Lab 13

Logistic Regression

Kristen Fitzgerald

2024-07-16

Question 1

Data were collected in an effort to relate the safety of certain vehicles to different aspects of those vehicles. This dataset has the following variables:

- `Unsafe`: binary safety designation (1 = below average (unsafe), 0 = average or above average (safe))
- `Type`: type of car (Large, Medium, Small, Sport/Utility, Sports)
- `Region`: manufacturing region (Asia, N America)
- `Weight`: integer value for car weight ranging from 1 to 6
- `Size`: size of car corresponding to Type (1 = Small/Sports, 2 = Medium, 3 = Large or Sport/Utility)

Part (a)

Build a logistic regression, predicting `Unsafe` using the variables `Region`, `Weight`, and `Size`. Treat `Weight` as a continuous variable. Treat `Region` and `Size` as categorical. Make sure to use the factor function for `Size`.

```
safety$Size <- as.factor(safety$Size)

safety.glm <- glm(Unsafe ~ Region + Weight + Size, data = safety,
                  family = binomial(link = "logit"))

summary(safety.glm)
```

```
##
## Call:
## glm(formula = Unsafe ~ Region + Weight + Size, family = binomial(link = "logit"),
##     data = safety)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)       2.7285     1.3949   1.956  0.05046 .
## RegionN America  -0.3775     0.5624  -0.671  0.50203
## Weight           -0.6678     0.4589  -1.455  0.14559
```

```
## Size2            -2.0200     0.6246  -3.234  0.00122 **
## Size3            -2.6785     0.8810  -3.040  0.00236 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 119.249  on 95  degrees of freedom
## Residual deviance:  84.004  on 91  degrees of freedom
## AIC: 94.004
##
## Number of Fisher Scoring iterations: 5
```

a. Which variables were significant at the 0.05 level?

Solution: The variables that were significant at the 0.05 level were `Size2` and `Size3`.

b. What is the concordance proportion for this model?

Solution: This model has a concordance proportion of 0.8543.

```
concordance(safety.glm)
```

```
## Call:
## concordance.lm(object = safety.glm)
##
## n= 96
## Concordance= 0.8482 se= 0.03897
## concordant discordant     tied.x     tied.y    tied.xy
##       1622        243        115       2273        307
```

Part (b)

Remove variables one at a time that have a p-value above 0.05.

```
safety.glm2 <- glm(Unsafe ~ Size, data = safety,
                   family = binomial(link = "logit"))

summary(safety.glm2)
```

```
##
## Call:
## glm(formula = Unsafe ~ Size, family = binomial(link = "logit"),
##     data = safety)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.6506     0.3561   1.827 0.067708 .
## Size2        -2.2192     0.6070  -3.656 0.000256 ***
## Size3        -3.3586     0.8125  -4.134 3.57e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 119.249  on 95  degrees of freedom
## Residual deviance:  86.629  on 93  degrees of freedom
## AIC: 92.629
##
## Number of Fisher Scoring iterations: 5
```

a. What variables are left in the model?

Solution: `Size` is the only variable left in the model.

b. What is the proportion of concordance with your final model?

Solution: The concordance proportion is 0.8182.

```
concordance(safety.glm2)
```

```
## Call:
## concordance.lm(object = safety.glm2)
##
## n= 96
## Concordance= 0.8182 se= 0.043
## concordant discordant     tied.x     tied.y    tied.xy
##       1392        132        456       1539       1041
```

c. What is the interpretation of the `Size` variable for comparing categories 1 to 3?

Solution: Cars of `Size2` decreases the expected odds of being unsafe by 89.13% compared to cars of `Size1`.

or

Cars of `Size2` increases the expected odds of being safe by 89.13% compared to cars of `Size1`.

Cars of `Size3` increases the expected odds of being safe by 96.52% when compared to cars of `Size1`.

```
100*(exp(cbind(coef(safety.glm2), confint(safety.glm2)))-1)
```

```
##                             2.5 %    97.5 %
## (Intercept)  91.66667  -2.84649 298.16962
## Size2       -89.13043 -96.98548 -66.42930
## Size3       -96.52174 -99.49411 -85.85478
```