STAT 714 fa 2023 Lec 05

Inference in linear models, general linear hypothesis

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Uniform minimum variance unbiased estimator.

BLUE : Best linear unbiased estimator.

1. UMVUE in the linear model under Normality
2. The general linear hypothesis, ad-hoc test
3. Likehood ratio test of the general linear hypothesis
4. Multiple testing in linear models

Assume throughout that $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

We know that $\mathbf{c}^T\hat{\mathbf{b}}$ is the BLUE for $\mathbf{c}^T\mathbf{b}$, but is it the UMVUE?

Result (UMVUE for an estimable contrast under Normality)

Let $\mathbf{c}^T\mathbf{b}$ be estimable in the model $\boxed{\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e},}$ where $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, and suppose $\hat{\mathbf{b}}$ satisfies $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$. Then $\mathbf{c}^T\hat{\mathbf{b}}$ is UMVUE for $\mathbf{c}^T\mathbf{b}$.

See Res 6.2 and Cor 6.2 of Monahan (2008). Recall Lehmann-Scheffé.

Recall definition of UMVUE and prove the result.

$$\underset{\sim}{y} \sim N\left(X\underset{\sim}{b},\ \sigma^2 I_n\right)$$

$$f(\underset{\sim}{y}; \underset{\sim}{b}, \sigma^2) = (2\pi)^{-n/2} \left|\sigma^2 I_n\right|^{-1/2} \exp\left[-\frac{1}{2}\left(\underset{\sim}{y} - X\underset{\sim}{b}\right)^T \left[\sigma^2 I_n\right]^{-1} \left(\underset{\sim}{y} - X\underset{\sim}{b}\right)\right]$$

$$= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\left(y^T y - 2y^T X b + b^T X^T X b\right)\right]$$

$$\left|\begin{matrix}\sigma^2 & & \\ & \ddots & \\ & & \sigma^2\end{matrix}\right|_{n \times n} = (\sigma^2)^n$$

$$= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left[-\frac{b^T X^T X b}{2\sigma^2}\right] \exp\left[y^T y\left(-\frac{1}{2\sigma^2}\right) + y^T X\, b\left(\frac{1}{\sigma^2}\right)\right]$$

$\Rightarrow \quad T = \left(y^T y,\; y^T X\right)$ is a compl. suff stat.

Recall:

$$f(y;\theta) = h(y)\, c(\theta) \exp\left[\sum_{j=1}^{k} t_j(y)\, w_j(\theta)\right]$$

$t_j(y) \quad j = 1, \dots, k$ are the complete suff. statistics.

If $c^T b$ is estimable then $c = X^T a$, so

$$c^T b = a^T X b.$$

Then

$$c^T \hat{b} = a^T X \hat{b} = a^T P_X y = a^T X (X^T X)^- X^T y$$

function of complete. suff stat.

$$\hat{\sigma}^2 = \frac{\|\hat{e}\|^2}{n}$$

Result (MLEs in Normal model)

Suppose $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, *where* $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, *and let* $\mathbf{c}^T\mathbf{b}$ *be estimable. The MLE of* $(\mathbf{c}^T\mathbf{b}, \sigma^2)$ *is given by* $(\mathbf{c}^T\hat{\mathbf{b}}, \|\hat{\mathbf{e}}\|^2/n)$, *where* $\hat{\mathbf{b}}$ *satisfies* $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$.

See Res 6.3 and Cor 6.3 of Monahan (2008).

Derive the result.

$$f(y; b, \sigma^2) = (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \|y - Xb\|^2 \right]$$

$$= L(b, \sigma^2; y)$$

$$\ell(b, \sigma^2; y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\|y - Xb\|^2$$

$$\frac{\partial}{\partial \underset{\sim}{b}} \ell(\underset{\sim}{b}, \sigma^2; \underset{\sim}{y}) = -X^T(\underset{\sim}{y} - X\underset{\sim}{b})\frac{1}{\sigma^2} \overset{set}{=} 0$$

$$\Leftrightarrow \boxed{X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y}}$$

MLE

$$\frac{\partial}{\partial \sigma^2} \ell(\hat{\underset{\sim}{b}}, \sigma^2; \underset{\sim}{y}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\|y - X\hat{b}\|^2 \overset{set}{=} 0$$

$$\Rightarrow \boxed{\hat{\sigma}^2_{mle} = \frac{1}{n}\|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2.}$$

$$= \frac{1}{n} \underset{\sim}{y}^T (I - P_X) \underset{\sim}{y}$$

$(\hat{b}, \hat{\sigma}^2)$ is an MLE for (b, σ^2) if

$$L(\hat{b}, \hat{\sigma}^2; \underset{\sim}{y}) \geq L(\underset{\sim}{b}, \sigma^2; \underset{\sim}{y}) \quad \forall \quad \underset{\sim}{b}, \sigma^2.$$

Since $(\hat{b}, \hat{\sigma}^2)$ is MLE for (b, σ^2)

then $(\underset{\sim}{c}^T \hat{\underset{\sim}{b}}, \hat{\sigma}^2)$ is MLE for $(\underset{\sim}{c}^T \underset{\sim}{b}, \sigma^2)$.

The general linear hypothesis

rank $X = r$

In the context of the model $\underset{n\times 1}{\mathbf{y}} = \underset{n\times p}{\mathbf{X}}\underset{p\times 1}{\mathbf{b}} + \underset{n\times 1}{\mathbf{e}}$, the *general linear hypothesis* is expressed

$$H_0\colon \underset{s\times p}{\mathbf{K}^T}\underset{p\times 1}{\mathbf{b}} = \underset{s\times 1}{\mathbf{m}} \quad \text{versus} \quad H_1\colon \mathbf{K}^T\mathbf{b} \neq \mathbf{m},$$

where $\mathbf{K}$ is a $p \times s$ matrix and $\mathbf{m}$ is a $s \times 1$ vector.

$c^T b$ is estimable $\iff c \in \text{Col}\, X^T$

The general linear hypothesis is *testable* provided that:

1. Each column of $\mathbf{K}$ defines an estimable contrast. Need $K = X^T A$, some A.
2. The columns of $\mathbf{K}$ are linearly independent (contrasts are non-redundant). rank $K = s$.

If one or the other condition does not hold, the hypothesis is *non-testable*.

See Def 6.1 of Monahan (2008).

Exercise: Consider the treatment effect model

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2), \quad j = 1, \dots, n_i,\ i = 1, \dots, 3$$

Formulate the general linear hypothesis for testing for a treatment effect.

$\underset{\sim}{y} = X\underset{\sim}{b} + \underset{\sim}{e}$, $\quad \underset{\sim}{e} \sim \text{Normal}\left(\underset{\sim}{0}, \sigma^2 I_n\right)$. $\quad y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

$$X = \begin{bmatrix} \underset{\sim}{1}_{n_1} & \underset{\sim}{1}_{n_1} & & \\ \underset{\sim}{1}_{n_2} & & \underset{\sim}{1}_{n_2} & \\ \underset{\sim}{1}_{n_3} & & & \underset{\sim}{1}_{n_3} \end{bmatrix} \qquad \underset{\sim}{b} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

Can we do this?

$$K^T = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow \quad H_0 : K^T \underset{\sim}{b} = \underset{\sim}{0} \ ?$$

Try $K^T = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$, $\underset{\sim}{m} = \underset{\sim}{0}$.

Then $K^T \underset{\sim}{b} = \begin{bmatrix} \alpha_1 - \alpha_2 \\ \alpha_2 - \alpha_3 \end{bmatrix}$. $\qquad H_0 : \begin{bmatrix} \alpha_1 - \alpha_2 \\ \alpha_2 - \alpha_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

estimable

Means there is no variation among the treatment means.

$$K^T \underset{\sim}{b} - \underset{\sim}{m} = \underset{\sim}{0} \iff \begin{bmatrix} \alpha_1 - \alpha_2 \\ \alpha_2 - \alpha_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

means $K = X^T A$ for some A

Result (Helper result for building test statistic for gen. lin. hyp.)

If H_0: $\mathbf{K}^T\mathbf{b} = \mathbf{m}$ is $\boxed{\text{testable,}}$ then the matrix $\underbrace{\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{K}}_{s \times s}$ is invertible.

See Res 6.4 of Monahan (2008).

Prove the result.

Show that $\operatorname{rank}\left(K^T(X^TX)^-K\right)$ is s.

Write $\underbrace{K^T(X^TX)^-K}$

$$
\begin{aligned}
K^T(X^TX)^-K &= A^T X (X^TX)^- X^T A \\
&= \underline{A^T} X (X^TX)^- X^T X (X^TX)^- X^T A \\
&= K^T (X^TX)^- X^T \quad X (X^TX)^- K
\end{aligned}
$$

$$= B^T B, \quad \text{where} \quad \underline{B = X(X^TX)^- K}.$$

Then $\text{rank}(B^TB) = \text{rank } B$.

$\underline{\text{Col } X^TX} = \text{Col } X^T \implies \text{rank}(X^TX) = \text{rank } X^T = \text{rank } X$

Now show that rank B is equal to s.

On one hand $\text{rank}(B) = \text{rank}\big(\underbrace{X(X^TX)^-}\underbrace{K}\big) \leq \text{rank } K = s$

On the other hand $\text{rank}(B) \geq \text{rank}(X^TB) = \text{rank}\big(X^TX(X^TX)^-K\big)$

$$= \text{rank}\big(X^T\underbrace{X(X^TX)^-}_{\text{g inv of } X^T}X^TA\big)$$

$$= \text{rank}(X^TA)$$

$$= \text{rank}(K)$$

$$= s$$

$\text{rank}(AB) \leq \min\{\text{rank } A, \text{rank } B\}$

So $s \leq \text{rank}(B) \leq s$.

Result (Chi-square results for quadratic forms in Normals.)

Let $\mathbf{y} \sim \text{Normal}(\boldsymbol{\mu}, \mathbf{V})$, $\mathbf{V}$ *a* $p \times p$ *positive definite matrix.*

1. *We have* $\mathbf{y}^T\mathbf{V}^{-1}\mathbf{y} \sim \chi^2_p(\phi = \boldsymbol{\mu}^T\mathbf{V}^{-1}\boldsymbol{\mu})$

pf: $\underset{\sim}{y}^T V^{-1} \underset{\sim}{y} = \underset{\sim}{y}^T R R \underset{\sim}{y}$ $V^{-1} = RR$, R symm, invertible

$= \underset{\sim}{y}^T R^T R \underset{\sim}{y}$

$= (R\underset{\sim}{y})^T (R\underset{\sim}{y})$, $R\underset{\sim}{y} \sim N(R\underset{\sim}{\mu}, \underbrace{RVR})$

Use results in lec 4

$\sim \chi^2_p(\phi = \underset{\sim}{\mu}^T R^T R \underset{\sim}{\mu})$

$= \chi^2_p(\phi = \underset{\sim}{\mu}^T V^{-1} \underset{\sim}{\mu})$

$R(RR)^{-1}R$
$\shortparallel$
$R R^{-1} R^{-1} R$
$\shortparallel$
I_p

so $R\underset{\sim}{y} \sim N(R\underset{\sim}{\mu}, I_p)$

$\text{Cov}(K^T\hat{\underset{\sim}{b}}) = \text{Cov}(A^T X (X^TX)^- X^T \underset{\sim}{y}) = \sigma^2 A^T X (X^TX)^- X^T X (X^TX)^- X^T A$

$= \sigma^2 A^T X (X^TX)^- X^T A$

$= \sigma^2 K^T (X^TX)^- K$

Let $\underset{\sim}{y} = X\underset{\sim}{b} + \underset{\sim}{e}$, $\underset{\sim}{e} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$.

Want to test

$$H_0: \underset{s\times p}{K^T} \underset{p\times 1}{\underset{\sim}{b}} = \underset{\sim}{m} \quad \text{vs} \quad H_1: K^T \underset{\sim}{b} \neq \underset{\sim}{m},$$

where $\underset{p\times s}{K} = X^T A$ for some A (K defines estimable contrasts).

Need rank $K = s$.

E.g.

$$\underset{\sim}{b} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} \qquad K^T = \begin{bmatrix} 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

$$K^T \underset{\sim}{b} = \underset{\sim}{0} \iff \begin{bmatrix} \alpha_1 - \alpha_2 \\ \alpha_2 - \alpha_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Can we find a test?

Formulate hypotheses as $H_0: K^T \underset{\sim}{b} - \underset{\sim}{m} = \underset{\sim}{0}$ vs $H_1: K^T \underset{\sim}{b} - \underset{\sim}{m} \neq \underset{\sim}{0}$.

Look at value of $K^T \hat{b} - \underset{\sim}{m}$. Ask how far from $\underset{\sim}{0}$.

We have

$$K^T \underset{\sim}{\hat{b}} - \underset{\sim}{m} \sim N\left(K^T \underset{\sim}{b} - \underset{\sim}{m}, \ \sigma^2 K^T (X^T X)^- K \right)$$

$$K^T \underset{\sim}{\hat{b}} = A^T X \underset{\sim}{\hat{b}} = A^T P_X \underset{\sim}{y} = A^T \left(P_X (X \underset{\sim}{b} + \underset{\sim}{e} \right)$$

$X^T X \underset{\sim}{\hat{b}} = X^T \underset{\sim}{y}$

$$= A^T X \underset{\sim}{b} + A^T P_X \underset{\sim}{e}$$

$$= K^T \underset{\sim}{b} + A^T X (X^T X)^- X^T \underset{\sim}{e}$$

$$= K^T \underset{\sim}{b} + K^T (X^TX)^- X^T \underset{\sim}{e}$$

$$\mathbb{E}\left(K^T \hat{\underset{\sim}{b}} - \underset{\sim}{m}\right) = K^T \underset{\sim}{b} - \underset{\sim}{m}$$

$$\begin{aligned}
\text{Cov}\left(K^T \hat{\underset{\sim}{b}} - \underset{\sim}{m}\right) &= \text{Cov}\left(K^T (X^TX)^- X^T \underset{\sim}{e}\right) \\
&= K^T (X^TX)^- X^T \, \text{Cov}(\underset{\sim}{e}) \, X (X^TX)^- K \\
&= \sigma^2 K^T (X^TX)^- X^T X (X^TX)^- K \\
&= \sigma^2 A^T \underbrace{X (X^TX)^- X^T X}_{X} (X^TX)^- X^T A \\
&= \sigma^2 A^T X (X^TX)^- X^T A \\
&= \sigma^2 K^T (X^TX)^- K
\end{aligned}$$

$$y_i = \sum x_{ij}\beta_j + e_i$$

$$\underset{\sim}{b} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad K^T = [1\ 0 \cdots 0]$$

$$K^T \underset{\sim}{b} = \underset{\sim}{0} \iff \beta_1 = 0.$$

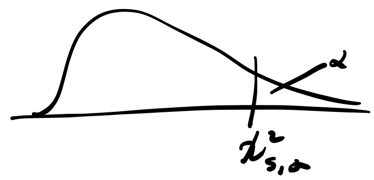
Under H_0: $K^T \underset{\sim}{b} - \underset{\sim}{m} = 0$, we have

$$K^T \hat{\underset{\sim}{b}} - \underset{\sim}{m} \sim N\left(\underset{\sim}{0}, \sigma^2 K^T (X^TX)^- K\right)$$

Idea: look at distance from $\underset{\sim}{0}$ of $K^T \hat{\underset{\sim}{b}} - \underset{\sim}{m}$.

$$\underbrace{\left(K^T \hat{\underset{\sim}{b}} - \underset{\sim}{m}\right)^T}_{\underset{\sim}{y}} \underbrace{\left[\sigma^2 K^T (X^TX)^- K\right]^{-1}}_{V^{-1}_{s\times s}} \left(K^T \hat{\underset{\sim}{b}} - \underset{\sim}{m}\right) \sim \begin{cases} \chi^2_s \text{ under } H_0 \\ \chi^2_s\left(\phi = \frac{1}{\sigma^2}\left(K^T \underset{\sim}{b} - \underset{\sim}{m}\right)^T \left[K^T (X^TX)^- K\right]^{-1} \left(K^T \underset{\sim}{b} - \underset{\sim}{m}\right)\right) \end{cases}$$

If I know σ^2, we test



Reject H_0 if

$$\underbrace{\big(\underset{s\times p}{K^T}\underset{p\times 1}{\hat{\underset{\sim}{b}}} - \underset{s\times 1}{\underset{\sim}{m}}\big)^T}_{1\times s} \underbrace{\left[\sigma^2 K^T(X^TX)^- K\right]^{-1}}_{s\times s} \underbrace{\big(K^T\hat{b} - \underset{\sim}{m}\big)}_{s\times 1} > \chi^2_{s,\alpha}$$

Replace σ^2 with $\hat\sigma^2$?

$$\frac{1}{\hat\sigma^2}\underbrace{\big(\underset{s\times p}{K^T}\underset{p\times 1}{\hat{\underset{\sim}{b}}} - \underset{s\times 1}{\underset{\sim}{m}}\big)^T}_{1\times s} \underbrace{\left[K^T(X^TX)^- K\right]^{-1}}_{s\times s} \underbrace{\big(K^T\hat{b} - \underset{\sim}{m}\big)}_{s\times 1}$$

$$\hat\sigma^2 = \frac{\|\hat{\underset{\sim}{e}}\|^2}{n-r} = \frac{\underset{\sim}{y}^T(I-P_X)\underset{\sim}{y}}{n-r}$$

divide by s:

$$\frac{\frac{1}{\sigma^2}\big(\underset{s\times p}{K^T}\underset{p\times 1}{\hat{\underset{\sim}{b}}} - \underset{s\times 1}{\underset{\sim}{m}}\big)^T \left[K^T(X^TX)^- K\right]^{-1}\big(K^T\hat{b} - \underset{\sim}{m}\big) \Big/ s}{\frac{\underset{\sim}{y}^T(I-P_X)\underset{\sim}{y}}{\sigma^2} \Big/ (n-r)}, \quad \text{where } r = \text{rank } X.$$

$$= \frac{\chi^2_s\left(\phi = (K^T\underset{\sim}{b}-m)^T\left[K^T(X^TX)^-K\right](K^T\underset{\sim}{b}-m)/\sigma^2\right)/s}{\chi^2_{n-r}/(n-r)}$$

$$\sim F_{s,n-r}\left(\phi = \frac{(K^T\underset{\sim}{b}-m)^T\left[K^T(X^TX)^-K\right](K^T\underset{\sim}{b}-m)}{\sigma^2}\right)$$

$$\boxed{\begin{array}{ll} A = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} & AGA = A \\ & GAG \neq G \\ G = I & \end{array}}$$

Now reject $H_0: \; K^T \underset{\sim}{b} - \underset{\sim}{m} = \underset{\sim}{0}$ if

$$\frac{\left(\underset{s\times p}{K^T}\underset{p\times 1}{\hat{\underset{\sim}{b}}} - \underset{s\times 1}{\underset{\sim}{m}}\right)^T \left[K^T (X^T X)^- K\right]^{-1} \left(K^T \hat{\underset{\sim}{b}} - \underset{\sim}{m}\right) / s}{\underset{\sim}{y}^T (I - P_X) \underset{\sim}{y} \; / \; (n-r)} > F_{s, n-r, \alpha}$$

The test has size α.

Exercise: Assume $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2\mathbf{I}_n)$ and consider the *ad-hoc rule*

$$F = \frac{(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})^T[\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{K}]^{-1}(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})/s}{\mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{y}/(n-r)} > F_{s,n-r,\alpha},$$

for rejecting testable H_0: $\mathbf{K}^T\mathbf{b} = \mathbf{m}$, where $s = \operatorname{rank}\mathbf{K}$ and $r = \operatorname{rank}\mathbf{X}$.

1. Give the size of the test.
2. Give an expression for the power of the test.
3. Check if the test is unbiased (has greater power over H_1 than under H_0).

$$\gamma(\underset{\sim}{b}) = P\left(F > F_{s,n-r,\alpha}\right),$$

$$F \sim F_{s,n-r}\left(\phi = \frac{(K^T\underset{\sim}{b} - m)^T\,[K^T(X^TX)^-K]\,(K^T\underset{\sim}{b} - m)}{\sigma^2}\right)$$

$$\underset{n\times p}{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & & x_{np} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{x}_1^T \\ \vdots \\ \underset{\sim}{x}_n^T \end{bmatrix} \qquad \underset{\sim}{b} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \qquad K^T = \underbrace{[\,0 \cdots 0 \;\; 1 \;\; 0 \cdots 0\,]}_{j^{th}} = \underset{\sim}{e}_j^T$$

$$\underset{\sim}{m} = 0$$

$$K^T \underset{\sim}{b} = \underset{\sim}{m} \implies \beta_j = 0.$$

Exercise: Consider the multiple linear regression model

$$Y_i = \cancel{\beta_0 +} \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad \varepsilon_i \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2), \quad i = 1, \dots, n.$$

Assume a full-rank design matrix.

$$H_0: K^T \underset{\sim}{b} = \underset{\sim}{m}$$

1. Give the size α ad-hoc rule for testing $\boxed{H_0: \beta_j = 0}$ for any $j = 1, \dots, p$.
2. Give an expression of the power of the test.

$$\hat{\underset{\sim}{b}} = \begin{bmatrix} \hat\beta_1 \\ \vdots \\ \hat\beta_p \end{bmatrix} = (X^TX)^{-1} X^T \underset{\sim}{y}$$

$$F = \frac{(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})^T[\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^- \mathbf{K}]^{-1}(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})/s}{\mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{y}/(n-r)} > F_{s,n-r,\alpha},$$

$$= [\underset{\sim}{e}_j^T \hat{\underset{\sim}{b}} - 0]^T [\underset{\sim}{e}_j^T (X^TX)^{-1} \underset{\sim}{e}_j]^{-1} [\underset{\sim}{e}_j^T \hat{\underset{\sim}{b}} - 0] / \hat\sigma^2$$

$$= \hat\beta_j \, [\, [(X^TX)^{-1}]_{jj} \,]^{-1} \, \hat\beta_j / \hat\sigma^2$$

$$= \frac{\hat{\beta}_j^2}{\hat{\sigma}^2 \left[(X^T X)^{-1} \right]_{jj}}$$

Reject H_0 when $\frac{\hat{\beta}_j^2}{\hat{\sigma}^2 \left[(X^T X)^{-1} \right]_{jj}} > F_{1, n-p, \alpha}$

or

$$\frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 [(X^T X)^{-1}_{jj}]}} < \underbrace{-\sqrt{F_{1,n-p,\alpha}}}_{t_{n-p,\alpha/2}} \quad \text{or} \quad \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 [(X^T X)^{-1}_{jj}]}} > \underbrace{\sqrt{F_{1,n-p,\alpha}}}_{t_{n-p,\alpha/2}}$$

Reject H_0 if $\frac{|\hat{\beta}_j|}{\hat{\sigma}\sqrt{[(X^T X)^{-1}]_{jj}}} > t_{n-p,\alpha/2}$.

$$T = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{[(X^T X)^{-1}]_{jj}}} \sim t_{n-p}(\phi = \quad) \qquad \gamma(\beta_j) = P(T > t_{n-p,\alpha/2}),$$

when $T \sim t_{n-p}(\phi = \frac{}{\sigma_{\beta_j}})$

$$F = \frac{(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})^T [\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^- \mathbf{K}]^{-1}(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})/s}{\mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{y}/(n-r)} > F_{s,n-r,\alpha},$$

$A = U_r \Sigma_r V_r^T$, Σ_r invertible

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix}^T$$

$G = V_r \Sigma_r^{-1} U_r^T$

$$\begin{aligned} AGA &= U_r \Sigma_r \underline{V_r^T (V_r} \Sigma_r^{-1} U_r^T) U_r \Sigma_r V_r^T \\ &= U_r \Sigma_r \Sigma_r^{-1} \Sigma_r V_r^T \\ &= U_r \Sigma_r V_r^T \end{aligned}$$

$GAG = V_r \Sigma_r^{-1} U_r^T U_r \Sigma_r V_r^T V_r \Sigma_r^{-1} U_r^T$

$= V_r \Sigma_r^{-1} U_r^T$

$= G$

1. UMVUE in the linear model under Normality
2. The general linear hypothesis, ad-hoc test
3. Likehood ratio test of the general linear hypothesis
4. Multiple testing in linear models

Likelihood ratio and LRT

Suppose $\theta \in \Theta$ is a parameter with likelihood $\mathcal{L}(\theta; \mathbf{y})$ for some data $\mathbf{y}$. Consider

$$\overset{\text{null space}}{H_0\colon \theta \in \Theta_0} \quad \text{vs} \quad H_1\colon \theta \in \overbrace{\Theta \setminus \Theta_0}^{\text{alternate space.}}. = \Theta_1$$

The *likelihood ratio* is defined as $\mathrm{LR}(\mathbf{y}; \Theta_0) = \dfrac{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta; \mathbf{y})}{\sup_{\theta \in \Theta} \mathcal{L}(\theta; \mathbf{y})}$.

The *likelihood ratio test* is the test which rejects H_0 when $\mathrm{LR}(\mathbf{y}; \Theta_0) < c$.

Choose c to give the test size α or find an equivalent test easier to calibrate.

Assume $\underset{\sim}{y} = X\underset{\sim}{b} + \underset{\sim}{e}$, $\underset{\sim}{e} \sim \text{Norm}(\underset{\sim}{0}, \sigma^2 I_n)$

OR $\underset{\sim}{y} \sim \text{Normal}(X\underset{\sim}{b}, \sigma^2 I_n)$.

Find LRT of

$$H_0: K^T \underset{\sim}{b} = \underset{\sim}{m} \quad \text{vs} \quad H_1: K^T\underset{\sim}{b} \neq \underset{\sim}{m}.$$

Likelihood:

$$L(\underset{\sim}{b}, \sigma^2; \underset{\sim}{y}) = (2\pi)^{-n/2} |\sigma^2 I_n|^{-1/2} \exp\left[-\frac{1}{2}(\underset{\sim}{y} - X\underset{\sim}{b})^T(\sigma^2 I_n)^{-1}(\underset{\sim}{y} - X\underset{\sim}{b})\right]$$

$$= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2}\frac{\|\underset{\sim}{y} - X\underset{\sim}{b}\|^2}{\sigma^2}\right]$$

$$LR(\underset{\sim}{y}; H_0) = \frac{\sup_{\{(\underset{\sim}{b},\sigma^2):\, K^T\underset{\sim}{b} = \underset{\sim}{m},\, \sigma^2 > 0\}} L(\underset{\sim}{b}, \sigma^2; \underset{\sim}{y})}{\sup_{\{(\underset{\sim}{b},\sigma^2):\, \underset{\sim}{b} \in \mathbb{R}^p,\, \sigma^2>0\}} L(\underset{\sim}{b}, \sigma^2; \underset{\sim}{y})} = \frac{L(\hat{\underset{\sim}{b}}_{null}, \hat{\sigma}^2_{null}; \underset{\sim}{y})}{L(\hat{\underset{\sim}{b}}_{mle}, \hat{\sigma}^2_{mle}; \underset{\sim}{y})},$$

where

$$(\hat{\underset{\sim}{b}}_{null}, \hat{\sigma}^2_{null}) = \underset{\{(\underset{\sim}{b},\sigma^2):\, K^T\underset{\sim}{b} = \underset{\sim}{m},\, \sigma^2>0\}}{\arg\max} L(\underset{\sim}{b}, \sigma^2; \underset{\sim}{y}).$$

From before, $\hat{\underset{\sim}{b}}_{mle} = \hat{\underset{\sim}{b}}$ (Least squares), $\hat{\sigma}^2_{mle} = \dfrac{\|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2}{n}$

Need to find $(\hat{\underset{\sim}{b}}_{null}, \hat{\sigma}^2_{null})$ with Lagrange.

Log-likelihood is

$$\ell(b, \sigma^2; y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\|y - Xb\|^2$$

We see that to get $\hat{b}_{null}$, we can focus on $\|y - Xb\|^2$.

$$\hat{b}_{null} = \underset{b:\, K^T b = m}{\operatorname{argmin}} \ \frac{1}{2}\|y - Xb\|^2 \qquad \hat{\sigma}^2_{null} = \frac{1}{n}\|y - X\hat{b}_{null}\|^2.$$

$$\mathcal{L}(b, \nu; y) = \frac{1}{2}\|y - Xb\|^2 + \nu^T(K^T b - m).$$

$$\frac{\partial}{\partial b}\mathcal{L}(\cdot) = \boxed{-X^T(y - Xb) + K\nu \overset{set}{=} 0}$$

$$\frac{\partial}{\partial \nu}\mathcal{L}(\cdot) = \boxed{K^T b - m \overset{set}{=} 0}$$

$$X^T X b + K\nu = X^T y$$

$$K^T b = m$$

$$\implies \begin{bmatrix} X^T X & K \\ K^T & 0 \end{bmatrix} \begin{bmatrix} b \\ \nu \end{bmatrix} = \begin{bmatrix} X^T y \\ m \end{bmatrix} \quad \text{RNEs.}$$

Let $(\hat{b}_{null}, \hat{\nu})$ solve the RNEs. I.e. let

$$\begin{bmatrix} X^TX & K \\ K^T & 0 \end{bmatrix} \begin{bmatrix} \hat{b}_{null} \\ \hat{\underset{\sim}{v}} \end{bmatrix} = \begin{bmatrix} X^T\underset{\sim}{y} \\ \underset{\sim}{m} \end{bmatrix}.$$

Then we can show that

$$\|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}_{null}\|^2 - \|y - X\hat{\underset{\sim}{b}}\|^2 = (K^T\hat{b} - m)^T [K^T(X^TX)^{-}K]^{-1}(K^T\hat{\underset{\sim}{b}} - \underset{\sim}{m}).$$

Thm 6.1 of Monahan.

LRT rejects when

$$\frac{h(\hat{b}_{null}, \hat{\sigma}^2_{null}; \underset{\sim}{y})}{h(\hat{b}_{mle}, \hat{\sigma}^2_{mle}; \underset{\sim}{y})} < c$$

$\iff$

$$\frac{\cancel{(2\pi)^{-n/2}} (\hat{\sigma}^2_{null})^{-n/2} \exp\left[-\frac{1}{2}\frac{\|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}_{null}\|^2}{\hat{\sigma}^2_{null}}\right]}{\cancel{(2\pi)^{-n/2}} (\hat{\sigma}^2_{mle})^{-n/2} \exp\left[-\frac{1}{2}\frac{\|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2}{\hat{\sigma}^2_{mle}}\right]} < c$$

$$\iff \left(\frac{\hat{\sigma}^2_{null}}{\hat{\sigma}^2_{mle}}\right)^{-n/2} \exp\left[-\frac{1}{2}\frac{\|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}_{null}\|^2}{\hat{\sigma}^2_{null}} + \frac{1}{2}\frac{\|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2}{\hat{\sigma}^2_{mle}}\right] < c$$

$\hat{\sigma}^2_{null} = \frac{1}{n}\|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}_{null}\|^2$

$\frac{1}{n}\|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2$

$$\iff \left(\frac{\hat{\sigma}^2_{null}}{\hat{\sigma}^2_{mle}}\right)^{-n/2} < c$$

$$\iff \frac{\hat{\sigma}^2_{null}}{\hat{\sigma}^2_{mle}} > c^{-2/n}$$

$$\iff \frac{\frac{1}{n}\| \underset{\sim}{y} - X\hat{b}_{null}\|^2}{\frac{1}{n}\| \underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2} > c^{-2/n}$$

$$\iff \frac{\| \underset{\sim}{y} - X\hat{\underset{\sim}{b}}_{null}\|^2}{\| \underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2} > c^{-2/n}$$

$$\iff \frac{\| \underset{\sim}{y} - X\hat{\underset{\sim}{b}}_{null}\|^2 - \| \underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2}{\| \underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2} > c^{-2/n} - 1$$

$$\iff \frac{\left(\| \underset{\sim}{y} - X\hat{\underset{\sim}{b}}_{null}\|^2 - \| \underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2\right)/s}{\| \underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2/(n-r)} > \underbrace{\left(c^{-2/n} - 1\right)\left(\frac{n-r}{s}\right)}_{c^*}$$

some work $\downarrow$

$$\iff \frac{\left(K^T\hat{\underset{\sim}{b}} - \underset{\sim}{m}\right)^T \left[K^T(X^TX)^- K\right]^{-1}\left(K^T\hat{\underset{\sim}{b}} - \underset{\sim}{m}\right)/s}{\underset{\sim}{y}^T(I - P_X)\underset{\sim}{y}/(n-r)} > c^* = F_{s,n-r,\alpha}$$

Result (LRT is equivalent to ad-hoc test of general linear hyp)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ *and* H_0: $\mathbf{K}^T\mathbf{b} = \mathbf{m}$ *be testable. Then the rule*

$$\text{Reject } H_0 \text{ if} \quad F := \frac{(\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}_{\text{null}}\|^2 - \|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\|^2)/s}{\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\|^2/(n-r)} > F_{s,n-r,\alpha}$$

is the size-α *LRT, where* $\hat{\mathbf{b}}_{\text{null}}$ *minimizes* $\|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2$ *subject to* $\mathbf{K}^T\mathbf{b} = \mathbf{m}$. *Also*

$$F = \frac{(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})^T[\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{K}]^{-1}(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})/s}{\mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{y}/(n-r)},$$

which has distribution $F_{s,n-r}\left(\phi = \frac{1}{\sigma^2}(\mathbf{K}^T\mathbf{b} - \mathbf{m})^T[\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{K}]^{-1}(\mathbf{K}^T\mathbf{b} - \mathbf{m})\right)$.

See Thm 6.1 of Monahan (2008).

Prove the result.

Full- reduced-model F test

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, and let $\mathbf{X} = [\mathbf{X}_0 \; \mathbf{X}_1]$, where $\mathbf{X}_0$ has rank r_0 and $\mathbf{X}$ has rank r. Then we can test

$$H_0\text{: “the columns in } \mathbf{X}_1 \text{ make no contribution to } \mathbf{y}\text{”}$$

at significance level α with the test

$$\left(\|\underset{\sim}{y} - X\hat{b}_{\text{full}}\|^2 - \|\underset{\sim}{y} - X\hat{b}_0\|^2\right)$$

$$\frac{(\text{SSE}_{\text{Red}} - \text{SSE}_{\text{Full}})/(r - r_0)}{\text{SSE}_{\text{Full}}/(n-r)} > F_{r_1, n-r, \alpha},$$

where $\text{SSE}_{\text{Red}} = \mathbf{y}^T(\mathbf{I} - \mathbf{P}_{\mathbf{X}_0})\mathbf{y}$ and $\text{SSE}_{\text{Full}} = \mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})\mathbf{y}$.

Why is this formulated in a funny way? We don't know if $\mathbf{b}_1$ is estimable.

Exercise: To justify the above test, reparameterize as $\mathbf{y} = \mathbf{W}\mathbf{d} + \mathbf{e}$:

- Let $\mathbf{W} = [\mathbf{W}_0 \; \mathbf{W}_1]$ be full-rank with $\text{Col}\,\mathbf{W}_0 = \text{Col}\,\mathbf{X}_0$ and $\text{Col}\,\mathbf{W}_1 = \text{Col}\,\mathbf{X}_1$.
- Show that LRT for H_0: $\mathbf{d}_1 = \mathbf{0}$ is equivalent to the full- reduced model F test.

$$\underset{\sim}{y} = \begin{bmatrix} W_0 & W_1 \end{bmatrix} \begin{bmatrix} \underset{\sim}{d_0} \\ \underset{\sim}{d_1} \end{bmatrix} + \underset{\sim}{e} = \begin{bmatrix} X_0 & X_1 \end{bmatrix} \begin{bmatrix} \underset{\sim}{b_0} \\ \underset{\sim}{b_1} \end{bmatrix} + \underset{\sim}{e}$$

estimable

maybe not estimable

Test $H_0: \underset{\sim}{d_1} = 0$

↳ LRT

$$\frac{\left(\underset{\sim}{y}^T (I - \overset{P_{X_0}}{P_{W_0}}) \underset{\sim}{y} - \underset{\sim}{y}^T (I - \overset{P_X}{P_W}) \underset{\sim}{y} \right) / \overset{\text{rank } X_1}{\text{rank } W_1}}{\underset{\sim}{y}^T (I - \underset{P_X}{P_W}) \underset{\sim}{y} / (n - \underbrace{\text{rank } W}_{\text{rank } X})}$$

$$(I - P_X)\underset{\sim}{y} = \begin{bmatrix} y_{11} - \bar{y}_{1\cdot} \\ \vdots \\ y_{1n} - \bar{y}_{1\cdot} \\ \vdots \\ y_{a1} - \bar{y}_{a\cdot} \\ \vdots \\ y_{an} - \bar{y}_{a\cdot} \end{bmatrix}$$

$$P_X \underset{\sim}{y} = \begin{bmatrix} \underset{\sim}{1}_n \bar{y}_{1\cdot} \\ \vdots \\ \underset{\sim}{1}_n \bar{y}_{a\cdot} \end{bmatrix} \qquad P_{X_0} \underset{\sim}{y} = \begin{bmatrix} \underset{\sim}{1}_n \bar{y}_{\cdot\cdot} \\ \vdots \\ \underset{\sim}{1}_n \bar{y}_{\cdot\cdot} \end{bmatrix}$$

$$(P_X - P_{X_0})\underset{\sim}{y} = \begin{bmatrix} \underset{\sim}{1}_n (\bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot}) \\ \vdots \\ \underset{\sim}{1}_n (\bar{y}_{a\cdot} - \bar{y}_{\cdot\cdot}) \end{bmatrix}$$

$$F = \frac{\overbrace{\left[\underset{\sim}{y}^T (I - P_{X_0}) \underset{\sim}{y} - \underset{\sim}{y}^T (I - P_X) \underset{\sim}{y} \right]}^{\underset{\sim}{y}^T (P_X - P_{X_0}) \underset{\sim}{y}} / (a-1)}{\underset{\sim}{y}^T (I - P_X) \underset{\sim}{y} / (na - a)} = \frac{n \sum_{i=1}^{a} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 / (a-1)}{\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot})^2 / (na-1)}$$

$$X = \begin{bmatrix} \underset{\sim}{1}_n & \underset{\sim}{1}_n & & \\ \vdots & & \ddots & \\ \underset{\sim}{1}_n & & & \underset{\sim}{1}_n \end{bmatrix} \qquad \underset{\sim}{b} = \begin{bmatrix} \mu \\ d_1 \\ \vdots \\ d_a \end{bmatrix}$$

X_0 $\quad$ X_1

$H_0: \boxed{K^T \qquad\qquad}$

Exercise: Consider the treatment effect model

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2), \quad j = 1, \ldots, n, \; i = 1, \ldots, a.$$

One can show that the size α LRT for whether there is any treatment effect is:

$$\frac{n \sum_{i=1}^{a} (\bar{y}_{1.} - \bar{y}_{..})^2 / (a-1)}{\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2 / (na - a)} > F_{a-1, n-a, \alpha}.$$

1. Give the noncentrality parameter as $\phi = n \cdot \text{SNR}$.
2. Let $\boxed{a = 4}$, $\boxed{\sigma = 1/2}$, and suppose $\boxed{\sum_{i=1}^{a} (\alpha_i - \bar{\alpha})^2 = 1}$. Find n such that the power of the F-test is at least 0.90 when using $\alpha = 0.05$.

$$\text{SNR} = \frac{1}{(1/2)^2} = 4$$

Find $\underline{\underline{ncp}}$:

$$F = \frac{\frac{1}{\sigma^2}\overbrace{\left[\underset{\sim}{y}^T(I-P_{X_0})\underset{\sim}{y} - \underset{\sim}{y}^T(I-P_X)\underset{\sim}{y}\right]}^{\underset{\sim}{y}^T(P_X-P_{X_0})\underset{\sim}{y}} / (a-1)}{\frac{1}{\sigma^2}\underset{\sim}{y}^T(I-P_X)\underset{\sim}{y} / (na-a)}$$

$$\frac{\underset{\sim}{y}^T(P_X-P_{X_0})\underset{\sim}{y}}{\sigma^2} \sim \chi^2_{a-1}\left(\phi = \frac{(\mathbb{E}\underset{\sim}{y})^T(P_X-P_{X_0})\mathbb{E}\underset{\sim}{y}}{\sigma^2}\right),$$

where

$$\frac{(\mathbb{E}\underset{\sim}{y})^T(P_X-P_{X_0})\mathbb{E}\underset{\sim}{y}}{\sigma^2} = \frac{1}{\sigma^2}(X\underset{\sim}{b})^T(P_X-P_{X_0})X\underset{\sim}{b}$$

$$= \frac{1}{\sigma}\left\|(P_X-P_{X_0})X\underset{\sim}{b}\right\|^2.$$

We have

$(P_X-P_{X_0})X\underset{\sim}{b}$:

$$P_X X\underset{\sim}{b} = Xb = \begin{bmatrix} \underset{\sim}{1}_n(\mu+\alpha_1) \\ \vdots \\ \underset{\sim}{1}_n(\mu+\alpha_a) \end{bmatrix}.$$

$$P_{X_0}X\underset{\sim}{b} = \underset{\sim}{1}_{na}\frac{n\sum_{i=1}^{a}(\mu+\alpha_i)}{na} = \underset{\sim}{1}_{na}\Big(\mu + \underbrace{\frac{1}{a}\sum_{i=1}^{a}\alpha_i}_{\bar{\alpha}}\Big) = \begin{bmatrix} \underset{\sim}{1}_n(\mu+\bar{\alpha}) \\ \vdots \\ \underset{\sim}{1}_n(\mu+\bar{\alpha}) \end{bmatrix}$$

$$(P_X-P_{X_0})X\underset{\sim}{b} = X\underset{\sim}{b} - P_{X_0}X\underset{\sim}{b} = \begin{bmatrix} \underset{\sim}{1}_n(\alpha_1-\bar{\alpha}) \\ \vdots \\ \underset{\sim}{1}_n(\alpha_a-\bar{\alpha}) \end{bmatrix}$$

$$\| (P_X - P_{X_0}) X b \|^2 = n \sum_{i=1}^{a} (\alpha_i - \bar{\alpha})^2 .$$

$$\frac{y^T (P_X - P_{X_0}) y}{\sigma^2} \sim \chi^2_{a-1} \left(\phi = \frac{n \sum_{i=1}^{a} (\alpha_i - \bar{\alpha})^2}{\sigma^2} \right)$$

$\Rightarrow$

$$\frac{n \sum_{i=1}^{a} (\bar{y}_{1.} - \bar{y}_{..})^2 / (a-1)}{\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2 / (na - a)} \sim F_{a-1,\, na-a} \left(\phi = \frac{n \sum_{i=1}^{a} (\alpha_i - \bar{\alpha})^2}{\sigma^2} \right).$$

$$\phi = n \cdot SNR$$

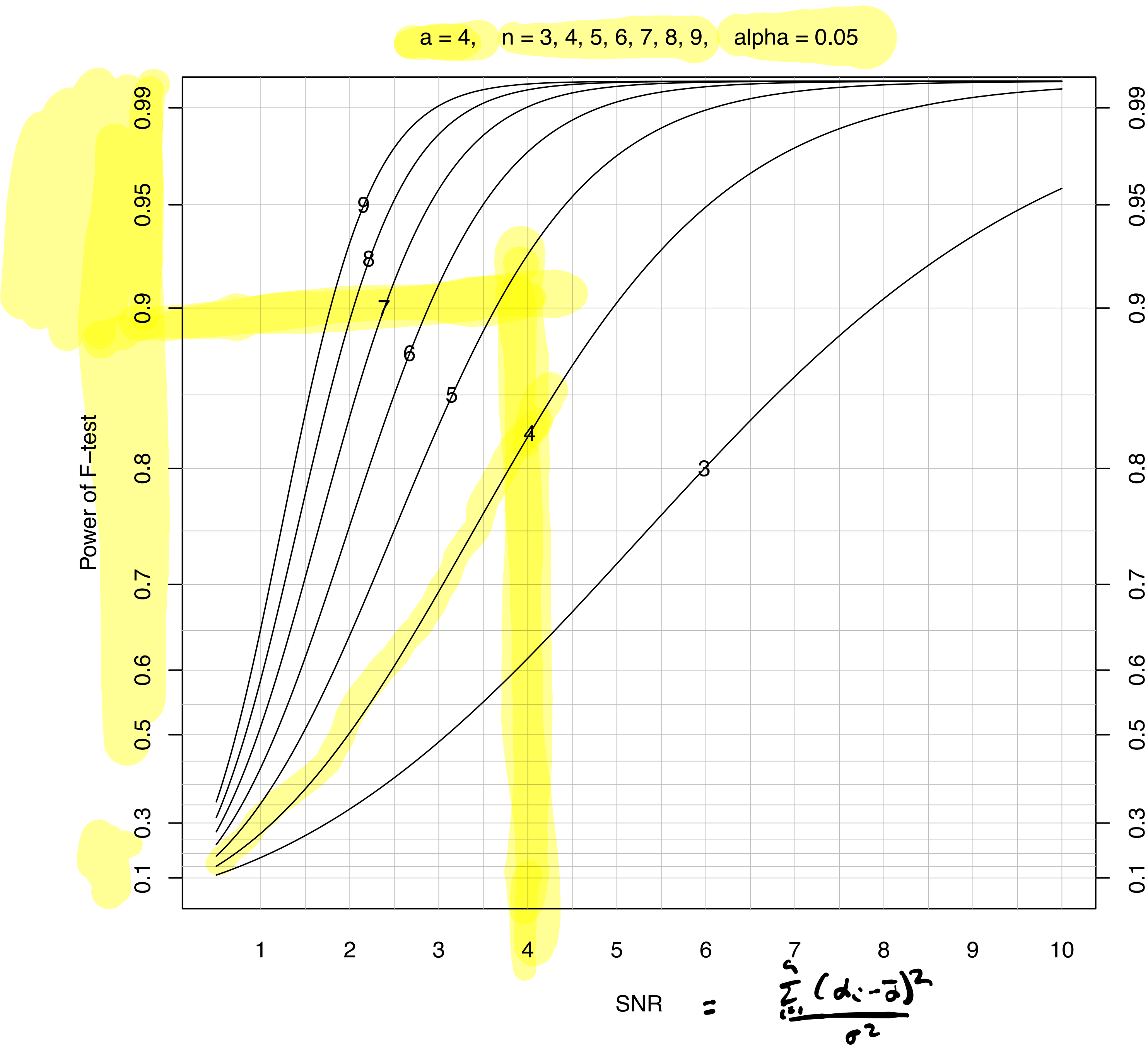
$$SNR = \frac{\sum_{i=1}^{a} (\alpha_i - \bar{\alpha})^2}{\sigma^2}$$

Mike power function:

$$\gamma(SNR) = P_{SNR}(\text{Reject } H_0)$$

$$= P\left(F > F_{a-1,\, na-a,\, \alpha}\right), \quad F \sim F_{a-1,\, na-a}(\phi = n \cdot SNR)$$

```
alpha <- 0.05
a <- 4
nn <- 3:9
snr <- seq(1/2,10,length = 500)
powF <- matrix(NA,length(nn),length(snr))

for(i in 1:length(nn)) powF[i,] <- 1-pf(qf(1-alpha,a-1,a*(nn[i]-1)),a-1,a*(nn[i]-1),nn[i]*snr)

plot(NA, xlim = range(snr), ylim = exp(exp(c(.1,.99))),
     yaxt = "n", xaxt = "n", ylab = "Power of F-test",xlab = "SNR")

at <- c(.1,.3,.5,.6,.7,.8,.9,.95,.99)
axis(side = 2, at = exp(exp(at)), labels = at)
axis(side = 4, at = exp(exp(at)), labels = at)
abline(h = exp(exp(c(seq(.1,.95, by = .05),.99))), lwd = .5, col = "gray")

axis(side = 1, at = 1:10, tick = FALSE)
abline(v = seq(1,10, by = 0.5), lwd = .5, col = "gray")

pow_at <- seq(.8,.95,length = length(nn))
for(i in 1:length(nn)){
  lines(exp(exp(powF[i,])) ~ snr)
  snr_pow <- sum(exp(exp(powF[i,])) < exp(exp(pow_at[i])))
  text(x = snr[snr_pow], y = exp(exp(pow_at[i])), label = nn[i])
}

mtext(side = 3, text = paste("a = ",a,",    n = ",paste(nn,collapse=", "),
                             ",    alpha = ",alpha,sep = ""), line = 1)
```


a = 4, n = 3, 4, 5, 6, 7, 8, 9, alpha = 0.05
Power of F–test
0.1
0.3
0.5
0.6
0.7
0.8
0.9
0.95
0.99
1
2
3
4
5
6
7
8
9
10
SNR
3
4
5
6
7
8
9
SNR $= \dfrac{\sum_{i=1}^{a}(\alpha_i-\bar{\alpha})^2}{\sigma^2}$

1.

STAT 714 fa 2023 Quiz 4

Let $Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijkl}$, $\varepsilon_{ijkl} \sim \text{Normal}(0, \sigma^2)$, for $i, j, k \in \{1, 2\}$ and $l = 1, \dots, n$, where the Y_{ijkl} are responses and the α_i, β_j, and γ_k represent treatment effects from factors A, B, and C, respectively. The model can be written in matrix form $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ with

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_n & \mathbf{1}_n & & \mathbf{1}_n & & \mathbf{1}_n & \\ \mathbf{1}_n & \mathbf{1}_n & & \mathbf{1}_n & & & \cancel{\mathbf{1}_n} \\ \mathbf{1}_n & \mathbf{1}_n & & & \cancel{\mathbf{1}_n} & \mathbf{1}_n & \\ \mathbf{1}_n & \mathbf{1}_n & & & \cancel{\mathbf{1}_n} & & \cancel{\mathbf{1}_n} \\ \mathbf{1}_n & & \cancel{\mathbf{1}_n} & \mathbf{1}_n & & \mathbf{1}_n & \\ \mathbf{1}_n & & \cancel{\mathbf{1}_n} & \mathbf{1}_n & & & \cancel{\mathbf{1}_n} \\ \mathbf{1}_n & & \cancel{\mathbf{1}_n} & & \cancel{\mathbf{1}_n} & \mathbf{1}_n & \\ \mathbf{1}_n & & \cancel{\mathbf{1}_n} & & \cancel{\mathbf{1}_n} & & \cancel{\mathbf{1}_n} \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \gamma_1 \\ \gamma_2 \end{bmatrix}.$$

~~1.~~ (a) (i) Give the matrix $\mathbf{K}^T$ and the vector $\mathbf{m}$ such that H_0: $\mathbf{K}^T\mathbf{b} = \mathbf{m}$ is a testable hypothesis expressing that factor A has no effect on the response mean.

$K^T = [\,0 \;\; 1 \;\; -1 \;\; 0 \;\; 0 \;\; 0 \;\; 0\,]$ $\qquad \underset{\sim}{m} = 0.$

This gives H_0: $\alpha_1 - \alpha_2 = 0.$

To see that $K^T\underset{\sim}{b}$ is estimable, add first $4n$ rows of X, then subtract last $4n$ rows.

(b) (ii) Give the numerator and denominator degrees of freedom of the relevant F-distribution.

num: 1

den: $N - \text{rank}\, X = 8n - 4$

~~2.~~ (b) Suppose you perform a full-reduced model F test to check whether *any* of the factors have an effect on the response mean.

(a) Give the test statistic in terms of some projection matrices.

Let X_0 be the first column of X. Then

$$F = \frac{[\underset{\sim}{y}^T(I - P_{X_0})\underset{\sim}{y} - \underset{\sim}{y}^T(I - P_X)\underset{\sim}{y}]/3}{\underset{\sim}{y}^T(I - P_X)\underset{\sim}{y}/(8n-4)}$$

$\underset{\sim}{y}^T\underbrace{(P_X - P_{X_0})}_{\text{rank } 3}\underset{\sim}{y}$

3 is the rank of X_1 when we write $X = [X_0 \; X_1]$.

(b) Give the numerator and denominator degrees of freedom of the relevant F distribution.

Can also see $\text{rank}(P_X - P_{X_0}) = 3$

num: 3

den: $8n - 4$

1. UMVUE in the linear model under Normality

2. The general linear hypothesis, ad-hoc test

3. Likehood ratio test of the general linear hypothesis

End exam II material

4. Multiple testing in linear models

$Y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$

Simultaneous confidence interval setup

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2\mathbf{I}_n)$ with rank $\mathbf{X} = r$ and let $\mathbf{c}_1^T\mathbf{b}, \dots, \mathbf{c}_K^T\mathbf{b}$ be estimable contrasts. We seek L_k and U_k, $k = 1, \dots, K$ such that

$$P\left(\bigcap_{k=1}^{K} \boxed{\{L_k \leq \mathbf{c}_k^T\mathbf{b} \leq U_k\}}\right) \geq 1 - \alpha.$$

The intersection event above is called *simultaneous coverage*.

Note $A_k = \{ L_k \leq \underset{\sim}{c}_k^T \underset{\sim}{b} \leq U_k \}$, my CI for $\underset{\sim}{c}_k^T \underset{\sim}{b}$ captures its target.

$$P\left(\bigcap_{k=1}^{K} A_k\right) = 1 - P\left(\left(\bigcap_{k=1}^{K} A_k\right)^c\right)$$

$$= 1 - P\left(\bigcup_{k=1}^{K} A_k^c\right)$$

Suppose $P(A_k) = 1 - \alpha^*$

$$\geq 1 - \sum_{k=1}^{K} P(A_k^c)$$

$$= 1 - K\alpha^*$$

strategy for achieving simultaneous cov. prob. $\geq 1-\alpha$ is to chose $\alpha^* = \dfrac{\alpha}{K}$.

Result (Simul. coverage guarantee for Bonferroni-adj. t- intervals)

Under the simultaneous confidence interval setup, the intervals

$$[L_k, U_k] = \left[\mathbf{c}_k^T\hat{\mathbf{b}} \pm t_{n-r,\alpha/(2K)}\hat{\sigma}\sqrt{\mathbf{c}_k^T(\mathbf{X}^T\mathbf{X})^-\mathbf{c}_k}\right], \quad k = 1, \ldots, K$$

ensure simultaneous coverage with probability at least $1 - \alpha$.

Prove the result.

Begin with

$$\underset{\sim}{c}_k^T \hat{\underset{\sim}{b}} \sim N\left(\underset{\sim}{c}_k^T \underset{\sim}{b},\ \sigma^2\, \underset{\sim}{c}_k^T (X^TX)^- \underset{\sim}{c}_k\right)$$

$\underset{\sim}{c}_k = X^T \underset{\sim}{a}_k$

$\hat{\underset{\sim}{b}}$ satisfies $(X^TX)\hat{\underset{\sim}{b}} = X^T\underset{\sim}{y} \Rightarrow X\hat{\underset{\sim}{b}} = P_X \underset{\sim}{y}$

$$\operatorname{Var}(c_r^T \hat{b}) = \operatorname{Var}(a_r^T X \hat{b})$$

$$= \operatorname{Var}(a_r^T P_X y)$$

$$= a_r^T P_X (\operatorname{Cov} y) P_X^T a_r$$

$$= a_r^T \underbrace{X(X^TX)^- X^T [\sigma^2 I] X}_{X} (X^TX)^- X^T a_r$$

$$= \sigma^2 a_r^T X (X^TX)^- X^T a_r$$

$$= \sigma^2 c_r^T (X^TX)^- c_r$$

$$\Rightarrow \quad \frac{c_r^T \hat{b} - c_r^T b}{\sqrt{\sigma^2 c_r^T (X^TX)^- c_r}} \sim N(0,1)$$

Replace σ^2 with $\hat{\sigma}^2$:

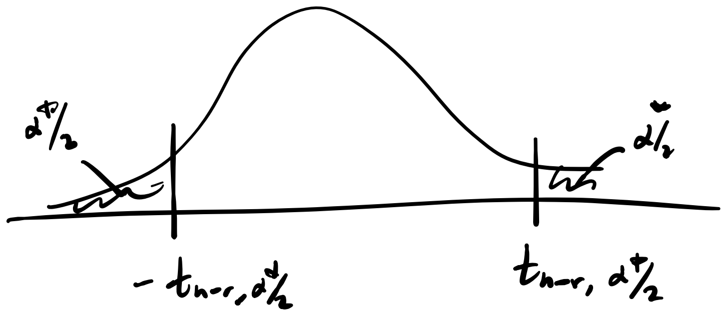
$$\frac{c_r^T \hat{b} - c_r^T b}{\sqrt{\hat{\sigma}^2 c_r^T (X^TX)^- c_r}} = \frac{\dfrac{c_r^T \hat{b} - c_r^T b}{\sqrt{\sigma^2 c_r^T (X^TX)^- c_r}}}{\sqrt{\dfrac{(n-r)\hat{\sigma}^2}{\sigma^2} \Big/ (n-r)}} = \frac{Z}{\sqrt{\chi^2_{n-r} / (n-r)}}$$

$r = \operatorname{rank} X$

$$\sim t_{n-r}.$$

$$\hat{\sigma}^2 = y^T (I - P_X) y \,/\, (n-r)$$

$$P\left(-t_{n-r,\alpha^*/2} \le \frac{\underset{\sim}{c}_r^T \hat{\underset{\sim}{b}} - \underset{\sim}{c}_r^T \underset{\sim}{b}}{\sqrt{\hat\sigma^2\, \underset{\sim}{c}_r^T (X^TX)^- \underset{\sim}{c}_r}} \le t_{n-r,\alpha^*/2}\right) = 1-\alpha^*$$



$$P\Big(\underset{\sim}{c}_r^T \hat{\underset{\sim}{b}} - t_{n-r,\alpha^*/2}\, \hat\sigma \sqrt{\underset{\sim}{c}_r^T (X^TX)^- \underset{\sim}{c}_r} \le c_r^T \underline{b} \le$$

$$\underset{\sim}{c}_r^T \hat{\underset{\sim}{b}} + t_{n-r,\alpha^*/2}\, \hat\sigma \sqrt{\underset{\sim}{c}_r^T (X^TX)^- \underset{\sim}{c}_r}\Big) = 1-\alpha^*$$

A $(1-\alpha^*)\cdot 100\%$ C.I. for $c_r^T \underline{b}$ is

$$\left[\underset{\sim}{c}_r^T \hat{\underset{\sim}{b}} \pm t_{n-r,\alpha^*/2}\, \hat\sigma \sqrt{\underset{\sim}{c}_r^T (X^TX)^- \underset{\sim}{c}_r}\right]$$

For simultaneous coverage of $\underset{\sim}{c}_1^T \underset{\sim}{b}_1, \ldots, \underset{\sim}{c}_K^T \underset{\sim}{b}_K$ with prob $1-\alpha$, use

$$\left[\underset{\sim}{c}_r^T \hat{\underset{\sim}{b}} \pm t_{n-r,\frac{\alpha}{2K}}\, \hat\sigma \sqrt{\underset{\sim}{c}_r^T (X^TX)^- \underset{\sim}{c}_r}\right] \qquad r=1,\ldots,K.$$

Result (Exact simul. coverage for max-abs intervals)

Under the simultaneous confidence interval setup, the intervals

$$[L_k, U_k] = \left[\mathbf{c}_k^T\hat{\mathbf{b}} \pm |t|^{\vee}_{n-r,\alpha}\hat{\sigma}\sqrt{\mathbf{c}_k^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}_k}\right], \quad k = 1, \dots, K$$

give simultaneous coverage probability $1-\alpha$, *where* $|t|^{\vee}_{n-r}$ *is the dist. such that*

$$\max_{1 \le k \le K}\left|\frac{\mathbf{c}_k^T\hat{\mathbf{b}} - \mathbf{c}_k^T\mathbf{b}}{\hat{\sigma}\sqrt{\mathbf{c}_k^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}_k}}\right| \sim |t|^{\vee}_{n-r}.$$

The upper α quantile $|t|^{\vee}_{n-r,\alpha}$ of $|t|^{\vee}_{n-r}$ can be found by Monte Carlo simulation.

Tukey's pairwise comparisons of means in one-way ANOVA is exactly this.

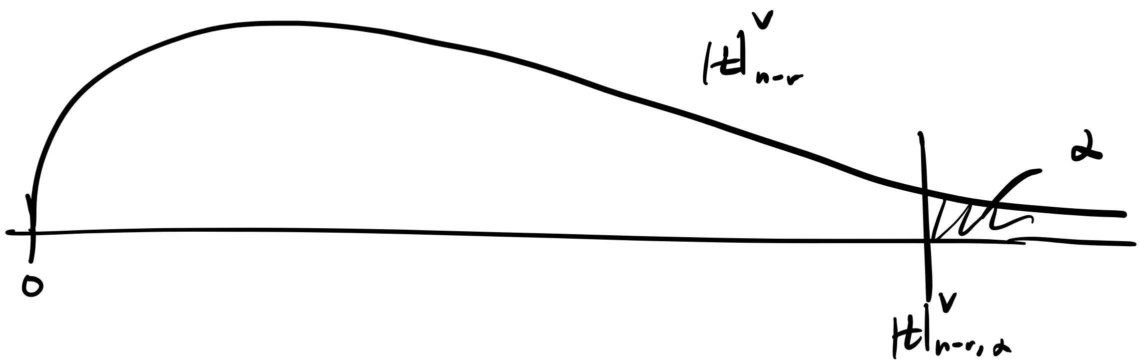
Prove the result.

$$\max_i \{a_i\} = \bigvee_{i=1}^{n} a_i \qquad a \vee b = \max\{a,b\}$$

$$\frac{\underset{\sim}{c}_k^T \hat{\underset{\sim}{b}} - \underset{\sim}{c}_k^T \underset{\sim}{b}}{\hat\sigma \sqrt{\underset{\sim}{c}_k^T (X^TX)^- \underset{\sim}{c}_k}} \sim t_{n-r}$$

"\vee"

$$\max_{1 \le k \le K} \left| \frac{\underset{\sim}{c}_k^T \hat{\underset{\sim}{b}} - \underset{\sim}{c}_k^T \underset{\sim}{b}}{\hat\sigma \sqrt{\underset{\sim}{c}_k^T (X^TX)^- \underset{\sim}{c}_k}} \right| \sim |t|^{\vee}_{n-r}$$



Then

$$P\left(\max_{1 \le k \le K} \left| \frac{\underset{\sim}{c}_k^T \hat{\underset{\sim}{b}} - \underset{\sim}{c}_k^T \underset{\sim}{b}}{\hat\sigma \sqrt{\underset{\sim}{c}_k^T (X^TX)^- \underset{\sim}{c}_k}} \right| \le |t|^{\vee}_{n-r,\alpha} \right) = 1-\alpha$$

$$= P\left(\bigcap_{1 \le k \le K} \left\{ \left| \frac{\underset{\sim}{c}_k^T \hat{\underset{\sim}{b}} - \underset{\sim}{c}_k^T \underset{\sim}{b}}{\hat\sigma \sqrt{\underset{\sim}{c}_k^T (X^TX)^- \underset{\sim}{c}_k}} \right| \le |t|^{\vee}_{n-r,\alpha} \right\} \right) = 1-\alpha$$

$$= P\left(\bigcap_{1\le k\le K}\left\{ -|t|^{\vee}_{n-r,\alpha} \le \frac{c_k^T\hat{b} - c_k^T b}{\hat\sigma\sqrt{c_k^T (X^TX)^- c_k}} \le |t|^{\vee}_{n-r,\alpha} \right\}\right) = 1-\alpha$$

$$= P\left(\bigcap_{1\le k\le K}\left\{ c_k^T\hat{b} + |t|^{\vee}_{n-r,\alpha}\,\hat\sigma\sqrt{c_k^T(X^TX)^- c_k} \le c_k^T b \le c_k^T\hat{b} + |t|^{\vee}_{n-r,\alpha}\,\hat\sigma\sqrt{c_k^T(X^TX)^- c_k} \right\}\right) = 1-\alpha$$

$$\Rightarrow P\left(\bigcap_{1\le k\le K}\left\{ c_k^T b \in \left[c_k^T\hat{b} \pm |t|^{\vee}_{n-r,\alpha}\,\hat\sigma\sqrt{c_k^T(X^TX)^- c_k} \right]\right\}\right) = 1-\alpha$$

obtain via Monte Carlo

How to get $|t|^{\vee}_{n-r,\alpha}$?

$$\max_{1\le k\le K}\left| \frac{c_k^T\hat{b} - c_k^T b}{\hat\sigma\sqrt{c_k^T(X^TX)^- c_k}} \right| = \max_{1\le k\le K}\left| \frac{\overbrace{\dfrac{c_k^T\hat{b} - c_k^T b}{\sigma\sqrt{c_k^T(X^TX)^- c_k}}}^{Z_k}}{\sqrt{\underbrace{(n-r)\dfrac{\hat\sigma^2}{\sigma^2}}\,/(n-r)}} \right|$$

$$= \max_{1 \le r \le k} \left| \frac{z_r}{\sqrt{\chi^2_{n-r}/(n-r)}} \right|$$

$z_1, \ldots, z_K$

Monte Carlo simulation for $|t|^{\vee}_{n-r}$

1. Note
$$\frac{\mathbf{c}_k^T\hat{\mathbf{b}} - \mathbf{c}_k^T\mathbf{b}}{\hat{\sigma}\sqrt{\mathbf{c}_k^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}_k}} = \frac{Z_k}{\sqrt{W/(n-r)}} \sim t_{n-r}, \quad k = 1,\dots,K,$$
where $W = \frac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-r}$ and $(Z_1,\dots,Z_K)^T = \frac{1}{\sigma}\mathbf{D}(\mathbf{C}^T\hat{\mathbf{b}} - \mathbf{C}^T\mathbf{b})$ with
$$\mathbf{C} = [\mathbf{c}_1 \;\dots\; \mathbf{c}_K] \quad \text{and} \quad \mathbf{D}^{-2} = \operatorname{diag}\left(\mathbf{c}_1^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}_1,\dots,\mathbf{c}_K^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}_K\right).$$
2. Since $\mathbf{C}^T\hat{\mathbf{b}} - \mathbf{C}^T\mathbf{b} \sim \text{Normal}(0, \sigma^2\mathbf{C}^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{C})$, we have
$$(Z_1,\dots,Z_K)^T \sim \text{Normal}(\mathbf{0}, \mathbf{D}\mathbf{C}^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{C}\mathbf{D}).$$
3. To obtain an MC approx. to $|t|^{\vee}_{n-r,\alpha}$, we generate many realizations of
$$\max\{|Z_1|,\dots,|Z_K|\}/\sqrt{W/(n-r)}, \quad W \text{ independent of } (Z_1,\dots,Z_K)^T,$$
and take the upper α quantile.

The covariance matrix $\mathbf{D}\mathbf{C}^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{C}\mathbf{D}$ may not be positive definite dwbh.

Monahan, J. F. (2008). *A primer on linear models*. CRC Press.