Azamat Salamatov

Marvin Johnson

MATH 1312 - 1

Homework 4

April 11, 2023,

Homework 4

57.

a - The dependent variable is fatality and the independent variable is age of the driver.

b - The dependent variable -> grocery bills, and the independent variable -> number of family members.

c. - The dependent variable is insurance premium and the independent variable is age of the applicant.

d. - The dependent variable is the utility bill and the independent variable is power consumption.

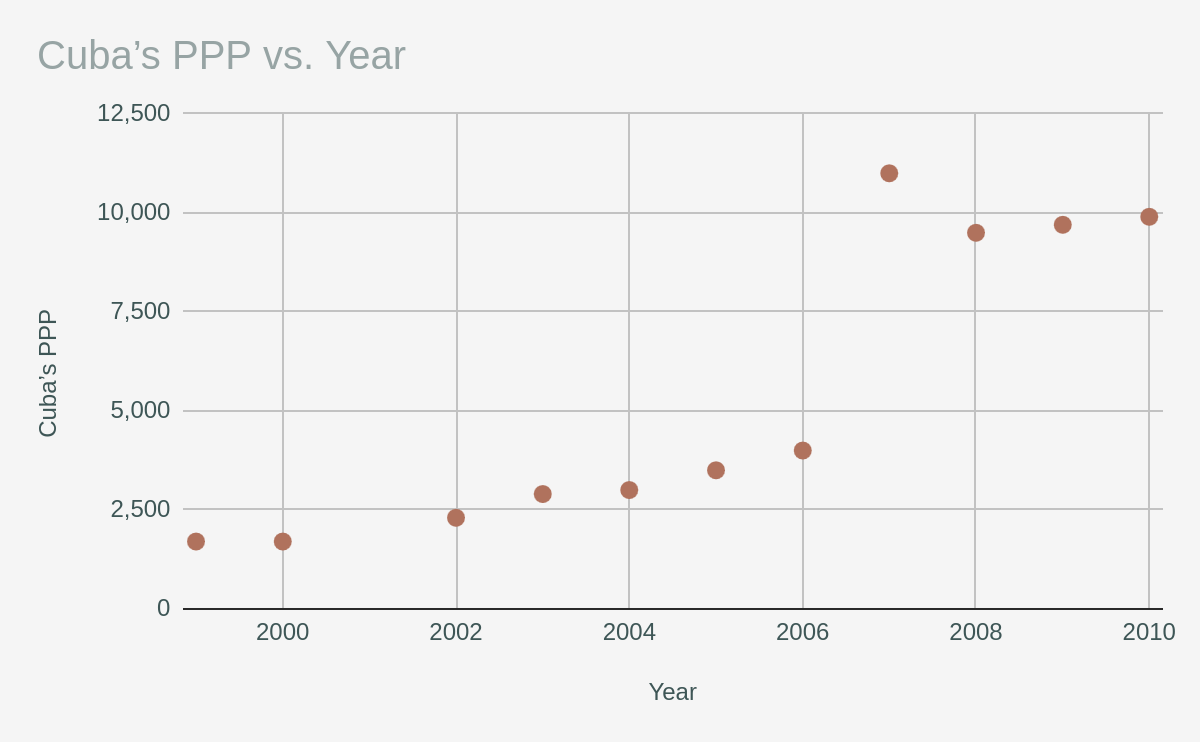
e. - The dependent variable is crime rate and the independent variable is higher education.

59.

The Gross Domestic Product Purchasing Power Parity is an indication of a country’s currency value compared to another country. Table 12.16 shows the GDP PPP of Cuba as compared to US dollars. Construct a scatter plot of the data.

65) Explain what it means when a correlation has an r2 of 0.72.

Solution: Here, it is provided that r2 = 0.72, it means that 72% of the variation in the dependent variable can be explained by the independent variable using the regression line.

67) a)

|  |  |  |
| --- | --- | --- |
| Age | Number of driver deaths per 100,000 | midpoint |
| 16 - 19 | 38 | 17.5 |
| 20 - 24 | 36 | 22 |
| 25 - 34 | 24 | 29.5 |
| 35 - 54 | 20 | 44.5 |
| 55 - 74 | 18 | 64.5 |
| 75+ | 28 | 80 |

b) and c)

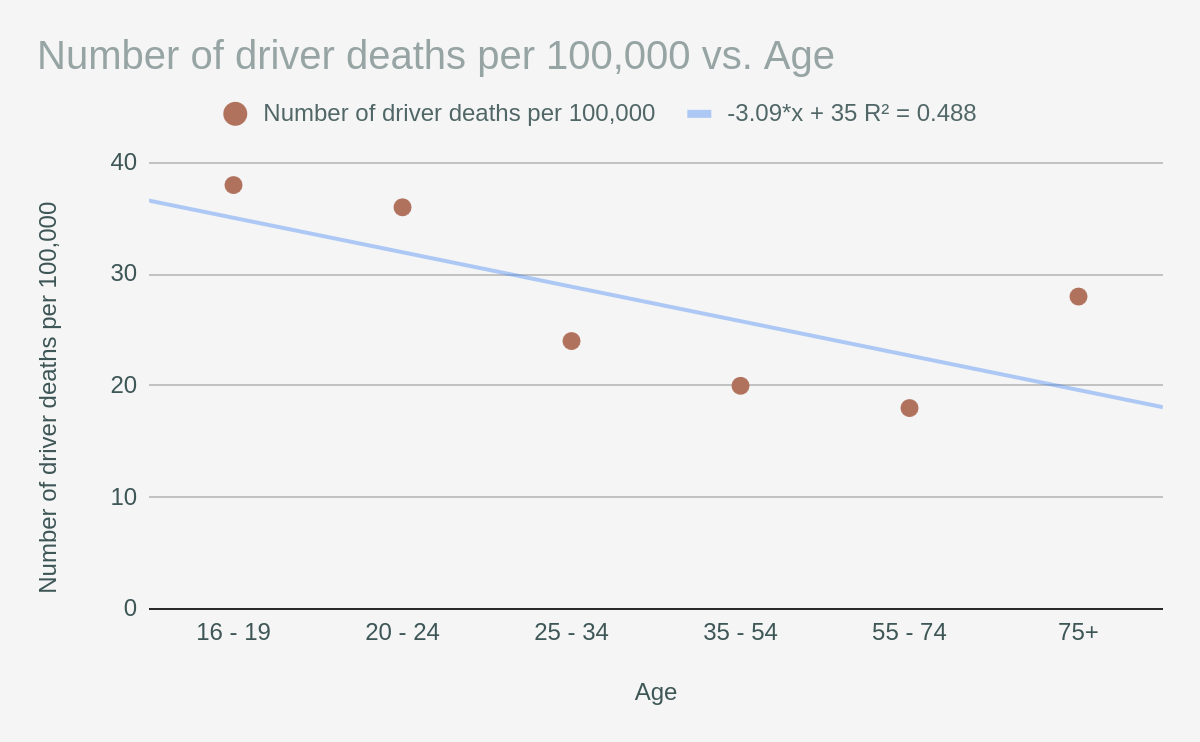
Equation: ŷ = -3.09\*x + 35, R2 = 0.4882

D) The R-squared in the chart is 0.4882

Calculating R:

√(o.4882) = 0.699

r = 0.699



Correlation Coefficient: 0.699

e) predict the number of deaths in the age of 40 and 60

x = 40

y = -3.09\*x + 35

y = -3.09\*40 + 35

y = -88.6

=================

x = 60

y = -3.09\*x + 35

y = -3.09\*60 + 35

y = -150.4

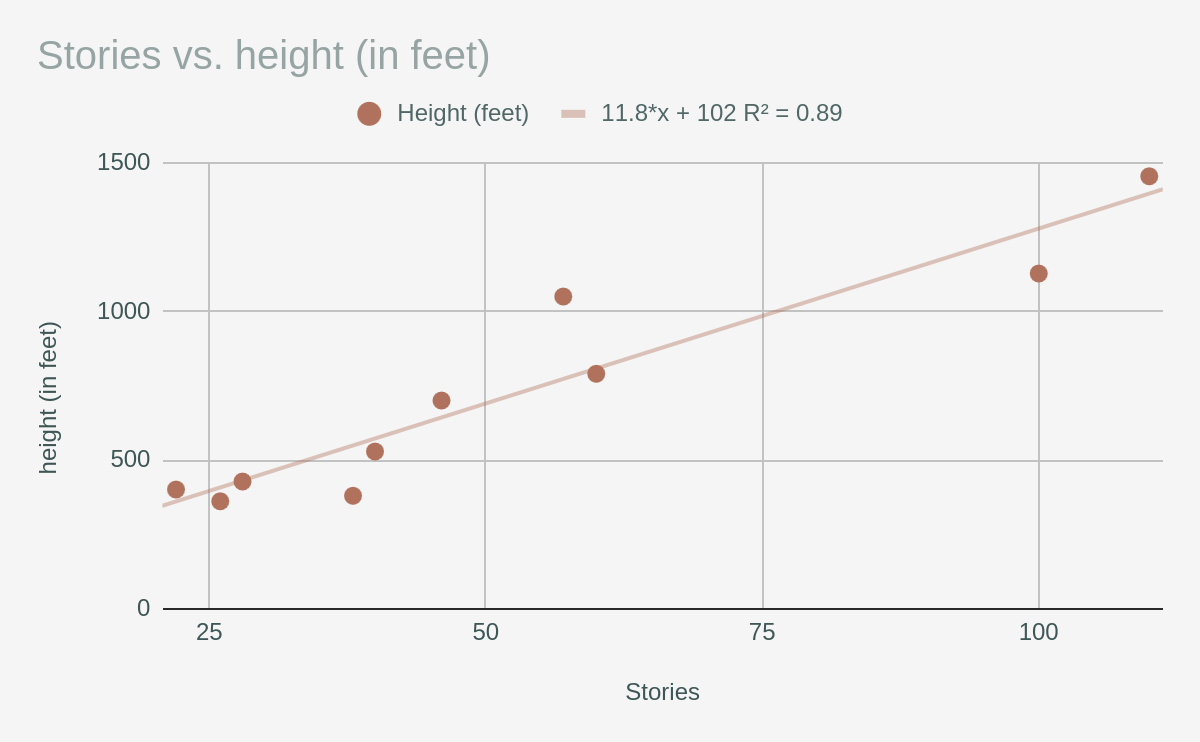
f) Relationship:

There is a negative linear relationship between Age and driver fatality rate

g) slope:

-3.09 is the slope of the linear regression from the Graph

72) a)

b) From the above scatterplot, it can conclude that there is a positive linear relationship between heights and stories.

c) From the output, the least squares line is ŷ = 102 + 11.8\*x

d) The R-squared in the chart is 0.89

Calculating R:

√(0.89) = 0.943

r = 0.943

Correlation Coefficient: 0.943, Yes it’s significant

e) 32 stories:

y = 102 + 11.8\*x

y = 102 + 11.8\*32

y = 479.6

—----------------------------------------------

94 stories:

y = 102 + 11.8\*x

y = 102 + 11.8\*94

y = 1211.2

f) Using given data, the Correlation Coefficient is significant. It means that the correlation coefficientis different from zero. Thus, there is sufficient evidence to conclude that there is significant linear relationship between the heights and stories.

g) Outlier using scatterplot: height and stories – (100, 1127) and (110, 1454)

h) Consider stories as 6.

y = 102 + 11.8\*x

y = 102 + 11.8\*6

y = 172.8

No, the least square line is not valid for six stories because this prediction can be considered as extrapolation.

i) Consider a single story. Using the estimated least square line in equation (2) in part c, the estimated height (in feet) for a single story is calculated below:

y = 102 + 11.8\*1 = 113.8

j) Using equation (2) in part c, the slope is 11.8.

69) As the number of degrees of freedom increases, the graph of the chi-square distribution looks more and more symmetrical.

Answer:

The chi-square distribution contains the sum of squared standard normal deviates where standard normal deviates are a random sample of standard normal distribution. If the number of degrees of freedom increases then the curve of chi-square distribution will become less skewed and more symmetric. The mean of the distribution will be equal to the degrees of freedom. Therefore, the statement is true.

86)

H0: Best ski area is independent of the level of the skier.

Ha: Best ski area is dependent of the level of the skier.

We assume level of significance = α = 0.05

Test statistic formula is given as below:

Chi square = ∑[(O – E)^2/E]

Where, O is observed frequencies and E is expected frequencies.

E = row total \* column total / Grand total

Number of rows = r = 3

Number of columns = c = 3

Degrees of freedom = df = (r – 1)\*(c – 1) = 2\*2 = 4

α = 0.05

Critical value = 9.487729037

Calculation tables for test statistic are given as below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Observed Frequencies |  |  |
| U.S. Ski Area | Beginner | Intermediate | Advance | Total |
| Tahoe | 20 | 30 | 40 | 90 |
| Utah | 10 | 30 | 60 | 100 |
| Colorado | 10 | 40 | 50 | 100 |
| Total | 40 | 100 | 150 | 290 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Expected Frequencies |  |  |
| U.S. Ski Area | Beginner | Intermediate | Advance | Total |
| Tahoe | 12.4137931 | 31.03448276 | 46.55172414 | 90 |
| Utah | 13.79310345 | 34.48275862 | 51.72413793 | 100 |
| Colorado | 13.79310345 | 34.48275862 | 51.72413793 | 100 |
| Total | 40 | 100 | 150 | 290 |

Calcuation through formula:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | (O - E)^2/E |  |
| U.S. Ski Area | Beginner | Intermediate | Advance |
| Tahoe | 4.636015326 | 0.03448275862 | 0.9220945083 |
| Utah | 1.043103448 | 0.5827586207 | 1.324137931 |
| Colorado | 1.043103448 | 0.8827586207 | 0.05747126437 |

Test Statistic = Chi square = ∑[(O – E)^2/E] = 10.52592593

P-value < α = 0.05

So, we reject the null hypothesis

There is not sufficient evidence to conclude that Best ski area is independent of the level of the skier.