Stat E-150 Section #8

Neha Dhawan

dhawan@g.harvard.edu

Office hours by request for one-one help

Getting Midterm back this week!

Regarding policy

- Grade the whole thing over
- Check solutions

- Extra credit opportunity this week!

Calculating for that B

- Need an **84**
- Grad students-What you need on the final:

$$= \frac{(84 - (\text{possible project grade} * .2)}{.8} \times 3$$

=> above answer -midterm grade –hw grade

- Undergrad students:

= (84x3) -midterm grade –hw grade

Nested F-test

- Test models that are ‘nested’ within other models
- The larger fuller model against the smaller reduced model nested within the big model
- Are the extra predictors in the larger model useful?

Example

Let's investigate whether the entry fee to enter an elite club at the door depends on age, gender. They recording 50 people trying to enter a particular club in Boston.

Model: $Fee = \beta_0 + \beta_1 Age + \beta_2 Gender + e$

Fuller complete Model (with all possible combinations):

$Fee = \beta_0 + \beta_1 Age + \beta_2 Gender + \beta_3 Age^2 + [\beta_4 Gender^2] + \beta_5 Age*Gender + e$

Question we ask is:

Is it worthwhile to convert our initial model into a complete second order model?

Nested Model: $Fee = \beta_0 + \beta_1 Age + \beta_2 Gender + e$

Full model: $Fee = \beta_0 + \beta_1 Age + \beta_2 Gender + \beta_3 Age^2 + \beta_4 Age*Gender + e$

Nested F-test!!

Nested F-test

- Lets us look at our additional predictors in the full model together

$$F = \frac{(SSModelfull - SSModelnested)/\#predictors\ tested}{SSEfull/(n-k-1)}$$

k- predictors in full model

n- sample size

predictors tested- number of predictors full model- predictors in nested model

Run each model separately and compare them:

Nested

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.458[a]	.210	.176	18.1681

a. Predictors: (Constant), Age, Gender

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4118.622	2	2059.311	6.239	.004[b]
	Residual	15513.702	47	330.079		
	Total	19632.324	49			

a. Dependent Variable: Price

b. Predictors: (Constant), Age, Gender

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.376	7.707		.698	.489
	Gender	24.572	7.050	.455	3.485	.001
	Age	.235	.248	.124	.948	.348

a. Dependent Variable: Price

Full

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6518.580	4	1629.645	5.592	.001[b]
	Residual	13113.744	45	291.417		
	Total	19632.324	49			

a. Dependent Variable: Price

b. Predictors: (Constant), age2, Gender, age_gender, Age

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.576[a]	.332	.273	17.0709

a. Predictors: (Constant), age2, Gender, age_gender, Age

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	12.167	16.701		.729	.470
	Gender	-20.778	18.097	-.384	-1.148	.257
	Age	.059	.937	.031	.063	.950
	age_gender	1.721	.629	.907	2.737	.009
	age2	-.002	.012	-.071	-.150	.881

a. Dependent Variable: Price

Nested F-test

H_0: β_i =0 for all predictors in subset

H_a: β_i =/0 for at least one predictor in subset

ANOVA[a]

Model		Sum of Squares	df	Mean Square
1	Regression	4118.622	2	2059.311
	Residual	15513.702	47	330.079
	Total	19632.324	49	

a. Dependent Variable: Price

b. Predictors: (Constant), Age, Gender

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6518.580	4	1629.645	5.592	.001[b]
	Residual	13113.744	45	291.417		
	Total	19632.324	49			

a. Dependent Variable: Price

b. Predictors: (Constant), age2, Gender, age_gender, Age

$$F = \frac{(SSModelfull - SSModelnested)/\#predictors\ tested}{SSEfull/(n-k-1)}$$

$$F = \frac{(6519 - 4117)/2}{13114/(50-4-1)} = 4.12$$

To find p-value have to look it up manually

http://www.socr.ucla.edu/applets.dir/f_table.html

Numerator df= # of predictors being tested
Denominator df= Residual df for full model

Numerator- 2; denominator- 45; critical value= 3.15

F= 4.12 > 3.15 thus $p<0.05$

F> cv $p<0.05$
F< cv $p>0.05$

Nested F-test

H_0: β_i =0 for all predictors in subset

H_a: β_i =/0 for at least one predictor in subset

F= 4.12 > 3.15 thus $p<0.05$

So we can reject our null, and conclude that one of the tested betas, or a combination of them, is/are useful in predicting **average** club entry fee, **after accounting for** Age and Gender

So final model?

Nested

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.458[a]	.210	.176	18.1681

a. Predictors: (Constant), Age, Gender

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4118.622	2	2059.311	6.239	.004[b]
	Residual	15513.702	47	330.079		
	Total	19632.324	49			

a. Dependent Variable: Price

b. Predictors: (Constant), Age, Gender

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.376	7.707		.698	.489
	Gender	24.572	7.050	.455	3.485	.001
	Age	.235	.248	.124	.948	.348

a. Dependent Variable: Price

Full

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6518.580	4	1629.645	5.592	.001[b]
	Residual	13113.744	45	291.417		
	Total	19632.324	49			

a. Dependent Variable: Price

b. Predictors: (Constant), age2, Gender, age_gender, Age

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.576[a]	.332	.273	17.0709

a. Predictors: (Constant), age2, Gender, age_gender, Age

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	12.167	16.701		.729	.470
	Gender	-20.778	18.097	-.384	-1.148	.257
	Age	.059	.937	.031	.063	.950
	age_gender	1.721	.629	.907	2.737	.009
	age2	-.002	.012	-.071	-.150	.881

a. Dependent Variable: Price

Picking the full model

- Higher order terms are significant even if other terms aren't
- R2 is good, std. error is low, model is more significant
- Can also plot graphs to see if should have started with quadratic term in the first place

Nested F-test

- Test models that are ‘nested’ within other models
- Lets us clump predictors together and test their significance
- Avoids Type 1 error when running multiple t-tests

The End!
Questions?

Email dhawan@g.harvard.edu for feedback and any future changes