CSC 423 Homework 6

Akhil Kumar Ramasagaram

May 26, 2016

Flight Response of Geese.

a)

```
load("rdata/PACGEESE.Rdata")
glm_model <- glm(as.integer(RESPONSE) ~ ALTITUDE + LATERAL, data = PACGEESE)
summary(glm_model)
```

```
##
## Call:
## glm(formula = as.integer(RESPONSE) ~ ALTITUDE + LATERAL, data = PACGEESE)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0006  -0.2714   0.1229   0.2318   1.3468
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.840750   0.035457  51.915  < 2e-16 ***
## ALTITUDE     0.026951   0.007765   3.471 0.000568 ***
## LATERAL     -0.028716   0.001615 -17.776  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1395234)
##
##     Null deviance: 109.95  on 463  degrees of freedom
## Residual deviance:  64.32  on 461  degrees of freedom
## AIC: 407.91
##
## Number of Fisher Scoring iterations: 2
```

As we can see both the explanotory variable has p-values less than α which is 0.01 which tells they are realy significant. b) & c)

```
summary(glm(as.integer(RESPONSE) ~ ALTITUDE, data = PACGEESE))
```

```
##
## Call:
## glm(formula = as.integer(RESPONSE) ~ ALTITUDE, data = PACGEESE)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7651  -0.5688   0.3132   0.3918   0.4469
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.52962    0.03999  38.250   <2e-16 ***
## ALTITUDE     0.02577    0.01007   2.559   0.0108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2346535)
##
##     Null deviance: 109.95  on 463  degrees of freedom
## Residual deviance: 108.41  on 462  degrees of freedom
## AIC: 648.14
##
## Number of Fisher Scoring iterations: 2
```

```
summary(glm(as.integer(RESPONSE) ~  LATERAL, data = PACGEESE))
```

```
##
## Call:
## glm(formula = as.integer(RESPONSE) ~ LATERAL, data = PACGEESE)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9241  -0.2932   0.1220   0.2143   1.3782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.928725   0.025088   76.88   <2e-16 ***
## LATERAL     -0.028668   0.001635  -17.54   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1428595)
##
##     Null deviance: 109.946  on 463  degrees of freedom
## Residual deviance:  66.001  on 462  degrees of freedom
## AIC: 417.88
##
## Number of Fisher Scoring iterations: 2
```

When $\alpha = 0.01$ flight response of the geese doesn't depends on altitude of the helicopter. But it depends on Altitude based on our test.

d)

```
pred <- predict(glm_model, newdata = data.frame("ALTITUDE" = 6, "LATERAL" = 3))
exp(pred)/(1 + exp(pred))
```

```
##         1
## 0.8717264
```

9.25 Gender Discrimination In Hiring.

a)

```
load("rdata/DISCRIM.Rdata")
glm_model <- glm(HIRE ~ ., data = DISCRIM)
summary(glm_model)
```

```
##
## Call:
## glm(formula = HIRE ~ ., data = DISCRIM)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.49572  -0.24532  -0.00046   0.15100   0.72310
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.84436    0.28128  -3.002 0.006179 **
## EDUC         0.10692    0.04230   2.528 0.018482 *
## EXP          0.08863    0.02069   4.285 0.000256 ***
## GENDER       0.48473    0.13814   3.509 0.001802 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1178536)
##
##     Null deviance: 6.1071  on 27  degrees of freedom
## Residual deviance: 2.8285  on 24  degrees of freedom
## AIC: 25.272
##
## Number of Fisher Scoring iterations: 2
```

Since, all variables have p-values less than α we can conlcude that this model is adequate to make predictions.

b)

```
summary(glm(HIRE ~ GENDER, data = DISCRIM))
```

```
##
## Call:
## glm(formula = HIRE ~ GENDER, data = DISCRIM)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.4615  -0.2654  -0.2000   0.5385   0.8000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.2000     0.1202   1.664    0.108
## GENDER        0.2615     0.1763   1.483    0.150
##
```

```
## (Dispersion parameter for gaussian family taken to be 0.216568)
##
##     Null deviance: 6.1071  on 27  degrees of freedom
## Residual deviance: 5.6308  on 26  degrees of freedom
## AIC: 40.55
##
## Number of Fisher Scoring iterations: 2
```

If yo loos at the p-values which is 0.15 which is more than alpha we can conclude that it is not an important indicator for predictions.

c)

```
ex = data.frame("EDUC" = 4,"EXP" =0,"GENDER" = 1)
preds = predict(glm_model, newdata = ex, type = "link", se.fit = TRUE)
critval = 1.96
logit.point = preds$fit
logit.lower = preds$fit - (critval * preds$se.fit)
logit.upper = preds$fit + (critval * preds$se.fit)
logit.point; logit.lower; logit.upper
```

```
##          1
## 0.06804796
```

```
##          1
## -0.1806073
```

```
##         1
## 0.3167032
```

These results can be interpreted as, 95 out of 100 samples with the following observations log odds lies between these values.