MA 214 Homework 6

12.8

a. $t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} = 1.35 \rightarrow$ p-value $= 0.096$

$df = n-p-1 = 25-2-1 = 22$ p-value > alpha = 0.05 → Fail reject the H_0.

This means that we do not have sufficient evidence to conclude that β_1 is greater than 0.

b. $t = \frac{\hat{\beta}_2 - \beta_{2,0}}{SE(\hat{\beta}_2)} = 3.41 \rightarrow$ p-value $= 0.0025$

p-value < alpha = 0.05 → Reject the H_0.

This means we have sufficient evidence to conclude that β_2 is not equal to 0 at the 0.05 significance level.

c. $\hat{\beta}_1 \pm t_{\frac{\alpha}{2}, df} \times SE(\hat{\beta}_1)$ $df = n-p-1 = 25-2-1 = 22$ $\frac{\alpha}{2} = 0.05 \rightarrow$ t value $= 1.717$

Lower Limit: $3.1 - 1.717 \times 2.3 = -0.85$ $(-0.85, 7.05)$

Upper Limit: $3.1 + 1.717 \times 2.3 = 7.05$

d. $df = 22$ $\alpha/2 = 0.005 \rightarrow$ t value $= 2.819$

Low Limit $= 0.92 - 2.819 \times 0.27 \approx 0.16$ $(0.16, 1.68)$

Upper Limit $= 0.92 + 2.819 \times 0.27 \approx 1.68$

12.18

a. β-estimate x_1: For every one unit increase in the proportion of the block that is low-density residential area, we expect the population density to increase 2.006 units, assuming the proportion of the block that is high-density residential area remains constant.

β-estimate x_2: For every one unit increase in the proportion of the block that is high-density residential area, we expect the population density to increase 5.006 units, assuming the proportion of the block that is low-density residential area remains constant.

The intercept term (-0.0304) represents the expected population density when both x_1 and x_2 are zero. In this context, it would represent the population density in a block that has no residential areas, which may not have a practical interpretation.

b. $R^2 = 0.686$ means that approximately 68.6% of the variability in population density can be explained by the proportion of the block that is low density residential area and the proportion of the block that is high-density residential area. The remaining 31.4% of the variability in population density is due to other factors not included in the model.

c. $H_0: \beta_1 = \beta_2 = 0$ vs. H_a: At least one of β_1 and β_2 is not equal to 0.

d. $F = \frac{(TSS - RSS)/p}{RSS/(n-p-1)} = \frac{(R^2/p)}{(1-R^2)/(n-p-1)}$

Given $R^2 = 0.686$, $p = 2$, $n = 125$

$F \approx 133.27$

e. The p-value for F-statistic of 133.27 is extremely small, essentially zero.

Reject the H_0.

This means that we have strong evidence to conclude that at least one of the predictors (x_1 or x_2) is significantly related to the response variable at the 0.01 significance level.

12.26

$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_{18} = 0$

H_a: At least one of $\beta_1, \beta_2 \dots \beta_{18}$ is not equal to 0.

$F = \dfrac{(R^2/p)}{(1-R^2)/(n-p-1)}$ Given: $R^2 = 0.95$, $p = 18$, $n = 20$

≈ 1.06

$\rightarrow$ p-value ≈ 0.656

$>$ alpha $= 0.05$

$\hookrightarrow$ Fail to reject the H_0

This means that we do not have strong evidence to conclude that at least one of the predictors is significantly related to the response variable (SAT score) at the 0.05 significance level.

In other words, despite the high R^2 value, the model does not provide strong evidence that the psychological and sociological factors are useful in predicting SAT scores, given the small sample size and the large number of predictors.

12.36

a. $\hat{y} = 52484 + 2941 \times 10 + 16880 \times 0 + 11108 \times 1$

$= 93002$

b. $\hat{y} = 52484 + 2941 \times 10 + 16880 \times 1 + 11108 \times 0$

$= 98774$

c. The predicted values given in parts a and b are point estimates, meaning they represent a single "best guess" for the salary of a UPA member given their years of experience, PhD status, and manager status. However, these point estimates do not account for the uncertainty associated with prediction.

A 95% prediction interval provides a range of values that is likely to contain the actual salary of a UPA member with a certain level of confidence. This interval accounts for the variability in the data and the uncertainty in the prediction.

In other words, a prediction interval gives a range in which we can be fairly confident the actual outcome will fall, assuming the model is correct. It provides more information than a point estimate by quantifying the uncertainty in the prediction.

12.174.

a. $\hat{y} = 90.1 - 1.84 \cdot x_1 + 0.285 \cdot x_2$

b. $R^2 = 91.6\%$.

This means that 91.6% of the variability in the response variable y can be explained by the predictors x_1 and x_2 in the model. The remaining 8.4% of the variability in y is due to other factors not included in the model.

c. $H_0: \beta_1 = \beta_2 = 0$

H_a: At least one of β_1 or β_2 is not equal to 0.

Given: F-statistic $= 64.91$ and the corresponding p-value is 0.001

Since the p-value (0.001) is less than the significance level ($\alpha = 0.05$), we reject the H_0. This means that we have strong evidence to conclude that at least one of the predictors (x_1 or x_2) is significantly related to the response variable y at the 0.05 significance level.

d. $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$

p-value = 0.001 -> Less than alpha = 0.05 -> reject the H_0 and conclude that x_1 is significantly related to y.

e. Given $S = 10.68$.

This means that, on average, the actual values of y deviate from the predicted values given by about 10.68 units. In other words, if we use the model to predict y based on the predictors x_1 and x_2, we can expect the prediction to be off by about 10.68 points on average.