# Review of concepts for the Normal Linear Model

- $Y_i \sim N(\mu_i, \sigma^2), i = 1, 2, \ldots, n$.
- $\mu_i = E(Y_i) = x_i^t \beta$ linear predictor so $g(\mu_i) = \mu_i$. $\beta$ is the vector of parameters
- The usual notation for the model is,

$$Y = X\beta + \epsilon$$

  $Y$ is the vector of observations, $X$ design matrix, $\epsilon$ is the vector of errors.
- MLR, ANOVA, ANCOVA models take this form.
- MLE of $\beta$ is $\hat{\beta} = (X^T X)^{-1} X^T Y$ (denote this as $b$)
- $\sigma^2$ is treated as a *nuisance parameter*.

- $\sigma^2$ can be estimated as

$$\hat{\sigma^2} = \frac{1}{(N-p)}(Y - Xb)^T(Y - Xb) = \frac{1}{(N-p)}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$$

where $\hat{Y}_i$ is the fitted value for observation $i$ or $\hat{Y}_i = x_i^T b$

- If $\hat{Y}$ is the vector of all fitted values, then $Y - \hat{Y}$ is the vector of *residuals*.
- **Deviance**

$$D = \frac{1}{\sigma^2}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 = \frac{1}{\sigma^2}(Y - Xb)^T(Y - Xb)$$

which depends on $\sigma^2$.

# Hypothesis tests on $\beta$

- $H_0 : \beta = \beta_0 = (\beta_1, \beta_2, \ldots, \beta_q)$. Model with $q - 1$ covariates.
- $H_1 : \beta = \beta_0 = (\beta_1, \beta_2, \ldots, \beta_p)$. Model with $p - 1$ covariates.
- Assume $p > q$ so model under $H_0$ is reduced compared to model for $H_1$.
- Fit model under $H_0$ ($H_1$) and compute sum of squares $S_0$ ($S_1$)
- Since model under $H_1$ has more parameters $S_0 > S_1$,
- Test $H_0$ with

$$F = \frac{(S_0 - S_1)}{(p-q)} \Big/ \frac{S_1}{(N-p)} = \frac{(D_0 - D_1)}{(p-q)} \Big/ \frac{D_1}{(N-p)} \sim F(p-q, N-p)$$

.

- Reject $H_0$ if $F_{obs} > F_{(1-\alpha)}(p-q, N-p)$

# Residuals

- Based on $\hat{e}_i = Y_i - \hat{Y}_i = Y_i - x_i^T b = obs - fitted$
- *Standardized residuals:* $r_i = \frac{\hat{e}_i}{\hat{\sigma}(1-h_{ii})^{1/2}}; i = 1, 2, \ldots, N$ where $h_i i$ is $i - th$ element on the diagonal of $H = X(X^T X)^{-1} X^T$.
- *Plots:* Normal probability plots, independence, homoscedasticity, constant variance, etc.
- *High leverage or influential:* $h_{ii} > 2(p/N)$, $DFITS_i = r_i(h_{ii}/(1 - h_{ii})^{1/2}$.
- *Cook's distance:* $D_i = (1/p)(DFITS)^2$. Large values means "influential".

$$D_i = \frac{1}{p}(b - b_{(i)})^T X^T X(b - b_{(i)})$$

and $b_{(i)}$ is the estimate of $\beta$ without observation $i$.