Multiple Regression

1. $\mathbb{E}(\mathbf{Y}|\mathbf{X}) = \mathbf{X}\beta$, $\mathrm{Var}(\mathbf{Y}|\mathbf{X}) = \sigma^2\mathbf{I}_n$
   Specially, in simple linear regression, $y_i = \beta_0 + \beta_1 x_i$,
   $\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$, $\mathbb{E}(\mathbf{Y}|\mathbf{X}) = \mathbf{1}_n\beta_0 + \mathbf{x}\beta_1$, $\mathrm{Var}(\mathbf{Y}|\mathbf{X}) = \sigma^2\mathbf{I}_n$

2. Least squares estimation: $Q = \sum_{i=1}^{n}\{y_i - \mathbb{E}(Y_i|\mathbf{X})\}^2 = \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{X}\beta + \beta^T\mathbf{X}^T\mathbf{X}\beta$
   Least squares *unbiased* estimator: $\hat{\boldsymbol{\beta}} = (\mathbf{X}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$
   $\mathbb{E}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \boldsymbol{\beta}$, $\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$, $\mathrm{Var}(\mathbf{c}^T\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}$

3. Vector of residuals: $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{y}$,
   where $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}\mathbf{X})^{-1}\mathbf{X}^T$ is $n \times n$, symmetric, idempotent, rank p,
   $(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{X} = \mathbf{0}$, $(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{P}_\mathbf{X} = \mathbf{0}$
   $\mathbb{E}(\mathbf{E}|\mathbf{X}) = (\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbb{E}(\mathbf{Y}|\mathbf{X}) = (\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$
   $\mathrm{Var}(\mathbf{E}|\mathbf{X}) = (\mathbf{I}_n - \mathbf{P}_\mathbf{X})\sigma^2\mathbf{I}_n(\mathbf{I}_n - \mathbf{P}_\mathbf{X}) = \sigma^2(\mathbf{I}_n - \mathbf{P}_\mathbf{X})$
   $\mathrm{RSS} = \mathbf{e}^T\mathbf{e} = \mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y}$, $\hat{\sigma}^2 = \frac{\mathrm{RSS}}{n-p} = \frac{\mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y}}{n-p}$

4. Alternative formulation (for models with an intercept)
   $\mathbb{E}(Y_i|\mathbf{X}) = \gamma + \beta_1(x_{i1} - \bar{x}_1) + \beta_2(x_{i2} - \bar{x}_2) + \cdots + \beta_q(x_{iq} - \bar{x}_q)$
   $\mathbb{E}(\mathbf{Y}|\mathbf{X}) = \gamma\mathbf{1}_n + \dot{\mathbf{X}}\dot{\boldsymbol{\beta}}$, where $\dot{\mathbf{X}}_{ij} = x_{ij} - \bar{x}_j$, $\dot{\boldsymbol{\beta}} = \begin{pmatrix}\beta_1 & \cdots & \beta_q\end{pmatrix}^T$, $\gamma = \beta_0 + \beta_1\bar{x}_1 + \cdots + \beta_q\bar{x}_q$
   Least squares *unbiased* estimators: $\hat{\gamma} = \bar{y}$, $\hat{\dot{\boldsymbol{\beta}}} = \left(\dot{\mathbf{X}}_T\dot{\mathbf{X}}\right)^{-1}\dot{\mathbf{X}}^T\mathbf{y}$
   $\mathrm{Var}\left(\hat{\dot{\boldsymbol{\beta}}}\middle|\mathbf{X}\right) = \sigma^2\left(\dot{\mathbf{X}}^T\dot{\mathbf{X}}\right)^{-1}$, $\mathrm{Var}\left(\hat{\gamma}|\mathbf{X}\right) = n^{-1}\sigma^2$, $\mathrm{cov}\left(\hat{\dot{\boldsymbol{\beta}}}, \hat{\gamma}\middle|\mathbf{X}\right) = \mathbf{0}$

5. Distributional results:
   - $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$
   - regression (model) SS $\mathbf{Y}^T\mathbf{P}_\mathbf{X}\mathbf{Y} \sim \sigma^2\chi^2(q, \sigma^{-2}\dot{\boldsymbol{\beta}}^T\dot{\mathbf{X}}^T\dot{\mathbf{X}}\dot{\boldsymbol{\beta}})$
   - RSS $\mathbf{Y}^T(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{Y} \sim \sigma^2\chi^2(n-q-1, 0)$
   - RSS and regression SS are independent
   - $\frac{\mathbf{c}^T\hat{\boldsymbol{\beta}} - \mathbf{c}^T\boldsymbol{\beta}}{\sigma\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}} \sim N(0,1)$ and
     $\frac{\mathbf{c}^T\hat{\boldsymbol{\beta}} - \mathbf{c}^T\boldsymbol{\beta}}{\hat{\sigma}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}} \sim t(n-p)$ (test hypotheses about linear funcs of the parameters)

6. 95% Confidence interval: $\mathbf{c}^T\hat{\boldsymbol{\beta}} \pm t_{0.025}\hat{\sigma}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}$
   CI for future response: $\mathbf{x}_*^T\hat{\boldsymbol{\beta}} \pm t_{0.025}\hat{\sigma}\sqrt{\mathbf{x}_*^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_*}$
   Prediction interval: $\mathbf{x}_*^T\hat{\boldsymbol{\beta}} \pm t_{0.025}\hat{\sigma}\sqrt{1 + \mathbf{x}_*^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_*}$

7. To test $\boldsymbol{\beta} = \mathbf{0}$,
   $F = \frac{\text{regression SS}}{\text{RSS}} \sim F(p, n-p)$ (chk simple linear regression, but with different SS)

8. To test a more general linear hypothesis about the coefficients of the model, $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$,
   Extra SS $= \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}\right)^T\left(\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T\right)^{-1}\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}\right)$
   $F = \frac{(\text{ESS for } H_0)/c}{\text{RMS}} = \frac{(\text{RSS under } H_0 - \text{RSS under full model})/c}{\text{RMS}} \sim F(k, n-p)$