Alanna Hazlett
3/18/24

MPle Continued

2.) (d.)

Source of variation	df	SS	MS
Regression	5-1 = 4	84.6244	21.1561
Error	113-5 = 108	120.0576	1.0816
Total	113-1 = 112	204.6820	***

$S = 1.04 \quad S^2 = 1.0816 = MS_{res}$

$MS_{res} = \frac{SS_{res}}{113-2} \quad 1.0816 \cdot 111 = 120.0576 = SS_{res}$

$F\text{ stat} = 19.56 \quad F\text{-stat} = \frac{MS_R}{MS_{res}} \quad 19.56 \cdot 1.0816 = 21.1561 = MS_R$

$SS_R = MS_R \cdot df_R = 21.1561 \cdot 4 = 84.6244 = SS_R$

$SS_T = SS_R + SS_{res} = 84.6244 + 120.0576 = 204.6820 = SS_T$

(e.) What is R^2 for this model? Interpret in context.

$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{res}}{SS_T} = \frac{84.6244}{204.6820} = 0.4134$

This is the proportion of variance in the response variable that is explained by the predictor variables.

About 41.34% of the variance in infection risk can be explained by Stay, Age, Xrays, and Services.

(f.) What is R^2_{adj} for this model?

$R^2_a = 1 - \left(\frac{n-1}{n-p}\right)\left(\frac{SS_{res}}{SS_T}\right) = 1 - \left(\frac{113-1}{113-5}\right)\left(\frac{120.0576}{204.6820}\right) = 1 - 0.60828$

$= 0.3917$

3.) ANOVA F statistic is significant, t-statistics for both predictors are insignificant. Does this warrant concern?

- If the F statistic is significant then we reject the null hypothesis of $\hat{\beta}_0 = \hat{\beta}_1 = \hat{\beta}_2 = 0$, our data support alternative hypothesis that at least one coefficient $\neq 0$.
- If t statistic is insignificant for the predictors we fail to reject the null hypothesis of $\hat{\beta}_j = 0$ for each predictor.

This is concerning, as these two tests should draw the same conclusions, either all reject the null hypotheses or all fail to reject the null hypotheses.

4.) $H = X(X'X)^{-1}X'$

Show that H is idempotent, so $HH = H$.

$H \cdot H$

$= (X(X'X)^{-1}X')(X(X'X)^{-1}X')$

$= X(X'X)^{-1}(X'X)(X'X)^{-1}X'$ $\quad (X'X)(X'X)^{-1} =$ identity matrix

$= X(X'X)^{-1}X'$

$H \times H = H \quad (X(X'X)^{-1}X')(X(X'X)^{-1}X') = (X(X'X)^{-1}X')$