Introduction to Statistics for Psychologists
Claremont McKenna College
Professor Cook

Descriptive Statistics: Central Tendency & Variability

1. Although both frequency and probability distributions are quite useful, more succinct methods of summarizing the outcomes of simple experiments are needed. In general, distributions are depicted well by two classes of measurement: **central tendency** & **dispersion** (i.e., variability/deviance).

   (a) Central tendency

      i. The **mode** describes the most frequently occurring score in a distribution of scores. Often symbolized as *Mo.*

      ii. The **median** describes the score about which 50% of the scores fall above and below. Symbolized as *Mdn*. The middle score or the midpoint between two middle scores. How is the median affected by extreme scores?

      iii. The **mean** describes the average of all the scores in your distribution. Symbolized as $\overline{X}$ (or less commonly, M). The formula is $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{N} = \frac{\sum X_i}{N}$, where X represents the measurement of some property (e.g., a participant's reaction time). How is the mean affected by extreme scores?

      iv. Physical analogies of the mean (as a segue into dispersion): $\sum_{i=1}^{n}\left(X_i - \overline{X}\right) = 0$

(b) Central tendency measures help describe scores that best represent a sample or a population; scores most central to the entire distribution. However, all scores help define the shape of a distribution. The **shape of a distribution** of scores will change depending on whether most scores are clustered near its center, in the negative direction, or in the positive direction. As such, skewed distributions will affect central-tendency measures as well as the preference for using.

i. **Symmetrical** distributions

ii. **Negatively-skewed** distributions $\left(Sk < 0\right)$

iii. **Positively-skewed** distributions $\left(Sk > 0\right)$

2. Measures of central tendency depict one aspect of a distribution (*viz.,* location). That value, however, is relatively uninformative about the distribution of scores unless it is supplemented with a measure of **dispersion** (i.e., how variable or "spread out" the scores are in a distribution). There are several ways to numerically represent dispersion, including:

(a) **Range**: *noninclusive* $\left(R_{ni} = X_{hi} - X_{lo}\right)$

(b) **Semi-interquartile range (SIR)** $\left(Q = \dfrac{Q_3 - Q_1}{2}\right)$

- Because the semi-interquartile range depends only on the 50% of the scores nearest the mean, the SIR is not affected much by outliers, or extreme scores.

(c) **Variance** is a measure of the **average squared deviations** about the mean.

$$\textit{Unbiased Sample}:\ SD^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{N-1} \ \ or\ \ \frac{SS}{N-1};$$

SS = Sum of Squared deviations OR Sum of Squares

$$\textit{Population}:\ \sigma^2 = \frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{N}$$

Although the *variance* is not as typically used as a descriptive statistic (whereas the *standard deviation* is), the variance is nevertheless used extensively in various statistical analyses. You must know it.

Because individual scores are deviations from the mean, sometime you might see the formula written in shorthand as $\frac{\sum d_i^2}{N-1}$ where *d* stands for *deviations.*

(d) **Standard deviation** is a measure of average variation; the square root of the average squared deviation (or variance).

$$\textit{Unbiased Sample}:\ SD = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{N-1}} \ \ or\ \ \sqrt{\frac{SS}{N-1}}$$

$$\textit{Population}:\ \sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{N}}$$

3. Although the median *could* be used as the measure of central tendency, in practice the median has historically been reserved for skewed distributions.

   (a) Medians and Average Deviations (AD) - $\left(A.D. = \frac{\sum_{i=1}^{n} | x_i - Mdn |}{N} \right)$

   (b) The mean and the variance are only two summary measures that characterize a distribution; other *moments* describe additional aspects of a distribution. We can understand the shape of distributions by considering the four moments of distributions

      i. First moment, ***mean***, measure of center

      ii. Second moment, ***variance***, measure of dispersion

      iii. Third moment, ***skewness,*** represents a measure of a distribution's symmetry. Asymmetrical distributions are skewed either positively or negatively:

      iv. Fourth moment, ***kurtosis***, represents a measure of a distribution's central peak and the fullness of its tails. **Leptokurtic** distributions have a positive peakedness, whereas **platykurtic** distributions have a negative peakedness.

Key terms, Concepts, Important Formula, and Study Tips

Measures of central tendency (mean, median, mode), Measures of dispersion/variability (variance, standard deviation), Samples statistics versus population parameters, Shapes of distributions (mesokurtic, platykurtic, leptokurtic, skewed positively, skewed negatively)

$$\left(R_{ni} = X_{hi} - X_{lo}\right) \qquad \Sigma\left(x_i - \overline{X}\right) = 0 \qquad \frac{\Sigma d_i^2}{N-1} \qquad \frac{\Sigma\left(X_i - \overline{X}\right)^2}{N-1} \qquad \frac{SS}{N-1}$$

$$\frac{\Sigma\left(X-\mu\right)^2}{N} \qquad \sqrt{\frac{\Sigma\left(X-\overline{X}\right)^2}{N-1}} \qquad \sqrt{\frac{SS}{N-1}} \qquad \sqrt{\frac{\Sigma\left(X-\mu\right)^2}{N}}$$

After reading this Topic, you should be able to:

1. explain the concept of central tendency
2. identify and calculate the three measures of central tendency
3. mathematically explain why the mean is analogous to a fulcrum
4. determine what measure of central tendency is best for describing different samples of data
5. locate the three measures of central tendency on symmetrical distribution
6. locate the three measures of central tendency on skewed distributions
7. explain why/when you should prefer one measure of central-tendency over another
8. explain the concept of dispersion; identify the 3 most commonly used measures of dispersion
9. explain the concept of variance
10. explain the concept of sums of squares
11. explain what the variance tells you about a distribution of scores and how that information is different from what a mean tells you about that same distribution
12. calculate the population variance when given sums of squares (definitional formula)
13. calculate the population standard deviation when given a set of scores from a population (definitional formula)
14. explain the relationship between the variance and the standard deviation
15. calculate the sample variance when given a sample of scores (definitional formula)
16. calculate the sample variance if you are given the sums of squares from a sample of scores
17. calculate the sample standard deviation if given a set of scores (definitional formula)
18. calculate the sample standard deviation if given the sums of squares from a sample of scores
19. explain the meaningfulness of the mean for nominal, ordinal, interval, and ratio data
20. identify the formulae for calculations of a sample's mean, variance, and standard deviation
21. identify the shapes of distributions
22. provide an example of rectangular, symmetrical, positively-skewed, and negatively-skewed distributions
23. identify a data set as either positively or negative skewed
24. identify a distribution as bimodal or multimodal
25. provide examples of symmetrical and skewed distributions
26. create and label a bar graph if given some sample means
27. understand that a "floor effect" describes a situation when most scores in a distribution are near the lower end (floor) of the distribution with fewer scores at the top end (ceiling); this results in positive skew
28. understand that a "ceiling effect" describes a situation when most scores in a distribution are near the higher end (ceiling) of the distribution with fewer scores at the bottom end (floor); this results in negative skew

Homework #2

Due: See Syllabus

See Homework Packet

Notes