Statistics

Lecture 3

Feb 19-8:47 AM

SG 5

x → data element

$\sum x$ → Sum of data elements

↑ Summation

n → Sample Size

$\bar{x}$ → x-bar → Sample Mean (Average)

$$\bar{x} = \frac{\sum x}{n}$$

Consider the Sample below

1, 2, 2, 2, 3

$n = 5$

Range = 3 − 1 = 2

Midrange = $\frac{3+1}{2} = 2$

Mode 2

Median 2

Value in the middle

"Data must be Sorted"

$\sum x = 1 + 2 + 2 + 2 + 3 = 10$

$\bar{x} = \frac{\sum x}{n} = \frac{10}{5} = 2$

when mean, mode, and median are the same, data dist. will be Symmetric.

Sep 15-5:06 PM

Consider the Sample below

1 6 6 8 8 9

1) $n = 6$ ✓

2) Range $= 9 - 1 = 8$

3) Midrange $= \frac{9+1}{2} = 5$

4) Mode $6 \& 8$

5) Median $\frac{6+8}{2} = 7$

Value in the middle

data must be Sorted

6) $\sum x = 1 + 6 + 6 + 8 + 8 + 9 = 38$

7) $\bar{x} = \frac{\sum x}{n} = \frac{38}{6} = 6.\overline{3}$

Round to

whole $\longrightarrow 6$

1-dec. $\longrightarrow 6.3$

Round-up to a whole #

7

Sep 15-5:14 PM

$n \rightarrow$ Sample Size

$x \rightarrow$ data elements

$\sum x \rightarrow$ Sum of data elements

$\sum x^2 \rightarrow$ Sum of data elements^2.

$\bar{x} \rightarrow$ Sample Mean

$s^2 \rightarrow$ Sample Variance

$$\bar{x} = \frac{\sum x}{n}$$

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

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

Sep 15-5:22 PM

Consider the Sample below

1 3 3 3 5

$n = 5$ Range $= 5 - 1 = 4$ Midrange $= \frac{5+1}{2} = 3$

Mode 3 Median 3

$\sum x = 1 + 3 + 3 + 3 + 5 = \boxed{15}$

$\sum x^2 = 1^2 + 3^2 + 3^2 + 3^2 + 5^2 = \boxed{53}$

$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = \boxed{3}$

$s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{5 \cdot 53 - 15^2}{5(5-1)} = \frac{40}{20} = \boxed{2}$

Sep 15-5:26 PM

Consider the Sample below

1 3 5 7 9 11

1) $n = 6$ 2) Range $= 10$ 3) Midrange $= 6$

4) Mode None 5) Median $\frac{5+7}{2} = 6$

6) $\sum x = 1 + 3 + 5 + 7 + 9 + 11 = \boxed{36}$

7) $\sum x^2 = 1^2 + 3^2 + 5^2 + 7^2 + 9^2 + 11^2 = \boxed{286}$

8) $\bar{x} = \frac{\sum x}{n} = \frac{36}{6} = \boxed{6}$

9) $s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}$

$= \frac{6 \cdot 286 - 36^2}{6(6-1)}$

$= \frac{420}{30} = \boxed{14}$

Sep 15-5:33 PM

$\bar{x}$ → Sample Mean

s^2 → Sample Variance

s → Sample Standard deviation

$$\bar{x} = \frac{\sum x}{n}$$

$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$

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

$$s = \sqrt{s^2}$$

$s^2 \geq 0$ Always

$s \geq 0$

Sep 15-5:46 PM

Given $n = 5$ $\quad \sum x = 15$ $\quad \sum x^2 = 47$

$$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = \boxed{3}$$

$$s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{5 \cdot 47 - 15^2}{5(5-1)} = \frac{10}{20} = \boxed{.5}$$

$$s = \sqrt{s^2} = \sqrt{.5} \approx \boxed{.707}$$

Sep 15-5:49 PM

Given $n=8$ $\sum x=40$ $\sum x^2=224$

$\min=2$ $\max=8$

Range $=8-2=\boxed{6}$

Midrange $=\frac{8+2}{2}=\boxed{5}$

$\bar{x}=\frac{\sum x}{n}=\frac{40}{8}=\boxed{5}$

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

$=\frac{8\cdot 224-40^2}{8(8-1)}=\frac{192}{56}$

$\approx\boxed{3.429}$

$s=\sqrt{s^2}$

$=\sqrt{3.429}\approx\boxed{1.852}$

How to estimate s:

$s\approx\frac{\text{Range}}{4}$ Range rule-of-thumb

Sep 15-5:54 PM

Sample variance is a measure of how spread out the data elements are from $\bar{x}$.

Standard deviation brings the variance to the same unit as the original data.

when s is small → data elements are close to $\bar{x}$.

when s is big → data elements are more spread out from $\bar{x}$.

when s is zero → All data elements are equal to $\bar{x}$.

Sep 15-6:13 PM

whenever mean, mode, and median are the same, data dist. will be symmetric and has a bell-shape graph.

Empirical Rule

About 68% of data elements are within $\bar{x} \pm s$

About 95% of data elements are within $\bar{x} \pm 2s$

Usual Range

About 99.7% of data elements are within $\bar{x} \pm 3s$

Sep 15-6:20 PM

I randomly selected 60 exams, scores had bell-shape dist. with $\bar{x} = 83$ and $s = 6$.

68% Range $\rightarrow \bar{x} \pm s = 83 \pm 6 \rightarrow$ 77 to 89

95% Range $\rightarrow \bar{x} \pm 2s = 83 \pm 2(6)$

$= 83 \pm 12 \rightarrow$ 71 to 95

Usual Range

2.5% | 95% | 2.5%

≈2 | 71 | 57 | 95 | ≈2

95%(60) = 57

99.7% Range $\rightarrow \bar{x} \pm 3s = 83 \pm 3(6)$

$= 83 \pm 18 \rightarrow$ 65 to 101

Sc 5 ✓

Sep 15-6:26 PM

SG 6

Z-Score Always Round to 3-dec. places

$$Z = \frac{x - \bar{x}}{S}$$

Z-Score calculates how many standard deviations is the element above or below the mean.

Z-Score allows us to compare data elements from different Sample.

Z-Score is the way to standardize data elements

If $-2 \leq Z \leq 2$ → data element is Usual

If $Z < -2$ or $Z > 2$ → data element is unusual

Sep 15-6:33 PM

Lisa got 90 on exam 1.

$\bar{x} = 82$ & $S = 5$

$$Z = \frac{x - \bar{x}}{S} = \frac{90 - 82}{5} = \frac{8}{5} = 1.6$$

Since $-2 \leq Z \leq 2$ → 90 was an usual Score.

I had a Z-Score of -2.2, what was my Score?

$$Z = \frac{x - \bar{x}}{S} \qquad -2.2 = \frac{x - 82}{5}$$

Cross-multiply $\quad x - 82 = 5(-2.2)$

$x - 82 = -11$

$x = 71$

Since $Z < -2$, my Score is unusual.

unusual | usual | unusual

-2 $\quad$ 2

Sep 15-6:39 PM

Maria makes $4000/mo. as a nurse.

$$z = \frac{x - \bar{x}}{s} = \frac{4000 - 3000}{400} = \frac{1000}{400} = 2.5$$

$\bar{x} = 3000$, $s = 400$

Mike makes $6000/mo. as a manager.

$$z = \frac{x - \bar{x}}{s} = \frac{6000 - 5000}{600} = \frac{1000}{600} \approx 1.66$$

$\bar{x} = 5000$, $s = 600$

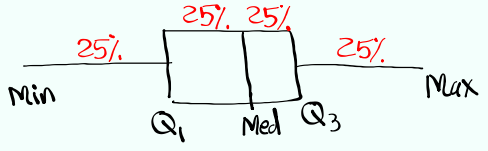
Who does better? **Maria**

Sep 15-6:45 PM

5-Number Summary

Min Q_1 Med Q_3 Max

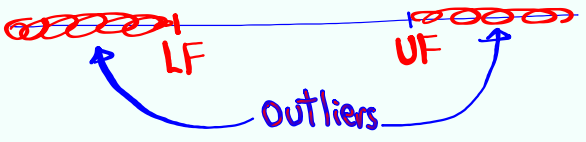
Box Plot



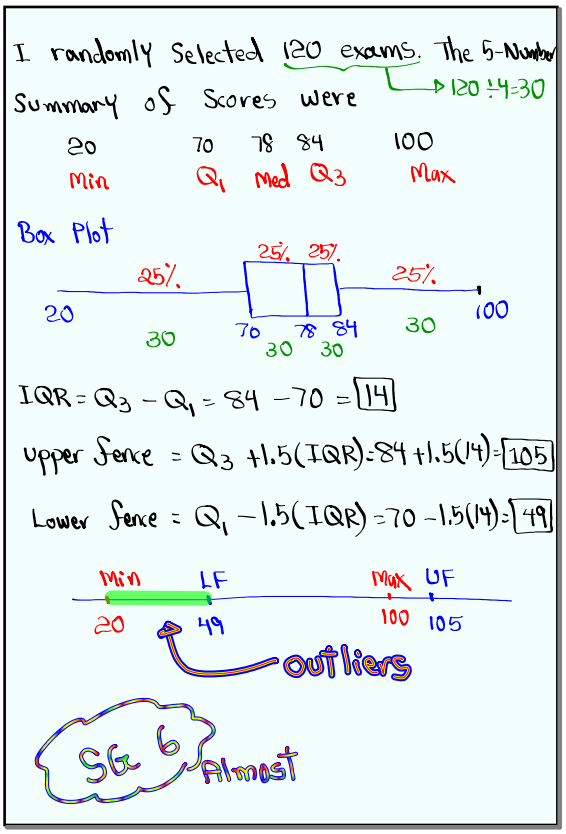
IQR (Inter-Quartile-Range) $= Q_3 - Q_1$

Upper fence $= Q_3 + 1.5(IQR)$

Lower fence $= Q_1 - 1.5(IQR)$



Sep 15-6:50 PM


I randomly Selected 120 exams. The 5-Number
Summary of Scores were
120 ÷ 4 = 30
20 70 78 84 100
Min Q1 Med Q3 Max
Box Plot
25% 25% 25% 25%
20 70 78 84 100
30 30 30 30
IQR = Q3 − Q1 = 84 − 70 = 14
Upper fence = Q3 + 1.5(IQR) = 84 + 1.5(14) = 105
Lower fence = Q1 − 1.5(IQR) = 70 − 1.5(14) = 49
Min LF Max UF
20 49 100 105
outliers
SG 6
Almost

Sep 15-6:56 PM