College of Science.
Department of Statistics & Operations Research

جامعة الملك سعود
King Saud University
كلية العلوم
قسم الإحصاء وبحوث العمليات

First Midterm Exam
Academic Year 1443-1444 Hijri- First Semester

Exam Information معلومات الامتحان			
Course name	تحليل انحدار		اسم المقرر
Course Code	332 إحص		رمز المقرر
Exam Date	2023-02-27	1444-08-07	تاريخ الامتحان
Exam Time	08: 00 AM		وقت الامتحان
Exam Duration	3 hours	ثلاث ساعات	مدة الامتحان
Classroom No.			رقم قاعة الاختبار
Instructor Name	د. وليد + د. محمود + د.ونام		اسم استاذ المقرر

Student Information معلومات الطالب		
Student's Name		اسم الطالب
ID number		الرقم الجامعي
Section No.		رقم الشعبة
Serial Number		الرقم التسلسلي

General Instructions:

- Your Exam consists of 8 PAGES (except this paper)
- Keep your mobile and smart watch out of the classroom.
-

تعليمات عامة:

- عدد صفحات الامتحان 8 صفحة. (باستثناء هذه الورقة)
- يجب إبقاء الهواتف والساعات الذكية خارج قاعة الامتحان.
- سيتم تصحيح ورقة المصحح الآلي فقط ولن ينظر لورقة الأسئلة لذا يلزمك التأكد من نقل اجاباتك بدقه.

هذا الجزء خاص بأستاذ المادة

This section is ONLY for instructor

#	Course Learning Outcomes (CLOs)	Related Question (s)	Points	Final Score
1	K1	Q1	6	10
2	K1	Q2	10	
3	S1	Q3	8	
4	K2	Q4	16	
5				40
6				
7				
8				

EXAM COVER PAGE

6

Problem 1 (6 points)

Suppose a sample of size 10 is used to estimate a simple linear regression model $Y = \beta_0 + \beta_1 X + \epsilon$ and obtain a 90% level confidence interval for the slope coefficient of (0.021, 0.045). Based on the given information, complete the following statements (keep three decimal digits during the calculations):

1. The point estimate for the slope is (1 points)
2. The standard error for the slope is (1 point)
3. The value of the test statistic for testing the slope is equal to 0 is (2 points)
4. The decision of the test in (3) at 5% level of significance is (2 points)

(1) $b_1 = \dfrac{0.021 + 0.045}{2} = 0.033$

(2) $b_1 - t(1-\alpha/2, n-2)\, s(b_1) = 0.021$

$t(1-0.5, 8)$

$t(0.95, 8) = 1.859548$

$0.033 - t(1-\frac{0.1}{2}, 8)\, s(b_1) = 0.021$

$0.033 - 1.859548\, s(b_1) = 0.021$

$-1.859548\, s(b_1) = 0.021 - 0.033$

$s(b_1) = \dfrac{-0.012}{-1.859548} = 0.00645$

(3) $T = \dfrac{b_1 - 0}{s(b_1)} = \dfrac{0.033}{0.00645} = \cancel{5.116}\ 5.116$

(4) decision of test in (3) 5% level sig $\alpha = 0.05$

$t(1-\alpha/2, n-2) = t(0.975, 8) = 2.306$

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$T = 5.116 \notin (-2.306, 2.306)$

we Reject H_0

8

Problem 3 (8 points)

A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data are as follows, where X is the year (coded) and Y is sales in thousands

i	1	2	3	4	5	6	7	8	9	10
X_i	0	1	2	3	4	5	6	7	8	9
Y_i	98	135	162	178	221	232	283	300	374	395

1. Fit the simple linear model **in matrix form** of the given data (1 points)
2. Fit the simple linear model **in matrix form** under the following transformations $Y' = \sqrt{Y}, Y' = \log_{10}(Y)$ and $Y' = \frac{1}{Y}$. (3 points)
3. Compare between the results in (1) and (2) using the coefficient of determination. (2 points)
4. What is the expected Y **for each model** when $X = 10$? (2 points)

احداثي

① $\hat{Y} = 91.56 + 32.50$

② $Y' = \sqrt{Y}$ $Y1 = sqrt(Y)$

model1 = lm(Y1 ~ X)

$\boxed{\widehat{\sqrt{Y}} = 10.261 + 1.076\,X}$

$Y' = \log_{10}(Y)$ $Y_2 = Log10(Y)$

model2 = lm(Y2 ~ X)

~~$\log_{10}(\hat{Y}) = 2.05236$~~

$\boxed{\log_{10}(\hat{Y}) = 2.05236 + 0.06369\,X}$

$Y' = \frac{1}{Y}$ $Y_3 = \frac{1}{Y}$

model3 = lm(Y3 ~ X)

$\boxed{\frac{1}{\hat{Y}} = 0.00838 + (-0.00074)\,X}$

③ compare between (1) and (2) using R^2 coefficient of determi

(1) $R^2 = 0.97977$

(2) $R^2_1 = 0.9891$

$R^2_2 = 0.9767$

$R^2_3 = 0.8836$

The best coefficient of determination is $R^2_1 = 0.9891$

④ 95 % CI for the slop
confint(model, Level=0.95)

$0.4428 < b_1 < 0.9027$

⑤ Test intercept is equal zero using t-test
$H_0: \beta_0 = 0$ vs $H_1: \beta_0 \neq 0$

$$T = \frac{b_0 - 0}{s(b_0)} = \frac{9.3}{3.435} = \underline{2.708}$$

$t(1-\alpha/2, n-2) = t(0.975, 10) = \underline{2.228}$

$T = 2.708 \notin (-2.228, 2.228)$
we Reject H_0

10

Problem 2 (10 points)

A college bookstore must order books two months before each semester starts. They believe that the number of books that will ultimately be sold for any particular course is related to the number of students registered for the course when the books are ordered. They would like to develop a linear regression equation to help plan how many books to order. From past records, the bookstore obtains the number of students registered, X, and the number of books actually sold for a course, Y, for 12 different semesters. These data are below.

Semester	Students (X)	Books (Y)
1	36	31
2	28	29
3	35	34
4	39	35
5	30	29
6	30	30
7	31	30
8	38	38
9	36	34
10	38	33
11	29	29
12	26	26

1. Estimate the simple linear regression equation. (1 point)
2. Calculate SSE, SSTO and SSR (3 points)
3. Calculate the variances of the estimators of the in part (a) (2 points)
4. Estimate 95% confidence interval for the slope β_1 of the model (2 points)
5. Test if the intercept is equal to zero using t test (2 points)

① $\hat{Y} = 9.3 + 0.6727X$ model = lm(Y ~ x)

② anova(model)

$SSE = 23.436$, $SSR = 99.564$

$SSTO = SSR + SSE$

$= 99.564 + 23.436 = 123$

③ summary(model)

$s(b_0) = 3.4346 \Rightarrow V(b_0) = 3.4346^2$

$V(b_0) = 11.7965$

$s(b_1) = 0.1032 \Rightarrow V(b_1) = 0.1032^2$

$V(b_1) = 0.01065$

④ what is the expected y for each model when $X = 10$

$\hat{Y} = 91.56 + 32.50(10) = \underline{416.56}$

$\sqrt{\hat{Y}} = 10.261 + 1.076(10) = 21.021^2 \Rightarrow \underline{441.88}$

$\text{Log}_{10}(\hat{Y}) = 2.05236 + 0.06369(10) = 10^{2.68926} = \underline{488.945}$

$\frac{1}{\hat{Y}} = 0.00838 + (-0.00074)(10) = \quad \underline{1020.41}$

$\hat{Y} = \frac{1}{0.00838 + (-0.00074)(10)} = \underline{1020.41}$

Problem 4 (Use datafile properties.txt) (16 points)

I. For the simple regression model use Matrix form to prove the following

ماهو داخل في البيانات

1. $E(\mathbf{b}) = \boldsymbol{\beta}$ (2 points)
2. Obtain the least square estimator which minimize $Q = e'e$. (3 points)

II. A commercial real estate company evaluates vacancy rates, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data are taken from 81 suburban commercial properties that are the newest, best located, most attractive, and expensive for five specific geographic areas. The considered variables are the age (X_1), operating expenses and taxes (X_2), vacancy rates (X_3), and rental rates (Y).

Consider the multiple regression model to the data for two predictor variables with normal error terms

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, i = 1, \ldots, n.$$

1. State the regression model **in matrix form**. (1 point)
2. State the estimated regression function. (1 point)
3. Interpret b_1, b_2 and b_3. (2 points)

اختبار الحل ؟

4. Test whether there is a regression relation; use $\alpha = 0.05$. State the alternatives, decision rule, and conclusion. (2points)
5. Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X_1; with X_2 given X_1; and with X_3 given X_1, X_2. (2 points)
6. Use the extra sum of squares (F^* test statistic) and $\alpha = 0.01$ to test whether X_3 can be dropped from the regression model given that X_1 and X_2 are retained. (State the alternatives, decision rule, and conclusion). (2 points)
7. A new property is to be evaluated, with $X_{h1} = 5, X_{h2} = 8.25$ and $X_{h3} = 0$. Obtain a 95% prediction interval for the rental rate for this property. (1 point)

$\boxed{\text{I}}$ 1. $E(b) = \beta$ $\quad = E[(x'x)^{-1}x'y]$

$= (x'x)^{-1}x'E(y) \quad = (x'x)^{-1}x'x\beta$

$= I\beta = \beta.$

2. obtain the LSM $\quad Q = e'e$

$e = (Y - X\beta)$

$Q = e'e \quad = (Y - X\beta)'(Y - X\beta) \quad = \cancel{Y'Y - \beta'X'}$

$= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta$

$= Y'Y - 2\beta'X'Y - 2Y'X\beta + \beta'X'X\beta$

$\dfrac{\partial Q}{\partial \beta} = -2X'Y - 2Y'X + X'X\beta$

$= -2X'Y - 2Y'X + X'Xb = 0$

$\Rightarrow \boxed{b = (X'X)^{-1}X'Y}$

① regression model in matrix

II

$\underset{n\times1}{Y} = \underset{n\times2}{X}\ \underset{4\times1}{\beta} + \underset{n\times1}{\varepsilon}$

② estimated regression function

$$\boxed{\hat{Y} = 11.54 + (-0.119)x_1 + 0.446x_2 + 2.62x_3}$$

③ interpret

b1 = The changes in value $\underline{Y}$ when x1 increase by one unit with x2, x3 no change

b2 = The changes in value $\underline{Y}$ when x2 increase by one unit with x1, x3 no change.

b3 = The changes in value Y when x3 increase by one unit with x1, x2 no change.

④ Test $\alpha = 0.05$

$H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$ => P-value = 8.31e-06 < 0.05 we Reject H_0.

$H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$ => P-value = 3.50e-09 < 0.05 we Reject H_0.

$H_0: \beta_3 = 0$ vs $H_1: \beta_3 \neq 0$ => P-value = 0.0353 < 0.05 we Reject H_0.

⑤ ANOVA

	Df	SS	MS	F
X1	1	14.819	14.819	8.118
X2	1	72.802	72.802	39.8828
X3	1	8.381	8.381	4.5916
Residual	77	140.556	1.825	
Total	80	236.558		

$SSR(X_1) = 14.819$, $SSR(X_2|X_1) = 72.802$, $SSR(X_3|X_1, X_2) = 8.381$

⑥ F test and $\underline{\alpha = 0.01}$ to test X_3 can be dropped

$H_0: \beta_3 = 0$ vs $H_1: \beta_3 \neq 0$

$F(0.99, 1, 77) = 6.9759$

$F_0 = 4.5916 < 6.9759$ we Accept H_0

Yes, we can Remove X_3

6

$s(b_1)=\sqrt{\frac{MSE}{Sxx}}$

$b_1 \in (0.021, 0.045)$

$b_1 \pm t_{0.95,8} \times s(b_1)$

Problem 1 (6 points)

Suppose a sample of size 10 is used to estimate a simple linear regression model $Y = \beta_0 + \beta_1 X + \epsilon$ and obtain a 90% level confidence interval for the slope coefficient of (0.021, 0.045). Based on the given information, complete the following statements (keep three decimal digits during the calculations):

1. The point estimate for the slope is (1 points)
2. The standard error for the slope is (1 point)
3. The value of the test statistic for testing the slope is equal to 0 is (2 points)
4. The decision of the test in (3) at 5% level of significance is (2 points)

(1) $b_1 = \frac{0.021 + 0.045}{2} = 0.033$

(2) $b_1 - t_{1-\frac{\alpha}{2}, n-2}\, s(b_1) = 0.021$

$0.033 - 1.859 \times s(b_1) = 0.021$

(3) $T = \frac{b_1}{s(b_1)} = \frac{0.033}{0.00645} = 5.11627$ $\quad s(b_1) = 0.00645$

(4) $H_0: B_1 = 0$ vs $H_1: B_1 \neq 0$

$t_{(0.975,8)} = 2.306$

$T = 5.11627 \notin (-2.306, 2.306)$ Then we reject Ho

King Saud University

جامعة الملك سعود

College of Science.
Department of Statistics & Operations Research

كلية العلوم
قسم الإحصاء وبحوث العمليات

First Midterm Exam
Academic Year 1443-1444 Hijri- First Semester

Exam Information معلومات الامتحان			
Course name	تحليل انحدار		اسم المقرر
Course Code	332 إحص		رمز المقرر
Exam Date	2023-02-27	1444-08-07	تاريخ الامتحان
Exam Time	08: 00 AM		وقت الامتحان
Exam Duration	3 hours	ثلاث ساعات	مدة الامتحان
Classroom No.			رقم قاعة الاختبار
Instructor Name	د. وليد + د. محمود + د.ونام		اسم استاذ المقرر

Student Information معلومات الطالب		
Student's Name	عمر عبدالرحيم الزهراني	اسم الطالب
ID number	442101310	الرقم الجامعي
Section No.		رقم الشعبة
Serial Number		الرقم التسلسلي

General Instructions:

- Your Exam consists of 8 PAGES (except this paper)
- Keep your mobile and smart watch out of the classroom.
-

تعليمات عامة:

- عدد صفحات الامتحان 8 صفحة. (باستثناء هذه الورقة)
- يجب إبقاء الهواتف والساعات الذكية خارج قاعة الامتحان.
- سيتم تصحيح ورقة المصحح الآلي فقط ولن ينظر لورقة الأسئلة لذا يلزمك التأكد من نقل اجاباتك بدقه.

هذا الجزء خاص بأستاذ المادة

This section is ONLY for instructor

#	Course Learning Outcomes (CLOs)	Related Question (s)	Points	Final Score
1	K1	Q1	6	27
2	K1	Q2	10	
3	S1	Q3	8	
4	K2	Q4	16	
5				40
6				
7				
8				

EXAM COVER PAGE

① Regression model in Matrix

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & X_{31} \\ 1 & X_{12} & X_{22} & X_{32} \\ 1 & X_{13} & X_{23} & X_{33} \\ 1 & X_{14} & X_{24} & X_{34} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \hat{Y}_3 \\ \hat{Y}_4 \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & X_{31} \\ 1 & X_{12} & X_{22} & X_{32} \\ 1 & X_{13} & X_{23} & X_{33} \\ 1 & X_{14} & X_{24} & X_{34} \end{bmatrix} \begin{bmatrix} 11.54 \\ -0.119 \\ 0.4461 \\ 2.6204 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

⑦ $X_{h1} = 5$, $X_{h2} = 8.25$, $X_{h3} = 0$ obtain 95 % Prediction interval

(11.899 , 17.353)

```
newx = data.frame(x1 = 5 , x2 = 8.25 , x3 = 0)
Predict (model, newx , level = 0.95, int = " Predict ")
```

8 / 0.05

Problem 2 (10 points)

A college bookstore must order books two months before each semester starts. They believe that the number of books that will ultimately be sold for any particular course is related to the number of students registered for the course when the books are ordered. They would like to develop a linear regression equation to help plan how many books to order. From past records, the bookstore obtains the number of students registered, X, and the number of books actually sold for a course, Y, for 12 different semesters. These data are below.

Semester (n)	Students (x)	Books (y)
1	36	31
2	28	29
3	35	34
4	39	35
5	30	29
6	30	30
7	31	30
8	38	38
9	36	34
10	38	33
11	29	29
12	26	26

1. Estimate the simple linear regression equation. (1 point)
2. Calculate SSE, SSTO and SSR (3 points)
3. Calculate the variances of the estimators of the in part (a) (2 points)
4. Estimate 95% confidence interval for the slope of the model (2 points)
5. Test if the intercept is equal to zero using t test (2 points)

(1) model = lm(y ~ x)
summary(model)

$\Rightarrow y = 9.3 + 0.6727x$

(2) anova(model) $\Rightarrow$ SSE = 23.436, SSTo = 123, SSR = 99.564

(3) $MSE = \frac{SSE}{n-2} = \frac{23.436}{10} = 2.344$ ✗

(4) $b_1 \pm t_{1-\frac{\alpha}{2}, n-2} \times s(b_1)$ or confint(model, level = 0.95) $\Rightarrow$

$\Rightarrow b_1 \in (0.4427547, 0.9026998)$

(5) $H_0: b_0 = 0$ vs $H_1: b_0 \neq 0$ (by using (summary(model)))

Pvalue = 0.022 < 0.05

7

Problem 3 (8 points)

A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data are as follows, where X is the year (coded) and Y is sales in thousands

i	1	2	3	4	5	6	7	8	9	10
X_i	0	1	2	3	4	5	6	7	8	9
Y_i	98	135	162	178	221	232	283	300	374	395

1. Fit the simple linear model **in matrix form** of the given data (1 points)
2. Fit the simple linear model **in matrix form** under the following transformations $Y'_1 = \sqrt{Y}, Y'_2 = \log_{10}(Y)$ and $Y'_3 = \frac{1}{Y}$. (3 points)
3. Compare between the results in (1) and (2) using the coefficient of determination. (2 points)
4. What is the expected Y **for each model** when $X = 10$? (2 points)

① $y = 91.564 + 32.497x$

② $> y_1 = sqrt(y)$

$> model1 = lm(y_1 \sim x)$

$> model1$

$\Rightarrow y_1 = 10.261 + 1.076(x)$

$\Rightarrow y_2 = log_{10}(y)$

$> model2 = lm(y_2 \sim x)$

$> model2$

$\Rightarrow y_2 = 2.05236 + 0.06389(x)$

$\Rightarrow y_3 = 1/y$

$> model3 = lm(y_3 \sim x)$

$> model 3$

$\Rightarrow y_3 = 0.00838 - 0.0007446(x)$

③

Transformation	
y	0.9797666
$\sqrt{y}$	0.9891359
$\log_{10} y$	0.9766626
$\frac{1}{y}$	0.8836173

$\sqrt{y}$ is the best

b

(1) $\begin{bmatrix} 81 & 637 & 784.74 & 6.56 \\ 637 & 8829 & 1274 & 33.55 \\ 784.74 & 1274 & 8136.5 & 52.995 \\ 6.56 & 33.55 & 52.995 & .98 \end{bmatrix}$?

(2) $y_i = 11.5406 - \cancel{0}.119 x_1 + 0.44 x_2 + 2.62 x_3$

(3) $b_1 = -0.119$?

$b_2 = 0.446$

$b_3 = 2.62$

(4) using T

$H_0: B_1 = 0$ vs $H_1: B_1 \neq 0$ Pvalue = 2e-16 reject

$H_0: B_2 = 0$ vs $H_1: B_2 \neq 0$ Pvalue = 8.31-06 > 0.05 reject

$H_0: B_3 = 0$ vs $H_1: B_3 \neq 0$ Pvalue = reject

(5) by using F test $H_0: B_1 = B_2 = B_3 = 0$ H_1: --- $\neq 0$

reject

(b)

$$\begin{bmatrix} n & \sum x_{i1} & \sum x_2 & \sum x_3 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 & \sum x_1x_3 \\ \sum x_2 & \sum x_1x_2 & \sum x_2x_3 & \sum x_2^2 \\ \sum x_3 & \sum x_1x_3 & \sum x_2x_3 & \sum x_3^2 \end{bmatrix}$$

Problem 4 (Use datafile properties.txt) (16 points)

I. For the simple regression model use Matrix form to prove the following

1. $E(\mathbf{b}) = \boldsymbol{\beta}$ (2 points)
2. Obtain the least square estimator which minimize $Q = \boldsymbol{e}'\boldsymbol{e}$. (3 points)

II. A commercial real estate company evaluates vacancy rates, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data are taken from 81 suburban commercial properties that are the newest, best located, most attractive, and expensive for five specific geographic areas. The considered variables are the age (X_1), operating expenses and taxes (X_2), vacancy rates (X_3), and rental rates (Y).
Consider the multiple regression model to the data for two predictor variables with normal error terms

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, i = 1, \dots, n.$$

1. State the regression model **in matrix form**. (1 point)
2. State the estimated regression function. (1 point)
3. Interpret b_1, b_2 and b_3. (2 points)
4. Test whether there is a regression relation; use $\alpha = 0.05$. State the alternatives, decision rule, and conclusion. (2points)
5. Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X_1; with X_2 given X_1; and with X_3 given X_1, X_2. (2 points)
6. Use the extra sum of squares (F^* test statistic) and $\alpha = 0.01$ to test whether X_3 can be dropped from the regression model given that X_1 and X_2 are retained. (State the alternatives, decision rule, and conclusion). (2 points)
7. A new property is to be evaluated, with $X_{h1} = 5, X_{h2} = 8.25$ and $X_{h3} = 0$, Obtain a 95% prediction interval for the rental rate for this property. (1 point)

(I)

(1) $E(\hat{b}) = E(X'X)^{-1}X'y = \cancel{(XX')^{-1}}\ X' E(y) = (XX)^{-1} X'XB$

$= IB = B$

(2) $Q = \sum e_i^2 = e'e = (y - yB)'(y - XB)$

$Q = y'y - B'X'y - y'XB + B'X'XB$

$\Rightarrow y'y - 2B'X'y + B'X'XB$ Then we diffrentiat

$\Rightarrow \frac{d}{dB}(Q) = \begin{bmatrix} \frac{dQ}{dB_0} \\ \frac{dQ}{dB_1} \end{bmatrix} \Rightarrow \frac{d}{dB}(Q) = -2X'y + 2X'XB$

(2)

ⓦ $y = 91.544 + 32.497(10) \Rightarrow y = 416.\cancel{5}34$

$y_1 = 10.261 + 1.076(10) \Rightarrow y_1 = 21.021$ $y = ?$

$y_2 = 2.05236 + 0.06369(10) \Rightarrow y_2 = 2.68926$ $y = ?$

$y_3 = 0.00838 + 0.0007446(10) \Rightarrow y_3 = 0.015826$ $y = ?$

جامعة الملك سعود
King Saud University

College of Science.
Department of Statistics & Operations Research

كلية العلوم
قسم الإحصاء وبحوث العمليات

First Midterm Exam
Academic Year 1443-1444 Hijri- First Semester

Exam Information معلومات الامتحان			
Course name	تحليل انحدار		اسم المقرر
Course Code	332 إحص		رمز المقرر
Exam Date	2023-02-27	1444-08-07	تاريخ الامتحان
Exam Time	08: 00 AM		وقت الامتحان
Exam Duration	3 hours	ثلاث ساعات	مدة الامتحان
Classroom No.			رقم قاعة الاختبار
Instructor Name	د. وليد + د. محمود + د.ونام		اسم استاذ المقرر

Student Information معلومات الطالب		
Student's Name	عبد الإله عبد الله الخليفي	اسم الطالب
ID number	441102185	الرقم الجامعي
Section No.		رقم الشعبة
Serial Number		الرقم التسلسلي

General Instructions:

- Your Exam consists of 8 PAGES (except this paper)
- Keep your mobile and smart watch out of the classroom.
-

تعليمات عامة:

- عدد صفحات الامتحان 8 صفحة. (باستثناء هذه الورقة)
- يجب إبقاء الهواتف والساعات الذكية خارج قاعة الامتحان.
- سيتم تصحيح ورقة المصحح الآلي فقط ولن ينظر لورقة الأسئلة لذا يلزمك التأكد من نقل اجاباتك بدقه.

هذا الجزء خاص بأستاذ المادة

This section is ONLY for instructor

#	Course Learning Outcomes (CLOs)	Related Question (s)	Points	Final Score
1	K1	Q1	6	37
2	K1	Q2	10	
3	S1	Q3	8	
4	K2	Q4	16	
5				40
6				
7				
8				

EXAM COVER PAGE

4

Problem 1 (6 points)

Suppose a sample of size 10 is used to estimate a simple linear regression model $Y = \beta_0 + \beta_1 X + \epsilon$ and obtain a 90% level confidence interval for the slope coefficient of (0.021, 0.045). Based on the given information, complete the following statements (keep three decimal digits during the calculations):

$SE(\beta_1) = $ [illegible]

1. The point estimate for the slope is (1 points)
2. The standard error for the slope is (1 point)
3. The value of the test statistic for testing the slope is equal to 0 is (2 points)
4. The decision of the test in (3) at 5% level of significance is (2 points)

1) $\frac{0.021+0.045}{2} = 0.033$

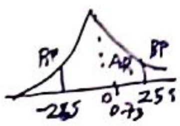
2) $S.E(\beta_1) = 0.045$

3) $t_0 = \frac{\beta_1 - 0}{S.E.(\beta_1)} = \frac{0.033}{0.045} = 0.733$

4)

$H_0: \beta_1 = 0$
$H_1: \beta_0 \neq 0$

qt(0.975, 1, 8) = 255.27



Since, 0.733 lies i AR. We don't reject H_0.

(10)

Problem 2 (10 points)

A college bookstore must order books two months before each semester starts. They believe that the number of books that will ultimately be sold for any particular course is related to the number of students registered for the course when the books are ordered. They would like to develop a linear regression equation to help plan how many books to order. From past records, the bookstore obtains the number of students registered, X, and the number of books actually sold for a course, Y, for 12 different semesters. These data are below.

Semester	Students X	Books Y
1	36	31
2	28	29
3	35	34
4	39	35
5	30	29
6	30	30
7	31	30
8	38	38
9	36	34
10	38	33
11	29	29
12	26	26

β_1 ~~SSE~~ $= \frac{MSe}{Sxx}$

$\beta_0 =$

1. Estimate the simple linear regression equation. (1 point)
2. Calculate SSE, SSTO and SSR (3 points)
3. Calculate the variances of the estimators of the in part (a) (2 points)
4. Estimate 95% confidence interval for the slope of the model (2 points)
5. Test if the intercept is equal to zero using t test (2 points)

1) >Model=lm(Y~X)

$Y = 9.3 + 0.6727X + \epsilon$

2) >anova(model)

SSE = 23.436

SSTO = 123

SSR = 99.564

3) ~~var~~ ~~SStudents~~ MSE $(X'X)^{-1}$

$Var[\beta_0] =$ ~~[illegible]~~ 117.96

$Var[\beta_1] =$ ~~[illegible]~~ 0.1065

4) 95% C.I. for β_1

```
> confint (model, lvl=0.95)
```

$\beta_1 \in (0.4428, 0.9027)$

5) 1- hypotheses:

$H_0: \beta_0 = 0$

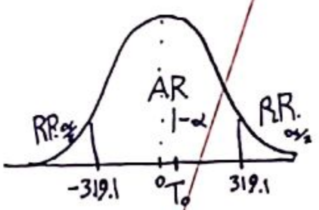
$H_1: \beta_0 \neq 0$

2- test Statistic:

```
> Summary(model)
```

$T_0 = \frac{\beta_0 - 0}{S.E.(\beta_0)} = 2.708$

3- Critical Region:



$t(0.975, 1, 10) = qt(0.957, 1, 10) = 319.1$

4- decsion and conclusion:

Since T_0 lies in acceptence Region we don't reject H_0.

So, the Intercept is significant.

7

Problem 3 (8 points)

A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data are as follows, where X is the year (coded) and Y is sales in thousands

i	1	2	3	4	5	6	7	8	9	10
X_i	0	1	2	3	4	5	6	7	8	9
Y_i	98	135	162	178	221	232	283	300	374	395

1. Fit the simple linear model **in matrix form** of the given data (1 points)
2. Fit the simple linear model **in matrix form** under the following transformations $Y' = \sqrt{Y}, Y' = \log_{10}(Y)$ and $Y' = \frac{1}{Y}$. (3 points)
3. Compare between the results in (1) and (2) using the coefficient of determination. (2 points)
4. What is the expected Y **for each model** when $X = 10$? (2 points)

~~1) > b = solve(t(x) %*% X) %*% t(x) %*% Y~~

~~$b = \begin{bmatrix} 91.5636 \\ 32.4970 \end{bmatrix}$~~

~~$Y = X\beta + \varepsilon$ ⟹ $Y = X \begin{bmatrix} 91.56 \\ 32.49 \end{bmatrix} + \varepsilon$~~

1) > Model = lm(Y~X)

$Y_i = 91.56 + 32.5 X_i + \varepsilon$

2) $Y1 = sqrt(Y)$

$Y2 = log(Y)$

$Y_3 = \frac{1}{Y}$

> Model1 = lm(Y1~X)

$Y' = 10.261 + 1.076X + \varepsilon$

> Model2 = lm(Y2~X)

$Y' = 4.726 + 0.1467X + \varepsilon$

> Model 3 = lm(Y3~X)

$Y' = 0.00838 + (-0.0007496)X + \varepsilon$

3) > Summary(model1)$r.squared

$R^2 = 0.9891$

> " " (Model2) " "

$R^2 = 0.9767$

> " " (Model3) " "

$R^2 = 0.8836$

> Sumary (Model) (original model)

$R^2 = 0.9788$

Since the (Model 1) has the highes R^2 at 0.9891 larger original model at 0.979 the best fit Model is Model 1 transformation $Y' = \sqrt{Y}$

⟶ > >

4) expected at $x=10$;

Model original;

$y = 91.56 + 32.5(10)$

$= 416.56$

Model1;

$y' = 10.261 + 1.076(10)$

$\sqrt{y'} = 21.021$

$y = 21.021^2 = \underline{441.882}$

Model 2;

$y' = 4.726 + 0.1467(10)$

$\log_{10}(y) = 6.193$

$y = 10^{6.193} = 1559552.503$

Model 3;

$y' = 0.00838 + (-0.0007446)(x)$

$\frac{1}{y} = 0.000934$

$y = \frac{1}{0.000934} = 1070.66$

16

Problem 4 (Use datafile properties.txt) (16 points)

I. For the simple regression model use Matrix form to prove the following

1. $E(\mathbf{b}) = \boldsymbol{\beta}$ (2 points)
2. Obtain the least square estimator which minimize $\boldsymbol{Q} = \boldsymbol{e}'\boldsymbol{e}$. (3 points) $(X'X)^{-1}$

II. A commercial real estate company evaluates vacancy rates, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data are taken from 81 suburban commercial properties that are the newest, best located, most attractive, and expensive for five specific geographic areas. The considered variables are the age (X_1), operating expenses and taxes (X_2), vacancy rates (X_3), and rental rates (Y).

Consider the multiple regression model to the data for two predictor variables with normal error terms

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, i = 1, \ldots, n.$$

1. State the regression model **in matrix form**. (1 point)
2. State the estimated regression function. (1 point)
3. Interpret b_1, b_2 and b_3. (2 points)
4. Test whether there is a regression relation; use $\alpha = 0.05$. State the alternatives, decision rule, and conclusion. (2points)
5. Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X_1; with X_2 given X_1; and with X_3 given X_1, X_2. (2 points)
6. Use the extra sum of squares (F^* test statistic) and $\alpha = 0.01$ to test whether X_3 can be dropped from the regression model given that X_1 and X_2 are retained. (State the alternatives, decision rule, and conclusion). (2 points)
7. A new property is to be evaluated, with $X_{h1} = 5, X_{h2} = 8.25$ and $X_{h3} = 0$, Obtain a 95% prediction interval for the rental rate for this property. (1 point)

I:- 1) $E(b) = E((X'X)^{-1}X'Y)) = (X'X)^{-1}X'E(Y) = \underbrace{(X'X)^{-1}X'X}_{I}\beta = I\beta = \beta$

2) $Q = e'e = (Y - X\beta)'(Y - X\beta)$

$= Y'Y - b'X'Y - Y'Xb + b'X'Xb$

$= Y'Y - 2b'X'Y + b'X'Xb$

$\frac{\partial Q}{\partial \beta} = -2X'Y + 2X'Xb = 0 \implies \beta = (X'X)^{-1}X'Y$

II:- 1) >Solve(t(x)%*%(x))%*%t(x)%*%(Y)

n=81

$Y_{81\times1} = X_{81\times4}\begin{bmatrix}11.541\\-0.1190\\0.4461\\2.6204\end{bmatrix} + \varepsilon_{81\times1} = Y_{n\times1}$

where:

$\underset{81\times1}{Y} = \begin{bmatrix}13.5\\ \vdots\\ 14.5\end{bmatrix}, \underset{81\times4}{X} = \begin{bmatrix}1 & 1 & 5.02 & 0.14\\ 1 & 14 & 8.19 & 0.27\\ \vdots & \vdots & \vdots & \vdots\\ 1 & 14 & 12.88 & 0.03\end{bmatrix}$

$\underset{4\times1}{\beta} = \begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{bmatrix}$ $\underset{81\times1}{\varepsilon} = \begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix} = 0$

2) $Y_i = 11.541 + (-0.1190)X_1 + (0.4461)X_2 + (2.6204)X_3 + \varepsilon$

—>>>

3) ~~[illegible]~~

b_1: Y (rental Rates) will decrese by 0.1190 for each Increase in X_1 (the age), if X_2 and X_3 remain the same.

b_2: Y (rental rates) Will increace by= 0.4461 for each increas in X_2 (expense & tax), if X_1 and X_3 remain the same.

b_3: Y (rental rates) will increase by 2.6204 for each Increase in X_3 (vacancy), if X_2 and X_1 remain the same.

---

4) 1- hypotheses:

~~[illegible]~~ | $H_0: \beta_0 = 0$
 $H_1: \beta_0 \neq 0$ | $H_0: \beta_1 = 0$
 $H_1: \beta_1 \neq 0$ | $H_0: \beta_2 = 0$
 $H_1: \beta_2 \neq 0$ | $H_0: \beta_3 = 0$
 $H_1: \beta_3 \neq 0$

t-

2- test statistic

Sine we want to test model linearity we will use ~~[illegible]~~ test

~~[illegible]~~ >summary (Model)

β_0:	β_1:	β_2:	β_3:
P.V. = 0.00...	P.V. = 0.00...	P.V. = 0.00...	P.V. 0.0353

3- Decsion P-value approach:

β_0:	β_1:	β_2:	β_3:
P.V. = 0.000... < 0.05	P.V. = 0.00... < 0.05	P.V. = 0.00 < 0.05	P.V. = 0.035 < 0.05

don't reject for all

4- conclusion:

all β are significant. the model is linear.

---

5) >anova(Model)

SSR(X_1) = 14.819 , df = 1 ,

SSR(X_2|X_1) = 72.802 , df = 1 ,

SSR(X_3|X_1, X_2) = 8.381 , df = 1 ,

→ Behind Page.

6) 1-the hypotheses;

$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$ $\qquad \alpha = 0.01$

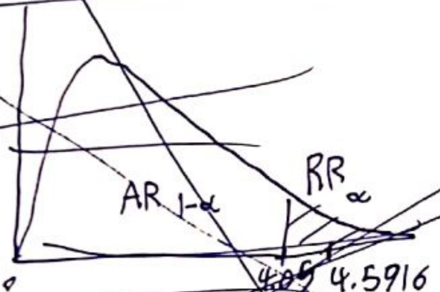
2-test statistic:

$$F^* = \frac{SSR(X_3 | X_1, X_2)}{1} \div \frac{SSE(X_1, X_2, X_3)}{n-p} = \frac{MSR(x_3)}{MSE} = \underline{4.5916}$$

$81-4$

3-Critical Region;

~~qf(0.99,3,77)= 4.046~~

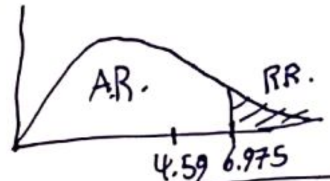


~~4-thedescion and conclusion;~~

~~since F* is in the rejection region we reject H_0~~

~~and claim β_3 is not significant. at 0.01~~

qf(0.99, 1, 77) = 6.975



4-thedecsion & conc.:

since F^* lies in A.R.

we reject H_0.

and claim β_3 is not signif

7) $X_{h1} = 5$, $X_{h2} = 8.25$, $X_{h3} = 0$

>Predict(model, newx, lvl=0.95, Aint=("Predict"))

$\hat{Y}$ fit= 14.626 , $\hat{Y}_h \in (11.899, 17.35287)$