MATH 204 Cheat Sheet

Simple Linear Regression

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

β_1 is the change in the mean of Y_i for a 1 unit increase in x_i,
β_0 is the mean when $x_i = 0$
$S_{XX} = \sum(x_i - \bar{x})^2$, $S_{YY} = \sum(y_i - \bar{y})^2$,
$S_{XY} = \sum(x_i - \bar{x})(y_i - \bar{y})$

Estimating σ^2

1. Standard deviation of $\hat{\beta}_1$: $\sigma_{\hat{\beta}_1} = \sqrt{var(\hat{\beta}_1)} = \sigma/\sqrt{S_{XX}}$
2. Variance of residuals: $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{SSE}{n-2}$
3. $SSE = S_{YY} - \hat{\beta}_1 S_{XY}$
4. $\hat{\sigma}_{\hat{\beta}_1} = \hat{\sigma}/\sqrt{S_{XX}}$

Inference about β_1

1. When the error terms are normal, $\hat{\beta}_1 \sim \mathcal{N}(\beta_1, \sigma^2/S_{XX})$
2. $T = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/S_{XX}} \sim t_{n-2}$

   $$\mathcal{H}_0 : \beta_1 = 0 \quad vs \quad \mathcal{H}_a : \beta_1 \neq 0$$

   $T_{obs} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{S_{XX}}}$
   Compare T_{obs} with the student distribution $t_{n-2,\alpha/2}$ to get RR.
3. Could get same conclusion from p-value, which illustrates the probability that our results occurred under $\mathcal{H}_0$.
4. Confidence interval for β_1: $\hat{\beta}_1 \pm t_{n-2,\alpha/2}\frac{\hat{\sigma}}{\sqrt{S_{XX}}}$.

ANOVA

1. $SS_{reg} = S_{YY} - SSE = \sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2 = \sum(\hat{y}_i - \bar{y})^2$
2. $T \sim t_v$, $\quad T^2 \sim \mathcal{F}(1, v)$, where the latter is the Fisher-Snedecor dis.
3. ANOVA table guide:
   - (X, Sum Sq) $= SS_{reg}$
   - (Residuals, Sum Sq) = SSE
   - (Residuals, Df) $= n - 2$
4. lm summary table
   - t-value (slope): $T_{obs} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}}$
   - F-statistic : T_{obs}^2
   - Residual std error: $\hat{\sigma}$

Correlation

1. corr(X,Y) = corr(Y,X)
2. $r = S_{XY}/\sqrt{S_{XX}S_{YY}}$ is an estimator for ρ (the true pop. correlation).
3. $(1-\alpha)100\%$ confidence interval for ρ: transform r to $z = 0.5\ln(\frac{1+r}{1-r})$. Build an interval:$z \pm \frac{z_{\alpha/2}}{\sqrt{n-3}} = (c_l, c_u)$, where $z_{\alpha/2}$ is from the standard Normal table. Then, the interval is

   $$\left(\frac{e^{2c_L} - 1}{e^{2c_L} + 1}, \frac{e^{2c_U} - 1}{e^{2c_U} + 1}\right)$$
4. Coefficient of determination: $R^2 = 1 - SSE/S_{YY}$

Estimating response

1. Mean response confidence interval:
   $\hat{y}_0 \pm t_{n-2,\alpha/2}\hat{\sigma}\sqrt{1/n + (x_0 - \bar{x})^2/S_{XX}}$
2. Individual value Y_0 confidence interval:
   $\hat{y}_0 \pm t_{n-2,\alpha/2}\hat{\sigma}\sqrt{1 + 1/n + (x_0 - \bar{x})^2/S_{XX}}$

Residual Analysis