LO 1. Define analysis of variance (ANOVA) as a statistical inference method that is used to determine - by simultaneously considering many groups at once - if the variability in the sample means is so large that it seems unlikely to be from chance alone.

LO 2. Recognize that the null hypothesis in ANOVA sets all means equal to each other, and the alternative hypothesis suggest that at least one mean is different.

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$

H_A : At least one mean is different

LO 3. List the conditions necessary for performing ANOVA

1. the observations should be independent within and across groups
2. the data within each group are nearly normal
3. the variability across the groups is about equal and use graphical diagnostics to check if these conditions are met.

LO 4. Recognize that the test statistic for ANOVA, the F statistic, is calculated as the ratio of the mean square between groups (MSG, variability between groups) and mean square error (MSE, variability within errors). Also recognize that the F statistic has a right skewed distribution with two different measures of degrees of freedom: one for the numerator ($df_G = k - 1$, where k is the number of groups) and one for the denominator ($df_E = n - k$, where n is the total sample size).

- Note that you won't be expected to calculate MSG or MSE from the raw data, but you should have a conceptual understanding of how they're calculated and what they measure.

LO 5. Describe why calculation of the p-value for ANOVA is always "one sided".

LO 6. Describe why conducting many t-tests for differences between each pair of means leads to an increased Type 1 Error rate, and we use a corrected significance level (Bonferroni correction, $\alpha^\star = \alpha / K$, where K is the number of comparisons being considered) to combat inflating this error rate.

- Note that $K = \frac{k(k-1)}{2}$, where k is the number of groups.

LO 7. Describe why it is possible to reject the null hypothesis in ANOVA but not find significant differences between groups when doing pairwise comparisons.

LO 8. Describe how bootstrap distributions are constructed, and recognize how they are different from sampling distributions.

LO 9. Construct bootstrap confidence intervals using one of the following methods:

- Percentile method: XX% confidence level is the middle XX% of the bootstrap distribution.
- Standard error method: If the standard error of the bootstrap distribution is known, and the distribution is nearly normal, the bootstrap interval can also be calculated as $point\ estimate \pm t^\star SE_{boot}$.

LO 10. Recognize that when the bootstrap distribution is extremely skewed and sparse, the bootstrap confidence interval may not be reliable.