F tests and the Extra Sum of Squares

Example:

$$\begin{aligned} Y &= \text{plaster hardness} \\ s &= \text{sand content} \\ f &= \text{fibre content} \end{aligned}$$

Model:

$$Y_i = \beta_0 + \beta_1 s_i + \beta_2 s_i^2 + \beta_3 f_i + \beta_4 f_i^2 + \beta_5 s_i f_i + \epsilon_i$$

In matrix form:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & s_1 & s_1^2 & f_1 & f_1^2 & s_1 f_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & s_n & s_n^2 & f_n & f_n^2 & s_n f_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_5 \end{bmatrix}$$

Hypotheses to test

Questions:

- Do we need the $S * F$ term?
- Do we need the F or F^2 terms?
- Do we need the S or S^2 terms?

To answer these questions we test

- $H_o : \beta_5 = 0$
- $H_o : \beta_3 = \beta_4 = 0$ (or perhaps $H_o : \beta_3 = \beta_4 = \beta_5 = 0$)
- $H_o : \beta_1 = \beta_2 = 0$

Technique: we fit a sequence of models:

(a) Original "full" model.

(b) The model with **no interactions**:

$$\mu_i = \beta_0 + \beta_1 s_i + \beta_2 s_i^2 + \beta_3 f_i + \beta_4 f_i^2$$

(c) The Sand only model:

$$\mu_i = \beta_0 + \beta_1 s_i + \beta_2 s_i^2$$

(d) The Fibre only model:

$$\mu_i = \beta_0 + \beta_3 f_i + \beta_4 f_i^2$$

(e) The "Empty" model:

$$\mu_i = \beta_0$$

Each model has design matrix which is submatrix of full design matrix:

$$Y = [\mathbf{1}|X_1|X_2|X_3] \left[\begin{array}{c} \beta_0 \\ \hline \beta_1 \\ \beta_2 \\ \hline \beta_3 \\ \beta_4 \\ \hline \beta_5 \end{array}\right] + \epsilon$$

The design matrices for the models a, b, c, d and e are given by

$$\begin{aligned} X_a &= X \\ X_b &= [\mathbf{1}|X_1|X_2] \\ X_c &= [\mathbf{1}|X_1] \\ X_d &= [\mathbf{1}|X_2] \\ X_e &= [\mathbf{1}] \end{aligned}$$

Note that $\mathbf{1}$ is a column vector of n 1s.

F tests

- Can compare two models easily if one is a special case of the other.
- Example: design matrix of first model is submatrix of second obtained by selecting subcolumns.
- Model (b) is a special case of (a), model (c) is a special case of (a) or (b) but models (c) and (d) are not comparable.

Comparing two models: General Theory

- Consider matrix X partitioned into pieces X_1 and X_2.

$$X = [X_1 | X_2]$$

- The Full Model is

$$\begin{aligned} Y &= X\beta + \epsilon \\ &= [X_1 | X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \epsilon \\ &= X_1\beta_1 + X_2\beta_2 + \epsilon \end{aligned}$$

- The Reduced model is

$$Y = X_1\beta_1 + \epsilon$$

- Difference is term $X_2\beta_2$ which is 0 IF the null hypothesis $H_o : \beta_2 = 0$ is true.

Dimensions

- β has p parameters.
- β_i has p_i parameters with $p_1 + p_2 = p$.

Testing $H_o : \beta_2 = 0$

- Fit both full and reduced models; get:

$$\begin{aligned} Y &= \hat{\mu}_F + \hat{\epsilon}_F \\ &= \hat{\mu}_R + \hat{\epsilon}_R \end{aligned}$$

- Subscript F refers to full model and R to reduced model.
- Get decomposition of data vector Y into sum of three perpendicular vectors:

$$Y = \hat{\mu}_R + (\hat{\mu}_F - \hat{\mu}_R) + \hat{\epsilon}_F$$

- I showed

$$\begin{aligned} \hat{\mu}_R &\perp \hat{\mu}_F - \hat{\mu}_R \\ \hat{\mu}_R &\perp \hat{\epsilon}_F \\ \hat{\mu}_F - \hat{\mu}_R &\perp \hat{\epsilon}_F \end{aligned}$$

Resulting ANOVA table

Source	Sum of Squares	Degrees of Freedom
X_1	$\|\|\hat{\mu}_R\|\|^2$	p_1
$X_2\|X_1$	$\|\|\hat{\mu}_F - \hat{\mu}_R\|\|^2$	p_2
Error	$\|\|\hat{\epsilon}\|\|^2$	$n - p$
Total (Unadjusted)	$\|\|Y\|\|^2$	n

F tests again

- In this table the notation $X_2|X_1$ means X_2 **adjusted for** X_1 or X_2 after fitting X_1.
- This table can now be used to test $H_o : \beta_2 = 0$ by computing

$$F = \frac{\mathrm{MS}(X_2|X_1)}{\mathrm{MSE}} = \frac{||\hat{\mu}_F - \hat{\mu}_R||^2/p_2}{||\hat{\epsilon}||^2/(n-p)}$$

- Get P value from $F_{p_2,n-p}$ distribution.
- P value usually computed by software.
- Do level α test by comparing P to α.

Another Formula for this F statistic

- Recall that
$$||\hat{\mu}_R||^2 + ||\hat{\epsilon}_R||^2 = ||Y||^2$$
and
$$||\hat{\mu}_R||^2 + ||\hat{\mu}_F - \hat{\mu}_R||^2 + ||\hat{\epsilon}_F||^2 = ||Y||^2$$
so that
$$||\hat{\epsilon}_R||^2 = ||\hat{\epsilon}_F||^2 + ||\hat{\mu}_F - \hat{\mu}_R||^2$$
- This makes
$$\begin{aligned} F &= \frac{(\mathrm{ESS}_R - \mathrm{ESS}_F)/p_2}{\mathrm{ESS}_F/(n-p)} \\ &= \frac{\mathrm{ExtraSS}/p_2}{\mathrm{ESS}_F/(n-p)} \end{aligned}$$

Remarks

1. If the errors are normal then
$$\frac{\mathrm{ESS}_F}{\sigma^2} \sim \chi^2_{n-p}$$
2. If the errors are normal **and** $H_o : \beta_2 = 0$ is true then
$$\frac{\mathrm{ExtraSS}}{\sigma^2} \sim \chi^2_{p_2}$$
3. ESS_F is independent of the Extra SS.
4. SO:
$$\frac{\mathrm{ExtraSS}/(\sigma^2 p_2)}{\mathrm{ESS}_F/(\sigma^2(n-p))} = F \sim F_{p_2,n-p}$$

Example of the above: Multiple Regression

- Hardness, Y_i, of plaster measured for each of 9 combinations of sand content and fibre content.
- Sand content s_i was set at one of 3 levels as was fibre content f_i.
- All possible combinations tried, each on two batches of plaster.
- Here is an excerpt of the data:

Sand	Fibre	Hardness	Strength
0	0	61	34
0	0	63	16
15	0	67	36
15	0	69	19
30	0	65	28

...

Models Fitted

- I fit submodels of the following "Full" model:

$$Y_i = \beta_0 + \beta_1 s_i + \beta_2 s_i^2 + \beta_3 f_i + \beta_4 f_i^2 + \beta_5 s_i f_I + \epsilon_i$$

- There are $2^5 = 32$ possible submodels of the full model
- Many of these 32 models are not sensible, such as

$$Y_i = \beta_0 + \beta_4 f_i^2 + \epsilon_i$$

or

$$Y_i = \beta_0 + \beta_5 s_i f_i + \epsilon_i$$

- Assume interaction term ($\beta_5 s_i f_i$) is negligible unless each of S and F have some effect .
- Assume that quadratic terms will probably not be present unless linear terms are present.
- This limits the set of potential reasonable models.
- Fit each; error sum of squares in following table:

Fitting various models

Model for μ	Error SS	Error df
Full	81.264	12
$\beta_0 + \beta_1 s_i + \beta_2 s_i^2 + \beta_3 f_i + \beta_4 f_i^2$	82.389	13
$\beta_0 + \beta_1 s_i + \beta_2 s_i^2 + \beta_3 f_i$	104.167	14
$\beta_0 + \beta_1 s_i + \beta_2 s_i^2$	169.500	15
$\beta_0 + \beta_1 s_i$	174.194	16
$\beta_0 + \beta_1 s_i + \beta_3 f_i + \beta_4 f_i^2$	87.083	14
$\beta_0 + \beta_1 f_i + \beta_2 f_i^2$	189.167	15
$\beta_0 + \beta_1 f_i$	210.944	16
$\beta_0 + \beta_1 s_i + \beta_3 f_i$	108.861	15

Hypotheses tested

- First question: do 2nd degree polynomial terms, that is, those involving β_2, β_4 and β_5 need to be included?
- Compare top line with model containing only $\beta_0 + \beta_1 s_i + \beta_3 f_i$.
- The extra SS is 108.861-81.264 on 3 degrees of freedom which gives a mean square of (108.861-81.264)/3= 9.199.
- The MSE is $81.264/12 = 6.772$.
- Gives an F-statistic of 9.199/6.772=1.358 on 3 numerator and 12 denominator degrees of freedom.
- P-value is 0.30 which is not significant.
- So we delete quadratic terms and consider the coefficients of S and F.

A t test and an F test

Q : Are the effects of S and F additive?

A : Test $H_o : \beta_5 = 0$.

- There are two methods to carry out such a test:
  1. A t test
  2. A F test.

Fact : the F test is equivalent to a two sided t test.

- The t test uses

$$t = \frac{\hat{\beta}_5 - 0}{\hat{\sigma}_{\hat{\beta}_5}} = \frac{\hat{\beta}_5}{\sqrt{\mathrm{MSE}}\sqrt{(X^T X)^{-1}_{66}}} \sim t_{1,n-6}$$

t tests

- See Distribution Theory section for linear combinations:

$$\beta_5 = \underbrace{[0,0,0,0,0,1]}_{x^T} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_5 \end{bmatrix}$$

  and $x^T(X^TX)^{-1}x$ is the lower right hand corner entry in $(X^TX)^{-1}$, that is, $(X^TX)^{-1}_{66}$.

- The F test uses

$$F = \frac{(\mathrm{ESS_R} - \mathrm{ESS}_F)/1}{\mathrm{ESS_{FULL}}/(n-6)} \sim F_{1,n-6} \quad (= t^2)$$

Testing for Main Effects

The Data

- Y = hardness of plaster. $n = 18$ batches.
- S = sand content. Values used 0%, 15% 30%.
- F = fibre content. Values used 0%, 25% 50%.
- Factorial design with 2 replicates.

Comparison of Models

"Full" model

$$Y_i = \beta_0 + \beta_1 s_i + \beta_2 s_i^2 + \beta_3 f_i + \beta_4 f_i^2 + \beta_5 s_i f_I + \epsilon_i$$

Fitted Models, ESS, df for error

Model for μ	ESS	Error df
Full	81.264	12
$\beta_0 + \beta_1 s_i + \beta_2 s_i^2 + \beta_3 f_i + \beta_4 f_i^2$	82.389	13
$\beta_0 + \beta_1 s_i + \beta_2 s_i^2 + \beta_3 f_i$	104.167	14
$\beta_0 + \beta_1 s_i + \beta_2 s_i^2$	169.500	15
$\beta_0 + \beta_1 s_i$	174.194	16
$\beta_0 + \beta_1 s_i + \beta_3 f_i + \beta_4 f_i^2$	87.083	14
$\beta_0 + \beta_1 f_i + \beta_2 f_i^2$	189.167	15
$\beta_0 + \beta_1 f_i$	210.944	16
$\beta_0 + \beta_1 s_i + \beta_3 f_i$	108.861	15
β_0(empty model)	276.278	17

Hypothesis tests:

1. Quadratic terms needed? $H_o : \beta_2 = \beta_4 = \beta_5 = 0$.
   - Extra SS = 108.861-81.264.
   - $F = [(108.861 - 81.264)/3]/[81.264/12] = 1.358$.
   - Degrees of freedom are $3, 12$ so $P = 0.30$, not significant.

3. Linear terms needed? There are several possible F-tests.
   3.1 Compare full model to empty model.

$$F = (276.278 - 81.264)/5/(81.264/12) = 5.76$$

   so P is about .006.
   3.2 Assume full model is now additive, linear model

$$\beta_0 + \beta_1 s_i + \beta_3 f_i.$$

   Then

$$F = [(276.278 - 108.861)/2]/[108.861/15] = 11.53$$

   and P is about 0.0009.
   3.3 Use estimate of σ^2 from full model
   3.4 But get extra SS from last comparison: $F = [(276.278 - 108.861)/2]/[81.264/12] = 12.36$ for a P value of 0.001

Conclusions

- Both Sand and Fibre influence Hardness.
- Linear terms in S and F are adequate.