1. F-test and ANOVA (cont'd)

① Type I Anova (sequential) The output depends on the order of the predictors in the mean function (of the input model)

Model 4 : $y \sim x^3 + x^2 + x$

For anova(Model 4), we have

the first row F-test : $H_0: E[Y|x] = \beta_0$ v.s $H_A: E[Y|x] = \beta_0 + \underline{\beta_3 x^3}$

the second row F-test: $H_0: E[Y|x] = \beta_0 + \beta_3 x^3$ v.s $H_A: E[Y|x] = \beta_0 + \beta_3 x^3 + \beta_2 x^2$

the third row F-test : $H_0: E[Y|x] = \beta_0 + \beta x^3 + \beta x^2$ v.s $H_A: E[Y|x] = \beta_0 + \beta_3 x^3 + \beta_2 x^2 + \beta_1 x$.

② Type II Anova

- Example 2.

- In example 2 (Type II Anova), the 3 F-tests give equivalent results to the 3 t-tests in the sumary output of Model 3.
(with quantitive predictors)

due to ① for these F-tests, the H_0 & H_A are the same as these t-tests. respectively.

② In this case

$$\underset{\substack{\wr H_0 \\ F\text{-dist}(1,\, n-(p+1))}}{F\text{-statistic}} = \underset{\substack{\wr H_0 \\ \sim \left(t\text{-dist},\ df = n-(p+1)\right)^2}}{(t\text{-statistic})^2}$$

- Note that generally, Type II Anova may not be the same as the t-tests.
i.g. when there are categorical (qualititive) predictors.
(This will be discussed in a later lecture).

- Type II Anova of Model 3 $y \sim x + x^2 + x^3$]
. . . . of model 4 $y \sim x^3 + x^2 + x$] give equavalent result,

(different from Type I Anova)

2. Prediction in linear Regression Model ('predict' function in R)

- Observed Data $(\underset{(p+1)\times 1}{x_i}, Y_i)$ $i=1,\cdots,n$ following a linear regression model.
- Suppose we have observed or will observe a New case with its predictors

given as $$\mathbf{x}^*_{(p+1)\times 1} = \begin{pmatrix} 1 \\ x_1^* \\ \vdots \\ x_p^* \end{pmatrix}_{(p+1)\times 1}$$

We want to predict the response of this observation, Y^* (unobserved yet).

- Under the linear regression model, we know

$$Y^* = (x^*)^T \underbrace{\beta}_{\text{true } \beta} + \underbrace{e^*}_{\text{statistical error}}$$

- Based on our OLS estimator $\hat{\beta}$ [from (x_i, Y_i), $i=1,\cdots,n$],

the predicted value of Y^* is

$$\hat{Y}^* = (x^*)^T \hat{\beta}.$$

- The standard error for the prediction of Y^* at x^* is

$$s.e.(\hat{Y}^*) = \sqrt{\widehat{Var}(\underbrace{(x^*)^T\hat{\beta}}_{\hat{Y}^*} \mid \mathbf{X}) + \widehat{Var}(e^*)}$$

from previous lecture,

$$Var(\hat{\beta}|x) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

$$Var((x^*)^T\hat{\beta} \mid x) = (x^*)^T \cdot Var(\hat{\beta}|x) \cdot (x^*)$$

$$= \sqrt{\hat{\sigma}^2 (x^*)^T(\mathbf{X}^T\mathbf{X})^{-1} x^* + \hat{\sigma}^2}$$

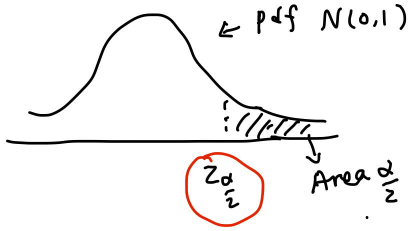
where $$\mathbf{X}_{n\times(p+1)} = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}_{n\times(p+1)}.$$

$\hat{\sigma}^2$ is the OLS estimator of σ^2 from observed data (x_i, Y_i), $i=1,\ldots,n$.

- Prediction confidence Interval (C. I.) with $(1-\alpha)$ level (e.g. $\alpha = 5\%$, 95% CI)

$$\left[\hat{y}^* - Z_{\frac{\alpha}{2}} \cdot s.e(\hat{y}^*) \ , \ \hat{y}^* + Z_{\frac{\alpha}{2}} \cdot s.e(\hat{y}^*) \right]$$

where



$$\frac{y^* - \hat{y}^*}{s.e.} \overset{\text{approximately}}{\sim} N(0,1) \quad \text{when } n \text{ is large}.$$

E.g. $\alpha = 5\%$ 95% CI is $\left[\hat{y}^* - 1.96\, s.e \ , \ \hat{y}^* + 1.96\, s.e \right]$ (approximately).

$Z_{0.025} \approx 1.96$

3. Simpson's paradox. (under linear Reg).

The regression coefficient of a predictor often changes sign when adding or removing another predictor.

E.g. $y \sim x_1$

$y \sim x_1 + x_2$

Example : Suppose the true model mean function is

$$E[Y \mid x_1, x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Fit $Y \sim X_1$, and the corresponding mean function of $Y \sim X_1$

$$E[Y \mid x_1] = E[\overbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \text{error}}^{Y} \mid x_1]$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 E(x_2 \mid x_1)$$

Case (1) If x_1 and x_2 are independent (uncorrelated) then $E(x_2 \mid x_1) = E(x_2)$.

$$\Rightarrow E[Y \mid x_1] = \beta_0 + \beta_2 E(x_2) + \beta_1 x_1$$

intercept of the mean of $Y \sim X_1$

slope of $Y \sim X_1$

<u>case (2)</u> If x_1 and x_2 are correlated,

Consider a simple case that $x_2 = \eta_0 + \eta_1 x_1 + \text{statistical error}$

$E[x_2 | x_1] = \eta_0 + \eta_1 x_1$, $(\eta_1 \neq 0)$

$$\Rightarrow \quad E[Y | x_1] = \beta_0 + \beta_1 x_1 + \beta_2 (\eta_0 + \eta_1 x_1)$$

$$= \underbrace{\beta_0 + \beta_2 \eta_0}_{\text{intercept of } Y \sim X_1} + \underbrace{(\beta_1 + \beta_2 \eta_1)}_{\text{slope of } Y \sim X_1} X_1$$

Now

slope for x_1 of $Y \sim X_1$: $\beta_1 + \beta_2 \eta_1$

slope for x_1 of $Y \sim X_1 + X_2$: β_1

It ~~can~~ may happen in practice that β_1 and $\beta_1 + \beta_2 \eta_1$ have different sign $\rightarrow$ so-called Simpson's paradox