Julia Nelson

Homework 01

"I pledge my honor that I have abided by the Stevens Honor system" JNelson

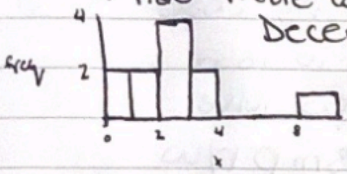
Problem 1

sample of 2 variables (X, Y)

(0.2, 1.1), (1.2, 2.3), (0.9, 1.1), (2.2, 3.6), (3.2, 0.1), (3.1, 4.8),
(2.3, 6.5), (1.5, 7.8), (3.0, 8.0), (2.6, 9.4), (9.0, 9.8)

(i) Histogram of X

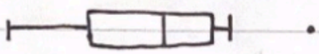
The distribution is skewed from its outlier + has little to no symmetry
Decently centered



Pie chart of Y

Slightly more evenly distributed
No outliers, but no center or symmetry

(ii) X box-plot



Sample Min: 0.2
1st Quartile: 1.2
2nd Quartile: 2.3
3rd Quartile: 3.1
Sample Max: 9.0 ← OUTLIER

Variance:

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \Rightarrow \ldots$

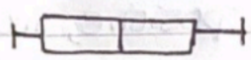
$\rightarrow \bar{x} = \frac{1}{11} \sum_{i=1}^{11} a_i \quad \rightarrow \sum_{i=1}^{n} (x_i - \bar{x})^2$

$= \frac{1}{11}(29.2) \quad (0.2-2.6545)^2 + (0.9-2.6545)^2 + \ldots$

$= 2.6545 \quad = 53.76727$

$\ldots \Rightarrow = \frac{53.76727}{11-1} = \boxed{5.3767}$

Y box plot



Sample Min: 0.1
1st Quartile: 1.1
2nd Quartile: 4.8 → No outliers
3rd Quartile: 8.0
Sample Max: 9.8

Variance:

$\bar{x} = \frac{1}{11} \sum_{i=1}^{11} a_i = \frac{1}{11}(54.5)$

$= 4.9545$

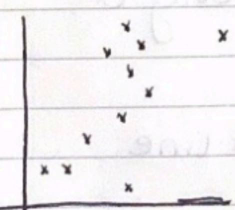
$S^2 = \frac{1}{11-1} \sum_{i=1}^{11} (x_i - \bar{x})^2$

$= \left(\frac{1}{10}\right)\left((0.1-4.9545)^2 + (1.1-4.9545)^2 + \ldots\right)$

$= 125.18727/10$

$= \boxed{12.51872}$

(iii)



$\sum_{i=1}^{11} (X - 2.655)(Y - 4.9545)$

$= 45.707$

$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$

$= 45.707 / \sqrt{53.767 \cdot 125.187}$

$= \boxed{0.5571}$

There is closer to a strong positive linear relationship between ~~them~~ variables.

(iv) There is one outlier of (X, Y)
it is (9, 9.8) because of its
outlying with x values

sum of X = 20.2 sum of Y = 44.7

mean of X = 2.02 mean of Y = 4.47

$S_x = 9.476$ $S_y = 99.361$

$\Sigma(x-\bar{x})^2$ $\Sigma(y-\bar{y})^2$

$n = 10$

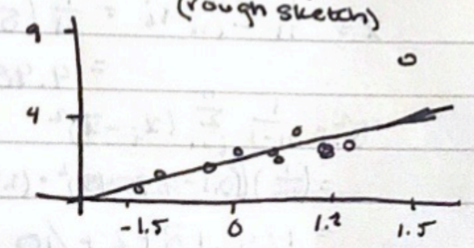
$\Sigma(x-\bar{x})(y-\bar{y}) = 11.886$

$r = 11.886 / \sqrt{9.476 \cdot 99.361} = \boxed{0.3874}$

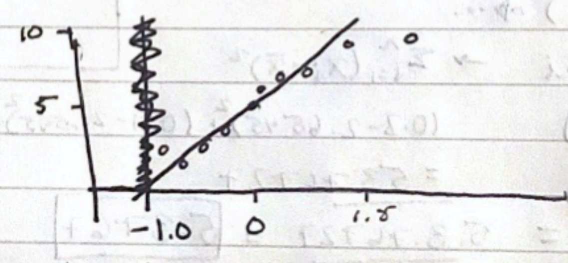
This one is closer to 0 so it is a weaker positive relationship b/w X and Y

(v) The first correlation coefficient was also positive, like the second. However, the first lays closer to +1 while the second lies closer to 0. This means without the outlier there is a weaker positive relationship b/w X and Y than w/ the outlier.

(vi) X QQ plot
(rough sketch)



Y QQ plot



Y is most likely to be Normal distribution because of the evenness / symmetry of the QQ plot compared to X's.
It is also more balanced around 0
Also has less deviation from the Normal line
(Not skewed)