

STATISTICS 3: QUIZ SOLUTIONS

Question 1

A tech researcher tracks the daily active screen time of a smartphone in hours and the percentage of battery used over a 16-hour period across 10 test days.

Daily Active Screen Time (h)	Battery Used (%)
2,5	38
3,0	44
4,2	55
5,0	62
5,5	68
6,1	75
7,0	82
7,8	90
8,2	93
9,1	99

Using Daily Active Screen Time (h) as variable x and Battery Used (%) as variable y , the researcher obtained $\hat{y} = 15,28 + 9,49x$ for the least squares linear regression line.

1. Explain the meaning of the regression coefficient 15,28.
2. Interpret the regression coefficient 9,49.
3. Calculate the predicted value of the percentage of Battery Used if Daily Active Screen Time is 6,5 hours.

Solution

1. The regression coefficient 15,28 is the y -intercept, that is, when the independent variable $x = 0$, the predicted variable $\hat{y} = 15,28$. You could interpret this as meaning that if the Daily Active Screen Time is zero, then the expected percentage of Battery Used is expected to be 15,28%. However, this may not be plausible or meaningful. The predicted value of y is valid only for values of x within the range of the data used to fit the model, that is, for x between 2,5 and 9,1.
2. The regression coefficient 9,49 is interpreted as a measure of the mean value of y for a given value of x . So, on average, for every 1h increase in Daily Active Screen Time (x), the percentage of Battery Used (y) increases by 9,49%. The positive sign indicates a positive linear relationship between x and y . As x increases, y increases at a constant rate of 9,49% per hour of x .
3. If $x = 6,5$, then $\hat{y} = 15,28 + 9,49(6,5) = 76,97$.
If Daily Active Screen Time is 6,5h, the predicted value of percentage of Battery Used is 77%.

Question 2

An environmental engineer measures the ambient outdoor temperature and the total energy output generated by a rooftop solar installation on 10 sunny days.

Temperature (°C)	Energy Output (kWh)
16	32,5
18	31,8
21	30,2
24	28,9
26	27,5
29	26,1
31	25,0
33	23,8
36	22,1
38	21,0

Using Temperature (°C) as variable x and Energy Output (kWh) as variable y , the engineer obtained $\hat{y} = 41,31 - 0,53x$ for the least squares linear regression line.

1. Explain the meaning of the regression coefficient 41,31.
2. Interpret the regression coefficient -0,53.
3. Calculate the predicted value of the Energy Output if the Temperature is 30°C.

Solution

1. The regression coefficient 41,31 is the y -intercept, that is, when the independent variable $x = 0$, the predicted variable $\hat{y} = 41,31$. You could interpret this as meaning that if the Temperature is 0°C, the Energy Output is expected to be 41,31 kWh. However, this may not be plausible or meaningful. The predicted value of y is valid only for values of x within the range of the data used to fit the model, that is, for x between 16 and 38.
2. The regression coefficient -0,53 is interpreted as a measure of the mean value of y for a given value of x . So, on average, for every 1°C increase in Temperature (x), the Energy Output (y) decreases by 0,53kWh. The negative sign indicates a negative linear relationship between x and y . As x increases, y decreases at a constant rate of 0,53kWh per degree.
3. If $x = 30$, then $\hat{y} = 41,31 - 0,53(30) = 25,41$.
If the Temperature is 30°C, the predicted value of Energy Output is 25,4kWh.

Question 3

A city planner studies the relationship between the distance of an Electric Vehicle (EV) fast-charging station from a freeway exit and the average number of charging sessions logged per day.

Distance from Freeway Exit (km)	Daily Charging Sessions
0,5	88
1,2	79
1,8	72
2,5	65
3,1	58
4,0	47
4,8	39
5,5	32
6,2	24
7,0	15

Using Distance from Freeway Exit (km) as variable x and the number of Daily Charging Sessions as variable y , the city planner obtained $\hat{y} = 92,55 - 11,11x$ for the least squares linear regression line.

1. Explain the meaning of the regression coefficient 92,55.
2. Interpret the regression coefficient -11,11.
3. Calculate the predicted number of Daily Charging Sessions if the Distance from Freeway Exit is 5km.

Solution

1. The regression coefficient 92,55 is the y -intercept, that is, when the independent variable $x = 0$, the predicted variable $\hat{y} = 92,55$. You could interpret this as meaning that if the Distance from Freeway Exit is 0km, the number of Daily Charging Sessions is expected to be 93. However, this may not be plausible or meaningful. The predicted value of y is valid only for values of x within the range of the data used to fit the model, that is, for x between 0,5 and 7,0.
2. The regression coefficient -11,11 is interpreted as a measure of the mean value of y for a given value of x . So, on average, for every 1km increase in Distance from Freeway Exit (x), the number of Daily Charging Sessions (y) decreases by 11. The negative sign indicates a negative linear relationship between x and y . As x increases, y decreases at a constant rate of 11,11 sessions per km.
3. If $x = 30$, then. $\hat{y} = 41,31 - 0,53(30) = 25,41$.
If the Temperature is 30°C, the predicted value of Energy Output is 25kWh.

Question 4

The research manager of an e-commerce business tracks its daily spend on targeted social media advertisements in R100s against the number of completed sales generated.

Daily Spend (R100s)	Completed Sales
12	45
15	58
20	72
25	90
28	102
32	115
38	138
42	150
45	162
50	180

Using Daily Spend (R100s) as variable x and number of Completed Sales as variable y , the research manager obtained $\hat{y} = 2,81 + 3,53x$ for the least squares linear regression line.

1. Explain the meaning of the regression coefficient 2,81.
2. Interpret the regression coefficient 3,53.
3. Calculate the predicted number of Completed Sales if the Daily Spend is R3 000.

Solution

1. The regression coefficient 2,81 is the y -intercept, that is, when the independent variable $x = 0$, the predicted variable $\hat{y} = 2,81$. You could interpret this as meaning that if the Daily Spend is zero, the number of Completed Sales is expected to be 2,81. However, this may not be plausible or meaningful. The predicted value of y is valid only for values of x within the range of the data used to fit the model, that is, for x between 12 and 50, i.e., R1 200 to R5 000.
2. The regression coefficient 3,53 is interpreted as a measure of the mean value of y for a given value of x . So, on average, for every R100 increase in Daily Spend (x), the number of Completed Sales (y) increases by 3,53. The positive sign indicates a positive linear relationship between x and y . As x increases, y increases at a constant rate of 3,53 completed sales per R100.
3. If $x = 30$, then. $\hat{y} = 2,81 + 3,53(30) = 108,7$.
If Daily Spend is R3 000, the predicted number of Completed Sales is 109.

Question 5

A network administrator measures the number of connected IoT devices (smart bulbs, security cameras, sensors) in a residential home and the total monthly background data used.

Number of Connected Devices	Monthly Data Used (GB)
8	12,4
12	18,1
15	22,5
18	27,0
22	33,2
25	37,8
30	45,1
34	51,0
38	57,2
42	63,5

Using Number of Connected Devices as variable x and Monthly Data Used (GB) as variable y , the network administrator obtained $\hat{y} = 0,10 + 1,50x$ for the least squares linear regression line.

1. Explain the meaning of the regression coefficient 0,10.
2. Interpret the regression coefficient 1,50.
3. Calculate the predicted amount of Monthly Data Used if the Number of Connected Devices is 20.

Solution

1. The regression coefficient 0,10 is the y -intercept, that is, when the independent variable $x = 0$, the predicted variable $\hat{y} = 0,10$. You could interpret this as meaning that if the Number of Connected Devices is zero, the Monthly Data Used is expected to be 0,10kWh. However, this may not be plausible or meaningful. The predicted value of y is valid only for values of x within the range of the data used to fit the model, that is, for x between 8 and 42.
2. The regression coefficient 1,50 is interpreted as a measure of the mean value of y for a given value of x . So, on average, for every unit increase in the Number of Connected Devices (x), the amount of Monthly Data Used (y) increases by 1,50 GB. The positive sign indicates a positive linear relationship between x and y . As x increases, y increases at a constant rate of 1,50GB per connected device.
3. If $x = 20$, then. $\hat{y} = 0,10 + 1,50(20) = 30,1$.
If the Number of Connected Devices is 20, the predicted amount of Monthly Data Used is 30,1GB.