

Faculty of Engineering

Department of Civil Engineering

2017-2018 Spring Semester

**CE 0337 Water Resources Management**

**Linear Regression Analysis Homework**

Prepared By  
1600006183 Tuğçe Altınkaynak

**Dr. K. Emre Can**

Linear regression is a way to model the relationship between two variables. Two-variable data table is given below. Age and glucose level are two variables that may have a relation.

|  |  |  |
| --- | --- | --- |
| N | X (Age) | Y (Glucose level) |
|
| 1 | 22 | 65 |
| 2 | 60 | 94 |
| 3 | 35 | 71 |
| 4 | 57 | 87 |
| 5 | 43 | 90 |
| 6 | 25 | 68 |
| 7 | 49 | 75 |

Assume a linear function is to be fitted to this data. Therefore, Y = A\*X + B

YEST (1) = A\*X (1) + B  
YEST (2) = A\*X (2) + B  
YEST (3) = A\*X (3) + B  
YEST (4) = A\*X (4) + B  
YEST (5) = A\*X (5) + B  
YEST (6) = A\*X (6) + B  
YEST (7) = A\*X (7) + B

YEST (1) = 22A + B  
YEST (2) = 60A + B  
YEST (3) = 35A + B  
YEST (4) = 57A + B  
YEST (5) = 43A + B  
YEST (6) = 25A + B  
YEST (7) = 49A + B

DIF (1) = Y (1) – YEST (1)  
DIF (2) = Y (2) – YEST (2)  
DIF (3) = Y (3) – YEST (3)  
DIF (4) = Y (4) – YEST (4)  
DIF (5) = Y (5) – YEST (5)  
DIF (6) = Y (6) – YEST (6)  
DIF (7) = Y (7) – YEST (7)

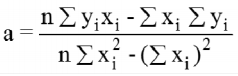
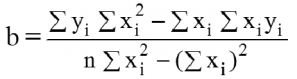
DIF (1) = 65 – YEST (1)  
DIF (2) = 94 – YEST (2)  
DIF (3) = 71 – YEST (3)  
DIF (4) = 87 – YEST (4)  
DIF (5) = 90 – YEST (5)  
DIF (6) = 68 – YEST (6)  
DIF (7) = 75 – YEST (7)

DIF (1) = 65 – 22A – B  
DIF (2) = 94 – 60A – B  
DIF (3) = 71 – 35A – B  
DIF (4) = 87 – 57A – B  
DIF (5) = 90 – 43A – B  
DIF (6) = 68 – 25A – B  
DIF (7) = 75 – 49A – B

Min Z = [DIF (1)]2 + [DIF (2)]2 + [DIF (3)]2 + [DIF (4)]2 + [DIF (5)]2 + [DIF (6)]2 + [DIF (7)]2

Min Z = [65 – 22A – B]2 + [94 – 60A – B]2 + [71 – 35A – B]2 + [87 – 57A – B]2 + [90 – 43A – B]2 + [68 – 25A – B]2 + [75 – 49A – B]2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| N | X | Y | XY | X2 | Y2 |
| 1 | 22 | 65 | 1430 | 484 | 4225 |
| 2 | 60 | 94 | 5640 | 3600 | 8836 |
| 3 | 35 | 71 | 2485 | 1225 | 5041 |
| 4 | 57 | 87 | 4959 | 3249 | 7569 |
| 5 | 43 | 90 | 3870 | 1849 | 8100 |
| 6 | 25 | 68 | 1700 | 625 | 4624 |
| 7 | 49 | 75 | 3675 | 2401 | 5625 |
| Σ | 291 | 550 | 23759 | 13433 | 44020 |

From the above table, ΣX = 291, ΣY = 550, ΣXY = 23759, ΣX2 = 13433, ΣY2 = 44020, and n = 7.  
  
 

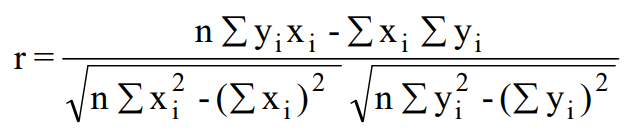
By substituting values, A and B are found.

**A** = (7\*23759 – 291\*550) / (7\*13433 – 2912)  
 = (166313 – 160050) / (94031 – 84681)  
 = 6263 / 9350  
 = 0.6698

**B** = (550\*13433 – 291\*23759) / (7\*13433 – 2912)  
 = (7388150 – 6913869) / (94031 – 84681)  
 = 474281 / 9350  
 = 50.725

**Y = 0.6698\*X + 50.725**

Correlation coeficient r,



**r** = (7\*23759 – 291\*550) / ((7\*13433 – 2912)0.5)\*((7\*44020 – 5502)0.5)  
 = (166313 – 160050) / (93500.5)\*(56400.5)  
 = 6263 / 96.70\*75.10  
 = 0.8625

***Comment on r:*** The range of values for the correlation coefficient is –1.0 to 1.0. R is equal to 1.0 means there is a perfect positive relationship between the two variables. R equals –1.0 means there is a perfect negative relationship between the two variables. If the correlation is 0, it means that there is no relationship between the two variables.

Correlation coefficient is calculated as 0.8625 for given data. R is a positive value and it is near to 1.0, so it means that there is high positive relationship between the two variables which are age and glucose level.

***Excel***

Linear regression analysis can be performed in some ways on Excel. After plotting graph, trendline can be added and equation can be displayed as shown below. Moreover, data analysis option can be used to get regression statistics.

|  |  |
| --- | --- |
| *Regression Statistics* | |
| **Multiple R** | **0,862456** |
| R Square | 0,743831 |
| Adjusted R Square | 0,692597 |
| Standard Error | 6,42494 |
| Observations | 7 |