**Logistic Regression – Assignment**

**1.** Using **Burns data set**to complete the following questions

In the Burns data set, 435 patients (adults ages 18-85) were grouped according to the size of the third-degree burns on their body.   
The explanatory variable is listed as the midpoint of set intervals: ln(area in square centimeters +1).   
The response in this data set is whether or not the patient survived (1 represents a survival).

**a.** Create a logistic regression model using area to estimate the probability of survival.   
**b.** Calculate the expected probabilities. Plot these probabilities against the median area.  
**c.** Interpret the model in terms of the odds ratio. Use the Wald statistic to create a 95% confidence interval for the odds ratio.  
**d.** Test H0: beta1 = 0 versus Ha: beta1 != 0 using both Wald's test and the likelihood ratio test. State your conclusion based on these tests.  
**e.** Calculate the Pearson and deviance residuals and construct normal probability plots for Pearson and deviance residuals, respectively.  
**f.** Conduct the Hosmer-Lemeshow goodness-of-fit tests to assess how well the model fits the data. Interpret the results.

**2.** Redo the above questions by **Burns\_alternative data set**, change part **(f)** to be "Conduct the Pearson chi-square and deviance goodness-of-fit tests for this model, respectively".

1. **Create a logistic regression model using area to estimate the probability of survival.**

burns. model <- glm (Survive ~ Area, family = binomial)

summary (burns. model)

Logistic Regression Model

A screenshot of a computer

Description automatically generated with medium confidence

1. **Calculate the expected Probabilities. Plot probabilities against median areas**

burns. prob <- predict (burns. model, type = "response")

burns. prob

Probability of survival Vs median area plot

Chart, line chart

Description automatically generated

There exists a negative relationship between area and probability of survival.

When area increases, the estimated probability of survival decrease.

1. Interpret the model in terms of the odds ratio. Use the Wald statistic to create a 95 % confidence interval for the odds ratio.

**Odds ratio = exp(β1)**

A picture containing text

Description automatically generated

**Interpretation in terms of odds ratio –**

**For every 1 square cm increase in area, the odds of probability of survival multiply by 0.2534978 times.**

**95% CI**

A picture containing text

Description automatically generated

odds ratio of confidence interval does not contain 1, which indicates that beta1 is not equal to 0

1. Test H0: beta1 = 0 versus Ha: beta1 != 0 using both Wald's test and the likelihood ratio test. State your conclusion based on these tests.

**Wald’s Test**

Ho: β1 Area = 0

Ha: β1 Area != 0

P-value = 2e^-16

P-value < 0.05

**Decision: Reject Ho**

**Conclusion: At 5% significance level, there is sufficient evidence to conclude that area has a significance effect on survival.**

**Likelihood Ratio Test**

Ho: β1 Area = 0

Ha: β1 Area != 0

A picture containing graphical user interface

Description automatically generated

Test statistic = G calc = 192.73

P value = 0

p-value < 0.05

**Decision: Reject H0**

**Conclusion: At 5% significance level, there is sufficient evidence to conclude that area has a significance effect on survival.**

1. Calculate the Pearson and deviance residuals and construct normal probability plots for Pearson and deviance residuals, respectively.

**Pearson Residuals**

A picture containing graphical user interface

Description automatically generated

**Normal Probability Plot**

A picture containing graphical user interface

Description automatically generated

Chart, line chart, scatter chart

Description automatically generated

Majority of the residuals follow an approximately straight line (except a very few) in the **normal probability plot, so we can assume normality.**

**Deviance Residual Plot**

Majority of the residuals follow an approximately straight line (except a very few) in the **normal probability plot, so we can assume normality.**

Chart

Description automatically generated

1. Conduct the Hosmer- Lemeshow goodness-of-fit tests to assess how well the model fits the data. Interpret the results.

**Ho: Logistic regression model adequately fits the data**

**Ha: Logistic regression model does not fit the data**

Graphical user interface, text, application

Description automatically generated

P – value = 0.8096

P – value > 0.05

**Decision: Fail to reject Ho.**

**Conclusion: There is insufficient evidence to say that the logistic regression does not adequately fits the data at 5% significance level.**

**PART 2**

1. **Create a logistic regression model using area to estimate the probability of survival.**

model. alternative <- glm (cbind (survived, dead) ~ area, family = binomial)

summary (model. alternative)

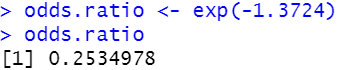
Logistic Regression Model

Text

Description automatically generated

1. Interpret the model in terms of the odds ratio. Use the Wald statistic to create a 95 % confidence interval for the odds ratio.

**Odds ratio = exp(β1)**



**Interpretation in terms of odds ratio –**

**For every 1 square cm increase in area, the odds of probability of survival multiply by 0.2534978 times.**

**95% CI**

Text

Description automatically generated

95 % CI for odds ratio = (0.1944018, 0.3305288)

odds ratio does not contain 1, which indicates beta1 is not equal to 0

1. Test H0: beta1 = 0 versus Ha: beta1 != 0 using both Wald's test and the likelihood ratio test. State your conclusion based on these tests.

**Wald’s Test**

Ho: β1 = 0

Ha: β1! = 0

Test statistic:

Text

Description automatically generated

**P-value < 0.05**

**Decision: Reject Ho**

**Conclusion: At 5% significance level, there is sufficient evidence to conclude that area has a significance effect on survival.**

**Likelihood Ratio Test**

Ho: β1 = 0

Ha: β1! = 0

Logo

Description automatically generated with low confidence

Test Statistic - G calc – 192.729

P value < 0.05

**Decision: Reject H0**

**Conclusion: At 5% significance level, there is sufficient evidence to conclude that area has a significance effect on survival.**

1. Calculate the Pearson and deviance residuals and construct normal probability plots for Pearson and deviance residuals, respectively.

**Pearson Residuals**

Text, letter

Description automatically generated with medium confidence

**Normal Probability Plot**

Majority of the residuals follow an approximately straight line (except a very few) in the **normal probability plot, so we can assume normality.**

Chart, line chart, scatter chart

Description automatically generated

**Deviance Residuals**

Text

Description automatically generated

**Residual Plot**

Majority of the residuals follow an approximately straight line (except a very few) in the **normal probability plot, so we can assume normality.**

Chart, scatter chart

Description automatically generated

1. Conduct the Pearson chi-square and deviance goodness-of-fit tests for this model, respectively".

H0: the logistic regression model provides an adequate fit to the data

Ha: the model does not adequately fit the data

**Pearson chi-square test**

A picture containing icon

Description automatically generated

df = n - (p+1) = 7

p value – 0.5362617

**Decision –- Fail to reject Ho**

**Conclusion – we have sufficient evidence to conclude that the logistic regression model provides an adequate fit to the data at 5% significance level.**

**Deviance Goodness Fit test**



df = n - (p+1) = 7

p value – 0.54147388

**Decision –- Fail to reject H0**

**Conclusion - We have insufficient evidence to conclude that the logistic regression model does not fit the data at 5% significance level.**