

Frequency Distributions (The 3rd Homework)

Name: Kyes

1. The lowest and highest scores are given below for different sets of scores. In each case, the scores are to be grouped into class intervals. For each, give (1) the range, (2) your choice of class interval width, (3) the score limits for the lowest interval, and (4) the score limits for the highest interval (do this directly without listing any of the intervals between the lowest and the highest):

- | | | |
|----|------------|---------------------------------|
| a. | 24, 70 | 46, 3, 24-26, 69-71 |
| b. | 27, 101 | 74, 5, 25-29, 100-104 |
| c. | 56, 69 | 13, 1, 56, 69 |
| d. | 187, 821 | 634, 50, 150-199, 800-849 |
| e. | 6.3, 21.9 | 15.6, 1.0, 6.0-6.9, 21.0-21.9 |
| f. | 1.27, 3.47 | 2.20, .20, 1.20-1.39, 3.40-3.59 |
| g. | 36, 62 | 26, 2, 36-37, 62-63 |

2. For each of the following intervals, give (1) the interval width, (2) the exact limits of the interval, and (3) the score limits and exact limits of the next higher interval (assume the scores are rounded to the nearest whole number or decimal place indicated unless otherwise specified):

- | | | |
|----|------------------------------------|--|
| a. | 10-14 | 5, 9.5-14.5, 15-19, 14.5-19.5 |
| b. | 20-39 | 20, 19.5-39.5, 40-59, 39.5-59.5 |
| c. | 2.5-2.74 | .25, 2.495-2.745, 2.75-2.99, 2.745-2.995 |
| d. | 1.0-1.9 | 1, .95-1.95, 2.0-2.9, 1.95-2.95 |
| e. | 30-40 (accurate to the nearest 10) | 11 , 25-45, 50-60, 45-65 |

3. The following is the cumulative frequency distribution for 30 scores on an anxiety test.

| Anxiety | f | Cum. f | Cum. % |
|---------|-----|----------|--------|
| 79 | 1 | 30 | 100 |
| 73 | 1 | 29 | 97 |
| 70 | 1 | 28 | 93 |
| 67 | 1 | 27 | 90 |
| 66 | 1 | 26 | 87 |
| 64 | 1 | 25 | 83 |
| 63 | 2 | 24 | 80 |
| 62 | 2 | 22 | 73 |
| 61 | 3 | 20 | 67 |
| 60 | 4 | 17 | 57 |
| 59 | 2 | 13 | 43 |
| 58 | 2 | 11 | 37 |
| 57 | 2 | 9 | 30 |
| 56 | 1 | 7 | 23 |
| 55 | 1 | 6 | 20 |
| 53 | 1 | 5 | 17 |
| 52 | 1 | 4 | 13 |
| 49 | 1 | 3 | 10 |
| 45 | 1 | 2 | 7 |
| 39 | 1 | 1 | 3 |

- A. Fill in the two blank columns (round the cumulative percentages to the nearest whole number).

- B. Find the cumulative percentage for each of the following scores: 67, 57, and 49.

90%, 30%, 10%

- C. Roughly two-thirds of the cases fall at or below which score?

61

- D. One-fifth of the cases fall at or below which score?

55

- E. Between what two scores is the “middle” of this distribution?

59 & 60

4. Imagine the data below are the GPAs for a sample of 60 sophomores at your university. Prepare a relative frequency distribution (use proportions), using an interval width of .30 and .90-1.19 as the score limits for the lowest interval.

| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|
| 3.08 | 1.81 | 3.63 | 2.52 | 2.97 | 3.48 | 1.00 | 2.70 | 2.95 | 3.29 |
| 1.40 | 2.39 | 4.00 | 2.69 | 2.92 | 3.34 | 3.00 | 3.37 | 3.01 | 2.11 |
| 2.36 | 3.23 | 2.99 | 2.61 | 3.02 | 3.27 | 2.65 | 3.89 | 1.60 | 2.31 |
| 3.93 | 2.98 | 3.59 | 3.04 | 2.88 | 3.76 | 2.28 | 3.25 | 3.14 | 2.85 |
| 3.45 | 3.20 | 1.94 | 3.80 | 2.58 | 3.26 | 2.06 | 3.99 | 3.06 | 2.40 |
| 2.44 | 2.81 | 3.68 | 3.03 | 3.30 | 3.54 | 3.39 | 3.10 | 3.18 | 2.74 |
| X | X | X | X | X | X | X | X | X | X |

| Score Limits | | frequency | proportion |
|--------------|------------------|-----------|------------|
| 3.90-4.19 | /// | 3 | .05 |
| 3.60-3.89 | //// | 5 | .08 |
| 3.30-3.59 | //////// | 8 | .13 |
| 3.00-3.29 | //////////////// | 16 | .27 |
| 2.70-2.99 | //////// | 10 | .17 |
| 2.40-2.69 | //// | 7 | .12 |
| 2.10-2.39 | //// | 5 | .08 |
| 1.80-2.09 | /// | 3 | .05 |
| 1.50-1.79 | / | 1 | .02 |
| 1.20-1.49 | / | 1 | .02 |
| 0.90-1.19 | / | 1 | .02 |
| | | N=60 | |

5. Select one quantitative variable (interval or ratio scale) from a data set from Green & Salkind (2005) and create a categorical variable (nominal or ordinal). Create a frequency table for the newly created categorical variable and explain what you found. Turn in the frequency table and the explanation next week.

| days with cold symptoms (initial data set) | days with cold symptoms (data set sorted) |
|--|---|
| 12 | 20 |
| 8 | 15 |
| 14 | 15 |
| 9 | 14 |
| 13 | 13 |
| 0 | 13 |
| 15 | 13 |
| 15 | 12 |
| 10 | 12 |
| 20 | 10 |
| 12 | 10 |
| 13 | 10 |
| 8 | 10 |
| 10 | 9 |
| 0 | 9 |
| 4 | 9 |
| 9 | 8 |
| 10 | 8 |
| 10 | 8 |
| 8 | 8 |
| 6 | 7 |
| 3 | 7 |
| 5 | 6 |
| 9 | 6 |
| 0 | 5 |
| 8 | 4 |
| 7 | 3 |
| 7 | 0 |
| 6 | 0 |
| 13 | 0 |

$$\text{Range} = 20 - 0 = 20$$

$$20/10 = 2; 20/20 = 1; 20/4 = 5$$

| $i = 5$ | | <i>frequency</i> |
|---------|----------|------------------|
| 20 – 24 | / | 1 |
| 15 – 19 | // | 2 |
| 10 – 14 | //////// | 10 |

| | | |
|-------|------------|----------|
| 5 – 9 | ////////// | 12 |
| 0 – 4 | //// | 5 |
| | | $n = 30$ |

0 – 4 = Healthy

5 – 9 = Okay

10 – 14 = Sick

15 – 19 = Chronic

20 – 24 = Say Your Prayers

| Category | <i>frequency</i> | Cum <i>f</i> | Cum % |
|------------------|------------------|--------------|-------|
| Say Your Prayers | 1 | 30 | 100 |
| Chronic | 2 | 29 | 97 |
| Sick | 10 | 27 | 90 |
| Okay | 12 | 17 | 57 |
| Healthy | 5 | 5 | 17 |
| | $n = 30$ | | |

When reducing quantitative data of interval or ratio scale to nominal (categorical) or ordinal scale, you just group your data and form class intervals, then label each interval. The downside of doing this is that you lose all sense of the actual score. When we group data we lose some specificity as to where each case falls in a given range of scores. When we place a label on a range of scores, we lose further specificity since we no longer show what scores qualify for a given category. For instance, case #1 had a score of 12. When we grouped the data into class intervals this case became a member of an interval of scores ranging from 10 – 14, which has a frequency of 10 but only two of those cases—including case #1—had actual scores of 12. So, of the 10 cases which make up the interval range 10 – 14 we have no idea of the actual score of each case. Further, by naming each interval we no longer know the score range of a category. So, if we were asked to add a case to our frequency table with the same score as case #1, we would have no idea what category to place it in.

(Data from, Green & Salkind (2005), Lesson 26 Data File 1, variable “days” defined as: “Number of days with cold symptoms in the second year”)

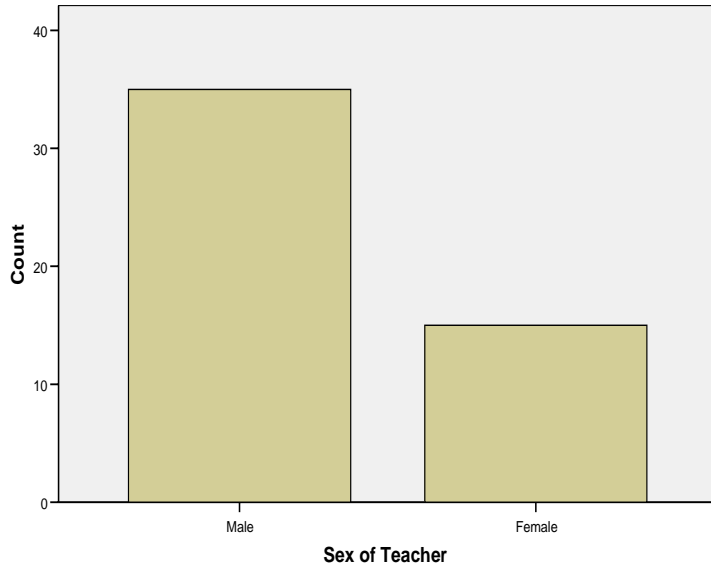
Graphing Homework (The 4th Homework) Name: Answer Key (Examples)

The goal of this assignment is to ensure you are able to create and modify graphic representation of data using SPSS by utilizing lessons 15A-17A (pp.74-94) in Green and Salkind (4th Ed.). Note to those with the newest edition (5th Ed.), these are lessons 16A-18A (pp.78-102). The lesson 16 and 17 data will be needed to create and modify graphs. **NOTE: THE LESSON 16 & 18 DATA FILE IS NOT ON THE DATA DISK!** I've created this file from past lessons to help with the learning process. **YOU WILL ONLY FIND THIS FILE ON ELECTRONIC RESERVES.**

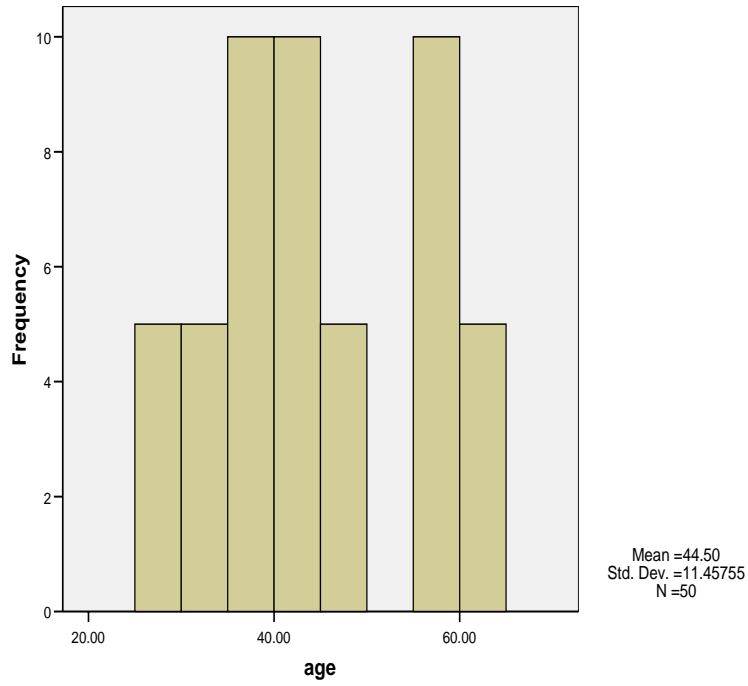
Bring a print out of a line graph, a bar graph, a histogram, a pie chart, and a boxplot. In addition, you should use the chart editor to modify the bar graph or histogram and include the revised graph as part of your output (six graphs total). Modifications should produce an APA formatted graph with an overall title and altered titles for both axes (the purpose is to demonstrate that you understand how to use the chart editor). Make sure you use the correct type of graph for the data you are displaying (i.e., no histograms for nominal data). You are welcome to try other graphing techniques if you are so inclined, I only ask that you include the six required graphs at the beginning of your homework.

Good luck and happy graphing ☺

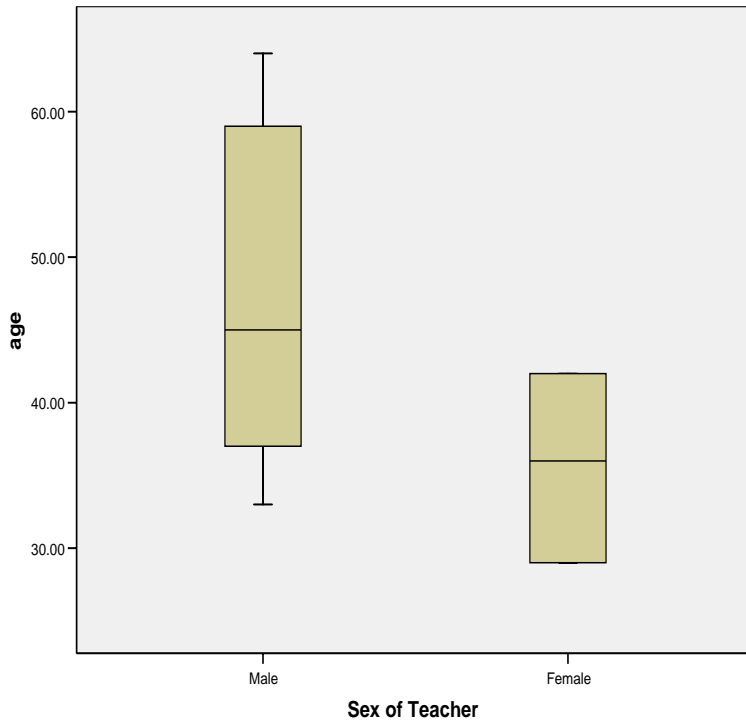
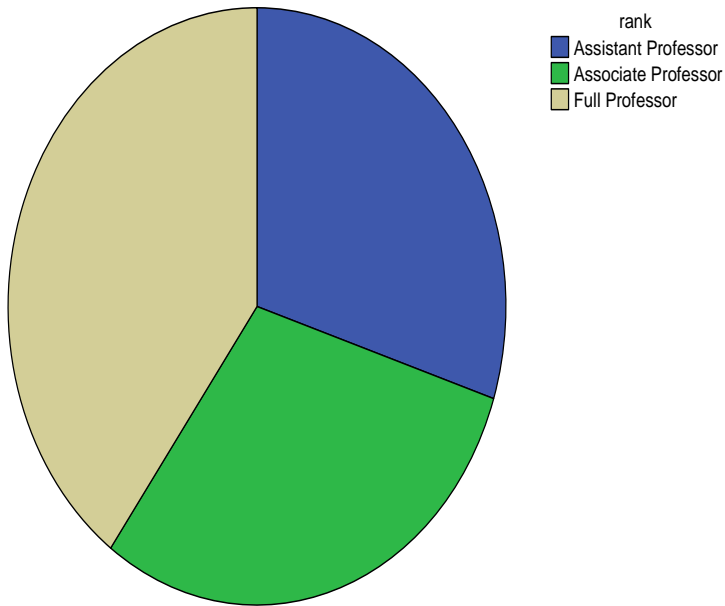
Professor Gender



Professor Age



Proffesor Rank



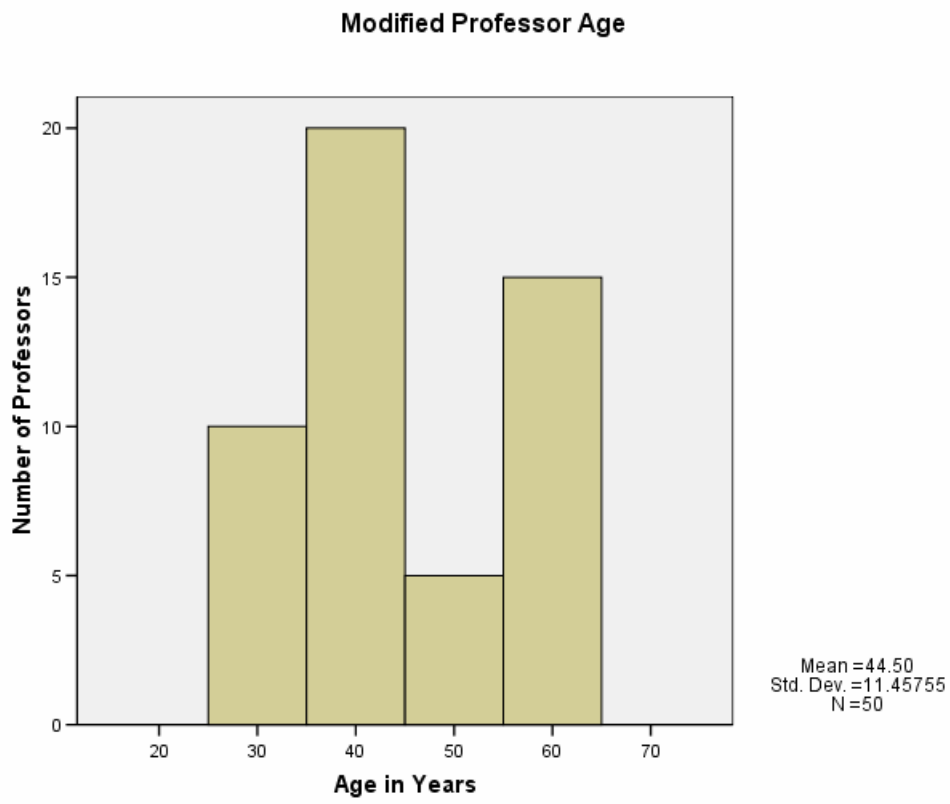


Figure 1. The frequency of professors by age.

Central Tendency Worksheet (the 5th Homework)

Name: _____ Answer Key

1. For each of the following sets of scores, find the mode, the median, and the mean:

a. 12, 10, 8, 22, 8 $Mo = 8, Mdn = 10, M = 12$

b. 14, 12, 25, 17 Mo does not exist, $Mdn = 15.5, M = 17$

c. 10, 6, 11, 15, 11, 13 $Mo = Mdn = M = 11$

2. In the following quotation, taken verbatim from a company newsletter, the author was attempting to provide statistical enlightenment:

One of the most misused words is the word “average.” It is often confused with “mean.” The difference is this: If five products sell for \$2, \$3, \$5, \$8, and \$67, the average price is \$17. The median, or mean, price is \$5, the \$5 price being the middle price—two prices are higher and two are lower. The average of a series may or may not be the middle.

Comment on the accuracy of the author’s remarks, sentence by sentence.

The first sentence is correct the word average is often misused.

The second sentence is also correct as far as it goes. The average, technically, is computed the same way as the arithmetic mean. However, because people often misuse the word average to refer to any measure of central tendency, it is important to maintain a distinction between the words “average” and “mean”.

The third sentence is also correct—except for the “difference” the author is trying to point out—as long as we assume that he meant to use the word “average” as being equal to the arithmetic mean. In this sentence, he has computed the arithmetic mean.

The fourth sentence is incorrect. In this sentence, the author equates the median with the mean. These are two separate measures of central tendency and may share the same value in certain cases—i.e. normal curve—however, in this case they do not; $Mdn = 5, M = 17$.

The fifth sentence is correct assuming that “average” = arithmetic mean and “middle” = median.

3. Comment on the probable shape of each of the following distribution:
- $M = 52, Mdn = 55, Mo = 60$ Negative Skew
 - $M = 78, Mdn = 78, Mo = 78$ Normal Curve
 - $M = 50, Mdn = 50, Mo = 60, 40$ Bi-modal Curve
 - $M = 28, Mdn = 26, Mo = 20$ Positive Skew
4. State the likely relative positions of the mean, median, and mode for the following distributions:
- Family income in a large city $Mo < Mdn < M$
 - Scores on a very easy exam $Mo > Mdn > M$
 - Heights of a large group of 25-year-old males $Mo = Mdn = M$
5. A newspaper editor once claimed that *more than half* of American families earned below average income. Could this claim possibly be correct? Explain.

Yes, providing the editor was using the word “average” to denote the arithmetic mean. Since half the scores in any distribution lay both above and below the median and income distribution curves tend to be positively skewed the arithmetic mean (“average”) would lie to the right of the median because extreme outliers in the distribution influence the arithmetic mean. The mode—scores with the greatest frequency—would lie to the left of the median. $Mo < Mdn < M$

6. Choose ONE variable in the “Teacher Scale Results” data set (in the Week TWO folder) and create a SPSS table with the mean, median, and mode and provide a description of the distribution (What kind of shape? Where is mean, median, or mode compared to the others) without looking at a graphic representation. Attach the table.

Table1: *Frequency of Teacher 1 Scores*

[DataSet1] G:\Documents\school\EMU\Kim research project\EDPS 621 files\teacher_scale_results_1.sav

Statistics

| teacher1 | | |
|----------|---------|--------|
| N | Valid | 50 |
| | Missing | 0 |
| Mean | | 3.4400 |
| Median | | 3.0000 |
| Mode | | 3.00 |

| teacher1 | | | | | |
|----------|------|-----------|---------|---------------|--------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 1.00 | 3 | 6.0 | 6.0 | 6.0 |
| | 2.00 | 9 | 18.0 | 18.0 | 24.0 |
| | 3.00 | 14 | 28.0 | 28.0 | 52.0 |
| | 4.00 | 11 | 22.0 | 22.0 | 74.0 |
| | 5.00 | 13 | 26.0 | 26.0 | 100.0 |
| Total | | 50 | 100.0 | 100.0 | |

This distribution of “Teacher 1” scores has a slight positive skew. The *Mo* and *Mdn* scores are the same (3.00) yet the *M* has a higher score (3.44). This indicates that the scores at the far right of the distribution have a greater degree of variance from the median than the scores at the far left of the distribution. This greater variance of the scores on the far right causes the higher score of the arithmetic mean, which pulls its location to the right of the median.

Variability Homework (the 6th Homework)

Name: _____ Answer Key _____

1. Compute the range, variance, and standard deviation.

3, 8, 2, 6, 0, 5

$$\text{Range} = X_{\max} - X_{\min} = 8 - 0 = 8$$

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

$$\bar{X} = (3+8+2+6+0+5)/6$$

$$\bar{X} = 24/6$$

$$\bar{X} = 4$$

$$(X - \bar{X})^2$$

$$(3-4)^2 = (-1)^2 = 1$$

$$(8-4)^2 = (4)^2 = 16$$

$$(2-4)^2 = (-2)^2 = 4$$

$$(6-4)^2 = (2)^2 = 4$$

$$(0-4)^2 = (-4)^2 = 16$$

$$(5-4)^2 = (1)^2 = 1$$

$$S^2 = \Sigma(X - \bar{X})^2/n$$

$$S^2 = (1+16+4+4+16+1)/6$$

$$S^2 = 42/6$$

$$S^2 = 7$$

$$SD = \sqrt{\Sigma(X - \bar{X})^2/n} = \sqrt{SS/n} \text{ and } S^2 = SS/n \text{ therefore } SD = \sqrt{S^2}$$

$$SD = \sqrt{S^2}$$

$$SD = \sqrt{7}$$

$$SD = 2.65$$

2. Given: $S^2 = 18$ and $SS = 900$. What is n ?

$$\begin{aligned} S^2 &= SS/n & 18 &= 900/n \\ 18 & & n &= 900/n \times n \\ 18 & & n &= 900 \\ n & & &= 900/18 \\ n & & &= 50 \end{aligned}$$

3. For each of the following statistics, what would be the effect of adding one point to every score in a distribution? What generalization do you make from this? (Do this without calculations).
- Mode – score value up one point
 - Median – score value up one point
 - Mean – score value up one point
 - Range – no change
 - Variance – no change
 - Standard Deviation – no change

When increasing all score by a constant, measures of central tendency will increase according to the value of that constant. However, there is no change in the variability of the scores.

4. If you wanted to *decrease* variance by adding a point to some (but not all) scores in a distribution, which scores would you modify? What would you do if you wanted to *increase* variance?

To decrease variance, add a point to any score below the mean. To increase variance, add a point to any score at or above the mean.

5. After you have computed the mean, median, range and standard deviation of a set of 40 scores, you discover that the lowest score is in error and should be even lower. Which of the statistics above will be affected by the correction? (Explain)

The median is not affected. Since we are not increasing the number of scores and are decreasing the value of a score already below the median, the middle score (50% of scores are above and 50% of scores are below) remains the same.

The range is affected. Since we are changing the value of X_{\min} and $\text{Range} = X_{\max} - X_{\min}$, we are affecting the calculation, which will result in an increased range value.

The mean is affected. Since we are changing the value of a score and all score values are used for the calculation of the mean, $M = \Sigma X/n$, the value of the mean will change; it will decrease.

The standard deviation is affected. Since we are changing the value of a score and all score values are used for the calculation of the standard deviation, $SD = \sqrt{\Sigma(X-\bar{X})^2/n}$, the value of the standard deviation will change; its value will increase.

6. Imagine that each of the following pairs of means and standard deviations was determined from scores on a 50-item test. With only this information, describe the probable shape of each distribution. (Assume a normal distribution unless you believe the information presented suggests otherwise.)

- | | |
|---------------------|----------------------|
| a. $M = 25, SD = 3$ | normal curve |
| b. $M = 48, SD = 4$ | negative (left) skew |
| c. $M = 50, SD = 0$ | no curve |

7. Determine the sum of squares (SS) corresponding to each of the following standard deviations ($n = 30$).

$$SD = \sqrt{\sum(X - \bar{X})^2 / n} = \sqrt{SS/n}$$

- a. 6
- $$6 = \sqrt{SS/30}$$
- $$6^2 = SS/30$$
- $$36 = SS/30$$
- $$30 \times 36 = SS/30 \times 30$$
- $$1080 = SS$$
- b. 4.5
- $$4.5 = \sqrt{SS/30}$$
- $$4.5^2 = SS/30$$
- $$20.25 = SS/30$$
- $$30 \times 20.25 = SS/30 \times 30$$
- $$607.5 = SS$$

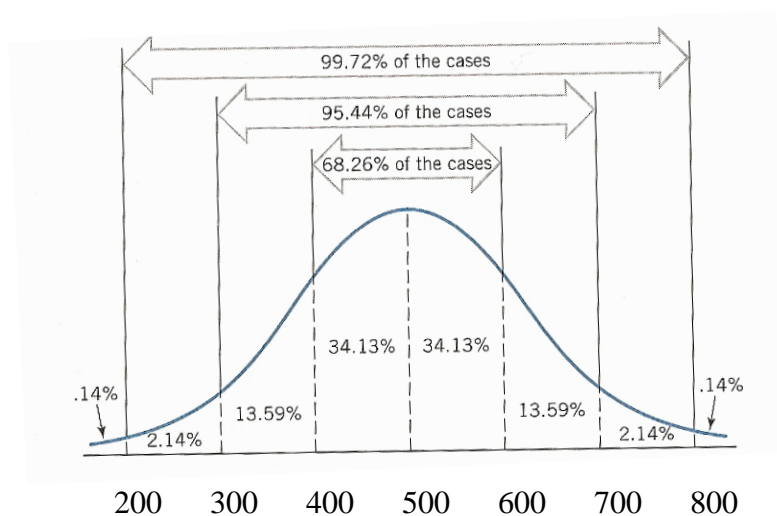
8. Given: $M = 500$ and $SD = 100$ for the SAT – V.

- a. What percentage of scores would you expect to fall between 400 and 600?

68.26%

- b. Between 200 and 800?

99.72%



Normal Distributions and Standard Scores (the 7th Homework_new)

Name: Answer Key

1. $M = 82$ and $SD = 12$ for a distribution of scores on a test of introversion-extroversion that is completed by a large group of college students (high scores are in the direction of introversion). Convert the following scores to a z score:

$$z = \frac{X - \bar{X}}{SD}$$

a. 60 $z = \frac{60 - 82}{12}$
 $z = \frac{-22}{12}$
 $z = -1.83$

b. 110 $z = \frac{110 - 82}{12}$
 $z = \frac{28}{12}$
 $z = 2.33$

2. Convert the scores in Question 1 to T scores.

$$T = 50 + 10z$$

a. 60 $z = -1.83$ $T = 50 + 10(-1.83)$
 $T = 50 + (-18.3)$
 $T = 50 - 18.3$
 $T = 31.7$
 $T = 32$

b. 110 $z = 2.33$ $T = 50 + 10(2.33)$
 $T = 50 + 23.3$
 $T = 73.3$
 $T = 73$

3. Convert the following z scores back to introversion-extroversion scores from the distribution of Question 1 (round answers to the nearest whole number)

$$z = \frac{X - \bar{X}}{SD}$$

a. 0

$$0 = \frac{X - 82}{12}$$

$$12 \times 0 = \frac{X - 82}{12} \times 12$$

$$0 = X - 82$$

$$82 + 0 = X - 82 + 82$$

$$82 = X$$

b. -.70

$$-.70 = \frac{X - 82}{12}$$

$$12 \times -.70 = \frac{X - 82}{12} \times 12$$

$$-8.4 = X - 82$$

$$82 + (-8.4) = X - 82 + 82$$

$$82 - 8.4 = X - 82 + 82$$

$$73.6 = X$$

$$74 = X$$

4. In a normal distribution, what proportion of cases fall (report to two decimal places):

a. Above $z = +1.00$? .1587
.16

b. Below $z = -2.00$? .0228
.02

5. The Maine Education Assessment (MEA), a test completed annually by all students in the state in select grades, has $M = 250$ and $SD = 50$. What MEA score separates the upper 30% of the cases from the lower 70%?

Find z using table A from the text. Use column 3 "Area Beyond z " and locate the proportion that is closest to 30% (.3000).

$$z = .52$$

$$z = X - \bar{X} / SD$$

$$.52 = X - 250 / 50$$

$$50 \times .52 = X - 250 / 50 \times 50$$

$$26 = X - 250$$

$$250 + 26 = X - 250 + 250$$

$$276 = X$$

6. Given an normal distribution with $M = 500$ and $SD = 100$, find the percentile ranks for scores of:

$$z = X - \bar{X} / SD$$

a. 400 $z = 400 - 500 / 100$
 $z = -100 / 100$
 $z = -1 = P_{16}$

b. 510 $z = 510 - 500 / 100$
 $z = 10 / 100$
 $z = .1 = P_{54}$

Normal Distributions and Standard Scores (the 7th Homework)

Name: Answer Key Class Number:

1. $M = 82$ and $SD = 12$ for a distribution of scores on a test of introversion-extroversion that is completed by a large group of college students (high scores are in the direction of introversion). Convert the following scores to a z score:

$$z = \frac{X - \bar{X}}{SD}$$

a. 60 $z = \frac{60 - 82}{12}$
 $z = \frac{-22}{12}$
 $z = -1.83$

b. 110 $z = \frac{110 - 82}{12}$
 $z = \frac{28}{12}$
 $z = 2.33$

2. Convert the scores in Question 1 to T scores.

$$T = 50 + 10z$$

a. 60 $z = -1.83$ $T = 50 + 10(-1.83)$
 $T = 50 + (-18.3)$
 $T = 50 - 18.3$
 $T = 31.7$
 $T = 31$

b. 110 $z = 2.33$ $T = 50 + 10(2.33)$
 $T = 50 + 23.3$
 $T = 73.3$
 $T = 73$

3. Convert the following z scores back to introversion-extroversion scores from the distribution of Question 1 (round answers to the nearest whole number)

$$z = \frac{X - \bar{X}}{SD}$$

a. 0

$$0 = \frac{X - 82}{12}$$

$$12 \times 0 = \frac{X - 82}{12} \times 12$$

$$0 = X - 82$$

$$82 + 0 = X - 82 + 82$$

$$82 = X$$

b. -.70

$$-.70 = \frac{X - 82}{12}$$

$$12 \times -.70 = \frac{X - 82}{12} \times 12$$

$$-8.4 = X - 82$$

$$82 + (-8.4) = X - 82 + 82$$

$$82 - 8.4 = X - 82 + 82$$

$$73.6 = X$$

$$73 = X$$

4. In a normal distribution, what proportion of cases fall (report to two decimal places):

a. Above $z = +1.00$? .1587

b. Below $z = -2.00$? .0228

5. The Maine Education Assessment (MEA), a test completed annually by all students in the state in select grades, has $M = 250$ and $SD = 50$. What MEA score separates the upper 30% of the cases from the lower 70%?

Find z using table A from the text. Use column 3 "Area Beyond z " and locate the proportion that is closest to 30% (.3000).

$$z = .52$$

$$z = X - \bar{X} / SD$$

$$.52 = X - 250 / 50$$

$$50 \times .52 = X - 250 / 50 \times 50$$

$$26 = X - 250$$

$$250 + 26 = X - 250 + 250$$

$$276 = X$$

6. Given an normal distribution with $M = 500$ and $SD = 100$, find the percentile ranks for scores of:

$$z = X - \bar{X} / SD$$

a. 400 $z = 400 - 500 / 100$
 $z = -100 / 100$
 $z = -1 = P_{16}$

b. 510 $z = 510 - 500 / 100$
 $z = 10 / 100$
 $z = .1 = P_{54}$

Normal Distributions and Standard Scores (the 7th Homework)

Name: _____ Class Number _____

1. $M = 82$ and $SD = 12$ for a distribution of scores on a test of introversion-extroversion that is completed by a large group of college students (high scores are in the direction of introversion). Convert the following scores to a z score:
 - a. 60

- b. 110

2. Convert the scores in **Question 1** to T scores.
 - a.

- b.

3. Convert the following z scores back to introversion-extroversion scores from the distribution of Question 1 (round answers to the nearest whole number)
- 0

b. -0.70

4. In a normal distribution, what proportion of cases fall (report to two decimal places):
- Above $z = +1.00$?

b. Below $z = -2.00$?

5. The Maine Education Assessment (MEA), a test completed annually by all students in the state in select grades, has $M = 250$ and $SD = 50$. What MEA score separates the upper 30% of the cases from the lower 70%?
6. Given a normal distribution with $M = 500$ and $SD = 100$, find the percentile ranks for scores of:
- 400
 - 510

Correlation (Pearson r)

Name: _____

1. Why is it important to inspect scatterplots?
2. Answer the following questions using the data below:

| X | Y |
|-----|-----|
| 11 | 12 |
| 9 | 8 |
| 8 | 10 |
| 6 | 7 |
| 4 | 4 |
| 3 | 6 |
| 1 | 2 |

- a. Draw (by hand) a scatterplot of the data.
- b. From the scatterplot, describe your impressions of the strength and direction of the association between X and Y .
- c. Do you detect any outliers or evidence of curvilinearity?
- d. Estimate the correlation between X and Y using the scatterplot.

c. Calculate the covariance of X and Y using the $X/2$ data. Show your work!

d. How is the covariance affected by this transformation?

e. Calculate the correlation of X and Y using $X/2$ data. Show your work!

f. How is the correlation affected by this transformation?

g. What generalizations do these results permit regarding the effect of linear transformations (e.g., halving each score) on the degree of linear association between two variables?

Regression

Name: _____

1. Use the following data table, documenting student performance on two quizzes, to answer the questions below.

| | Keith | Bill | Charlie | Brian | Mick |
|----------------|--------------|-------------|----------------|--------------|-------------|
| Quiz 1 (X) | 8 | 9 | 4 | 2 | 2 |
| Quiz 2 (Y) | 10 | 6 | 8 | 5 | 1 |

- a. Draw the scatter plot and estimate the least-squares regression line for prediction Y from X .
- b. Use a straight edge with the regression line to estimate (to one decimal place) the predicted Y score (Y') of each of the five individuals.
- c. Use the answers from Problem 1b to determine the error in prediction for each of the five individuals.
- d. Use the answers from Problem 1c to compute the error sum of squares.
- e. If any other line were used for prediction, how would the error sum of squares compare with your answer to Problem 1d? Why?

2. The following are the summary statistics for the scores given in Problem 1:

$$\bar{X} = 5.00, S_x = 2.97, \bar{Y} = 6.00, S_y = 3.03, r = +.62$$

- a. From these values, compute intercept and slope for the Y -on- X regression equation; state the regression equation.

- b. Obtain predicted scores (Y') for Keith, Bill, Charlie, Brian, and Mick. Compare your answers with those obtained in Problem 1b; explain and discrepancies.

- c. Compute the mean of the predicted scores and compare with the summary statistics above. What important generalization (within the limits of round error) emerges from this comparison?

- d. Compute the sum of the prediction errors for these five individuals, and state the generalization that this sum illustrates (within the limits of rounding error).

3. From the records of a large HMO, an investigator is able to gather a sizable sample of males for whom heights both at age 10 and as adults are known. The following are the summary statistics for this sample:

| Height at age 10 | Adult height |
|-------------------------|---------------------|
| $\bar{X} = 48.3$ | $\bar{Y} = 67.3$ |
| $S_x = 3.1$ | $S_y = 4.1$ |
| $r = +.71$ | |

- a. Use the values above to compute intercept and slope for predicting adult height from height at age 10 (round to the second decimal place); state the regression equation, using the form of Formula (8.1).
- b. With this regression equation, predict the adult height for the following 10-year-olds: Jean P. (42.5 in.), Albert B. (55.3 in.), and Burrhus S. (50.1 in.).
- c. Consider Jean's predicted adult height. In what sense is that value a mean? Explain.

4. Gayle falls one standard deviation above the mean of X . What is the correlation between X and Y if her predicted score on Y falls:
- One standard deviation above \bar{Y} ?
 - One-third of a standard deviation below \bar{Y} ?
 - Three quarters of a standard deviation above \bar{Y} ?
 - One-fifth of a standard deviation below \bar{Y} ?

Subject: Syllabus & Assignments for EDPS 621

From: Kyung-Hee Kim <kkim7@emich.edu>

Date: Thu, 30 Aug 2007 09:21:23 -0400 (EDT)

BCC:

Dear EDPS 621 Students,

This will be my first of many messages to you, my future students, in EDPS 621.

You will need to complete the attachment which will be the first Professionalism assignment and 10% of your Professionalism Grade. Please fill out the information in order, and email the attachment with a recent photograph of yourself (showing your face) to me by September 5 (before the class starts) (by September 10 for Monday Class). I will have your information posted as soon as possible on the E-reserve for this class.

Please Note: This information will be shared with your fellow classmates so it will be public information. I will also attach my vitae to introduce myself to you.

I will attach the second Professionalism assignment, Pre-assessment, which will be 10% of your Professionalism Grade. You will need to submit this assignment by beginning of the first class (September 5 for Wednesday Class; by September 10 for Monday Class). You do not have to study for the pre-assessment. You just have to guess if you do not know the answer. However, please do not say, "I do not know" or "I do not have any idea" You should answer the questions as best as you can so that you can receive some points.

I will attach the first Homework assignment, reading the article (Learning From Each Other), which will be 10% of your Homework Grade. You will also need to submit this assignment (a paragraph of reflection of the reading) by beginning of the first class (September 5 for Wednesday Class; by September 10 for Monday Class).

I will also attach the syllabus for my EDPS 621 class here. Please read VERY carefully and ask questions on the first class.

I look forward to this semester. I am expecting to receive great work from all of you because it is my sincerest hope that you all earn an "A" in this class.

Thank you.

Dr. Kim

| | |
|---|--|
| introduce yourself_621_Wednesday.doc | Content-Type: application/msword Content-Encoding: base64 |
|---|--|

| | |
|---|---|
| 2007-8-28-Dr. Kim_Vitae For EDPS 621.pdf | Content-Type: application/pdf Content-Encoding: base64 |
|---|---|

| | |
|------------------------------------|---|
| Pre Assessment_EDPS_621.pdf | Content-Type: application/pdf Content-Encoding: base64 |
|------------------------------------|---|

| |
|--|
| edps_621_syllabus_fall_2007.pdf |
|--|

| |
|---|
| Content-Type: application/pdf Content-Encoding: base64 |
|---|

| |
|---|
| KIM_learning from each other.pdf |
|---|

| |
|---|
| Content-Type: application/pdf Content-Encoding: base64 |
|---|

Review Worksheet (The 2nd Homework)Name: Keys

1. Solve the following:
- $9 + 2(9 - 2) \div 2^2$
- ,
- $(8 * 7)^2 \div 3 + 7$
- ,
- $(4 \div 9 + 3)^2 + (4 - 5)^2$

| | | |
|-------------------------|------------------------|--------------------------------|
| $9 + 2(9 - 2) \div 2^2$ | $(8 * 7)^2 \div 3 + 7$ | $(4 \div 9 + 3)^2 + (4 - 5)^2$ |
| $9 + 2(7) \div 2^2$ | $(56)^2 \div 3 + 7$ | $(.44 + 3)^2 + (4 - 5)^2$ |
| $9 + 2(7) \div 4$ | $3136 \div 3 + 7$ | $(3.44)^2 + (-1)^2$ |
| $9 + 14 \div 4$ | $1045.33 + 7$ | $11.864 + 1$ |
| $9 + 3.5$ | 1052.33 | 12.864 |
| 12.5 | | |

2. Does the sum of the squares equal the square of the sums (e.g.,
- $(2 + 3 + 4)^2 =$
- or
- $\neq 2^2 + 3^2 + 4^2$
-)? Explain.

No the sum of the squares does not equal the square of the sums.

| | | |
|-----------------|-------------------|------|
| $(2 + 3 + 4)^2$ | $2^2 + 3^2 + 4^2$ | |
| $(9)^2$ | $4 + 9 + 16$ | |
| 81 | \neq | 29 |

3. Compute the following:
- $\frac{1}{2} + \frac{1}{4}$
- ,
- $\frac{1}{2} * \frac{1}{4}$
- ,
- $\frac{1}{2} \div \frac{1}{4}$
- ,
- $(\frac{1}{2})^2$

| | | | |
|-----------------------------|-----------------------------|--------------------------------|-----------------------------|
| $\frac{1}{2} + \frac{1}{4}$ | $\frac{1}{2} * \frac{1}{4}$ | $\frac{1}{2} \div \frac{1}{4}$ | $(\frac{1}{2})^2$ |
| $\frac{2}{4} + \frac{1}{4}$ | $\frac{1}{8}$ | $\frac{1}{2} * \frac{4}{1}$ | $\frac{1}{2} * \frac{1}{2}$ |
| $\frac{3}{4}$ | | $\frac{4}{2}$ | $\frac{1}{4}$ |
| | | 2 | |

4. Convert the following fractions to decimals:
- $\frac{1}{4}$
- ,
- $\frac{1}{16}$
- ,
- $\frac{1}{25}$

$$\frac{1}{4} = .25 \qquad \frac{1}{16} = .0625 \qquad \frac{1}{25} = .04$$

5. Compute the square root of the following:
- $\frac{1}{25}$
- , 324, 432

$$\sqrt{\frac{1}{25}} = .2 \qquad \sqrt{324} = 18 \qquad \sqrt{432} = 20.785$$

6. Compute the following:
- $7 + (-13)$
- ,
- $(-3) * 8$
- ,
- $(-3) * (-8)$
- ,
- $(-6)^3$

$$7 + (-13) = -6 \qquad (-3) * 8 = -24 \qquad (-3) * (-8) = 24 \qquad (-6)^3 = -216$$

7. Solve the following equations:
- $7x + 12 = 33$
- ,
- $3x \div 4 = 8$
- ,
- $x - \frac{1}{4}x = 12$

| | | |
|-----------------------------|-------------------------|-----------------------------------|
| $7x + 12 = 33$ | $3x \div 4 = 8$ | $x - \frac{1}{4}x = 12$ |
| $7x + 12 + (-12) = 33 - 12$ | $3x \div 4 * 4 = 8 * 4$ | $(x - \frac{1}{4}x) * 4 = 12 * 4$ |
| $7x = 21$ | $3x = 32$ | $4x - x = 48$ |
| $7x \div 7 = 21 \div 7$ | $3x \div 3 = 32 \div 3$ | $3x = 48$ |
| $x = 3$ | $x = 10.67$ | $3x \div 3 = 48 \div 3$ |
| | | $x = 16$ |

8. Indicate which scale of measurement each of the following variables reflects:
- Ratio reaction time
 - Ordinal urbanicity (where 1 = urban, 2 = suburban, and 3 = rural: in order of how rural? Or how urban?)
 - Nominal the numbers on soccer players' jerseys
 - Interval Scholastic Assessment Test (SAT) score
 - Nominal type of psychotherapy
 - Ordinal university ranking (in terms of library holdings)
 - Ratio class size
 - Nominal religious affiliation (1 = Protestant, 2 = Catholic, 3 = Jewish, etc.)
 - Ordinal restaurant rating (* to *****)
 - Nominal astrological sign
 - Ratio miles per gallon

9. Which of the variables from question 8 are qualitative variables and which are quantitative (identify with the letter above)?

Qualitative = b, c, e, h, i, j

Quantitative = a, d, g, k

f = could be qualitative (1 to 4 ranks) or quantitative (5 or more ranks)

10. For the three questions that follow, illustrate your reasoning with a variable from the list in question 8 (use the letters above to illustrate).

- a. Can a ratio variable be reduced to an ordinal variable?

Yes. Variable "a" could be reduced to a simple ranking: 1st, 2nd, 3rd, etc.

- b. Can an ordinal variable be promoted to a ratio variable?

No. Variable "i" cannot become ratio because it lacks a real zero and equal intervals.

- c. Can an ordinal variable be reduced to a nominal variable?

Yes. Variable "f" can be reduced to nominal (dummy coding: below average: 0; above average: 1).

11. Round the following numbers as specified:

- a. To the nearest whole number: 8.545, -43.2, 123.04, .095

8.545 = 9 -43.2 = -43 123.04 = 123 .095 = 0

- b. To the nearest tenth: 27.33, 1.9288, -.38, 4.9746

27.33 = 27.3 1.9288 = 1.9 -.38 = -.4 4.9746 = 5

- c. To the nearest hundredth: -31.519, 76.0048, .82951, 40.7442

$$-31.519 = -31.52$$

$$76.0048 = 76$$

$$.82951 = .83$$

$$40.7442 = 40.74$$