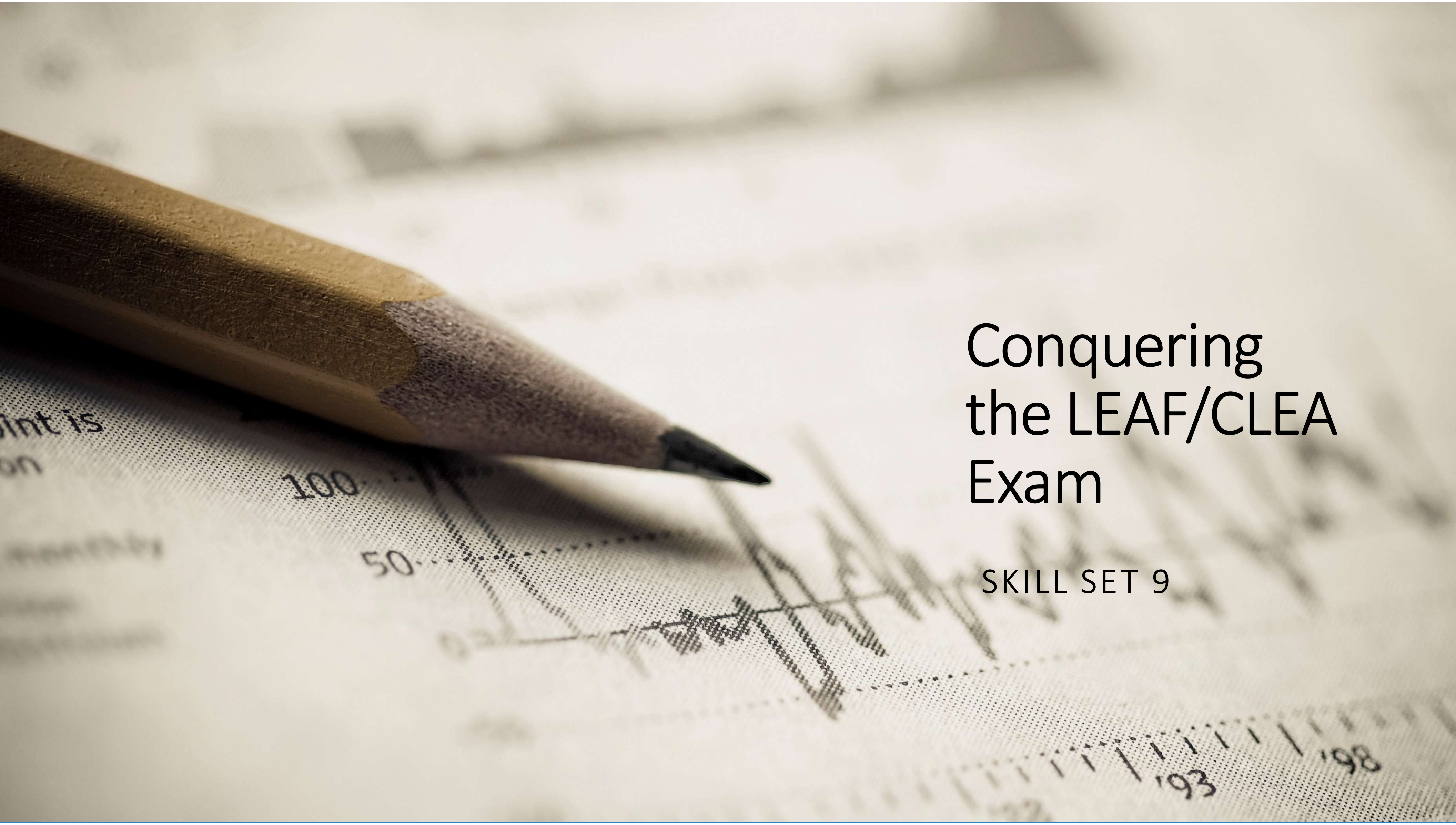


Conquering the LEAF/CLEA Exam

SKILL SET 9

About the Instructor/Course

- Instructor – Jenny Zawitz Jennifer.Zawitz@gmail.com
- CLEA Study Guide: https://iaca.net/wp-content/uploads/2021/06/CLEA-Skill-Sets_Study-Resources-051821.pdf
- LEAF Study Guide: https://iaca.net/wp-content/uploads/2021/06/en_LEAF-Core-Competencies_Study-Resources.pdf
- Exploring Crime Analysis: Readings on Essential Skills (3rd Edition) - IACA
- Each month will cover a different section of the study guide
- Intended as a supplement NOT a substitute for the texts and the Essential Skills classes
    - This course will help you focus your studying, but the courses and text will provide the actual understanding you need to pass the tests

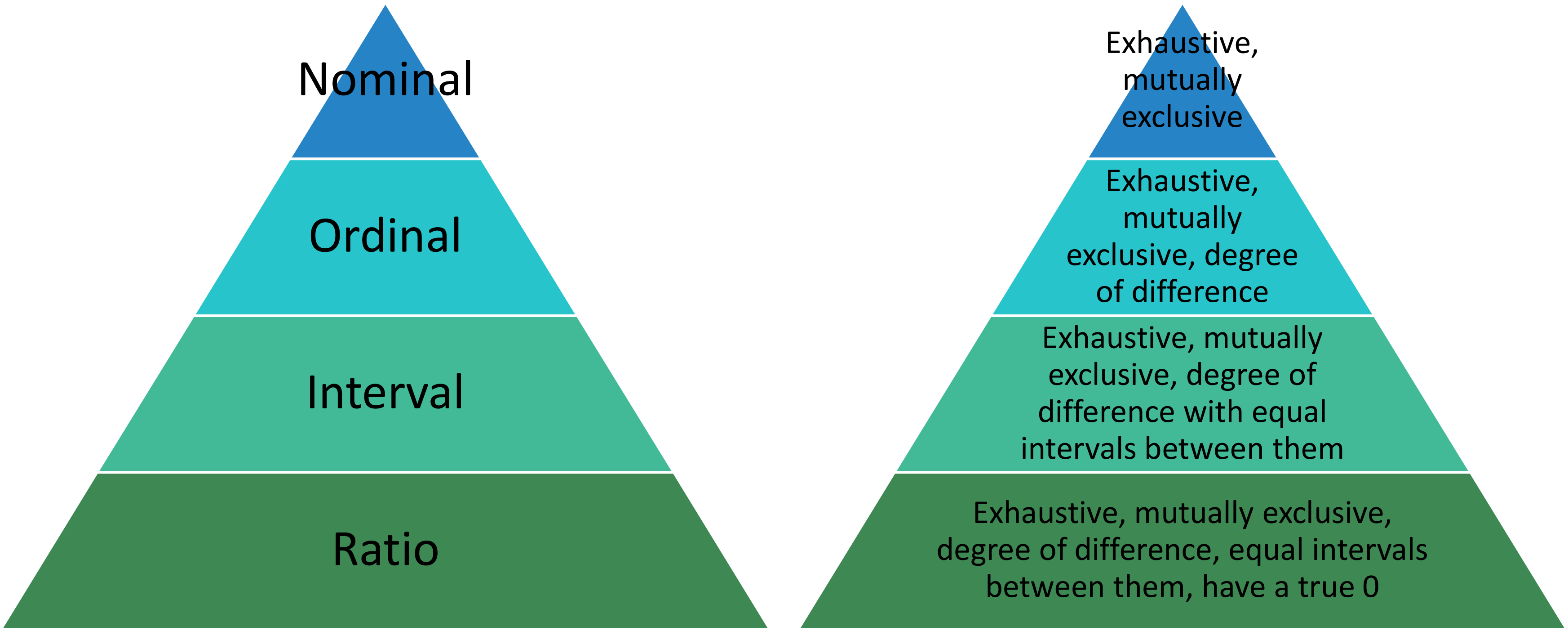
Refresh: Descriptive Statistics

SKILL SET 8, CHAPTER 8

Statistics General

- Science of collecting and organizing data and then drawing conclusions based on that data
- Three types
- Descriptive – summarize large amounts of information in an efficient and easily understood manner
  - Ex: Measures of central tendency – mean, median, mode
- Multivariate – allow comparisons among factors by isolating the effect of one factor or variable from others that may distort conclusions
  - Ex: Regression, correlation analysis
- Inferential – suggest statements about a population based on a sample drawn from the population (covered in July)
  - Statistical significance, t-test, p-value

Levels of Measurement

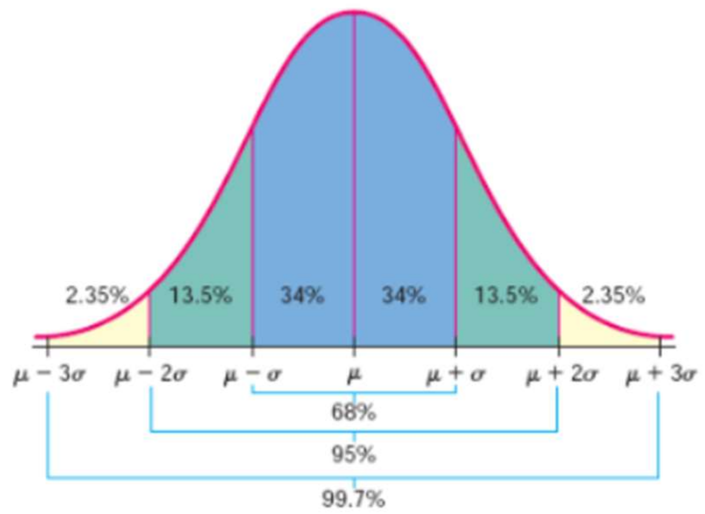


Measures of Central Tendency

- Mean: Average. Distribution of data can only have one mean.
  - Distribution impacted by outliers or extreme scores.
  - Useful with interval and ratio data
- Median: Midpoint or middle score of the distribution.
  - Point where 50% of scores are above the median and 50% are below.
  - Strength in that it is not impacted by extreme scores
  - If odd number of cases, rank order the scores and determine the middle case
    - (n + 1) / 2 with n being the total number of cases
  - If even number of cases, average the two cases at the midpoint.
- Mode: most frequent score
  - Can be unimodal (one mode) or multimodal (more than one mode)
  - Best used with nominal data.

Normal Distribution and Skewness

- Unimodal distribution can be normal, positive, or negative
- Normal distribution (bell shaped curve) means that all scores are evenly distributed throughout the distribution. No skewness or extreme scores.
  - Standard Deviation: measures the average distance that each data item is away from the mean of all the data in the distribution. Attempt to create a curve from the data. Demonstrates how scores compare with each other and comparison between two distributions.
  - Approximately 68% of all cases are within 1 SD, 95% are within 2 SD, and 99.7 are within 3 SD.
- Skewness: shows the spread of scores weighted to one side of the mean (positive or negative)
  - The farther apart the measures, the more skewed the distribution



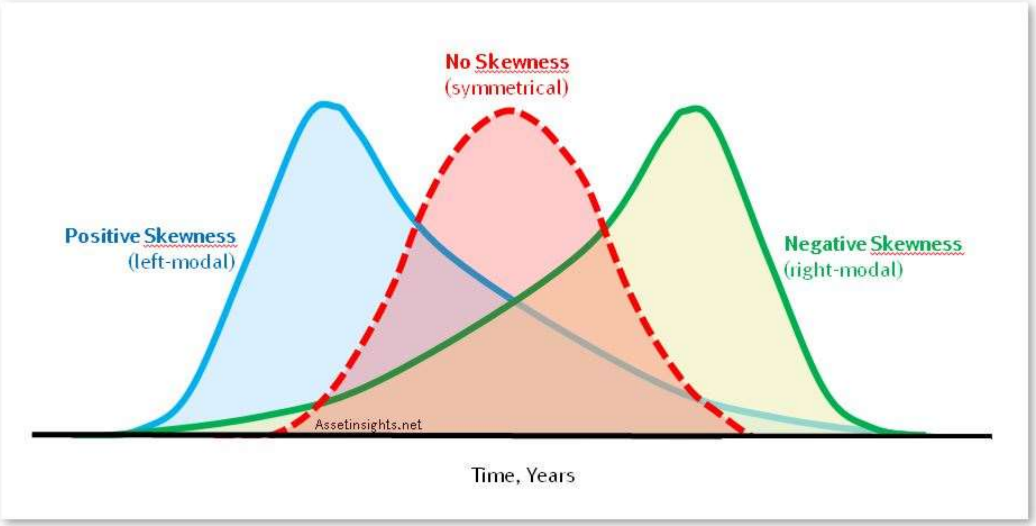
Skewness

POSITIVELY SKEWED

- Scores weighted to the left (hump located to the left and the tail moves right)
- Right tail is longer because a few scores have values much higher than the rest

NEGATIVELY SKEWED

- Scores weighted to the right (hump located to the right and the tail moves left)
- Left tail is longer because a few scores have values much lower than the rest

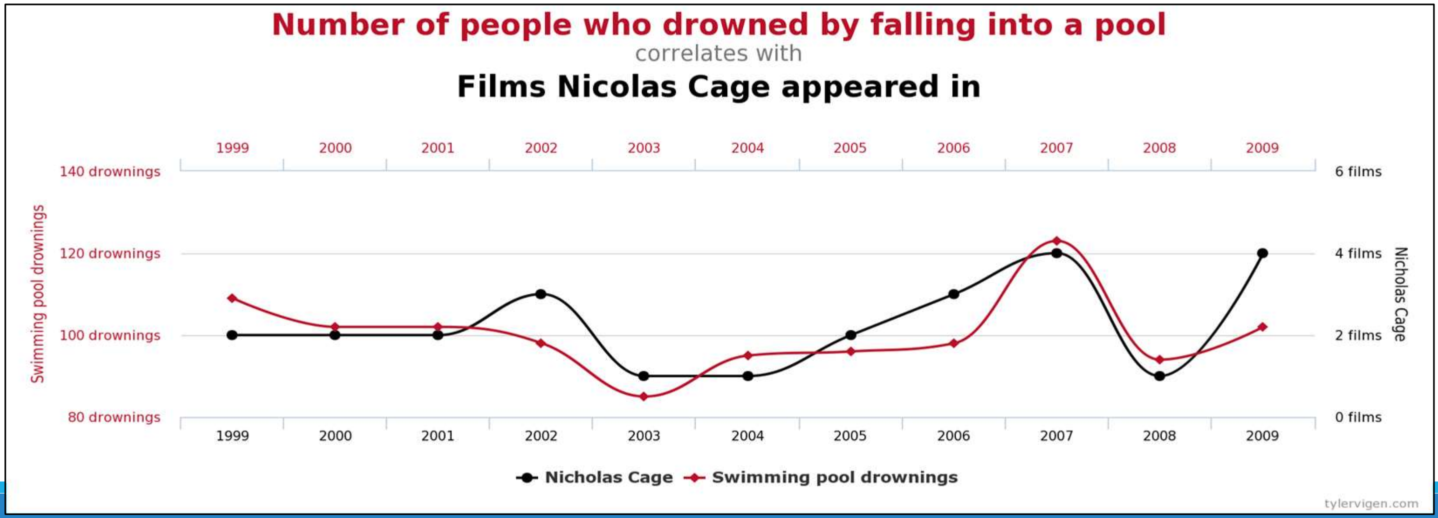


Advanced Statistical Concepts

SKILL SET 9, CHAPTER 9

Inferential Statistics

- Descriptive statistics describe the characteristics of a particular data set. Inferential statistics are used to draw or infer conclusions about a larger population based on a sample.
- Relationship between two or more variables to see how they relate to each other.
- Any measures using descriptive statistics to describe an entire population are called parameters while measures calculated from the sample (not referencing population) are called statistics.



Vigen, T. (2014), https://www.tylervigen.com/

Population and Sampling

- Reminder: goal is to generalize findings from a sample to the population as a whole
- Population: total possible set of an item in a group (people, incidents, addresses). Size will vary.
- Sample: sub-set of observations taken from a population.
- Statistical generalization: ability of a sub-set or sample of data taken from a population to correctly identify a set of characteristics of that population.
- Sampling frame: list from which the sample is drawn.
- Sample size: the number of elements taken from the population. The larger the sample, the more likely it will be to resemble the population.
- Variability: the amount of differences among a variable in the sample and the extent to which it is spread through the distribution or clustered. Generally, more variability requires a larger sample.
- Both variability and sample size will influence the likelihood that your sample will resemble the population.

Normal Distribution and Sampling

- Reminder: normal distribution is unimodal (one peak); with a mean, median, and mode being relatively equal; falling in the center of the distribution (no skew); and symmetrical.
- Large samples (100+) will typically be normally distributed.
- Central Limit Theorem: distribution of means of an independent and randomly selected sample will approximate the normal distribution regardless of the underlying distribution of the parent population. Can assume large samples are normally distributed.
- Point estimate: value from a sample which best estimates or approximates the true value of the population.
- Statistical inferences are accomplished through hypothesis testing – process by which you identify a research question, determine the statistical technique that can be used to answer questions, carry out testing, and make determinations about the data based on the sample.

Probability vs. Nonprobability Sampling Strategies

- Probability samples: must utilize some form of random selection. All elements of the population or sample frame must have an equal chance of being selected.
  - Simple Random Sampling
  - Systematic Random Sampling
  - Stratified Random Sampling
  - Cluster Sampling
- Nonprobability samples: do not use random selection to create the sample. These can be representative of a population, but statistical inference tools cannot be applied to them.
  - Voluntary Response Sampling (people volunteer themselves for the research)
  - Convenience Sampling (individuals most accessible to the researchers)
  - Purposive Sampling (judgment sampling – researcher selects sample most useful to research)
  - Snowball Sampling (currently enrolled participants recruit other participants)
  - Not covered in the book, but I was curious so I thought you might be too. ☺

Types of Probability Sampling

- Random Selection: most powerful method for sampling. Sample is randomly selected when every subject in a population has equal chance of being selected into the sample.
- Simple Random Sample: Sample collected through random selection. Assign each element an equal chance of getting selected.
  - Can be done with or without replacement.
  - In replacement sampling, after each element is selected, it is "put back" into the pool and may be chosen again.
  - In sampling without replacement, after an element is selected, it is taken out of the pool which gives other elements a better chance of getting picked. (most often)
  - Who cares? No one if the population is large enough. However, samples with replacement are not independent (knowledge of one variable does not convey knowledge about another variable). With small populations, this can create issues with covariance or the degree to which two values change together.

Types of Probability Sampling

- Systematic Random Sampling: type of simple random sampling used when an analyst does not have a sampling frame.
- Take the first element/piece of data and select every nth element after the initial one.
- Ex: select the first vehicle larceny incident, and then select every 5^{th} incident to get a sample of the population.
- "n" is generally selected by dividing the total number of cases by the number needed for a sample. This is called the sample interval.
- Following the determination of "n", select a random number and start with this case or element in your analysis. If your n = 5 and you randomly selected number 7, you would start at the seventh case in your population, select that, count 5 from there and select the 12^{th} case, and so on.
- If your n is not a whole number, you select between the two elements/cases where the number falls.
- Note that elements in systematic random sampling cannot be sorted in any meaningful way because they will be affected by periodicity (may make the sample biased and not representative of the pop.)

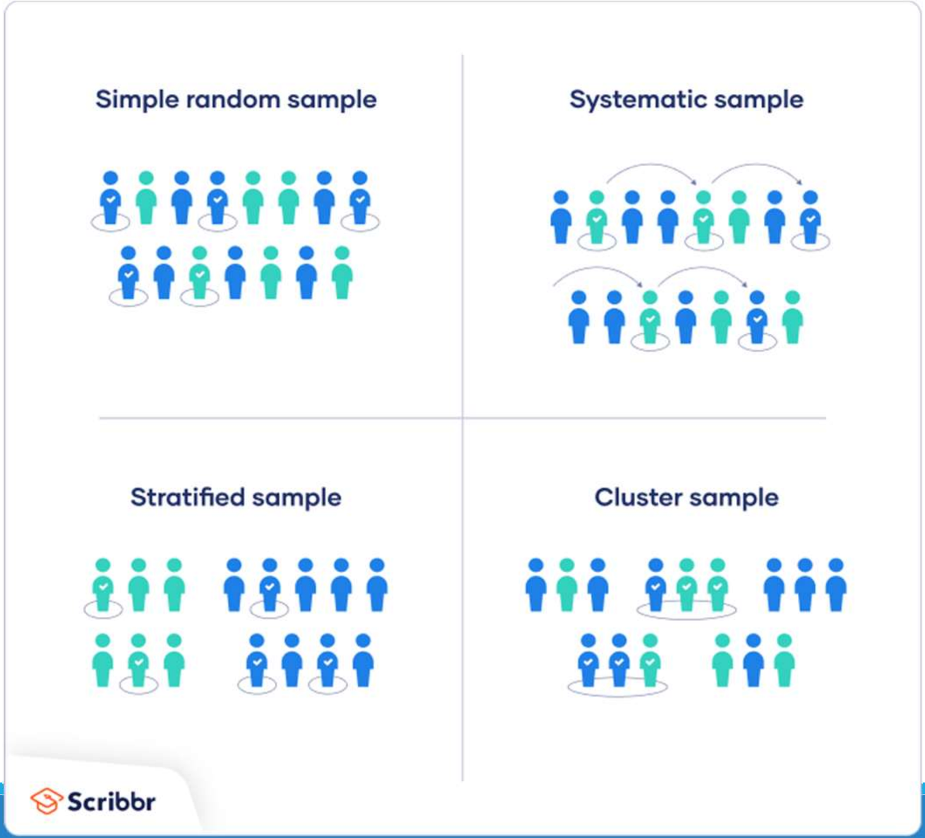
Types of Probability Sampling

- Stratified Random Sampling: uses information known about the population prior to sampling to create strata from which the elements are randomly selected.
- Can be more efficient as strata are created by separating portions of the population into groups according to characteristics of interest.
- Ensures each group of elements has enough samples (rape, robbery, homicide)
- Disproportionate Stratified Sampling – ensure that cases from a smaller strata are sufficiently represented and has the potential to increase precision when sampling fixed inputs
- Proportionate Stratified Sample: sample size of each stratum is determined by its proportion in the population. (see page 189 for a good example of this)
- Proportionate reduces selection bias and error associated with randomization but they require that the population be broken into strata that are mutually exclusive. Can't have one element in two strata.

Types of Probability Sampling

- Cluster Sampling: used when naturally occurring groups with a high degree of homogeneity need to be sampled.
- Useful when sampling frame isn't available
- Overcome limitations when attempting to sample large populations
- Identify clusters (naturally occurring aggregations of elements of a population) to be studied and then randomly select elements to be sampled. Each element can only be in one cluster.
  - Sampling police officers through clusters of police agencies or inmates in clusters of jail
- Single-stage cluster sample: survey all of the elements in a cluster
- Multi-stage cluster sample: first randomly select among the elements (agencies/jails) and then select elements within those clusters (officers/inmates) to sample.
- Cluster sampling is efficient but may increase sampling error (likely divergence of sample statistic from population) because of the complexity.

Types of Sampling



https://www.scribbr.com/methodology/sampling-methods/

Hypothesis Testing

- Hypothesis: specific educated guess about what will occur.
- Hypothesis testing is a form of statistical inference where the analysts uses data to test a hypothesis.
- Dependent variable is the variable that is influenced
- Independent variable is the variable doing the influencing (the thing you are changing)
- Control variables are variables collected to help test the idea that only the IV is influencing the DV (stronger correlation)
- Hypotheses can be directional statements (specify how the IV will influence the DV) or non-directional statements (simply suggest a difference but not identify the direction of the impact)
- Hypotheses cannot be proved. Rather null hypothesis (proposes no difference) vs. research hypotheses (states that there will be a difference) are tested.
- Null must be falsifiable (shown to be incorrect). If have data to demonstrate the null is false with a high enough level of certainty, can reject the null and say the research is possible.

Reasoning Deductive vs. Inductive

- Deductive Reasoning – take a broad generalization (theory) and use this to develop predictions/hypotheses about specific actions, objects, or ideas.
  - Hypotheses can be tested against data and potentially found to be false.
  - Top-down reasoning. Work from large idea to specific prediction.
  - Prior to 2012, UCR definition of rape required offender to be male. Therefore, anyone committing a rape had to be male. Deduction is that rapists can only be male.
- Inductive Reasoning – uses specific knowledge of actions, things, or ideas to induce general properties of a group or class to which those things belong.
  - Bottom up reasoning – Work from specific characteristics about an individual to draw inferences about the group as a whole
  - Example above – look at rape cases and realize all offenders were male. The conclusion would be that all men are rapists (clearly not correct).
  - Possible for more general observation to be false despite specific observation being correct.
  - The Problem of Induction – impossible to draw general conclusions from a series of observations, i.e. Black Swan events. Past performance is no guarantee of future results.

Type I vs. Type II Errors

- Type I Error
  - incorrectly reject the null when it is true.
- Type II Error
  - incorrectly accept the null when it is false.
- Reminder: you cannot prove the hypothesis but rather attempt to balance the errors.
- Likely will have to make a choice to increase the risk of Type I or Type II error.
- Use this determination to establish a level of statistical significance.

When: / If you:	The null is true	The null is false
Reject the null	Type I Error	Correct
Fail to reject the null	Correct	Type II Error

P-Value and Statistical Significance

- P-value is the probability of analyses generating the observed results given that the null hypothesis is correct. Use to determine statistical significance
- One safeguard to reduce error. May be overemphasized is social sciences for publication impact.
- Includes sample size, variability of data, and effect size (statistic that estimates the impact of the IV on the DV) to provide a number of how consistent the data is.
- Finding is considered statistically significant if the p-value exceeds a designated significance level (p-value).
- Social science often set p value at $p<.05$ meaning that a p value of less than that is statistically significant.
- $p<.05$ means that 95% degree certainty the results was not due to randomness caused by the sampling process.
- Can be one-tailed or two-tailed. One-tailed tests if the sample mean is above or below (not both a certain p-value) vs. two-tailed which tests if the sample mean is <u>either</u> above or below a set p-value

Notes on Statistical Significance

- A hypothesis test simply evaluates two mutually exclusive statements regarding a population using sample data.
- If a finding is statistically significant, it does not prove an alternative hypothesis and applies only to a sample taken from the population.
- Finding a non-significance simply indicates that the data from the sample is not sufficient to reject the null hypothesis.
- Probability distribution assigns likelihood to an outcome of a statistical test which allows us to provide context around our findings.

Means Testing: Parametric and Non-Parametric Statistics

- Parametric statistics make the assumption that sample data comes from a population that is normally distributed. Mean, median, and mode are roughly equal. Use Shapiro-Wilk test of normality and Q-Q Plot to determine if parametric tests are appropriate.
- Non-Parametric statistics make fewer assumptions about the distribution underlying the sample. Can be used with ordinal (rank) data or continuous data that can be converted.
  - Best to use non-parametric tests when:
    - Data is ordinal (rank)
    - Data have significant outliers
    - Data is continuous, but limits cannot be detected (tests for narcotics when determining if they are there or not)
    - Count data is not normally distributed (esp. if data has zeros)
- Parametric tests: T-Test (one sample, two sample, and paired) and ANOVA (one way or two way). Non-Parametric tests: Mann-Whitney U test and One-way ANOVA using ranks.

Parametric Tests – T-Test

- T-test: Statistical tests that compares two means. Helps you determine if the differences between the means are based on chance. Ex: compare experimental group mean to control group mean to see if a new drug is effective on life expectancy (know the population life expectancy).
- P-value applied to help provide a guard against random variations in the data causing the analyst to draw the wrong conclusions.
- One-tailed t-test determines direction (sample mean is significantly smaller or larger). Can help determine if sample mean comes from a population with either a known or hypothesized mean. As a rule, as sample size increases, the t-distribution becomes more similar to the normal distribution.
- Independent sample t-test is designed to determine if two different means, independent from one another, come from different populations. (ex: avg. shooting score of officers with annual refresher compared to avg. shooting score of receive refresher every third year)
- Paired samples t-test used when sample means are not independent. Ex: check the shooting scores of the same individuals at the time they exited the academy with their mean scores later.

Analysis of Variance (ANOVA)

- Useful when analyst wants to compare means from three or more independent samples that come from one or more categorical groups. More flexibility than t-test because more than 2 means can be tested.
- One-way ANOVA = similar to a t-test but for more than two groups.
- Seek to compare multiple levels of one factor, though can look at two factors each with multiple levels.
- Ex: examine impact of education and rank on morale for police officers.
- Two-way ANOVA: determine if there is an interaction between two independent variables on the dependent variable.
  - Main effect are the two independent variables (education and rank).
  - Interaction effect is how the independent variables interact with each other to impact the dependent variable.

Assumptions Met Prior to T-Test/ANOVA

T-Test

1) Dependent variable must be continuous.
2) Dependent variable should be normally distributed or close
3) There should be no significant outliers
4) The dependent variables should be independent of one another.

ANOVA

1) Dependent variable must be continuous.
2) Dependent variable should be normally distributed or close.
3) Observations must be independent of each other.
4) Independent variable is categorical and has three or more independent groups.

Non-Parametric Tests

- Wilcoxon-Mann-Whitney U-Test – form of an independent samples t-test for data with non-parametric distribution.
- Used to compare two groups where the dependent variables are either continuous or ordinal
- Frequently more useful for analysts because of the ordinal data.
- Can accommodate ordinal data because it is rank-order test. Treats all data as ranks. Very helpful in dealing with issues like outliers.
- Tests the difference of the <u>medians or distributions</u> of the two groups rather than the means.
- Again dependent variable needs to be continuous or ordinal, must be an independent variable composed of two categorical groups independent of each other, and the observations are mutually exclusive.
- Wilcoxon Signed-rank Test – can be used for dependent observations. Makes the assumption that data are from the same population and paired, the data is categorical, and the pairs are independent of each other. Useful when t-test data not near normalcy.

Correlation – Statistical Association

- Degree to which two or more variables change in unison.
- Positive correlation: as one variable increases/decreases, the other increases/decreases
- Negative correlation: as one variable increases, the other decreases. (bivariate correlation)
- Correlation does not equal causation but indicates dependence
- Strength of the relationship is measured through the correlation coefficient which ranges from 1 (values go together) to -1 (as one value increases the other decreases). A value of 0 means there is no correlation.
- Correlation coefficient is the degree of dependence between two or more variables, Pearson's r^2. Used to determine strength and direction of association between two variables.
- Variables should be continuous (interval or ratio), two variables form a bivariate normal distribution, no outliers, and the relationship should be linear.
- Assumes data is homoscedastic – variance of dependent variable is equal across the regression line.

Spearman's rank-order correlation coefficient and Chi-Square Test

- Spearman's correlation – uses the ranking of the variables to minimize the impact of others. Can be used for two continuous variables, two ordinal variables, or one continuous variable and one ordinal variable.
- In addition to the requirements for Pearson's r, Spearman's correlations requires the variables be paired observations (stops and calls coming from the same neighborhood) and there must be a monotonic relationship (as one variable increases, the other variable either increases or decreases). It does not need to be linear.
- Chi-square test for independence is appropriate for nominal/dichotomous variables. Goal is to determine if the two variables are statistically independent (knowledge of one variable does not convey any knowledge regarding the properties of another variable).
- Ex: demonstrated relationship between traffic stops and rate of emergency calls. Use chi-square test to determine if the race of the person stopped had an impact on vehicle searches.

Regression Analysis

- Method of modelling the relationship between a dependent variable and one or more independent variables.
- Often used in the course of a prediction but can help identify specific independent variables related to the dependent variable.
- Linear regression demonstrates the existence of a linear relationship to predict the value of the dependent variable based on the value of the independent variable.
- Both variables must be continuous and linear regression cannot accommodate more than two variables.
- Logistic regression: model outcome of categorical dependent variables (type of arrest, demographic data)
- Poisson Regression: model count data (includes zeros). Creates positively skewed data.
- Negative Binomial Regression: used when the dependent variable is a count (number of car larcenies) and is over-dispersed. Ex: number of shoplifting arrests at retail places based on the characteristics of the business (security cameras, etc.)

Conclusions

- Read the books and take the classes to strengthen understanding.
- Check out Statistics for People Who Hate Statistics.
- Try to apply the things learned to your every day work to “make them stick”.
- Use the study guides.
  - https://iaca.net/about-clea/ (links for program outline and study guides here)
  - https://iaca.net/about-leaf/ (links for program outline and study guides here)
- Next month: Spreadsheet Operations (Skill Set 11)

Any questions?