PHC 6091

Dr. Hu

4/09/2020

**Background:** Dataset exam2.sas7dat is the data of an observational study designed to compare the effectiveness of a new drug, which is viewed by the investigators as the active treatment vs. a standard inhaler, for treating asthma. The patient follow-up period begins at the hospitalization at which treatment for asthma is initiated, using either the new drug or an inhaler. The follow-up period (since the initiation of the interventions) is one year. Variables in the study include:

Severity = An index for asthma severity, with higher values indicating great severity.

QOL = quality of life index

Hosp1yr = a binary 0-1 variable indicating the presence of at least one hospitalization during the 1 year follow-up period

Treatment = a binary 0-1 indicator variable where 1 indicates that a patient received the active experimental drug and 0 indicates the patient used the inhaler (standard of care).

QOLCat = a 4-level ordered categorical variable, with higher scores representing a higher QOL

Num\_hosp = the number of hospital, emergency department, or urgent care admissions during the 1 year follow-up period.

Age1 = Age in years

Male = 0-1 indicator variable for male sex

Hispanic = 0-1 indicator variable for Hispanic ethnicity (0=non-Hispanic; 1=Hispanic)

MaternalEd = Maternal education in years

Urban = 0-1 indicator for residence in an urban area (0=rural, 1= urban)

FamIncome = Annual family income in $

**Question 1: Examine is the effect of the QOL category (QoLCat) on severity. Please answer the following questions** (25 pts)

1. Using One-way analysis of variance (ANOVA) model to test if QoLCat is a statistically significant factor on severity. (15 pts)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ANOVA** | | | | | |
| Severity | | | | | |
|  | Sum of Squares | df | Mean Square | F | Sig. |
| Between Groups | 1153.659 | 3 | 384.553 | 35.772 | .000 |
| Within Groups | 8546.319 | 795 | 10.750 |  |  |
| Total | 9699.977 | 798 |  |  |  |

Ho= QoLCat is **not** statistically significant factor on severity.

H1= QoLCat is **a** statistically significant factor on severity.

**F= 35.772; p-value <0 .000**

Alpha = 0.05, H0 can be rejected since the p-value <0.000, suggesting that there is evidence that QoLCat is **a** statistically significant factor on severity.

1. What is the degree of freedom for the QoLCat effect? (2 pts)

The degree of freedom for the QoLCat effect is 4-1= **3.**

1. What is the total sum of square (SS) for severity that is explained by QoLCat? (2 pts)

The total sum of square for severity that is explained by QoLCat is **1153.659.**

1153.659+ 8546.319 = 9699.977.

(c) What is the mean square (MS) for severity that is explained by QoLCat? (2 pts)

The mean square (MS) for severity that is explained by QoLCat is **384.553.**

1. What is the degree of freedom for the error term? (2 pts)

The degree of freedom of the error term is 799-4= **795.**

1. What is the mean square error (MSE)? (2 pts)

The mean square error is **10.750.**

1. Conduct a hypothesis test to assess the effect of QoLCat on severity. (5 pts)

Ho= QoLCat is **not** statistically significant factor on severity.

H1= QoLCat is **a** statistically significant factor on severity.

**F= 35.772; p-value <0 .000**

Alpha = 0.05, H0 can be rejected since the p-value <0.000, suggesting that there is evidence that QoLCat is **a** statistically significant factor on severity.

**2**. Obtain the estimated mean of severity for each of the four QoL levels. (4 pts)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Descriptives** | | | | | | | | | | |
| Severity | | | | | | | | | | |
|  | | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum | Between- Component Variance |
| Lower Bound | Upper Bound |
| 1 | | 112 | 19.17 | 2.956 | .279 | 18.62 | 19.72 | 12 | 26 |  |
| 2 | | 312 | 17.76 | 3.205 | .181 | 17.41 | 18.12 | 4 | 27 |  |
| 3 | | 272 | 16.38 | 3.125 | .189 | 16.01 | 16.75 | 9 | 26 |  |
| 4 | | 103 | 15.14 | 4.126 | .407 | 14.33 | 15.94 | 5 | 27 |  |
| Total | | 799 | 17.15 | 3.486 | .123 | 16.91 | 17.39 | 4 | 27 |  |
| Model | Fixed Effects |  |  | 3.279 | .116 | 16.92 | 17.38 |  |  |  |
| Random Effects |  |  |  | .793 | 14.63 | 19.67 |  |  | 2.018 |

The estimated mean of severity for each of the four QoL levels are **19.17, 17.76,16.38, and 15.14 respectively**.

1. Using the Tukey-Kramer method to conducted the pairwise comparisons between different levels of QoLCat that are adjusted for multiple comparisons. Which pairs have significant difference in the mean of severity? (Hint: The total number of unique pairwise comparisons is 6). (6 pts)

K=4 because there are four levels of QOLCat. 4\*(4-1)/2=6

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Multiple Comparisons** | | | | | | | |
| Dependent Variable: Severity | | | | | | | |
|  | (I) QOLCat | (J) QOLCat | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|  | Lower Bound | Upper Bound |
| Tukey HSD | 1 | 2 | 1.407\* | .361 | .001 | .48 | 2.34 |
| 3 | 2.791\* | .368 | .000 | 1.84 | 3.74 |
| 4 | 4.034\* | .448 | .000 | 2.88 | 5.19 |
| 2 | 1 | -1.407\* | .361 | .001 | -2.34 | -.48 |
| 3 | 1.384\* | .272 | .000 | .68 | 2.08 |
| 4 | 2.627\* | .373 | .000 | 1.67 | 3.59 |
| 3 | 1 | -2.791\* | .368 | .000 | -3.74 | -1.84 |
| 2 | -1.384\* | .272 | .000 | -2.08 | -.68 |
| 4 | 1.243\* | .379 | .006 | .27 | 2.22 |
| 4 | 1 | -4.034\* | .448 | .000 | -5.19 | -2.88 |
| 2 | -2.627\* | .373 | .000 | -3.59 | -1.67 |
| 3 | -1.243\* | .379 | .006 | -2.22 | -.27 |
| \*. The mean difference is significant at the 0.05 level.  **A screenshot of a cell phone  Description automatically generated** | | | | | | | |

**All p-values listed above are below 0.05. Reject H0. All** the pair listed (1-2,1-3,1-4,2-1,2-3,2-4,3-1,3-2,3-4,4-1,4-2,4-3) have a statistically significant difference in the mean of severity. Mean differences and CI for each is listed above.

**Question 2: Using Two-way ANOVA model to test if Treatment and QoLCat are a statistically significant factors on severity** (31 pts)

1. In the two-way ANOVA model, you first include the two main effect and the interaction between the two main effects. (11 pts)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter |
| Corrected Model | 2323.571a | 7 | 331.939 | 35.595 | .000 | .240 | 249.165 |
| Intercept | 179210.294 | 1 | 179210.294 | 19217.400 | .000 | .960 | 19217.400 |
| QOLCat | 1065.544 | 3 | 355.181 | 38.087 | .000 | .126 | 114.262 |
| Treatment | 854.473 | 1 | 854.473 | 91.628 | .000 | .104 | 91.628 |
| QOLCat \* Treatment | 177.312 | 3 | 59.104 | 6.338 | .000 | .023 | 19.014 |
| Error | 7376.406 | 791 | 9.325 |  |  |  |  |
| Total | 244709.000 | 799 |  |  |  |  |  |
| Corrected Total | 9699.977 | 798 |  |  |  |  |  |



In this case alphai= QOLCat, betaj = Treatment and (alpha\*beta)ij is the interaction effect between

QOLCat and treatment.

(a) Conduct an overall F test to examine if this ANOVA model is significantly better than the null model. (3 pts)

Ho= The ANOVA model is **not** significantly better than the null model.

H1= The ANOVA model **is** significantly better than the null model.

**F= 35.595; p-value <0 .000**

Alpha = 0.05, H0 can be rejected since the p-value <0.000, suggesting that the ANOVA model is significantly better than the null model.

(b) In this ANOVA model, what are the degrees of freedom for (i) Treatment, (ii) QoLCat, (iii) the

interaction term, and (iv) the error term? (4 pts)

The degrees of freedom are the following:

* Treatment = **1**
* QOLCat = **3**
* The interaction term (QOLCat \* Treatment) = **3**
* The error term = **791**

1. In this ANOVA model, what are the sum of squares for (i) Treatment, (ii) QoLCat, (iii) the

interaction term, and (iv) the error term? (4 pts)

The sum of square for the terms are the following:

* Treatment = **854.473**
* QOLCat = **1065.544**
* The interaction term (QOLCat \* Treatment) = **177.312**
* The error term = **7376.406**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter |
| Corrected Model | 2323.571a | 7 | 331.939 | 35.595 | .000 | .240 | 249.165 |
| Intercept | 179210.294 | 1 | 179210.294 | 19217.400 | .000 | .960 | 19217.400 |
| QOLCat | 1065.544 | 3 | 355.181 | 38.087 | .000 | .126 | 114.262 |
| Treatment | 854.473 | 1 | 854.473 | 91.628 | .000 | .104 | 91.628 |
| QOLCat \* Treatment | 177.312 | 3 | 59.104 | 6.338 | .000 | .023 | 19.014 |
| Error | 7376.406 | 791 | 9.325 |  |  |  |  |
| Total | 244709.000 | 799 |  |  |  |  |  |
| Corrected Total | 9699.977 | 798 |  |  |  |  |  |

2. Conduct a type III F test to examine if the interaction term is statistically significant. (5 pts)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter |
| Corrected Model | 2323.571a | 7 | 331.939 | 35.595 | .000 | .240 | 249.165 |
| Intercept | 179210.294 | 1 | 179210.294 | 19217.400 | .000 | .960 | 19217.400 |
| QOLCat | 1065.544 | 3 | 355.181 | 38.087 | .000 | .126 | 114.262 |
| Treatment | 854.473 | 1 | 854.473 | 91.628 | .000 | .104 | 91.628 |
| QOLCat \* Treatment | 177.312 | 3 | 59.104 | 6.338 | .000 | .023 | 19.014 |
| Error | 7376.406 | 791 | 9.325 |  |  |  |  |
| Total | 244709.000 | 799 |  |  |  |  |  |
| Corrected Total | 9699.977 | 798 |  |  |  |  |  |

Ho= The interaction term is **not** statistically significant.

H1= The interaction term **is** statistically significant.

**F= 6.338; p-value <0 .000**

Alpha = 0.05, H0 can be rejected since the p-value <0.000, suggesting that the interaction term (QOLCAT \* Treatment) is statistically significant.

3. How many terms should be included in the final two-way ANONA model? What are they? (3 pts)

**All three terms** should be included in the model because they are all statistically significant with p-

values <0.000. The names of these terms are QOLCat, treatment, and QOLCAT \* Treatment (the

interaction term.

4. Conduct the type III F test for two main effects (Treatment, QolCat) in the final ANOVA model. (6 pts)

Ho= The term QOLCat is **not** statistically significant.

H1= The QOLCat term **is** statistically significant.

**F= 38.087; p-value <0 .000**

Alpha = 0.05, H0 can be rejected since the p-value <0.000, suggesting that QOLCat **is** statistically significant.

Ho= The term treatment is **not** statistically significant.

H1= The term treatment **is** statistically significant.

**F= 91.628; p-value <0 .000**

Alpha = 0.05, H0 can be rejected since the p-value <0.000, suggesting that treatment **is** statistically significant.

5. Based on the final ANOVA model, what is the estimated mean for each combination of Treatment and QoLCat? (Hint: there are 8 combinations in total) (6 pts)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Estimates** | | | | | |
| Dependent Variable: Severity | | | | | |
| QOLCat | Treatment | Mean | Std. Error | 95% Confidence Interval | |
| Lower Bound | Upper Bound |
| 1 | 0 | 18.889 | .360 | 18.182 | 19.595 |
| 1 | 19.675 | .483 | 18.727 | 20.623 |
| 2 | 0 | 16.974 | .219 | 16.545 | 17.404 |
| 1 | 19.077 | .282 | 18.523 | 19.631 |
| 3 | 0 | 15.469 | .240 | 14.998 | 15.940 |
| 1 | 17.718 | .291 | 17.147 | 18.290 |
| 4 | 0 | 13.158 | .404 | 12.364 | 13.952 |
| 1 | 17.587 | .450 | 16.703 | 18.471 |

The estimated means in their respective orders are **18.889,19.675,16.974,19.077,15.469,17.718,13.158 and 17.587.**

**Question 3: Polynomial model of QOL**

Now, using severity as the outcome variable, build the polynomial model using quality of life index (QOL) as the independent variable. (11 pts)

1. Start the polynomial model with the linear and quadratic terms of QOL. (6 pts)

Y=β0+ β1X+ β2X2+E

Severity= β0+ QOLX+ QOLX2+E

Severity= 18.895+.079X+(-0.002) X2+E

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Summaryc** | | | | | | | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
| R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .472a | .223 | .222 | 3.076 | .223 | 228.316 | 1 | 797 | .000 |
| 2 | .485b | .236 | .234 | 3.052 | .013 | 13.350 | 1 | 796 | .000 |
| a. Predictors: (Constant), QOL | | | | | | | | | | |
| b. Predictors: (Constant), QOL, qol2 | | | | | | | | | | |
| c. Dependent Variable: Severity | | | | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta | VIF |
| 1 | (Constant) | 24.412 | .493 |  | 49.542 | .000 |  |
| QOL | -.134 | .009 | -.472 | -15.110 | .000 | 1.000 |
| 2 | (Constant) | 18.895 | 1.587 |  | 11.905 | .000 |  |
| QOL | .079 | .059 | .279 | 1.343 | .180 | 44.988 |
| qol2 | -.002 | .001 | -.759 | -3.654 | .000 | 44.988 |

|  |
| --- |
| a. Dependent Variable: Severity |

(a) Are there any influential points? If so, remove them in the analyses below. If not, please use the entire dataset (you do not need to worry about the outliers here). (2 pts)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Residuals Statisticsa** | | | | | |
|  | Minimum | Maximum | Mean | Std. Deviation | N |
| Cook's Distance | .000 | .076 | .001 | .005 | 799 |
| a. Dependent Variable: Severity | | | | | |

Using Cook’s distance point, the maximum value shows 0.076, which is less than 1 suggesting there are no serious influence points in the dataset.

(b). What is the Pearson’s correlation coefficient between the linear term (QOL) and the quadratic terms (QOL2)? (2 pts)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Correlations** | | | | |
|  | | Severity | QOL | qol2 |
| Pearson Correlation | Severity | 1.000 | -.472 | -.483 |
| QOL | -.472 | 1.000 | .989 |
| qol2 | -.483 | .989 | 1.000 |

The Pearson’s correlation coefficient between the linear term and quadratic term is 0.989**.**

(c)Using rule of thumb for variation inflation factor (VIF) to see if there is a collinearity issue. (2 pts)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta | VIF |
| 1 | (Constant) | 24.412 | .493 |  | 49.542 | .000 |  |
| QOL | -.134 | .009 | -.472 | -15.110 | .000 | 1.000 |
| 2 | (Constant) | 18.895 | 1.587 |  | 11.905 | .000 |  |
| QOL | .079 | .059 | .279 | 1.343 | .180 | 44.988 |
| qol2 | -.002 | .001 | -.759 | -3.654 | .000 | 44.988 |

|  |
| --- |
| a. Dependent Variable: Severity |

With VIF> 10, there is a collinearity issue between QOL and QOL2.

2.If there is a severe collinearity issue, what is the remedy? Rerun your model using this method and then test significances of the corresponding linear and quadratic terms. (5 pts)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | T | Sig. |
| B | Std. Error | Beta | VIF |
| 1 | (Constant) | 17.150 | .109 |  | 157.606 | .000 |  |
| qolcenter | -.134 | .009 | -.472 | -15.110 | .000 | 1.000 |
| 2 | (Constant) | 17.444 | .135 |  | 129.453 | .000 |  |
| qolcenter | -.133 | .009 | -.468 | -15.080 | .000 | 1.001 |
| qolcenter2 | -.002 | .001 | -.113 | -3.654 | .000 | 1.001 |

Since this is a quadratic equation, we must center the predictor/ independent variable by subtracting the mean from each observation using the equation below

𝑌=𝛽0+𝛽1𝑋− 𝑋+𝛽2𝑋− 𝑋2+𝐸

Using the new terms, the VIF for qolcenter and qolcenter2 is 1.001, which is less than ten. This solves the issue of collinearity.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 2159.975 | 1 | 2159.975 | 228.316 | .000b |
| Residual | 7540.002 | 797 | 9.460 |  |  |
| Total | 9699.977 | 798 |  |  |  |
| 2 | Regression | 2284.349 | 2 | 1142.175 | 122.602 | .000c |
| Residual | 7415.628 | 796 | 9.316 |  |  |
| Total | 9699.977 | 798 |  |  |  |
| a. Dependent Variable: Severity | | | | | | |
| b. Predictors: (Constant), qolcenter | | | | | | |
| c. Predictors: (Constant), qolcenter, qolcenter2 | | | | | | |

Ho= The linear (qolcenter) and quadratic terms (qolcenter2) **does not significantly** add to the model.

H1= The linear (qolcenter) and quadratic terms (qolcenter2) **does significantly** add to the model.

F= **122.602** and **p-value <0.000**

At alpha =0.05, H0 can be rejected since the p-value <0.000, the linear (qolcenter) and quadratic terms (qolcenter2) **does significantly** add to the model.

The final model of the data should be

Y=β0+ β1X+ β2X2+E

Severity= β0+ QOLX+ QOLX2+E

Severity= 17.444+(-0.133)X+(-0.002) X2+E

**4. Logistic regression model**

Now, using Hosp1yr (a binary 0-1 variable indicating the presence of at least one hospitalization during the 1 year follow-up period) as the outcome variable and Treatment as the independent variable (exposure) to build up logistic regression models. (33 pts)

1. Conduct a simple logistic regression using Treatment as the only independent variable. (10 pts)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | B | S.E. | Wald | df | Sig. |
| Step 1a | Treatment(1) | .349 | .147 | 5.658 | 1 | .017 |
| Intercept | .033 | .091 | 0.132 | 1 | .717 |

a. Variable(s) entered on step 1: Treatment.

𝑙ogitp=log𝑝1−𝑝=β0+β1𝑋1+⋯β𝑘𝑋𝑘

Hos1yr= .033 + (0.349)Treatment

(a)Conduct an overall F test to examine if the model is significantly better than the null model. (2 pts)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type III Test (Testing against the Null) a** | | | | |
| Source | Numerator df | Denominator df | F | Sig. |
| Intercept | 1 | 799 | 938.848 | .000 |
| Treatment | 1 | 799 | 5.750 | .017 |
| a. Dependent Variable: Hosp1yr. | | | | |

Ho= The model is **not** statistically better than the null.

H1= The model **is** statistically better than the null .

**F= 5.750; p-value =0.017**

Alpha = 0.05, H0 can be rejected since the p-value =0.017, suggesting the model (with terms) **is** statistically better than the null.

(b) What are the values of AIC and deviance (-2log likelihood) for the model? (1 pt)

|  |  |
| --- | --- |
| **Information Criteriaa** | |
| -2 Log Likelihood | 1096.330 |
| Akaike's Information Criterion (AIC) | 1100.330 |
| Schwarz's Bayesian Criterion (BIC) | 1109.700 |
| The information criteria are displayed in smaller-is-better form. | |
| 1. Dependent Variable: Hosp1yr. | |

The values of AIC and deviance (-2log likelihood) for the model are 1100and 1096 respectively.

(d)Conduct the Wald test for the effect of Treatment on hospitalization. (2 pts)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | B | S.E. | Wald | df | Sig. |
| Step 1a | Treatment(1) | .349 | .147 | 5.658 | 1 | .017 |
| Intercept | .033 | .091 | 0.132 | 1 | .717 |

a. Variable(s) entered on step 1: Treatment.

Ho= Treatment has **no** effect on hospitalization.

H1= Treatment has an effecton hospitalization.

**Wald= 5.658; p-value =0.017**

Alpha = 0.05, H0 can be rejected since the p-value =0.017, treatment has an effect on hospitalization. An individual who receives the medication is more likely to have at least one hospitalization.

(d) Express the simple logistic regression model using the resultant parameter coefficients. (2 pt)

𝑙ogitp=log𝑝1−𝑝=β0+β1𝑋1+⋯β𝑘𝑋𝑘

Hos1yr= .033 + (0.349) Treatment

(e ) What is the unadjusted odds ratio (OR) and its 95% confidence interval (CI) of hospitalization between patients treated with the experimental drug and the standard of care? (3 pts)

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk Estimate** | | | |
|  | Value | 95% Confidence Interval | |
| Lower | Upper |
| Odds Ratio for Treatment (0 / 1) | 1.417 | 1.063 | 1.889 |
| For cohort Hosp1yr = 0 | 1.212 | 1.031 | 1.425 |
| For cohort Hosp1yr = 1 | .855 | .754 | .971 |
| N of Valid Cases | 799 |  |  |

Unadjusted odds ratio is 1.417 CI 95% (1.063, 1.889)

2.Test if gender is a confounder for the relationship between hospitalization and Treatment. (5 pts)

**Variables in the Equation**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | B | S.E. | Wald | df | Sig. |
|
|
| Step 1a | Treatment(1) | .367 | .148 | 6.142 | 1 | .013 |
| Male(1) | -.573 | .144 | 15.794 | 1 | .000 |
| Intercept | .322 | .117 | 7.574 | 1 | .006 |

a. Variable(s) entered on step 1: Treatment, Male.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | B | S.E. | Wald | df | Sig. |
| Step 1a | Treatment(1) | .349 | .147 | 5.658 | 1 | .017 |
| Intercept | .033 | .091 | 0.132 | 1 | .717 |

a. Variable(s) entered on step 1: Treatment.

Ho= Gender is **no**t a cofounder.

H1= Gender is a cofounder.

Gender is **not** a cofounder because the percent change is beta is within the acceptable range, less than 10 % as measured by beta one (0.367-0.349/0.349) \*100= 5.16%

1. Test if gender is an effect modifier for the relationship between hospitalization and Treatment. (5 pts)

**Variables in the Equation**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | B | Std. Error | Wald | df | Sig. |
|
| Step 1a | Treatment(1) | .357 | .217 | 2.700 | 1 | .101 |
| Male(1) | -.581 | .183 | 10.034 | 1 | .002 |
| Male\*Treatment | .02 | .297 | 0.005 | 1 | .946 |
| Constant | .327 | .131 | 6.255 | 1 | .012 |

a. Variable(s) entered on step 1: Treatment, Male, Male\* Treatment.

Ho= Gender is **no**t an effect modifier for the relationship between hospitalization and treatment.

H1= Gender is an effect modifier for the relationship between hospitalization and treatment.

**Wald= 0.005; p-value =0.946**

Alpha = 0.05, H0 fails to be rejected since the p-value =0.946, since gender is **no**t an effect modifier for the relationship between hospitalization and treatment.

4.Conduct a multiple logistic regression model for the relationship between Treatment & hospitalization adjusting for the following covariates: age, Hispanic race, and residency (urban/rural). (5 pts)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model Fitting Information** | | | | | | |
| Model | Model Fitting Criteria | | | Likelihood Ratio Tests | | |
| AIC | BIC | -2 Log Likelihood | Chi-Square | df | Sig. |
| Intercept Only | 1090.737 | 1095.420 | 1088.737 |  |  |  |
| Final | 1024.176 | 1047.593 | 1014.176 | 74.561 | 4 | .000 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Goodness-of-Fit** | | | |
|  | Chi-Square | df | Sig. |
| Pearson | 782.596 | 761 | .286 |
| Deviance | 1001.464 | 761 | .000 |

|  |  |
| --- | --- |
| **Pseudo R-Square** | |
| Cox and Snell | .089 |
| Nagelkerke | .119 |
| McFadden | .068 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Likelihood Ratio Tests** | | | | | | |
| Effect | Model Fitting Criteria | | | Likelihood Ratio Tests | | |
| AIC of Reduced Model | BIC of Reduced Model | -2 Log Likelihood of Reduced Model | Chi-Square | df | Sig. |
| Intercept | 1024.176 | 1047.593 | 1014.176a | .000 | 0 | . |
| age1 | 1032.493 | 1051.226 | 1024.493 | 10.317 | 1 | .001 |
| Treatment | 1031.889 | 1050.622 | 1023.889 | 9.713 | 1 | .002 |
| Hispanic | 1077.713 | 1096.447 | 1069.713 | 55.537 | 1 | .000 |
| Urban | 1034.772 | 1053.505 | 1026.772 | 12.596 | 1 | .000 |
| The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0. | | | | | | |
| a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom. | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Hosp1yra | | B | Std. Error | Wald | df | Sig. |
|
| 0 | Intercept | -3.846 | .829 | 21.529 | 1 | .000 |
| age1 | .051 | .016 | 10.180 | 1 | .001 |
| [Treatment=0] | .480 | .155 | 9.584 | 1 | .002 |
| [Treatment=1] | 0b | . | . | 0 | . |
| [Hispanic=0] | 1.312 | .185 | 50.234 | 1 | .000 |
| [Hispanic=1] | 0b | . | . | 0 | . |
| [Urban=0] | -.622 | .177 | 12.302 | 1 | .000 |
| [Urban=1] | 0b | . | . | 0 | . |

(a) Conduct the partial Wald test for the treatment effect on hospitalization in this model. (2 pts)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | | | | | |
| Hosp1yra | | B | Std. Error | Wald | df | Sig. | Exp(B) | 95% Confidence Interval for Exp(B) | |
| Lower Bound | Upper Bound |
| 0 | Intercept | -3.846 | .829 | 21.529 | 1 | .000 |  |  |  |
| age1 | .051 | .016 | 10.180 | 1 | .001 | 1.052 | 1.020 | 1.086 |
| [Treatment=0] | .480 | .155 | 9.584 | 1 | .002 | 1.616 | 1.193 | 2.190 |
| [Treatment=1] | 0b | . | . | 0 | . | . | . | . |
| [Hispanic=0] | 1.312 | .185 | 50.234 | 1 | .000 | 3.714 | 2.584 | 5.338 |
| [Hispanic=1] | 0b | . | . | 0 | . | . | . | . |
| [Urban=0] | -.622 | .177 | 12.302 | 1 | .000 | .537 | .379 | .760 |
| [Urban=1] | 0b | . | . | 0 | . | . | . | . |
| a. The reference category is: 1. | | | | | | | | | |
| b. This parameter is set to zero because it is redundant. | | | | | | | | | |

Ho= Treatment has **no** effect on hospitalization in this model.

H1= Treatment has an effecton hospitalization in this model.

**Wald= 9.584; p-value =0.002**

Alpha = 0.05, H0 can be rejected since the p-value =0.002, treatment has an effect on hospitalization in this model.

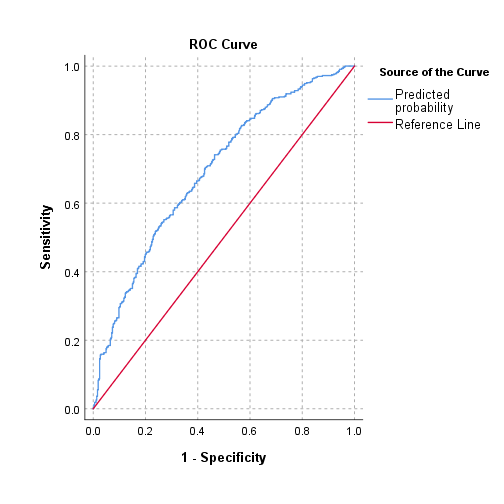
(b) What is the adjusted OR of hospitalization comparing patients treated with the new drug and the standard of care? (2 pts)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hosp1yra | | B | Std. Error | Wald | df | Sig. | Exp(B) | 95% Confidence Interval for Exp(B) | |
| Lower Bound | Upper Bound |
| 0 | Intercept | -3.846 | .829 | 21.529 | 1 | .000 |  |  |  |
| age1 | .051 | .016 | 10.180 | 1 | .001 | 1.052 | 1.020 | 1.086 |
| [Treatment=0] | .480 | .155 | 9.584 | 1 | .002 | 1.616 | 1.193 | 2.190 |
| [Treatment=1] | 0b | . | . | 0 | . | . | . | . |
| [Hispanic=0] | 1.312 | .185 | 50.234 | 1 | .000 | 3.714 | 2.584 | 5.338 |
| [Hispanic=1] | 0b | . | . | 0 | . | . | . | . |
| [Urban=0] | -.622 | .177 | 12.302 | 1 | .000 | .537 | .379 | .760 |
| [Urban=1] | 0b | . | . | 0 | . | . | . | . |
| a. The reference category is: 1. | | | | | | | | | |

The adjusted odds ratio is 1.616 CI 95% (1.193,2.190)

5. Evaluate the predictive accuracy of the model you fit in part 4 using the ROC curve analysis. (8 pts)

(a) Plot the ROC curve using the predicted probability based on the model you fit in part 4. (3 pts)

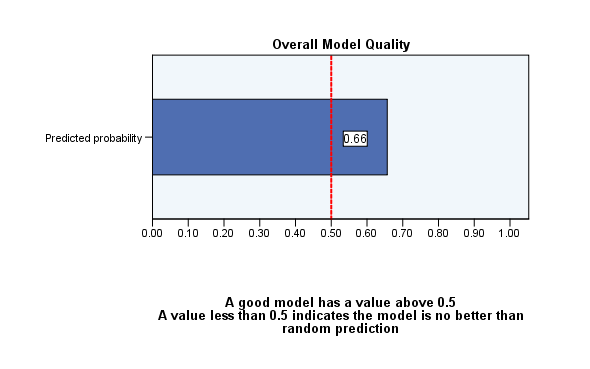


(b) What is the area under the curve (AUC)? (2 pt)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Area Under the ROC Curve** | | | | | |
| Test Result Variable(s) | Area | Std. Errora | Asymptotic Sig.b | Asymptotic 95% Confidence Interval | |
| Lower Bound | Upper Bound |
| Predicted probability | .693 | .019 | .000 | .657 | .729 |
| The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased. | | | | | |
| a. Under the nonparametric assumption | | | | | |
| b. Null hypothesis: true area = 0.5 | | | | | |

The area under the curve is 0.693.

(c)Explain the value of AUC. (3 pts)



|  |
| --- |
| Note: Use caution in interpreting this chart since it only reflects a general measure of overall model quality. The model quality can be considered "good" even if the correct prediction rate for positive responses does not meet the specified minimum probability. Use the classification table to examine correct prediction rates. |

The AUC is equal to 0.693, suggesting that it is a good model because the value is greater than 0.5. This suggest that around 69% of hospitalizations are being classified correctly based on the treatment.