Statistics 7_9b_practice.pdf	Michael Hallstone, Ph.D. hallston@hawaii.edu

Lecture 7-9b: Practice Problems

Practice

Everything that appears in the lecture notes are fair game for the test. They are the best "study guide" I can provide. It is impossible to provide a "list" that is more comprehensive than the lecture notes above. However, here are a few additional practice exercises or practice concepts.

There are also other "practice problems" embedded in the body of the lecture you should do as well. Basically the answers to my examples in the lecture *can and should be used* as practice problems.

There are two practice exercises for you. One asks you to practice with MCT. The answers follow. The next asks you to compute measures of dispersion and the answers also follow.

Practice with Measures of Central Tendency (group #1)

Below is a data set with some fake variables.

Do the following with each of the four variables:

- Classify each variable as continuous vs. discrete and nominal, ordinal, interval, or ratio
- Calculate all three measures of central tendency: mean, median, mode
- Choose the best measure of central tendency for each variable. <u>You can compute these here, but don't have to as I have plenty of practice problems computing these below. If you just want to look at my answers and then choose the best measure of central tendency that is fine. The goal in this first set of problems is to choose the best measure of central tendency not computation.</u>
- Choose the best measure of dispersion for each variable. <u>Note: you do NOT have to compute them. I have done that for you.</u> <u>There are plenty of practice problems computing measures of dispersion below. The goal here is to choose the best measure of dispersion.</u>

Age: age in years

Gender: 1=male 2=female

Opinion: "I do most of the housework" 1= disagree 2= neutral 3=agree

Yearly Income: annual income in $

	Age	Gender	Opinion	yearly income
X1	21	1	1	$2,500
X2	21	1	1	$3,000
X3	21	1	1	$10,000
X4	22	1	1	$12,000
X5	22	1	1	$12,000
X6	22	1	1	$15,000
X7	22	1	2	$19,000
X8	23	1	2	$20,000
X9	23	2	2	$20,000
x10	23	2	2	$21,000
x11	23	2	2	$22,000
x 12	24	2	2	$22,000
x13	25	2	3	$24,000
x14	27	2	3	$25,000
x15	29	2	3	$32,000
x16	31	2	3	$35,000
x17	45	2	3	$35,000
Σx	424	26	33	$329,500
skewness	2.8	-0.1	0.1	-0.1
mean	24.94	1.53	1.94	$19,382.35
median	23	2	2	$20,000.00
mode	22	2	1	$12,000.00 *see note below
variance	34.8	0.3	0.7	$92,922,794.12
standard dev	5.9	0.5	0.8	$9,639.65

*A few of these variables (age, opinion, and income) are “multi-modal distributions” and have more than one mode. What you see in the table above is the computer picking out the lowest mode by default. For example for income there are actually four total modes here and this is the lowest one. The three additional modes are $20,000, $22,000, and $35,000. Also on a test, I would NEVER give you a multi-modal distribution on a test and ask you to pick out “the mode.”

Measures of Central Tendency (using SPSS)

Statistics

		age in years	gender	opinon	yearly income
N	Valid	17	17	17	17
	Missing	0	0	0	0
Mean		24.94	1.53	1.94	$19,382.35
Median		23.00	2.00	2.00	$20,000.00
Mode		22[a]	2	1[a]	$12,000[a]
Std. Deviation		5.900	.514	.827	$9,639.647
Variance		34.809	.265	.684	9.292E7
Skewness		2.778	-.130	.117	-.077
Std. Error of Skewness		.550	.550	.550	.550
Range		24	1	2	$32,500

a. Multiple modes exist. The smallest value is shown

Answers to Practice with Measures of Central Tendency (group #1)

- Classify each variable as continuous vs. discrete and nominal, ordinal, interval, or ratio

Age: could be discrete or continuous depending upon how you measure it. Theoretically it is continuous. It is ratio.

Gender: discrete and nominal

Opinion: discrete and ordinal

Income: theoretically continuous, ratio

- Choose the best measure of central tendency for each variable.

Age: skewness is too high so mean is out. Use median or mode

Gender: it is nominal so mode is only choice

Opinion: it is ordinal so mode or median. Mean is out.

Income: skewness is okay, so mean is probably best choice and mode is probably not a very good choice as it is "too low" compared to rest of the numbers. The median would be okay too.

- Choose the best measure of dispersion for each variable.

Age: any would suffice although SD is usually best. If you wanted to get technical the skewness is too high so mean is a poor measure of central tendency and the SD and variance are based upon the mean, so perhaps they are not so good. But they do have more information than the range.

Gender: it is nominal so range is only choice

Opinion: it is ordinal so range is only choice

Income: any would suffice although SD is usually best.

More practice computing measures of central tendency (group #2)

Below are some variables I took from the survey templates on homework #1. Classify each of the variables as nominal, ordinal, interval, or ratio and choose the best measure of central tendency (mean, median, and mode) and dispersion (range, variance, standard deviation).

I do not provide the skewness for these set of variables, but I do for "group #3" below.

Therefore if the variable is interval or ratio you could choose mean, median, or mode. You would not choose the mean if the skewness was larger than +2 or smaller than -2. For ratio level variables you should probably choose the standard deviation as a measure of central tendency, but the variance or even range is acceptable. But the range is a pretty limited number, so in the real world you should choose standard deviation or variance. But since it's hard to explain what the variance means in plain English, I'd choose standard deviation.

Use these tables from lecture 7-9:

Which measure of central tendency to use?

measure	**when used**	**advantages and disadvantages**
Mean	at least Interval or ratio & no extremes "no extremes" =-2<skewness<+2	incorporates all the data, use in other statistics, influenced by extremes
median	at least ordinal data	good when have extreme values
Mode	any-- only choice for nominal	quick to get

Another Table

Type of Variable		**Averages**	**Measures of Dispersion**
NOMINAL		only choice is MODE	Only choice range and it's suspect
ORDINAL		MEDIAN OR MODE	Range only choice
INTERVAL/RATIO	**extremes in data** skewness<-2 or skewness >+2	MEDIAN (or mode) **not mean!!!**	Use SD or Variance
	no extremes in data (-2<skewness<+2)	MEAN, MEDIAN, OR MODE probably should use mean	Use SD or Variance

1. I support the death penalty

Yes (1) No (0)

2. What do you think costs more? (circle only one)

The death penalty (1) "Life" in prison (40 years) (2) I think they both cost about the same (3)

3. About how much money does it cost the government to convict and execute a murder? _____ (enter exact $ amount)

4. About how much money does it cost the government to keep a murder in prison for 40 years? _____ (enter exact $ amount)

5. Prison sentences are a good way to discourage people from using illegal drugs.

Yes (1) No (0)

6. Keeping drug illegal is the best way to lower the societal costs of drug abuse.

agree (1) neutral (2) disagree (3)

7. About how much money does it cost per year to keep a drug user in prison? _____ (enter exact $ amount)

8. About how much money does Hawaii spend per year keeping non-violent drug users in prison? _____ (enter exact $ amount)

9. The United States should have universal health care insurance

Yes (1) No (0)

10. If the US had universal health care insurance, the scheme should be run by

private insurance companies (1) the government (2) it should be a mix of both government and private insurance (3)

11. How much money should health care insurance cost the average American per month? _____(enter exact $ amount. If you think it should be free put zero)

12. How much should the annual deductible be for insurance (i.e. what is the maximum "out of pocket costs" per year per person for health care?)_____ (enter exact amount. . If you think it should be free put zero)

13. All K-12 education should be privatized in the US

Yes (1) No (0)

14. In general, kindergarten teachers who work for private schools would be better teachers than kindergarten teachers who work for public schools.

agree (1) neutral (2) disagree (3)

15. How much money per year should a kindergarten teacher in Hawaii make? _____(enter exact $ amount)

16. How much money per year should a high school teacher in Hawaii make? _____(enter exact $ amount)

17. Government workers should be unionized in the US

Yes (1) No (0)

18. In general, private “government” workers would be better at their jobs than unionized government workers.

agree (1) neutral (2) disagree (3)

19. How much money per year should a firefighter in Hawaii make? _____(enter exact $ amount)

20. How much money per year should a police officer in Hawaii make? _____(enter exact $ amount)

21. All families in Hawaii should have a disaster preparedness kit.

Yes (1) No (0)

22. Hawaii will be struck by a major hurricane in the next ten years.

agree (1) neutral (2) disagree (3)

22a. How many days of food should a disaster kit have for each person in the family? _____(enter exact # of days of food per person)

23. How many days of water should a disaster kit have for each person in the family? _____(enter exact # of days of water per person)

24. What is your favorite way to surf? (circle only one)

1=Shortboard 2=Longboard 3=Bodyboard 4=Bodysurf 5=Piapo Board

25. Where do you usually surf on the island? (circle only one)

1=West Side 2=South Shore 3=East Side 4=North Shore

26. What is the biggest wave you have ever surfed? _____ (exact height of the FACE of the wave)

27. How many years have you been a surfer? _____ (enter the exact number of years)

27a.College students should be required to take statistics.

Yes (1) No (0)

28. I agree with Mark Twain– there are three kinds of lies: lies, damn lies, and statistics.

Strongly Disagree (1) Disagree (2) Neutral (3) Agree (4) Strongly Agree (5)

29. How many statistics courses should college student be required to take _____(enter exact number of courses)

30. How many math courses should college student be required to take _____(enter exact number of courses)

31. A basic knowledge of statistics is important to being an informed citizen

Disagree (1) Neutral (2) Agree (3)

32. Age (in exact years) _____

33. Gender: _____ female _____ male

34. What is your favorite shade of brown? _____dark brown _____brown _____tan (please choose one)

35. About how much money per month do you spend on transportation? _____(enter an exact estimate)

36. About how many minutes is your daily commute? _____(enter an exact estimate)

37. Are you currently an enrolled student at UHWO? _____yes _____no

38. Approximate monthly income (US dollars) $_____

Answers to "More practice computing measures of central tendency (group #2)"

1. I support the death penalty

Yes (1) No (0)

This variable is nominal thus the only choice is mode and range.

2. What do you think costs more? (circle only one)

The death penalty (1) "Life" in prison (40 years) (2) I think they both cost about the same (3)

This variable is nominal thus the only choice is mode and range.

3. About how much money does it cost the government to convict and execute a murder? _____ (enter exact $ amount)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

4. About how much money does it cost the government to keep a murder in prison for 40 years? _____ (enter exact $ amount)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how

spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

5. Prison sentences are a good way to discourage people from using illegal drugs.

Yes (1) No (0)

This variable is nominal thus the only choice is mode and range.

6. Keeping drug illegal is the best way to lower the societal costs of drug abuse.

agree (1) neutral (2) disagree (3)

This variable is ordinal thus you could choose median or mode, but not the mean. For a measure of dispersion you would need to choose range. Both mean and standard deviation and variance are not okay to choose because an ordinal variable is not a real number and you can't do math on it. The mean requires math, so does the standard deviation and variance which are both based upon the mean.

7. About how much money does it cost per year to keep a drug user in prison? _____ (enter exact $ amount)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

8. About how much money does Hawaii spend per year keeping non-violent drug users in prison? _____ (enter exact $ amount)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it

communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

9. The United States should have universal health care insurance

Yes (1) No (0)

This variable is nominal thus the only choice is mode and range.

10. If the US had universal health care insurance, the scheme should be run by

private insurance companies (1) the government (2) it should be a mix of both government and private insurance (3)

This variable is nominal thus the only choice is mode and range.

11. How much money should health care insurance cost the average American per month? _____(enter exact $ amount. If you think it should be free put zero)

This variable is ratio, thus you can choose mean, median, or mode. You would worry about skewness or extremes for the mean though. So as long as the skewness is between -2 and +2 there are no extremes and you should use the mean. For dispersion, although range, variance and standard deviation are technically okay, you should choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

12. How much should the annual deductible be for insurance (i.e. what is the maximum "out of pocket costs" per year per person for health care?)_____ (enter exact amount. . If you think it should be free put zero)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how

spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

13. All K-12 education should be privatized in the US

Yes (1) No (0)

This variable is nominal thus the only choice is mode and range.

14. In general, kindergarten teachers who work for private schools would be better teachers than kindergarten teachers who work for public schools.

agree (1) neutral (2) disagree (3)

This variable is ordinal thus you could choose median or mode, but not the mean. For a measure of dispersion you would need to choose range. Both mean and standard deviation and variance are not okay to choose because an ordinal variable is not a real number and you can't do math on it. The mean requires math, so does the standard deviation and variance which are both based upon the mean.

15. How much money per year should a kindergarten teacher in Hawaii make? _____(enter exact $ amount)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

16. How much money per year should a high school teacher in Hawaii make? _____(enter exact $ amount)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it

communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

17. Government workers should be unionized in the US

Yes (1) No (0)

This variable is nominal thus the only choice is mode and range.

18. In general, private "government" workers would be better at their jobs than unionized government workers.

agree (1) neutral (2) disagree (3)

This variable is ordinal thus you could choose median or mode, but not the mean. For a measure of dispersion you would need to choose range. Both mean and standard deviation and variance are not okay to choose because an ordinal variable is not a real number and you can't do math on it. The mean requires math, so does the standard deviation and variance which are both based upon the mean.

19. How much money per year should a firefighter in Hawaii make? _____(enter exact $ amount)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

20. How much money per year should a police officer in Hawaii make? _____(enter exact $ amount)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally**

speaking any of the three measures of central tendency would an appropriate choice." Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

21. All families in Hawaii should have a disaster preparedness kit.

Yes (1) No (0)

This variable is nominal thus the only choice is mode and range.

22. Hawaii will be struck by a major hurricane in the next ten years.

agree (1) neutral (2) disagree (3)

This variable is ordinal thus you could choose median or mode, but not the mean. For a measure of dispersion you would need to choose range. Both mean and standard deviation and variance are not okay to choose because an ordinal variable is not a real number and you can't do math on it. The mean requires math, so does the standard deviation and variance which are both based upon the mean.

22a. How many days of food should a disaster kit have for each person in the family? _____(enter exact # of days of food per person)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

23. How many days of water should a disaster kit have for each person in the family? _____(enter exact # of days of water per person)

This variable is ratio, thus you can choose mean, median, or mode. You would worry about skewness or extremes for the mean though. So as long as the skewness is between -2 and +2 there are no extremes and you should use the mean. For dispersion, although range, variance and standard deviation are technically okay, you should choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

24. What is your favorite way to surf? (circle only one)

1=Shortboard 2=Longboard 3=Bodyboard 4=Bodysurf 5=Piapo Board

This variable is nominal thus the only choice is mode and range.

25. Where do you usually surf on the island? (circle only one)

1=West Side 2=South Shore 3=East Side 4=North Shore

This variable is nominal thus the only choice is mode and range.

26. What is the biggest wave you have ever surfed? _____ (exact height of the FACE of the wave)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

27. How many years have you been a surfer? _____ (enter the exact number of years)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it

communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

27a. College students should be required to take statistics.

Yes (1) No (0)

This variable is nominal thus the only choice is mode and range.

28. I agree with Mark Twain– there are three kinds of lies: lies, damn lies, and statistics.

Strongly Disagree (1) Disagree (2) Neutral (3) Agree (4) Strongly Agree (5)

This variable is ordinal thus you could choose median or mode, but not the mean. For a measure of dispersion you would need to choose range. Both mean and standard deviation and variance are not okay to choose because an ordinal variable is not a real number and you can't do math on it. The mean requires math, so does the standard deviation and variance which are both based upon the mean.

29. How many statistics courses should college student be required to take _____(enter exact number

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

30. How many math courses should college student be required to take _____(enter exact number of courses)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are

extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

31. A basic knowledge of statistics is important to being an informed citizen

Disagree (1) Neutral (2) Agree (3)

This variable is ordinal thus you could choose median or mode, but not the mean. For a measure of dispersion you would need to choose range. Both mean and standard deviation and variance are not okay to choose because an ordinal variable is not a real number and you can't do math on it. The mean requires math, so does the standard deviation and variance which are both based upon the mean.

32. Age (in exact years) _____

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

33. Gender: _____ female _____ male

This variable is nominal thus the only choice is mode and range.

34. What is your favorite shade of brown? _____dark brown _____brown _____tan (please choose one)

This variable is nominal thus the only choice is mode and range.

35. About how much money per month do you spend on transportation? _____(enter an exact estimate)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

36. About how many minutes is your daily commute? _____(enter an exact estimate)

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally speaking any of the three measures of central tendency would an appropriate choice."** Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)
In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.
For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

37. Are you currently an enrolled student at UHWO? _____yes _____no

This variable is nominal thus the only choice is mode and range.

38. Approximate monthly income (US dollars) $_____

This variable is ratio, thus you can choose mean, median, or mode**.** You would worry about skewness or extremes for the mean though**.** So as long as the skewness is between -2 and +2 there are no extremes so you can use any or all of them, depending on your application. **For a test in this class, the correct answer would be "generally**

speaking any of the three measures of central tendency would an appropriate choice." Obviously, if there are extremes you should not use the mean either on a test or in the real world. In the real world if you had a ratio variable with extremes, you should chose the median or mode (or report either one or both if you had a reason.)

In the real world you should use the mean as most people don't remember what the median and mode are. However if you have a specific reason for using one over the other – meaning you want to choose one because you feel it communicates an important part of your presentation to you audience – that would be okay. Or you can simply report all three and allow your audience to decide which one they like.

For dispersion, range, variance and standard deviation a similar rule applies for a test - all three are technically okay. We do not worry about extremes in the data for measures of dispersion as they measure "spread-out-ed-ness." If you go too deep, you can argue that if there are extremes in the data and standard deviation is base on the mean, then it's sort of inaccurate. While technically true, in the real world we sort of ask, "well what's the alternative?" I've never seen a standard deviation based upon the median, although come to think of it I like that idea! But it would only be for statistics geeks and have limited applicability. in the real world you'd be trying to communicate how spread out the data are to a general audience who probably does not remember what the standard deviation formula means. That's why in class I repeat it means "on average how much are these data spread out from the mean." So again in the real world should mostly choose the standard deviation. The range is a pretty limited number and since it's hard to explain what the variance means in plain English, choose standard deviation.

Even more practice computing Measures of Central Tendency (group#3)

Below are more variables for you to practice computing the mean, median and mode. Compute them in the left hand column and check your answers in the right hand column. The sample size (n) is the number of data points in each example (x1, x2, x3, etc.); so for the first one # of children there are 5 data points and thus n=5.

- Compute mean, median, and mode
- Choose the best measure of central tendency. (Note that all of these variables are Ratio in scale. The skewness is provided. When the skewness is "extreme" that "throws off" the mean; thus the mean would be a poor choice. So when we have extremes in the data, either the median or the mode would be a good choice – just don't choose the mean.)
- By the way, I'm sorry about the page breaks in the middle of tables…I'm very bad at making webpages.

	# of children
x1	1
x2	1
x3	3
x4	4
x5	11
sum X=	
mean	
median	
mode	
skewness	1.7

	# of children
x1	1
x2	1
x3	3
x4	4
x5	11
sum X=	20
mean	4
median	3
mode	1

	# of arrests
x1	2
x2	2
x3	4
x4	5
x5	87
sum X=	
mean	
median	
mode	
skewness	2.2

	# of arrests
x1	2
x2	2
x3	4
x4	5
x5	87
sum X=	100
mean	20
median	4
mode	2

skewed! No mean!

	# of cars
x1	1
x2	1
x3	3
x4	4
x5	66
sum X=	
mean	
median	
mode	
skewness	2.2

	# of cars
x1	1
x2	1
x3	3
x4	4
x5	66
sum X=	75
mean	15
median	3
mode	1

skewed! No mean!

	# years at job
x1	0
x2	5
x3	14
x4	22
x5	34
sum X=	
mean	
median	
mode	
skewness	0.5

	# years at job
x1	0
x2	5
x3	14
x4	22
x5	34
sum X=	75
mean	15
median	14
mode	#N/A

	# of years
x1	3
x2	3
x3	6
x4	7
x5	9
sum X=	
mean	
median	
mode	
skewness	0.2

	# of years
x1	3
x2	3
x3	6
x4	7
x5	9
sum X=	28
mean	5.6
median	6
mode	3

	# of jobs
x1	0
x2	1
x3	1
x4	2
x5	3
sum X=	
mean	
median	
mode	
skewness	0.4

	# of jobs
x1	0
x2	1
x3	1
x4	2
x5	3
sum X=	7
mean	1.4
median	1
mode	1

	# of arrests
x1	0
x2	0
x3	1
x4	1
x5	98
sum X=	
mean	
median	
mode	
skewness	2.2

	# of arrests
x1	0
x2	0
x3	1
x4	1
x5	98
sum X=	100
mean	20
median	1
mode	Both 1 and 0 but spss shown only the smallest number (0) and that multiple modes exist
	skewed! No mean

	# convictions
x1	1
x2	2
x3	3
x4	3
x5	21
sum X=	
mean	
median	
mode	
skewness	2.2

	# convictions
x1	1
x2	2
x3	3
x4	3
x5	21
sum X=	30
mean	6
median	3
mode	3
	skewed! No mean

	# of DUI arrests
x1	1
x2	2
x3	3
x4	3
x5	11
sum X=	
mean	
median	
mode	
skewness	2.0

	# of DUI arrests
x1	1
x2	2
x3	3
x4	3
x5	11
sum X=	20
mean	4
median	3
mode	3

	# of careers
x1	2
x2	2
x3	4
sum X=	
mean	
median	
mode	
skewness	1.7

	# of careers
x1	2
x2	2
x3	4
sum X=	8
mean	2.666666667
median	2
mode	2

	# of homes
x1	1
x2	1
x3	2
sum X=	
mean	
median	
mode	
skewness	1.7

	# of homes
x1	1
x2	1
x3	2
sum X=	4
mean	1.333333333
median	1
mode	1

	income
x1	1,000
x2	1,000
x3	4,000
sum X=	
mean	
median	
mode	
skewness	1.7

	income
x1	1000
x2	1000
x3	4000
sum X=	6000
mean	2000
median	1000
mode	1000

	# of court appearances
x1	1,000
x2	1,000
x3	4,000
sum X=	
mean	
median	
mode	
skewness	1.7

	# of court appearances
x1	1000
x2	1000
x3	4000
sum X=	6000
mean	2000
median	1000
mode	1000

	# of computers
x1	2
x2	2
x3	8
sum X=	
mean	
median	
mode	
skewness	1.7

	# of computers
x1	2
x2	2
x3	8
sum X=	12
mean	4
median	2
mode	2

	temperature
x1	72
x2	72
x3	84
sum X=	
mean	
median	
mode	
skewness	1.7

	temperature
x1	72
x2	72
x3	84
sum X=	228
mean	76
median	72
mode	72

Practice Variance and Standard Deviation and Range (group#4)

Here are many variables w/ n=4 (and then N=4) to practice computing range, standard deviation, and variance. The answers follow.

The formula for the range is the same for sample and population data.

For the variance and standard deviation remember there are different formulas for sample and population data. The first section of problems has sample data and the second section of problems has population data. (Recall, the only difference between the formulas is the bottom part of the fraction – either n-1 or N. That means you do not have to do MOST of the math twice. Just divide by n-1 in the sample section and by N in the population section.)

Formulas for standard deviation and variance

$\sigma^2 = \frac{\sum (x-\mu)^2}{N}$ =population variance

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

$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$ = population standard deviation

$S = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$ = sample standard deviation

problems using sample data					
data points	# of children	mean	x-mean	(x-mean)2	
x1	2				
x2	3				
x3	3				
x4	4				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# of cars	mean	x-mean	(x-mean)2	
x1	2				
x2	2				
x3	2				
x4	4				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range

data points	# of jobs	mean	x-mean	(x-mean)2	
x1	1				
x2	3				
x3	5				

x4	7				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# of years	mean	x-mean	(x-mean)2	
x1	2				
x2	4				
x3	5				
x4	7				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# of brothers	mean	x-mean	(x-mean)2	
x1	0				
x2	1				
x3	2				
x4	4				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range

data points	# of sisters	mean	x-mean	(x-mean)2	
x1	1				
x2	1				
x3	2				
x4	4				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# of arrests	mean	x-mean	(x-mean)2	
x1	0				
x2	3				
x3	5				
x4	8				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# of felonies	mean	x-mean	(x-mean)2	
x1	3				
x2	5				
x3	6				
x4	10				

mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# of DUIs	mean	x-mean	(x-mean)2	
x1	0				
x2	1				
x3	1				
x4	10				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# prison terms	mean	x-mean	(x-mean)2	
x1	0				
x2	1				
x3	1				
x4	14				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range

data points	# prison guards	mean	x-mean	(x-mean)2	
x1	1				
x2	2				
x3	3				
x4	14				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# nite shifts	mean	x-mean	(x-mean)2	
x1	1				
x2	2				
x3	7				
x4	14				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# day shifts	mean	x-mean	(x-mean)2	
x1	0				
x2	2				
x3	6				
x4	12				

mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# swing shifts	mean	x-mean	(x-mean)2	
x1	0				
x2	2				
x3	6				
x4	12				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
data points	# days vacation	mean	x-mean	(x-mean)2	
x1	0				
x2	5				
x3	5				
x4	10				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range

data points	# of holidays	mean	x-mean	(x-mean)2	
x1	6				
x2	7				
x3	7				
x4	8				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/n-1=		variance
					standard dev
					range
population data problems					
data points	# of children	mean	x-mean	(x-mean)2	
x1	2				
x2	3				
x3	3				
x4	4				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# of cars	mean	x-mean	(x-mean)2	
x1	2				
x2	2				

x3	2				
x4	4				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range

data points	# of jobs	mean	x-mean	(x-mean)2	
x1	1				
x2	3				
x3	5				
x4	7				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# of years	mean	x-mean	(x-mean)2	
x1	2				
x2	4				
x3	5				
x4	7				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range

data points	# of brothers	mean	x-mean	(x-mean)2	
x1	0				
x2	1				
x3	2				
x4	4				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# of sisters	mean	x-mean	(x-mean)2	
x1	1				
x2	1				
x3	2				
x4	4				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# of arrests	mean	x-mean	(x-mean)2	
x1	0				
x2	3				
x3	5				
x4	8				

mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# of felonies	mean	x-mean	(x-mean)2	
x1	3				
x2	5				
x3	6				
x4	10				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# of DUIs	mean	x-mean	(x-mean)2	
x1	0				
x2	1				
x3	1				
x4	10				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range

data points	# prison terms	mean	x-mean	(x-mean)2	
x1	0				
x2	1				
x3	1				
x4	14				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# prison guards	mean	x-mean	(x-mean)2	
x1	1				
x2	2				
x3	3				
x4	14				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# nite shifts	mean	x-mean	(x-mean)2	
x1	1				
x2	2				
x3	7				
x4	14				

mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# day shifts	mean	x-mean	(x-mean)2	
x1	0				
x2	2				
x3	6				
x4	12				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# swing shifts	mean	x-mean	(x-mean)2	
x1	0				
x2	2				
x3	6				
x4	12				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range

data points	# days vacation	mean	x-mean	(x-mean)2	
x1	0				
x2	5				
x3	5				
x4	10				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range
data points	# of holidays	mean	x-mean	(x-mean)2	
x1	6				
x2	7				
x3	7				
x4	8				
mean			Σ (x-mean)2=		
			Σ (x-mean)2/N=		variance
					standard dev
					range

Answers for Variance and Standard Deviation and Range (group #4)

Recall the sample data problems come first and the population data problems come second.

problems using sample data					
data points	# of children	mean	x-mean	(x-mean)2	
x1	2	3	-1	1	
x2	3	3	0	0	
x3	3	3	0	0	
x4	4	3	1	1	
sumX=	12		Σ (x-mean)2=	2	
mean	3		Σ (x-mean)2/n-1=	0.67	variance
				0.82	standard dev
				2.00	range
data points	# of cars	mean	x-mean	(x-mean)2	
x1	2	2.5	-0.5	0.25	
x2	2	2.5	-0.5	0.25	
x3	2	2.5	-0.5	0.25	
x4	4	2.5	1.5	2.25	
sumX=	10		Σ (x-mean)2=	3	
mean	2.5		Σ (x-mean)2/n-1=	1.00	variance
				1.00	standard dev
				2.00	range

data points	# of jobs	mean	x-mean	(x-mean)2	
x1	1	4	-3	9	
x2	3	4	-1	1	
x3	5	4	1	1	
x4	7	4	3	9	
sumX=	16		Σ (x-mean)2=	20	
mean	4		Σ (x-mean)2/n-1=	6.67	variance
				2.58	standard dev
				6.00	range
data points	# of years	mean	x-mean	(x-mean)2	
x1	2	4.5	-2.5	6.25	
x2	4	4.5	-0.5	0.25	
x3	5	4.5	0.5	0.25	
x4	7	4.5	2.5	6.25	
sumX=	18		Σ (x-mean)2=	13	
mean	4.5		Σ (x-mean)2/n-1=	4.33	variance
				2.08	standard dev
				5.00	range
data points	# of brothers	mean	x-mean	(x-mean)2	
x1	0	1.75	-1.75	3.0625	
x2	1	1.75	-0.75	0.5625	
x3	2	1.75	0.25	0.0625	

x4	4	1.75	2.25	5.0625	
sumX=	7		Σ (x-mean)2=	8.75	
mean	1.75		Σ (x-mean)2/n-1=	2.92	variance
				1.71	standard dev
				4.00	range
data points	# of sisters	mean	x-mean	(x-mean)2	
x1	1	2	-1	1	
x2	1	2	-1	1	
x3	2	2	0	0	
x4	4	2	2	4	
sumX=	8		Σ (x-mean)2=	6	
mean	2		Σ (x-mean)2/n-1=	2.00	variance
				1.41	standard dev
				3.00	range
data points	# of arrests	mean	x-mean	(x-mean)2	
x1	0	4	-4	16	
x2	3	4	-1	1	
x3	5	4	1	1	
x4	8	4	4	16	
sumX=	16		Σ (x-mean)2=	34	
mean	4		Σ (x-mean)2/n-1=	11.33	variance
				3.37	standard dev

				8.00	range
data points	# of felonies	mean	x-mean	(x-mean)2	
x1	3	6	-3	9	
x2	5	6	-1	1	
x3	6	6	0	0	
x4	10	6	4	16	
sumX=	24		Σ (x-mean)2=	26	
mean	6		Σ (x-mean)2/n-1=	8.67	variance
				2.94	standard dev
				7.00	range
data points	# of DUIs	mean	x-mean	(x-mean)2	
x1	0	3	-3	9	
x2	1	3	-2	4	
x3	1	3	-2	4	
x4	10	3	7	49	
sumX=	12		Σ (x-mean)2=	66	
mean	3		Σ (x-mean)2/n-1=	22.00	variance
				4.69	standard dev
				10.00	range
data points	# prison terms	mean	x-mean	(x-mean)2	
x1	0	4	-4	16	

x2	1	4	-3	9	
x3	1	4	-3	9	
x4	14	4	10	100	
sumX=	16		Σ (x-mean)2=	134	
mean	4		Σ (x-mean)2/n-1=	44.67	variance
				6.68	standard dev
				14.00	range
data points	# prison guards	mean	x-mean	(x-mean)2	
x1	1	5	-4	16	
x2	2	5	-3	9	
x3	3	5	-2	4	
x4	14	5	9	81	
sumX=	20		Σ (x-mean)2=	110	
mean	5		Σ (x-mean)2/n-1=	36.67	variance
				6.06	standard dev
				13.00	range
data points	# nite shifts	mean	x-mean	(x-mean)2	
x1	1	6	-5	25	
x2	2	6	-4	16	
x3	7	6	1	1	
x4	14	6	8	64	
sumX=	24		Σ (x-mean)2=	106	

mean	6		Σ (x-mean)2/n-1=	35.33	variance
				5.94	standard dev
				13.00	range
data points	# day shifts	mean	x-mean	(x-mean)2	
x1	0	5	-5	25	
x2	2	5	-3	9	
x3	6	5	1	1	
x4	12	5	7	49	
sumX=	20		Σ (x-mean)2=	84	
mean	5		Σ (x-mean)2/n-1=	28.00	variance
				5.29	standard dev
				12.00	range
data points	# swing shifts	mean	x-mean	(x-mean)2	
x1	0	5	-5	25	
x2	2	5	-3	9	
x3	6	5	1	1	
x4	12	5	7	49	
sumX=	20		Σ (x-mean)2=	84	
mean	5		Σ (x-mean)2/n-1=	28.00	variance
				5.29	standard dev
				12.00	range

data points	# days vacation	mean	x-mean	(x-mean)2	
x1	0	5	-5	25	
x2	5	5	0	0	
x3	5	5	0	0	
x4	10	5	5	25	
sumX=	20		Σ (x-mean)2=	50	
mean	5		Σ (x-mean)2/n-1=	16.67	variance
				4.08	standard dev
				10.00	range
data points	# of holidays	mean	x-mean	(x-mean)2	
x1	6	7	-1	1	
x2	7	7	0	0	
x3	7	7	0	0	
x4	8	7	1	1	
sumX=	28		Σ (x-mean)2=	2	
mean	7		Σ (x-mean)2/n-1=	0.67	variance
				0.82	standard dev
				2.00	range
population data problems					
data points	# of children	mean	x-mean	(x-mean)2	
x1	2	3	-1	1	

x2	3	3	0	0	
x3	3	3	0	0	
x4	4	3	1	1	
sumX=	12		Σ (x-mean)2=	2	
mean	3		Σ (x-mean)2/N=	0.50	variance
				0.71	standard dev
				2.00	range
data points	# of cars	mean	x-mean	(x-mean)2	
x1	2	2.5	-0.5	0.25	
x2	2	2.5	-0.5	0.25	
x3	2	2.5	-0.5	0.25	
x4	4	2.5	1.5	2.25	
sumX=	10		Σ (x-mean)2=	3	
mean	2.5		Σ (x-mean)2/N=	0.75	variance
				0.87	standard dev
				2.00	range

data points	# of jobs	mean	x-mean	(x-mean)2	
x1	1	4	-3	9	
x2	3	4	-1	1	
x3	5	4	1	1	
x4	7	4	3	9	
sumX=	16		Σ (x-mean)2=	20	

mean	4		Σ (x-mean)2/N=	5.00	variance
				2.24	standard dev
				6.00	range
data points	# of years	mean	x-mean	(x-mean)2	
x1	2	4.5	-2.5	6.25	
x2	4	4.5	-0.5	0.25	
x3	5	4.5	0.5	0.25	
x4	7	4.5	2.5	6.25	
sumX=	18		Σ (x-mean)2=	13	
mean	4.5		Σ (x-mean)2/N=	3.25	variance
				1.80	standard dev
				5.00	range
data points	# of brothers	mean	x-mean	(x-mean)2	
x1	0	1.75	-1.75	3.0625	
x2	1	1.75	-0.75	0.5625	
x3	2	1.75	0.25	0.0625	
x4	4	1.75	2.25	5.0625	
sumX=	7		Σ (x-mean)2=	8.75	
mean	1.75		Σ (x-mean)2/N=	2.19	variance
				1.48	standard dev
				4.00	range

data points	# of sisters	mean	x-mean	(x-mean)2	
x1	1	2	-1	1	
x2	1	2	-1	1	
x3	2	2	0	0	
x4	4	2	2	4	
sumX=	8		Σ (x-mean)2=	6	
mean	2		Σ (x-mean)2/N=	1.50	variance
				1.22	standard dev
				3.00	range
data points	# of arrests	mean	x-mean	(x-mean)2	
x1	0	4	-4	16	
x2	3	4	-1	1	
x3	5	4	1	1	
x4	8	4	4	16	
sumX=	16		Σ (x-mean)2=	34	
mean	4		Σ (x-mean)2/N=	8.50	variance
				2.92	standard dev
				8.00	range
data points	# of felonies	mean	x-mean	(x-mean)2	
x1	3	6	-3	9	
x2	5	6	-1	1	
x3	6	6	0	0	
x4	10	6	4	16	

sumX=	24		Σ (x-mean)2=	26	
mean	6		Σ (x-mean)2/N=	6.50	variance
				2.55	standard dev
				7.00	range
data points	# of DUIs	mean	x-mean	(x-mean)2	
x1	0	3	-3	9	
x2	1	3	-2	4	
x3	1	3	-2	4	
x4	10	3	7	49	
sumX=	12		Σ (x-mean)2=	66	
mean	3		Σ (x-mean)2/N=	16.50	variance
				4.06	standard dev
				10.00	range
data points	# prison terms	mean	x-mean	(x-mean)2	
x1	0	4	-4	16	
x2	1	4	-3	9	
x3	1	4	-3	9	
x4	14	4	10	100	
sumX=	16		Σ (x-mean)2=	134	
mean	4		Σ (x-mean)2/N=	33.50	variance
				5.79	standard dev
				14.00	range

data points	# prison guards	mean	x-mean	(x-mean)2	
x1	1	5	-4	16	
x2	2	5	-3	9	
x3	3	5	-2	4	
x4	14	5	9	81	
sumX=	20		Σ (x-mean)2=	110	
mean	5		Σ (x-mean)2/N=	27.50	variance
				5.24	standard dev
				13.00	range
data points	# nite shifts	mean	x-mean	(x-mean)2	
x1	1	6	-5	25	
x2	2	6	-4	16	
x3	7	6	1	1	
x4	14	6	8	64	
sumX=	24		Σ (x-mean)2=	106	
mean	6		Σ (x-mean)2/N=	26.50	variance
				5.15	standard dev
				13.00	range
data points	# day shifts	mean	x-mean	(x-mean)2	
x1	0	5	-5	25	
x2	2	5	-3	9	
x3	6	5	1	1	
x4	12	5	7	49	

sumX=	20		Σ (x-mean)2=	84	
mean	5		Σ (x-mean)2/N=	21.00	variance
				4.58	standard dev
				12.00	range
data points	# swing shifts	mean	x-mean	(x-mean)2	
x1	0	5	-5	25	
x2	2	5	-3	9	
x3	6	5	1	1	
x4	12	5	7	49	
sumX=	20		Σ (x-mean)2=	84	
mean	5		Σ (x-mean)2/N=	21.00	variance
				4.58	standard dev
				12.00	range
data points	# days vacation	mean	x-mean	(x-mean)2	
x1	0	5	-5	25	
x2	5	5	0	0	
x3	5	5	0	0	
x4	10	5	5	25	
sumX=	20		Σ (x-mean)2=	50	
mean	5		Σ (x-mean)2/N=	12.50	variance
				3.54	standard dev
				10.00	range

data points	# of holidays	mean	x-mean	(x-mean)2	
x1	6	7	-1	1	
x2	7	7	0	0	
x3	7	7	0	0	
x4	8	7	1	1	
sumX=	28		Σ (x-mean)2=	2	
mean	7		Σ (x-mean)2/N=	0.50	variance
				0.71	standard dev
				2.00	range