

PSY 216 Exam 1 Review

Statistics

- summarize large sets of data,
- determine if a relation exists between two or more variables, and
- infer if differences exist between different conditions in an experiment

Experiments are a type of research design in which all variables except the independent and dependent variables are held constant. People are randomly assigned to the various conditions. Experiments allow statements of causality to be made.

- independent variable (IV) -- variable manipulated by the researcher
- dependent variable (DV) -- variable measured by the researcher

Quasi-experiments assign people to conditions based on a subject-variable. No statements of causality.

- subject variable -- a characteristic of a person that is used to assign people to conditions

Population is the entire group of people or things that the researcher wants his/her results to apply to. Sample is a subset of the population; the sample is used because the population is too large to study.

- Sample should be random and large for it to represent the population

Several types of statistics

- Descriptive statistics
 - Measures of central tendency tell us about the average or typical value
 - Measures of dispersion tell us how similar the scores are to each other
 - Measures of relation tell us if two or more variables are related to each other
- Inferential statistics tell us if one group in a study is likely to be different from another

The level of measurement of a variable is the characteristics of the variable which determine which statistical tests can be performed on the variable

- Nominal -- categories in no order; can only compare to see if values are equal
- Ordinal -- categories in a particular order; can compare to see if values are equal and if one value is larger than another
- Interval -- categories in a particular order with a unit of measure; can compare to see if values are equal, if one value is larger than another, and perform addition and subtraction
- Ratio -- categories in a particular order with a unit of measure and a true zero point; can compare to see if values are equal, if one value is larger than another, and perform addition, subtraction, multiplication, and division

Discrete (discontinuous) variables can take on only a limited set of values, typically whole numbers; continuous variables can take on any value

- Continuous variables cannot be measured precisely; true limits: observation $\pm \frac{1}{2}$ X unit of measure

X and Y represent variables; X_i represents a particular observation of the variable X; ΣX is the sum of all the observations of variable X

Exploratory Data Analysis (EDA) allows you to

- find possible incorrect observations
- determine if the data are reasonable
- see trends in the data

EDA starts by sorting the data and then applying a graphical or tabular organization technique

- Stem and leaf plots separate each observation into two parts:
 - Stem is the most significant digits; want approximately 10 unique stems
 - Leaf is the single digit after the stem; each leaf is listed next to its stem
- Frequency distribution -- table showing the number of observations in each class
 - Class -- group of scores that are conceptually similar to each other
- Cumulative frequency distribution -- table showing the number of observations in a given class and all lower classes
 - Useful for calculating percentile rank -- the percentage of scores at or below a given score

Normal curve (Gaussian, bell-shaped) is the shape of the distribution of many psychological variables

- Distributions show frequency (or percentage or proportion) of scores on the Y axis and the variable of interest on the X axis
- Mean at the peak; ± 1 standard deviations at the inflection points; symmetrical about the mean

Skewed distribution have one tail (end of the distribution) that is longer than the other

- Positively skewed -- long tail points to the + end of the X axis
- Negatively skewed -- long tail points to the - end of the X axis

Kurtosis is related to the spread (dispersion) of the distribution relative to a normal distribution

- Mesokurtic -- no deviation from normality
- Leptokurtic -- less dispersion than normal
- Platykurtic -- more dispersion than normal

Measures of central tendency tell you about the average or typical value of a data set

- Mode -- score that occurs most frequently; bimodal -- two modes (or values that occur significantly more than the others); multimodal -- more than two modes
 - Use only with nominally scaled data as it is insensitive to changes in the data
- Median -- score that is in the middle of the data set; 50th percentile
 - Sort the data
 - Find the middle score using $(N + 1) / 2$ where N is the number of points
 - If N is odd, then the score at position $(N + 1) / 2$ is the median
 - If N is even, then take the average of the score just above the score at position $(N + 1) / 2$ and the score just below position $(N + 1) / 2$
 - Use with ordinal scaled data or interval or ratio scaled data that is highly skewed
- Mean -- the arithmetic average of the data ($\Sigma X / N$); the number, m, that makes $\Sigma(X - m) = 0$; the number, m, that makes $\Sigma(X - m)^2$ a minimum
 - Use with interval or ratio scaled data that is not highly skewed; most psychologists would use it with non-skewed ordinal scaled data too.
 - Preferred measure of central tendency as it is sensitive to every score in the data set

Non-skewed, normal distribution: mean = median = mode; positively skewed distributions usually mean > median; negatively skewed distributions usually mean < median

Measures of dispersion tell us how similar the scores are to each other

- Range -- difference between the largest and smallest scores
 - Used with ordinal scaled data or when audience is not technically sophisticated
 - It is insensitive to changes in most of the data
- Semi-interquartile range -- one half the difference between the third and first quartiles = $(75^{\text{th}} \text{ percentile} - 25^{\text{th}} \text{ percentile}) / 2$
 - interquartile range = difference between third and first quartiles
 - Used when data are highly skewed
- Standard deviation (σ) -- the mean of the squared deviations from the mean
 - Variance = standard deviation² = σ^2
 - Preferred measure of dispersion as it is sensitive to every score

s^3 is a measure of skewness

- $s^3 < 0$ = negative skew
- $s^3 > 0$ = positive skew
- $s^3 = 0$ = no skew

s^4 is a measure of kurtosis

- $s^4 > 3$ = leptokurtic
- $s^4 = 3$ = mesokurtic
- $s^4 < 3$ = platykurtic