and Rubinfeld, \bar{R}^2 can equal R². (90-10-17-MC)

- C10. According to Pindyck and Rubinfeld, is R² or \bar{R}^2 is a better measure of fit for a particular model? Why? (91-10-40a-1)

- C11. A sample of size 20 is fitted to a multiple linear regression model relating a dependent variable to five independent variables. The resulting F ratio used to test the hypothesis:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 \text{ is equal to } 21.$$

Determine R^2 .

- A. 35/41 B. 7/8 C. 15/17 D. 9/10 E. 126/139 (93S-120-10)

- C12. Twenty pairs of observations (X_i, Y_i) are represented by the following model:

$$Y_i = \alpha + \beta X_i + \epsilon_i, \quad i = 1, \dots, 20,$$

where the ϵ_i are independent, normally distributed random variables with mean 0 and variance σ^2 . You have determined:

$$\sum_i X_i = 100 \quad \sum_i Y_i = 120 \quad \sum_i X_i^2 = 2,100 \quad \sum_i Y_i^2 = 7,120 \quad \sum_i X_i Y_i = 2,520 \quad \hat{\beta} = 1.2$$

Calculate the F ratio for testing $H_0: \beta = 0$.

- A. 2 B. 4 C. 6 D. 8 E. 10 (93S-120-8)

- C13. You are given:

Source	Sum of Squares	Degrees of Freedom
Regression	1,115.11	2
Error	138.89	5
Total	1,254.00	7

Determine the coefficient of determination corrected for degrees of freedom (\bar{R}^2).

- A. .84 B. .89 C. .93 D. .97 E. 1.00 (95F-120-9)

- C14. You fit a simple linear regression model to five pairs of observations. The residuals for the first four observations are .4, -.3, 0, -.7, and the estimated variance of the dependent variable (Y) is:

$$\frac{\sum (Y_i - \bar{Y}_i)^2}{N - 1} = 1.5$$

Calculate R^2 .

- A. .82 B. .84 C. .86 D. .88 E. .90 (96S-120-3)

- C15. You fit a simple linear regression model to seven observations. You determine that $ESS = 218.680$ and $F = 2.088$. Calculate R^2 .

- A. .3 B. .4 C. .5 D. .6 E. .7 (98S-120-3)

- C16. You fit the regression model $Y_j = \alpha + \beta X_j + \epsilon_j$ to 11 observations. You are given that $R^2 = .85$. Determine \bar{R}^2 , the corrected R^2 .

- A. .77 B. .79 C. .80 D. .81 E. .83 (98S-120-4)

- C17. An actuary is developing a model to predict insurance sales. She considers a regression model with 8 predictors and takes a random sample of 27 observations on these 8 variables and insurance sales. You are given that $ESS = 76,893$ and $TSS = 192,068$. Calculate \bar{R}^2 , the corrected (adjusted) coefficient of determination.

A. .1 B. .4 C. .5 D. .6 E. .9 (98F-120-4)

- C18. You model the relationship between the unit price (Y) and the number of units sold (X) of a certain item by the piecewise linear regression model:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 (X_i - 50) D_i + \varepsilon_i,$$

where $D_i = 1$ if $X_i \geq 50$ and $D_i = 0$ if $X_i < 50$. You have collected the following weekly data:

Week (i)	1	2	3	4	5	6	7	8
Unit price \$ (Y_i)	26	44	45	14	48	36	25	38
Units sold (X_i)	65	34	40	80	30	57	72	48

You are given that $\hat{\beta}_1 = 58.914$ and $\hat{\beta}_2 = -.389$. Determine $\hat{\beta}_3$.

A. -.5 B. -.4 C. -.3 D. -.2 E. -.1 (98F-120-8)

- C19. You fit the following model to 20 observations:

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

You determine that $R^2 = .64$. Calculate the value of the F statistic used to test for a linear relationship.

A. < 30 B. ≥ 30 but < 33 C. ≥ 33 but < 36 D. ≥ 36 but < 39 E. ≥ 39 (00S-4-1)

- C20. You fit the following model to 48 observations:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

You are given:

Source of Variation	Degrees of Freedom	Sum of Squares
Regression	3	103,658
Error	44	69,204

Calculate \bar{R}^2 , the corrected R^2 .

A. .57 B. .58 C. .59 D. .60 E. .61 (00S-4-31)

- C21. Which of the following is not an objection to the use of R^2 to compare the validity of regression results under alternative specifications of a multiple linear regression model?

- A. The F statistic used to test the null hypothesis that none of the explanatory variables helps explain variation of Y about its mean is a function of R^2 and degrees of freedom.
- B. Increasing the number of independent variables in the regression equation can never lower R^2 and is likely to raise it.
- C. When the model is constrained to have zero intercept, the ratio of regression sum of squares to total sum of squares need not lie within the range $[0, 1]$.
- D. Subtracting the value of one of the independent variables from both sides of the regression equation can change the value of R^2 while leaving the residuals unaffected.
- E. Because R^2 is interpreted assuming the model is correct, it provides no direct procedure for comparing alternative specifications. (02F-4-30)