HW4 self

Chong Zhang

November 9, 2018

A. Buying a Car

a. What is the formula of the estimated probability for a family to buy a car given its income and the age of the car the family owns?

$$Ln(\frac{p}{1-p}) = -4.739 + 0.0677 * income + 0.5986 * age$$

$$p = \frac{e^{-4.739+0.0677*income+0.5986*age}}{1+e^{-4.739+0.0677*income+0.5986*age}}$$

b. Identify the coefficient for the regressor age and interpret.

income: 0.06773 age: 0.59863

c. Perform a hypothesis testing procedure for the statistical significance of the regression coefficients corresponding to the income factor. Interpret at alpha = 0.05.

```
z = 0.06773/0.02806
2*pnorm(z,lower.tail = F)
```

```
## [1] 0.01578902
```

d. Estimate the probability that a family with $70,000.00 an income whose oldest car is 4 years old buys a new car in the following year. **0.916**

e. What can you say about the reliability of the statistical significance of the regression coefficients for this data example? Provide supporting arguments for your answer. **Both intercept and income are significant, however age is not.**

B. Simulation Example

a. Analyze the data using Poisson regression; obtain both point estimates and confidence intervals for the regression coefficients. Provide the R output and interpret in relation to the true parameters for the model you generated data from.

```
x=1:10/10
mu = exp(3+0.2*x)
y = rpois(10, mu)
z = as.data.frame(cbind(y,x))
method = glm(formula = y~x, data = z, family = 'poisson')
fitted(method)
```

```
##        1        2        3        4        5        6        7        8
## 20.52529 20.75700 20.99132 21.22830 21.46794 21.71029 21.95538 22.20323
##        9       10
## 22.45388 22.70736
```

```
summary(method)
```

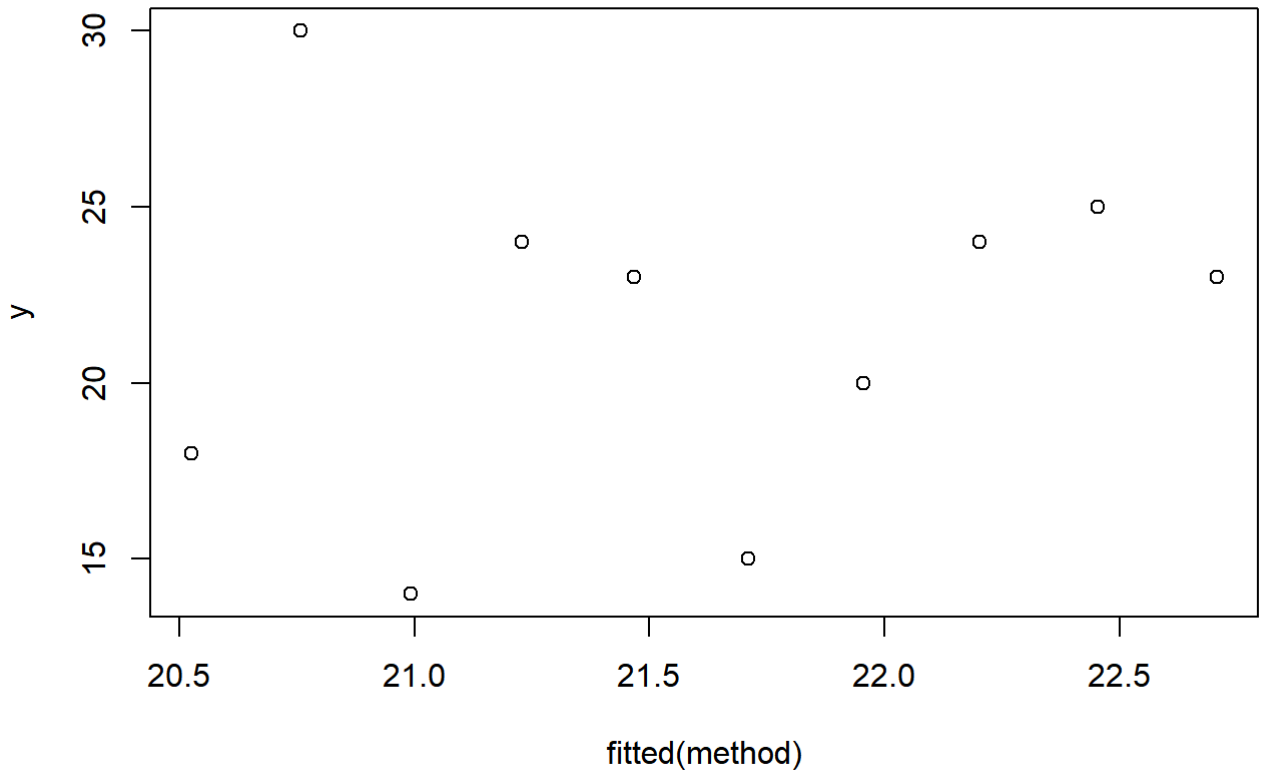
```
##
## Call:
## glm(formula = y ~ x, family = "poisson", data = z)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6252  -0.5330   0.1941   0.4898   1.9007
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.0104     0.1490  20.208   <2e-16 ***
## x             0.1123     0.2370   0.474    0.636
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 10.1885  on 9  degrees of freedom
## Residual deviance:  9.9641  on 8  degrees of freedom
## AIC: 62.904
##
## Number of Fisher Scoring iterations: 4
```

```
confint(method)
```

```
## Waiting for profiling to be done...
```

```
##                 2.5 %    97.5 %
## (Intercept)  2.7103903 3.2947774
## x           -0.3522009 0.5775319
```

```
plot(fitted(method), y)
```



b. Repeat the data generation process for 100 times. You can use the for loop in R for repeating the data generation. Set a seed to 1 before the loop in order to reproduce random numbers.

```
set.seed(1)
intcover = 0
xcover=0

for (i in 1:100){
  N=10
  x=1:N/N
  mu = exp(3+0.2*x)
  y=rpois(N,mu)
  model = glm(y~x,family = 'poisson')
  CI = confint(model,level = 0.95)
  intcover = intcover +(CI[1,1]<3 &CI[1,2]>3)
  xcover = xcover+(CI[2,1]<0.2 & CI[2,2]>0.2)
}
intcover
```

```
## [1] 97
```

```
xcover
```

```
## [1] 96
```

How many times the estimated confidence intervals include the true parameters? Interpret. **97 times are included.**

c. Generate data as in part (a) except that this time you will generate with for . Repeat 100 times. How many times the estimated confidence intervals include the true parameters? Interpret.

```
set.seed(1)
intcover = 0
xcover=0

for (i in 1:100){
  N=1000
  x=1:N/N
  mu = exp(3+0.2*x)
  y=rpois(N,mu)
  model = glm(y~x,family = 'poisson')
  CI = confint(model,level = 0.95)
  intcover = intcover +(CI[1,1]<3 &CI[1,2]>3)
  xcover = xcover+(CI[2,1]<0.2 & CI[2,2]>0.2)
}
intcover
```

```
## [1] 93
```

```
xcover
```

```
## [1] 95
```

93 time are included.

d. Are the answers for part (b) and (c) different? Why? They are differnt. Because the sample number of part b is not large enough as part c. Possion regression requires large sample size to make acurate statistic inferrence.