Basic Overview: Part 2

Descriptive statistics: Describe the data with graphs, words, and numbers that indicate the middle and spread of the data

Inferential statistics: Uses what we know about a sample to *infer* about the larger population, using the concept of probability

- Probability & the Normal distribution
- Null Hypothesis Testing
- t tests (one sample, t test of r vs. 0, independent samples, dependent samples)
- ANOVA (one way ANOVA, factorial ANOVA, repeated measures ANOVA)
- Chi square

Two types of distributions

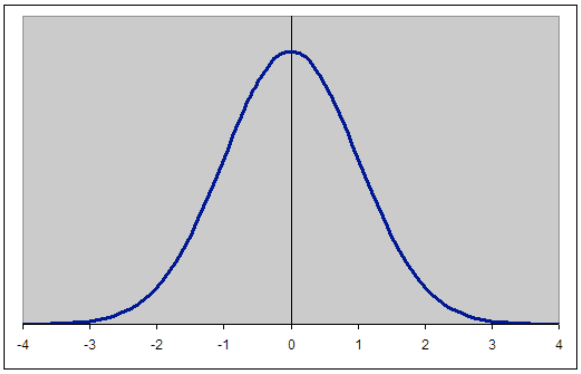
1. **Empirical distributions**: Scores that come from observations (real data).
2. **Theoretical distributions**: Hypothetical scores that come from math formulas and logic.
   - As the number of observations increase, the empirical distribution comes closer and closer to the theoretical distribution.

Probability:

- Basic question: what are the odds?
- Varies from 0 (impossible) to 1 (certain)
- A ratio of success or failure to total number of events
  - Example: If I role a die, there is 1/6 chance that I will role a 3.
- Empirical approach to probability: observe events; look at ratio of success
- Theoretical approach to probability: find the area under a theoretical curve

Normal distribution:

- A bell-shaped, theoretical distribution that predicts the frequency of how often events will occur by chance.
- All normal distributions have the same overall shape - bell shaped, symmetric, single peak.
- Mean = μ, Std dev = σ.
- The standard deviation controls the spread of the normal curve.



Sampling issues

- Whenever we use a sample, there is a certain amount of uncertainty. We use sampling distributions to measure this uncertainty.
- Types of samples:
  - **Random sample** - A subset of a population chosen so that all samples have an equal chance of being selected.
  - **Biased sample** - Not all from the population have an equal chance of being selected.
  - **Non-random research samples**
- Why does the sample matter? Determines whether or not we can generalize to others in the population.

Sampling distributions

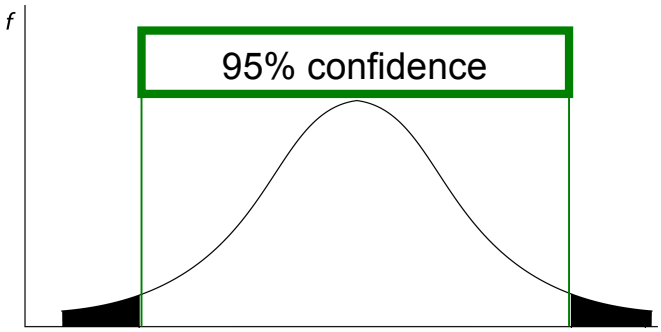
- Last week we talked about 1 sample, and from this, we calculate statistics such as M, S, r, etc.
- How confident can we be that this represents the whole population?
    - We draw a second sample, and do things again. Then we draw a third sample, and a fourth, and a fifth. If we do this over and over again, we get a **sampling distribution**.
- Definition: a theoretical distribution of a statistic (like the mean or standard deviation) based on all possible random samples drawn from the same population.
- **Expected value:** The mean of a sampling distribution (the mean of all the sample means)
- **Standard error:** The standard deviation of a sampling distribution
- When there are enough samples, the distribution is normally distributed

t distributions

- t distributions are a type of curve distinguished by their degrees of freedom (df).We draw a second sample, and do things again.
    - Degrees of freedom range from 1 to (infinity symbol); this determines what t distribution you use.
    - A benefit of t distributions: identifying extreme scores
    - Main question: what's the chance that you're sample is different from the population? Or, how confident are you that your sample represents the population and is not an extreme score?

Confidence intervals

- Upper and lower limits are calculated for the mean of an immeasurable population, using a sample.
    - → Based on your sample, how certain are you that the population mean lies in the interval you calculate?
- We often like to get the 95% CI
    - I am 95% certain that the mean of the population lies within this interval.



Null hypothesis testing

- A process that produces probabilities that are accurate when the null hypothesis is true.
- Testing something you claim based on a sample, and using probability to compare it to the population.
- We test for no effect
    - **Null hypothesis**: H_0 – the equality hypothesis; there is no relationship
    - **Alternative hypothesis**: H_a – there is a relationship or effect in the population
- If there is a direction to your hypothesis, then use a 1 tailed test. If there is no direction to your hypothesis, then use a 2 tailed test.

T tests and ANOVA

- If we want to compare our sample to a population value, we can use a 1 sample t test. If we want to compare 2 samples to see if they are similar or different. In this case, use a 2 sample (or independent samples) t test.
- Types of t tests:

    - **One Sample t-test**: compares a sample mean to a specified value or a population mean.
    - **t-test of r vs 0**: determines whether or not two variables are related (testing correlations).
    - **Two sample- independent t-test**: a test that compares two group means (groups are not correlated; one sample is independent of the other).
    - **Two sample- correlated (dependent) t-test**: a test that compares two group means (groups are correlated in some way; one group is dependent on the other's mean).
- Our t test tells us whether there is a significant difference. An effect size tells us how large a difference there really is.
- Analysis of Variance = **ANOVA**
    - An inferential statistics technique that lets us compare means, variances, and interactions among variables.
    - Examines whether two or more population means are equal
        - Used when comparing more than two independent groups.
        - If we compare two groups, we should get the same results as using a t test.
        - In one-way ANOVA, there is 1 independent variable and multiple groups (levels, treatments). We compare the means at the same time
- Types of ANOVAs:
    - **One way ANOVA** - one independent variable with multiple levels
    - **Factorial ANOVA** - two independent variables with multiple levels
    - **Repeated measures ANOVA** - correlated samples within ANOVA
    - **Factorial ANOVA** is an experimental design with two or more independent variables
        - Factor = independent variable
        - Each variable can have two or more levels
        - An advantage: we can look at *the effect of each variable*, AND the *interaction* between the two variables
- For each research question there is:
    - A null hypothesis
    - An alternative hypothesis
    - A calculated test statistic
    - A critical value that must be compared to calculated value
    - A conclusion
- Assumptions of t & F
    - **Normality** – the populations are normally distributed
    - **Homogeneity of variance** – the groups have similar amounts of variance within the groups
    - **Independence** – a person is only in one group and not related to the person in the other group