

Review for Midterm Exam

1. What kinds of questions can we answer with statistical methods? (empirical)
2. Define
 - a. Empirical – questions based on observable data
 - b. non-empirical – questions based on personal opinion
3. What are the **Six Steps of the Research Process**?
 - a. Identify the problem
 - b. Do a literature review to formulate hypothesis
 - c. Design the research study (subjects, measurements, procedures, etc.)
 - d. Carry out the study and collect the data
 - e. Analyze the data
 - f. Interpret the data
4. Define Descriptive Statistics. (Descriptive statistics describe characteristics of the group or sample)
5. Define Inferential statistics (sample statistics are used to draw inferences about the population)
6. What is the best way of assuring that a sample is representative of the population and why? (Random sampling - because each member of the population has an equal and independent chance of being included in the sample).
7. How do we choose a random sample? (use a table of random numbers or a statistical program)
8. State and describe three types of variables
 - a. Dependent variable – outcome variable
 - b. Independent variable- variable being manipulated
 - c. Control variable – variable being measured because it may also have an interfering or intervening effect on the dependent variable if not accounted for.
9. Name and describe the four scales of measurement.
 - a. Nominal – Categorical variables (gender, ethnicity, research condition – numbers represent group membership)
 - b. Ordinal – ranks (numbers indicate relative rank but not amount of difference, i.e. percentiles)
 - c. Interval – equal units of measurement (numbers represent amount of difference, zero point is arbitrary, i.e. – Fahrenheit or Celsius scale)
 - d. Ratio – equal units of measurement and an absolute zero (Kelvin scale, height and weight).
10. How do we begin to organize and make sense of our data (use of group frequency distributions and graphs).
11. What is a Histogram – a histogram is a bar graph in which the frequency is graphed on the ordinate and the class interval is graphed on the abscissa.
12. What is a Frequency Polygon – a line graph where the mid point of each class interval is graphed along the abscissa and the frequency on the ordinate.
13. Stem and Leaf Plot – a way of visually representing the shape of the distribution by charting the tens in the Stem and the ones in the leaf of the plot.

14. Ogive – An ogive is a cumulative frequency line graph. You chart the cumulative percent on the ordinate and the scores on the abscissa.
15. Describe the following types of Distributions
 - a. Normal Distribution – symmetrical, two halves are mirror images
 - b. Negatively Skewed Distribution – There are more scores in the high range and fewer in the low range, the tail points towards negative infinity
 - c. Positively Skewed Distribution – There are more lower scores and fewer higher scores, the tail points towards positive infinity
16. What are the three characteristics that describe a distribution?
 - a. Shape
 - (1) Symmetric– normal,
 - (2) Skewed - negatively skewed, positively skewed
 - (3) Modality – number of relative peaks it exhibits
 - (4) Kurtosis – amount of peakedness (flat peaks – platykurtic; very peaked – leptokurtic; medium peaks – mesokurtic).
 - b. Central Tendency
 - c. Dispersion or Variability
17. Identify and Describe the Measures of Central Tendency
 - a. Mean – arithmetic average
 - b. Median – Counting average or the 50th percentile or the point below which 50% of the scores fall
 - c. Mode – most frequently occurring score
18. What are the characteristics of the mean?
 - a. most stable from sample to sample
 - b. affected by all scores in the distribution
 - c. requires interval level or higher data
 - d. adding or subtracting a constant to/from each score increases/decreases the mean by the same amount
 - e. most useful for inferential statistics
 - f. the sum of the deviation scores about the mean equals zero
19. What are characteristics of the median
 - a. Useful for ordinal data or higher (use for percentile ranks)
 - b. Adding or subtracting a constant from each score increases or decreases the median by the constant.
 - c. Less influenced by extreme scores, therefore more useful than the mean for skewed distributions.
20. Describe characteristics of the mode
 - a. useful for categorical data only
 - b. is highly effected by extreme scores
21. What is the relationship of the three measures of central tendency in a symmetrical distribution? (They will all have the same value)
22. What is the relationship between the three measures of central tendency in a Positive skewed distribution? (The mean will be larger than the median and the mode will have the lowest value - $\bar{X} - M = +.$)

23. What is the relationship between the three measures of central tendency in a Negatively skewed distribution? (The Median will be larger than the mean and the mode will be the largest value - $\bar{X} - M = -$.)
24. How do we calculate the mean from a Frequency distribution?

$$\frac{\bar{X}_1(F_1) + \bar{X}_2(F_2) + \bar{X}_3(F_3)}{F_1 + F_2 + F_3}$$
25. How do we calculate the mean of Combined Groups having different N's? (We take the mean of means or in other words we use the weighted Mean Formula -
$$\frac{\bar{X}_1(N_1) + \bar{X}_2(N_2) + \bar{X}_3(N_3)}{N_1 + N_2 + N_3}$$
26. What does Variability tell us? (How different the scores are from one another.)
27. What are the measures of Variability?
- Range – the highest score minus the lowest score
 - Variance - $S^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$ or $S^2 = \frac{\sum \chi^2}{N - 1}$
 - Standard Deviation $S = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$ or $S = \sqrt{\frac{\sum \chi^2}{N - 1}}$
28. What is the standard deviation? (The standard deviation is the average distance from the mean).
29. How do you interpret the standard deviation? (The larger the standard deviation the more spread out the scores are).
30. What are the properties of the Normal Distribution?
- Unimodal
 - Symmetrical – equally balanced, mirror images on both sides
 - Asymptotic – the tales extend to infinity and never touch the abscissa.
 - Mean, median and mode are all the same value
 - 68% of the area under a normal curve is within ± 1 standard deviation from the mean
 - 95% of the area is within ± 2 standard deviations from the mean
 - 99.9% of the area is within ± 3 standard deviations from the mean.
31. What are Z scores? (Z scores are standard scores that tell you in standard deviation terms how far a score is from the mean)

$$Z = \frac{X - \bar{X}}{S}$$
32. What is the mean of a distribution of Z scores? (The mean of a distribution of Z scores is always 0.)
33. What is the formula for finding the raw score when the Z score is known?

$$X = ZS + \bar{X} .$$
34. What is the correlation between two variables called? (Bivariate Correlation)
35. What is a Positive Correlation? (A positive correlation is also known as a Direct correlation and describes the relationship between two variables where High values of one variable are associated with High values of the other)

variable or low values of one variable are associated with low values of the other variable.)

36. What is a Negative Correlation (A negative correlation is also known as an indirect correlation and describes the relationship between two variables where high values of one variable are associated with low values of another variable.)
37. How do we measure correlation statistically (We compute the Pearson Correlation Coefficient = r_{xy}).
38. What is the formula for the Covariance?
$$Cov_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N - 1} \text{ or } \frac{\sum xy}{N - 1}$$
39. What does the Covariance represent? (This represents Average cross-product deviation between X and Y.
40. How do we compute the r_{xy} ?

Deviation Formula

$$r_{xy} = \frac{Cov_{xy}}{S_x S_y} \text{ or } \frac{\frac{\sum xy}{N - 1}}{S_x S_y} \text{ or } \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N - 1}$$

Raw Score Formula p.81 in Salkind

$$r_{xy} = \frac{N(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

41. What is the range of the Pearson Product Moment Correlation Coefficient? (The correlation ranges from -1 to +1).
42. What does ± 1 represent? (± 1 represents a perfect positive or negative correlation).
43. What does a correlation coefficient of 0 represent? (A 0 correlation coefficient represents the absence of relationship between two variables).
44. In order to compute the r_{xy} what must the relationship between the two variables be? (The relationship must be a linear relationship which means that a scatterplot drawn of the two variables should approximate a straight line.)
45. What is the Coefficient of Determination? (the percentage of variance in one variable that is accounted for by the variance in the other variable = r^2). P. 89 Salkind
46. What is the Coefficient of Alienation? (The coefficient of alienation also called the coefficient of non-determination refers to the amount of unexplained variance or the variance in one variable that is not accounted for by the variance in the other variable = $1 - r^2$.) P. 89 Salkind
47. What does the term Monotonic mean? (This describes a relationship that is not strictly linear but at the same time is not curvilinear it may plateau off at some point or the rate of change may not be consistent but it does not reverse directions).

48. What statistic should be used to compute the correlation between two variables that are monotonically related? (The Spearman Rank Correlation Coefficient).
49. How do you compute the Spearman Rank Correlation Coefficient?
 - a. Convert the data to ranks
 - b. Compute the Pearson Correlation Coefficient on the ranks.