1. Review of hypothesis testing:

a. Variance vs. Standard Deviation vs. Standard Error

Up to this point you've learned variances, standard deviations, and standard errors, which can be a lot to keep track of. There's a few things you should remember: 1) both your underlying data and your estimators have a variance. The variance of your estimator is related to the variance of your underlying data, but it decreases as the number of observations n grow; 2) The square root of a variance of the underlying variable is called the standard deviation, while the square root of the variance of your estimator is called the standard error; 3) We frequently don't observe the true variance σ^2, so we use sample variances s^2 or $\hat{\sigma}^2$.

b. Hypothesis Testing and Confidence Interval Review

Just to help keep all of these tests straight:

Test Type	Test Statistic	Distribution
Population mean e.g. $H_0 : \mu = \mu_0$	$t = \frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}}$	$t \sim t_{n-1}$
Difference in population means e.g. $H_0 : \mu_1 - \mu_2 = \mu_0$	$t = \frac{(\bar{x}_1-\bar{x}_2)-\mu_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	$t \sim t_{n_1+n_2-2}$
Population proportion e.g. $H_0 : p = p_0$	$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$z \sim N(0,1)$
Difference in population proportions e.g. $H_0 : p_1 - p_2 = p_0$	$z = \frac{(\hat{p}_1-\hat{p}_2)-p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}+\frac{\hat{p}(1-\hat{p})}{n_2}}}$	$z \sim N(0,1)$
True regression parameter (k variables) e.g. $H_0 : \beta = \beta_0$	$t = \frac{\hat{\beta}-\beta_0}{SE(\hat{\beta})}$	$t \sim t_{n-k-1}$
Multiple restrictions in regression (q restrictions, k total variables in UR model)	$F = \frac{\left(R^2_{UR}-R^2_R\right)/q}{\left(1-R^2_{UR}\right)/(n-k-1)}$ $F = \frac{(SSR_R-SSR_{UR})/q}{(SSR_{UR})/(n-k_{UR}-1)}$	$F \sim F_{q,n-k-1}$

And one for confidence intervals:

Confidence Interval for:	Standard Error	Distribution of c
Population mean (non-binary) $\left[\bar{x} - cSE(\bar{x}), \bar{x} + cSE(\bar{x})\right]$	$\sqrt{\frac{s^2}{n}}$	t_{n-1}
Difference in population means (non-binary) $\left[\hat{D} - cSE(\hat{D}), \hat{D} + cSE(\hat{D})\right]$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$t_{n_1+n_2-2}$
Population mean/proportion (binary) $\left[\hat{p} - cSE(\hat{p}), \hat{p} + cSE(\hat{p})\right]$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$z \sim N(0,1)$
Difference in population proportions (binary) $\left[\hat{D} - cSE(\hat{D}), \hat{D} + cSE(\hat{D})\right]$	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$z \sim N(0,1)$
Regression population parameter $\left[\hat{\beta} - cSE(\hat{\beta}), \hat{\beta} + cSE(\hat{\beta})\right]$	$\sqrt{\frac{SSR/(n-k-1)}{[SST_j(1-R_j^2)]}}$	t_{n-k-1}

c. Interpretation of $\hat{\beta}$ significance

- Example with reject: say the null and alternative hypotheses are $H_0 : \beta_j = 0$ and $H_1 : \beta_j > 0$ respectively, and I reject H_0 based on the p-value, and find $\beta_j > 0$. Then I would conclude: I reject H_0 in favor of the alternative H_1, and $\hat{\beta}_j$ is statistically greater than zero at the α% significance level (when we reject a null such as $H_0 : \beta_j = 0$, we usually say our variable x_j is statistically significant). You should then add "this suggests that the variable x_j has a statistically significant and positive relationship with y **holding all other variables** constant.

- Example with fail to reject: say the null and alternative hypothesis are $H_0 : \beta_j = -1$ and $H_1 : \beta_j \neq -1$ where β_j is the effect of air pollution on housing prices. We estimate the effect to be $\hat{\beta}_j = -0.954$. Doing all the math, we get a t-stat of 0.393, and we see that we cannot reject H_0. But there are many other values for β_j that we cannot reject. For example, if our null was $H_0 : \beta_j = -0.9$, we would get a t-statistic of -0.462, which cannot be rejected either. But $\beta_j = -1$ and $\beta_j = -0.9$ can't both we true. So it makes no sense to say that we "accept" either of these hypothesis. All we can say is that the data do not allow us to reject either of these hypotheses at the 5% (or indeed the 10%) significance level, or that there is not statistical evidence against the null hypothesis that $\beta_j = -1$ or $\beta_j = -0.9$. So to conclude, you would say "I fail to reject the null in favor of the alternative. There is no statistical evidence that the impact of air pollution on *prices* is not -1 **holding all other variables** constant.

d. Using R output for other β hypothesis tests

R output gives t-statistics and p-values for two-sided $H_0 : \beta = 0$. What if you want to test a different hypothesis? You have all the information you need!

- $t = \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})}$
- Plug in $\hat{\beta}$ and $SE\left(\hat{\beta}\right)$ from the regression output, and put in whatever your β_0 is under H_0
- Compare to the appropriate critical value, whether two-sided or one-sided

2. Introducing p-values

To date we have selected a specific significance level (significance level = the probability of rejecting the null when the null is true), which then gives us a critical value and a rejection region. This means that in theory two different economists could select different significance levels to test the same hypothesis (because of their different tolerance for rejecting the null when the null is true), and wind up with different conclusions.

- Ex: We have some data, and we want to test the null hypothesis that some parameter is zero. With 40 degrees of freedom, we obtain a test statistic of 1.85. One researcher chooses a signficance level of 5% with an associated critical value of 2.021. He doesn't report his t-statistic, and he fails to reject! Alternatively, another researcher chooses the 10%, with an associated critical value of 1.68. He doesn't report his t-statistic, and he rejects!

We can also do hypothesis testing by using p-values. I provide 3 definitions that say the same thing using different words (the reason being that p-values are tricky and different definitions can appeal to different people)

Definition. *The p-value is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis.*

Definition. *The p-value is the smallest significance level at which the null hypothesis would be rejected.*

Definition. *The p-value is the probability of obtaining a value of the test statistic as extreme or more extreme than the one actually obtained from the sample under the null (i.e if the null is true).*

Therefore small p-values close to zero constitute strong evidence against the null H_0 whereas large p-values close to one constitute weak evidence against the null H_0. The decision rule is as follows:

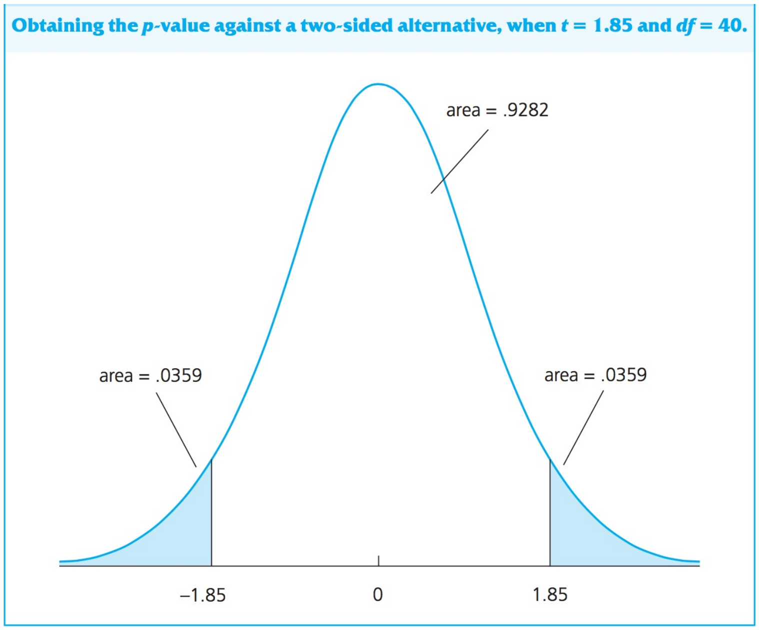
- H_0 is rejected if p-value $< \alpha$
- H_0 is not rejected if p-value $> \alpha$

Example: Suppose we calculate a test statistic of $t = 1.85$, with 40 degrees of freedom (two sided). We can find the p-value

$$pvalue = P(|T| > 1.85|H_0) = 2P(T > 1.85) = 2(0.0359) = 0.0718$$

where $P(T > 1.85)$ is the area to the right of 1.85 in a t-distribution with 40 degrees of freedom. This means that under the null hypothesis, we would observe an absolute value of the t-statistic as large as 1.85 about 7.2% of the time. Knowing this p-value we can carry out the test at any level. Indeed, if we chose a significance level of 5% then we fail to reject H_0. Conversely, if we chose a significance level of 10%, we will reject H_0. In other words if we decide that our tolerance for rejecting the null when the null is true is 5%, and the probability we get a test statistic more extreme than the one in our sample is 7.2%, then we fail to reject the null because we aren't comfortable with "getting it wrong" more than 5% of time.

It is helpful to have the following picture from Woolridge (p.777) in mind:



If the alternative is $H_1 : \beta_j < 0$ then we want the p-value $= P(T < t) = P(T > t)$ because the t-distribution is symmetric about 0. This can be obtained as one-half of the p-value for the two tailed test.

3. Hypothesis test about a single linear combination of parameters

Take the following population model we saw in class:

$$log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$$

where:

$$\begin{aligned} jc &= \text{number of years attending a two-year college} \\ univ &= \text{number of years at a four year college} \\ exper &= \text{months in the workforce} \end{aligned}$$

We want to test whether one year of junior college is worth one year at a university. So our null:

$$H_0 : \beta_1 = \beta_2 \quad H_1 : \beta_1 \neq \beta_2$$

We can rewrite the test statsistic

$$H_0 : \beta_1 - \beta_2 = 0 \quad H_1 : \beta_1 - \beta_2 \neq 0$$

Then our test-statistic becomes:

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)}$$

This is looking very similar to what we did with the difference in means. The issue here is that the formula for $se(\hat{\beta}_1 - \hat{\beta}_2)$ is complex. Some statistical packages have features that allow you to obtain this formula, and then calculate the test statistic.[1] What we did in class, and detail here, is

[1]In R you can use the "car" package to run the command: "linearHypothesis" like so: *linearHypothesis(modelr, "jc = univ")*.

an approach that is simpler, and virtually any statistical package can do it. Define a new parameter $\theta_1 = \beta_1 - \beta_2$. Then our null becomes:

$$H_0 : \theta_1 = 0 \quad H_1 : \theta_1 \neq 0$$

Because $\theta_1 = \beta_1 - \beta_2$, we can also write $\beta_1 = \theta_1 + \beta_2$. Plugging this into our population regression equation:

$$\begin{aligned}
lwage &= \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exp + u \\
&= \beta_0 + (\theta_1 + \beta_2) jc + \beta_2 univ + \beta_3 exper + u \\
&= \beta_0 + \theta_1 jc + \beta_2 (jc + univ) + \beta_3 exper + u \\
&= \beta_0 + \theta_1 jc + \beta_2 (totcoll) + \beta_3 exper + u \qquad \text{Defined } totalcoll = jc + univ
\end{aligned}$$

We can estimate this model! We can easily get the standard error of $\hat{\theta}_1$, and compute the t-statistic, and reject/fail to reject the null.

4. Hypothesis test for multiple parameters: The F-Test

Now suppose we estimate a slightly different model from above:

$$lwage = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exp + \beta_4 sublg + \beta_5 lgcity + u$$

where *sublg* is an indicator for whether an individual works in a suburb of a large city and *lgcity* is an indicator for whether an individual works in a large city .

Suppose we want to test that both β_4 and β_5 are zero, i.e. that in general working in or near a big city has no effect on wages. If we did these two individual t-tests, we'd see that one is statistically different from zero and one isn't. So do we reject or fail to reject the null hypothesis that *both* are equal to zero? The F-test will tell us.

When setting up the F-test, we will follow all the same steps as before. However, there are a few caveats we introduce first:

1. The null is stated as :

$$H_0 : \beta_4 = 0 \;\;\&\;\; \beta_5 = 0$$

   If we apply these two restrictions we obtain a **restricted model** (where the restrictions are given by the values under the null). The original regression equation is then referred to as the **unrestricted model**:

$$\begin{aligned}
\text{Unrestricted model: } & lwage = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exp + \beta_4 sublg + \beta_5 lgcity + u \\
\text{Restricted model: } & lwage = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exp + u
\end{aligned}$$

2. To calculate the F-test we will need to estimate both these models separately and record the associated SSR. The two formulas we have for the F-stat are:

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n - k_{UR} - 1)}$$

$$F = \frac{\left(R^2_{UR} - R^2_R\right)/q}{\left(1 - R^2_{UR}\right)/(n - k_{UR} - 1)}$$

Where q is the number of restrictions. As always, n is the number of observations, which should be the same for both regressions. k is the number of parameters in the unrestricted model
The easiest way to remember the F-stat formula, is to remember that the F is measuring the relative increase in the SSR by moving from the unrestricted to the restricted model.

3. The F-stat is always positive. If you compute a negative F on the exam, check your math. Because OLS estimates are chosen to minimize the sum of squared residuals, the SSR always increases when variables are dropped from the model (this is algebraically true)

4. A large F statistic means that the unrestricted model has a lot more explanatory power than the restricted model (this becomes apparent in the formula for the F-stat). So essentially the F statistic is helping us decide whether the increase in the SSR in going from the unrestricted model to the restricted model is large enough to warrant rejecting the null hypothesis.

Below are the R results for the first UR model , and the results for the second R model:

```
Call:
lm(formula = lwage ~ jc + univ + exper + sublg + lgcity, data = twoyear)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.09025 -0.27929  0.00423  0.28029  1.79032 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.4656944  0.0211152  69.414  < 2e-16 ***
jc          0.0657284  0.0068181   9.640  < 2e-16 ***
univ        0.0762304  0.0023071  33.041  < 2e-16 ***
exper       0.0049255  0.0001572  31.334  < 2e-16 ***
sublg       0.1008531  0.0186367   5.412 6.46e-08 ***
lgcity      0.0181336  0.0179366   1.011    0.312    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4293 on 6757 degrees of freedom
Multiple R-squared:  0.2258,    Adjusted R-squared:  0.2253 
F-statistic: 394.2 on 5 and 6757 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = lwage ~ jc + univ + exper, data = twoyear)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.10362 -0.28132  0.00551  0.28518  1.78167 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.4723256  0.0210602  69.910   <2e-16 ***
jc          0.0666967  0.0068288   9.767   <2e-16 ***
univ        0.0768762  0.0023087  33.298   <2e-16 ***
exper       0.0049442  0.0001575  31.397   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4301 on 6759 degrees of freedom
Multiple R-squared:  0.2224,	Adjusted R-squared:  0.2221 
F-statistic: 644.5 on 3 and 6759 DF,  p-value: < 2.2e-16
```

Step 1: Define hypotheses:

$$H_0 : \beta_4 = 0 \quad \& \quad \beta_5 = 0$$
$$H_1 : \beta_4 \neq 0 \quad \&/or \quad \beta_5 \neq 0$$

Step 2: Compute test statistic:

The two formulas we have for the F-stat are:

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n - k_{UR} - 1)}$$
$$F = \frac{\left(R^2_{UR} - R^2_R\right)/q}{\left(1 - R^2_{UR}\right)/(n - k_{UR} - 1)}$$

Where q is the number of restrictions: we're testing (1) $\beta_4 = 0$ and (2) $\beta_5 = 0$, so in this example $q = 2$. As always, n is the number of observations, which should be the same for both regressions. We can calculate SSR and R^2 using R.

```
> SSR_U<-sum(modelur$residuals^2)
> SSR_U
[1] 1245.106
> SSR_R<-sum(modelr$residuals^2)
> SSR_R
[1] 1250.544
> summary(modelur)$r.squared
[1] 0.225823
> summary(modelr)$r.squared
[1] 0.222442
> nobs(modelur)
[1] 6763
> nobs(modelr)
[1] 6763
```

Let's check that we know what to plug in for each formula and verify that they will equal the same thing:

$$
\begin{aligned}
SSR_R &= 1250.544 \\
SSR_{UR} &= 1245.106 \\
R^2_R &= 0.222442 \\
R^2_{UR} &= 0.225823 \\
n &= 6763 \\
k_{UR} &= 5 \\
q &= 2
\end{aligned}
$$

Then

$$
\begin{aligned}
F &= \frac{(1250.544 - 1245.106)/2}{1245.106/(6763 - 5 - 1)} = 14.75 \\
F &= \frac{(0.225823 - 0.222442)/2}{(1 - 0.225823)/(6763 - 5 - 1)} = 14.75
\end{aligned}
$$

Step 3: Choose significance level (α) and find the critical value:

Let's use the standard significance level in economics, 5%. To find the critical value here, we need the numerator degrees of freedom and the denominator degrees of freedom:

$$
\begin{aligned}
\text{Numerator d.o.f.} \quad & q = 2 \\
\text{Denominator d.o.f.} \quad & n - k_{UR} - 1 = 6763 - 5 - 1 = 6757 \\
& c_{.05} = 2.997
\end{aligned}
$$

Then we need to know that the F statistic is distributed as an F random variable with $(q, n-k-1)$ degrees of freedom:

$$F \sim F_{q,n-k-1}$$

The distribution of $F_{q,n-k-1}$ is readily tabulated and available in statistical tables/online calculator.

Step 4: Reject the null hypothesis or fail to reject it: Rejection rule:

$$F > c$$

If our F is large, larger than our critical value, this means that the R^2_{UR} is relatively higher than R^2_R. In this example, that means that including the large city proximity variables greatly increases the amount of variation in log wages explained by the model. If our F isn't large, it means that the R^2 for the two models are pretty close to each other, and the large city variables don't have a lot of explanatory power.

With this F-stat and this critical value, do we Reject the null or Fail to reject the null?

Step 5: Interpret

If we reject the null in favor of the alternative: There is statistical evidence at the 5% level that, conditional on years at a two year college, years at a four year college, and months in the workforce, working in a suburb of a large city or working in a large city affects wages. In other words we say that working in a suburb of a large city and working in a large city are jointly statistically significant.

If we fail to reject the null in favor of the alternative: There is no statistical evidence at the 5% level that, conditional on years at a two year college, years at a four year college, and months in the workforce, working in a suburb of a large city or working in a large city affects wages. In other words we say that the variables are jointly insiginficant, which often justifies dropping them from the model.

Conclusion about F-stat

A final word about F-statistics from Woolridge p.148: "the F statistic is often useful for testing exclusion of a group of variables when the variables in the group are highly correlated. For example, suppose we want to test whether firm performance affects the salaries of chief executive officers. There are many ways to measure firm performance, and it probably would not be clear ahead of time which measures would be most important. Since measures of firm performance are likely to be highly correlated, hoping to find individually significant measures might be asking too much due to multicollinearity (the partial effect will be difficult to uncover). But an F-test can be used to determine whether, as a group, the firm performance variables affect salary."