Busn 210 Fun With Stats ➔ Chapter 3 = Mean and Standard Deviation

Average GPA
stock return

1. Means tell us:
   a. What a "typical value" is
   b. Measure of central tendency
   c. A way of "seeing all the data at once" using a typical value
   d. A typical value is useful because it can help us to make decisions

2. Standard Deviations tell us:
   a. Does the mean fairly represent the data?
   b. How clustered the data is around the mean?
   c. How reliable the mean is
   d. Conceptually, it is an average (mean) of all the deviations
      1. A deviation is simply how far a particular value is from the mean (is it below the mean or above?)

Sample Mean

For ungrouped data (data not in a frequency distribution)

$$\text{Sample Mean} = \frac{\text{Sum of all the values in the sample}}{\text{Number of values in the sample}}$$

$$\bar{X} = \frac{\Sigma X}{n}$$

Define Variables & Symbols		
$\bar{X}$	=	Sample Mean = "X bar"
n	=	Total number of observations
X	=	A particular value
Σ	=	Indicates the operation of adding = "sigma"
ΣX	=	Sum of the X values

Standard Deviation

Measures of Dispersion:
"How Big is the Variability in the Data?"

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(X-\mu)^2}{N}} \qquad s = \sqrt{s^2} = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}}$$

Define Variables & Symbols		
$\sigma = \sqrt{\sigma^2}$	=	Pop. Standard Deviation
N	=	Total number of observations
X	=	A particular value
μ	=	Pop. Mean

Define Variables & Symbols		
$s = \sqrt{s^2}$	=	Sample Standard Deviation
n	=	Total number of observations (n as denominator tends to underestimate variance)
$n-1$	=	Provides the appropriate correction for underestimation
X	=	A particular value
$\bar{X}$	=	Sample Mean

The primary use of the statistic s^2 is to estimate σ^2, therefore (n-1) is necessary to get a better representation of the deviation or dispersion in the data (n tends to underestimate)

3. Skew tells us:
    a. The relationship between:
        i. Mean (Arithmetic Mean)
        ii. Median (The one in the middle positionally)
        iii. Mode (The one that occurs most frequently)

Skewness

No Skew	
Classes (Scores)	**Frequency**
5 up to 10	6
10 up to 15	7
15 up to 20	10
20 up to 25	14
25 up to 30	26
30 up to 35	14
35 up to 40	10
40 up to 45	7
45 up to 50	6
Total	100

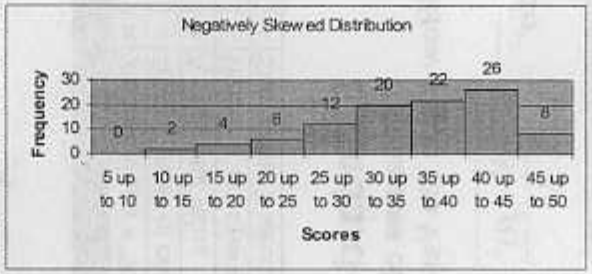
Mean	27
Median	27
Mode	27

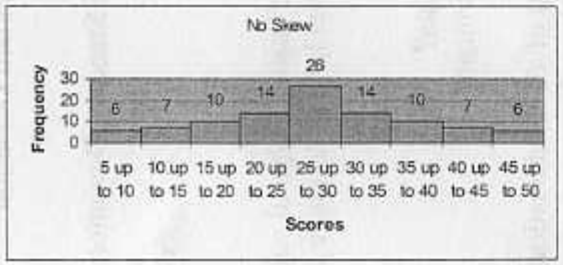
Positively Skewed Distribution	
Classes (Scores)	**Frequency**
5 up to 10	8
10 up to 15	26
15 up to 20	22
20 up to 25	20
25 up to 30	12
30 up to 35	6
35 up to 40	4
40 up to 45	2
45 up to 50	0
Total	100

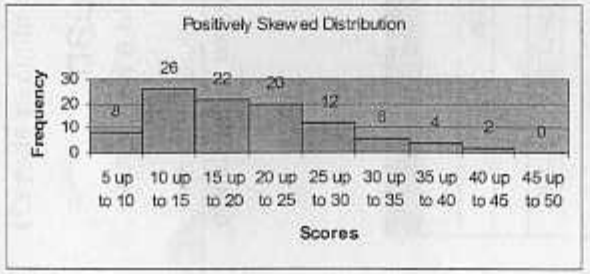
Mean	19.17
Median	18
Mode	13

Negatively Skewed Distribution	
Classes (Scores)	**Frequency**
5 up to 10	0
10 up to 15	2
15 up to 20	4
20 up to 25	6
25 up to 30	12
30 up to 35	20
35 up to 40	22
40 up to 45	26
45 up to 50	8
Total	100

Mean	34.57
Median	36
Mode	43







Here we have a sample of the ages of people that have attended the recital. We would like to know how old is a typical recital attendee (the sponsor wants to know if the recital is worth sponsoring - are the attendees an age that the sponsor can target?). Then we would like to know whether the typical age (mean) is reliable - are the ages clustered around the mean? or are they spread out? - thus we will need to calculate the standard deviation for a sample.

Ages
15
27
51
58
62
63
75
37
43

X	(X-Xbar)	(X-Xbar)^2
15		
27		
51		
58		
62		
63		
75		
37		
43		
Total		

$\bar{x} = \frac{\Sigma x}{n} = \frac{431}{9} = 47.\overline{8}$

431

① Does the mean fairly represent our data?

② How clustered is data around mean?

③ what is spread of data - what is dispersion

$n = 9$

Range = $75 - 15 = 60$

$\bar{x}$ = sample mean

n = # of items

$n-1$ = reduces underestimation

x = particular value

Σ = sigma = Add

$(x - \bar{x})$ = deviation

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

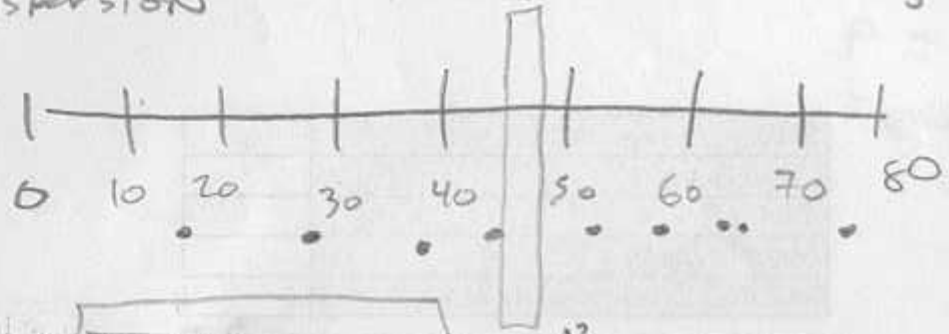
$\sqrt{s^2} = s$ = {sample standard deviation

$366.86 = s^2 = \frac{2934.89 \text{ years}^2}{8}$

$\sqrt{s^2} = 19.15$ years

SUM	
COUNT = n	
n - 1 =	
Sample Mean = Xbar =	
ROUND to hundredth (2) =	

Sample Variance = s^2 =		$YEARS^2$
Sample Standard Deviation = s =		years
AVERAGE		
VAR		
STDEV		



15 - 47.9 = -32.9	$-32.9^{\wedge 2} = 1082.41$
27 - 47.9 = -20.9	$-20.9^{\wedge 2} = 436.81$
51 - 47.9 = 3.1	$3.1^{\wedge 2} = 9.61$
58 - 47.9 = 10.1	$10.1^{\wedge 2} = 102.01$
62 - 47.9 = 14.1	$14.1^{\wedge 2} = 198.81$
63 - 47.9 = 15.1	$15.1^{\wedge 2} = 228.01$
75 - 47.9 = 27.1	$27.1^{\wedge 2} = 734.41$
37 - 47.9 = -10.9	$-10.9^{\wedge 2} = 118.81$
43 - 47.9 = -4.9	$-4.9^{\wedge 2} = 24.01$
	$s = 2934.89$ years^2

Here we have a sample of the ages of people who own iPods. We would like to know how old is a typical owner? (Apple needs to know who to target for advertising and promotion and recital sponsoring). Then we would like to know whether the typical age (mean) is reliable - are the ages clustered around the mean? or are they spread out? - thus we will need to calculate the standard deviation for a sample.

Ages
20
15
10
34
43
12
17
19
29

$\bar{X} = \frac{\Sigma X}{n} = \frac{199}{9} = 22.1$

	X	(X-Xbar)	(X-Xbar)^2
	20		
	15		
	10		
	34		
	43		
	12		
	17		
	19		
	29		
Total			

199

$n = 9$

Range =

SUM	
COUNT = n	
n - 1 =	
Sample Mean = Xbar =	
ROUND to hundredth (2) =	

Sample Variance = s^2 =		$YEARS^2$
Sample Standard Deviation = s =		years
AVERAGE		
VAR		
STDEV		

$s^2 = \frac{964.88\ years^2}{8}$

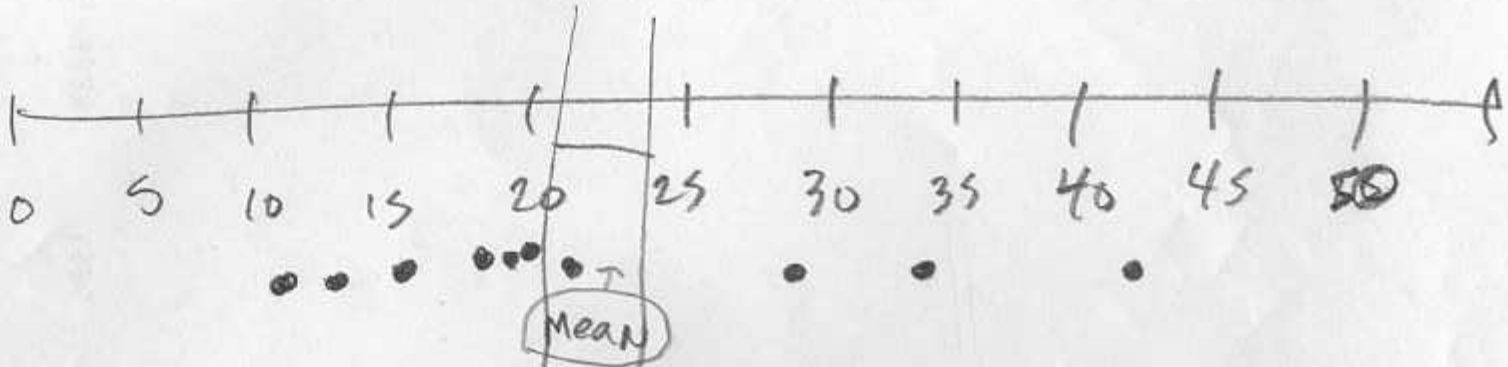
$s^2 = 120.61\ years^2$

$s = \sqrt{s^2} = \sqrt{120.61\ years^2} = 10.98\ years$

iPod user age
20
15
10
34
43
12
17
19
29

$$\bar{x} = \frac{\Sigma x}{n} = \frac{199}{9} = 22.\overline{11}$$

Does this mean fairly represent data?



Is Data clustered around mean?

Range = 43 − 10 = 33

$$s^2 = \frac{964.88 \text{ years}^2}{8} = 120.61 \text{ years}^2$$

SD

$$s = 10.98 \text{ years}$$

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

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

x = particular value

n = count

$n-1$ = correction factor for underestimation → $\frac{10}{3} = 3\frac{1}{3}$, $\frac{10}{3-1} = 5$

$\bar{x}$ = mean

s = sample standard deviation

$\sum$ = sigma = sum

iPod

From our sample data the typical age for iPod users is 22.11 years. The range is 43-10=33. The standard deviation is 10.98 years. The data seems to be somewhat spread out - not too clustered around mean. The mean perhaps is a bit above the mode because a few large values are pulling it up (skew = +). 68% of values should lie between ≈ 11 & 33 years. So maybe the mean does not fairly represent the data.

Recital

From our sample results the typical age for the recital is 47.88 years.

data not too clustered around mean.

The range is 75-15 60 years.

The standard deviation is 19.15 years

This standard deviation is large. The mean of 47.88 may not fairly represent the data. The sponsor would have to decide. 68% of values typically fall within ± 1 standard deviation. $48 \pm 19 \Rightarrow$ 29 to 67 years. If this age group is good for the sponsors target audience, then...

recital
≠
iPod

If apple wants to sponsor a recital with its target group $\bar{X}$ = mean $\approx$ 22, SD = s = 11, then Apple will not want to sponsor recital. However if Apple is trying to gain sales in a group with ages = $\bar{X} \approx 48$ and s = 19 years, then they should sponsor recital - especially if they want to gain access to the particular type of music @ recital...