Econ 7026 Econometric Theory (I) Fall 2021
National Taiwan University Chih-Sheng Hsieh

Econ 7026 Problem Set 3

- Problem set 3 is due on 2021/11/29 (collected on NTU Cool).
- The program codes should be attached in your answer.
- You are encouraged to form groups for discussion, but should submit an individual answer.
- The answers can be hand-written or machine-typed and you can submit it electronically as **PDF** files.
- Your answers to questions II 5. and 6. are limited to two pages long (excluding Tables).

I. Let $x_1, \cdots, x_n$ be iid Bernoulli(p).

1. Write down the likelihood function of p
2. Derive the MLE estimator of p, $\hat{p}_{mle}$
3. Derive the variance of MLE, $var(\hat{p}_{mle})$.

II. The random variable Y comes from i.i.d. Poisson distribution, i.e.,

$$P(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

1. Obtain the log-likelihood function for a random sample of n observations, i.e., $(y_1, y_2, \cdots, y_n)$.
2. Obtain the maximum likelihood estimator of λ. → 我写∵無法手算，∴用 numerical method than stick to mle
3. Explain how to estimate the variance of the MLE estimator for λ.
4. → Now 全部變成什 β
   If we formulate λ by $\exp(X\beta)$, then we obtain $\log(E(Y|X)) = X\beta$, the so-called Poisson regression. Please write down the log-likelihood function of β and discuss why a numerical method such as Newton method is needed to obtain the MLE of β?
5. Many colleges and universities in U.S. have "Greek letter organizations," i.e., fraternities (兄弟會) and sororities (姐妹會). These organizations are often known for their social events, especially parties with alcohol. We have unique data from one U.S. university in which fraternity and sorority members were surveyed. Please use the data "collegehookup" downloaded from NTU Cool to study how the number of hookups is related to student's characteristics, i.e., provide estimation result and interpretation. You should use the Poission regression as the dependent variable is count number. Try to code the maximum likelihood estimator by yourself. Variable descriptions are in the Table below.

6. Following the previous question, please use the same data to study if there exist peer effects from the same fraternity (sorority) house on hookup.

↳同一兄弟會,行為會比較像?

非10人 → 隨配!!

用其他9人的平均 hookup

當作 peer effect variable

Variable	Definition
Y ← hookup sum	Number of hookups in this semester → 在一起幾個
greek group	Group index of greek house → 哪一個兄弟會
greek house	Index of greek house
Gender	Male=1 & Female=0
Age	Age
Hisp	Dummy for Hispanic
Black	Dummy for Black
Asian	Dummy for Asian
Native	Dummy for Native American
Mideast	Dummy for Mideast
BMI	Body BMI index
BMI2	Body BMI index square
college dad	Dummy of whether father has a college degree or not
college mom	Dummy of whether mother has a college degree or not
hookup highschool	Number of hookups in high school
ParentsDivorce	Dummy of whether parents are divorced or not

I. Let $x_1, \cdots, x_n$ be iid Bernoulli(p).

1. Write down the likelihood function of p
2. Derive the MLE estimator of p, $\hat{p}_{mle}$
3. Derive the variance of MLE, $var(\hat{p}_{mle})$.

1. $f(x|p) = p^{x} \cdot (1-p)^{1-x}$

$\Rightarrow L(p|x) = \prod_{i=1}^{n} p^{x_i} \cdot (1-p)^{1-x_i}$ #

2. $\ln L(p|x_i) = \sum_{i=1}^{n} x_i \ln p + \sum_{i=1}^{n} (1-x_i) \ln(1-p)$

FOC : $\dfrac{\sum x_i}{\hat{p}} - \dfrac{\sum(1-x_i)}{1-\hat{p}} = 0$

$\Rightarrow (1-\hat{p}) \sum x_i = \hat{p} \sum (1-x_i)$

$\Rightarrow \hat{p}_{MLE} = \dfrac{\sum(x_i)}{\sum(x_i) + \sum(1-x_i)} = \dfrac{\sum x_i}{n}$ #

3. $Var(\hat{p}) = Var\left(\dfrac{\sum x_i}{n}\right) = \dfrac{1}{n^2} \sum Var(x_i) = \dfrac{n p(1-p)}{n^2} = \dfrac{p(1-p)}{n}$ #

II. The random variable Y comes from i.i.d. Poisson distribution, i.e.,

$$P(Y=y)=\frac{\lambda^y e^{-\lambda}}{y!}$$

1. Obtain the log-likelihood function for a random sample of n observations, i.e., $(y_1, y_2, \cdots, y_n)$.
2. Obtain the maximum likelihood estimator of λ.
3. Explain how to estimate the variance of the MLE estimator for λ.
4. If we formulate λ by $\exp(X\beta)$, then we obtain $\log(E(Y|X)) = X\beta$, the so-called Poisson regression. Please write down the log-likelihood function of β and discuss why a numerical method such as Newton method is needed to obtain the MLE of β?

1. $$\ln \prod_{i=1}^{n} L(\lambda|y) = \ln \prod_{i=1}^{n} \frac{\lambda^{y_i} \cdot e^{-\lambda}}{y_i!} = \sum_{i=1}^{n} \ln \frac{\lambda^{y_i} \cdot e^{-\lambda}}{y_i!}$$
$$= \sum_{i=1}^{n} \left[y_i \ln(\lambda) - \lambda - \ln(y_i!) \right]$$
$$= \sum_{i=1}^{n} y_i \ln(\lambda) - n\lambda - \sum_{i=1}^{n} \ln(y_i!)$$ ※

2. FOC to λ: $\frac{\sum y_i}{\lambda} - n = 0 \Rightarrow \hat{\lambda} = \frac{\sum y_i}{n}$ ※

3. $Var(\hat{\lambda}) = \frac{1}{n^2} \sum Var(y_i) = \frac{1}{n^2} \cdot n\lambda = \frac{\lambda}{n}$ ※

4. $$\ln \prod_{i=1}^{n} L(\beta|X,Y) = \ln \prod_{i=1}^{n} \frac{e^{y_i \beta x_i} \cdot e^{-e^{x_i\beta}}}{y!} = \sum_{i=1}^{n} \left[y_i \beta x_i - e^{x_i\beta} - \ln(y_i!) \right]$$

FOC to β: $\sum x_i y_i - \sum e^{x_i\beta} \cdot x_i = 0$

Because this is an non-linear func., so we cannot find the explicit form of $\hat{\beta}_{ML}$, and thus we need to use the numerical method to approximate the value. ※

Econometrics_HW3

Lo, Wei-Chun r10323027

11/25/2021

Question II.

5.

```r
rm(list=ls())
library(tidyverse)
library(readxl)
set.seed(5678)

student_sample <- read_excel("/Users/alexlo/Desktop/collegehookup.xlsx")
student_sample <- select(student_sample, hookup_sum, everything())

#m1 <- glm(hookup_sum ~ Gender + Age + Hisp + Black + Asian + Native +
            #Mideast +BMI + BMI2 + college_dad +college_mom + hookup_highschool +
            #ParentsDivorce + Siblings, family="poisson", data=student_sample)
#summary(m1)

m2 <- glm(hookup_sum ~ Hisp + Black + BMI2 + college_dad +college_mom +
            hookup_highschool + ParentsDivorce + Siblings,
          family="poisson", data=student_sample)
summary(m2)$coefficients
```

```
##                      Estimate   Std. Error   z value     Pr(>|z|)
## (Intercept)       -0.91557365 0.2953353151 -3.100116 1.934450e-03
## Hisp              -0.90615158 0.1546341644 -5.859970 4.629505e-09
## Black             -1.20053125 0.3811374439 -3.149864 1.633463e-03
## BMI2               0.00114419 0.0002532424  4.518161 6.237905e-06
## college_dad        0.66501931 0.1558594693  4.266788 1.983074e-05
## college_mom       -0.64850270 0.1177516658 -5.507376 3.642217e-08
## hookup_highschool  0.05434702 0.0031685791 17.151857 6.087517e-66
## ParentsDivorce     0.64659476 0.1011388984  6.393136 1.625172e-10
## Siblings           0.19212042 0.0314672805  6.105403 1.025418e-09
```

First, I used 'hookup_sum' as the dependent variables and every other variables as the independent variables to run the Poisson regression model. Then, I found 'Gender', 'Age', 'Asian', 'Native', and 'Mideast' are statistically insignificant, so I removed these variables and run the regression model once again.

Now, all the independent variables are statistically significant, and moreover, 'Black' and 'Hisp' has the strongest effect on 'hookup_sum'. By the way, among all the removed variables, three of them are dummy variables about different race. Therefore, maybe culture would be an interesting and potential approach to conduct the follow-up research.

6.

```
student_sample2 <- student_sample %>%
  group_by(greek_group) %>%
  mutate(peer_effects = (sum(hookup_sum) - hookup_sum)/length(greek_group)) %>%
  select(hookup_sum, peer_effects, everything())

m3 <- glm(hookup_sum ~ peer_effects + Hisp + Black + BMI2 + college_dad +
            college_mom + hookup_highschool + ParentsDivorce + Siblings,
          family="poisson", data=student_sample2)
summary(m3)$coefficients
```

```
##                       Estimate   Std. Error    z value     Pr(>|z|)
## (Intercept)       -0.987518151 0.3045554977 -3.2424900 1.184901e-03
## peer_effects       0.020743153 0.0222219483  0.9334534 3.505859e-01
## Hisp              -0.901475035 0.1549285285 -5.8186510 5.932446e-09
## Black             -1.226181806 0.3824132201 -3.2064315 1.343923e-03
## BMI2               0.001125904 0.0002536811  4.4382632 9.068770e-06
## college_dad        0.669261898 0.1562370042  4.2836324 1.838666e-05
## college_mom       -0.645873162 0.1178645018 -5.4797938 4.258219e-08
## hookup_highschool  0.054711389 0.0031990665 17.1022983 1.426656e-65
## ParentsDivorce     0.638763357 0.1014039266  6.2991975 2.991906e-10
## Siblings           0.196358873 0.0317724122  6.1801689 6.403306e-10
```

To test the peer effects from the same fraternity house on hookup, I add a new independent variable 'peer_effects', which is the average number of hookup in the same fraternity house after reducing the observation's own hookup number.

However, the regression result shows that this variable is statistically insignificant, which implies peer effects has no effect on the number of hookups in the semester.