HW07q2

Samuel Swain

2022-11-30

Question 2

Data

```
dat = expand.grid(factory=c("East", "West"), accident=c("No", "Yes"))
dat$y = c(645,1275, 28,31)
tab = matrix(dat$y, nrow=2,
dimnames=list(factory=c("East", "West"), accident=c("No", "Yes")))
```

2(a)

```
fit_2a = glm(y ~ factory + accident, poisson, dat)
fit_2a
```

```
##
## Call:  glm(formula = y ~ factory + accident, family = poisson, data = dat)
##
## Coefficients:
## (Intercept)  factoryWest  accidentYes
##       6.481        0.663       -3.483
##
## Degrees of Freedom: 3 Total (i.e. Null);  1 Residual
## Null Deviance:       2423
## Residual Deviance: 4.678     AIC: 38.43
```

2(b)

```
predict(fit_2a, newdata = data.frame(factory = factor("West"),
   accident = factor("Yes")), type = "response")
```

```
##        1
## 38.93583
```

To attain the result manually, we can use the equation bellow:

$$e^{(6.481+0.663+-3.483)} = 38.9$$

2(c)

```
fit_2c = glm(y ~ factory*accident, poisson, dat)
fit_2c
```

```
##
## Call:  glm(formula = y ~ factory * accident, family = poisson, data = dat)
##
## Coefficients:
##             (Intercept)              factoryWest              accidentYes
##                  6.4693                   0.6815                  -3.1370
## factoryWest:accidentYes
##                 -0.5797
##
## Degrees of Freedom: 3 Total (i.e. Null);  0 Residual
## Null Deviance:       2423
## Residual Deviance: 1.061e-13     AIC: 35.75
```

```
predict(fit_2c, newdata = data.frame(factory = factor("West"),
   accident = factor("Yes")), type = "response")
```

```
##  1
## 31
```

To attain the result manually, we can use the equation bellow:

$$e^{(6.4693+0.6815-3.1370-0.5797)} = 31.0$$

2(d)

We get a residual deviance of 0 because our predicted model is the saturated model. Our model can adjust to fit any row in the data frame perfectly. Thus, when we subtract the log-likelihood of the saturated model from the log-likelihood of the predicted model, we will get zero.

2(e)

```
summary(fit_2c)
```

```
##
## Call:
## glm(formula = y ~ factory * accident, family = poisson, data = dat)
##
## Deviance Residuals:
## [1]  0  0  0  0
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)              6.46925    0.03937 164.299   <2e-16 ***
## factoryWest              0.68145    0.04832  14.103   <2e-16 ***
## accidentYes             -3.13705    0.19304 -16.251   <2e-16 ***
## factoryWest:accidentYes -0.57967    0.26515  -2.186   0.0288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2.4235e+03  on 3  degrees of freedom
## Residual deviance: 1.0614e-13  on 0  degrees of freedom
## AIC: 35.749
##
## Number of Fisher Scoring iterations: 3
```

The z-value for the interaction term is -2.186. This is less than -1.645. We can reject the null that $\beta_{factory*accident} = 0$ and conclude $\beta_{factory*accident} \neq 0$.

2(f)

The result from question 2 part e tells us the west factory on average is less likely to have accidents.

2(g)

```
d = drop1(fit_2c, test="Chisq")
d
```

```
## Single term deletions
##
## Model:
## y ~ factory * accident
##                  Df Deviance    AIC   LRT Pr(>Chi)
## <none>                 0.000 35.749
## factory:accident  1    4.678 38.427 4.678  0.03055 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
p_val_2g = d$`Pr(>Chi)`[2]
chi_sq_2g = qchisq(p_val_2g, 1, lower.tail = FALSE)
```

Using the likelihood ratio test to evaluate the interaction, we get a chi-squared value of 4.6779804 and a p-value of 0.0305516.

2(h)

```
east_accident_0 = predict(fit_2c, newdata = data.frame(factory = factor("East"), accident = factor("No")), t
ype = "link")
east_accident_1 = predict(fit_2c, newdata = data.frame(factory = factor("East"), accident = factor("Yes")),
 type = "link")

west_accident_0 = predict(fit_2c, newdata = data.frame(factory = factor("West"), accident = factor("No")), t
ype = "link")
west_accident_1 = predict(fit_2c, newdata = data.frame(factory = factor("West"), accident = factor("Yes")),
 type = "link")

log_odds_east_2h = east_accident_1 - east_accident_0
log_odds_west_2h = west_accident_1 - west_accident_0
```

- The log odds of an accident in the east is -3.1370458
- The log odds of an accident in the west is -3.7167143

2(i)

```
c = log_odds_east_2h
d = log_odds_west_2h - log_odds_east_2h
cat("c: ", c, "\n", "d: ", d)
```

```
## c:  -3.137046
##  d:  -0.5796684
```

$$log(\frac{\pi_{1|i}}{1-\pi_{1|i}}) = log(\frac{\pi_{1|i}}{\pi_{0|i}})$$
$$= log(m_{i0}) + log(\frac{m_{i1}}{m_{i0}}) * west$$
$$= -3.1370 \; - \; 0.5797 * west$$

2(j)

```
fit_2j = glm(accident ~ factory, binomial, dat, weights = y)
summary(fit_2j)
```

```
##
## Call:
## glm(formula = accident ~ factory, family = binomial, data = dat,
##     weights = y)
##
## Deviance Residuals:
##      1       2       3       4
## -7.404  -7.827  13.344  15.229
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.1370     0.1930 -16.251   <2e-16 ***
## factoryWest  -0.5797     0.2651  -2.186   0.0288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 530.73  on 3  degrees of freedom
## Residual deviance: 526.06  on 2  degrees of freedom
## AIC: 530.06
##
## Number of Fisher Scoring iterations: 6
```

2(k)

I would expect the estimated logistic regression to be just the intercept: $log(\frac{\pi_{1|i}}{1-\pi_{1|i}}) = -3.1370$