# SUBMODELS (NESTED MODELS)
# AND
# ANALYSIS OF VARIANCE
# OF REGRESSION MODELS

Data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

Assumptions:
- Linear mean function
- Constant conditional variance
- Independence of observations
- Normality of conditional distributions

Our model has 3 parameters:

$E(Y|x) = \eta_0 + \eta_1 x$ (Two parameters: $\eta_0$ and $\eta_1$)
$Var(Y|x) = \sigma^2$ (One parameter: $\sigma$)

We will call this the *full model*.

Many hypothesis tests on coefficients can be reformulated as test of the full model against a *submodel* : a special case of the full model obtained by specifying certain of the parameters or certain relationships between parameters.



*Examples*:

a. NH: $\eta_1 = 1$
AH: $\eta_1 \neq 1$

What submodel does NH correspond to?

How many parameters does it have?

AH corresponds to the full model (with three parameters, including $\eta_1$).

b. NH: $\eta_1 = 0$
AH: $\eta_1 \neq 0$

AH <-> full model.

NH <-> ?

Number of parameters?

c. NH: $\eta_0 = 0$
AH: $\eta_0 \neq 0$

AH <-> full model.

NH <->?

Number of parameters?

We can go the other way:

*Examples*: For the submodel given, what is the null hypothesis of the corresponding hypothesis test?

d. $E(Y|x) = 2 + \eta_1 x$
$\text{Var}(Y|x) = \sigma^2$.

e. $E(Y|x) = \eta_0 + \eta_0 x$
$\text{Var}(Y|x) = \sigma^2$



As with the full model, we can "fit" a submodel using least squares:

*Example 1*: Submodel:

$$E(Y|x) = 2 + \eta_1 x$$

$$\text{Var}(Y|x) = \sigma^2,$$

Consider lines $y = 2 + h_1 x$

[picture!]

Minimize

$$RSS(h_1) = \sum d_i^2 = \sum [y_i - (2 + h_1 x_i)]^2$$

to get $\hat{\eta}_1$.

*Note*: For this example, $y_i - (2 + h_1 x_I) = (y_i - 2) - h_1 x_i$, so fitting this model is equivalent to fitting the model

$$E(Y|x) = \eta_1 x$$
$$\text{Var}(Y|x) = \sigma^2$$

to the transformed data

$(x_1, y_1 - 2), (x_2, y_2 - 2), \ldots , (x_n, y_n - 2)$

*Example 2*: Submodel

$$E(Y|x) = \eta_0$$

$$Var(Y|x) = \sigma^2$$

Minimize

$$RSS(h_0) = \sum d_i^2 = \sum (y_i - h_0)^2$$

[picture!]

Details:

Result: $h_0 = \bar{y}$ -- the same as the univariate estimate.

*Note*: This is also the same as setting $\hat{\eta}_1 = 0$ in the least squares fit for the full model.

*Caution*: The result is *not* always this nice, as the exercise below shows.



*Exercise*: Try finding the least squares fit for the submodel

$$E(Y|x) = \eta_1 x$$

$$Var(Y|x) = \sigma^2$$

("Regression through the origin")

You should get a *different* formula for $\hat{\eta}_1$ from that obtained by setting $\hat{\eta}_0 = 0$ in the formula for the least squares fit for the full model.

*Generalizing*: If we fit a submodel by Least Squares, we can define the residual sum of squares for the submodel:

$$RSS_{sub} = \sum(y_i - \hat{y}_{i,sub})^2 = \sum \hat{e}_{i,sub}^2$$

where

$$\hat{y}_{i,sub} = \hat{E}_{sub}(Y|x)$$

is the fitted value for the *submodel* and

$$\hat{e}_{i,sub} = y_i - \hat{y}_{i,sub}$$

*Example:* For the submodel

$$E(Y|x) = \eta_0$$
$$Var(Y|x) = \sigma^2,$$

$\hat{y}_{i,sub} = \bar{y}$ for each i, so

$$RSS_{sub} = \sum(y_i - \bar{y})^2 = SYY = (n-1)\, s_Y,$$

where $s_Y$ is the sample standard deviation of Y.



**General Properties**: (Stated without proof; true for multiple regression as well as simple regression)

i. $RSS_{sub}$ is a multiple of a $\chi^2$ distribution, with

ii. degrees of freedom $df_{sub}$ = n - (# of coefficients estimated), and

iii. $\hat{\sigma}_{sub}{}^2 = \dfrac{RSS_{sub}}{df_{sub}}$ is an estimate of $\sigma^2$ for the submodel.

This will allow us to do hypothesis tests comparing a submodel with the full model.

**Another Perspective**:

We want a way to help decide whether the full model is significantly better than the full model.

$RSS_{sub}$ - $RSS_{full}$ can be considered a measure of how much better the full model is than the submodel.

(Why is this difference always ≥ 0?).

But $RSS_{sub}$ - $RSS_{full}$ depends on the scale of the data,

so $\dfrac{RSS_{sub} - RSS_{full}}{RSS_{full}}$ is a better measure.

*Example* (to help build intuition): The submodel

$$E(Y|x) = \eta_0$$
$$Var(Y|x) = \sigma^2$$

Testing this model against the full model is equivalent to performing a hypothesis test with

NH: $\eta_1 = 0$
AH: $\eta_1 \neq 0$.



This hypothesis test uses the t-statistic

$$t = \frac{\hat{\eta}_1}{s.e.(\hat{\eta}_1)} = \frac{SXY\big/SXX}{\hat{\sigma}\big/\sqrt{SXX}} \sim t(n-2),$$

where here $\hat{\sigma} = \hat{\sigma}_{full}$ is the estimate of σ for the *full* model.

Note that

$$t^2 = \frac{(SXY)^2\big/(SXX)^2}{\hat{\sigma}^2\big/SXX} = \frac{(SXY)^2}{\hat{\sigma}^2(SXX)}$$

*Recall*:

$$RSS_{full} = SYY - \frac{(SXY)^2}{SXX}$$

$RSS_{sub}$ = SYY  (in this particular example)

Thus

$$RSS_{sub} - RSS_{full} = \frac{(SXY)^2}{SXX}.$$

so

$$t^2 = \frac{RSS_{sub} - RSS_{full}}{\hat{\sigma}^2}$$

$$= \frac{RSS_{sub} - RSS_{full}}{RSS_{full}/(n-2)}$$

$$= \frac{RSS_{sub} - RSS_{full}}{RSS_{full}} \div \text{(n-2)},$$

which is just a constant times our earlier measure

$\dfrac{RSS_{sub} - RSS_{full}}{RSS_{full}}$ of how much better the full model is

than the submodel.



**F Distributions**

*Recall*: A t(k) random variable has the distribution of a random variable of the form

where

Thus

$t(k)^2 \sim$

Also,

$Z^2 \sim$

Definition: An *F-distribution $F(\nu_1, \nu_2)$ with $\nu_1$ degrees of freedom in the numerator and $\nu_2$ degrees of freedom in the denominator* is the distribution of a random variable of the form

$$\frac{W/\nu_1}{U/\nu_2} \qquad \text{where} \quad W \sim \chi^2(\nu_1)$$

$$U \sim \chi^2(\nu_2)$$

and U and W are independent.

Thus:

$t(k)^2 \sim F(1, k)$;

i.e., the square of a t(k) random variable is an F(1,k) random variable – so any two-sided t-test could also be formulated as an F-test.

In particular, the hypothesis test with hypotheses

NH: $\eta_1 = 0$
AH: $\eta_1 \neq 0$

could be done using the F-statistic $t^2$ instead of the t-statistic .

Recall that in this example,

$$t^2 = \frac{RSS_{sub} - RSS_{full}}{RSS_{full}} \div (n-2),$$

which we have seen *does* make sense as a measure of whether the full model (corresponding to AH) is better than the submodel (corresponding to NH).

*Example*: Forbes data.



**Still another look at the F-statistic $t^2$:**

$$F = \frac{RSS_{sub} - RSS_{full}}{RSS_{full}/(n-2)}$$

$$= \frac{(RSS_{sub} - RSS_{full})/(df_{sub} - df_{full})}{RSS_{full}/df_{full}},$$

since $df_{sub} - df_{full} = (n - 1) - (n - 2) = 1$.

i.e., F is the ratio of:

Numerator:
the difference between the residual sum of squares for the submodel and the RSS for the full model

Denominator:
the residual sum of squares for the full model

*But*:
with each divided by its degrees of freedom to "weight" them appropriately to get a tractable distribution.

This is also just a constant times $\dfrac{RSS_{sub} - RSS_{full}}{RSS_{full}}$,
which is a reasonable measure of how much better the full model is than the submodel in fitting the data.

**This generalizes**: Whenever we have a submodel (in multiple linear regression as well as simple linear regression),

a. $RSS_{sub}$ (hence $\hat{\sigma}^2{}_{sub}$) will be a constant times a $\chi^2$ distribution, with degrees of freedom $df_{sub}$, which we then also refer to as the degrees of freedom of $RSS_{sub}$ and of $\hat{\sigma}^2{}_{sub}$.

b. $\dfrac{(RSS_{sub} - RSS_{full})/(df_{sub} - df_{full})}{\hat{\sigma}_{full}{}^2}$

$$= \frac{(RSS_{sub} - RSS_{full})/(df_{sub} - df_{full})}{RSS_{full}/df_{full}}$$

$\sim$ $F(df_{sub} - df_{full}, df_{full})$.



Rewriting the F-statistic,

$$\frac{(RSS_{sub} - RSS_{full})/(df_{sub} - df_{full})}{RSS_{full}/df_{full}}$$

$$= \left(\frac{RSS_{sub} - RSS_{full}}{RSS_{full}}\right)\left(\frac{df_{full}}{df_{fsub} - df_{full}}\right)$$

is just a constant multiple of $\dfrac{RSS_{sub} - RSS_{full}}{RSS_{full}}$, which is a reasonable measure of how much better the full model is than the submodel in fitting the data.

Thus we can use this F statistic for the hypothesis test

NH: Submodel
AH: Full model