Partial (sequential) F tests

$$Y = \underline{\beta_0 + \beta_1 X_1} + \varepsilon \qquad (1)$$

$$Y = \underline{\beta_0 + \beta_1 X_1} + \beta_2 X_2 + \varepsilon \qquad (2)$$

A sequential F test is a partial F test for the addition of a single term to an existing model

$\rightarrow + \beta_2 X_2$

F test

1. $H_0 : \dfrac{\sigma^2_{addition}}{\sigma^2_{Error,\ larger\ model\ (2)}} = 1$ vs $H_A : \dfrac{\sigma^2_a}{\sigma^2_\varepsilon} > 1$

   (in variance terms)

   equiv. $H_0 : \beta_2 = 0$ vs $H_A : \beta_2 \neq 0$

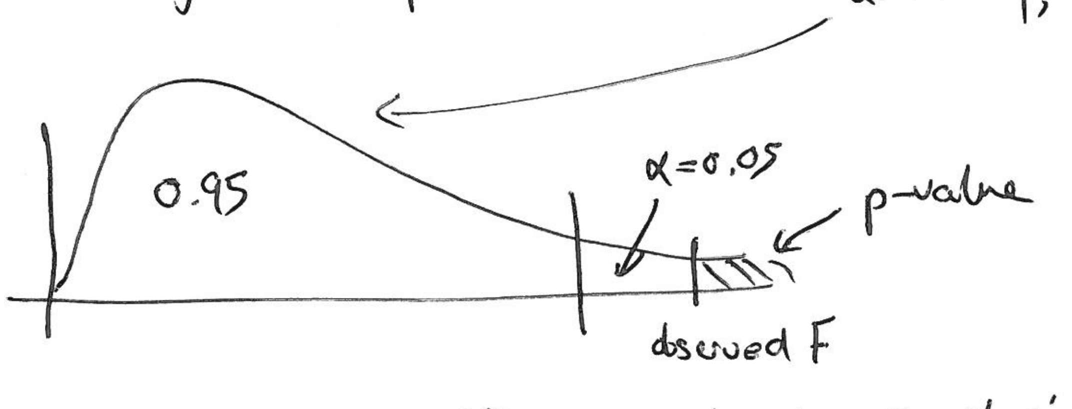
   (in mean terms)

2. Test statistic $F = \dfrac{MS_{addition}}{MS_{error/residual}} \sim F_{addition\ df,\ error\ df}$

   from model (2)

3. Decision rule $\alpha = 0.05$

   reject H_0 if observed $F > F_{addition\ df,\ error\ df}$

4. 

5. Compare p-value with α & draw a conclusion

Partial (nested) F test for a group of terms in a nested model

Model A $\quad Y = \underline{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta x_3} + \varepsilon \quad$ "base" model

Model B $\quad Y = \underbrace{\qquad\qquad}_{\text{"base" model}} + \underbrace{\beta_4 x_4 + \beta_5 x_5}_{\text{additions}} + \varepsilon$

"expanded" model

"base" model A is nested inside the "expanded" model B

model A is a subset of model B

1. $H_0: \dfrac{\sigma^2_{\text{addition}}}{\sigma^2_{\text{error, larger model B}}} = 1$

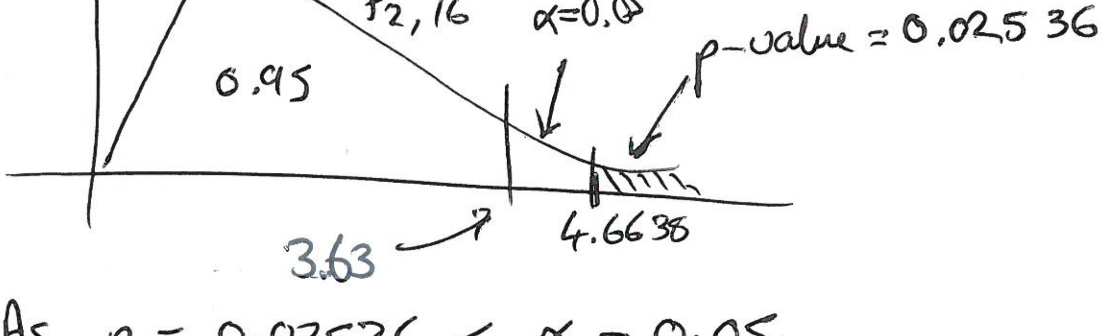
   vs $H_A \quad \dfrac{\sigma^2_{\text{addition}}}{\sigma^2_{\text{error}}} > 1$

   or equivalently $H_0: \beta_4 = \beta_5 = 0$ vs H_A: not all of (both) $\beta_4, \beta_5 = 0$

2. Test statistic $F = \dfrac{MS_{\text{addition}}}{MSE_{(\text{model B})}} \sim F_{2,16}$

3. $\alpha = 0.05$ reject H_0 if $p < \alpha$

4. obs $F_{2,16} = 4.6638$



5. As $p = 0.02536 < \alpha = 0.05$

   reject H_0 in favour of H_A & conclude that at least one of the two additional terms is a significant addition to the base model

The partial (nested) F test is the most general of these tests for nested models, but the differences in names here is just jargon!

A sequential F test is just the special case of a nested test where we are adding a single additional (they are both still partial F tests)

The overall F test for a multiple regression model is also just a special case of the nested F test → here the additional terms are all of the x variables on top of a null model;

Model (A) : $Y = \beta_0 + \varepsilon$ ← null or mean model

Model (B) : $Y = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2 + \dots \beta_k x_k}_{\text{addition!}} + \varepsilon$

In all of these (partial) F tests for nested models, we will tend to prefer the simpler "base" model over the more complicated "expanded" model whenever we fail to reject H_0 (with some caveats → the null model is not always the best "base" model → it will depend on the research question)

→ but these F tests and the ANOVA table are key in deciding what belongs in the model, we will use these as an approach for refining models

What has changed from SLR to MR?

→ plot() – is model appropriate?
are the underlying assumptions OK?

→ these are pretty much the same as earlier

→ anova() – is model adequate?
does it (and all parts of it) have significant explanatory power?

→ this has definitely changed

→ summary() – used to try the research question, once we have the right model

→ again, much the same as earlier

→ predict() – we will also see that this hasn't changed much