Name: Jake Bohman

Username: jbohman 1

jbohman 1

(1)
a) 2
b) 4
c) 1
d) 3
e) 5

(2) a) false; since both predictors cover the same information, the coefficient will have to change to cover fully that which previously was shared.

b) False, since the model has only established correlation, not causation. Just because we observe $\hat{\beta_1} = 5.7$ does not mean that we can automatically cause a jump of 5.7 units by increasing x_1 by 1.

c) true

(3) a) False, we can only conclude that at least one mean is different from the others, since the alternative hypothesis is H_A: At least one $\beta_i \neq 0$, not every $\beta_i \neq 0$.

b) true

c) false, since a pairwise analysis man natt tell us this.

(4) a) false, decreasing the significance level increases the probability of a false positive, which is Type 2 not Type 1. Correct the statement by changing Type 1 to Type 2.

b) true, since with larger sample sizes the CI for the mean will be smaller.

c) true, since the correlation coefficient measures association.

(5) Analyzing the studentized residuals can help us analyze the residuals without the model being influenced by them, allowing us to more easily identify outliers if they are high leverage, while standardized residuals include the data point in the estimated standard deviation of the error.

(6) I would use Cook's Distance, since it takes into account both standardized residuals and leverage, allowing us to identify influential points, which are captured by a balance of the two, rather than just one.

⑦ d) $\delta_{FE} = n-k-1 = 76-5-1 = 70$ degrees of freedom (error)

b) The reference level is General Mills, since it has no coefficient.

---

⑧ The estimated group effect is $(11.364 - 96.364) - 106.97 = -11.97$ calories

In other words, the mean amount of calories in Qvaker Oats cereal (95.0) is 11.97 less than the overall mean (106.97).

---

⑨ The coefficient for the interaction term between sugars and protein shows the effect that each has on the other's effect on calories per serving, allowing for adjustments in manufacturer.

---

⑩ a) Can the proportion of full-time instructional staff be used to predict average cost of tuition?

$Y = \beta_0 + \beta_1 X + \epsilon$, where Y is the cost of tuition in dollars, and X is the percentage of full-time instructional staff employed.

$H_0: \beta_1 = 0$

b) Is the mean cost of tuition different between different types of colleges?

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$, where Y is the average cost of tuition in dollars,

$X_1 = \begin{cases} 1 & \text{if liberal arts} \\ 0 & \text{otherwise} \end{cases}$, $X_2 = \begin{cases} 1 & \text{if community college} \\ 0 & \text{if otherwise} \end{cases}$,

$X_3 = \begin{cases} 1 & \text{if vocational} \\ 0 & \text{otherwise} \end{cases}$, $X_4 = \begin{cases} 1 & \text{if affiliated} \\ 0 & \text{otherwise} \end{cases}$

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

c) Is the following model significant with 95% confidence for estimating average cost of tuition?

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$, where Y is the average cost of tuition, $X_1 = \begin{cases} 1 & \text{if private} \\ 0 & \text{if public} \end{cases}$, $X_2 = \%$ full time instructional staff

$H_0: p > 0.05$

11) a) Constant variance: the residuals plot shows that the variation is different between cereals, not meeting this condition.

normality: the normal quantile plot has large tails on both ends, not meeting this condition

randomness: given as met in the description

independence: there is no reason to think one cereal has an effect on the others

* Since there is near zero mean for every group, we can conduct tests for the mean of each group.

b) $H_0: M_G = M_K = M_N = M_P = M_Q = M_R$ Null: All of the mean calories are the same.

M_i = mean calories for cereal i

$H_A:$ At least one M_i is not equal to the rest   Alt: At least one mean calories is different.

c) I believe that the test will reject the null hypothesis in favor of the alternative, because looking at the R summary output shows Nabisco as having an average calories 24.7 less than General Mills, with standard errors of 8.5 and 4.0 respectively, it is unlikely that the CIs for the means overlap enough.

12) The conclusions are not contradictory, since Person A makes a conclusion about Year and Miles put together, while Person B is making two separate conclusions about Year and miles independently. It is hard to compare Person A's value of 0.26 and B's of 0.77 and -0.34 because A's is adjusted R^2 and B's are both r.