Module 2 Practice problem and Homework answers

Practice problems Set A, page 1
religion: Christian, Islamic, Hindu, Buddhist, Sikh, Jewish, other religion, not religious

Answer: mode -- these data are measured at the nominal level

years of work experience in current job: 0, 1, 2, 3, 4, 5, 6, etc.

Answer: mean (without extreme values present) – these data are measured at the ratio level, so as long as there are no extreme values present, the mean is the appropriate measure of central tendency.

years of work experience in current job: 0, 1 – 3, 4 – 6, 7 – 10, 1 or more

Answer: median – these data are measured at the ordinal level

Country of origin of 20 passengers on a flight to Beijing

Answer: mode – these data are measured at the nominal level

monthly income

Answer: mean (without extreme values present) – these data are measured at the ratio level, so as long as there are no extreme values present, the mean is the appropriate measure of central tendency.

opinion scale: completely agree, somewhat agree, somewhat disagree, completely disagree

Answer: median – these data are measured at the ordinal level

Practice Problem Set A, page 2

Age

--measure of central tendency: median – Because there are extreme values in this list (the 2, 4 and 11 year olds), the median is a more appropriate measure in this case than the mean.

--value of the median: 49.5, which is the midpoint between 47 and 52

Happiness

--Measure of central tendency: median – Because these data are measured at the ordinal level, the median is the best choice

--value of the median: pretty happy – the middle two values when this variable is sorted in order are both pretty happy

Number of pets

--Measure of central tendency: mean – Because there are no extreme values in this list, the mean is the best choice.

--value of the mean: 1.4

Religion

--Measure of central tendency: mode – Because this variable is measured nominally, the mode is the only measure of central tendency that we can calculate.

--value of the mode: Christian

Practice Problem Set B, page 8

standard deviation for person B

Scores for B	Mean	Deviation	Deviation squared
4	4.5	-0.5	0.25
5	4.5	0.5	0.25
4	4.5	-0.5	0.25
5	4.5	0.5	0.25
4	4.5	-0.5	0.25
5	4.5	0.5	0.25
			$\Sigma = 1.5$

$$s = \sqrt{\frac{1.5}{6}} = 0.5$$

standard deviation for person C

Scores for C	Mean	Deviation	Deviation squared
6	4.5	1.5	2.25
3	4.5	-1.5	2.25
5	4.5	0.5	0.25
2	4.5	-2.5	6.25
9	4.5	4.5	20.25
2	4.5	-2.5	6.25
			$\Sigma = 37.5$

$$s = \sqrt{\frac{37.5}{6}} = 2.5$$

Practice Problem Set B, page 8, continued
standard deviation for person D

Scores for D	Mean	Deviation	Deviation squared
14	5.2	8.8	77.44
0	5.2	-5.2	27.04
9	5.2	3.8	14.44
0	5.2	-5.2	27.04
8	5.2	2.8	7.84
0	5.2	-5.2	27.04
			Σ = 180.8

$$s = \sqrt{\frac{180.8}{6}} = 5.5$$

Homework problems
page 1

$$mean = \frac{\sum x}{N} = \frac{125}{17} = 7.4$$

median: 6 (when the data are ordered from smallest to largest, the bottom 8 cases are lower than 6, and the top 8 cases are higher than 6)

mode: 1 is the value that occurs most frequently

The mode is most representative since half of the respondents had either 0 or 1 drinks. The median and mean represent only one respondent, who is atypical of the rest of the group.

page 2

gender

--measure of central tendency: mode – this variable is measured at the nominal level
--value of the mode: female, since there are more females than males

education

--measure of central tendency: median – this is an ordinal variable
--value of the median – high school diploma, since the middle two values when the data are sorted from low to high both have this value

political orientation

--measure of central tendency: median – this is an ordinal variable
--value of the median: moderate, since the middle two values when the data are sorted from strongly conservative to strongly liberal (or vice versa) are both moderate

political optimism scale

--measure of central tendency: mean – Because there are no extreme values in this list of interval/ratio-level data, we are safe to calculate a mean.
--$mean = \frac{\sum x}{N} = \frac{535}{20} = 26.75$

rallies attended

--measure of central tendency: median – Because of the extreme values (22, 38 & 50), the median is the most appropriate measure of central tendency
--value of the median: 3

page 3

VR for group A (mode = SUV)

$$VR = 1 - \left(\frac{5}{17}\right) = 0.71$$

VR for group B (mode = sedan)

$$VR = 1 - \left(\frac{7}{17}\right) = 0.59$$

VR for group C (mode = sedan)

$$VR = 1 - \left(\frac{4}{17}\right) = 0.76$$

Homework problems, page 3 continued
Group B has the lowest VR score, and therefore the lowest variation, and Group C has the highest VR score, and therefore the highest variation. This means that there is more similarity regarding types of car drive among group B than group C, and group A is in the middle.

page 4

Semester 1	Mean	Deviation	| Deviation |	Deviation squared
70	81	-11	11	121
83	81	2	2	4
76	81	-5	5	25
65	81	-16	16	256
100	81	19	19	361
87	81	6	6	36
74	81	-7	7	49
75	81	-6	6	36
90	81	9	9	81
70	81	-11	11	121
73	81	-8	8	64
91	81	10	10	100
71	81	-10	10	100
93	81	12	12	144
97	81	16	16	256
Σ = 1215			Σ = 148	Σ = 1754

$$\text{mean} = \frac{\sum x}{N} = \frac{1215}{15} = 81$$

median: 76: When the data are sorted from lowest to highest (or vice versa), there are 7 values below 76, and 7 values above 76.

mode: there are 2 scores of 70 in the data, while all other scores are represented only once. Therefore, 70 is the mode.

Range: The lowest value is 65, and the highest is 100. So, the range is 65 – 100.

$$\text{mean deviation} = \frac{\sum|(x-\bar{x})|}{N} = \frac{148}{15} = 9.9$$

$$\text{standard deviation} = \sqrt{\frac{\sum(x-\bar{x})^2}{N}} = \sqrt{\frac{1754}{15}} = \sqrt{116.9} = 10.8$$

Homework problems, page 4 continued

Semester 2	Mean	Deviation	| Deviation |	Deviation squared
87	89.1	-2.1	2.1	4.41
86	89.1	-3.1	3.1	9.61
90	89.1	0.9	0.9	0.81
86	89.1	-3.1	3.1	9.61
93	89.1	3.9	3.9	15.21
94	89.1	4.9	4.9	24.01
81	89.1	-8.1	8.1	65.61
90	89.1	0.9	0.9	0.81
94	89.1	4.9	4.9	24.01
91	89.1	1.9	1.9	3.61
95	89.1	5.9	5.9	34.81
78	89.1	-11.1	11.1	123.21
90	89.1	0.9	0.9	0.81
90	89.1	0.9	0.9	0.81
92	89.1	2.9	2.9	8.41
Σ = 1337	Σ = 1337		Σ = 55.5	Σ = 325.75

$$\text{mean} = \frac{\sum x}{N} = \frac{1337}{15} = 89.1$$

median: 90: When the data are sorted from lowest to highest (or vice versa), there are seven values below 90, and seven values above 90.

mode: there are four scores of 90 in the data, while all other scores are represented fewer than four times. Therefore, 90 is the mode.

Range: The lowest value is 78, and the highest is 95. So, the range is 78 - 95.

$$\text{mean deviation} = \frac{\sum|(x - \bar{x})|}{N} = \frac{55.5}{15} = 3.7$$

$$\text{standard deviation} = \sqrt{\frac{\sum(x - \bar{x})}{N}} = \sqrt{\frac{325.75}{15}} = \sqrt{21.7} = 4.7$$

The students in semester 2 had higher central tendency scores than the students in semester 1 had, so we can conclude that the Semester 2 students did better on the exam.

Homework, page 4 continued
It seems more likely that the students hoping to earn an A+ would be in favor of the method used in Semester 1 since the maximum values (the top of the range) in this semester were higher than in Semester 2.

Because the measure of dispersion (range, mean deviation, and standard deviation) decreased from Semester 1 to Semester 2, we can conclude that scores in Semester 2 were more consistent than they were in Semester 1. Therefore, the professor was successful.

page 5
The 2010 – 2011 and 2012 – 2013 seasons had snowfall amounts that fell within the 7.5 to 10.0 ideal range.

$$\text{mean} = \frac{\sum x}{N} = \frac{39.6}{5} = 7.9$$

median = 76.8, since two values are lower and two values are higher

mode: there is no mode because all 5 values are distinct.

standard deviation

snowfall amounts	Mean	Deviation	Deviation squared
6.8	7.9	-1.1	1.21
7.8	7.9	-0.1	0.01
5.9	7.9	-2.0	4.00
8.9	7.9	1.0	1.00
10.2	7.9	2.3	5.29
			Σ = 11.51

$$\text{standard deviation} = \sqrt{\frac{\sum x}{N}} = \sqrt{\frac{11.51}{5}} = \sqrt{2.3} = 1.5$$

If we replace 6.8 with 4.8, **the median won't change** because both values are below the original median. The median would change if we had replaced a value below the original median with a value above the original median, or vice versa. (Not sure? Try it out!)

However, by replacing 6.8 with 4.8, we're reducing the value of $\sum x$, which will **decrease the value of the mean**.

Finally, by replacing a value in the dataset that is closer to the original mean than the value that is replacing it, we're introducing a more extreme value, or more variation within the dataset. Because dispersion measures variation, introducing this more extreme value **increases the standard deviation**.