**Tutorial 10**

**Data Analytics**

In this tutorial, you are going to work on some statistical methods using SQL.

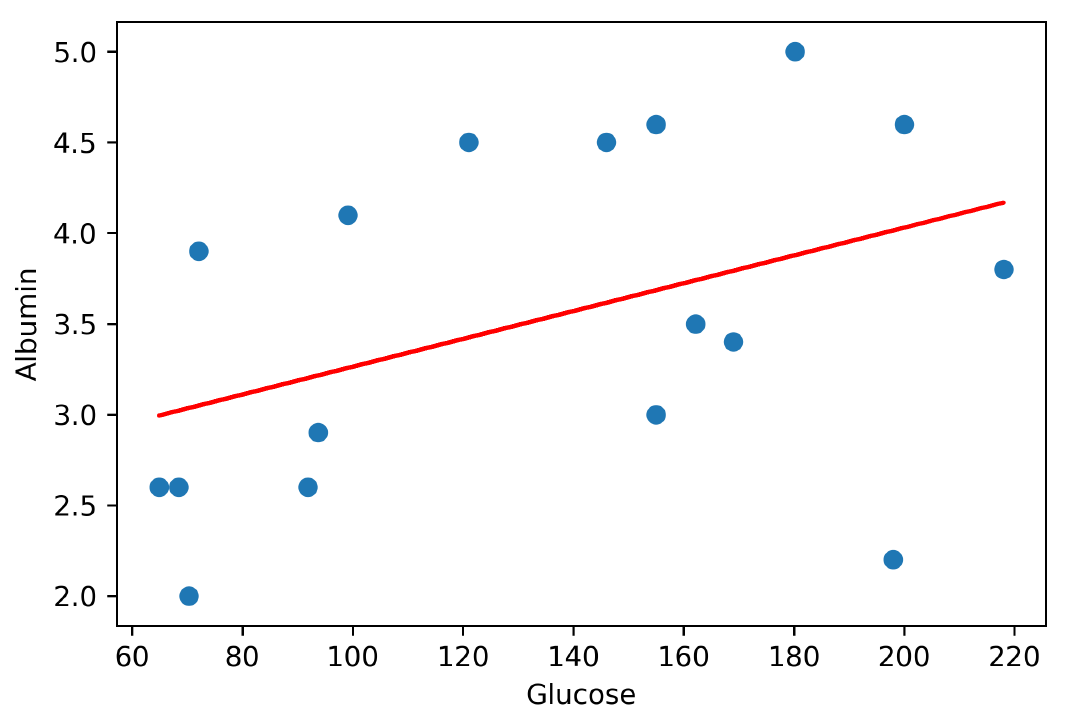


Figure 1 - Linear Regression: Glucose-Albumin

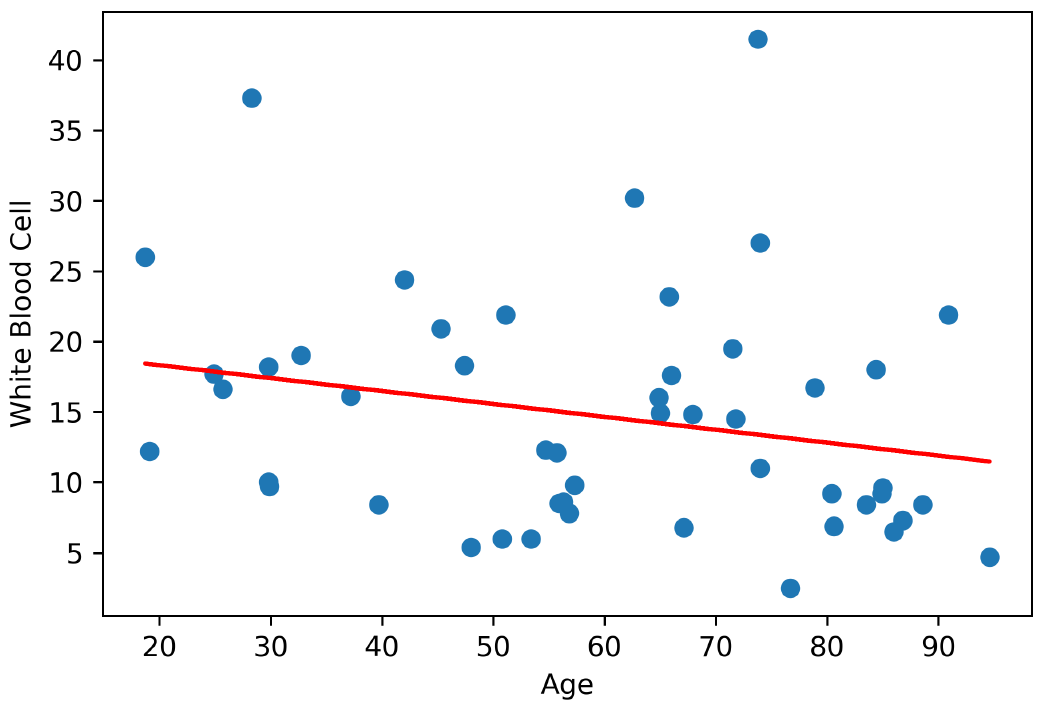
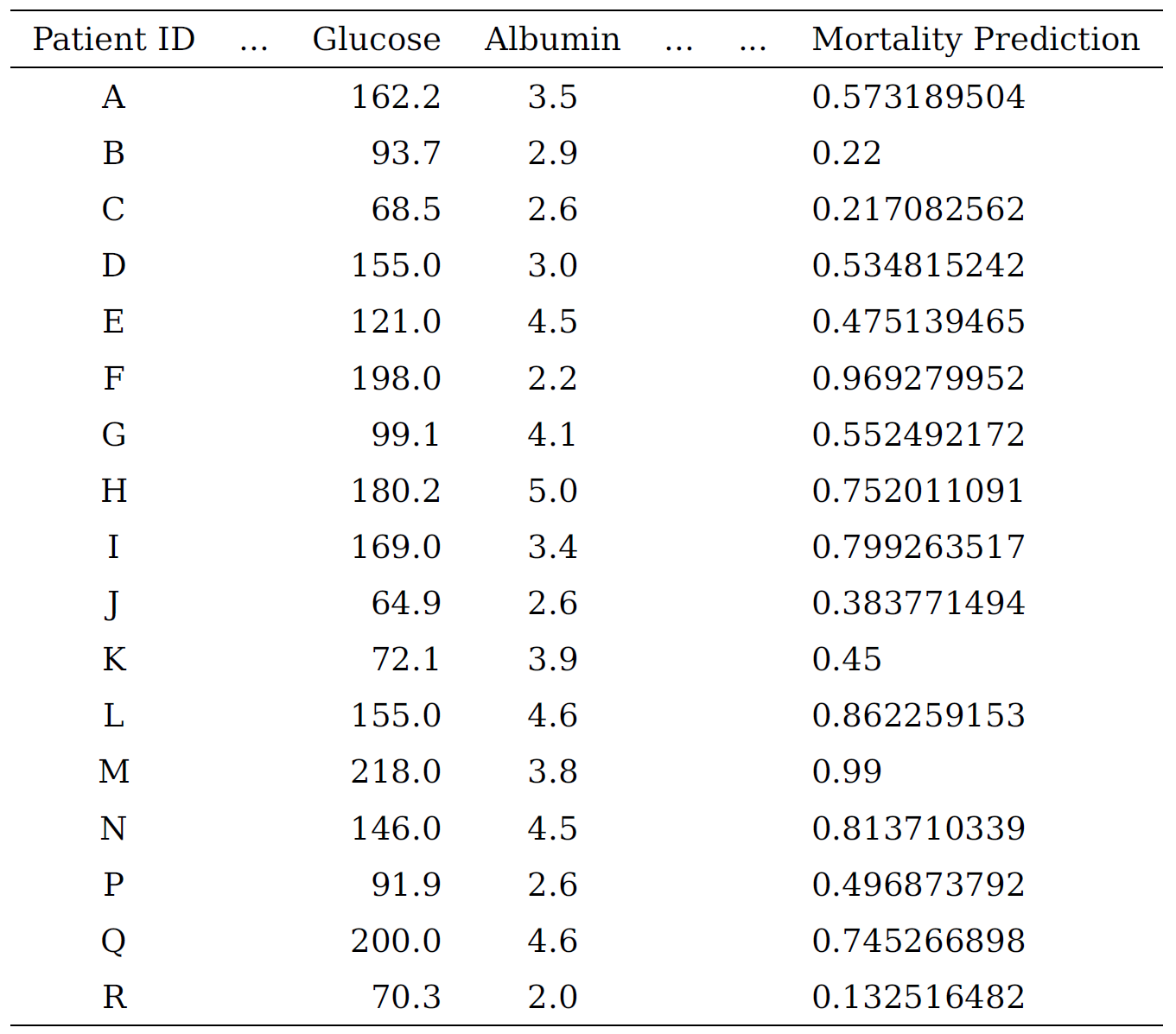


Figure 2 - Linear Regression: Age-WBC (White Blood Cell)

Figure 1 and Figure 2 show the simple linear regression models for Glucose-Albumin, and Age-WBC (White Blood Cell). The Glucose-Albumin data consists of 17 data points, as shown in the table below.

Table 1 - Emergency Patient Extended Fact Table



Whereas the Age-WBC data consists of 49 data points, in the format of (Age, WBC), which are as follows:

{(74, 27), (56.3, 8.6), (55.9, 8.5), (80.4, 9.2), (56.8, 7.8), (29.8, 18.2), (19.1, 12.2), (32.7, 19), (28.3, 37.3), (73.8, 41.5), (62.7, 30.2), (25.7, 16.6), (66, 17.6), (64.9, 16), (37.2, 16.1), (94.6, 4.7), (39.7, 8.4), (29.9, 9.7), (48, 5.4), (42, 24.4), (71.5, 19.5), (47.4, 18.3), (24.9, 17.7), (50.8, 6), (45.3, 20.9), (76.7, 2.5), (55.7, 12.1), (86.8, 7.3), (83.5, 8.4), (84.4, 18), (18.7, 26), (85, 9.6), (57.3, 9.8), (80.6, 6.9), (65.8, 23.2), (67.9, 14.8), (84.9, 9.2), (67.1, 6.8), (51.1., 21.9), (88.6, 8.4), (29.8, 10), (65, 14.9), (74, 11), (78.9, 16.7), (53.4, 6), (86, 6.5), (71.8, 14.5), (54.7, 12.3), (90.9, 21.9)}

Using the SQL commands, calculate the intercept and slope for both datasets. To check your solutions, the intercept and slope for the Glucose-Albumin are 2.4963 and 0.0077, respectively. The intercept and slope for the Age-WBC are 20.1455 and -0.0918, respectively.

# Tasks

1. Create a table for Glucose-Albumin that contains Glucose and Albumin attributes. After you have created the table, insert some values in the table using the 17 data points of Glucose and Albumin from Table 1.

-- Create GlucoseAlbumin table

Create TABLE GlucoseAlbumin

(

Glucose NUMERIC(18,6) NOT NULL,

Albumin NUMERIC(18,6) NOT NULL

);

-- Insert the data

insert into GlucoseAlbumin (Glucose, Albumin) values (162.2, 3.5);

insert into GlucoseAlbumin (Glucose, Albumin) values (93.7, 2.9);

insert into GlucoseAlbumin (Glucose, Albumin) values (68.5, 2.6);

insert into GlucoseAlbumin (Glucose, Albumin) values (155.0, 3.0);

insert into GlucoseAlbumin (Glucose, Albumin) values (121.0, 4.5);

insert into GlucoseAlbumin (Glucose, Albumin) values (198.0, 2.2);

insert into GlucoseAlbumin (Glucose, Albumin) values (99.1, 4.1);

insert into GlucoseAlbumin (Glucose, Albumin) values (180.2, 5.0);

insert into GlucoseAlbumin (Glucose, Albumin) values (169.0, 3.4);

insert into GlucoseAlbumin (Glucose, Albumin) values (64.9, 2.6);

insert into GlucoseAlbumin (Glucose, Albumin) values (72.1, 3.9);

insert into GlucoseAlbumin (Glucose, Albumin) values (155.0, 4.6);

insert into GlucoseAlbumin (Glucose, Albumin) values (218.0, 3.8);

insert into GlucoseAlbumin (Glucose, Albumin) values (146.0, 4.5);

insert into GlucoseAlbumin (Glucose, Albumin) values (91.9, 2.6);

insert into GlucoseAlbumin (Glucose, Albumin) values (200.0, 4.6);

insert into GlucoseAlbumin (Glucose, Albumin) values (70.3, 2.0);

1. Calculate the slope for Glucose-Albumin.

-- Calculate the slope

select

sum((Glucose - x\_bar) \* (Albumin - y\_bar)) /

sum((Glucose - x\_bar) \* (Glucose - x\_bar)) as slope

from (

select Glucose, avg(Glucose) over () as x\_bar, Albumin, avg(Albumin) over () as y\_bar

from GlucoseAlbumin group by Glucose, Albumin) g;

1. Calculate the slope and intercept for Glucose-Albumin.

-- Calculate the slope and intercept

select slope,

y\_bar\_max - x\_bar\_max \* slope as intercept

from (

select sum((Glucose - x\_bar) \* (Albumin - y\_bar)) / sum((Glucose - x\_bar) \* (Glucose - x\_bar)) as slope,

max(x\_bar) as x\_bar\_max,

max(y\_bar) as y\_bar\_max

from (

select Glucose, avg(Glucose) over () as x\_bar,

Albumin, avg(Albumin) over () as y\_bar

from GlucoseAlbumin group by Glucose, Albumin) s

);

1. Create Age WBC table and insert the data into the newly created table.

-- Create Age WBC table

Create TABLE Age\_WBC

(

Age NUMERIC(3,0) NOT NULL,

WBC NUMERIC(18,6) NOT NULL

);

-- Insert the data

insert into Age\_WBC (Age, WBC) values (74,27);

insert into Age\_WBC (Age, WBC) values (56.3, 8.6);

insert into Age\_WBC (Age, WBC) values (55.9, 8.5);

insert into Age\_WBC (Age, WBC) values (80.4, 9.2);

insert into Age\_WBC (Age, WBC) values (56.8, 7.8);

insert into Age\_WBC (Age, WBC) values (29.8, 18.2);

insert into Age\_WBC (Age, WBC) values (19.1, 12.2);

insert into Age\_WBC (Age, WBC) values (32.7, 19);

insert into Age\_WBC (Age, WBC) values (28.3, 37.3);

insert into Age\_WBC (Age, WBC) values (73.8, 41.5);

insert into Age\_WBC (Age, WBC) values (62.7, 30.2);

insert into Age\_WBC (Age, WBC) values (25.7, 16.6);

insert into Age\_WBC (Age, WBC) values (66, 17.6);

insert into Age\_WBC (Age, WBC) values (64.9, 16);

insert into Age\_WBC (Age, WBC) values (37.2, 16.1);

insert into Age\_WBC (Age, WBC) values (94.6, 4.7);

insert into Age\_WBC (Age, WBC) values (39.7, 8.4);

insert into Age\_WBC (Age, WBC) values (29.9, 9.7);

insert into Age\_WBC (Age, WBC) values (48, 5.4);

insert into Age\_WBC (Age, WBC) values (42, 24.4);

insert into Age\_WBC (Age, WBC) values (71.5, 19.5);

insert into Age\_WBC (Age, WBC) values (47.4, 18.3);

insert into Age\_WBC (Age, WBC) values (24.9, 17.7);

insert into Age\_WBC (Age, WBC) values (50.8, 6);

insert into Age\_WBC (Age, WBC) values (45.3, 20.9);

insert into Age\_WBC (Age, WBC) values (76.7, 2.5);

insert into Age\_WBC (Age, WBC) values (55.7, 12.1);

insert into Age\_WBC (Age, WBC) values (86.8, 7.3);

insert into Age\_WBC (Age, WBC) values (83.5, 8.4);

insert into Age\_WBC (Age, WBC) values (84.4, 18);

insert into Age\_WBC (Age, WBC) values (18.7, 26);

insert into Age\_WBC (Age, WBC) values (85, 9.6);

insert into Age\_WBC (Age, WBC) values (57.3, 9.8);

insert into Age\_WBC (Age, WBC) values (80.6, 6.9);

insert into Age\_WBC (Age, WBC) values (65.8, 23.2);

insert into Age\_WBC (Age, WBC) values (67.9, 14.8);

insert into Age\_WBC (Age, WBC) values (84.9, 9.2);

insert into Age\_WBC (Age, WBC) values (67.1, 6.8);

insert into Age\_WBC (Age, WBC) values (51.1, 21.9);

insert into Age\_WBC (Age, WBC) values (88.6, 8.4);

insert into Age\_WBC (Age, WBC) values (29.8, 10);

insert into Age\_WBC (Age, WBC) values (65, 14.9);

insert into Age\_WBC (Age, WBC) values (74, 11);

insert into Age\_WBC (Age, WBC) values (78.9, 16.7);

insert into Age\_WBC (Age, WBC) values (53.4, 6);

insert into Age\_WBC (Age, WBC) values (86, 6.5);

insert into Age\_WBC (Age, WBC) values (71.8, 14.5);

insert into Age\_WBC (Age, WBC) values (54.7, 12.3);

insert into Age\_WBC (Age, WBC) values (90.9, 21.9);

1. Calculate the slope and intercept for Age WBC.

-- Calculate the slope and intercept

select slope,

y\_bar\_max - x\_bar\_max \* slope as intercept

from (

select sum((Age - x\_bar) \* (WBC - y\_bar)) / sum((Age - x\_bar) \* (Age - x\_bar)) as slope,

max(x\_bar) as x\_bar\_max,

max(y\_bar) as y\_bar\_max

from (

select Age, avg(Age) over () as x\_bar,

WBC, avg(WBC) over () as y\_bar

from Age\_WBC group by Age, WBC) s);

1. Observe your calculations of the intercept and slope for both datasets.

**THE END**