Lecture 14: August 1, 2022 - Z-score, t-test, and Anova

Hypothesis tests

X^2 (from Friday) and t-test (today) are simple enough to work out with a calculator, so a small data version of one or the other or both is likely to be on the final

- Friday

```
+ $X^2$ (Chi-squared) test
+ For categorical variables
+ aka as cross-tabs because of the format
+ worked through a $X^2$ test problem together
+ This is on Problem Set 5 and final
```

- Today

```
+ Concepts of Z-score, t-test, and ANOVA
+ Where each one is appropriate
+  Brief discussion of the formulas for all three
+ We will work through a paired sample t-test together as our second example of hypothesi
+ Paired t-test is on Problem Set 5 and final
```

Hypothesis test uses

- X^2 (Chi squared): Categorical variables with counts
- t-test: compares the means of two groups
- z-score: continuous, normally distributed variables
- ANOVA (Analysis of Variance)
- Lots of others!

Hypothesis test uses

- X^2 (Chi squared): Categorical variables with counts

```
    + $X^2$ test of goodness of fit - tells whether the sample data is representative of the
    + $X^2$ test of indendependence (we dit this) - tells us if two categorical variables are
```

- t-test: compares the means of two groups

```
+ pairwise comparison
+ one sample: comparing one group against a standard value
+ two-sample or independent t-test: compares two groups from different populations
+ paired t-test: compares a single group as in before and after comparison
+ Two tailed test: tells if they are different, either greater or less
+ One tailed test: tells if one group is specifically greater or less, bot not either
+ *degrees of freedom* is *n - 1*
+ When t-test degrees of freedom $>$ 30, it converges on the z-score
```

- z-score: continuous, normally distributed variables

```
+ Continuous variables
+ normal distribution
+ known population standard deviation
+ "known" ~ accepted estimate
+ Central Limit Theorem can get us to normal
+ Use if: if the population standard deviation is known and sample size > 30
```

- ANOVA (Analysis of Variance)

```
+ tests difference of means between 3 or more indendepnt groups
+ This is often used to test removing variables one at a time from a multi-variable model
+ Uses the *F-test*
+ The same F-test in regression results - model fit
```

Deeper look at t-tests

- Paired sample t-test
- $t = \frac{x_{diff}}{s_{diff}/\sqrt{n}}$