Summative Quiz 1 (An Overview of Multiple Regression for Estimation, Adjustment, and Basic Prediction, and Multiple Linear Regression) Solutions:

1. Which of the following is true about an observational prospective cohort study being designed to study the association between cigarette smoking (smoking ≥ 1 pack of cigarettes per day, smoking < 1 pack of cigarettes per day, or no smoking) to be assessed on 3/1/19, and catching a cold between 3/2/19 and 3/31/19?

   The 2010 National Hospital Ambulatory Medical Care Survey (NHAMCS) is a national (United States) sample survey of visits to hospital outpatient and emergency departments. This survey was conducted by the National Center for Health Statistics. In this exercise, multiple linear regression will be used to examine factors associated with patient waiting time (*in minutes*) of persons admitted to the Emergency Departments (EDs) of participating hospitals in 2010. The average waiting time reported by the over 27,000 survey participants is 56.3 minutes (s= 78.7 minutes, with a range of 0 to 1,335 minutes).
   The linear regression that relates average time to visitors' sex, racial identity (white (reference), black, other), insurance type (private vs not private) , and systolic blood pressure (mmHg) is of the form:

$$\hat{y} = \hat{\beta}_o + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5$$

   where x_1 = 1 for female (0 for male); x_2 = 1 for black, x_3 = 1 for other racial identity; x_4= 1 for private insurance, 0 for non-private; x_5 = systolic blood pressure (mmHg)
   The regression equation with the actual intercept and slope estimates is:

$$\hat{y} = 62.6 + \ -2.4x_1 + 18.8x_2 + 1.7x_3 + \ -7.9x_4 + \ -0.05x_5$$

   Based on these results, what is the estimated mean waiting time for females who identify as white, who do not have private insurance, with systolic blood pressures of 125 mmHg? (rounded to the nearest integer)

   Answer: 54 mmHg.

   Reasoning: $\hat{y} = 62.6 - 2.4(1) + 18.8(0) + 1.7(0) - 7.9(0) + \ \ -0.05 * 125 =$ 53.95 ≈ 54.0 minutes

2. (this item references the same linear regression results as item #1)

   Based on these results, what is the estimated mean difference in waiting times for the group from question 1 compared to males who identify as Black, who have private insurance and systolic blood pressures of 115 mmHg? (rounded to 1 decimal place)

   Answer: -13.8 minutes
   Reasoning: For this seconds group,

   $\hat{y} = 62.6 - 2.4(0) + 18.8(1) + 0.7(0) - 7.9(1) + \quad -0.05 * 115 = 67.75 \ \approx$ 67.8 minutes.

   So the mean difference is 54.0– 67.8 = -13.8.

3. (this item references the same linear regression results as item #1)

   What is the average difference in waiting times for those who identify as Black compared to those who identify as White, adjusted for sex, insurance type, and systolic blood pressure?

   Answer: 18.8 minutes

   Reasoning: $\hat{\beta}_2 = 18.8$ minutes, and $\hat{\beta}_2$ is the estimated adjusted mean difference in waiting times between those who identify as Black and those who identify as White, adjusted for sex, insurance type, and systolic blood pressure.

4. (this item references the same linear regression results as item #1)

   The standard error for $\hat{\beta}_2$, $\widehat{SE}\left(\hat{\beta}_2\right)$, =1.3. Estimate a 95% CI for the adjusted mean difference in waiting times for those who identify as Black compared to those who identify as White, adjusted for sex, insurance type, and systolic blood pressure.

   Answer: $(16.2, 21.4)$ minutes

   Reasoning: $18.8 \pm 2 * 1.3 = (16.2, 21.4)$ minutes

5. (this item references the same linear regression results as item #1)

   What additional information is needed to determine whether the relationship between waiting times and racial identify is confounded by at least one of the following: sex, insurance type, and/or systolic blood pressure?

Answer: The unadjusted mean differences in waiting times for those who identify as Black, and other, respectively, compared to those who identify as White.

6. (this item references the same linear regression results as item #1)

   The adjusted R^2 for this model is 1.1 %. What is the best explanation for this given that some of these adjusted associations are statistically significant?

   Answer: While there are differences in the mean waiting times for different predictor groups, there is still much variability in the individual waiting times around their respective group means.

7. The researchers refit the above model detailed in question 1, treating systolic blood pressure as categorical (4 quartiles) instead of continuous. The resulting model is of the form:

$$\hat{y} = \hat{\beta}_o + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 + \hat{\beta}_7 x_7$$

   where $x_1 - x_4$ are the same as with the previous model, and x_5-x_7 are indicators of systolic blood pressure quartiles 2 -4, respectively. (quartile 1 is the reference group)

   What is the null hypothesis, in terms of the slopes (betas), for testing whether systolic blood pressure is a statistically significant predictor of waiting times?

   Answer: $H_0: \beta_5 = \beta_6 = \beta_7 = 0$

8. The p-value from the partial F-test for systolic blood pressure quartiles from the model detailed in item 7 is 0.42. What conclusion should the researchers make?

   Answer: After accounting for sex, racial identity, and insurance type, SBP is not a statistically significant predictor of waiting times.

9. The researchers decided to drop systolic blood pressure from their model, but add an interaction term between sex and insurance type. The results are as follows:

$$\hat{y} = 55.5 + -1.3x_1 + -6.8x_2 + -1.7x_3 + 17.7x_4 + 2.0x_5$$

where x_1 = 1 for female (0 for male); x_2= 1 for private insurance, 0 for non-private; x_3=x_1*x_2; x_4 = 1 for black, x_5 = 1 for other racial identity.

Based on this model, what is the estimated mean waiting time for females with private insurance who identify as Black?

Answer: 63.4 minutes

Reasoning:

For this group, x_1=1, x_2 =1, and hence the interaction term x_3=x_1*x_2 = 1*1 = 1, and:

$$\hat{y} = 55.5 + -1.3(1) + -6.8(1) + -1.7(1 \times 1) + 17.7(1) + 2.0(0) = 63.4\ minutes$$

10. The researchers decided to drop systolic blood pressure from their model, but add an interaction term between sex and insurance type. The results are as follows:

$$\hat{y} = 55.5 + -1.3x_1 + -6.8x_2 + -1.7x_3 + 17.7x_4 + 2.0x_5$$

where x_1 = 1 for female (0 for male); x_2= 1 for private insurance, 0 for non-private; x_3=x_1*x_2; x_4 = 1 for black, x_5 = 1 for other racial identity.

Based on this model, what is the estimated mean waiting time for males with private insurance who identify as Black?

Answer: 66.4 minutes

Reasoning:

For this group, x_1=0, x_2 =1, and hence the interaction term x_3=x_1*x_2 = 0*1 = 0, and:

$$\hat{y} = 55.5 + -1.3(0) + -6.8(1) + -1.70(0) + 17.7(1) + 2.0(0) = 66.4\ minutes$$

11. The researchers decided to drop systolic blood pressure from their model, but add an interaction term between sex and insurance type. The results are as follows:

$$\hat{y} = 55.5 + \ -1.3x_1 + \ -6.8x_2 + \ -1.7x_3 + \ 17.7x_4 + \ 2.0x_5$$

where x_1 = 1 for female (0 for male); x_2= 1 for private insurance, 0 for non-private; x_3=x_1*x_2; x_4 = 1 for black, x_5 = 1 for other racial identity.

Based on this model, what is the estimated mean difference in waiting times for females with private insurance compared to females without private insurance (adjusted for race)?

Answer: -8.5 minutes

Reasoning:

When sex is female, x_1=1 and x_3 = 1*x_2. As such, the equation becomes….

$$\hat{y} = 55.5 + \ -1.3(1) + \ -6.8(x_2) + \ -1.7(1 \times x_2) + \ 17.7x_4 + \ 2.0x_5$$

$$\hat{y} = 55.5 + \ -1.3 + \ (-6.8 + -1.7)x_2 \ + \ 17.7x_4 + \ 2.0x_5$$

So the overall slope of insurance type for females (and hence the adjusted mean difference in waiting times females with private insurance compared to females without private insurance) is -6.8+ -1.7 = -8.5 minutes

12. The researchers decided to drop systolic blood pressure from their model, but add an interaction term between sex and insurance type. The results are as follows:

$$\hat{y} = 55.5 + \ -1.3x_1 + \ -6.8x_2 + \ -1.7x_3 + \ 17.7x_4 + \ 2.0x_5$$

where x_1 = 1 for female (0 for male); x_2= 1 for private insurance, 0 for non-private; x_3=x_1*x_2; x_4 = 1 for black, x_5 = 1 for other racial identity.

Based on this model, what is the estimated mean difference in waiting times for males with private insurance compared to males without private insurance (adjusted for race)?

Answer: -6.8 minutes

Reasoning:

When sex is female, x_1=0 and x_3 = 0*x_2 =0. As such, the equation becomes….

$$\hat{y} = 55.5 + \ -1.3(0) + \ -6.8(x_2) + \ -1.7(1 \times 0) + \ 17.7x_4 + \ 2.0x_5$$

So the slope of insurance (x_2) for males (and hence the adjusted mean difference in waiting times females with private insurance compared to males without private insurance) is -6.8 minutes.

13. The researchers decided to drop systolic blood pressure from their model, but add an interaction term between sex and insurance type. The results are as follows:

$$\hat{y} = 55.5 + \ -1.3x_1 + \ -6.8x_2 + \ -1.7x_3 + \ 17.7x_4 + \ 2.0x_5$$

where x_1 = 1 for female (0 for male); x_2= 1 for private insurance, 0 for non-private; x_3=x_1*x_2; x_4 = 1 for black, x_5 = 1 for other racial identity.

The standard error for the slope of the interaction term (x_3) is 2.1. Is the (race-adjusted) relationship between wait times and insurance type statistically significantly ($\alpha = 0.05$) modified by sex?

Answer: No.

Reasoning: The 95% CI for the true value of β_3 is -1.7±2(2.1) → (-5.9, 2.5). As this confidence interval includes the null value of 0, the resulting p-value for testing $H_o\!: \beta_3 = 0$ will be greater than 0.05.