

**College of Sciences**

**Department of**

**Statistics and Operations Research**

**كلية العلوم**

**قسم الإحصاء وبحوث العمليات**

**Midterm Exam**

|  |  |  |
| --- | --- | --- |
| **Wednesday, 03 / 07 / 1444 H** | **STAT 332** | **Academic year 1444 H** |
| **7:00 – 09:00 pm** | **Regression Analysis** | **Second Semester** |

30

|  |  |  |
| --- | --- | --- |
| Student’s Name |  | اسم الطالب |
| ID number |  | الرقم الجامعي |
| Section No. |  | رقم الشعبة |
| Classroom No. |  | رقم قاعة الاختبار |
| Teacher’s Name |  | اسم أستاذ المقرر |
| Roll Number |  | رقم التحضير |

Important Notices: Answer the following questions (3 questions, two pages) Answer a new question in a new page.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |
| --- |
| **Question One (8 Marks)** |

Suppose that is a data set to which we fit a simple regression model,

1. If , find the least squared estimate of the of , next, discuss the unbiasedness and derive its variance.
2. Ifprove that:
3. The sum of residuals is equal to zero.
4. The total sum of square error is equal to the sum the regression square error of the plus the sum of square error.
5. When , find the least squared estimate of the of , next, discuss the unbiasedness and derive its variance.

\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*

1. When Prove that
2. The sum of residuals is equal to zero

1. The total sum of square error is equal to the sum the regression square error of the plus the sum of square error 



|  |
| --- |
| **Question Two (12 Marks)** |

> X=c(3.4,1.8,4.6,2.3,3.1,5.5,0.7,3.0,2.6,4.3,2.1,1.1,6.1,4.8,3.8)

> X

 [1] 3.4 1.8 4.6 2.3 3.1 5.5 0.7 3.0 2.6 4.3 2.1 1.1 6.1 4.8 3.8

> Y=c(26.2,17.8,31.3,23.1,27.5,36.0,14.1,22.3,19.6,31.3,24.0,17.3,43.2,36.4,26.1)

> Y

 [1] 26.2 17.8 31.3 23.1 27.5 36.0 14.1 22.3 19.6 31.3 24.0 17.3 43.2 36.4 26.1

> model=lm(Y~X)

1. Estimate the regression line and interpret the coefficients.

> lm(Y~X)

Call:

lm(formula = Y ~ X)

Coefficients:

(Intercept) X

 10.278 4.919

The regression model is given by

 => The mean response Y (the amount of fire damage) is equal 10.278 thousand SR at X (the distance between the residence and the nearest fire station) is zero. (The part of Y does not depend on X)

 => The mean of response Y (the amount of fire damage) increases by 4.919 thousand SR when X (the distance between the residence and the nearest fire station) increases by one mile.

1. Calculate the estimate of the standard deviation of fire damage for all homes the same distance from the fire station is (in SR1000)

> anova(model)

Analysis of Variance Table

Response: Y

 Df Sum Sq Mean Sq F value Pr(>F)

X 1 841.77 841.77 156.89 1.248e-08 \*\*\*

Residuals 13 69.75 5.37

1. Construct 90% confidence intervals for the model coefficients and explain the results.

> confint(model,level=0.9)

 5 % 95 %

(Intercept) 7.762711 12.793146

X 4.223801 5.614861

This means that when the distance between the residence increases by one mile, the damage increases from 4.22 to 5.6 thousand SR.

This means there the amount of damage that does not depend on the residence distance is between 7.76 to 12.79 thousand SR.

1. Test the linearity by using ANOVA.

Hypothesis:

F-statistic =

> qf(0.95,1,13) [1] 4.667193

The critical value =

Since , we reject , so the model in linear

1. Find 95% confidence interval for the mean fire damage for all house 3.5 miles from the fire station is (in SR1000).

Confidence interval for

> newx=data.frame(X=3.5)

> predict (model,newx,level=0.95,interval="confidence")

 fit lwr upr

1 27.49559 26.1901 28.80107

1. Calculate the residual corresponding to the of the damage at the house of 5.5 miles (in SR1000).

This is the residual corresponding the value number 5 which X = 5.5 and Y = 36.0

> summary(model)$res

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | **6** | 7 |
| -0.80365 | -1.33272 | -1.60685 | 1.507611 | 1.972146 | **-1.33425** | 0.37854 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| -2.73592 | -3.46819 | -0.13105 | 3.391477 | 1.610808 | 2.914154 | 2.509284 |
| 15 |   |   |   |   |   |   |
| -2.87139 |   |   |   |   |   |   |

|  |
| --- |
| **Question Three (10 Marks)** |

1. Fit the model 

*> x=c(10,10,15,15,20,20,25,25,25,30,35,35)*

*> y=c(73,78,85,88,90,91,87,86,91,75,65,63)*

*>model=lm(y~x)*

*>summary(model)*

*Coefficients:*

 *Estimate Std. Error t value Pr(>|t|)*

*(Intercept) 92.0025 7.7638 11.850 3.29e-07 \*\*\**

*x -0.4982 0.3292 -1.514 0.161*

*=92.002-0.498*

1. Test the linearity using t-test.

*Test the hypothesis 𝐻0: 𝛽1=0. VS 0: 𝛽10*

*>anova(model)*

*Response: y*

 *Df Sum Sq Mean Sq F value Pr(>F)*

*x 1 204.27 204.273 2.2908 0.1611*

*Residuals 10 891.73 89.173*

*Since p-value=0.1611> 0.05*

*The decision: reject 𝐻1 and accept 𝐻0 (there is no linear relation between X and Y.*

1. Discuss the lack of fit test.

*𝐻0: The regression function is linear (no lack of fit)*

*H1: The regression function is not linear (lack of fit)*

*> install.packages("olsrr")*

*>library(olsrr)*

*> ols\_pure\_error\_anova(model)*

*Lack of Fit F Test*

*--------------*

*Response : y*

*Predictor: x*

 *Analysis of Variance Table*

*----------------------------------------------------------------------*

 *DF Sum Sq Mean Sq F Value Pr(>F)*

*----------------------------------------------------------------------*

*x 1 204.2734 204.2734 36.58628 0.0001238284*

*Residual 10 891.7266 89.17266*

 *Lack of fit 4 858.2266 214.5566 38.42806 0.0002061043*

 *Pure Error 6 33.50 5.583333*

*----------------------------------------------------------------------*

**Another package**

*>install.libraries("EnvStats")*

*>install.packages("EnvStats")*

*>library(EnvStats)*

*>anovaPE(model)*

*Df Sum Sq Mean Sq F value Pr(>F)*

*x 1 204.27 204.273 36.586 0.0009245 \*\*\**

*Lack of Fit 4 858.23 214.557 38.428 0.0002061 \*\*\**

*Pure Error 6 33.50 5.583*

*p-value = 0.0002061<0.05*

*We reject 𝐻0 (there is a lack of fit)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. A linear regression was run on a set of data. You are given only the following partial information from R:

Coefficients:

 Estimate Std. Error t value

(Intercept) 293.89 5.62 52.294

 X (1)-1.7069 0.13 -13.13

Analysis of Variance Table

Response: y

 Df Sum Sq Mean Sq F value

X 1 7621.667 (2) (4) 172.3969

Residuals 5 (3) 221.05 44.21

Complete the table as follows

 (1)

 (2)

(3)

(4)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**With My Best Wishes and Regards ..**

 **Dr. Mahmoud Bakr**