

There will be more homework on the F-test  
(Wackerly Ch. 11; see also Faraway Ch. 3)

Q: Why an F-test?

A: The F-test that shows up in the standard regression summary is just one specific type of F-test. It is a test of the fitted model for your regression against the "null model": the model with no predictor variables.

If you chose any reasonable predictor variables, the answer will be that your model is better than the null model and the F-stat will be significant: low p-value.

This is a very weak test.

Q: So what is an F-test good for?

A: The F-test can be used to compare two nested models. Your model and the null model

There will be more homework on the ...  
(Wednesday Ch 11; see also Learning 10.2)

Q: Why an F-test?

A: The F-test that shows up in the standard regression summary is just one special type of F-test. It is a test of the fitted model for your regression against the "null model"; the model with no predictor variables. If you check any reasonable predictor variables the answer will be that your model is better than the null model and the F-test will be significant: too good to be true.

This is a very weak test.

Q: So what is an F-test good for?

A: The F-test can be used to compare two nested models. You would want the null model

are a special case of this.

Q: What does this mean?

A: Let's consider an example.

We have 100 data points  $(X_i, Y_i)$

and we fit two models: a linear

model:  $Y = \beta_0 + \beta_1 X + \epsilon$  and a cubic

model:  $Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 X^3 + \epsilon$ .

How does the sum of squared residuals for the linear model compare to the RSS for the cubic model?

The cubic model will always fit ~~better~~ (or at least as well ~~&~~ as) ~~the~~ than the linear model. This is because the cubic model ~~is~~

contains the linear model: if the cubic model estimates are  $\hat{\alpha}_2 = 0, \hat{\alpha}_3 = 0$  and

$\hat{\alpha}_1 = \hat{\beta}_1$  and  $\hat{\alpha}_0 = \hat{\beta}_0$  then the cubic model

is:  $Y = \hat{\alpha}_0 + \hat{\alpha}_1 X + \epsilon$  a linear model.

Q: In the regression model, we assume that the error terms are all having mean zero and constant variance. We also assume that the error terms are independent.

A: Remember that the least squares estimates are only defined in the context of the given regression. It is not clear that the regression without these assumptions would be significant. In other words, it can be

The cubic model with errors is common. Or at least as well as of the linear model. This is because the cubic model is contained in the linear model: if the cubic model expands as  $\beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$  then the cubic model is a linear model.

that Area and  $S_{\text{CR2}}$  are collectively significant but ~~we~~ each is not individually significant in the presence of the other.

The way to justify dropping all 3 at once is a relative F-test.

---

⚠ The F-test can only be used to compare NESTED models. (Homework)

Q: What if we have to compare 2 models that are NOT nested?

A: This will be treated in Ch. 10, model selection. There are many methods.

---

Explanation vs Prediction.

We have a data set  $(X_i, Y_i)$   $i=1, \dots, n$ .

We fit a model  $Y = \beta_0 + \beta_1 X + \epsilon$

We get  $\hat{\beta}_0, \hat{\beta}_1$ .

that there are some differences

experimentally and we need to be careful

experimentally and we need to be careful

The way to find the difference is to

a relative difference



The first part of the experiment

compare the two results and see if they are

Q: What is the difference between the two

that we get from the experiment?

A: This will be the difference between the two

there are many ways to do it

Experimentally we can find the

We have a way to find the difference

We get a result from the experiment

We get a result from the experiment

For any  $X_i$ , we have a "fitted value"

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i.$$

This has lots of uses, e.g.  $e_i = Y_i - \hat{Y}_i$

and  $R^2 = \text{Cor}(\hat{Y}, Y)^2$

Is this  $\hat{Y}_i$  a "prediction" of  $Y_i$ ?

Usually we say "no": here is why.

in obtaining  $\hat{\beta}_0, \hat{\beta}_1$  we used the data point  $(X_i, Y_i)$ . So we already saw this value  $Y_i$  and it is "cheating" to use this information in "predicting"  $Y_i$ .

So instead we say that our equation

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \quad \text{explains the values}$$

$(X_i, Y_i) \quad i=1, \dots, n$  that we used to estimate the  $\hat{\beta}_0, \hat{\beta}_1$ .

If we have new data say one point

For any  $\epsilon > 0$ , we can find  $\delta > 0$  such that

$$\|f(x) - f(y)\| < \epsilon \text{ whenever } \|x - y\| < \delta$$

This can be done by using the definition of continuity.

$$\text{and } \|x - y\| < \delta \implies \|f(x) - f(y)\| < \epsilon$$

is this  $\delta$  a "function" of  $\epsilon$ ?

Usually we say "yes" because we can

obtain  $\delta$  from  $\epsilon$  and we can use the same

point  $(x, y)$  to see that  $\delta$  works for all

values of  $\epsilon$  and this is "clearing" the

way for a "function" of  $\delta$ .

So instead we can say that  $\delta$  is a function

$$\delta = \delta(\epsilon) \text{ and we can write } \|f(x) - f(y)\| < \epsilon$$

whenever  $\|x - y\| < \delta(\epsilon)$  and we can

estimate the value of  $\delta(\epsilon)$ .

If we have a function  $f$  and we want

$(X_{n+1}, Y_{n+1})$  that we did NOT use  
in creating the estimates  $\hat{\beta}_0, \hat{\beta}_1$ .

Then  $\hat{Y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{n+1}$  is a prediction

of  $Y_{n+1}$ , because  $(X_{n+1}, Y_{n+1})$  was not used  
to create the estimates.

Anytime we estimate or predict, we should  
have an error measure.

Two types of predictions: "prediction of a  
mean response" and "prediction of a single  
future observation"

(See Faraway p. 51 - 2).

The error measures are different, but the  
point estimate is the same.

(X, Y) that we did NOT use

in creating the estimator  $\hat{\beta}$ .

Then  $\hat{Y}_i = \hat{\beta}' X_i$  is a prediction

of  $Y_i$ , because  $(X_i, Y_i)$  was not used

to create the estimator.

And since the estimator is unbiased, we should

have an error variance.

Two types of predictions: "prediction of a  
mean response" and "prediction of a single

future observation."

(See Faraway p. 21-22)

The error measure we mention, but the  
point estimate is the same.