

Assignment - 4

Problems from chapter 8

8.2

Given number of concurrent were is recorded

17.2	22.1	18.5	17.2	18.6	14.8	21.7	15.8	16.3	22.8
24.1	13.3	16.2	17.5	19.0	23.9	14.8	22.2	21.7	20.7
13.5	15.8	13.1	16.1	21.9	23.9	19.3	12.0	19.9	19.4
15.4	16.7	19.5	16.2	16.9	17.1	20.2	13.4	19.8	13.7
19.7	18.7	17.6	15.9	15.2	17.1	15.0	18.8	21.6	11.9

- (a) Compute the Sample mean, Variance and Standard deviation of the number of concurrent users.

$$\text{Mean } \bar{x} = \frac{\sum x_i}{n}$$

$$= \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_n}{n}$$

$$= \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_{50}}{50}$$

$$\bar{x} = \frac{17.2 + 22.1 + 18.5 + 17.2 + \dots + 21.6 + 11.9}{50}$$

$$\bar{x} = \frac{89.77}{50}$$

$$\bar{x} = 17.954$$

$$\text{mean} = 17.954.$$

$$\text{Sample Variance } s^2 = \frac{\sum_{i=1}^{50} x_i^2 - (80)x^2}{(n-1)}$$

Sec Parthish Mandemula
1002022847.

$$S^2 = \frac{(11.2)^2 + (22.1)^2 + (18.5)^2 + \dots + (11.9)^2}{(50-1)} - \frac{50(13.954)^2}{(50-1)}$$

$$S^2 = \frac{488.444}{49}$$

$$S^2 = 9.9682$$

$$S^2 = 9.97$$

we know that, Standard deviation

$$\begin{aligned} &= \sqrt{\text{Variance}} \\ &= \sqrt{9.9682} \\ &= 3.1573 \\ &\approx 3.16 \end{aligned}$$

Standard deviation = 3.16

- b) Estimate the Standard Error of the Sample mean

Standard Error of Sample mean

$$\begin{aligned} S(\bar{x}) &= S/\sqrt{n} \\ &= 3.1573 \\ &/\sqrt{50} \\ &= 0.4465 \\ &\approx 0.447 \end{aligned}$$

Standard Error = 0.447

- c) Compute the five-point Summary and construct a boxplot

Sai Parthish Mandemula
1002022843

five point Summary = (min, Q₁, M, Q₃, max)

By Sorting the given values

minimum = 11.4

maximum = 24.1

Q₁ = 13th Smallest = 15.8

Q₃ = 37th Smallest = 19.9

Median = It is between 25 and 26. So the values are 17.5 & 17.6

$$= 17.55$$

∴ five point Summary = (11.9, 15.8, 17.55, 19.9, 24.1)

(d) Compute the Interquartile range. Are there any outliers?

Inter quartile range = Q₃ - Q₁

$$= 19.9 - 15.8 \\ = \underline{\underline{4.1}}$$

$$\begin{aligned} Q_3 + 1.5(IQR) &= 19.9 + (1.5)(4.1) \\ &= 26.05 \end{aligned}$$

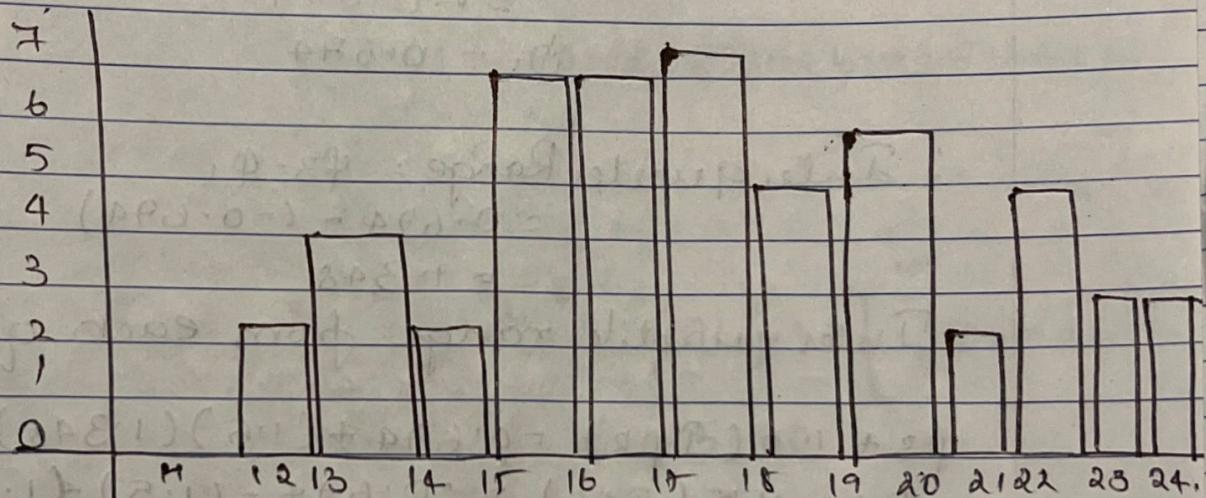
$$\begin{aligned} Q_1 - 1.5(IQR) &= 15.8 - (1.5)(4.1) \\ &= 9.65 \end{aligned}$$

Therefore, in the values we have are starting from 11.9 and ends at 24.1.

SciParthish Mandumula
1002022847.

i. we got the values 9.65 and 26.05 less than or greater than it
Hence, it has no outliers.

- ② It is reported that the number of concurrent users follows approximately Normal distribution. Does this histogram support their claim?



This histogram does not support their claim because it does not have a curve shaped or a bell shaped curve. So, it doesn't support assumption for normal distribution.

6.4) Given that

$$\begin{aligned}\Phi_3 \text{ is } 75^{\text{th}} \text{ percentile} \\ P(Z \leq z) = 0.75 \\ \Rightarrow \Phi(z) = 0.75 \\ z = \Phi^{-1}(0.75)\end{aligned}$$

Φ_1 is 25^{th} percentile

$$\begin{aligned}P(Z \leq x) = 0.25 \\ \Phi(x) = 0.25 \\ x = \Phi^{-1}(0.25) \\ = -0.674 \\ \Phi_1 = 0.674\end{aligned}$$

$$\begin{aligned}\therefore \text{Interquartile Range} &= \Phi_3 - \Phi_1 \\ &= 0.674 - (-0.674) \\ &= 1.348\end{aligned}$$

Inter quartile range from each quartile

$$\begin{aligned}\Phi_3 + 1.5(IQR) &= 0.674 + (1.5)(1.348) \approx 2.3 \\ \Phi_1 - 1.5(IQR) &= -0.674 - (1.5)(1.348) \approx -2.3 \\ P(Z \leq 2.3) - P(Z \leq -2.3) \\ &= \Phi(2.3) - \Phi(-2.3) \\ &\Rightarrow 0.9965 - 0.0035 \\ &= 0.993\end{aligned}$$

The probability of normal distribution variable is 0.993

8.9) The given dataset represents the number of new computer accounts registered during ten consecutive days.

43, 37, 50, 51, 58, 105, 52, 45, 45, 10

@ Compute the mean, median, quartiles and Standard deviation.

$$\textcircled{1} \text{ mean: } \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad n=10$$

$$\bar{x} = \frac{43 + 37 + 50 + 51 + 58 + 105 + 52 + 45 + 45 + 10}{10}$$

$$\Rightarrow \frac{496}{10} = 49.6$$

$$\textcircled{2} \text{ median} = \text{Avg of } \left(\frac{n}{2} \right)^{\text{th}} \text{ & } \left[\left(\frac{n}{2} \right) + 1 \right]^{\text{th}} \text{ position.}$$

$$= \frac{5^{\text{th}} + 6^{\text{th}}}{2} \Rightarrow \frac{45 + 50}{2}$$

$$\text{median} = 47.5 = Q_2$$

Q_1 = median of first half of the data.

$$\Rightarrow \text{median of } 10, 37, 43, 45, 45 \quad Q_1 = 43.$$

Q_3 = median of second half of the data

= median of 50, 51, 52, 58, 105

$$Q_3 = 52$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

$$\begin{aligned}
 &= \sqrt{\frac{(10r + 37r + \dots + 105r) - (10)(49.6)}{10}} \\
 &= \sqrt{\frac{4960.4}{10}} \\
 \sigma &= \underline{22.27}
 \end{aligned}$$

b) Check for the outliers using the $1.5(\text{IQR})$

$$\begin{aligned}
 \text{IQR} &= Q_3 - Q_1 \\
 &= 52 - 43 \\
 &= 9
 \end{aligned}$$

$$\begin{aligned}
 Q_1 - 1.5(\text{IQR}) &= 43 - 1.5(9) \\
 &= 43 - 13.5 \\
 &= 29.5
 \end{aligned}$$

$$\begin{aligned}
 Q_3 + 1.5(\text{IQR}) &= 52 + 1.5(9) \\
 &= 52 + 13.5 \\
 &= 65.5
 \end{aligned}$$

$$Q_1 - 1.5(\text{IQR}) \leq x \leq Q_3 + 1.5(\text{IQR})$$

$$29.5 \leq x \leq 65.5$$

Outlier = $10 \notin [29.5, 65.5]$

c) Delete the detected outliers and Compute the mean, median, quartiles, and standard deviation again.

37, 43, 45, 47, 50, 51, 52, 58

$$\begin{aligned}
 n &= 8 \\
 \text{mean} &= \frac{\sum x_i}{n}
 \end{aligned}$$

Sai Parthish Mandemula
1002022843

$$37 + 43 + \dots + 58 = \frac{381}{8} - \bar{x} = 47.625$$

$$\text{median is } \frac{45+50}{2} = 47.5$$

Quartiles is $Q_1 = 43$ (25 percent)
 $Q_3 = 51$ (75 percent)

Standard deviation is

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

$$\Rightarrow \sqrt{(37-47.625)^2 + (43-47.625)^2 + \dots + (58-47.625)^2}$$

$$\Rightarrow \sqrt{\frac{241.875}{7}} \Rightarrow \sqrt{41.6964} \Rightarrow 6.457$$

Standard deviation = 6.457

- d) Make a conclusion about the effect of outliers on basic descriptive statistics

The mean ~~for~~ and the standard deviation is decreased when we remove the outliers. It describe the data that are close in value. The data will be related accurately with the median. If we want to categorize or group the data the mode is the best one to do.