(sequential) ANOVA $\overset{cf.}{\longleftrightarrow}$ t-tests (LNp.2-23) & F-test (LNp.2-25)

LM, LNp.8-18~19

- anova($y{\sim}1{+}A{+}B{+}A{:}B$), A: 3 levels, B: 4 levels ⇒ $A{:}B$ 2×3=6 dummy var.

2 dummy var. 3 dummy var.

drop one from the full model

1) test ω:model **1** ($y{\sim}1$) against Ω:model **2** ($y{\sim}1{+}A$) [$df_\omega - df_\Omega = 2$]

2) test ω:model **2** ($y{\sim}1{+}A$) against Ω:model **4** ($y{\sim}1{+}A{+}B$) [$df_\omega - df_\Omega = 3$]

3) test ω:model **4** ($y{\sim}1{+}A{+}B$) against Ω:model **5** ($y{\sim}1{+}A{+}B{+}A{:}B$) [$df_\omega - df_\Omega = 6$]

- $$F = \frac{(RSS_\omega - RSS_\Omega)/(df_\omega - df_\Omega)}{RSS_{\text{model 5}}/df_{\text{model 5}}} \sim F_{df_\omega - df_\Omega, df_{\text{model 5}}}$$ cf.

$\hat{\sigma}^2_{\text{full model}}$ ← full model

- invariant to the choice of dummy variables since they generate same ω and Ω

- ANOVA could have different results when the order of effect sequence is changed, e.g., anova($y{\sim}1{+}B{+}A{+}A{:}B$):

cf. α) test ω:model **1** ($y{\sim}1$) against Ω:model **3** ($y{\sim}1{+}B$) [$df_\omega - df_\Omega = 3$]

cf. β) test ω:model **3** ($y{\sim}1{+}B$) against Ω:model **4** ($y{\sim}1{+}B{+}A$) [$df_\omega - df_\Omega = 2$]

χ) test ω:model **4** ($y{\sim}1{+}B{+}A$) against Ω:model **5** ($y{\sim}1{+}B{+}A{+}A{:}B$) [$df_\omega - df_\Omega = 6$]

- anova($y{\sim}1{+}A{+}B{+}A{:}B$) and anova($y{\sim}1{+}B{+}A{+}A{:}B$) will have **identical** results when orthogonality exists between the three groups of effects: span$\{d_i^A\}$, span$\{d_j^B\}$, span$\{d_{ij}^{A:B}\}$, because in the case, $RSS_\omega - RSS_\Omega$ would equal for 1) and β), 2) and α), 3) and χ) → furthermore, also identical to the drop-one t-tests & F-tests

➢ consider the *full model*: $\overset{cf.}{\longleftrightarrow}$ submodels

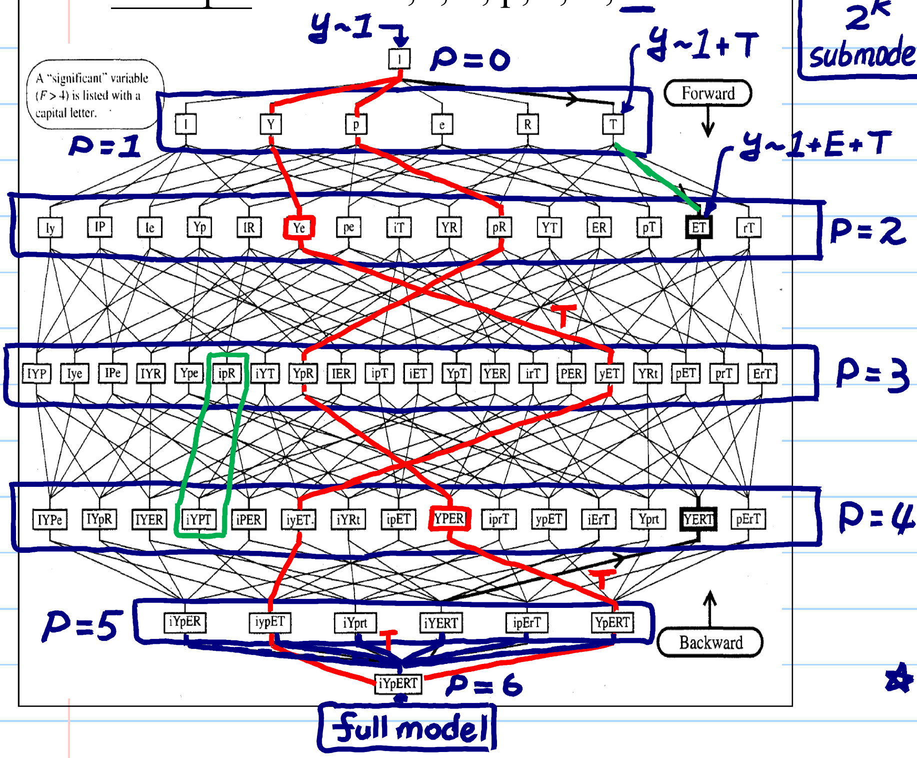
$$y = \beta_0 + \beta_1 g_1(x_1, \ldots, x_m) + \beta_2 g_2(x_1, \ldots, x_m) + \cdots + \beta_k g_k(x_1, \ldots, x_m) + \epsilon$$

For $1 \le i \le k$, should the term $\beta_i g_i$ be included in the *final fitted model*? ← main purpose of performing t- & F-tests

— : sequential (Type I ANOVA)

— : drop one (Type III ANOVA)

Example: 6 effects, I, Y, p, e, R, T



➢ (sub-)model: a model with a subset of all k terms, e.g.,

2^k submodels

- $\{1, g_1, g_2\}$, → $y{\sim}1{+}g_1{+}g_2$
- $\{1, g_2, g_4, g_5, g_k\}$, … → $y{\sim}1{+}g_2{+}g_4{+}g_5{+}g_k$

➢ hierarchical structure of all sub-models (see graph)

- p = # of terms in a sub-model
- # of different sub-models=2^k
- connecting line: model nesting

★ can be generated to "groups" of effects

LM. LNp.5-8~9 → Orthogonality

- **Q**: consider the two models: $Y = X_1\beta_1 + \varepsilon$ $\quad Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

  $X_1(X_1^TX_1)^{-1}X_1^T$ (hat matrix, a projection matrix)

  $HX_2\beta_2 \quad (I-H)X_2\beta_2$

  model **1**: $y = \beta_0 + \beta_1 x_1 + \varepsilon$, $\quad$ model **2**: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

  In general, $\hat{\beta}_1$, in the 2 models are not identical (of course, test H_0: $\beta_1 = 0$ not identical neither)

  an exception: when x_1 and x_2 are orthogonal

fitted model = model **1**, What if $X_1^TX_2 = 0$?
true model = model **2**
$E(\hat{\beta}_1) = \beta_1 + (X_1^TX_1)^{-1}X_1^TX_2\beta_2$

- $Y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon$, where $\beta = [\beta_1 \ \ \beta_2]^T$ and $X = [X_1 \ \ X_2]$ with the property

  $X_1^TX_2 = 0 \Rightarrow X_1$ and X_2 are orthogonal $\quad \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (X^TX)^{-1}X^TY = \begin{bmatrix} (X_1^TX_1)^{-1}X_1^TY \\ (X_2^TX_2)^{-1}X_2^TY \end{bmatrix}$

  span$\{X_1\} \perp$ span$\{X_2\}$

  $$X^TX = \begin{pmatrix} X_1^TX_1 & X_1^TX_2 \\ X_2^TX_1 & X_2^TX_2 \end{pmatrix} = \begin{pmatrix} X_1^TX_1 & 0 \\ 0 & X_2^TX_2 \end{pmatrix} \Rightarrow (X^TX)^{-1} = \begin{pmatrix} (X_1^TX_1)^{-1} & 0 \\ 0 & (X_2^TX_2)^{-1} \end{pmatrix}$$

- Estimation: $\hat{\beta}_1 = (X_1^TX_1)^{-1}X_1^TY$, $\hat{\beta}_2 = (X_2^TX_2)^{-1}X_2^TY$, and $\hat{\beta}_1, \hat{\beta}_2$ independent

  $\Rightarrow$ note that $\hat{\beta}_1$ will be the same regardless of whether X_2 is in the model or not (and vise versa). Under model $Y = X_1\beta_1 + \varepsilon \Rightarrow \hat{\beta}_1 = (X_1^TX_1)^{-1}X_1^TY$

  Q: what if only two predictors, say some x_i in X_1 and some x_j in X_2, are orthogonal?

- Randomization: In an exp't, suppose that true model is $Y = X\beta + Z\gamma + \varepsilon$, but Z cannot be measured or may not even be suspected $\Rightarrow E(\hat{\beta}) = \beta + (X^TX)^{-1}X^TZ\gamma \Rightarrow$ $\boxed{X, Z \text{ indep.}}$

  Q: what's the best way of controlling X to make X and Z as orthogonal as possible?

- Generalization.

$$Y = \beta_0 \mathbb{1} + X_1\beta_1 + X_2\beta_2 + \cdots + X_k\beta_k + \varepsilon$$

$\boxed{W_0 = \text{span}\{\mathbb{1}\}}$ $\boxed{W_1 = \text{span}\{X_1\}}$ $\boxed{W_2 = \text{span}\{X_2\}}$ $\boxed{W_k = \text{span}\{X_k\}}$

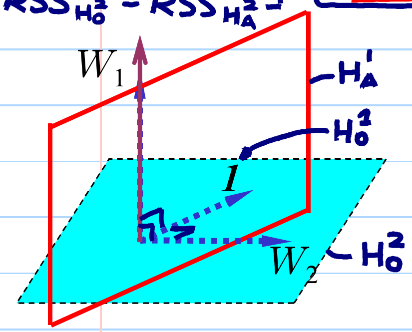
Suppose that $W_0 \perp W_1 \perp W_2 \perp \cdots \perp W_k$ $\Rightarrow X^TX = \begin{bmatrix} n & & & & 0 \\ & X_1^TX_1 & & & \\ & & X_2^TX_2 & & \\ & & & \ddots & \\ 0 & & & & X_k^TX_k \end{bmatrix}$

orthogonal

$\Downarrow$

$\hat{\beta}_i = (X_i^TX_i)^{-1}X_i^TY$

$RSS_{H_0^1} - RSS_{H_A^1}$
$RSS_{H_0^2} - RSS_{H_A^2}$ $\Big] \boxed{\|P_{W_1}(y)\|^2}$



H_0^1: span$\{\mathbb{1}\}$ $\quad H_A^1$: span$\{\mathbb{1}, W_1\}$

H_0^2: span$\{\mathbb{1}, W_2\}$ $\quad H_A^2$: span$\{\mathbb{1}, W_1, W_2\}$

H_0: $y = \beta_0\mathbb{1} + X_2\beta_2 + X_5\beta_5 + \cdots + \varepsilon$ $\quad$ irrelevant

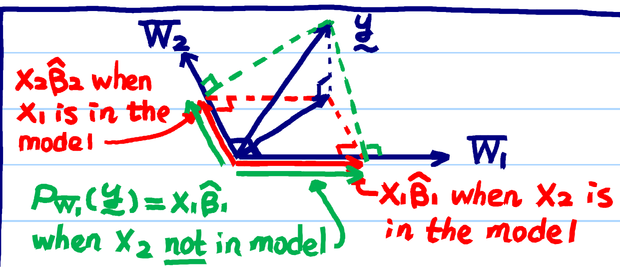
$\qquad\qquad \|$

H_A: $y = \beta_0\mathbb{1} + X_2\beta_2 + X_5\beta_5 + \cdots + X_1\beta_1 + \varepsilon$

(LNp.25) $RSS_{H_0} - RSS_{H_A} = \|P_{W_1}(y)\|^2$

$$\|y - \bar{y}\mathbb{1}\|^2 = \|P_{\Omega}(y)\|^2 + \|P_{\Omega^\perp}(y)\|^2$$
$$= \|P_{W_1}(y)\|^2 + \|P_{W_2}(y)\|^2 + \cdots + \|P_{W_k}(y)\|^2 + \|P_{\Omega^\perp}(y)\|^2$$

not true if not orthogonal



❖ **Reading**: Textbook, 1.4~1.6, 1.8

Some Properties of (Multivariate) Normal Distribution

(N1) Linear transformation of normal is still normal

$E^*[(AZ+c)(AZ+c)^T] = E^*[(AZ)(AZ)^T] = E^*[AZZ^TA^T]$

n×1 vector, n×n matrix

$$\underline{Z} \sim N(\mu, \Sigma) \Rightarrow AZ + c \sim N(A\mu + c, A\Sigma A^T).$$

$cov(Z) = \begin{bmatrix} cov(Z_1) & cov(Z_1, Z_2) \\ T & cov(Z_2) \end{bmatrix}$

(N2) When 1st and 2nd moments are given, the normal distribution is specified.

i.e., mean vector & variance-covariance matrix

(N3) $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim$ normal, and Z_1, Z_2 uncorrelated (i.e., $cov(Z_1, Z_2) = \mathbf{0}$)

$\Rightarrow Z_1, Z_2$ independent

Z can be generalized to k

By (N3)

normal $W = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} Z$, $cov(W_1, W_2) = E^*(W_1 W_2^T) = E^*(A_1 Z Z^T A_2^T) = A_1 E^*(ZZ^T) A_2^T$

(N4) $Z \sim N(\mu, \Sigma)$, $W_1 = A_1 Z$, $W_2 = A_2 Z$

Z can be generalized to k & $A_i \Sigma A_j^T = 0$, $i \neq j$

By (N3) (N4)

$\Rightarrow W_1, W_2$ are independent iff $A_1 \Sigma A_2^T = \mathbf{0}$. → If $\Sigma = \sigma^2 I$, then $A_1 \Sigma A_2^T = 0 \Leftrightarrow A_1 A_2^T = 0$

(N5) $Z \sim N(\mu, \Sigma)$, $W_1 = A_1 Z$, $W_2 = A_2 Z$, $\ldots$, $W_k = A_k Z$, and $cov(W_i, W_j) = \mathbf{0}$ for $1 \le i < j \le k$, $\Rightarrow W_1^T W_1, W_2^T W_2, \ldots, W_k^T W_k$ are mutually independent.

length² of W_i

useful for the independence between sums of squares

(N6) Z: an $n \times 1$ random vector and $Z \sim N(\mu, \Sigma)$, then

$\Sigma^{-1} = (\Sigma^{-1/2})^T (\Sigma^{-1/2})$

– if Σ is non-singular, $(Z - \mu)^T \Sigma^{-1} (Z - \mu) \sim \chi^2_n$

$\Sigma^{-1/2}(Z - \mu) \sim N(0, I)$ ← standardization

– if Σ is singular and has rank r ($< n$), → The possible vectors of Z only occupy an r-dim subspace of $\mathbb{R}^n$

let Σ^- be a generalized inverse of Σ (i.e., $\Sigma\Sigma^-\Sigma = \Sigma$), then

not unique

$$(Z - \mu)^T \Sigma^- (Z - \mu) \sim \chi^2_r$$