Statistics
Summer 2021
Lecture 2

Class Quiz 1

1) Simplify: $\dfrac{82-70}{\sqrt{16}} = \dfrac{12}{4} = \boxed{3}$

2) Solve $2.5x - 12 = 8$

$2.5x = 8 + 12$

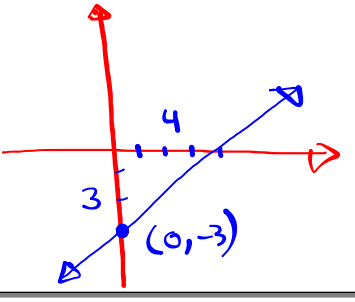
$2.5x = 20$

$x = \dfrac{20}{2.5}$ $\boxed{x = 8}$

3) Graph $y = \dfrac{3}{4}x - 3$

Y-Int $(0, -3)$

$m = \dfrac{3}{4}$



SG 1 – SG 4

SG 5 – SG 8

Basic Computations in Statistics

n → Sample Size

x → Data element

$\sum x$ → Summation of data elements

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

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

Consider the Sample below

2, 5, 5, 5, 8

$n = \boxed{5}$ Range $= 8 - 2 = \boxed{6}$ Midrange $= \frac{8+2}{2} = \boxed{5}$

Mode $= \boxed{5}$ $\sum x = 2+5+5+5+8 = \boxed{25}$

$$\bar{x} = \frac{\sum x}{n} = \frac{25}{5} = \boxed{5}$$

Consider the Sample below

1, 3, 3, 3, 5, 5, 5, 9

1) $n = \boxed{8}$

2) Range $= 9 - 1 = \boxed{8}$

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

4) Mode $= 3 \& 5$

Bimodal

5) $\sum x = 1 + 3 + 3 + 3 + 5 + 5 + 5 + 9 = \boxed{34}$

6) $\bar{x} = \frac{\sum x}{n} = \frac{34}{8} = \boxed{4.25}$

$n \rightarrow$ Sample Size

$x \rightarrow$ Data element

$x^2 \rightarrow$ Data element^2

$\sum x \rightarrow$ Summation of data elements

$\sum x^2 \rightarrow$ Square data elements, then add them

$\bar{x} \rightarrow$ Sample Mean $\qquad \bar{x} = \frac{\sum x}{n}$

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

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

Consider the Sample below

2, 3, 5, 8, 12

1) $n = \boxed{5}$ 2) Range $= 12 - 2 = \boxed{10}$ 3) Midrange $= \frac{12+2}{2} = \boxed{7}$

4) Mode = None 5) $\Sigma x = 2+3+5+8+12 = \boxed{30}$

6) $\bar{x} = \frac{\Sigma x}{n} = \frac{30}{5} = \boxed{6}$ 7) $\Sigma x^2 = 2^2+3^2+5^2+8^2+12^2 = \boxed{246}$

8) $S^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)} = \frac{5 \cdot 246 - 30^2}{5(5-1)} = \frac{330}{20} = \frac{33}{2} = \boxed{16.5}$

S → Sample Standard deviation

$S = \sqrt{S^2}$ Standard Dev. $= \sqrt{\text{Variance}}$

$S = \sqrt{16.5} \approx \boxed{4.062}$

Consider the Sample below

1, 2, 2, 3, 3, 4, 4, 11

1) $n = \boxed{8}$ 2) Range $= 11 - 1 = \boxed{10}$ 3) Midrange $= \frac{11+1}{2} = \boxed{6}$

4) Mode = 2, 3, & 4 Trimodal 5) $\Sigma x = 1+2+2+3+3+4+4+11 = \boxed{30}$

6) $\bar{x} = \frac{\Sigma x}{n} = \frac{30}{8} = \boxed{3.75}$ 7) $\Sigma x^2 = 1^2+2^2+2^2+3^2+3^2+4^2+4^2+11^2 = \boxed{180}$

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

$= \frac{8 \cdot 180 - 30^2}{8(8-1)} = \frac{540}{56} = \boxed{\frac{135}{14}} = 9.643$

9) $S = \sqrt{S^2}$

$= \sqrt{9.643}$

$S \approx \boxed{3.105}$

How to estimate Sample Standard deviation:

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

The range rule-of-thumb

Exam 1: Min 50 , Max 100

Estimate S $\quad S \approx \frac{\text{Range}}{4} = \frac{100-50}{4} = \frac{50}{4}$

$= \boxed{12.5}$

What is Standard deviation?

- Non-negative
- It indicates how data elements are Spread with respect to $\bar{x}$.

When S is Small $\Rightarrow$ Data elements are close to $\bar{x}$.

When S is big $\Rightarrow$ Data elements are more Spread out from $\bar{x}$.

When S is Zero $\Rightarrow$

- No deviation from $\bar{x}$
- All data elements are the Same
- Data elements $= \bar{x}$.

Empirical Rule:

This rule is best when data distribution is symmetric. (Mean = Mode = Median)

1) 68% Range $\Rightarrow \bar{x} \pm s$

2) 95% Range $\Rightarrow \bar{x} \pm 2s$ ⇐ Usual Range

3) 99.7% Range $\Rightarrow \bar{x} \pm 3s$

Exam 1 Scores have a symmetric distribution with $\bar{x} = 82$ and $s = 6$.

Use empirical rule to find

1) 68% Range $\Rightarrow \bar{x} \pm s = 82 \pm 6 \Rightarrow$ 76 to 88

2) 95% Range $\Rightarrow \bar{x} \pm 2s = 82 \pm 2(6) \Rightarrow$ 70 to 94

Usual Range

3) 99.7% Range $\Rightarrow \bar{x} \pm 3s = 82 \pm 3(6) \Rightarrow$ 64 to 100

Salaries of 200 nurses randomly selected had a symmetric dist with mean of \$6400 and Stand. dev. of \$300.

1) 68% Range

$\bar{x} \pm S = 6400 \pm 300$

$\Rightarrow$ 6100 to 6700

Usual Range

2) 95% Range

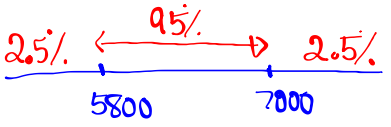
$\bar{x} \pm 2S$

$= 6400 \pm 2(300)$

$= 6400 \pm 600$

$\Rightarrow$ 5800 to 7000

3) How many of them make more than \$7000?



2.5% of 200 = .025(200) = 5

Z – Score

- Round 3-decimals
- $Z = \frac{x - \bar{x}}{S}$
- It is a method to standardize data element.
- It allows us to compare data element from different samples

When $-2 \leq Z \leq 2 \Rightarrow$ Data element is usual

When $Z < -2$ or $Z > 2 \Rightarrow$ unusual data element

Z-Score indicates how many standard deviation is the data element above or below the mean.

Exam 1 : $\bar{x} = 84$, $S = 8$

Osvaldo got 90.

$-2 \le Z \le 2$ Usual Score

$$Z = \frac{x - \bar{x}}{S} = \frac{90 - 84}{8} = \frac{6}{8} = \boxed{.75}$$

Exam 2: $\bar{x} = 74$, $S = 5$

Osvaldo got 85.

$Z < -2$ or $Z > 2$ Unusual Score

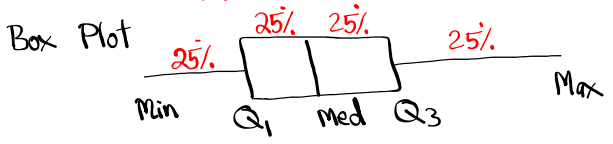
$$Z = \frac{x - \bar{x}}{S} = \frac{85 - 74}{5} = \frac{11}{5} = \boxed{2.2}$$

5- Number Summary

Data must be sorted

Min , Q_1 , Med. , Q_3 , Max

First Quartile — Median — Third Quartile



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

Upper Fence = $Q_3 + 1.5(IQR)$

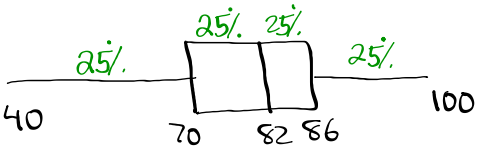
Lower Fence = $Q_1 - 1.5(IQR)$

any data element below lower fence or above the upper fence is an outlier.

Exam 1 results have a 5-number Summary of 40, 70, 82, 86, and 100.

Min = 40, Q_1 = 70, Med = 82, Q_3 = 86, Max = 100

Draw Box Plot



$IQR = Q_3 - Q_1 = 86 - 70 = 16$

Upper Fence $= Q_3 + 1.5(IQR) = 86 + 1.5(16) = \boxed{110}$

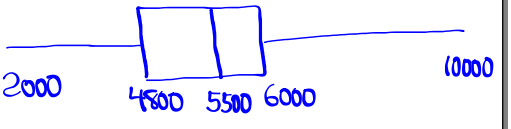
Lower Fence $= Q_1 - 1.5(IQR) = 70 - 1.5(16) = \boxed{46}$

Outliers $\boxed{40 \text{ to } 46}$

240 randomly selected nurses had a 5-number Summary for their salaries given below

2000, 4800, 5500, 6000, and 10000

Draw Box Plot



$IQR = Q_3 - Q_1 = 6000 - 4800 = 1200$

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

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

Outliers $\boxed{2000 - 3000}$ OR $\boxed{7800 - 10000}$

Class QZ 2

Complete the chart below, then draw Freq. Polygon. clearly label.

Class MP	class F	Cum. F
24	2	
30	8	
36	15	
42	5	