STATS: Exam #1 Material

1. **Provided definition of independent and dependent variable that your mother could fully understand.**

Independent variables: variables experimenters have control over & what is being manipulated.

- The variable that is changed or controlled in a scientific experiment to test the effects of the DV.
- Ex: the number of cookies someone eats

Dependent variables reflect the effects of changes in the independent variables.

- The variable being tested and measured in a scientific experiment. The effect being measured.
- Ex: Their performance on a demanding physical test

2. **Explain what a paradigm is and how hypotheses are constructed relative to that paradigm.**

A **paradigm** can have a <u>number of theories within its framework</u> & paradigm acts as a reference point for the theory. These two concepts operate with each other but have their differences.

- Paradigms contain all the distinct, established patterns, theories, common methods and standards that allow us to recognize an experimental result as belonging to a field or not.
- The paradigm is the "world view" as a set of assumptions about how things work.
- Theory is a system of ideas intended to explain something (based on the paradigm)
- Hypothesis- Has to be testable. A testable statement about the relationship between variables.
- Theorie are influenced by your paradigm, hypothesis generated to test certain aspects of your theory, and the results from your hypothesis can influence your theory. Hypothesis work under the assumptions of your paradigm.
- If you fail to reject the null→ something is wrong in theory, hypothesis, paradigm, or chance. Often times we assume our data is faulty but it is likely the theory.

3. **What does it mean to measure something?**

Measurement is the process of quantifying properties of an object by comparing them with a standard unit.

- The assignment of numeral to objects or events according to rules
  - Using numbers to represent what we are studying
- Nothing more or less than the attempt to discover or estimate such numerical relations to one another

4. **What are the measures of central tendency and their relationship to different types of distributions (Gaussian, bimodal, skewed, uniform)?**
   a. **What is skewness and kurtosis?**

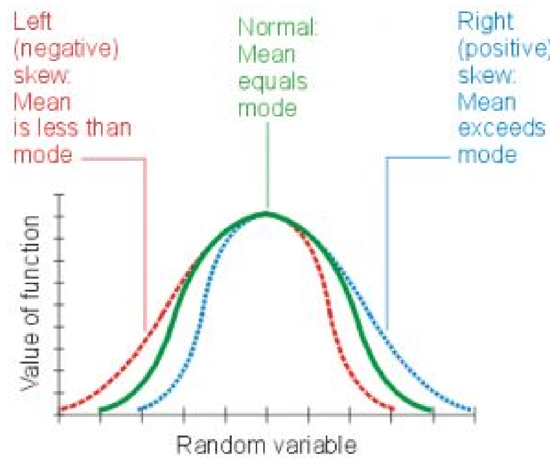
Mean → the balancing point of a set of scores; the average

- One value of the distribution that balances the rest out
- We can draw conclusions from this value

- the mean is the average of the data set, and can be used on interval or ratio data. It is a parameter defining the normal Gaussian distribution, and is the most stable measure of central tendency from sample to sample. However, it is highly influenced by outliers, as it is based on all scores from the sample, and therefore is not reliably representative in skewed distributions.

Median → value at which ½ of ordered scores fall above & ½ fall below
- This is used instead of mean when data is NOT symmetric (aka skewed) bc the mean is sensitive to outliers & is generally pulled towards where the outliers lie which may end up inflating or deflating the results

Mode → most frequently occurring score
- Visually, data represents what occurs most often
- When there's a lot of data → frequency tables
- It can also be used with ordinal, interval, and ratio data. It is the least stable measure from sample to sample, but can be useful in describing bimodal distributions (distributions with two modes – two separate, most frequent data points).



Skewness → asymmetry in a statistical distribution, in which the curve appears skewed either to the left or to the right. It can be quantified to define the extent to which a distribution differs from a normal distribution.

Kurtosis → 4th moment in statistics
- a measure of the combined weight of the tails relative to the rest of the distribution.
- (+) kurtosis → fat tails
- (-) kurtosis → relatively thin tails

Add some bimodal notes

5. **What the reasons for using randomization to select a sample for an experiment?**
   a. **What are the dangers of random sampling without replacement**

- To remain unbiased.
- Randomized experimental design yields the most accurate analysis of the effect of an intervention
- purpose of taking a random sample from a lot or population and computing a statistic, such as the mean from the data, is to approximate the mean of the population.
- Random sampling allows for generalizability of results to the broader population by creating samples that are, in theory, representative of the population if done correctly.
- **Danger** → it will not necessarily be a true representation of the population?
- Without replacement, sampling becomes less and less random, as each following member of the pool has a higher probability of being chosen than the last.

In sampling without replacement, the two sample values aren't independent. Each sample unit of the population has only one chance to be selected in the sample.

- Ex: if one draws a simple random sample such that no unit occurs more than one time in the sample, the sample is drawn *without replacement*

6. **Provide a verbal definition of variance.**

Variance is the measurement of dispersion within a set of scores. If the variance were to be zero, than all of the scores would be identical. Variance allows us to find differences in our data & the importance of those differences.

a. **Be able to explain the differences between metrics of variance.**

Biased vs unbiased variance

- Biased → only N is in the denominator it underestimates the variance of the population
- Unbiased → corrects the underestimation by subtracting 1 from n; (n -1)

b. **Explain the concept of degrees of freedom. (df)**

'The number of deviations from the mean that are free to vary' = df

We use df when we DO NOT know the population mean, therefore, we must calculate variability from the mean of our sample which means that we will lose one degree of freedom

- It allows us to remain unbiased statistics

c. **What is the benefit of the coefficient of variation over the raw standard deviation**

coefficient of variation (CV) is a **better** risk measure **than** the **standard deviation** alone bc it adjusts for the size of the project.

CV is the ratio of the standard deviation to the **mean**. (usually expressed in %)

- Higher CV → more variability, scores are more dispersed
- Lower CV → more precise estimate

Without units, it allows for comparison between distributions of values whose scales of measurement are not comparable.

(i.e. a data set with a standard deviation that is 15% of the mean shows less variation than one with a standard deviation that is 25% of the mean).

i. **Under what circumstance with the coefficient of variation be inappropriate to calculate. How might you correct for that problem?**

Inappropriate to calculate when the mean of a sample population is zero. In other words, the sum of all values above and below zero are equal to each other. In this circumstance, the formula for COV is useless because it would place a zero in the denominator

- Even if the mean is close to zero the CV can be misleading

d. **How are the mean and standard deviation related?**

- **The SD is a measure of variability from the mean.**

7. **What are conditional probabilities and how do the impact determining the probability of the outcome of an experiment?**

conditional probability is the probability of one event occurring with some relationship to one or more other events. For example:

- Event A is that it is raining outside, and it has a 0.3 (30%) chance of raining today.
- Event B is that you will need to go outside, and that has a probability of 0.5 (50%).

A conditional probability would look at these two events in relationship with one another, such as the probability that it is both raining *and* you will need to go outside

Conditional probabilities are a measure of probability on some event given that another event has already occurred. This occurs when probability is measured without replacement

8. **Why the central limit theorem is so useful?**
   a. **How does the central limit theorem apply when the underlying distribution you are selecting from is uniform?**

central limit theorem states that if you have a population with mean μ & SD σ & take large random samples from the population **with replacement** then the distribution of the sample means will be approximately normally distributed. This will hold true regardless of whether the source population is normal or skewed, provided the sample size is sufficiently large (usually $n \geq 30$).
If the population is normal, then the theorem holds true even for samples smaller than 30. In fact, this also holds true even if the population is binomial, provided that $\min(np, n(1-p)) \geq 5$, where n is the sample size and p is the probability of success in the population. This means that we can use the normal probability model to quantify uncertainty when making inferences about a population mean based on the sample mean.

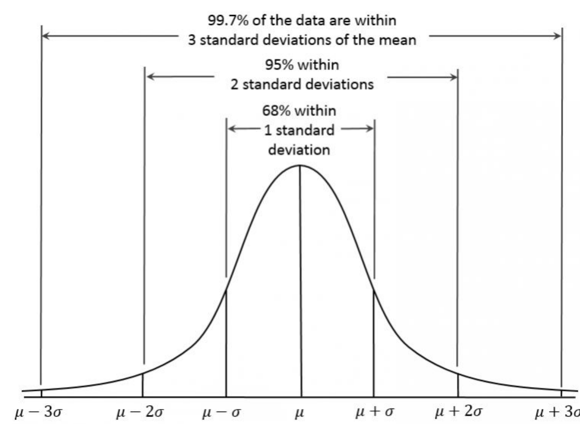
Uniform distribution → "rectangle distribution"

- Distribution that has constant probability; there are an infinite number of possible uniform distributions
- not particularly useful in describing much of the randomness we see in the natural world. Its claim to fame is instead its usefulness in random number generation → useful for assigning treatments

9. **What are Z-scores?**
   a. **Explain how you can use the z-score to calculate percentiles.**

A z-score shows you the distance between an observed score & the mean in units of SDs when the SD of the population is KNOWN

- We calculate the scores based on differences between a score from the distribution & the mean of the distribution, relative to the SD
- A z-score is thus a standardized score, as the number of standard deviations is a measure that can be compared between data sets of differing measures.
- Centile scores are intuitive ways of summarizing a person's location in a larger set of scores



68% of the observations fall within 1 SD of the mean
95% fall within 2 SD
99.7% fall within 3 SD.

- That means the probability of observing an outcome greater than 3 SDs from the mean is very low: 0.3%

Assumptions for **One-sample Z test:**

- Must know the null hypothesis distribution (normal usually)
- dependent variable must be measured on an interval or ratio scale
- sample was drawn randomly
- variable measured has a normal distribution in the population
- standard deviation for the sampled population is the same as that of the comparison population

10. Explain why normality is so important to parametric statistics.

normality tests are used to determine if a data set is well-modeled by a normal distribution & to compute how likely it is for a random variable underlying the data set to be normally distributed.

It's important bc Most statistical tests rest upon the assumption of normality. Deviations from normality, render those statistical tests inaccurate, so it is important to know if your data are normal or non-normal.

Tests that rely upon the assumption of normality are called **parametric tests**
If your data is not normal, then you would use statistical tests that do not rely upon the assumption of normality, call non-parametric tests.

Non-parametric tests are less powerful than parametric tests, which means the non-parametric tests have less ability to detect real differences or variability in your data. → You want to conduct parametric tests because you want to increase your chances of finding significant results.

- Look at the frequencies of variables whereas parametric tests are looking at specific values of dependent variables for individuals.

11. What is sampling error?

Sampling error (standard error)→ sample might have different properties than population

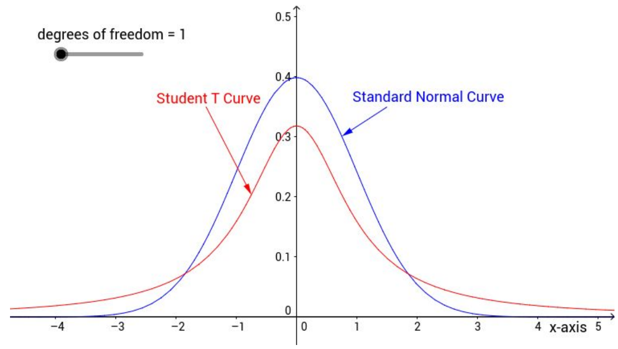
- A way to measure the accuracy with which a sample represents the population
- The bigger the sample size → SMALLER standard error
  - Experimental group may contain individuals that naturally higher performance on task than control such that it may show a mean difference that will be mistaken for a systematic effect.

Because sample means do not vary as much as individual data points, the standard deviation for the sample will be less than the standard deviation for the population. The standard error accounts for this variability in standard deviation. As sample size goes up, the standard error goes down, and vice versa. The standard error is calculated by dividing the population standard deviation by the square root of the sample size. This relation with the standard deviation means that when the standard deviation of the population is higher, so is the standard error (the more individuals vary, the more the groups will vary), and vice versa. On the other hand, the larger the sample size, the less the sample means will vary, and the standard error will be lower.

a. **How are we estimating it from experiment?**

12. Explain how the t-distribution is related to the normal distribution

- *t*-distribution is used when the population variance is unknown
- estimating the variance from sample leads to greater uncertainty and a more spread out distribution, as can be seen by the *t*-distributions heavier tails.
- *t*-distribution approaches normal distribution as the degrees of freedom increase.
- It also shows the absolute and relative error when the normal approximation is used.



13. What are the steps involved the null hypothesis testing?

1. State the hypothesis
2. Set the criteria decision
3. Collect data and compute sample statistics
4. Make a decision (accept or reject the null)

a. **What is type I and type II error?**
   i. **How do they relate to each other?**
   ii. **How are they related to setting an alpha?**

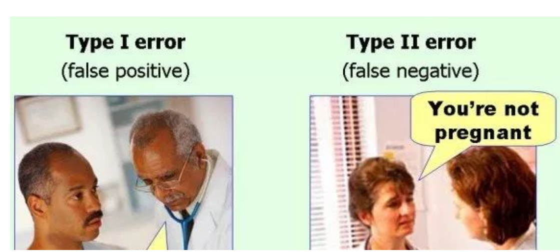
Type I errors are represented by the alpha level, which represents the area of the null hypothesis distribution above the criterion level. Type II errors are represented by beta, which represents the area of the alternative hypothesis distribution beneath the criterion level. Therefore, alpha represents the probability of making a type I error if the null is true. Beta represents the probability of making a type II error if the null is false.

null hypothesis: a hypothesis that is possibly false or has no effect or effects are due to chance. Often, during a test, the scientist will study another branch of the idea that may work, which is called an alternative hypothesis

- If false → variance between means will be greater than what we would observe if it were true.
- Accept (fail to reject) null → results are due to chance,dont retain the null ($p > 0.05$)
- Reject null → results are less likely due to chance, chance of Dr. Null beating ME is low → less than *alpha (the amount of risk you're willing to take), it is NOT the null*

Alternative hypothesis

- Observed data is a real reflection of real association
- Reject the null → $p < 0.05$
- The larger the z-score the lower the p value



Type I Error

- When we reject the null hypo although it is true; think you have an effect but YOU DONT
  - Experimental was entirely ineffective; falsely claiming statistical significance for results
  - We are more concerned with having a type I error because having a false positive is worse than a false negative
  - Minimize → by having bigger alpha value
  - Reduce alpha error → reduce error (ex:from .05 to 0.005)
    - This tends to create type II error

Type II Error

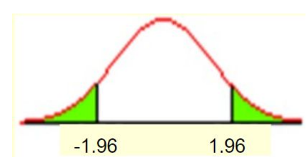
- Accept the null hypothesis even though it is false (less dangerous bc it's more conservative, you missed the effect but didnt find it)
  b. Not providing false-hope
  c. Experimental manipulation was at least somewhat effective
  d. To reduce this type of error → perform one-tailed instead of two-tailed test
    i. P-value is doubled when you do a one-tailed test
    ii. Bigger sample size will increase strength of manipulations

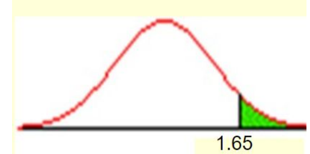
  e. When should we use a directional hypothesis?

Logically, all the time. A directional alternative hypothesis is one which states the direction of change in relation to the null. For example, if a null hypothesis states that no difference in height will be present between men and women, a directional alternative hypothesis would state that men will be TALLER than women. Often times, psychologists test nondirectional hypotheses, making for a less conservative analysis, as a change in either direction can result in a significant result. Directional hypotheses are more conservative, as only change in the predicted direction can result in a significant result.

Ways to specify H0 and Ha (w/ example)

- **Null hypothesis:** depressed w/ treatment **=** depressed without treatment
  - No true difference exists between treatment vs no treatment.
  - Any observed difference is due to sampling error
- **Alternative hypothesis**:
  - NON-DIRECTIONAL (two-tailed, zcrit = 1.96) → (assuming there is a difference but not sure if it will be greater or less than pop. mean
    - Depressed w/ treatment $\neq$ depressed without treatment
      - Difference between well-being exists → not sure which will be greater
  - DIRECTIONAL (one-tailed, zcrit = 1.64) → anticipate in advance that mean will be $\geq$ *or* $\leq$ than pop mean)
    - Depressed w/ treatment $>$ depressed w/out treatment
      - Prediction is made beforehand





Direction vs Non-directional Summary

- In z & t tests, the way hypotheses are specified can have an effect on criteria for evaluating our statistical test
- If you specify your hypothesis → one-tailed t-test
- If you specify difference but NOT direction → two-tailed t-test

- In practice, most researchers use 2-tailed tests by default. This allows for possibility that they could be wrong & effect could be in opposite direction of what was initially predicted

f. Why can we not accept the alternative hypothesis?

Our results could be due to chance fluctuations in random sampling.
We cannot accept the alternative hypothesis because, even if there is a change significant enough to reject the null, this does not rule out other factors besides that stated with the alternative hypothesis in causing this change.

g. What are the benefits and drawbacks of null hypothesis significance testing?

Pros:
- They provide an accepted convention for statistical analysis
- The techniques are tried and tested

Cons:
- Commonly misinterpreted
- *P*-values take no account of any hypothesis other than the null.

h. Take the position of Fisher and explain to Neyman and Pearson, the problem of NHST when you set an alpha of .05 and the exact probability of your experiment turned out to be .051.

Fisher thought that the p-value could be interpreted as *a continuous measure of evidence against the null hypothesis*. There is no particular fixed value at which the results become 'significant'.
- To Fisher a p=.049 & p=.051 constitute an identical amount of evidence against the null hypothesis

Neyman & Pearson thought you could use p-value as part of *a formalized decision making process*.
- you have to either reject the null hypothesis, or fail to reject the null hypothesis
- No marginal …. Either it's significant or is not

14. What are t-tests?

t-test is an analysis of two populations means; a t-test with two samples is commonly used with small sample sizes, testing the difference between the samples when the variances of two normal distributions are not known.

a. Explain the assumptions underlying the t-test.

Assumptions: One-sample *t* test and the confidence interval for the population mean:
- Independent random sampling
- Normal distribution
- Standard deviation of the sampled population equals that of the comparison population
- A bootstrapped confidence interval

b. Why can't we just compare t-values between experiments?

Different experiments are using different samples. Chance that either value is very unrepresentative of population, or experimental manipulation.

c. **Explain the logic behind correcting the degrees of freedom for specific violations of the t-test.**

Homogeneity of variance:

- When this assumption is violated (i.e. about 3 to 1 ratio for variance for your two samples), it's unethical to go back and try to sample more people to try to meet this assumption (considered p-hacking). Also can't transform one group to make more normal, and not the other You would instead go back and correct the df based on the degree of variance mismatch for sample size. Essentially penalizes df based on proportional balance issues with variance of samples.

d. **What are confidence intervals?**

Confidence interval → provides a range of values which is likely to contain the population parameter of interest.
Confidence intervals describe the probability that a certain percentage of the data (i.e. 95%) will fall between a certain range of values.

e. **Describe a scenario where t-test would not produce a significant result but there are obvious differences between the two groups.**

If there are not enough degrees of freedom (the sample size is not large enough), then the standard error will be larger, and it will be more difficult to get a t-value high enough to reach the threshold level.

f. **Explain the rationale for the pooled variance calculation necessary for both the independent sample t-test and hedges G.**

Hedges g=sample statistic (biased, but small and usually ignored) of Cohen's **d**

It is assumed that both populations have the same variance, and so both sample variances you have are estimates of the same population variance. When sample sizes are not the same, the larger sample has the better estimate of the population variance, and thus you should weight the values, and use the pooled variance as the population variance estimate for both samples (should be a better estimation than either sample on their own, because it considers a greater number of people)

g. **Explain the rationale for denominator of the independent sample t-test.**

This denominator accounts for sample sizes that are too small (not enough degrees of freedom). With smaller sample sizes, the standard error will be larger (the standard error of the difference is the denominator) , thus making it more difficult to get a t-value high enough to reach the criterion threshold. This denominator calculation keeps the calculation more conservative.

h. Explain when it is appropriate to transform your dependent variables. In addition, explain which transformation you would use to correct a negative vs positive skew.

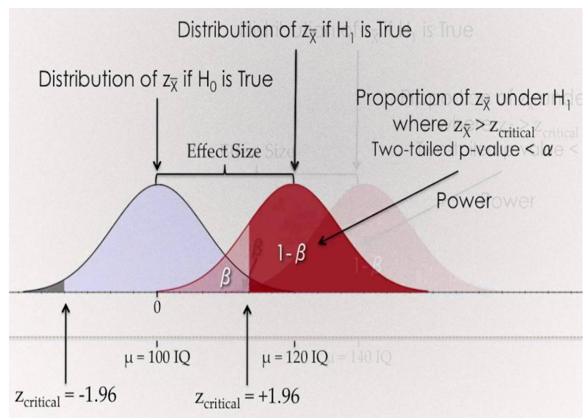
When your sample has violated the assumption of normality, it is appropriate to transform the data. Make it normally distributed.
Positively skewed (floor): Most often, would correct with a **log transformation**
Negatively skewed (ceiling): ??

Idk--be able to argue which one based on what he gives, find one that results in least amount of variance?

15. Explain power and effect size both verbally and graphically.

Effect Size → magnitude of effect in population; can be conceptualized in many ways (ex: Cohens D)

- You need alpha & sample size in order to predict it
- Correlation is a measure of effect size
- Has direct influence on statistical power
- Less likely to make type II error



Statistical power → probability of rejecting the null when it is false → a correct decision

- Strongly influenced by effect size & sample size
- Large sample size → smaller SEM → larger t
    - With larger N. sampling error becomes less problematic & true differences are easier to detect

Inversely related to type II error → more power = less likely to type II error

a. How to sample size relate to power and effect size.

As sample size increases:

- Power increases
    - N in numerator when calculating □, so larger □ value, more likely to reject null when effect is real
    - Also, as N increases, SE decreases, less overlap in curves (skinnier curves, more chance of rejection)
- Effect Size increases
    - As you get larger □, **d** increases--means of curves moving away from each other, more likely to find higher effect sizes when the average *t* value is higher

b. How should we set power for experiment?

A priori we set it beforehand.
power analysis: we run a power analysis in order to predict the power of a statistical test before we run the experiment

- With this you can find the power associate with each possible effect size for a given combo of alpha & sample size.
- With a power analysis we can decide whether it's worthwhile to run the experiment
- Find □, use power tables. Having a greater average *t* value means more power (more likely to find effect if it's there, i.e. NOT commit Type II error)
    - Helps you decide how many people you need to obtain an adequate amount of power

c. What are some factors that would hurt the power of an experiment?

Power depends on sample size

Power also depends on variance: smaller variance yields higher power.

- Larger samples, smaller variance

Smaller effect size

- Small effects harder to detect, because the means are closer together, there will be a large amount of overlap (larger samples means less overlap, and more likely detection of even small effects)

d. Explain a scenario where you can have a large effect size, and non-significant result.

- You run an experiment that has a very small sample (maybe a special population, could only get 10 people). The effect of a treatment is very large, but because there is a great deal of variance, due the small sample, the test is underpowered, and results in Type II Error: non-significance. There is too much overlap between the null and alternative (wide curve vs skinny curve from variance)
