Announcements

- Don't forget to put your Homework Code on your homework. Codes linked to course website. A few of you might receive an email from me changing your code.
- Today:
  - Additional explanation of normal probability plots (AKA q-q plots), on white board
  - Review hypothesis testing
  - Learn about chi-square, F and *t* distributions
  - Learn how to test regression slope (on white board)

General Steps of Hypothesis (Significance) Testing

Steps in Any Hypothesis Test

1. Determine the null and alternative hypotheses.
2. Verify necessary data conditions, and if met, summarize the data into an appropriate test statistic.
3. Assuming the null hypothesis is true, find the *p*-value.
4. Decide whether or not the result is statistically significant based on the *p*-value.
5. Report the conclusion in the context of the situation.

Rejection Region Approach (instead of p-values)

Alternative Step 3:

Find a *rejection region* instead of a *p-value*

Alternate Step 4:

The result is *statistically significant* if the test statistic falls into the rejection region.

Step 1 – general: Determine the hypotheses.

- **Null hypothesis**—hypothesis that says nothing is happening, status quo, no relationship, chance only, parameter equals a specific value (called "null value").
- **Alternative (research) hypothesis** —hypothesis is usually the reason data being collected; researcher suspects status quo belief is incorrect or that there is a relationship or change, or that the "null value" is not correct.

Step 2. Collect data and summarize with a test statistic.

Decision in hypothesis test based on single summary of data – the **test statistic.** Often this is a ***standardized version of the point estimate***.

Step 3. Determine how unlikely test statistic would be if null hypothesis true.

If null hypothesis true, how likely to observe sample results of this magnitude or larger (in direction of the alternative) just by chance? **… called *p*-value.**

Step 4. Make a Statistical Decision.

Choice 1: *p*-value *not* small enough to convincingly rule out chance. (Usually use 0.05 or 5%, so if $p > 0.05$)

- **Cannot reject the null hypothesis**
- **No statistically significant difference** or relationship

Choice 2: *p*-value small enough to convincingly rule out chance. (Usually use 0.05, so if $p \leq 0.05$)

- **Reject the null hypothesis**
- **Accept the alternative hypothesis**
- There is a **statistically significant difference**

How small is small enough?

Standard of **5% or 0.05** is called **level of significance**.

Sometimes use a different value, like 0.01 or 0.10

Step 5. Make a Conclusion in the Context of the Situation.

- It is important to answer the research question of interest.
- Don’t just stop with whether or not to reject the null hypothesis – explain what that means in context.

Caution: Real Importance versus Statistical Significance

A **statistically significant** relationship or difference ***does not necessarily mean an important one***.

Whether results are statistically significant or not, it is helpful to examine a confidence interval so that you can determine the ***magnitude*** of the effect.

From **width** of the confidence interval, also learn how much **uncertainty** there was in sample results.

Example of men’s heights in spring versus fall.

Steps for Testing Hypotheses About One Mean (simplest case)

Step 1: Determine null and alternative hypotheses

1. H_0: $\mu = \mu_0$ versus H_a: $\mu \neq \mu_0$ (two-sided)
2. H_0: $\mu \geq \mu_0$ versus H_a: $\mu < \mu_0$ (one-sided)
3. H_0: $\mu \leq \mu_0$ versus H_a: $\mu > \mu_0$ (one-sided)

Often H_0 for a one-sided test is written as H_0: $\mu = \mu_0$.
The *p*-value is computed assuming H_0 is true,
and μ_0 is the value used for that computation.

Example: Is Mean Temp 98.6? One sample t-test

- It's always been stated that "normal" body temperature is 98.6. Is that true?
- Many people think it's actually lower. Speculation that it came from rounding to 37 degrees C.
- Data: 16 donors at a blood bank, all under age 30.
- Define μ = ***population*** mean body temperature for **all** healthy people under 30.

Step 1: Determine null and alternative hypotheses

$$H_0: \mu = 98.6 \text{ versus } H_a: \mu < 98.6 \text{ (one-sided)}$$

Step 2: Verify Necessary Data Conditions
For one mean

Situation 1: Population of measurements of interest is **approximately normal**, and a random sample of any size is measured. In practice, use method if shape is not notably skewed or no extreme outliers.

Situation 2: Population of measurements of interest is **not approximately normal**, but a ***large*** **random sample** ($n \geq 30$) is measured. If extreme outliers or extreme skewness, better to have a larger sample.

Continuing Step 2: The Test Statistic

The ***t*-statistic** is a standardized score and is the test statistic for measuring the difference between the sample mean and the null hypothesis value of the population mean:

$$t = \frac{\text{sample mean} - \text{null value}}{\text{standard error}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

This *t*-statistic has (approx) a *t*-distribution with df = *n* - 1.

Step 2 for the Example

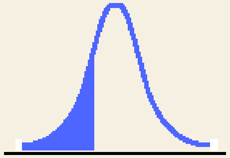
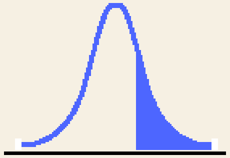
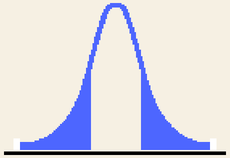
- Histogram of values looks okay (not shown)
- Sample mean = 98.2 degrees, *s* = 0.497 degrees

$$t = \frac{\text{sample mean} - \text{null value}}{\text{standard error}} = \frac{98.2 - 98.6}{0.497 / \sqrt{16}} = -3.22$$

This *t*-statistic has (approx) a *t*-distribution with df = 16 - 1.

Step 3: Assuming H_0 true, Find the *p*-value

- For H_a ***less than***, the *p*-value is the area below *t*, even if *t* is positive.
- For H_a ***greater than***, the *p*-value is the area above *t*, even if *t* is negative.
- For H_a ***two-sided***, *p*-value is 2 × area above |*t*|.

Statement of H_a		*p*-Value Area	*t*-Curve Region
$\mu < \mu_0$	(less than)	Area to the left of t (even if $t > 0$)	
$\mu > \mu_0$	(greater than)	Area to the right of t (even if $t < 0$)	
$\mu \neq \mu_0$	(not equal)	$2 \times$ area to the right of $\lvert t \rvert$	

Results from R

```
> t.test(BodyTemp$Temp, alternative='less', mu=98.6,
conf.level=.95)
       One Sample t-test
data:  BodyTemp$Temp
t = -3.2215, df = 15, p-value = 0.002853
alternative hypothesis: true mean is less than 98.6
95 percent confidence interval:
     -Inf 98.41767
sample estimates:
mean of x
     98.2
```

REDO with alternative='two.sided' (or omit) to get two-sided confidence interval:

```
95 percent confidence interval:
 97.93535 98.46465
```

Steps 4 and 5: Decide Whether or Not the Result is Statistically Significant based on the *p*-value and Report the Conclusion in the Context of the Situation

These two steps remain the same for all of the hypothesis tests we will cover.

Choose a level of significance α, and **reject H_0 if the *p*-value is less than (or equal to) α.**

Otherwise, conclude that there is not enough evidence to support the alternative hypothesis.

Standard is to use $\alpha = 0.05$

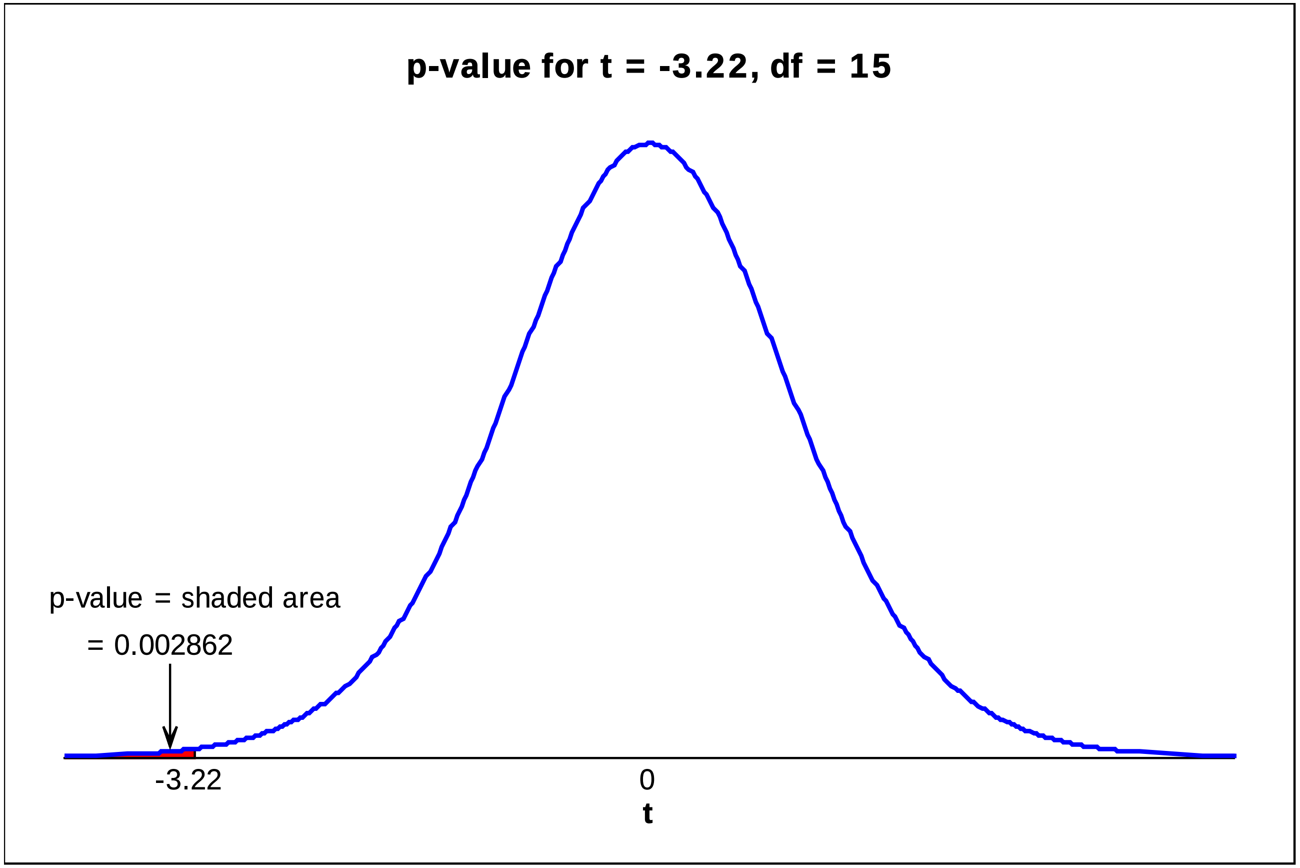
Step 4: Decide whether or not the result is statistically significant based on the *p*-value

Example: *p-value = .002853*, Using $\alpha = 0.05$ as the level of significance criterion, the results are **statistically significant** because the *p*-value of the test is clearly less than 0.05. In other words, we can reject the null hypothesis.

Step 5: Report the Conclusion in Context

We can conclude, based on these data, that the **population** mean body temperature for adults under age 30 is less than 98.6 degrees Fahrenheit.

P-value = area below test statistic of -3.22

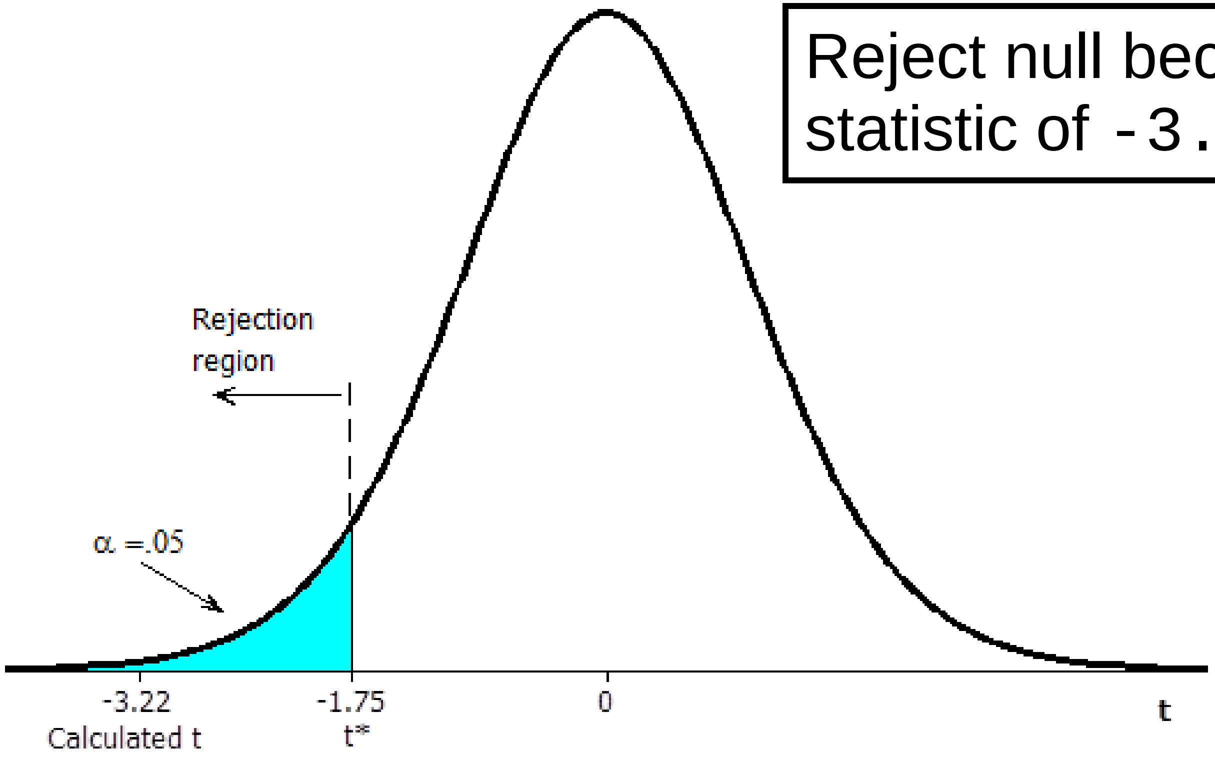


Rejection Region Method:

Find *t* with area 0.05 below it for *t* with df = 15; Using R:

```
> qt(c(.05), df=15)
[1] -1.75305
```

Reject null because test statistic of `-3.22 < -1.75`



Relationship Between Two-Sided Tests and Confidence Intervals

(C.I. formula on white board.) For two-sided tests:
H_0: parameter = null value *and* H_a: parameter ≠ null value

- If the null value is ***covered*** by a $(1 - \alpha)100\%$ confidence interval, the null hypothesis is **not rejected** and the test is **not statistically significant** at level α.

- If the null value is ***not covered*** by a $(1 - \alpha)100\%$ confidence interval, the null hypothesis is **rejected** and the test is **statistically significant** at level α.

Note: 95% confidence interval ⇔ 5% significance level
99% confidence interval ⇔ 1% significance level

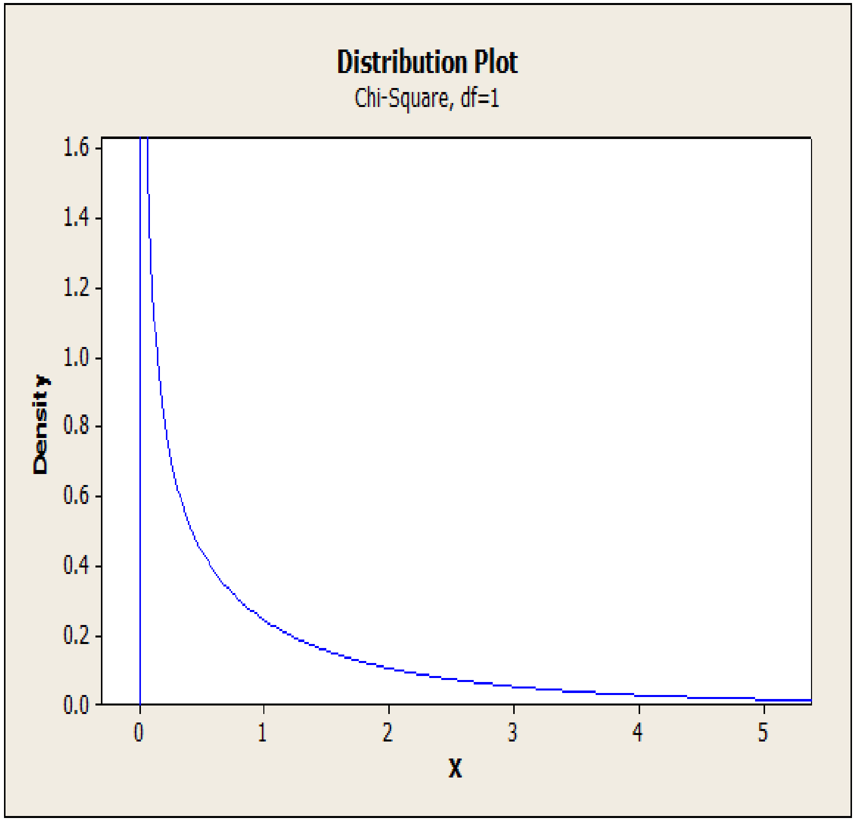
F Distribution and F Tests

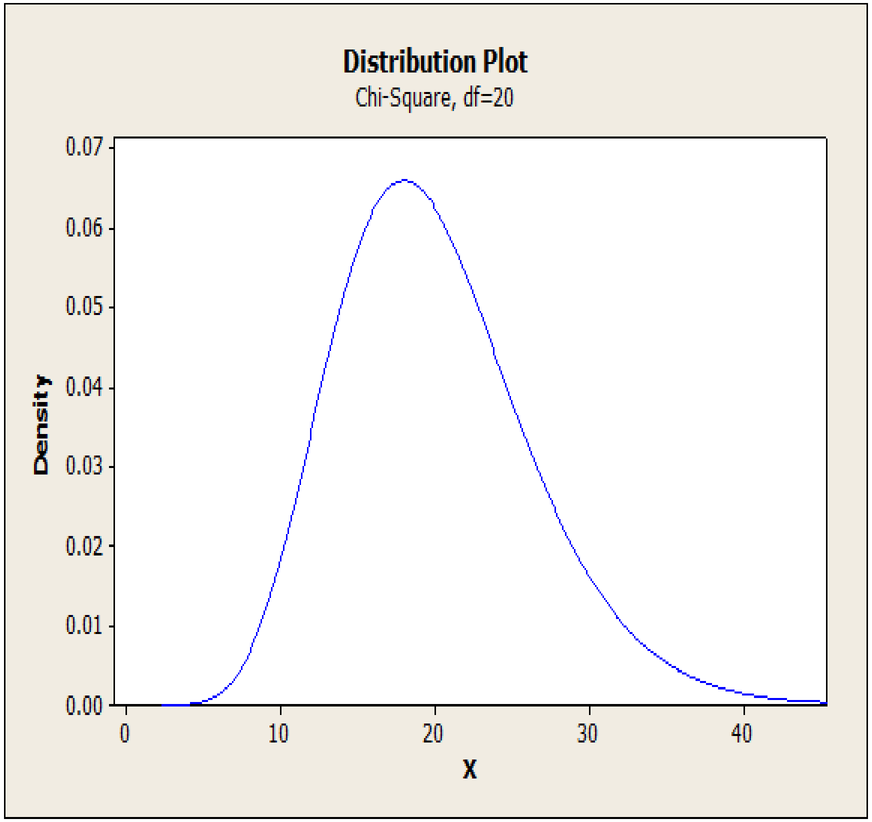
- In regression, tests are not always about a single parameter.
- Sometimes need to compare two sources of variability.
- The resulting *test statistic*, when the null hypothesis is true, has an **F Distribution**.
- What is an F Distribution??

Various distributions, all derived from starting with Normal distribution

- Individual $Y \sim N(\mu, \sigma^2)$
- Standardized version $z = \frac{Y - \mu}{\sigma} \sim N(0, 1)$
  - Called the *standard normal distribution*
- $Z^2 \sim$ Chi-square with df = 1
- Sum of k independent chi-square variables ~ chi-square with df = k

Examples of chi-square distributions, df = 1 and df = 20
Mean for chi-square = df





F Distribution

Suppose

- $X_1 \sim$ chi-square(k_1)
- $X_2 \sim$ chi-square(k_2)
- X_1 and X_2 are independent

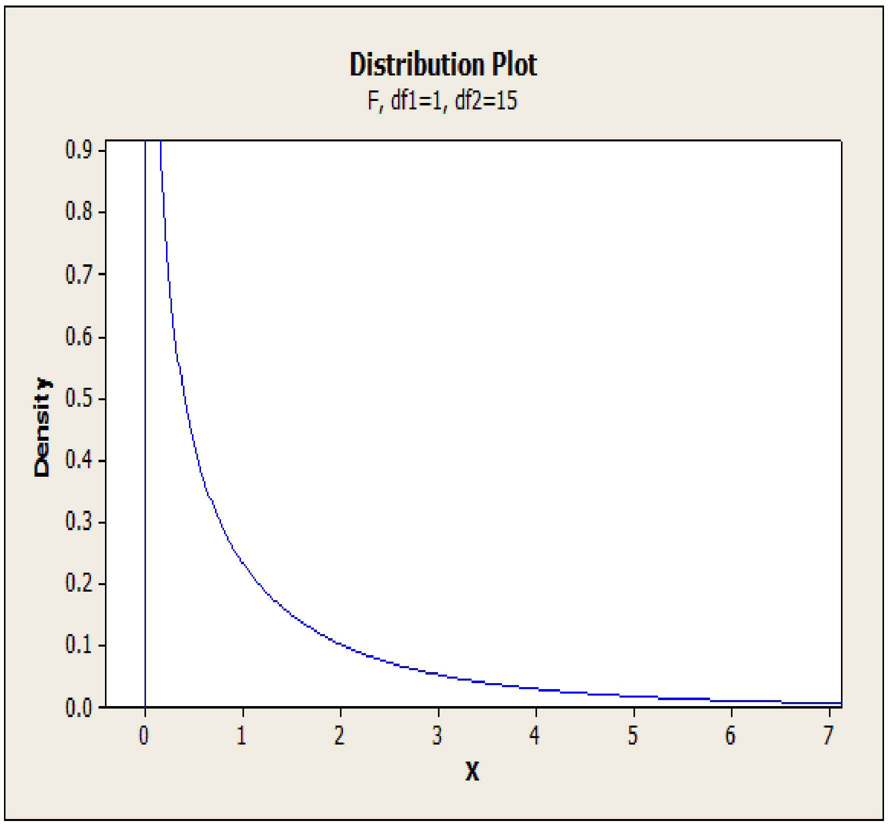
Then the ratio $(X_1/k_1) \div (X_2/k_2) \sim F(k_1, k_2)$
where

k_1 = *numerator degrees of freedom*

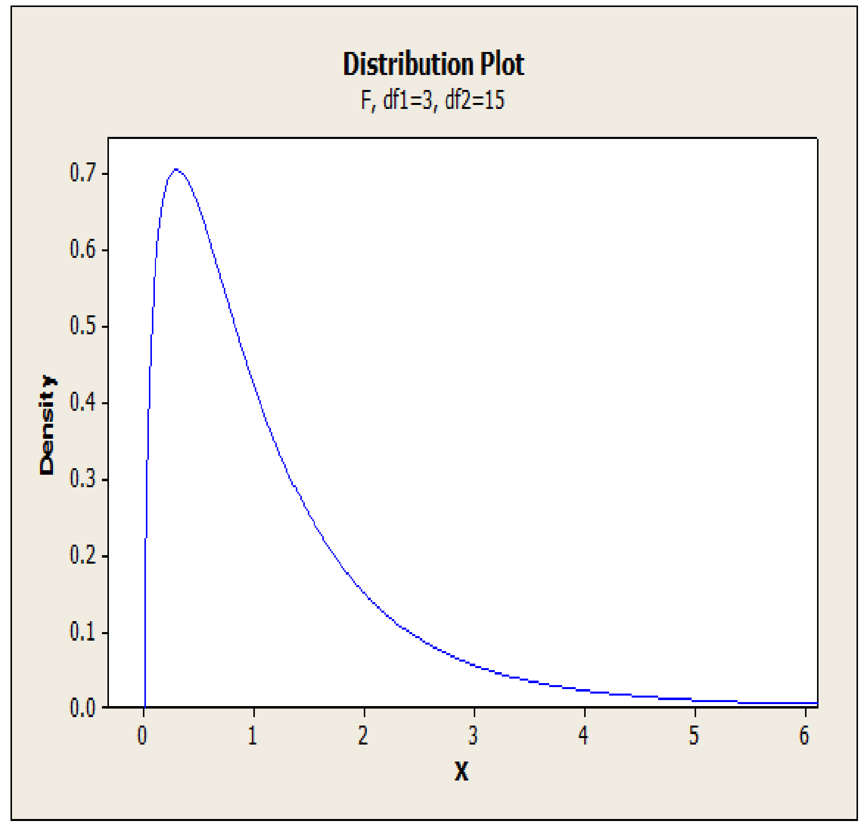
k_2 = *denominator degrees of freedom*

Examples of F Distributions
Mean of F distribution is $k_2/(k_2-2)$

Df = 1 and 15



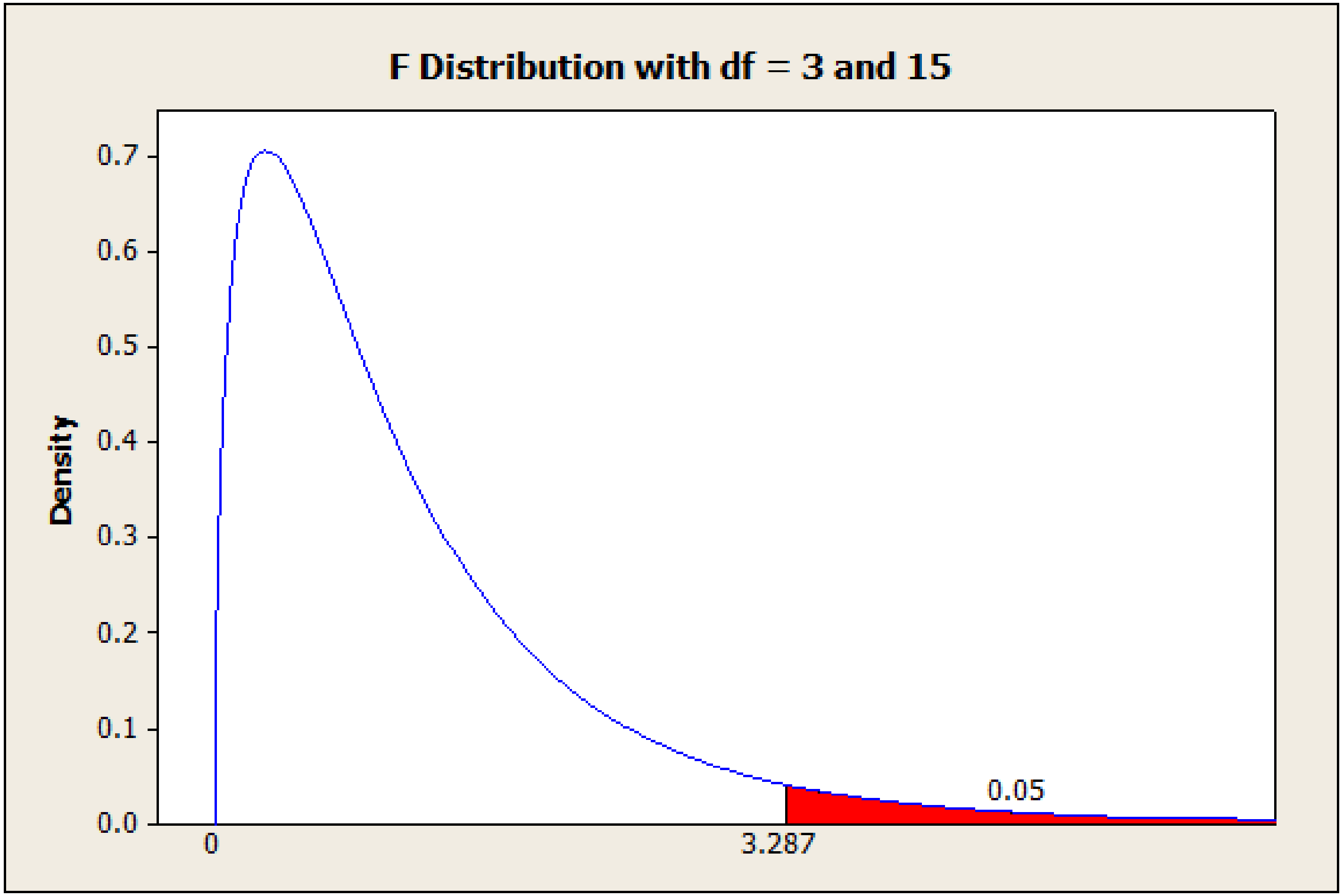
Df = 3 and 15



Example of a rejection region

Suppose F test statistic has df = 3, 15

Reject null if F > 3.287 (for $\alpha = .05$)



Relationship between t and F

- If a random variable $W \sim t(k)$ then
  $W^2 \sim F(1, k)$
- This comes from the fact that a t is formed by a ratio of N(0, 1) and square root of a chi-square/k. (Show on white board.)
- Section 2.1: Test for regression slope – on white board; example on next slide.

Highway Sign Data R Output

```
> summary(HWModel)
Call:
lm(formula = Distance ~ Age, data = Highway)
[OMITTED PART]
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 576.6819    23.4709  24.570  < 2e-16 ***
Age          -3.0068     0.4243  -7.086 1.04e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

Residual standard error: 49.76 on 28 degrees of freedom
Multiple R-squared: 0.642,     Adjusted R-squared: 0.6292
F-statistic: 50.21 on 1 and 28 DF,  p-value: 1.041e-07
```