- **key terms/concepts:**
  - **sampling distribution**
    - a probability distribution composed of all values of some sample statistic that would occur if all possible samples of a fixed size (N) with replacement were drawn from a pop’n of X.
    - pop’n distribution and also a pop’n distribution of sampling (random) errors
    - serves the same fct as the freq. distribution, provide the benchmark for comparing statistical results and used to make judgement about the value of pop’n parameters/
  - **binomial distribution**
    - sampling distribution that shows the probability of obtaining X successes in N trials.
    - there are only two possible outcomes on each trial
    - trials are independent (probability of an event does not change across trials)
    - they are complementary events, thus mutually exclusive and exhaustive. Ex:s tossing coin
    - \[ p(X) = \frac{N!}{X!(N-X)!}p^xq^{N-x} \]
      - N = # of trials
      - X = # of specified outcomes
      - p = probability of the event occurring (positive)
      - q = probability of the event not occurring
      - N – X = # of negative outcomes
      - any # with an exponent of 0 is equal to 1
  - **central limit theorem**
    - if we have a big N, distribution approaches normal & can safely estimate the standard error of the mean
    - can’t use for sampling distribution of correlations, cuz not means, in that case need to convert r’s to Zr’s to correct skew
  - **standard error of the mean**
    - standard deviation of the sample means, a parameter
    - \[ \sigma_{\bar{X}} = \sqrt{\frac{\sigma_X^2}{N}} = \frac{\sigma_X}{\sqrt{N}} \]
      - \( \sigma_{\bar{X}} \) std dev. of pop’n x
      - \( \frac{\sigma_X}{\sqrt{N}} \) sq root of sample size
  - **sampling error**
    - random error
    - the difference between the value of an unbiased estimator (i.e., sample statistic) and the value of the population parameter. Sampling error is assumed to be random (nonsystematic).
  - **confidence intervals**
    - a range of values centered around a sample statistic with a specific probability (e.g 95) or degree of confidence (e.g 95%) of including the value of the pop’n parameter
    - Theoretically, a confidence interval (CI) can be developed based on the sampling distribution of any sample statistic that is used as an estimator of a pop’n parameter
      - sample mean is the statistic (estimator) being used and the pop’n mean is the
parameter being estimated, but remember that CIs are not restricted to means. CIs can also be used with correlations, variances…

- same percentage (e.g., 95%) of such intervals will contain the value of the population mean

%CI = lower limit ≤ μ_x ≤ upper limit

- Calculate CI
  1) Need alpha: Alpha = 1 – % Confidence/100 ex: alpha of 95% CI is 0.05
  2) df= n-1 for t distribution
  3) 2 tailed, t_{crit} use Table C (tail #, df, alpha)
  4) get standard error of the mean (  ) or unbiased estimate of mean (  )
  5) choose one of the following for limit:
     - pop’n variance known, use Z:
       Upper limit of CI = \overline{X} + (2 - tailed critical Z)(\sigma_{\overline{X}})
       Lower limit of CI = \overline{X} - (2 - tailed critical Z)(\sigma_{\overline{X}})
     - pop’n variance estimated, use t:
       Upper limit of CI = \overline{X} + (2 - tailed critical t)(\hat{S}_{\overline{X}})
       Lower limit of CI = \overline{X} - (2 - tailed critical t)(\hat{S}_{\overline{X}})
  6) conclude: ex: The 95% CI is 101.82 ≤ μ_x ≤ 114.18

- p-value
  - a measure of how likely we are to get a certain sample result or a result “more extreme,” assuming H_0 is true
  - by comparing alpha and p-value, decided to reject the null or not

- alpha
  - level of significance set by the researcher, a priori, it is the probability of rejecting H_0 when is H_0 true
  - overall alpha level is not effected by one or two tail test. In 2-tailed, the null is harder to reject cuz the value needs to be more extreme to be significant.

- Beta
  - probability of making a type II error, affect by sample size and alpha

- Type I and Type II errors
  - Type I error: Reject H_0 when it is actually true, concludes that there is an effect when there is not an effect
    - p(make Type I error) = alpha ex: alpha = .05, make Type I error 5% of time
    - minimize type I error by using more conservative alpha, tradeoff is greater chance pf type II error
  - Type II error = Fail to reject H_0 when it is false, misses an effect
    - p(making a Type II error) = Beta
    - minimize this error by more liberal alpha (larger)
  - 4 times more conservative about type 1 error
- power
  - probability we will correctly reject the null hypo., 1-Beta, and is influenced by diff. btw sample & pop’n mean, sample size, variability in pop’n, alpha and directionality
  - power ↑ as the difference between 2 pop’n ↑
  - directional test more powerful than non-dir. test

- one- vs. two-tailed
  - one-tailed (directional hypo): probability that an outcome will occur at one end of the sampling distribution
  - two-tailed (non-directional hypo): outcome will occur at either extreme of a sampling distribution
  - for one-tailed, pay attention to the sign of z, critical z smaller for 1-tailed, need to be more extreme to get significant result)

- directional vs. non-directional probabilities and hypotheses
  - non-directional hypotheses: If null and alternative hypo involve a single specific value, the hypo are non-directional.
  - directional hypotheses: If null and alternative hypo both involve a range of values, the hypotheses are directional.

- null and alternative hypotheses
  - null: states that a pop’n parameter is equal to a specific value or a range of values, must include an equal sign
  - alternative: states that the pop’n parameter is not equal to the specific value or range of values, complement of the Null Hypothesis
  - null & alternative are mutually exclusive and exhaustive (include all possible outcomes)
  - hypotheses can be directional or non-directional

- null hypothesis significance testing
  - never directly test alternative, either reject null or fail to

- how to reject the null and fail to reject the null
  - neither H₀ nor H₁ can be proved or disproved, evaluate on the basis of alpha
- Have sufficient evidence to reject the null hypothesis when the probability associated with our outcome (p-value) is less than $\alpha$.
- Comparing probability of our result to $\alpha$ (a priori):
  
  $p \leq \alpha \quad$ Reject $H_0$
  
  $p > \alpha \quad$ Fail to reject $H_0$

**One-sample tests**

- Not trying to estimate the pop’n parameter, try to show the sample is not representative of the population (special/different).

**One-sample Z test for means**

- Pop’n distribution of $X$ is normal and the pop’n variance of $X$ ($\sigma_X^2$) is known.

\[
Z = \frac{\bar{X} - \mu_X}{\sigma_X} \quad \& \quad \mu_X = \mu_X \quad \text{and} \quad \sigma_X = \frac{\sigma_X^2}{n} = \frac{\sigma_X}{\sqrt{n}} \quad \rightarrow \quad Z = \frac{\bar{X} - \mu_X}{\sigma_X} \quad \text{or} \quad Z = \frac{\bar{X} - \mu_X}{\sigma_X} \Rightarrow \frac{\bar{X}}{\sqrt{n}}
\]

**One-sample t test for means:**

- Test whether the mean of one variable differs from a constant (e.g., does the mean grade of 72 for a sample of students differ significantly from the passing grade of 70)?
- Parametric test assuming a normal distribution, pop’n parameter unknown, when its assumptions are met more powerful than corresponding 2-sample nonparametric tests.
- T-test used when sample sizes are small (e.g., < 30), computation of t differs for independent vs. dependent samples, but inference is the same.

\[
t(df) = \frac{\bar{X} - \mu_X}{\hat{S}_X} \leftarrow \text{unbiased estimate based on sample}
\]

\[
\mu_X = \mu_X \quad \hat{S}_X = \sqrt{\frac{S_X^2}{n}} = \frac{S_X}{\sqrt{n}}
\]

\[
t(df) = \frac{\bar{X} - \mu_X}{\sqrt{\frac{S_X^2}{n}}} \quad \text{or} \quad t(df) = \frac{\bar{X} - \mu_X}{\hat{S}_X} \quad \text{df} = n - 1
\]

**T distribution**

- Like z-distribution, theoretical symmetrical, mean=0, but SD depends on the sample size.
  - The larger the SD, the greater the error in predicting the pop’n mean of means.
  - $\downarrow$ SD, $\text{df} \uparrow$

- With a smaller sample size, have a platykurtic distribution.
  - As the sample size $\uparrow$, $t_{\text{crit}} \downarrow$

- There’s a different t distribution for each df, as df $\uparrow$, t approaches standard normal Z distribution.

**Probability testing and creating confidence intervals**

\[
t(df) = \frac{\bar{X} - \mu_X}{\sqrt{\frac{S_X^2}{n}}} \quad \text{or} \quad t(df) = \frac{\bar{X} - \mu_X}{\sqrt{\frac{S_X^2}{n - 1}}}
\]

**Degrees of freedom:**
- # of variables left open to vary

- one-sample t test for correlations
  - rho = 0, normal distr., use t test cuz variance of the sampling dist. of $r_{xy}$ is estimated

- one-sample z test for correlations
  - rho = any value but 0, distribution skewed, use r to z transformation, eq see next def

- Fisher r to z transformation
  - used for correlations, since can’t use the Central Limit Theorem, b/c it’s a sampling distribution of r, not means
  - convert the r’s to Zr’s, correct the skew and get a good estimate of the variance
  - creates a z-distribution and shows us where our sample’s r stands
  - $r_{xy}$ to $Z_r$, rho $xy$ to $Z_{r_{xy}}$ to $Z_{r_{xy}}$

1. When should you use each of the following tests? Be specific.

1) Independent Samples z test
   - to determine if the samples were drawn from two different populations or from the same population
   - If the population distribution of X is not normal, we rely on the central limit theorem
     - pop’n distribution of is normal and the pop’n variances are known
     $$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_{X_1} - \mu_{X_2})}{\sigma_{X_1 - X_2}}$$
     $$\text{Mean: } \mu_{X_1 - X_2} = \mu_{X_1} - \mu_{X_2}$$
     $$\text{Variance: } \sigma_{X_1 - X_2}^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 = \frac{\sigma_{X_1}^2}{n_1} + \frac{\sigma_{X_2}^2}{n_2}$$
     (sq it to get eq for std. dev.)

2) Independent Samples t test:
   - used to compare the means of two independently sampled groups ex., do those working in high noise differ on a performance variable compared to those working in low noise
   - pop’n variances are not known, estimates variance and the std deviation
   - calculation set table: $X_1$, $X_1^2$, $X_2$, $X_2^2$, sum of each, mean of $X_1$, mean of $X_2$
   - $t = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_{X_1} - \mu_{X_2})}{\hat{S}_{X_1 - X_2}}$
     - The pop’n diff. is zero
     - df = $(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$
     - Estimated variance, sq. root for std. dev.
     - when $n_1 = n_2$, sq. root for std. dev.
     $$\hat{S}_{X_1 - X_2}^2 = \frac{(SS_{X_1} + SS_{X_2})}{(n_1 - 1) + (n_2 - 1)} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$$
     $$\hat{S}_{X_1 - X_2}^2 = \hat{S}_{X_1}^2 + \hat{S}_{X_2}^2 = \frac{n_1}{n} \hat{S}_{X_1}^2 + \frac{n_2}{n} \hat{S}_{X_2}^2$$
Computational Indep t test for unequal or equal n

\[
\begin{align*}
  t &= \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{x_1} - \mu_{x_2})}{\sqrt{\frac{\sum X_1^2 - (\sum X_1)^2}{n_1} + \frac{\sum X_2^2 - (\sum X_2)^2}{n_2}} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}
\end{align*}
\]

Indep t test for equal n

\[
\begin{align*}
  t &= \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{x_1} - \mu_{x_2})}{\sqrt{\frac{\sum X_1^2 - (\sum X_1)^2}{n} + \frac{\sum X_2^2 - (\sum X_2)^2}{n}}} \frac{1}{n(n-1)}
\end{align*}
\]

For both: \(df = (n_1 + n_2) - 2\)

3) Dependent Samples t test (correlated t test for diff. btw 2 means)
- compare means where the two groups are correlated, as in before-after, repeated measures, matched-pairs
- benefit: control indiv. differences, ↓ standard error of differences, which ↑ power (more than indept t-ratio)
- method: find the differences (D) btw the scores by each pair of correlated scores, then treat each D as a raw score to conduct a one-sample t test on the mean of the differences
  - When setting up the table, look @ alternative to match the final sign of D, critical t have to match the hypo.
  - Table setup: \(X_1, X_2, D, D^2\), sum of each

\[
\begin{align*}
  t &= \frac{\bar{D} - \mu_D}{\hat{S}_D} = \frac{\bar{D} - \mu_D}{\frac{\hat{S}_D^2}{\sqrt{n}}}
\end{align*}
\]

where \(df = n - 1\)

\[
\begin{align*}
  \bar{D} &= \frac{\sum D}{n} = \text{sample mean of the difference scores}
  \\
  \mu_D &= \mu_D = \text{population mean of difference scores}
  \\
  \hat{S}_D &= \frac{\hat{S}_D^2}{\sqrt{n}} = \text{standard error of the mean difference}
  \\
  n &= \text{number of pairs (difference scores)}
\end{align*}
\]

4) Mann Whitney U test
- 2 independent samples, ordinal DV (or I/R DV with violated parametric assumptions; DV is reduced from I/R to ordinal)
- to det if the 2 distributions of scores are the same (null hypo) vs. diff (alternative hypo) in terms of location.
- when 2 distributions of scores are symmetrical, the Mann-Whitney U is a test of the null hypo that the pop'n medians are equal vs. the alternative hypo that the pop'n medians are not mentioned are non-directional but hypo can also be directional, stated in terms of DV
- Steps 1) rank all scores from highest to lowest or reverse, when ≥ ppl have the same score, assign the average of the ranks to those people
  2) add the sum of the ranks for each group
  3) Assign R1 to group with greater rank sum
- Table R: \(\alpha, N1\) (# of subj. in the grp), N2, and tailed (1 or 2), to reject null, \(U_{\text{observed}}\) must
equal or fall outside these limits.

Formulas
\[ U = N_1N_2 + \frac{N_1(N_1 + 1)}{2} - R_1 \]
\[ U' = N_1N_2 + \frac{N_2(N_2 + 1)}{2} - R_2 \]

5) Wilcoxon Matched Pairs Signed Ranks T test:
- referred to as Wilcoxon T test
- nonparametric analog to the t-test for variables with an even number of categories (data are obtained via repeated-measures or matched-pairs design)
- similar to Mann-Whitney U, except Wilcoxon T is for 2 correlated samples or ordinal DV (I/R DV with violated parametric assumptions; DV is reduced from I/R to ordinal)
- As with the Mann-Whitney U, used to determine if the 2 distributions of scores are the same (null hypo) vs. different (alternative hypo) in terms of location
- Table: X1, X2, D, Ranks (ignore sign), Neg Rank, Pos. rank, sum
- Steps:
  1) obtain diff btw paired scores
  2) If pair has difference of zero, drop it from the analysis & reduce N accordingly
     ➢ Rank differences from lowest to highest, ignoring sign of the difference
  3) sum ranks separately for positive diff & for negative diff
  4) smaller sums is the observed T statistic
  5) critical value for the T statistic in Table S: alpha, N (# of difference scores), tail (1 or 2)
  6) If \( T_{obs} \leq T_{crit} \), Reject H0

6) Chi-Square test (both the Goodness of Fit test and the Test of Independence)

**Goodness of Fit test**
- nominal, one variable, mutually exclusive categories, independent, use Table Q (\( \alpha \) and df)
- non-directional hypo, 2-tailed
- \( \chi^2 \) observed and \( \chi^2 \) critical always positive
- Table: Oi, Ei= N/k, Oi – Ei, (Oi – Ei)^2/ Ei, sum of each

\[ \chi^2_{observed} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \text{ df } = k - 1, k = \text{ number of categories} \]

Oi = Obs freq for category i (data) ex: # of children choose a toy
Ei = Expected freq for category i (you calculate): ex: null expect equal # for all, if 30 total 5 grp 6 in each
In general, Ei = \( \pi_i \times N \)
If the probability the same for all categories, then \( \pi_i = 1/k \) for each category and Ei = 1/k x N = N/k for each category.
If the probability is not the same for all categories, then \( \pi_i \) must be determined for each category

**Test of Independence**
- to determine if 2 categorical variables are independent vs. related
- null hypothesis states that the 2 variables are independent
  ➢ Probabilities for cells (categories) are obtained under the assumption that null hypothesis is true (i.e., that the 2 variables are independent)

\[ \chi^2_{observed} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \text{ df } = (r - 1)(c - 1) \]
where \( r \) = \# of rows and \( c \) = \# of columns
for each cell, \( E_{ij} \) diff. probability (cross chart product/N)

- \( O_{ij} = \) Observed frequency (data)
- \( E_{ij} = \) Expected frequency (must calculate) = \( p_{ij} \times N = \frac{R_iC_j}{N}\),
  where \( R_i = \) row total for row \( i \) and \( C_j = \) column total for column \( j \).
  - Recall from probability that \( P(A \text{ and } B) = P(A) \times P(B) \) when events are independent.
  - In this case, \( p_{ij} = P(R_i \text{ and } C_j) = P(R_i) \times P(C_j) \).
  - Therefore, \( E_{ij} = P(R_i) \times P(C_j) \times N \).
  - Since \( P(R_i) = \frac{R_i}{N} \) and \( P(C_j) = \frac{C_j}{N} \), we get \( E_{ij} = \frac{R_iC_j}{N} \).

7) Analysis of Variance (ANOVA):
- alternative to test difference of means between independent samples (one-way ANOVA) or testing the difference of means for paired samples (repeated measure ANOVA)
- when your IV is continuous, you use regression, when it’s categorical you use ANOVA
  - Each IV (factor) consist \( \geq \) levels
  - When samples are independent, ANOVA conducted is between-subjects ANOVA.
    - When samples are correlated, ANOVA conducted is either within-subjects or repeated-measures ANOVA.
  - Hypo. non-directional, stated in words
- used to uncover the main and interaction effects of categorical independent variables (called “factors”) on an interval dependent variable
- It is not a test of differences in variances, but rather assumes relative homogeneity of variances
- not to describe individual scores, to see if differences between groups is significant
- ANOVA F-test is a function of (variance): size of the difference between group means, sample sizes in each group (larger sample sizes give more reliable info), variances of the dependent variable in each group
- component of variance: sum-of-squares (\( SS_T \)), Between-groups sum-of-squares (\( SS_B \)) and Within-groups sum-of-squares (\( SS_w \))
  - \( SS_T = SS_B + SS_w \) \( df_T = df_B + df_w \)
  - mean squares (\( k = \# \) of grps, \( N = \# \) of subjects) always positive, where \( MS_w = \) pure error, \( MS_B = IV + error \)

\[
MS_B = \frac{SS_B}{df_B} = \frac{SS_B}{k-1} \quad MS_w = \frac{SS_w}{df_w} = \frac{SS_w}{N-k} \quad df_B = k - 1 \quad df_w = N - k \\
\]

- \( F \) always positive, use Table J \((df_B/df_w)\), light row \( \alpha = .05 \), dark row \( \alpha = .01 \)
  - \( F = (\text{treatment variance} + \text{error variance}) / \text{error variance} \)
  - \( F_{observed} = \frac{MS_B}{MS_w} = \frac{\text{Between - groups variance estimate}}{\text{Within - groups variance estimate}} \)
- If \( F \) sig. proceed to HSD

\[
HSD = q_{critical} \sqrt{\frac{MS_w}{n}} \quad -q_{critical} \text{ from Table L, need } k, \alpha, df_w \\
- n = \# \text{ of scores per group} \]

- \( df_T = N - 1 \)
2. What are the **assumptions** of each of the following tests? Be specific.

1) **Independent Samples z test**
   - interval-level or ratio-level DV
   - approx. normal distribution (use z-score to show relative standing)
   - test hypotheses about specific population parameters
   - homogeneity of variance (equal population variances) in certain situations involving more
   than one population

2) **Independent Samples t test**
   - interval-level or ratio-level DV
   - approximately normal distribution of the measure in the two groups
   - roughly similar variances: homogeneity of variance
   - observation in each sample are indept. of e.o

3) **Dependent Samples t test (both the repeated measures and the matched-pairs)**

4) **Mann Whitney U test**
   - parametric assumptions violated, test are about entire distributions of scores, not specific
   parameters
   - ordinal DV, 2 independent samples (or I/R DV with violated parametric assumptions; DV is
   reduced from I/R to ordinal)
   - when 2 distributions of scores are symmetrical, the Mann-Whitney U is a test of the null hyp
   that the pop'n medians are equal vs. the alternative hypo that the pop'n medians are not equal.
   - mentioned are non-directional but hypo can also be directional

5) **Wilcoxon Matched Pairs Signed Ranks T test**
   - parametric assumptions violated, test are about entire distributions of scores, not specific
   parameters
   - ordinal DV, 2 correlated samples

6) **Chi-Square test (both the Goodness of Fit test and the Test of Independence)**
   - both: parametric assumptions violated, test are about entire distributions of scores, not specific
   parameters
   - Goodness of Fit test: omnibus test to det. if observed matches hypothesized (expected)
   frequency distribution [null hypo. says match & is non-directional], mutually exclusive
   categories
     - Categories are mutually exclusive, indiv must belong in only one category
     - Observations are independent: probability of being placed in a category is not affected
       by the categorization of other individuals
     - Sufficient sample size:
       - If df is greater than one (df > 1), none of the expected frequencies is less than 5
       - If df is equal to one (df = 1), none of the expected frequencies is less than 10
   - Test of Independence

7) **Analysis of Variance (ANOVA)**
   - sampling error is distributed normally around the mean in each treatment population (w/in)
   - homogeneity of variances: similar variances on the dependent variable (outer)
3. What does an omnibus test tell us? Which tests are omnibus tests? Why would we want to conduct a follow up (post hoc) test after we get a significant omnibus test statistic?

- Omnibus test: use to compare difference between many means simultaneously.
  - ANOVA, Chi-Square test
- Conduct post hoc cuz want to know which means are significantly different from the others.
  - HSD (post hocs not as powerful as ANOVA)

4. If you're comparing the effects of an independent variable with 3 or more different levels on the dependent variable, should you do multiple pairwise tests? Why or why not? If not multiple pairwise tests, what should you do instead?

- no, when samples are correlated, then ANOVA conducted is either within-subjects or repeated-measures ANOVA.
- conduct between-subjects ANOVA

5. In an ANOVA, to obtain a significant F statistic, what should the relationship be between the within-groups variance and the between-groups variance? What are within- and between-groups variances?

- Significant F when between-groups variance is greater than within-groups variance
- Between-groups variance: how much variance there is between all the group means. Influenced by treatment effect and random error.
- Within-groups variance: how much variance there is due to random error

6. How do you compute expected frequencies?

7. How do you determine the critical region for the following tests?

1) Independent Samples z test:
- set the boundary of alpha, compare our sample score as a specific spot where our alpha is.
- critical z smaller for 1-tailed, need to be more extreme to get significant result

2. For 1-tailed tests, check the signs:
   a) If \( |\text{observed}| > |\text{critical}| \), Reject H0
   b) If \( |\text{observed}| < |\text{critical}| \), Fail to Reject H0

1. For 2-tailed tests
   a) If \( |\text{observed}| > |\text{critical}| \), Reject H0
   b) If \( |\text{observed}| < |\text{critical}| \), Fail to Reject H0

2) Independent Samples t test
- Table C (alpha, tail, df) \( t_{\text{obs}} > t_{\text{crit}} \) difference of means is significant at that level (alpha)
- probability of our result (observed t) is less than or equal to alpha \( (p \leq \alpha) \), reject H0

3) Dependent Samples t test (both the repeated measures and the matched-pairs)
- Table H \( (r_{xy}, \rho_{xy}) t_{\text{obs}} > t_{\text{crit}} \) difference of means is significant at that level (alpha)
4) Mann Whitney U test  
- Table R (alpha, N1, N2, 1 or tailed), to reject null, $U_{\text{observed}}$ must equal or fall outside these limits.

5) Wilcoxon Matched Pairs Signed Ranks T test  
- Table S (alpha, N, tail), If $T_{\text{obs}} \leq T_{\text{crit}}$, Reject $H_0$

6) Chi-Square test (both the Goodness of Fit test and the Test of Independence)  
- Table Q (alpha, df) Goodness of Fit: $\chi^2_{\text{observed}} \geq \chi^2_{\text{critical}}$, reject $H_0$

7) Analysis of Variance (ANOVA)  
- Table J (dfB, dfw, apha) $F_{\text{obs}} > F_{\text{crit}}$ Reject $H_0$, means at least one diff, btw grps, proceed to post-hoc test – HSD (table L)  
  ➢ statistically sig if it equals or exceeds the HSD.

If you're given an observed statistic for one of the tests, be able to determine if it is significant and what it means.

8. What are we plotting when we create a sampling distribution for the difference between 2 means?  
- The difference between 2 sample means (lab 9).

9. Be able to go the steps of hypothesis testing for all the tests.

10. Compare the advantages and disadvantages of independent samples t-tests and dependent samples t-tests.  
- Dependent t test control for individual differences.

11. What are the factors that affect significance of a two-sample test?  
- sample size, variability in the pop’n (larger std. error, obv. t smaller), diff. btw pop’n means

12. Know the degrees of freedom that go along with each test.  
- t test: df= n-1  
- correlated t test (rho= 0): df= n-2  
- Independent t (difference btw means): df = $(n_1-1)+(n_2-1) = n_1 + n_2 - 2$  
- Dependent t (correlated t for diff btw means): df = n - 1  
- ANOVA (sum of squares SS$_T$): df$_T$ = df$_B$ + df$_w$ = N-1  
- ANOVA (mean sq. MS): df$_B$ = k - 1  
  df$_w$ = N - k  
  df$_T$ = N - 1  
- Chi-Sq (goodness of fit): df = k - 1  
- Chi-Sq (test of independent): df = $(r - 1) (c - 1)$

13. How do we interpret our findings from a one-way ANOVA?  
- When conducting a one-way ANOVA, we identify 3 forms of variance: total variance, between-groups variance, within-groups variance. We can identify if there is at least one significant difference between group means.

14. What are main effects and interactions in ANOVA? How do we know if we have either one?
In two-way ANOVA, we can examine simultaneously the effects of 2 factors. Allows the examination of main effects and interactions. Advantages:

- Study effects of 2 variables with fewer subjects
- Partition effect for each indept var. and variance due to interaction of the 2 factors
- Greater power than single factor experiments, ↓ estimate of the variance due to error, which ↑F-ratio

- Main effects: main effect reflects the effect of a single factor independent of the other factors and interactions (direct effect of an ID var on the DV)
  - probability of F < .05 for any independent, variable does have an effect on the dependent
- Interactions: represent individual variables cannot be considered in isolation when discussing the relation to the dependent variable (joint effect of ≥ ID var on the DV)
  - probability of F < .05 for any such combination, interaction of the combination does have an effect on the dependent

Others:
- value for critical region reflect type of null hypo selected
- type of sampling distribution & significance level det. critical region
- observed value is the result of the statistical test performed, determined by your data.
- critical value determined by alpha, which is set by the experimenter, a priori
- sample size ↑, accuracy of pop'n estimate ↑, spread of distribution ↓ (SD), smaller the diff.
  - between the means has to be to be stat. sig.
    - less overlap of sampling distribution, greater power
- Type I error (approx)= 1-(1- α)^c [c is the number of tests conducted]
  - reduced alpha level for each test, ↓ power (1-Beta)
- any score w/in grp= pop’n mean + T + E = \( X_{ij} = \mu + \alpha_j + \epsilon_{ij} \)

- F-test, also called the F-ratio. The F-test is an overall test of the null hypothesis that group means on the dependent variable do not differ. It is used to test the significance of each main and interaction effect
  - If the computed F score is greater than 1, then there is more variation between groups than within groups, from which we infer that the grouping variable does make a difference. If the F score is enough above 1, it will be found to be significant in a table of F values. If the computed F value is around 1.0, differences in group means are only random variations
  - Two-Way ANOVAs: factorial design ex: 2X3 = 2 levels of one variable, 3 levels of the other= 6 combinations
  - Compare mean w/in IV (ex: high vs. low exp), if diff then there's an effect
  - Additive effect: both IVs to demonstrate main effects at the same time, two parallel lines
  - How to spot interactions: See if pattern of effects is different between rows or columns
  - Don’t forget: possible to have an interaction regardless of the pattern of main effects

- 2 IV ANOVA:
  - Calculation steps:
    1) SS for Factor A using SS_B type eq.
    2) SS for Factor B using SS_B type eq.
    3) SS for Factor A & B using SS_B type eq.: [sum of (each group squared)/# subj in each group] – (SS_A + SS_B)
    \( Df_{AB} = (Level\ #\ for\ Factor\ A -1)(\ Level\ #\ for\ Factor\ B -1) \)
    one problem go thru the steps of hypothesis testing, and given some info about a research study, select the appropriate test statistic and compute it.