

Chapter 3 – Measurement and Descriptive Statistics

Chapter Outline

- I. Frequency Distributions
 - A. Definition: tally of the number of times each score on a single variable occurs.
 - B. Approximately normally distributed: there is a small number of scores for the low and high values and most of the scores occur in the middle values (distribution exhibits a “normal curve”).
 - C. Not normally distributed: distribution does not exhibit a normal curve.
 - 1. Negatively skewed: tail of the curve (extreme scores) is elongated on the low end (left side).
 - 2. Positively skewed: tail of the curve (extreme scores) is elongated on the high end (right side).
- II. Levels of Measurement
 - A. Measurement: the assignment of numbers or symbols to different characteristics (values) of the variables.
 - B. Nominal Variables: numerals assigned to each category stand for a name of category.
 - 1. Categories have no implied order or value.
 - 2. Categories are distinct and non-overlapping.
 - 3. Other terms for nominal variables:
 - a. Categorical
 - b. Qualitative
 - c. Discrete
 - C. Dichotomous Variables: have only two levels or categories.
 - 1. May or may not have an implied order
 - 2. Other terms for dichotomous variables:
 - a. dummy variables
 - b. discrete variables
 - c. categorical variables
 - D. Ordinal Variables: mutually exclusive categories that are ordered from low to high, but the intervals between categories may not be equal.
 - 1. Also includes ordered variables with only a few categories (2-4)
 - 2. Distribution of the scores is not normally distributed.
 - 3. Other terms for ordinal variables:
 - a. Ranks
 - b. Categorical
 - E. Approximately Normal (or Scale) Variables: levels or scores are ordered from low to high and the frequencies of the scores are approximately normally distributed.
 - 1. May be continuous (have an infinite number of possible values within a range).
 - 2. If not continuous, should have at least five ordered values or levels.
 - 3. Other terms for approximately normal variables:

- a. interval – have ordered categories that are equally spaced
- b. ratio – have ordered categories that are equally spaced and have a true zero
- c. continuous
- d. dimensional
- e. quantitative

F. How to Distinguish Between the Types of Measurement

- 1. Nominal versus ordinal variables:
 - a. Only two levels = treat as nominal in SPSS
 - b. Three or more categories and not ordered = nominal
 - c. Three or more categories and ordered = ordinal
- 2. Ordinal versus normal (scale) variables:
 - a. Five or more ordered levels with equal intervals and approximately normal distribution = normal
 - b. Three or more ordered levels with unequal intervals and not normally distributed = ordinal

III. Descriptive Statistics

- A. Frequency Tables: tabulates the number of occurrences of each level of a variable as well as the number of missing values; also calculates the valid percent and cumulative percent for each level.
 - 1. Nominal data: order of categories in table is arbitrary; cumulative percent column is not useful
 - 2. Ordinal or approximately normal data: order of categories in tables is shown from low to high; cumulative percent column is useful.
- B. Bar Charts: creates discrete (not connected) columns to illustrate the frequency distribution; appropriate for nominal data.
- C. Histograms: similar to a bar chart, but there are no spaces between the bars which indicates a continuous variable underlying the scores.
- D. Frequency Polygons: connects points between the categories; best used with approximately normal data (but can be used with ordinal data).
- E. Box and Whiskers Plot: useful for ordinal and normal data; gives a graphical representation of the distribution of scores.
 - 1. Box: middle 50% of cases (those between the 25th and 75th percentiles)
 - 2. Whiskers: represent the expected range of scores.
 - 3. Outliers: scores that fall outside the box and whiskers.
- F. Measures of Central Tendency
 - 1. Mean: the arithmetic average; statistic of choice for normally distributed data.
 - 2. Median: the middle score; appropriate measure for ordinal data or data that is skewed.
 - 3. Mode: the most common category; can be used with any type of data, but is the least precise information about central tendency.
- G. Measures of Variability: tells about the spread or dispersion of scores.

1. Range: highest score minus the lowest score; does not give an indication of spread of scores for ordered data.
2. Standard Deviation: most common measure of variability; based upon the deviation of each score from the mean of all scores; most appropriate for normally distributed data.
3. Interquartile Range: the distance between the 25th and 75th percentiles (as shown in the box plot); appropriate for ordinal data.
4. Nominal Data: variability measures are not appropriate; rather look at the number of categories and the frequency counts.

H. Conclusions About Measurement and the Use of Statistics

1. Normal data: utilize means and standard deviations for parametric statistics.
2. Ordinal data: utilize median and nonparametric tests.
3. Nominal data: utilize mode or count.

IV. The Normal Curve

A. Properties of the Normal Curve: the normal curve is theoretically formed by counting an “infinite” number of occurrences of a variable.

1. Unimodal – the distribution has one hump which is in the middle of the distribution.
2. The mean, median and mode are equal.
3. The curve is symmetric (not skewed).
4. The range is infinite (the extremes never touch the X axis).
5. The curve is not too peaked or too flat and is neither too short nor too long (does not exhibit kurtosis).

B. Non-Normally Shaped Distributions

1. Skewness: one tail of the frequency distribution is longer than the other.
2. Mean and median are different.

C. Kurtosis

1. Refers to the shape of the curve.
2. Leptokurtic (positive kurtosis): frequency distribution is more peaked than normal.
3. Platykurtic (negative kurtosis): frequency distribution is flatter than normal.

D. Area Under the Normal Curve (Figure 3.10)

1. The normal curve is a probability distribution whose area is equal to 1.0 and portions of the curve are fractions of 1.0.
2. Areas of the curves can be divided in terms of standard deviations.
 - a. 34% of area under the normal curve is between the mean and 1 standard deviation above or below the mean (thus, 68% of the area under the normal curve is within 1 standard deviation to the left and right of the mean).
 - b. 13.5% of the area under the normal curve is accounted for by adding a second standard deviation to the first (thus,

95% of the area under the normal is within 2 standard deviations to the left and right of the mean).

- c. 5% of the area under the normal curve falls beyond 2 standard deviations to the left and right of the mean (thus, this is why values not falling within 2 standard deviations of the mean are seen as relatively rare events).

E. The Standard Normal Curve

1. A normal curve converted so the mean is equal to 0 and the standard deviation is equal to 1.
2. This conversion allows comparison of normal curves with different means and standard deviations.
3. z scores = units of the standard normal distribution
 - a. standard scores = term for raw scores that are converted to the standard normal curve.