CatData HW6

Matthew Vanaman

05-15-2019

A4.31 (5 points)

A4.16-A4.17, A5.4 (10 points)

4.3

(a): In the linear probability model, the coefficient is treated the same way as it is in linear regression. That is, it represents a slope such that for every unit increase x (decade, in this case), there is a corresponding unit change in y (the probability of pitching a complete game, in this case, the change of which = -0.0694).

(b):

$$\hat{\pi} = 0.7578 - 0.0694(12), \\ = -0.075.$$

Percentages can't be negative! So not possible.

(c):

$$\hat{\pi} = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}.$$

$$\hat{\pi} = \frac{e^{1.148-0.315(12)}}{1+e^{1.148-0.315(12)}}, \\ = 0.0671071.$$

This is much more plausible.

4.5

```
Call:
glm(formula = TD ~ Temp, family = binomial, data = shuttle)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0611  -0.7613  -0.3783   0.4524   2.2175

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  15.0429     7.3786   2.039   0.0415 *
Temp         -0.2322     0.1082  -2.145   0.0320 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 28.267  on 22  degrees of freedom
```

```
Residual deviance: 20.315  on 21  degrees of freedom
AIC: 24.315

Number of Fisher Scoring iterations: 5
```

(a):

The model comes out to be 15.04 + (-0.23)x. When temperature is 0, the log odds are 15.04. The log odds decrease by 0.23 with every unit increase in temperature.

(b):

$$\hat{\pi} = \frac{e^{15.04-0.23(31)}}{1+e^{15.04-0.23(31)}}.$$
$$= 0.9996331.$$

(c):

You need to solve for x:

$$0.50 = \frac{e^{15.04-0.23(x)}}{1+e^{15.04-0.23(x)}}.$$

The temperature comes out to be 64.8. To get the linear approximation (i.e. rate of change), take the derivative of the logistic function above (with respect to x) to get its probability density function:

$$f'(x) = \frac{\beta e^{\alpha+\beta x}}{(1+e^{\alpha-\beta(x)})^2}$$
$$= \frac{15.04(e^{15.04-0.23(64.8)})}{(1+e^{15.04-0.23(64.8)})^2}$$
$$= -0.0580407.$$

At 64.8 degrees, for every unit increase in temperature, there is a 0.058 decrease in probability.

(d):

Since the coefficient is the log odds ratio, you exponentiate it to get the odds. Exp(-0.2321627) = 0.7928171. This seems kind of difficult to interpret though, so it might be easier to take $1/OR = 1/0.79 = 1.27$ and say that for every unit increase in temperature, the odds of the o-rings *working correctly* increase by 1.27 - that is, every unit increase in temperature means there is 27% greater odds of the orings working the way they're supposed to (or 27% greater probability if we think events are rare enough to warrent using the odds ratio as an approximation of the risk ratio).

(e):

The Wald test is:

$$z^2 = \left(\frac{\beta}{SE}\right)^2,$$
$$= \left(\frac{-0.2321627}{0.1082}\right)^2,$$
$$= 4.6039476.$$

z^2 approximates the chi-square; the p-value is 0.0318984 with one degree of freedom.

The likelihood ratio test is:

$$-2(\ell_0 - \ell_1),$$
$$-2(-14.1335764 - 10.1575963),$$
$$7.95196.$$

χ^2 p-value at 1 degree of freedom is 0.0048035. Temperature is probably important here.

A4.15

(a)

The CMH test is:

$$\frac{[\sum_k (n_{11k} - \mu_{11k})]^2}{\sum_k \mathrm{Var}(n_{11k})},$$

where:

$$\mu_{11k} = \frac{n_{1+k} n_{+1k}}{n_{++k}},$$

and:

$$\mathrm{Var}_{11k} = \frac{n_{1+k} n_{2+k} n_{+1k} n_{+2k}}{n_{++k}^2 (n_{++k} - 1)}.$$

The CMH test statistic comes out to be 7.8149199 with 1 df, and a p-value of 0.0051817. We reject the null hypothesis that merit pay decisions are independent of race.