Homework #2: Descriptive Statistics and Distributions

Homework Hints: Please round all of your final answers to 2 decimals (e.g., 1.75).

1. Understanding measures of *central tendency*
    a. Describe and explain the difference between the mean, median, and mode.
    b. Generate your own example (neither from your book nor from lecture) in which the median would be the preferred measure of central tendency.

2. Understanding measures of *dispersion*
    a. Describe the concepts of the *variance* and *standard deviation*.
    b. Explain why the standard deviation is more often used as a descriptive statistic than the variance; why might the standard deviation be more useful?

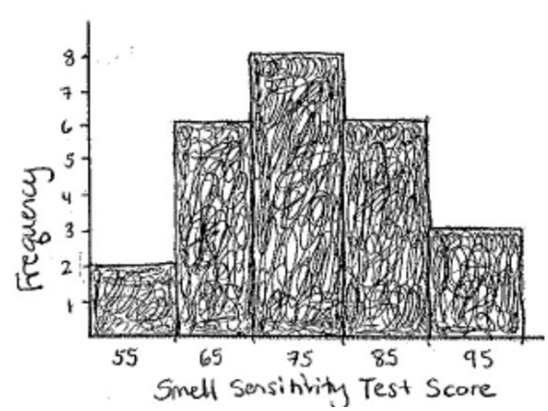
3. A psychologist interested in political behavior measured the square footage of the desks in the offices of four U.S. Governors and four chief executive officers (CEOs) of major U.S. corporations. The measurements for the governors were 44, 36, 52, and 40 square feet. The measures for the CEOs were 32, 60, 48, and 36 square feet.
    a. Calculate the *sample* (not estimated for the unknown population) means, variance, and descriptive standard deviations for both sets of scores.
    b. Verbalize to yourself, in *conceptual* terms rather than computational terms, what you have just done. (this is just a mental exercise)
    c. If you want to do a BONUS item, calculate the estimated population variance for the governor's desks.

4. You calculate the variance of a distribution of scores to be -4.26. Why must this value be incorrect?

5. The sum of individual deviations from the sample mean is equal to what value?

6. Provided below are the scores on a test of sensitivity to smell taken by 25 chefs attending a national conference: 96, 83, 59, 64, 73, 74, 80, 68, 87, 67, 64, 92, 76, 71, 68, 50, 85, 75, 81, 70, 76, 91, 69, 83, 75

    a. The average/mean of the data is about 75 points. Describe the general shape the histogram (based on grouped scores) below; this description should also provide information about the mode.



    b. If the shape of the distribution was skewed negatively, what would it look like? Would most chefs' sensitivities be less than or greater than the average sensitivity?

Homework #2:

Answers

1. **Understanding measures of central tendency**
   - **a.** The mean is the average of a set of numbers; which represents the sum of the scores divided by the total number of scores. The mean is the most commonly used measure of central tendency because it accounts for every score in the distribution and does not very much from sample to sample. Specifically, the mean is less sensitive to sampling variation than are other measures of central tendency. The median represents the middle score of a distribution; the median is easy to identify when scores are organized from lowest to highest. The median is also sometimes used in lieu of the mean when outlying scores change the shape of the distribution to become skewed either positively or negatively. In such cases, the mean of the sample will not be the best measure of central tendency to describe the center of the distribution. Finally, the mode is the most commonly occurring value in a distribution; the mode is often used to analyze variables with a nominal scale of measurement.

   - **b.** One example of when the median would be the preferred measure of central tendency to represent a data set of equal-interval variables with an outlying score. A data set best represented by the median would be the following set of scores for the shoe sizes of 9-year-old boys: 5, 6, 6, 7, 7, 7, 7, 7, 8, and 15.

2. **Understanding measures of dispersion**

   - **a.** The sample *variance* represents the average squared deviation of scores from the mean. The sample variance is calculated by squaring each score's individual deviation from the mean, adding these squared values, and dividing by the total number of scores. $SD^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n} \; or \; \frac{SS}{n}$. By extension, the standard deviation represents the average deviation of scores from the mean and is calculated by taking the square root of the variance, $SD = \sqrt{SD^2}$. When estimating population variance from a sample, use n-1.

   - **b.** The *standard deviation* is more often used as a descriptive statistic than the variance because the standard deviation offers a more direct representation of the deviations from the mean; the deviations are not squared. In other words, when people talk about data, they talk about actual measured units (rather than squared units), the standard deviation is common parlance.

3. **Governor's Desks vs. CEOs' Desks**

a. The mean for governors' desks $\overline{X} = (44 + 36 + 52 + 40)/4 = 43$ square feet
The mean for CEOs' desks: $\overline{X} = (32 + 60 + 48 + 36)/4 = 44$ square feet

Depending on how you want to think of variance:

$$\text{sample variance} = SD^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n} = \frac{SS}{n}$$

Depending on how you want to think of the standard deviation:

$$\text{sample standard dev} = SD = \sqrt{SD^2} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}} = \sqrt{\frac{SS}{n}}$$

The variance and standard deviation for governors' desks:

$$SS = (44 - 43 = 1)^2 + (-7)^2 + (9)^2 + (-3)^2 = 140$$
$$SD^2 = 140/4 = 35$$
$$SD = \sqrt{35} = 5.92$$

The variance and standard deviation for CEOs' desks:

$$SS = (32 - 44 = -12)^2 + (16)^2 + (4)^2 + (-8)^2 = 480$$
$$SD^2 = 480/4 = 120$$
$$SD = \sqrt{120} = 10.95$$

b. Although the means for the two samples are similar (e.g., 43 vs. 44 square feet), the standard deviation for the CEOs' desks is nearly twice as much as for the governors' desks. This means that there is a broader range of desk sizes among CEOs than among governors. In other words, governors' desk sizes are more similar to each other than CEOs' desks are similar to each other. This could potentially indicate that there is a greater range of incomes among CEOs than among governors, due to larger slush funds, etc.

c. BONUS: If you wanted to calculate the estimated population variance, you would use n-1. Using the governors' desks:

estimated pop. variance/standard deviation

$$S^2 = SS / \text{n} - 1$$
$$S^2 = 140/3 = 46.67$$
$$S = \sqrt{46.67} = 6.83$$

4. **Calculate a Negative Variance?**

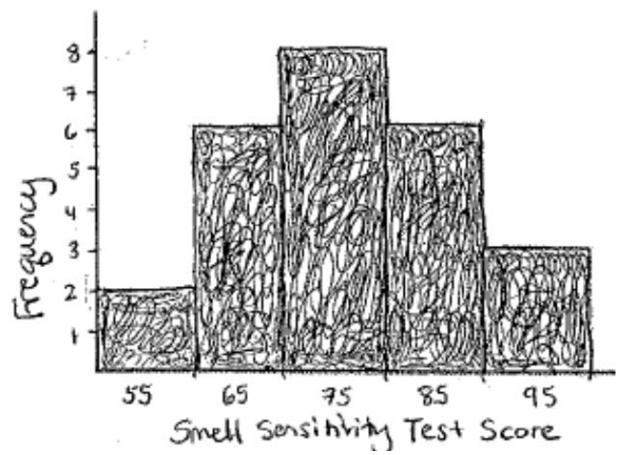
Calculating a negative variance measure must have occurred in error because the variance represents the average SQUARED deviation from the mean. All squared values have to be positive.

5. **Deviations**

The sum of all individual deviations from the mean is $\sum_{i=1}^{n}\left(X_i - \overline{X}\right) = 0$. The mean serves as a fulcrum of the distribution that splits the distribution into two equal parts based on the magnitude of the scores and on the variability of the scores in the sample.

6. **Sensitivity to Smell**

   a. This grouped (grouped in 10-point bins) histogram is *unimodal* in shape. There is only *one* mode, or most frequently occurring, score. In this instance, you can think of the class interval with midpoint 75 as the mode. The distribution is not exactly symmetrical, but is fairly symmetrical, because the shape of the distribution to the right and the left of the middle bar almost mirror each other.



   b. If the smell sensitivity data happened to produce a graph with a distribution that was skewed negatively, it would not look so symmetrical as it currently does. Rather, most of the scores would be on the high (right) end of the distribution and fewer scores would be on the low (left) side. Because the extreme low scores pull the mean of negative distributions to the left (negatively), most of the chef's would have higher sensitivities than the average of all chefs' sensitivities.