To Summarize ... we now know the multiple linear regression model

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \quad , \quad \underset{\sim}{\varepsilon} \sim \text{normal}(\underset{\sim}{0}, \sigma^2 I)$$

$\underset{\sim}{y}$ $n \times 1$ response vector

$X$ $n \times p$ covariate / design matrix

* $X_{i0} = 1$ for $i = 1, \ldots, n$

~ know how to:

- estimate $\underset{\sim}{\beta}$
- estimate $\sigma^2$
- estimate $E[\underset{\sim}{y}]$
- predict future $y_n$
- test if $\beta_k = 0$ (t-test)
- test if $\beta_1 = \ldots = \beta_{p-1} = 0$ (F-test)

] when $p=2$, these were equivalent

* we can construct a test of $\beta_k = 0$ based on an F-distributed test statistic

* we can construct tests of same $\beta_k = 0$ all at once

# Extra Sums of Squares

Given a model $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$, $\underset{\sim}{\varepsilon} \sim \text{normal}(\underset{\sim}{0}, \sigma^2 I_n)$

we can obtain $\underset{\sim}{b} = (X'X)^{-1}X'\underset{\sim}{y}$ to estimate $\underset{\sim}{\beta}$ and

corresponding residuals $\underset{\sim}{e} = \underset{\sim}{Y} - X\underset{\sim}{b}$, $\hat{\underset{\sim}{Y}} = X\underset{\sim}{b}$

SSTO: $\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, SSE: $\sum_{i=1}^{n} e_i^2$, SSR: $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$

SSTO = SSE + SSR

Depend on the covariates in the model

Suppose we had a dataset that contained a response $Y$ and three covariates $X_1, X_2, X_3$

Example: $Y_i$ body fat % for subject $i$
$X_{i1}$ tricep fold thickness of subject $i$
$X_{i2}$ thigh circumference of subject $i$
$X_{i3}$ midarm circumference of subject $i$

Define $SSE(X_1)$ to be the error sum of squares when $X_1$ is the only covariate in the model

Define $SSR(X_1)$ to be the regression sum of squares when $X_1$ is the only covariate in the model

Define $SSE(X_1, X_2)$ to be the error sum of squares when $X_1$ and $X_2$ are covariates in the model

Define $SSR(X_1, X_2)$ to be the regression sum of squares when $X_1$ and $X_2$ are covariates in the model

Define $SSR(X_2 | X_1) = SSR(X_1, X_2) - SSR(X_1)$ ] one example of extra sum of squares

marginal effect of adding $X_2$ to the regression model when $X_1$ is already in the model

$$SSTO = SSR(X_1) + SSE(X_1)$$
$$= SSR(X_1, X_2) + SSE(X_1, X_2)$$

$$SSR(X_2 | X_1) = SSR(X_1, X_2) - SSR(X_1)$$

$$\Longrightarrow \underbrace{SSTO}_{\text{total variability in } Y} = \underbrace{SSR(X_1)}_{\text{variability attributable to } X_1} + \underbrace{SSR(X_2 | X_1)}_{\text{residual variability after accounting for } X_1 \text{ that is attributable to } X_2} + \underbrace{SSE(X_1, X_2)}_{\text{remainder after } X_1, X_2 \text{ have been accounted for}}$$

$$SSTO = SSR(X_1) + SSR(X_2 | X_1) + SSR(X_3 | X_2, X_1) + SSE(X_1, X_2, X_3)$$

where $SSR(X_3 | X_1, X_2) = SSR(X_3, X_2, X_1) - SSR(X_1, X_2)$