BRAC UNIVERSITY
Inspiring Excellence

STA201 Assignment 2 (Summer 2022)

Question 1

Find an estimate of the Variance, Standard Deviation & Coefficient of Variation. of the following data for the marks obtained in a test by 92 students.

Marks (X)	$0 \leq X < 10$	$10 \leq X < 20$	$20 \leq X < 30$	$30 \leq X < 40$	$40 \leq X < 50$
Frequency (f)	8	18	24	25	17

Answer:

Marks(x)	**f**	**x**	x^2	**fx**	fx^2
$0 \leq X < 10$	8	5	25	40	200
$10 \leq X < 20$	18	15	225	270	4050
$20 \leq X < 30$	24	25	625	600	15000
$30 \leq X < 40$	25	35	1225	875	30625
$40 \leq X < 50$	17	45	2025	765	34425
Total	92		4125	2550	84300

Variance, $S^2 = \frac{1}{n-1}[\Sigma fx^2 - \frac{(\Sigma fx)^2}{n}]$

$= \frac{1}{92-1}[84300 - \frac{(2550)^2}{92}$

$= 149.677$

SD= $\sqrt{Variance}$

$= 12.234$

AM= 2550/92

$= 27.717$

CV= $\frac{12.234}{27.717}$ * 100%

= 44.138%

Question 2

For a distribution Karl Pearson's coefficient of skewness is 0.64, Variance is 40 and mean is $\sqrt{(X + Y + Z)}$ Find mode and median. [Here X,Y,Z = First 3 digits of you Student ID]

Answer:
Suppose, the sum of the first 3 digits of your student ID is 6.

$Mean, \sqrt{6} = 2.4494$

SD= $\sqrt{40} = 6.324$

Skp=0.64

Now,

$$\frac{Mean-Mode}{SD} = 0.64$$

Mode = -1.59796

And,

$$\frac{3(Mean-Median)}{SD} = 0.64$$

Median = 1.1003

Question 3

The owner of a used car dealership is interested in researching how an automobile's age and selling price are related. Below is a sample of 12 used vehicles that the dealership sold during the course of the previous year.

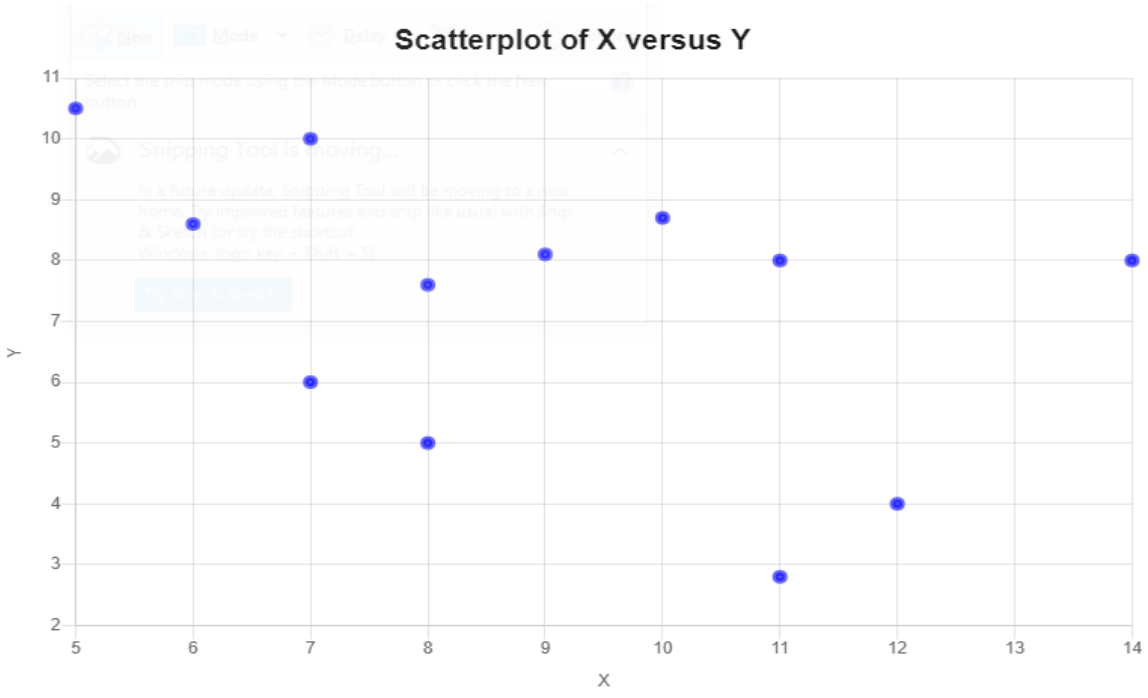
Age (years)	9	7	11	12	8	7	8	11	10	14	6	5
Price (thousand $)	8.1	X-4	2.8	X-6	$\frac{X}{2}$	X	7.6	$\frac{X}{2}+3$	8.7	X-2	8.6	$X+\frac{1}{2}$

Here, X= 10

(a) Draw a scatter diagram and comment on the relation between the age of the car and its selling price.
(b) Determine the Pearson correlation coefficient and the coefficient of determination and interpret it.

Age (years)	9	7	11	12	8	7	8	11	10	14	6	5
Price (thousand $)	8.1	6	2.8	4	5	10	7.6	8	8.7	8	8.6	10.5

(a)



From the scatter diagram, a negative correlation between the age of a car and its selling price can be observed.

(b)

	Age (years) (X)	Price (thousand $) (Y)	X^2	Y^2	XY
	9	8.1	81	65.61	72.9
	7	6	49	36	42
	11	2.8	121	7.84	30.8
	12	4	144	16	48
	8	5	64	25	40
	7	10	49	100	70
	8	7.6	64	57.76	60.8
	11	8	121	64	88
	10	8.7	100	75.69	87
	14	8	196	64	112
	6	8.6	36	73.96	51.6
	5	10.5	25	110.25	52.5
Sum:	108	87.3	1050	696.11	755.6

$$r = \frac{\sum_{i=1}^{12}(x_i y_i) - n\bullet\bar{x}\bullet\bar{y}}{\sqrt{\left(\sum_{i=1}^{12} x_i^{\,2} - n\bar{x}^{2}\right)\times\left(\sum_{i=1}^{12} y_i^{\,2} - n\bar{y}^{2}\right)}}$$

= -0.44

There is a moderate negative correlation between the age of a used car and its selling price.

$r^2 = 0.1936 = 19.36\%$

19.36% of the variation in the price of a used car can be explained by the age of the car.

Question 4

The iodine value is the amount of Iodine necessary to saturate a sample of 100 g of oil.
In the table below the first row states Iodine and second row states oil respectively.

132.0	129.0	120.0	113.2	105.0	92.0	84.0	83.2	88.4	59.0	80.0	81.5	71.0
46.0	48.0	51.0	52.1	54.0	52.0	59.0	58.7	61.6	64.0	61.4	54.6	58.8

x	y	xy	x^2	y^2
132	46	6072	17424	2116
129	48	6192	16641	2304
120	51	6120	14400	2601
113.2	52.1	5897.72	12814.24	2714.41
105	54	5670	11025	2916
92	52	4784	8464	2704
84	59	4956	7056	3481
83.2	58.7	4883.84	6922.24	3445.69
88.4	61.6	5445.44	7814.56	3794.56
59	64	3776	3481	4096
80	61.4	4912	6400	3769.96
81.5	54.6	4449.9	6642.25	2981.16
71	58.8	4174.8	5041	3457.44
Sum= 1238.3	Sum= 721.2	Sum= 67333.7	Sum= 124125.29	Sum= 40381.22

a. Determine the regression equation of 100 g of oil on the amount of Iodine.
Answer:

$$b_1 = \frac{n(\sum(x_i y_i) - (\sum x_i)(\sum y_i)}{n\left(\sum x_i^2\right) - (\sum x_i)^2}$$

$$= -0.22$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$= 76.43$$

$\hat{y} = b_0 + b_1 x$

$\hat{y} = 76.43 - 0.22\ x$

b. Interpret the model.

Answer:

b_0 = 76.43 means that the overall amount of oil will be 76.43, when the amount of iodine is zero.
b_1 = -0.22 means that the average overall amount of oil will decrease by 0.22 units, when the amount of iodine will increase by one unit.

c. What is the predicted price when the amount of Iodine is 88.
Answer:

$\hat{y} = 76.43 - 0.22\ x$
$= 76.43 - (0.22 * 88)$
$= 57.07$

d. Comment on the goodness of fit of the model.
Answer:

Total Sum of Squares (SST) = $\sum (y_i - \bar{y})^2 = \sum y_i^2 - (y_i)^2/n$

Error Sum of Squares (SSE) = $\sum (y_i - \hat{y}_i)^2 = \sum y_i^2 - b_0 \sum y_i - b_1 \sum x_i y_i$

Goodness of fit of the model, $r^2 = 1 - \frac{SSE}{SST}$
$= 0.81$

Comment: 81% variation in amount of oil saturated can be explained by variation in amount of iodine.

Question 5

A study on a range of automotive lubricants reported the following data on oxidation-induction time (min) for various commercial oils:

Sample 1:
87 103 130 160 180 195 132 145 211 105 145 153
152 138 87 99 93 119 129

Sample 2:
99 102 110 33 56 112 130 111 124 155 201 209 103 66 84 75
107 202 59

For which sample of commercial oils, the relative variability of oxidation-induction time is higher?

Answer:

For Sample 1,

x	x^2
87	7569
103	10609
130	16900
160	25600
180	32400
195	38025
132	17424
145	21025
211	44521
105	11025
145	21025
153	23409
152	23104
138	19044
87	7569
99	9801
93	8649
119	14161
129	16641
Sum= 2563	Sum= 368501

For Sample 2,

x	x^2
99	9801
102	10404
110	12100
33	1089
56	3136
112	12544
130	16900
111	12321
124	15376
155	24025
201	40401
209	43681
103	10609
66	4356
84	7056
75	5625
107	11449
202	40804
59	3481
Sum= 2138	Sum= 285158

For Sample 1,

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{n}\right]$$

$$= \frac{1}{18}\left[368501 - \frac{(2563)^2}{19}\right.$$

$$= 1264.77$$

$$s = 35.56$$

$$C_{v(1)} = \frac{s}{\bar{x}} * 100$$

$$= (35.56/134.89) * 100$$

$= 26.36\%$

For Sample 2,

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{n}\right]$$

$$= \frac{1}{18}\left[285158 - \frac{(2138)^2}{19}\right.$$

$$= 2476.49$$

$$s = 49.76$$

$$C_{v(2)} = \frac{s}{\bar{x}} * 100$$

$$= (49.76/112.52) * 100$$

$$= 44.22\%$$

Here, $C_{v(2)} > C_{v(1)}$

Thus, Sample 2 is higher.