#4, 11b)c), 12c)

$\bar{x} = 45.3 \rightarrow 44.80$

11.b)

	Salary ($)	# of employees
25	20-30	12
35	30-40	24
45	40-50	32 ← 50th & 51st
55	50-60	19
65	60-70	9
75	70-80	3
	80-90	0
	90-100	1

$\frac{100}{2} = 50$

∴ between 50th + 51st #

∴ median ≈ 45

12.c)

Ages	Frequency
16.5-22.5	~~IIII~~
22.5-28.5	I
28.5-34.5	II
34.5-40.5	
40.5-46.5	II
46.5-52.5	III

etc.

MDM 4UI

Data Management

Lesson 47: Standard Deviation & Variance (Measures of Spread)

Learning Goal:
Students can calculate the standard deviation and variance of various sets of data.

Oct 21-8:21 PM

INTRODUCTION:

The measures of central tendency indicate the central values of a set of data (i.e. mean, median & mode). Often you will also want to know how closely the data cluster around these centres.

The **"measures of spread"** or **"dispersion"** of a data set are quantities that indicate how closely a set of data clusters around its centre. There are several different measures associated with measure of spread.

We will be discussing five different measures of spread throughout the course. They are...

1. ***Standard Deviation
2. ***Variance
3. Interquartile Range (I.Q.R.)
4. Percentiles
5. Z-scores

STANDARD DEVIATION & VARIANCE:

A ***deviation*** is the difference between an individual value in a set of data and the mean for the data.

STANDARD DEVIATION is the square root of the mean of the squares of the deviations.

Population vs. **Sample**

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

VARIANCE is the square of the standard deviations.

Population vs. **Sample**

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

EX. 1. Use means and standard deviations to compare the distribution of heights for the two basketball teams listed in the table below.

$\mu = 165$ $\mu = 165$

FALCONS				RAVENS			
PLAYER	**Height (cm)**	$x-\mu$	$(x-\mu)^2$	**PLAYER**	**Height (cm)**	$x-\mu$	$(x-\mu)^2$
Laura	*183*	18	324	*Sam*	*166*	1	1
Jamie	*165*	0	0	*Shannon*	*163*	-2	4
Deepa	*148*	-17	289	*Tracy*	*168*	3	9
Colleen	*146*	-19	361	*Claudette*	*161*	-4	16
Ingrid	*181*	16	256	*Maria*	*165*	0	0
Justiss	*178*	13	169	*Amy*	*166*	1	1
Sheila	*154*	-11	121	*Selena*	*166*	1	1
			$\boxed{1520}$				$\boxed{32}$

$$\sigma = \sqrt{\frac{\boxed{\sum (x-\mu)^2}}{N}} = \sqrt{\frac{1520}{7}} \doteq 14.7$$

$$\sigma = \sqrt{\frac{\boxed{\sum (x-\mu)^2}}{N}} = \sqrt{\frac{32}{7}} \doteq 2.1$$

Conclusion: What does this tell us?

- Falcons' heights are more spread out (more extremes → e.g. really tall + short)
  ↳ less consistency ⇒ ∴ higher standard deviation
- Ravens' heights are clustered around the mean (less extremes)
  ↳ more consistent ⇒ ∴ smaller standard deviation

If you are working with grouped data, you can estimate the standard deviation by using the midpoint of each interval of the exact value. The following formulas represent this change...

Population vs. **Sample**

$$\sigma = \sqrt{\frac{\sum f_i (m_i - \mu)^2}{N}}$$

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

EX. 2. A group of children were asked how many hours a day they spend watching television. The table below summarizes their responses. Find the mean, standard deviation and variance for these data.

$\bar{x} = 2.72$

Number of Hours	Number of Children, f_i	m_i	$(m_i - \bar{x})$	$(m_i - \bar{x})^2$	$f_i(m_i - \bar{x})^2$
0 – 1	1	0.5	−2.22	4.9284	4.9284
1 – 2	4	1.5	−1.22	1.4884	5.9536
2 – 3	7	2.5	−0.22	0.0484	0.3388
3 – 4	3	3.5	0.78	0.6084	1.8252
4 – 5	2	4.5	1.78	3.1684	6.3368
5 – 6	1	5.5	2.78	7.7284	7.7284
	18				27.1112

$$s = \sqrt{\frac{27.1112}{18-1}}$$

$$= \sqrt{\frac{27.1112}{17}}$$

$$= 1.26$$

$\therefore$ Variance $= (1.26)^2$

$\doteq 1.5876$

Homework: p. 148 #1, 6a, 7a