Coursera Regression Models Quiz 4

Cheng-Han Yu

August 13, 2015

Question 1

Consider the space shuttle data `?shuttle` in the `MASS` library. Consider modeling the use of the autolander as the outcome (variable name `use`). 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).

Solution:

```
library(MASS)
data(shuttle)
str(shuttle)
## 'data.frame':    256 obs. of  7 variables:
##  $ stability: Factor w/ 2 levels "stab","xstab": 2 2 2 2 2 2 2 2 2 2 ...
##  $ error    : Factor w/ 4 levels "LX","MM","SS",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ sign     : Factor w/ 2 levels "nn","pp": 2 2 2 2 2 2 1 1 1 1 ...
##  $ wind     : Factor w/ 2 levels "head","tail": 1 1 1 2 2 2 1 1 1 2 ...
##  $ magn     : Factor w/ 4 levels "Light","Medium",..: 1 2 4 1 2 4 1 2 4 1 ...
##  $ vis      : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ use      : Factor w/ 2 levels "auto","noauto": 1 1 1 1 1 1 1 1 1 1 ...
shuttle$usebin <- as.numeric(shuttle$use == "auto") # create a binary variable
fit <- glm(usebin ~ factor(wind) - 1, family = "binomial", data = shuttle)
Coef <- coef(summary(fit))
coef.odds <- exp(c(Coef[1, 1], Coef[2, 1]))
(odds.ratio <- coef.odds[1] / coef.odds[2]) # "head" is the reference
## [1] 0.9686888
```

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.

Solution:

```
fit2 <- glm(usebin ~ factor(wind) + factor(magn) - 1, family = "binomial",
            data = shuttle)
(Coef2 <- coef(summary(fit2)))
##                          Estimate Std. Error       z value  Pr(>|z|)
## factor(wind)head    3.635093e-01  0.2840608  1.279688e+00 0.2006547
## factor(wind)tail    3.955180e-01  0.2843987  1.390717e+00 0.1643114
## factor(magn)Medium -1.009525e-15  0.3599481 -2.804642e-15 1.0000000
## factor(magn)Out    -3.795136e-01  0.3567709 -1.063746e+00 0.2874438
## factor(magn)Strong -6.441258e-02  0.3589560 -1.794442e-01 0.8575889
coef2.odds <- exp(c(Coef2[1, 1], Coef2[2, 1]))
(odds2.ratio <- coef2.odds[1] / coef2.odds[2]) # "head" is the reference
## [1] 0.9684981
```

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?

Solution:

```
fit1 <- glm(I(1 - usebin) ~ factor(wind) - 1, family = "binomial",
            data = shuttle)
summary(fit1)$coef
##                     Estimate Std. Error   z value  Pr(>|z|)
## factor(wind)head -0.2513144  0.1781742 -1.410499 0.1583925
## factor(wind)tail -0.2831263  0.1785510 -1.585689 0.1128099
```

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).

Solution:

```
data(InsectSprays)
str(InsectSprays)
## 'data.frame':    72 obs. of  2 variables:
##  $ count: num  10 7 20 14 14 12 10 23 17 20 ...
##  $ spray: Factor w/ 6 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...
fit4 <- glm(count ~ factor(spray), family = "poisson", data = InsectSprays)
(Coef4 <- coef(summary(fit4))) # "A" is the reference
##                    Estimate Std. Error    z value      Pr(>|z|)
## (Intercept)      2.67414865  0.0758098 35.2744434 1.448048e-272
## factor(spray)B  0.05588046  0.1057445  0.5284477  5.971887e-01
## factor(spray)C -1.94017947  0.2138857 -9.0711059  1.178151e-19
## factor(spray)D -1.08151786  0.1506528 -7.1788745  7.028761e-13
## factor(spray)E -1.42138568  0.1719205 -8.2676928  1.365763e-16
## factor(spray)F  0.13926207  0.1036683  1.3433422  1.791612e-01
exp(Coef4[1, 1]) / exp(Coef4[1, 1] + Coef4[2, 1])
## [1] 0.9456522
```

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.)

Solution:

```
fit5 <- glm(count ~ factor(spray) + offset(log(rep(sum(count), length(count)))),
            family = "poisson", data = InsectSprays)
```

```
fit5_10 <- glm(count ~ factor(spray) +
                   offset(log(10) + log(rep(sum(count), length(count)))),
               family = "poisson", data = InsectSprays)
coef(summary(fit5))
##                   Estimate Std. Error     z value     Pr(>|z|)
## (Intercept)    -3.85380927  0.0758098 -50.8352356 0.000000e+00
## factor(spray)B  0.05588046  0.1057445   0.5284477 5.971887e-01
## factor(spray)C -1.94017947  0.2138857  -9.0711059 1.178151e-19
## factor(spray)D -1.08151786  0.1506528  -7.1788745 7.028761e-13
## factor(spray)E -1.42138568  0.1719205  -8.2676928 1.365763e-16
## factor(spray)F  0.13926207  0.1036683   1.3433422 1.791612e-01
coef(summary(fit5_10))
##                   Estimate Std. Error     z value     Pr(>|z|)
## (Intercept)    -6.15639436  0.0758098 -81.2084191 0.000000e+00
## factor(spray)B  0.05588046  0.1057445   0.5284477 5.971887e-01
## factor(spray)C -1.94017947  0.2138857  -9.0711059 1.178151e-19
## factor(spray)D -1.08151786  0.1506528  -7.1788745 7.028761e-13
## factor(spray)E -1.42138568  0.1719205  -8.2676928 1.365763e-16
## factor(spray)F  0.13926207  0.1036683   1.3433422 1.791612e-01
```

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?

Solution:

```
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)
knots <- 0
splineTerms <- sapply(knots, function(knot) (x > knot) * (x - knot))
(xMat <- cbind(1, x, splineTerms))
##           x
##  [1,] 1 -5 0
##  [2,] 1 -4 0
##  [3,] 1 -3 0
##  [4,] 1 -2 0
##  [5,] 1 -1 0
##  [6,] 1  0 0
##  [7,] 1  1 1
##  [8,] 1  2 2
##  [9,] 1  3 3
## [10,] 1  4 4
## [11,] 1  5 5
(fit6 <- lm(y ~ xMat - 1))
##
## Call:
## lm(formula = y ~ xMat - 1)
##
## Coefficients:
##    xMat    xMatx     xMat
```

```
## -0.1826  -1.0242   2.0372
yhat <- predict(fit6)
plot(x, y, frame = FALSE, pch = 21, bg = "lightblue", cex = 2)
lines(x, yhat, col = "red", lwd = 2)
```

```
fit6$coef[2] + fit6$coef[3]
##    xMatx
## 1.013067
```