Math 455 Class 9 February 9

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 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 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 <u>nested</u> models. Your model and 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: $Y = \beta_0 + \beta_1 X + \varepsilon$ and a cubic model: $Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 X^3 + \varepsilon$.

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) than the linear model. This is because the cubic model 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: $Y = \hat{\alpha}_0 + \hat{\alpha}_1 X + \varepsilon$. a linear model.

that Area and Scw2 are <u>collectively</u> significant ~~but~~ ~~we~~ each is not <u>individually</u> significant in the presence of the other.

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

⚠ The F-~~st~~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 $Y = \beta_0 + \beta_1 X + \varepsilon$

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

that then [illegible] collection

[illegible]

significant [illegible] presence [illegible]

The way to [illegible]

a relative [illegible]

⚠ The [illegible]

compare NESTED [illegible]

Q: What if [illegible]

that are NOT [illegible]

A: This will be [illegible]

There are many [illegible]

Explanation [illegible]

We have a [illegible]

We fit a model [illegible]

We get [illegible]

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

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

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

<u>explains</u> the values (X_i, Y_i) $i = 1, \ldots, n$ that we used to estimate the $\hat{\beta}_0, \hat{\beta}_1$.

If we have <u>new</u> data say one point

(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 <u>prediction</u> of Y_{n+1}, because (X_{n+1}, Y_{n+1}) was not used to create the estimates.

Any time 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_{new}, Y_{new}) that we did NOT use

in creating the estimates $\hat{\beta}_0, \hat{\beta}_1$.

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

of Y_{new}, because (X_{new}, Y_{new}) was not used

to create the estimates.

Any time 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. 41-2)

The error measures are different, but the

point estimate is the same.