

PAST CAS AND SoA EXAMINATION QUESTIONS

A. General Linear Model

- A1. Observations y_1, y_2, \dots, y_n are assumed to come from a model with $E(y_i) = \theta + 2 \ln x_i$, where θ is an unknown parameter and x_1, x_2, \dots, x_n are given constants. What is the least-squares estimate of θ ?

A. $\sum_{i=1}^n y_i/N$ B. $\sum_{i=1}^n x_i/N$ C. $\sum_{i=1}^n \ln(x_i/N)$
 D. $\sum_{i=1}^n y_i \ln(x_i/N)$ E. $\sum_{i=1}^n [y_i - 2 \ln(x_i)]/N$ (82S-2-28)

- A2. For the regression model $y_i = \beta \ln x_i + \varepsilon_i$, $i = 1, 2$, assume that $x_1 = e$ and $x_2 = e^2$ and that ε_1 and ε_2 are independent random variables with mean zero and unknown variance σ^2 . What is the least-squares estimator for β based on (x_1, y_1) and (x_2, y_2) ?

A. $(y_1 + 2y_2)/3$ B. $(y_1 + y_2)/3$ C. $(y_1 + 2y_2)/5$ D. $[\exp(y_1) + \exp(y_2)]/3$
 E. $[\exp(y_1) + \exp(2y_2)]/3$ (83S-2-9)

- A3. You are given the following model:

$$Y = \beta X^2 + \varepsilon,$$

where ε is a random error term with mean 0 and variance σ^2 . Determine the least-squares estimator of β .

A. $\frac{\sum X_i Y_i}{\sum X^2}$ B. $\frac{\sum X_i^2 Y_i}{\sum X_i^4}$ C. $\frac{\sum (X_i - \bar{X})^2 Y_i^2}{\sum (X_i - \bar{X})^2}$ D. $\frac{\sum (X_i - \bar{X})^2 Y_i}{\sum (X_i - \bar{X})^4}$
 E. $\frac{\sum (X_i - \bar{X})^2 Y_i^2}{\sum (X_i - \bar{X})^4}$ (84F-3-10)

- A4. You are representing eight observations (X_i, Y_i) by the following model:

$$Y = \alpha \beta^x u,$$

where u is a random error term such that $\ln u$ is normally distributed with mean 0 and variance σ^2 . You have determined

$\sum_{i=1}^8 X_i = 36.00$	$\sum_{i=1}^8 \ln X_i = 10.60$	$\sum_{i=1}^8 X_i^2 = 204.00$	$\sum_{i=1}^8 (\ln X_i)^2 = 17.52$
$\sum_{i=1}^8 Y_i = 30.50$	$\sum_{i=1}^8 \ln Y_i = 8.61$	$\sum_{i=1}^8 X_i Y_i = 184.00$	$\sum_{i=1}^8 (\ln X_i)(\ln Y_i) = 15.14$
$\sum_{i=1}^8 X_i (\ln Y_i) = 52.36$			

You are to minimize $S = \sum_{i=1}^8 (\ln u_i)^2 = 17.52$. Calculate $\hat{\beta}$.

A. .32 B. 1.07 C. $e^{-.32}$ D. $e^{1.07}$ E. $e^{1.11}$ (87S-120-8)

A5. You are given:

i	1	2	3	4	5
Y _i	4	9	36	64	144
X _{i1}	1	2	6	3	4
X _{i2}	2	3	1	5	6

You are to represent these data by the following model:

$$\sqrt{Y_i} = \frac{1}{\sqrt{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i}}$$

where ϵ_i is a random error term with mean 0 and variance σ^2 . You are to fit the model by minimizing $\sum \epsilon_i^2$. You have determined:

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 6 & 3 & 4 \\ 2 & 3 & 1 & 5 & 6 \end{pmatrix}$$

$$(X'X)^{-1} = \frac{1}{1,263} \begin{pmatrix} 2,141 & -299 & -274 \\ -299 & 86 & 7 \\ -274 & 7 & 74 \end{pmatrix}$$

Calculate the least-squares estimate of β_1 .

- A. -.07 B. -.05 C. .65 D. 1.29 E. 1.59 (87F-120-10)

A6. Let $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be pairs of observations. The curve $y = \beta e^x$ is to be fitted to this data set. What is the least-squares estimate for β ?

- A. $\left[\sum_{i=1}^n y_i \right] / \left[\sum_{i=1}^n \exp(x_i) \right]$ B. $\left[\sum_{i=1}^n y_i \exp x_i \right] / \left[\sum_{i=1}^n \exp(2x_i) \right]$
 C. $\left[\sum_{i=1}^n y_i \exp(x_i) \right] / \left[\sum_{i=1}^n \exp(x_i) \right]$ D. $\left[\sum_{i=1}^n y_i \right] / \left[\sum_{i=1}^n \exp(2x_i) \right]$
 E. $\left[\sum_{i=1}^n y_i \right] / \left[\exp(x_i) \right]$ (88S-110-15)

A7. A study was made of the number (Y) of bacteria present in a culture after X days. The following data were collected:

X	0	1	2	3	4	5	6
Y	32	47	65	92	132	190	275
ln Y	3.47	3.85	4.17	4.52	4.88	5.25	5.62

You are given the following model: $Y = \alpha (\exp\{\beta X\})u$, where u is a random error term such that

ln u is normally distributed with mean 0 and variance σ^2 . You are to minimize $S = \sum_{i=0}^6 (\ln u_i)^2$.

Calculate $\hat{\alpha}$.

- A. 1 B. 3 C. 11 D. 21 E. 32 (88F-120-10)

$$y = \alpha \exp(\beta x) \epsilon$$

$$\ln y = \ln \alpha + \beta x + \ln \epsilon$$

A8. You are given the following regression model:

$$Y = \alpha \exp(2X) + \epsilon$$

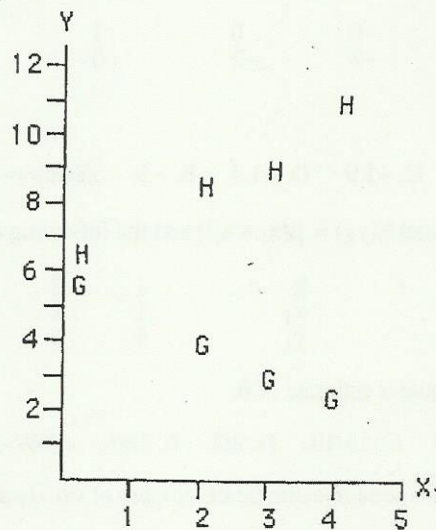
where ϵ is a random error term with mean 0 and variance σ^2 . Which of the following is the correct expression for the least-squares estimate of α ?

- A. $\exp(\bar{Y} - 2\bar{X})$ B. $\frac{\sum Y_i \exp(2X_i)}{\sum \exp(4X_i)}$ C. $\frac{\sum Y_i \exp[2(X_i - \bar{X})]}{\sum \exp[4(X_i - \bar{X})]}$
 D. $\frac{\bar{Y} \sum \exp(2X_i)}{\sum \exp(4X_i)}$ D. $\frac{\bar{Y}}{\sum \exp(2X_i)}$ (91S-120-12)

A9. Let $(x_1, y_1), \dots, (x_n, y_n)$ be n pairs of observations. The curve $y = \beta \ln x$ is to be fitted to this data set. What is the least-squares estimate for β ?

- A. $\left[\sum_{i=1}^n y_i \ln x_i \right] / \left[\sum_{i=1}^n 2 \ln x_i \right]$ B. $\left[\sum_{i=1}^n y_i \ln x_i \right] / \sum_{i=1}^n (\ln x_i)^2$
 C. $\left[\sum_{i=1}^n y_i \ln x_i \right] / \left[\sum_{i=1}^n \ln x_i \right]$ D. $\left[\sum_{i=1}^n y_i \right] / \left[\sum_{i=1}^n (\ln x_i)^2 \right]$ E. $\left[\sum_{i=1}^n y_i \right] / \left[\sum_{i=1}^n \ln x_i \right]$
 (92S-110-7)

A10. You are studying the moisture levels of two foods, G and H, as a function of shelf time. The graph below plots moisture level (Y) against shelf time (X_1).



You define $X_2 = 0$ for food G and $X_2 = 1$ for food H. You use the following linear regression model to fit the eight data points in the graph:

$$Y_i = \beta_0 + \beta_1 X_{i2} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i, i = 1, \dots, 8.$$

where the ϵ_i 's are independent normally distributed with mean 0 and variance σ^2 . Which of the following is true?

- A. $\hat{\beta}_1 < 0, \hat{\beta}_3 < 0$ B. $\hat{\beta}_1 < 0, \hat{\beta}_3 > 0$ C. $\hat{\beta}_1 = 0, \hat{\beta}_3 = 0$ D. $\hat{\beta}_1 > 0, \hat{\beta}_3 < 0$
 E. $\hat{\beta}_1 > 0, \hat{\beta}_3 > 0$ (93S-120-12)

A11. You are given the following model:

$$Y_t = \exp[-(\beta_0 + \beta_1 X_t + \epsilon_t)]$$

Determine the expressions for the least-squares estimates of β_0 and β_1 .

A. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ $\hat{\beta}_1 = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2}$

B. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ $\hat{\beta}_1 = \frac{\sum (X_t - \bar{X}) \ln(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2}$

C. $\hat{\beta}_0 = (-1/N) \sum \ln Y_t - \hat{\beta}_1 \bar{X}$ $\hat{\beta}_1 = \frac{-\sum (X_t - \bar{X}) \ln Y_t}{\sum (X_t - \bar{X})^2}$

D. $\hat{\beta}_0 = (-1/N) \sum \ln Y_t - \hat{\beta}_1 \bar{X}$ $\hat{\beta}_1 = \frac{\sum \ln(X_t - \bar{X}) \ln(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2}$

E. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ $\hat{\beta}_1 = \frac{-\sum (X_t - \bar{X}) \ln Y_t}{\sum (X_t - \bar{X})^2}$ (Sample-120-3)

A12. You use the method of least squares to fit the model $y_t = \beta_0 + \beta_1(x_t + x_t^2) + \epsilon_t$ to the following data:

x_i	0	0	1	1	2	2
y_t	-3	-2	0	1	8	10

Determine $\hat{\beta}_0$.

- A. -2.9 B. -2.4 C. -1.9 D. -1.4 E. -9 (Sample-120-10)

A13. You are given the model $E(y_k) = \beta(x_1 + x_1^2)$ and the following data:

k	1	2	3
x_i	1	2	3
y_i	4	8	14

Calculate the least-squares estimate of θ .

- A. 13/92 B. 28/23 C. 13/10 D. 9/2 E. 56/5 (96W-110-9)

A14. The following model is used to estimate the amount of fire damage Y (in thousands):

$$\hat{Y}_i = 8 + 5X_{1i} + 2(X_{1i} - 4)X_{2i} + 9X_{3i} - 2X_{1i}X_{3i}, \text{ where}$$

X_{1i} = the distance from the nearest fire station, in kilometers

X_{2i} = 1 if $X_{1i} \geq 4$ and = 0 otherwise

X_{3i} = 1 if the city is A and = 0 if the city is B

For the fires that took place at least 4 kilometers from the fire station, determine the distance from the nearest fire station for which the average fire damage for city A and city B is the same.

- A. 4.25 B. 4.50 C. 8.50 D. 28.00 E. 29.00 (98S-120-7)

Pindyck/Rubinfeld 5

A15. You fit the model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + \varepsilon_i$ to the following data:

X_{2i}	1	1	-1	-1	0	0
X_{3i}	1	-1	1	-1	1	-1
Y_i	28	13	0	29	44	36

Determine the upper limit of the symmetric 95% confidence interval for β_2 .

- A. 22 B. 25 C. 30 D. 37 E. 45 (98F-120-6)

B. Tests for Hypotheses Involving More Than One Parameter

B1. You are determining the relationship of salary (Y) to experience (X_2) for both men ($X_3 = 1$) and women ($X_3 = 0$). You fit the model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + \epsilon_i$ to a set of observations for a sample of employees. You are given:

<u>Source of Variation</u>	<u>Degrees of Freedom (df)</u>	<u>Sum of Squares (SS)</u>
Regression	3	330.0117
Error	7	12.8156

You also fit the model $Y_{ii} = \beta_1^* + \beta_2^* X_{2i} + \epsilon_i^*$ to the observations. You are given:

<u>Source of Variation</u>	<u>Degrees of Freedom (df)</u>	<u>Sum of Squares (SS)</u>
Regression	1	315.0992
Error	9	27.7281

Determine the F ratio to test whether the linear relationship between salary and experience is identical for men and women

- A. .6 B. 2.0 C. 3.5 D. 4.1 E. 6.2 (95F-120-5) (Sample-4-35)

B2. You wish to find a model to predict insurance sales, using 27 observations and 8 variables labeled X_2, X_3, \dots, X_9 . The analysis of variance tables for two different models from these data follow. Model A contains all 8 independent variables; model B contains X_2 and X_3 only.

	<u>Model A</u>		
<u>Source</u>	<u>SS</u>	<u>df</u>	<u>MS</u>
Regression	115,175	8	14,397
Error	76,893	18	4,272
Total	192,068	26	

	<u>Model B</u>		
<u>Source</u>	<u>SS</u>	<u>df</u>	<u>MS</u>
Regression	65,597	2	32,798
Error	126,471	24	5,270
Total	192,068	26	

Calculate the F ratio for testing the hypothesis $H_0: \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$.

- A. 5.8 B. 4.5 C. 2.6 D. 1.9 E. 1.6 (9881-120-6) (Sample-4-12)

B3. You fit the model $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ to 10 observations. You are given that $RSS = 61.3$ and $TSS = 128$. You then fit the model $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$, with an additional variable X_4 , to the same data. For this new model, you determine that $RSS = 65.6$ and $TSS = 128$. Calculate the value of the F statistic to test $H_0: \beta_4 = 0$.

- A. .01 B. .41 C. 1.76 D. 4.30 E. 10.40 (98S-120-8)

- B4. You fit two linear regression models to two independent data sets as follows:

Data Set I: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$, where $i = 1, \dots, 14$

Data Set II: $Y_j = \alpha_1 + \alpha_2 X_{2j} + \alpha_3 X_{3j} + \epsilon_j$, where $j = 1, \dots, 16$

You wish to test the hypothesis H_0 that the two models are the same. You determine the following:

- i) $ESS = 14.0$ for data set I
- ii) $ESS = 464.4$ for data set II
- iii) $ESS = 1,270.6$ for the common regression model fit to the pooled data set of 30 observations

Determine the value of the F statistic for the Chow test for testing H_0 .

- A. 13 B. 14 C. 15 D. 20 E. 22 (98F-120-7)

- B5. The following models are fitted to 30 observations:

Model I: $Y = \beta_1 + \beta_2 X_2 + \epsilon$

Model II: $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$

You are given

$$\sum (Y - \bar{Y})^2 = 160 \quad \sum (X_2 - \bar{X}_2)^2 = 10$$

For model I, $\hat{\beta}_2 = -2$ and for model II, $R^2 = .70$. Determine the value of the F statistic used to test that β_3 and β_4 are jointly equal to zero.

- A. < 15 B. ≥ 15 but < 18 C. ≥ 18 but < 21 D. ≥ 21 but < 24 E. ≥ 24
(00S-4-9)

- B6. You are given the following two regression models, each based on a different population of data:

Model A: $Y_i = A_1 + A_2 X_{2i} + A_3 X_{3i} + \epsilon_i$, where $i = 1, 2, \dots, 30$

Model B: $Y_j = B_1 + B_2 X_{2j} + B_3 X_{3j} + \epsilon_j$, where $j = 1, 2, \dots, 50$

You assume that the variances of the two models are equal and pool the data into one model:

Model G: $Y_p = G_1 + G_2 X_{2p} + G_3 X_{3p} + \epsilon_p$, where $p = 1, 2, \dots, 80$

You calculate R_{model}^2 and the error sum of squares, denoted as ESS_{model} , for all three models. Which of the following is the F statistic for testing the hypothesis that model A is identical to model B?

- A. $F_{3,74} = \frac{(ESS_G - ESS_A - ESS_B)/3}{(ESS_A + ESS_B)/74}$ B. $F_{6,77} = \frac{(ESS_G - ESS_A - ESS_B)/6}{(ESS_A + ESS_B)/77}$
- C. $F_{6,74} = \frac{(ESS_G - ESS_A - ESS_B)/6}{(ESS_A + ESS_B)/74}$ D. $F_{3,74} = \frac{(R_G^2 - R_A^2 - R_B^2)/3}{(R_A^2 + R_B^2)/74}$
- E. $F_{6,77} = \frac{(R_G^2 - R_A^2 - R_B^2)/6}{(R_A^2 + R_B^2)/77}$ (00F-4-21)

- B7. A professor ran an experiment in three sections of a psychology course to show that the more digits in a number, the more difficult it is to remember. The following variables were used in a multiple regression:

X_2 equals the number of digits in the number.
 X_3 is 1 if student was in section 1, 0 otherwise.
 X_4 is 1 if student was in section 2, 0 otherwise.
 Y is the percentage of students correctly remembering the number.

You are given:

- A total of 42 students participated in the study.
- The regression equation $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + \beta_4 X_3 + \beta_5 X_4 + \varepsilon$ was fit to the data and resulted in $R^2 = .940$
- A second regression equation $Y = \gamma_1 + \gamma_2 X_2 + \gamma_3 X_2^2 + \varepsilon^*$ was fit to the data and resulted in $R^2 = .915$.

Determine the value of the F statistic used to test whether class section is a significant variable.

- A. 5.4 B. 7.3 C. 7.7 D. 7.9 E. 8.3 (01S-4-5)

B8.

Three models have been fit to 20 observations:

$$\begin{array}{ll} \text{Model I:} & Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \\ \text{Model II:} & Y = \beta_1 + \beta_2 (X_2 + X_3) + \varepsilon \\ \text{Model III:} & Y - X_3 = \beta_1 + \beta_2 (X_2 - X_3) + \varepsilon \end{array}$$

You are given:

Model	I	II	III
ESS	484	925	982

Calculate the value of the F statistic used to test the hypothesis $H_0: \beta_2 + \beta_3 = 1$.

- A. < 15 B. ≥ 15 but < 16 C. ≥ 16 but < 17 D. ≥ 17 but < 18 E. ≥ 18 (01F-4-21)

B9.

For the multiple regression model $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$, you are given:

- $N = 3,120$
- $TSS = 15,000$
- $H_0: \beta_4 = \beta_5 = \beta_6 = 0$
- $R_{UR}^2 = .38$
- $RSS_R = 5,565$

Determine the value of the F statistic for testing H_0 .

- A. < 10 B. ≥ 10 but < 12 C. ≥ 12 but < 14 D. ≥ 14 but < 16 E. ≥ 16 (02F-4-27)

- B10. At the beginning of each of the past 5 years, an actuary has forecast the annual claims for a group of insureds. The table below shows the forecasts (X) and the actual claims (Y). A two-variable linear regression model is used to analyze the data.

t	X_t	Y_t
1	475	254
2	254	463
3	463	515
4	515	567
5	567	605

You are given:

- i) The null hypothesis is $H_0: \alpha = 0, \beta = 1$.
- ii) The unrestricted model fit yields $ESS = 69,843$.

Which of the following is true regarding the F test of the null hypothesis?

- A. The null hypothesis is not rejected at the .05 significance level.
 - B. The null hypothesis is rejected at the .05 significance level, but not at the .01 level.
 - C. The numerator has 3 degrees of freedom.
 - D. The denominator has 2 degrees of freedom.
 - E. The F statistic cannot be determined from the information given. (03F-4-20)
- B11. You are given the following information about a linear regression model:
- i) The unit of measurement is a region, and the number of regions in the study is 37.
 - ii) The dependent variable is a measure of workers compensation frequency, while the three independent variables are a measure of employment, a measure of unemployment rate and a dummy variable indicating the presence or absence of vigorous cost containment efforts.
 - iii) The model is fitted separately to the group of 18 largest regions and to the group of 19 smallest regions (by population). The ESS resulting from the first fit is 4,053, while the ESS resulting from the second fit is 2,087.
 - iv) The model is fitted to all 37 regions, and the resulting ESS is 10,374.

The null hypothesis to be tested is that the pooling of the regions into one group is appropriate. Which of the following is true?

- A. The F statistic has 4 numerator degrees of freedom and 29 denominator degrees of freedom, and it is statistically significant at the 5% significance level.
- B. The F statistic has 4 numerator degrees of freedom and 29 denominator degrees of freedom, and it is not statistically significant at the 5% significance level.
- C. The F statistic has 4 numerator degrees of freedom and 33 denominator degrees of freedom, and it is not statistically significant at the 5% significance level.
- D. The F statistic has 8 numerator degrees of freedom and 33 denominator degrees of freedom, and it is statistically significant at the 5% significance level.
- E. The F statistic has 8 numerator degrees of freedom and 33 denominator degrees of freedom, and it is not statistically significant at the 5% significance level. (04F-4-19-1)

B12. You use the model $Y = \alpha + \beta X + \varepsilon$ to analyze the following data:

i	X_i	Y_i
1	1	2.8
2	2	2.9
3	3	3.6
4	4	4.7
5	5	6.2

You are given:

- i) The null hypothesis is $H_0: \alpha = \beta$.
- ii) The unrestricted model fit yields $\hat{\alpha} = 1.46$ and $\hat{\beta} = .86$.
- iii) The restricted model fit yields $\hat{\alpha} = \hat{\beta} = .993$.

Which of the following is true regarding the F test of the null hypothesis?

- A. The null hypothesis is not rejected at the 5% significance level.
- B. The null hypothesis is rejected at the 5% significance level but not at the 1% level.
- C. The numerator has two degrees of freedom.
- D. The denominator has two degrees of freedom.
- E. The F statistic cannot be determined from the information given. (05-ASM-7)

C1. Other Topics

C1. You analyzed your monthly electricity bills and estimated the following relationship:

$$\hat{Y}_t = 90 + 30X_{t1} + .15x_{t2} + 20X_{t1}X_{t3} + 10X_{t4}, \text{ where}$$

X_{t1} equals 1 if the month is January, February, June, July, August, or December and 0 otherwise.

X_{t2} equals the year minus 1987.

X_{t3} equals 1 if the temperature exceeded 90°F for two or more days in the month, or if a snowstorm occurred in the month and 0 otherwise.

X_{t4} equals the number of snowstorms in the month.

Calculate your expected average monthly electricity bill for February and March 1995, given that one snowstorm occurs in February and two snowstorms occur in March.

- A. 131 B. 141 C. 151 D. 156 E. 161 (Sample-120-7)

C2. An insurance company uses a model to predict Y_i = daily phone sales for personnel located in New York, Chicago, and Portland. The model is:

$$\hat{Y} = 12 + 3X_i + 2(X_i - 5)D_{1i} - 2(X_i - 2)D_{2i} + (X_i - 5)D_{3i}, \text{ where}$$

X_i = salesperson's years of experience

D_{1i} = 1 if $X_i \geq 5$ and = 0 otherwise

D_{2i} = 1 if New York and = 0 otherwise and D_{3i} = 1 if Chicago and = 0 otherwise

Calculate the predicted phone sales for a salesperson with 7 years of experience located in Chicago.

- A. 32 B. 33 C. 35 D. 37 E. 39 (96S-120-9)

C3. You are given the following model.

$$\begin{aligned} E(C_t) &= \beta_1 + \beta_2 Y_t & 0 < t \leq t_0 \\ E(C_t) &= (\beta_1 - \beta_3 Y_{t_0}) + (\beta_2 + \beta_3) Y_t & t_0 < t \leq t_1 \\ E(C_t) &= (\beta_1 - \beta_3 Y_{t_0} - \beta_4 Y_{t_1}) + (\beta_2 + \beta_3 + \beta_4) Y_t & t > t_1 \end{aligned}$$

Which of the following is true?

- A. This model is not considered a spline function.
 B. This model is discontinuous, with two structural breaks.
 C. This model is continuous, with one structural break.
 D. Dummy variables are used to account for shifts in the intercept.
 E. This model is continuous, with two structural breaks. (9881-120-13)

- C4. Your claims manager has asserted that a procedural change in the claims department implemented on January 1, 1997 immediately reduced claim severity by 20 percent. You use a multiple regression model to test this assertion. For the dependent variable (Y) you calculate the average claim costs on closed claims by year during 1990–99. You define the variable X as the year. You also define a variable D as:

$$D = \begin{cases} 0 & \text{for years 1996 and prior} \\ 1 & \text{for years 1997 and later} \end{cases}$$

Assuming a lognormal error component and constant inflation over the entire period, which of the following models would be used to test the assertion?

- A. $Y = \alpha_1^D \beta_1^X \epsilon$ B. $Y = \alpha_1 \alpha_2^D \beta_1^X \epsilon$ C. $Y = \alpha_1 \beta_1^X \beta_2^{XD} \epsilon$ D. $Y = \alpha_1 \alpha_2^D \beta_1^X \beta_2^{XD} \epsilon$
 E. $Y = \alpha_1 \alpha_2^D X^{\beta_1} \epsilon$ (01S-4-24)

- C5. You study the impact of education and number of children on the wages of working women using the following model:

$$Y = a + b_1 E + b_2 F + c_1 G + c_2 H + \epsilon, \text{ where}$$

- $Y = \ln(\text{wages})$
 $E = \begin{cases} 1 & \text{if the woman has not completed high school} \\ 0 & \text{if the woman has completed high school} \\ -1 & \text{if the woman has postsecondary education} \end{cases}$
 $F = \begin{cases} 1 & \text{if the woman has completed high school} \\ 0 & \text{if the woman has not completed high school} \\ -1 & \text{if the woman has postsecondary education} \end{cases}$
 $G = \begin{cases} 1 & \text{if the woman has no children} \\ 0 & \text{if the woman has 1 or 2 children} \\ -1 & \text{if the woman has more than 2 children} \end{cases}$
 $H = \begin{cases} 1 & \text{if the woman has 1 or 2 children} \\ 0 & \text{if the woman has no children} \\ -1 & \text{if the woman has more than 2 children} \end{cases}$

Determine the expected difference between $\ln(\text{wages})$ of a working woman who has postsecondary education and more than 2 children and $\ln(\text{wages})$ of the average for all working women.

- A. $a - b_1 - b_2$ B. $b_1 + b_2$ C. $-b_1 - b_2$ D. $a - b_1 - b_2 + c_2$ E. $-b_1 - b_2 - c_1 - c_2$
 (02F-4-20)

- C6. You are given:

- i) Y_{ij} is the loss for the j th insured in the i th group in year t .
 ii) \bar{Y}_{it} is the mean loss in the i th group in year t .
 iii) $X_{ij} = \begin{cases} 0 & \text{if the } j\text{th insured is in the first group } (i = 1) \\ 1 & \text{if the } j\text{th insured is in the second group } (i = 2) \end{cases}$
 iv) $Y_{2ij} = \delta + \phi Y_{1ij} + \theta X_{ij} + \epsilon_{ij}$, where $i = 1, 2$ and $j = 1, 2, \dots, n$
 v) $\bar{Y}_{21} = 30, \bar{Y}_{22} = 37, \bar{Y}_{11} = 40, \bar{Y}_{12} = 41$
 vi) $\hat{\phi} = .75$

Determine the least-squares estimate of θ .

- A. 5.25 B. 5.50 C. 5.75 D. 6.00 E. 6.25 (03F-4-9)