Chapters 13

Sample correlation coefficient

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

$$= \frac{\sum xy - \left(\sum x\right)\left(\sum y\right)/n}{\sqrt{\left[\sum x^2 - \left(\sum x\right)^2/n\right]\left[\sum y^2 - \left(\sum y\right)^2/n\right]}} = \frac{Sxy}{\sqrt{Sxx\,Syy}}$$

For testing H_0: ρ=0 vs. H_A: ρ≠0

Test statistic $t_{n-2} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$ has df = n-2

If $t_{n-2} > t_{\alpha/2,n-2}$ then reject H_0.

Estimated regression model $\hat{y}_i = b_0 + b_1 x$

The Least Square Estimates are

$$b_1 = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = \frac{Sxy}{Sxx}$$

And $b_0 = \bar{y} - b_1\bar{x}$

Error $e_i = Y_i - \hat{Y}_i$

Total Sum of Squares

$$SST = Syy = \sum(y-\bar{y})^2$$

$$= \sum_{1}^{n} y_i^2 - n\bar{y}^2$$

Regression Sum of Squares $SSR = \sum(\hat{y}-\bar{y})^2$

Error Sum of Squares $SSE = SST - SSR$

$$SSE = Syy - \frac{(Sxy)^2}{Sxx}$$

(ALSO $SSE = \sum(y-\hat{y})^2$)

$$Sxy = \sum xy - \left(\sum x\right)\left(\sum y\right)/n$$

Coefficient of Determination

R- Squared = $R^2 = \frac{SSR}{SST}$

$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} = r^2$$

Standard Error of the Estimate

$$s_\varepsilon = \sqrt{\frac{SSE}{n-1-1}}$$

Standard Error of the Slope

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{\sum(x-\bar{x})^2}} = \frac{s_\varepsilon}{\sqrt{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}} = \frac{s_\varepsilon}{\sqrt{Sxx}}$$

For testing H_0: $\beta_1 = 0$ vs. H_1: $\beta_1 \neq 0$

The test statistic & **C.I. for the slope**

$$t_{n-2} = \frac{b_1 - \beta_1}{s_{b_1}} \quad \& \quad b_1 \pm t_{\alpha/2} s_{b_1}$$

C.I. for the **mean** of y given a particular x_p

$$\hat{y} \pm t_{\alpha/2} s_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x-\bar{x})^2}}$$

P.I. for an **Individual value of y** given a particular x_p

$$\hat{y} \pm t_{\alpha/2} s_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x-\bar{x})^2}}$$

For testing auto correlation H_0: ρ=0 vs. H_A: ρ≠0

Durbin-Watson Test statistic

$$d = \frac{\sum_{t=1}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

, If $d < d_L$ reject H_0.

If $d > d_U$ don't reject H_0. Inconclusive otherwise.

Chapters 14

Estimated multiple regression model

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

Two variable model is

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

$e_i = y_i - \hat{y}_i$ residuals from regression model

Proportion of variation in *y* explained by all ***x*** variables adjusted for sample size and the number of ***x*** variables.

$$R_A^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-k-1}\right)$$

For testing

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$

H_A: at least one $\beta_i \neq 0$

Test statistic $$F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{MSR}{MSE}$$

with $df_1 = k$ and $df_2 = n-k-1$

Contribution of X_j given other X variables

$$SSR(X_j \mid \text{All except } X_j) = SSR(\text{All}) - SSR(\text{All except } X_j)$$

Coefficient of Partial Determination X_j given other X variables already included

$$r^2_{YX_j \bullet (\text{All except } X_j)} = \frac{SSR(X_j \mid \text{All except } X_j)}{SST - SSR(\text{All}) + SSR(X_j \mid \text{All except } X_j)}$$

For testing H_0: $\beta_i = 0$ vs. H_A: $\beta_i \neq 0$

The test statistic & C.I. for the slope β_i are

$$t_{n-k-1} = \frac{b_i - 0}{s_{b_i}} \quad \text{or} \quad F_{1,n-k-1} = \frac{SSR(X_j \mid \text{All except } X_j)}{MSE}$$

$$t^2_{n-k-1} = F_{1,n-k-1}$$

C.I. for the slope β_i is $b_i \pm t_{\alpha/2} s_{b_i}$

The estimate of the standard error of the regression model

$$s_\varepsilon = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{MSE}$$

Contribution of X_j given other X variables

$$SSR(X_j \mid \text{All except } X_j) = SSR(\text{All}) - SSR(\text{All except } X_j)$$

For testing H_0: $\beta_c = \beta_d = \ldots = \beta_m = 0$

against H_A: at least one $\beta_i \neq 0$

Test statistic

$$F = \frac{\frac{SSR_{Full} - SSR_{Reduced}}{m}}{\frac{SSE}{n-k-1}} = \frac{\frac{SSE_{Reduced} - SSE_{Full}}{m}}{MSE}$$

with $df_1 = m$ and $df_2 = n-k-1$