REVIEW TOPICS

- 3 • Orthogonal columns 11
- 4 • Random errors vs. residuals 11
- 2 • ~~"Linear parametric test" vs. F-test~~ 12
- • Prediction vs. confidence 2
- 5? • CI for combination of parameters 10
- 6? • Gauss-Markov theorem 8
- 1 • ~~Interpreting CI → last Q. practice midterm~~ $\beta_2 - \beta_3$ 12

- Effect of categorical vs. continuous 3
-

3. An experiment was conducted to explore the relationship between the *lifetime* (measured in days) and sexual activity of fruitflys.

   125 fruit flys were divided randomly into 5 treatment groups, each of 25 flys. Each treatment was designed to simulate a different level of sexual activity, with levels: *none*, *one*, *many*, *low* and *high*.

   The *thorax length* of each male was also measured as this was known to affect lifetime.

   One observation in the *many* group was lost.

   The following model was fit:

$$\text{Lifetime}_i = \beta_0 + \beta_1 \text{Thorax Length}_i + \beta_2 \text{one}_i + \beta_3 \text{low}_i + \beta_4 \text{many}_i + \beta_5 \text{high}_i + \epsilon_i$$

   where *one*, *low*, *many*, and *high* are indicator variables for the respective treatment groups.

   Results from the least squares fit are given below.

$$\hat{\beta} = \begin{pmatrix} -48.8 \\ 134.3 \\ 2.6 \\ -7.0 \\ 4.1 \\ -20.0 \end{pmatrix}, \qquad (X^T X)^{-1} = \begin{pmatrix} 1.06 & -1.22 & -0.05 & -0.04 & -0.07 & -0.08 \\ -1.22 & 1.46 & 0.02 & -0.00 & 0.03 & 0.05 \\ -0.05 & 0.02 & 0.08 & 0.04 & 0.04 & 0.04 \\ -0.04 & -0.00 & 0.04 & 0.08 & 0.04 & 0.04 \\ -0.07 & 0.03 & 0.04 & 0.04 & 0.08 & 0.04 \\ -0.08 & 0.05 & 0.04 & 0.04 & 0.04 & 0.08 \end{pmatrix}$$

$$\hat{\sigma} = 10.54,$$

   (a) Conduct a t-test of the null hypothesis that $\beta_1 = 0$. (6)

Question 3 continues on the next page...

(b) For two flys with the same thorax length, show the difference between the expected lifetime for a fly in the *one* treatment group and the expected lifetime for a fly in the *low* treatment group is $\boxed{\beta_2 - \beta_3.}$ (3)

$E(\text{lifetime} \mid \underline{\text{one}}) -$

$E(\text{" "} \mid \text{low}) = \beta_2 - \beta_3$

(c) Construct a 95% confidence interval for $\beta_2 - \beta_3$. (6)

$(3.7, 15.5)$

(d) Interpret your interval from (c) in the context of the study. (2)

on average / mean / expected → this is a confidence interval

With 95% confidence, the expected lifetime for a fly in the "one" treatment group

is between 3.7 & 15.5 days
units

longer than the expected lifetime for a fly in the "low" treatment group, holding thorax length constant.

If this was a prediction interval: the lifetime of

A 95% prediction interval for a fly in the "low" treatment group, with thorax length 2 mm is _____ to _____ days.

Two common CI interpretations:

- difference in mean between two levels of a categorical variable
- the change in mean response when a continuous variable increase by 1 unit.
- the change in slope / "relationship" "interactions"

① Interpreting a CI on combination of β

② Linear Parametric Hypotheses

Full model $y = X\beta + \varepsilon$

Two approaches to F-test where we "notations"

impose some linear restrictions on β. ← Null hypothesis

Reduced model : $y = X_\omega \beta_\omega + \varepsilon$

some smaller design matrix & β parameter vector

Linear parametric hypothesis $\underbrace{K^T \beta = m}_{\text{HW \#4}}$

Example of linear parametric hypothesis

Midterm 2016 Q3

Full model: $\text{Lifetime}_i = \beta_0 + \beta_1 \text{Thorax}_i + \beta_2 \text{one}_i + \beta_3 \text{low}_i + \beta_4 \text{many}_i + \beta_5 \text{high}_i + \varepsilon_i$

H_0: $\boxed{\beta_2 - \beta_3 = 0}$ 1 constraint $\beta_2 = \beta_3$

Reduced model: Substitute $\beta_2 = \beta_3$ in full model + simplify

Linear parametric hypothesis: $K^T \beta = m$

$=$

$$K^T \beta = m$$

$$\underbrace{\begin{pmatrix} 0 & 0 & 1 & -1 & 0 & 0 \\ & & & & & \\ & & & & & \end{pmatrix}}_{p \text{ columns}} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} 0 \\ \\ \end{pmatrix}$$

1 column

this should never have any β's in it

$H_0: \beta_2 - \beta_3 = 0$

$H_0: \beta_2 = \beta_3$ DON'T DO THIS

$\cancel{K^T = (0\ 0\ 1\ 0\ 0\ 0\ 0)}$

$\cancel{m = (0\ 0\ \beta_3\ 0\ 0\ 0)}$

$$H_0: \left.\begin{array}{l} \beta_1 - \beta_3 = 0 \\ \beta_2 - \beta_1 = 0 \end{array}\right\}$$ two contraints

$K^T_{2 \times p}$

$\beta_1 = \beta_3$

$\boxed{\beta_2 = \beta_1 = \beta_3}$

A case where K^T would have more than one row.

$\cancel{\emptyset}$ If all columns are orthogonal:

$$\underset{n\times p}{X} = \begin{pmatrix} \underset{\sim}{x}_1 & \underset{\sim}{x}_2 & \cdots & \underset{\sim}{x}_p \end{pmatrix}$$

$$X^T X = \begin{pmatrix} & & 0 \\ & & \\ 0 & & \end{pmatrix}$$

Estimates don't change if you drop a term in the model.

$$(X^T X)^{-1} = \begin{pmatrix} & & 0 \\ & & \\ 0 & & \end{pmatrix}$$

If a couple of columns are orthogonal:

$$\underset{\sim}{x}_i^T \underset{\sim}{x}_j = 0 \quad \overset{??}{\Longrightarrow} \quad \hat{\beta}_i \text{ doesn't change}$$

if we drop $\underset{\sim}{x}_j$

HW#3: after some centering true for $\underset{\sim}{x}_3$ $\underset{\sim}{x}_5$

To decide about ?? in previous page, maybe approach like:

$$X^T X = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \qquad A = \begin{pmatrix} \ddots & 0 \\ 0 & \ddots \end{pmatrix}$$

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]^{-1} = \left[\begin{array}{c|c} A^{-1} + A^{-1}\left(\qquad \right)A^{-1} & \\ \hline & \end{array}\right]$$