**37252 Regression and Linear Models**

**Lab 8: Analysis of Categorical RVs**

This lab is marked out of 20.

Please save your file in PDF format with name

**37252\_Lab8\_Surname\_FirstName**

**Due: 12 noon Wednesday 8 May 2024**

In this week’s lab we investigate field goal attempts in the NFL. The data are available in **37252\_Lab8\_ data.csv** which can be downloaded from Canvas.

The variables we consider are summarised in the table below.

|  |  |  |
| --- | --- | --- |
| **Name** | **Role** | **Description** |
|  | response | successful field goal attempt: 1 (yes), 0 (no) |
|  | predictor | game time quarter (1, 2, 3, 4) |
|  | predictor | kicking distance |

**Link between and**

We being by looking at the relationship between successful field goal attempts and game time quarter via a cross-tabulation.

> library('gmodels')

> NFLdat <- read.csv("~/2024\_37252/Labs/Lab8/37252\_Lab8\_data.csv")

> CrossTable(NFLdat$qtr, NFLdat$good, expected = T, chisq = T)

R output is displayed below.

Cell Contents

|-------------------------|

| N |

| Expected N |

| Chi-square contribution |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 1026

| NFLdat$good

NFLdat$qtr | 0 | 1 | Row Total |

-------------|-----------|-----------|-----------|

1 | 22 | 192 | 214 |

| 28.784 | 185.216 | |

| 1.599 | 0.248 | |

| 0.103 | 0.897 | 0.209 |

| 0.159 | 0.216 | |

| 0.021 | 0.187 | |

------------|-----------|-----------|-----------|

2 | 59 | 316 | 375 |

| 50.439 | 324.561 | |

| 1.453 | 0.226 | |

| 0.157 | 0.843 | 0.365 |

| 0.428 | 0.356 | |

| 0.058 | 0.308 | |

------------|-----------|-----------|-----------|

3 | 28 | 154 | 182 |

| 24.480 | 157.520 | |

| 0.506 | 0.079 | |

| 0.154 | 0.846 | 0.177 |

| 0.203 | 0.173 | |

| 0.027 | 0.150 | |

------------|-----------|-----------|-----------|

4 | 29 | 226 | 255 |

| 34.298 | 220.702 | |

| 0.818 | 0.127 | |

| 0.114 | 0.886 | 0.249 |

| 0.210 | 0.255 | |

| 0.028 | 0.220 | |

------------|-----------|-----------|-----------|

Column Total| 138 | 888 | 1026 |

| 0.135 | 0.865 | |

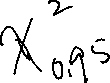
------------|-----------|-----------|-----------|

Statistics for All Table Factors

Pearson's Chi-squared test

------------------------------------------------------------

Chi^2 = 5.056845 d.f. = 3 p = 0.1676817



1. Based on the test output, what do you conclude in relation to the research question *“is field goal success linked to game time quarter”*?Write down the null and alternative hypotheses **[1 mark]**, the test statistic and p-value **[1 mark]**, the result of the test **[1 mark]** and your conclusion using a minimum of mathematical language **[1 mark]**.

**Hypotheses**

H0: and are independent (not associated with each other)

HA:  and are dependent (associated with each other) **[1 mark]**

**Test statistic and p-value**

Test stat

P-value: **[1 mark]**

**Test result**

Since we retain the null hypothesis **[1 mark]**.

**Conclusion**

Field goal success and game time quarter are not strongly associated **[1 mark]**.

1. Using the count data (not the row probabilities), calculate the odds of successful field goal attempts () in the first quarter of game time () **[1 mark]**.

**[1 mark]**

1. Using the count data (not the row probabilities), calculate the odds of successful field goal attempts () in the fourth quarter of game time () **[1 mark]**.

**[1 mark]**

1. Calculate the odds ratio of field goal success for first quarter of game time versus last quarter of game time (fourth quarter as reference) **[1 mark]**. Describe this ratio using a minimum of mathematical language **[1 mark]**.



**[1 mark]**

The odds of field goal success in first quarter is about 1.12 times the odds of field goal success in last quarter **[1 mark]**.



**Link between and**

We now want to explore whether there is any association between field goal success () and kicking distance (). As is continuous we will need to recode into quantiles (we will use quartiles for this example). Then we create a cross-tabulation between and .

> NFLdat$distanceBin <- CutQ(NFLdat$distance)

> CrossTable(NFLdat$distanceBin, NFLdat$good, expected = T, chisq = T)

Total Observations in Table: 1026

| NFLdat$good

NFLdat$distanceBin| 0 | 1 | Row Total |

-----------------|-----------|-----------|-----------|

Q1 | 5 | 258 | 263 |

| 35.374 | 227.626 | |

| 26.081 | 4.053 | |

| 0.019 | 0.981 | 0.256 |

| 0.036 | 0.291 | |

| 0.005 | 0.251 | |

-----------------|-----------|-----------|-----------|

Q2 | 19 | 254 | 273 |

| 36.719 | 236.281 | |

| 8.551 | 1.329 | |

| 0.070 | 0.930 | 0.266 |

| 0.138 | 0.286 | |

| 0.019 | 0.248 | |

-----------------|-----------|-----------|-----------|

Q3 | 29 | 206 | 235 |

| 31.608 | 203.392 | |

| 0.215 | 0.033 | |

| 0.123 | 0.877 | 0.229 |

| 0.210 | 0.232 | |

| 0.028 | 0.201 | |

----------------|-----------|-----------|-----------|

Q4 | 85 | 170 | 255 |

| 34.298 | 220.702 | |

| 74.950 | 11.648 | |

| 0.333 | 0.667 | 0.249 |

| 0.616 | 0.191 | |

| 0.083 | 0.166 | |

----------------|-----------|-----------|-----------|

Column Total | 138 | 888 | 1026 |

| 0.135 | 0.865 | |

----------------|-----------|-----------|-----------|

Statistics for All Table Factors

Pearson's Chi-squared test

------------------------------------------------------------

Chi^2 = 126.8604 d.f. = 3 p = 2.56808e-27

1. Based on the test output, what do you conclude in relation to the research question *“is field goal success linked to kicking distance quartile”*?Write down the null and alternative hypotheses **[1 mark]**, the test statistic and p-value **[1 mark]**, the result of the test **[1 mark]** and your conclusion using a minimum of mathematical language **[1 mark]**.

**Hypotheses**

H0: and are independent (not associated with each other)

HA:  and are dependent (associated with each other) **[1 mark]**

**Test statistic and p-value**

Test stat

P-value: **[1 mark]**

**Test result**

Since we reject the null hypothesis **[1 mark]**.

**Conclusion**

Field goal success and kicking distance quartile are strongly associated **[1 mark]**.

1. For each of the four kicking distance quartiles , calculate the odds of successful field goal attempts () **[4 marks]**.

**[1 mark]**

**[1 mark]**

**[1 mark]**

**[1 mark]**

1. Using as the reference group, calculate the odds ratios for the other kicking distance quartiles **[3 marks]** and comment on what this tells you about the relationship between successful field goal attempts and kicking distance quartile **[1 mark]**.

**[1 mark]**

**[1 mark]**

**[1 mark]**

The odds ratio of field goal success declines as kicking distance quartile increases **[1 mark].**