Chapter 11

- p^ is for categorical data and ybar is for quantitative data
- Normal Model for the sampling distribution of the means:
    - Used when you have the population mean and you want to get information about the sample mean
    - μ = ybar
    - SD(ybar)= $\sigma/\sqrt{n}$
        - Means have a smaller standard deviation than individuals
        - σ is the standard deviation of the population
- If the population is N(μ, σ) then the sample distribution for the means is N(μ, $\sigma/\sqrt{n}$)
- More degrees of freedom, the closer it is to a normal distribution
- Confidence intervals with means:
    - ybar ± t^*_{df}SE(ybar)
    - SE(ybar)= $s/\sqrt{n}$
    - Use T-table
- If S is larger then you can replace σ with s
- Degrees of freedom= n-1
- Find T in the same way as Z, except use standard error instead
- When you are centering the distribution on the sample mean use standard error
    - SE(ybar)= $s/\sqrt{n}$
- T-models are unimodal, symmetric, and bell-shaped
    - Less degrees of freedom means narrower peak and less normal-looking
- T-model assumptions
    - Independence
    - Randomization
    - 10% of population condition
- T-model conditions
    - Nearly normal condition
        - When $n<15$ data must be normal
        - When $15<n<40$ data must be unimodal and reasonably symmetric
        - When $n>40$ it is usually fine unless extremely skewed
        - **HE SAYS $n \geq 25$ IS LARGE ENOUGH**
    - Random
    - Unlike Z tables, T tables go from right to left
- One sample t-test
    - T_{df} = (ybar-μ_0)/ SE(ybar)
    - CENTER IT AT THE HYPOTHESIS, NOT THE SAMPLE MEAN
- When trying to guess a sample size, use the z-value instead of the t-value and if the sample is not very large plug it back in the t-model

- Assume Z: ME= $Z^* \times \sigma/\sqrt{n}$ and calculate n_1
    - If n<60:
        - Then use the n for the T-value and find: ME = t_{n1-1}^* x $s/\sqrt{n_2}$
- Calculator: Stat -> t-statistics

Chapter 12

- When looking at the t-table round down for degrees of freedom
- Two means:
  - $\mu_0 - \mu_n$
  - If the sample means come from independent samples, the varience of their sum or differences is the sum of their variences
  - $SD(\mu_0 - \mu_n) = \sqrt{((\sigma^2/n_1) + (\sigma^2/n_2))}$
  - Two sample t-test: the ratio of the difference in the means from our samples to its standard error and compare that ratio to a critical value from a student's t-model
  - DIFFERENT SAMPLES
    - For standard error always use T-distribution
    - $H_0 = \mu_1 - \mu_2 = \Delta_0$
    - $t = ((ybar_1 - ybar_2) - (\mu_1 - \mu_2)) / SE$
    - $SE(ybar_1 - ybar_2) = \sqrt{((s_1^2/n_1) + (s_2^2/n_2))}$
    - Degrees of freedom: the smaller of n_1-1 and n_2-1
    - Assumptions:
      - Independence
        - Randomization condition
        - 10% condition
      - Nearly normal condition
        - When $n<15$ data must be normal
        - When $15<n<40$ data must be unimodal and reasonable symmetric
        - When $n>40$ it is usually fine unless extremely skewed
        - **In general $n_1 + n_2 > 40$**
      - **Independent groups assumption**
  - Two sample t-interval
    - $(ybar_1 - ybar_2) \pm t^*_{df} \times SE$
    - $SE = \sqrt{((s_1^2/n_1) + (s_2^2/n_2))}$
- For t-tests you can pool the data and combine variances using:
  - $SE_{pooled} = \sqrt{((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2)/ (n_1 + n_2 -2)}$
  - $df = N_1 + N_2 - 2$
  - **Equal variance assumption**: the variances of the two populations from which the samples have been draw from are equal
- Cannot use two sample t-test for paired data because it is not independent
- Paired t-test: one sample t-test for the mean of the pairwise differences (dependent samples)
- SAME SAMPLE
  - H_0: $\mu = \Delta_0$
  - $t = (dbar - \Delta_0) / SE$
  - $SE = s_d / \sqrt{n}$