Idea of Larger ("Full") and Smaller ("Reduced") model

Concepts of variable added "IN ORDER" vs. "LAST"

Distinguish # of terms from # of variables

of ***terms*** > # of ***variables*** if adding higher powers of a continuous variable, or if a variable is categorical and represented by indicator terms.

Same ideas apply in logistic/Poisson regression later

G&S only deal with adding/deleting 1 term; if assessing the addition/deletion of a categorical variable, or a "block" of continuous variables, need a more general approach.

References

- G&Spp 65-68 and pp79-80 (models that differ by 1 term)
- ***NWNW § 7.1-7.3 or my Course 697 notes on these***
  (use same username and password to access these)
- KKMN Ch 9 or my 1999 Course 678 notes on these

I suggest Notes for NKNW Ch 7 -- more compact; NKNW have a good diagram for Extra Sums of Squares; my notes avoid notation for #s of variables in the larger & smaller models.

Below (because I had done the work anyway!) I summarize the KKMN chapter, using their notation, and with an annotated example.

Example -- following KKMN presentation / notation

§ = section of KKMN Chapter 9

Number of terms (not counting intercept):

M O D E L

§	**Reduced**	**Full**	Test **Statistic**	df
2	0	k	F(model)* ("Overall F-test")	k, (n-1 -k)
3	p	p + 1	t** or F(partial)	1, (n-1 - p-1)
4	p	p + k	F(partial)***	k, (n-1 - p-k)

TEST Statistic	**obtained from ...**
*F(model) ("Overall F-test")	1st line of summary Anova table
** t	$t = \hat{\beta}_{p+1} / SE[\hat{\beta}_{p+1}]$

$$F(\text{partial}) = \frac{\text{Diff. in Reg SS / k (= diff. in \# terms)}}{\text{Mean Sq Error("Residual") in Larger Model}}$$

= **square of t if k = 1 additional term**

Example

Berkeley data: prediction of height at 18 **HT18** (cm) from:-

weight (kg) at age 2 **WT2**

height (cm) at age 2 **WT2**

gender (1=girl, 0=boy) **GENDER**

Fitted Equation

HT18 = 61.84 + 0.04 WT2 + 1.33 HT2 – 12.00 GENDER

Summary

Mean of Response	172.72	R-Square	0.75
		Adj R-Sq	0.74
		Root MSE	4.68

Analysis of Variance

("**Overall**-F" test)
^ **SINGULAR**

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	3	3628.12	1209.37	55.33	0.0001
Error	54	1180.25	21.86		
C Total	57	4808.38			

test of All 3 β's = 0
versus
AT LEAST ONE β is **NOT** 0

Type III Tests ("Partial-F" tests)

^ **PLURAL**

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
WT2	1	0.14	0.14	0.01	**0.9363**
HT2	1	726.07	726.07	33.22	0.0001
GENDER	1	1952.56	1952.56	89.34	0.0001

test of β WT2 | HT2 GENDER = 0

test of β HT2 | WT2 GENDER = 0

test of β GENDER | WT2 HT2 = 0

Each F-test tests contribution of the TERM in question, **GIVEN THAT THE OTHER TERMS ARE ALREADY INCLUDED** i.e. TESTS its contribution as LAST TERM in model

Parameter Estimates

Variable	DF	Estimate	Std Error	T Stat	Prob >|T|
WT2	1	0.04	0.50	0.08	**0.9363**
HT2	1	1.33	0.23	5.76	0.0001
GENDER	1	-12.00	1.27	-9.45	0.0001

F = 0.01 = square of 0.08

F = 33.22 = square of 5.76

F = 89.34 = square of -9.45

Each is a "**TERM ENTERED LAST**" test. The order in which term is entered or "clicked" into model doesn't matter

Type I Tests

Again, "Partial-F" tests, but now **SEQUENTIAL !!**

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
WT2	1	935.54	935.54	42.80	0.0001
				^^^^^	^^^^^^
	test of β **WT2**				**= 0**
HT2	1	740.02	740.02	33.86	0.0001
				^^^^^	^^^^^^
	test of β **HT2 | WT2**				**= 0**
GENDER	1	1952.56	1952.56	89.34	0.0001
				^^^^^	^^^^^^
	test of β **GENDER | WT2 HT2**				**= 0**

F-test tests contribution of TERM, GIVEN THAT TERMS **BEFORE IT IN THE LIST** ARE ALREADY INCLUDED

Q: **Can't remember Type I from Type III?** (and software doesn't indicate which). How can one tell if partial F-tests are "variables entered last" tests or "variables entered in THAT PARTICULAR ORDER" tests?

A: If Sums of Squares associated with the individual terms add up to the "model" or "Regression" Sums of Squares, the partial-F tests refer to variables entered in THAT ORDER

Sums of Squares associated with each "VARIABLE ADDED LAST" can add up to MORE, or to LESS, than the "Regression" Sums of Squares.

Can you think of when they might add up to LESS THAN, MORE THAN, or EXACTLY the Model SS?

Multiple Partial F Test (SINGULAR)

```
HT18 =   11.3
       -  0.60 WT9        |
       +  1.21 HT9        |
       -  0.07 LG9        |  from age 9
       -  0.04 ST9        |
       - 11.24 GENDER     |
       +  0.76 WT2        |
       +  0.19 HT2        |  from age 2
```

Mean Response: 172.72 ; R^2 0.91;RMSE 2.89; Adj R^2 0.90.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	7	4389.93	627.13	74.94	0.0001
Error	50	418.44	**8.37**		
C Total	57	4808.38			

Type I Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
WT9	1	179.21	179.21	21.41	0.0001
HT9	1	2304.17	2304.17	275.33	0.0001
LG9	1	20.62	20.62	2.46	0.1228
ST9	1	248.66	248.66	29.71	0.0001
GENDER	1	1554.48	1554.48	185.75	0.0001
WT2	1	**74.34**	74.34	8.88	0.0044
HT2	1	**8.44**	8.44	1.01	0.3200
	---	-------			
	2	**82.78**			

Value of age 2 data, once already have age 9 data:

$$F = \frac{82.78 \,/\, 2}{8.37} = \frac{41.39}{8.37} = 4.94$$

F(2,50) table

0.95	*0.99*	*0.999*
3.18 ^	5.06	7.96

Type I Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
WT2	1	935.54	935.54	111.79	0.0001
HT2	1	740.02	740.02	88.43	0.0001
GENDER	1	1952.56	1952.56	233.31	0.0001
WT9	1	**0.08**	0.08	0.01	0.9238
HT9	1	**749.24**	749.24	89.53	0.0001
LG9	1	**1.50**	1.50	0.18	0.6736
ST9	1	**10.99**	10.99	1.31	0.2572
	---	------			
	4	**761.81**			

Value of age 9 data, once already have age 2 data:

$$F = \frac{761.81 \,/\, 4}{8.37} = \frac{190.45}{8.37} = 22.75$$

F(4,50) table

0.95	*0.99*	*0.999*
2.56	3.72	5.46

Double Check

Running the bigger & smaller models separately:

			SS(regression) Sum of Squares
	Source	DF	
with...			
all 7 terms	Model	7	4389.93
WT2 HT2 GENDER (earlier)	Model	3	3628.12
		---	-------
	Difference	4	761.81