A summary document for all statistical tests Covered in this class

1. Are you testing something about a single variable (examples, average age of customers equals 50 years or proportion of male equals 0.5, etc.)?
    a. **Yes** - then is the variable being tested an interval or continuous (such as age, income, attitude measured on a 7-point scale, etc.)?
        i. **Yes** – then, do you know population variance?
            1. **Yes** – then conduct one-sample z-test.
            2. **No** – then conduct one-sample t-test
    b. **No** -then, is the variable binary such as sex (male/female) or multinomial such as marital status with 5 categories - single, married, widowed, divorced, or separated)?
        i. **Yes** – then, do you want to do the theoretically correct test?
            1. **Yes** – then use Chi-square test
            2. **No** – then *if your variable is binary and you have large sample size* in each group, you may conduct one sample t-test? Note that in this case, if the variable is coded as 1/0, the t-test tests the proportion of 1's.
2. Are you testing relationships and /or patterns etc. between 2 variables?
    a. **Yes**, then *are both variables interval or continuous* (such as age, income, attitude measured on a scale, etc.)?
        i. **Yes**, then if you want to test just the linear (straight-line based) relationship between the two variables, use Pearson correlation or simple regression.
            1. If you want correlation but the relation between the two variables is not linear but monotonic, use Spearman's rank order correlation
            2. **If** you want to test for both linear and non-linear relationships between Y and X using regression, then include powers of X (such as X^2 and X^3 , etc.) along with X in the regression equation.
    b. **Yes**, then are *both variables categorical* (such as marital status and Gender)?
        i. **Yes**, then if you want to find whether *any* association exists between the two variables, conduct Cross-tab/Chi-square test.
    c. **Yes**, then is one variable binary (such as buyer/non-buyer) and the other is interval/continuous (such as income)?
        i. **Yes**, then if you want to test whether the average of the interval variable (income) is different between the two groups of the binary variable (buyer vs. non-buyer), then use 2-sample t-test.
    d. **Yes**, then is one variable multinomial (such as marital status) and the other is interval/continuous (such as income)?
        i. **Yes**, then use ANOVA with the interval variable as dependent and the multinomial variable as a factor (independent) to test if the average of income is different across all the levels of marital status.
            1. If ANOVA is significant, then do appropriate follow-up tests (such as Tukey's, Dunnets's, etc.)

3. Are you predicting/explaining/modeling one target (dependent) variable using many other variables as input (independent) variables?
    a. **Yes**, then if your target (dependent) is interval/continuous use Multiple Regression.
    b. **Yes**, then if your target (dependent) is binary (yes/no) use Logistic Regression.