**Question 1**

Consider the space shuttle data ?shuttle in the MASS library. Consider modeling the use of the autolander as the outcome (variable nameuse). Fit a logistic regression model with autolander (variable auto) use (labeled as "auto" 1) versus not (0) as predicted by wind sign (variable wind). Give the estimated odds ratio for autolander use comparing head winds, labeled as "head" in the variable headwind (numerator) to tail winds (denominator).

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| -0.031 |  |  |  |
| 1.327 |  |  |  |
| 0.031 |  |  |  |
| 0.969 | Correct | 1.00 |  |
| Total |  | 1.00 / 1.00 |  |

**Question Explanation**

library(MASS)

data(shuttle)

## Make our own variables just for illustration

shuttle$auto <- 1 \* (shuttle$use == "auto")

shuttle$headwind <- 1 \* (shuttle$wind == "head")

fit <- glm(auto ~ headwind, data = shuttle, family = binomial)

exp(coef(fit))

## (Intercept) headwind

## 1.3273 0.9687

## Another way without redifing variables

fit <- glm(relevel(use, "noauto") ~ relevel(wind, "tail"), data = shuttle, family = binomial)

exp(coef(fit))

## (Intercept) relevel(wind, "tail")head

## 1.3273 0.9687

**Question 2**

Consider the previous problem. Give the estimated odds ratio for autolander use comparing head winds (numerator) to tail winds (denominator) adjusting for wind strength from the variable magn.

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| 1.00 |  |  |  |
| 1.485 |  |  |  |
| 0.969 | Correct | 1.00 |  |
| 0.684 |  |  |  |
| Total |  | 1.00 / 1.00 |  |

**Question Explanation**The estimate doesn't change with the inclusion of wind strength

shuttle$auto <- 1 \* (shuttle$use == "auto")

shuttle$headwind <- 1 \* (shuttle$wind == "head")

fit <- glm(auto ~ headwind + magn, data = shuttle, family = binomial)

exp(coef(fit))

## (Intercept) headwind magnMedium magnOut magnStrong

## 1.4852 0.9685 1.0000 0.6842 0.9376

## Another way without redifing variables

fit <- glm(relevel(use, "noauto") ~ relevel(wind, "tail") + magn, data = shuttle,

family = binomial)

exp(coef(fit))

## (Intercept) relevel(wind, "tail")head

## 1.4852 0.9685

## magnMedium magnOut

## 1.0000 0.6842

## magnStrong

## 0.9376

**Question 3**

If you fit a logistic regression model to a binary variable, for example use of the autolander, then fit a logistic regression model for one minus the outcome (not using the autolander) what happens to the coefficients?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| The coefficients change in a non-linear fashion. |  |  |  |
| The intercept changes sign, but the other coefficients don't. |  |  |  |
| The coefficients reverse their signs. | Correct | 1.00 |  |
| The coefficients get inverted (one over their previous value). |  |  |  |
| Total |  | 1.00 / 1.00 |  |

**Question Explanation**Remember that the coefficients are on the log scale. So changing the sign changes the numerator and denominator for the exponent.

**Question 4**

Consider the insect spray data InsectSprays. Fit a Poisson model using spray as a factor level. Report the estimated relative rate comapring spray A (numerator) to spray B (denominator).

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| 0.9457 | Correct | 1.00 |  |
| 0.321 |  |  |  |
| -0.056 |  |  |  |
| 0.136 |  |  |  |
| Total |  | 1.00 / 1.00 |  |

**Question Explanation**

fit <- glm(count ~ relevel(spray, "B"), data = InsectSprays, family = poisson)

exp(coef(fit))[2]

## relevel(spray, "B")A

## 0.9457

**Question 5**

Consider a Poisson glm with an offset, *t*. So, for example, a model of the form glm(count ~ x + offset(t), family = poisson) where x is a factor variable comparing a treatment (1) to a control (0) and t is the natural log of a monitoring time. What is impact of the coefficient for x if we fit the model glm(count ~ x + offset(t2), family = poisson) where t2 <- log(10) + t? In other words, what happens to the coefficients if we change the units of the offset variable. (Note, adding log(10) on the log scale is multiplying by 10 on the original scale.)

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| The coefficient estimate is multiplied by 10. |  |  |  |
| The coefficient is subtracted by log(10). |  |  |  |
| The coefficient estimate is divided by 10. |  |  |  |
| The coefficient estimate is unchanged | Correct | 1.00 |  |
| Total |  | 1.00 / 1.00 |  |

**Question Explanation**

Note, the coefficients are unchanged, except the intercept, which is shifted by log(10). Recall that, except the intercept, all of the coefficients are interpretted as log relative rates when holding the other variables or offset constant. Thus, a unit change in the offset would cancel out. This is not true of the intercept, which is interperted as the log rate (not relative rate) with all of the covariates set to 0.

**Question 6**

Consider the data

x <- -5:5

y <- c(5.12, 3.93, 2.67, 1.87, 0.52, 0.08, 0.93, 2.05, 2.54, 3.87, 4.97)

Using a knot point at 0, fit a linear model that looks like a hockey stick with two lines meeting at x=0. Include an intercept term, x and the knot point term. What is the estimated slope of the line after 0?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| 1.013 | Correct | 1.00 |  |
| 2.037 |  |  |  |
| -0.183 |  |  |  |
| -1.024 |  |  |  |
| Total |  | 1.00 / 1.00 |  |

**Question Explanation**

z <- (x > 0) \* x

fit <- lm(y ~ x + z)

sum(coef(fit)[2:3])

## [1] 1.013