**37252 Regression and Linear Models**

**Lab 10: Multiple Logistic Regression**

This lab is marked out of 27.

Please save your file in PDF format with name

**37252\_Lab10\_Surname\_FirstName**

**Due: 12 noon Wednesday 24 May 2023**

In this week’s lab we continue our example from the last two weeks. The data are available in **37252\_Lab10\_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 |

**Multiple logistic regression**

Recall from Lab 9 we used three binary double variables in order to incorporate the 4-state variable in our model. These were coded as

Now we construct a multiple logistic regression model as response and with and as predictors.

> NFLdat <- read.csv("~/2023\_37252/Labs/Lab10/37252\_Lab10\_data.csv")

> NFLdat$qtr <- as.factor(NFLdat$qtr)

> NFLdat$qtr <- relevel(NFLdat$qtr, ref = "4")

> mod1 <- glm(good ~ qtr + distance, family = "binomial", data = NFLdat)

> summary(mod1)

> mod\_null <- glm(good ~ 1, family = "binomial", data = NFLdat)

> anova(mod\_null, mod1, test="LRT")

R output are shown below

Call:

glm(formula = good ~ qtr + distance, family = "binomial", data = NFLdat)

Deviance Residuals:

 Min 1Q Median 3Q Max

-2.9167 0.2023 0.3438 0.5701 1.2644

Coefficients:

 Estimate Std. Error z value Pr(>|z|)

(Intercept) 6.89651 0.57886 11.914 <2e-16 \*\*\*

qtr1 -0.03316 0.31661 -0.105 0.917

qtr2 -0.25645 0.26063 -0.984 0.325

qtr3 -0.36008 0.30220 -1.192 0.233

distance -0.12004 0.01240 -9.679 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

 Null deviance: 810.25 on 1025 degrees of freedom

Residual deviance: 678.47 on 1021 degrees of freedom

AIC: 688.47

Number of Fisher Scoring iterations: 6

Analysis of Deviance Table

Model 1: good ~ 1

Model 2: good ~ qtr + distance

 Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 1025 810.25

2 1021 678.47 4 131.78 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> -2\*logLik(mod1)

'log Lik.' 678.4709 (df=5)

> round(PseudoR2(mod1, c("CoxSnell", "Nagelkerke")),3)

 CoxSnell Nagelkerke

 0.121 0.221

> hoslem.test(NFLdat$good, fitted(mod1), g=10)

 Hosmer and Lemeshow goodness of fit (GOF) test

data: NFLdat$good, fitted(mod1)

X-squared = 13.892, df = 8, p-value = 0.08463

1. Write down the fitted logistic regression model in log-odds scale, odds scale and probability scale **[3 marks]**.

Let

**Log-odds scale**

**[1 mark]**

**Odds scale**

**[1 mark]**

**Probability scale**

**[1 mark]**

1. Using 0.05 significance level, test if the regression is significant. 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 a conclusion in non-mathematical language **[1 marks]**.

**Hypotheses**

H0:

HA: at least one where **[1 mark]**

**Test statistic and p-value**

Test stat

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

**Test result**

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

**Conclusion**

The regression is statistically-significant **[1 mark]**.

1. Provide interpretations of on the log-odds scale and of on the odds scale **[4 marks]**.

 is predicted log-odds of kick success from zero distance in quarter 4 **[1 mark]**.

 is predicted difference in log-odds of kick success in quarter 1 compared to quarter 4, holding distance constant **[1 mark]**.

 is predicted multiple of odds of kick success in quarter 2 compared to quarter 4, holding distance constant **[1 mark]**.

 is predicted multiple of odds of kick success for every extra yard of distance **[1 mark]**.

1. Using 0.05 significance level, test if the model provides an adequate fit to the data. 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 a conclusion in non-mathematical language **[1 mark]**.

**Hypotheses**

H0: the predicted probabilities match the observed probabilities

HA: the predicted probabilities do not match the observed probabilities **[1 mark]**

**Test statistic and p-value**

Test stat

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

**Test result**

Since we do not reject the null hypothesis **[1 mark]**.

**Conclusion**

The model does provide an adequate fit to the data **[1 mark]**.

Modify the model to allow for interaction between and .

> mod2 <- glm(good ~ qtr + distance + distance\*qtr, family = "binomial", data = NFLdat)

R output is displayed below.

Coefficients:

 Estimate Std. Error z value Pr(>|z|)

(Intercept) 7.89578 1.34888 5.854 4.81e-09 \*\*\*

qtr1 -0.95276 1.88744 -0.505 0.614

qtr2 -1.88761 1.55196 -1.216 0.224

qtr3 -0.70040 1.95129 -0.359 0.720

distance -0.14277 0.03002 -4.755 1.98e-06 \*\*\*

qtr1:distance 0.02085 0.04281 0.487 0.626

qtr2:distance 0.03707 0.03459 1.072 0.284

qtr3:distance 0.00749 0.04390 0.171 0.865

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

 Null deviance: 810.25 on 1025 degrees of freedom

Residual deviance: 676.95 on 1018 degrees of freedom

AIC: 692.95

Number of Fisher Scoring iterations: 6

> anova(mod1, mod2, test = "LRT")

Analysis of Deviance Table

Model 1: good ~ qtr + distance

Model 2: good ~ qtr + distance + distance \* qtr

 Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 1021 678.47

2 1018 676.95 3 1.5184 0.678

> confint.default(mod2)

 2.5 % 97.5 %

(Intercept) 5.25202215 10.53953365

qtr1 -4.65207908 2.74655366

qtr2 -4.92940058 1.15418630

qtr3 -4.52486599 3.12406435

distance -0.20161847 -0.08392207

qtr1:distance -0.06305836 0.10476917

qtr2:distance -0.03072679 0.10486529

qtr3:distance -0.07854476 0.09352399

1. Write down the fitted logistic regression model in log-odds scale, odds scale and probability scale for quarter 2 **[3 marks]**.

**Log-odds scale**

**[1 mark]**

**Odds scale**

**[1 mark]**

**Probability scale**

**[1 mark]**

1. Using 0.05 significance level, test if the interaction term is significant. 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 a conclusion in non-mathematical language **[1 mark]**.

**Hypotheses**

H0:

HA: at least one where **[1 mark]**

**Test statistic and p-value**

Test stat

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

**Test result**

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

**Conclusion**

No evidence of statistically-significant interaction effect between game quarter and kick distance **[1 mark]**.

1. Provide an interpretation of on the log-odds scale and on the odds-scale **[2 marks]**.

**Log-odds scale**

Note that is the difference in the intercept and is the difference in the slope coefficient

**Odds scale**

**Interpretations**

 is predicted difference in the sensitivity of log-odds to kick distance for quarter 3 compared to quarter 4 **[1 mark]**.

 is predicted multiple of the odds ratio for a unit increase in distance for quarter 3 compared to quarter 4 **[1 mark]**.

1. Using 0.05 significance level, test if . Write down the null and alternative hypotheses **[1 mark]**, the test decision with reason for your answer **[1 mark]** and a conclusion in non-mathematical language **[1 mark]**.

**Hypotheses**

H0:

HA: **[1 mark]**

**Test decision**

Reject the null as is outside the 95% CI for given by **[2 marks]**.

**Conclusion**

There is significant evidence that **[1 mark]**.