Statistics

Lecture 4

Feb 19-8:47 AM

Consider the Sample below

2 3 3 3 5 5 8 10

1) $n = 8$

2) Min. = 2 , Max. = 10

3) Range = Max − Min = 10 − 2 = $\boxed{8}$

4) Midrange = $\frac{\text{Max} + \text{Min}}{2}$

$= \frac{10+2}{2} = 6$

5) Mode = 3

6) $\sum x = 2 + 3 + 3 + 3 + 5 + 5 + 8 + 10 = \boxed{39}$

7) $\sum x^2 = 2^2 + 3^2 + 3^2 + 3^2 + 5^2 + 5^2 + 8^2 + 10^2 = \boxed{245}$

8) $\bar{x} = \frac{\sum x}{n} = \frac{39}{8} = \boxed{4.875}$

Sample Mean

Round to	
whole	5
1-decimal	4.9
2-decimal	4.88

Feb 24-1:48 PM

9) $S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{8 \cdot 245 - 39^2}{8(8-1)}$

Sample Variance

$= \frac{439}{56} = 7.839$

10) $S = \sqrt{S^2} = \sqrt{7.839} = 2.799|8$

Sample Standard Deviation

$\approx \boxed{2.800}$ 3-decimal places

→ whole $S \approx 3$

How to estimate S:

$$S \approx \frac{\text{Range}}{4}$$

Range Rule-of-thumb

$S \approx \frac{8}{4} = \boxed{2}$

Feb 24-1:56 PM

Consider the Sample below

1 2 3 3 3

5 5 5 8 11

1) $n = \boxed{10}$

2) Range $= 11 - 1 = \boxed{10}$

3) Midrange $= \frac{11+1}{2} = \boxed{6}$

4) Mode: 3 & 5 Bimodal

5) $\sum x = 46$

6) $\sum x^2 = 292$

7) $\bar{x} = \frac{\sum x}{n} = \frac{46}{10} = \boxed{4.6}$

Sample Mean

8) $S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}$

Sample Variance $= \frac{10 \cdot 292 - 46^2}{10(10-1)}$

$= \frac{804}{90}$

$= 8.9\bar{3}$

$\approx \boxed{8.933}$

9) $S = \sqrt{S^2} = \sqrt{8.933} \approx \boxed{2.989}$

Sample Standard Deviation

10) Estimate S

Range Rule-of-thumb

$S \approx \frac{\text{Range}}{4} = \frac{10}{4} = \boxed{2.5}$

$S \geq 0$

$S^2 \geq 0$

Feb 24-2:03 PM

What is Sample Standard deviation?

It is a non-negative numerical value that indicates how data elements are spread from the mean.

If S is small ⇒ data elements are close to the mean.

If S is big ⇒ data elements are more spread out from the mean.

If S is zero ⇒ All data elements are the same as the mean ⇒ No deviation from the mean.

Feb 24-2:17 PM

Consider the sample below

5 5 5 5 5 5

$n = 6$ $\quad \sum x = 30$ $\quad \sum x^2 = 150$

$\bar{x} = \frac{\sum x}{n} = \frac{30}{6} = \boxed{5}$

$$S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{6 \cdot 150 - 30^2}{6(6-1)} = \frac{900-900}{30} = \frac{0}{30} = \boxed{0}$$

$S = \sqrt{S^2} = \sqrt{0} = \boxed{0}$

Since $S = 0$ ⇒ All data elements are equal to $\bar{x} = 5$

Feb 24-2:22 PM

Empirical Rule

this works best when data distribution is symmetric and has a bell-shape graph.

This happen when Mean = Mode = Median.

About 68% of elements are between $\bar{x} \pm S$.

About 95% " " " " $\bar{x} \pm 2S$

Usual Range

About 99.7% " " " " $\bar{x} \pm 3S$

Feb 24-2:28 PM

I Surveyed 180 students.

Mean & standard deviation of their age was 32 years and 7 years.

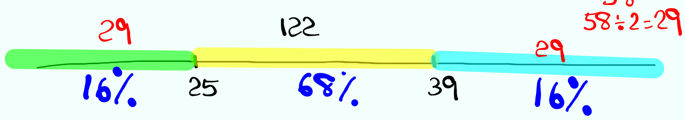
These ages had a symmetric dist.

$\bar{x} = 32$ $S = 7$ $n = 180$

68% Range $\Rightarrow \bar{x} \pm S = 32 \pm 7 \Rightarrow$ 25 to 39

68% of 180 = .68(180) ≈ 122

180
−122
58
58 ÷ 2 = 29



what % are above 25? 68% + 16% = 84%

How many were above 25? 122 + 29 = 151

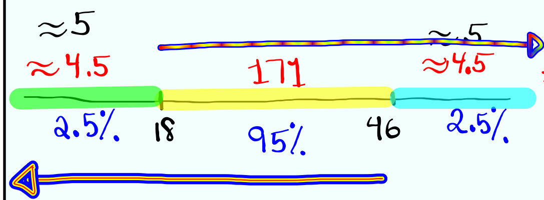
Feb 24-2:34 PM

Usual Range => 95% Range => $\bar{x} \pm 2S$

$32 \pm 2(7)$

$= 32 \pm 14$

$\Rightarrow$ 18 to 46



How many are between 18 & 46?

95% of 180 = .95(180)

100% − 95% = 5%

5% ÷ 2 = 2.5%

what % of students were below 46?

95% + 2.5% ≈ 97.5%

How many of them were above 18?

171 + 5 ≈ 176

S & S

Feb 24-2:41 PM

5 - Number Summary

Data must be Sorted

Min Q_1 Median Q_3 Max

this divides data elements in 4 group.

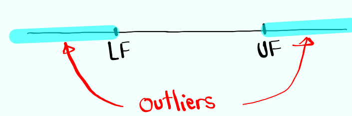
Each group is 25% of total Sample

IQR (Inter-Quartile-Range) = $Q_3 - Q_1$

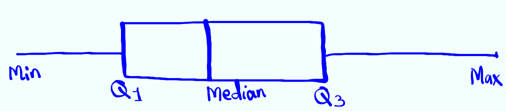
Upper Fence = $Q_3 + 1.5(IQR)$

Lower Fence = $Q_1 - 1.5(IQR)$

Any data element greater than the upper fence or Smaller than the lower fence is called Outlier.



Draw Box Plot



Feb 24-3:04 PM

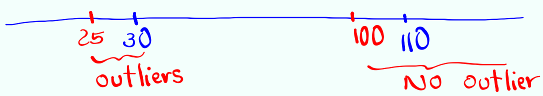
I randomly Selected 40 exams, here are 5-Number Summary of the Scores

25 (Min.), 60 (Q_1), 75 (Med.), 80 (Q_3), 100 (Max.)

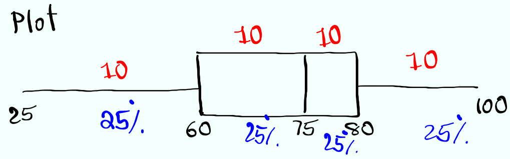
$IQR = Q_3 - Q_1 = 80 - 60 = 20$

Upper fence $= Q_3 + 1.5(IQR) = 80 + 1.5(20) = 110$

Lower fence $= Q_1 - 1.5(IQR) = 60 - 1.5(20) = 30$



Box Plot



$40 \div 4 = 10$

How many Scored above 60? 30

What % Scored below 80? 75%

Feb 24-3:12 PM

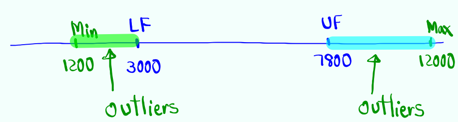
120 nurses had the following 5-Number Summary for their monthly Salaries:

1200 (Min), 4800 (Q_1), 5500 (Med.), 6000 (Q_3), 12000 (Max)

$IQR = Q_3 - Q_1 = 6000 - 4800 = \boxed{1200}$

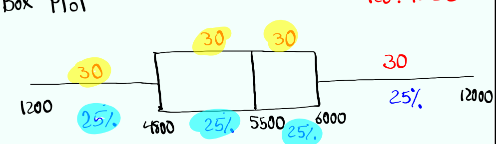
Upper fence $= Q_3 + 1.5(IQR) = 6000 + 1.5(1200) = \boxed{7800}$

Lower fence $= Q_1 - 1.5(IQR) = 4800 - 1.5(1200) = \boxed{3000}$



Box Plot

$120 \div 4 = 30$



How many nurses were making below \$6000? 90

What % of these nurses were making below \$6000? 75%

Feb 24-3:20 PM

Dianne made 88 on exam 1 and 78 on exam 2.

we want to Compare these 2 Scores.

we need Z-Scores (Standardize the Score)

$Z = \frac{x - \bar{x}}{S}$ Always round to 3-decimal places.

Exam 1: $\bar{x} = 85$, $S = 5$ $\quad Z = \frac{88-85}{5} = \frac{3}{5} = \boxed{.6}$

Exam 2: $\bar{x} = 70$, $S = 4$ $\quad Z = \frac{78-70}{4} = \boxed{2}$

Dianne did better on exam 2.

$Z < -2$		$-2 \le Z \le 2$		$Z > 2$
unusual	-2	usual	2	unusual

Feb 24-3:31 PM

Allen makes \$8000/month as a Salesman.

Carol makes \$20000/month as a dentist.

who is doing better?

Sales $\bar{x} = 5000$, $S = 1000$

$Z = \frac{8000 - 5000}{1000} = \boxed{3}$ — Allen's salary is unusual.

Dentist $\bar{x} = 18000$ $S = 2000$

$Z = \frac{20000 - 18000}{2000} = \boxed{1}$ — Carol's Salary is usual.

SG 6

Feb 24-3:38 PM