General Linear Test

13.1

Example

Y = Diastolic Blood Pressure

X_1 = Age

X_2 = BMI (weight in kg / height^2 in metres)
25+ is "overweight"

X_3 = Average Daily Salt intake

X_4 = Average Calories/day

X_5 = % Calories from refined sugar / Corn Syrup etc

X_6 = % of calories from fat

X_7 = Reported average daily minutes of exercise

X_8 = Oxygen consumption (treadmill test)

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_8 x_{i8} + \varepsilon_i$$

Lots of potential questions, like

Controlling for all other variables, is salt intake a risk factor for high blood pressure

$H_0: \beta_3 = 0$

- There's a t-test
- 2-sided

Could ask similar questions about all the other variables

13.2

Some questions are about single linear combinations of β values, like

Controlling for all other variables, which is worse, calories from fat or calories from refined sugar? $H_0: \beta_5 = \beta_6$, test with

$$T = \frac{a'\hat{\beta} - h}{s\sqrt{a'(X'X)^{-1}a}}$$

But some questions ask about more than one linear combination at the same time, like

- Once you allow for age, do any of the other variables predict blood pressure?

$$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$$

- Controlling for Age, BMI & diet, does exercise matter?

$$H_0: \beta_7 = \beta_8 = 0$$

All of these (& in fact all the hypotheses we will test) are examples of

$$H_0: \underset{q\times(k+1)}{C}\ \underset{(k+1)\times 1}{\beta} = \underset{q\times 1}{t}$$

13.3

Natural to ask for C & t matrices on quizzes, final exam

- Allowing for age, do the other variables matter?

$$\overset{C}{\begin{bmatrix} 0&0&1&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&1 \end{bmatrix}} \beta = \begin{pmatrix} 0\\0\\0\\0\\0\\0\\0 \end{pmatrix}$$

- Controlling for others, does exercise matter?

$$\begin{pmatrix} 0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&1 \end{pmatrix} \beta = \begin{pmatrix} 0\\0 \end{pmatrix}$$

- Which is worse, calories from fat or calories from sugar?

$$\begin{pmatrix} 0&0&0&0&0&1&-1&0&0 \end{pmatrix} \begin{pmatrix} \beta_0\\ \beta_1\\ \beta_2\\ \beta_3\\ \beta_4\\ \beta_5\\ \beta_6\\ \beta_7\\ \beta_8 \end{pmatrix} = 0$$

- Any of the t-tests
- The overall test

So a general test of $H_0: C\beta = t$ is useful

Recall if $W \sim N_p(\mu, \Sigma)$, then $(W-\mu)'\Sigma^{-1}(W-\mu) \sim \chi^2(p)$

13.4

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I_n)$$

$$\hat{\beta} = (X'X)^{-1}X'Y \sim N_{k+1}(\beta, \sigma^2(X'X)^{-1})$$

So

$$C\hat{\beta} \sim N_q(C\beta, \sigma^2 C(X'X)^{-1}C')$$

And if $H_0: C\beta = t$ is true, then $E(C\hat{\beta}) = t$ and a kind of (squared) <u>DISTANCE</u> between $C\hat{\beta}$ & t

$$(C\hat{\beta} - t)'(\sigma^2 C(X'X)^{-1}C')^{-1}(C\hat{\beta} - t) \sim \chi^2(q)$$

Independent of $SSE = \hat{\varepsilon}'\hat{\varepsilon}$, $\quad \frac{SSE}{\sigma^2} \sim \chi^2(n-k-1)$

And

$$F = \frac{(C\hat{\beta} - t)'(\sigma^2 C(X'X)^{-1}C')^{-1}(C\hat{\beta} - t)/q}{\frac{SSE}{\sigma^2}/(n-k-1)}$$

$$= \frac{\frac{1}{\sigma^2}(C\hat{\beta} - t)'(C(X'X)^{-1}C')^{-1}(C\hat{\beta} - t)}{\frac{1}{\sigma^2}\, q\, s^2}$$

$\uparrow$ $MSE = \frac{SSE}{n-k-1}$

$$\sim F(q, n-k-1)$$

Full - Reduced version next. Get notation right.