CatData HW 4

Matthew Vanaman

03/18/19

3.15 (do not verify for 3.15d)

3.9

(a)

Answer: $\text{logit}[\hat{P}(Y = 1)] = 0.0532x - 3.5561$. In this example, x is income, Y is possession of travel card (0 = does not possess, 1 = does possess).

(b)

Answer: $\hat{\beta}$ means, in this case, that as income increases, the probability of possessing a travel card increases (because $\hat{\beta}$ is positive).

(c)

Answer: First, we know that $\hat{\pi}$ denotes $P(Y = 1)$, and that the logit function is $\log\left[\dfrac{P(Y=1)}{1 - P(Y=1)}\right]$. When you substitute 0.50 in for $P(Y = 1)$, you get 0:

```
log(0.50 / (1 - 0.50))
```

```
[1] 0
```

To show when the etsimated π is 0.5, we just need to solve for x:

```
solveset(Eq(-3.556 + 0.0532*x)), 0.50
```

```
({66.8421052631579}, 0.5)
```

Therefore, when x = 66.84, $\hat{\pi} = 0.50$. According to the model, there is a 50/50 shot of owning a travel card for those who make around 66.86 million lira.

```
italianData <- as.data.frame(matrix(c(24, 27, 28, 29, 30, 31, 32, 33,
                                      34, 35, 38, 39, 40, 41, 42, 45,
                                      48, 49, 50, 52, 59, 60, 65, 68,
                                      70, 79, 80, 84, 94, 120, 130,

                                      1, 1, 5, 3, 9, 5, 8, 1,
                                      7, 1, 3, 2, 5, 2, 2, 1,
                                      1, 1, 10, 1, 1, 5, 6, 3,
                                      5, 1, 1, 1, 1, 6, 1,

                                      0, 0, 2, 0, 1, 1, 0, 0,
                                      1, 1, 1, 0, 0, 0, 0, 1,
                                      0, 0, 2, 0, 0, 2, 6, 3,
```

```r
                                    3, 0, 0, 0, 0, 6, 1), ncol = 3))
colnames(italianData) <- c("Inc", 'No.Cases', 'CreditCards')
knitr::kable(italianData)
```

Inc	No.Cases	CreditCards
24	1	0
27	1	0
28	5	2
29	3	0
30	9	1
31	5	1
32	8	0
33	1	0
34	7	1
35	1	1
38	3	1
39	2	0
40	5	0
41	2	0
42	2	0
45	1	1
48	1	0
49	1	0
50	10	2
52	1	0
59	1	0
60	5	2
65	6	6
68	3	3
70	5	3
79	1	0
80	1	0
84	1	0
94	1	0
120	6	6
130	1	1

3.15

(a)

Answer: The intercept in this case is whites (white = 0). To calculate the estimate population mean, just exponentiate the y-intercept by itself (or more accurately, the effect of white when black = 0):

```r
exp(-2.38 + (1.733 * 0))
```

```
[1] 0.09255058
```

To get the estimated population means for black, add black to the model:

```r
exp(-2.38 + (1.733 * 1))
```

```
[1] 0.5236143
```

(b)

The wald confidence interval for poisson data is $\hat{\beta} \pm z_{\alpha/2}(SE)$. $\hat{\beta}$ is the log of the ratio between the two means. Therefore:

```r
# CI Lower:
log(0.522/0.092) - 1.96 * 0.147
```

```
[1] 1.447759
```

```r
# CI Upper:
log(0.522/0.092) + 1.96 * 0.147
```

```
[1] 2.023999
```

(c)

The negative binomial is model is more believable because overdispersion is likely biasing our Poisson coefficients. We know we have overdispersion because in Poisson models, the mean should equal the variance, but in our case the variances for both blacks (1.150) and whites (0.155) are much, much greater taen their group means (0.522 and 0.092, respectively).

(d)

$\hat{D}$, the dispersion index of the negative binomial model, is 4.94, which indicates a huge discrepancy between the observed dispersion and the dispersion assumed by the Poisson model (i.e. $D = 0$). If $\hat{D}$ had turned out to be close to zero, then Poisson would be appropriate.

Question 3, AD Data

```r
q3Data <- as.data.frame(matrix(c(730,130,860,
                     100, 40, 140,
                     830,170,1000), ncol = 3))
colnames(q3Data) <- c("NLD", 'LD', 'Margin')
rownames(q3Data) <- c('NAD', 'AD','Margin')
knitr::kable(q3Data)
```

	NLD	LD	Margin
NAD	730	100	830
AD	130	40	170
Margin	860	140	1000