STAT 426 Assignment 2

Due Tuesday, September 7, 11:59 pm.

Submit through Moodle.

Name: Brianna Diaz

Netid: bdiaz22

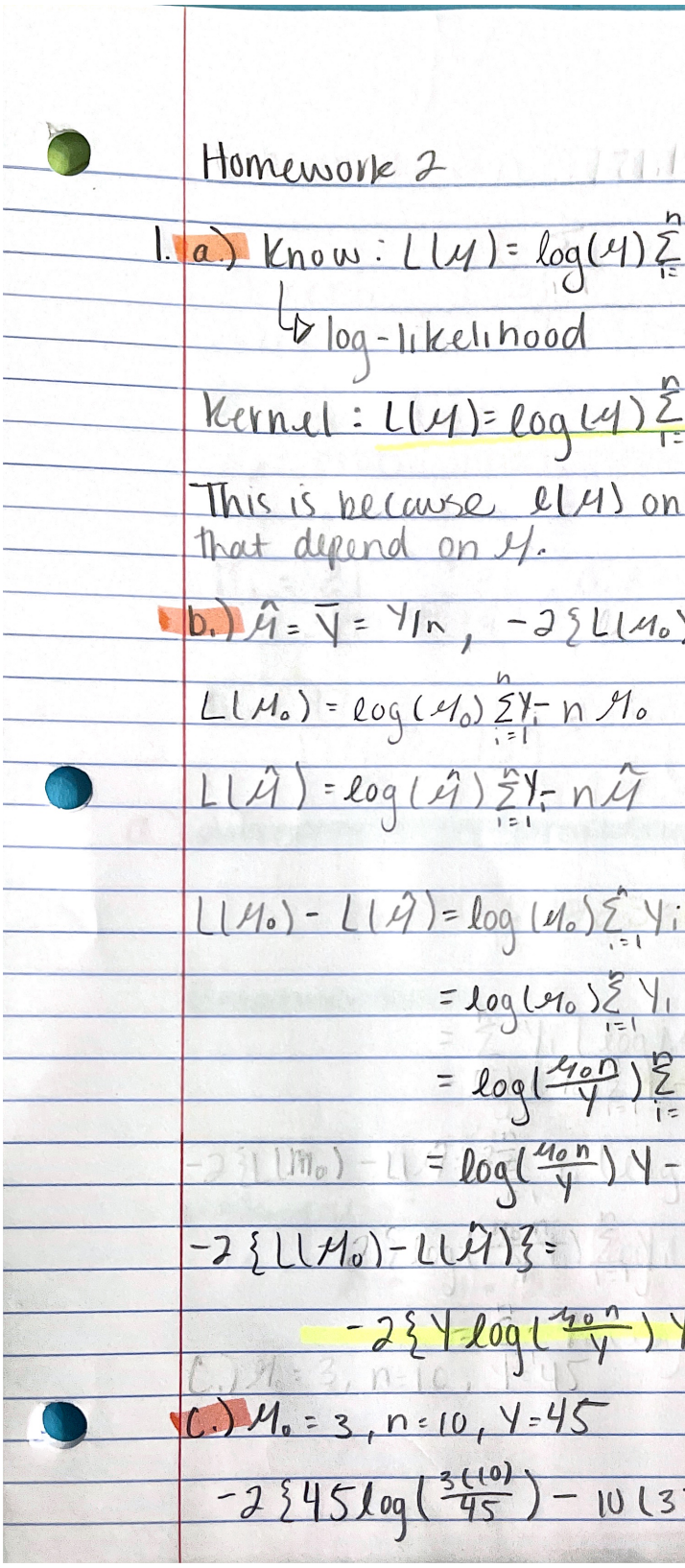
Submit your work both as an R markdown (*.Rmd) document and as a pdf, along with any files needed to run the code. Embed your answers to each problem in the document below after the question statement. If you have hand-written work, please scan or take pictures of it and include in a pdf file, ideally combined with your pdf output file from R Markdown. Be sure to show your work.

Problem 1 (6 pts)

Let $Y_1, Y_2, \ldots, Y_n$ be independent observations from the Poisson distribution with mean μ, which is unknown. Recall that if $Y = \sum_{i=1}^{n} Y_i$, then $Y \sim \text{Poisson}(n\mu)$.

(a) Show that the kernel of the log-likelihood, $L(\mu)$, depends only on Y, μ and n.

Answer: All work for this problem on separate sheet.



(b) Recall from Assignment 1 that the maximum likelihood estimator is $\hat{\mu} = \bar{Y} = Y/n$. Obtain the mathematical form of the log-likelihood ratio test statistic $-2\{L(\mu_0) - L(\hat{\mu})\}$,

which we would use to test the null hypothesis, $H_0 : \mu = \mu_0$. Simplify the expression to show that it is a function of Y, n, and μ_0 only.

Answer:

(c) Suppose in a sample of $n = 10$ we observe $Y = 45$. Find the p-value for the log-likelihood ratio test of $H_0 : \mu = 3$ versus $H_A : \mu \neq 3$ using the chi-square approximation for the null distribution of the test statistic.

Answer:

```
pchisq(-14.15, df = 1, log = FALSE)
```

```
## [1] 0
```

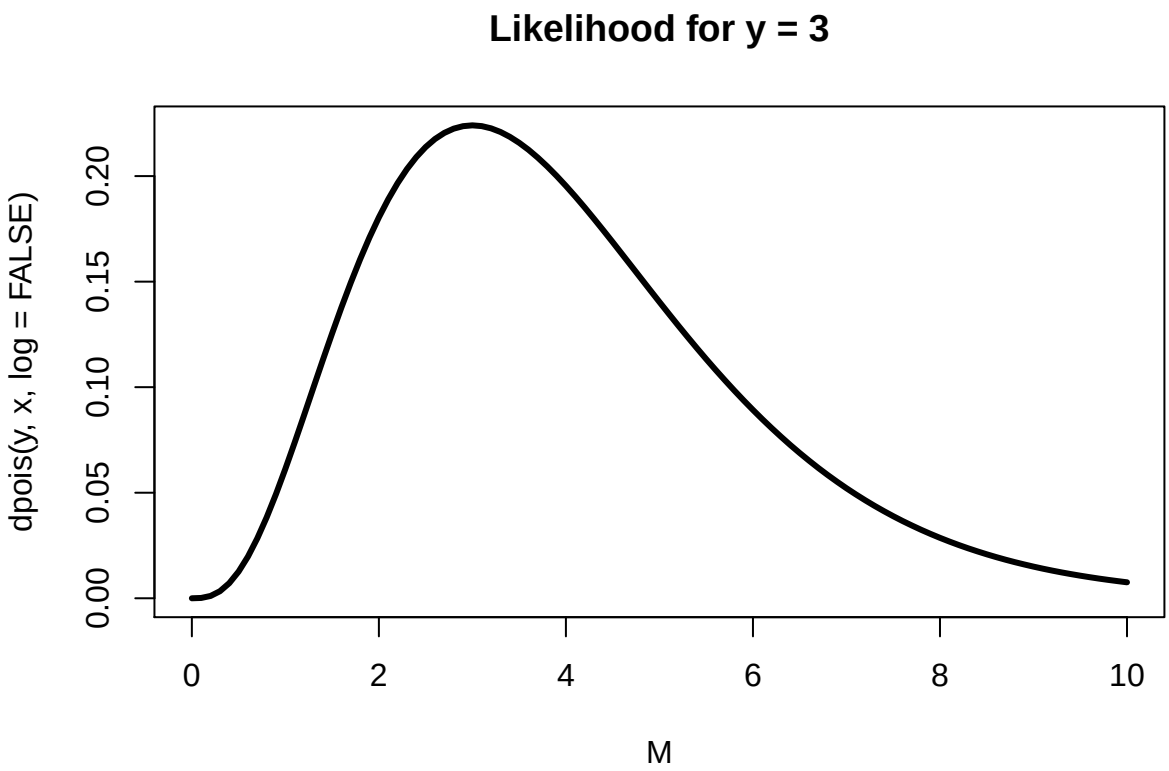
```
#We would reject not reject the null.
```

Problem 2 (6 pts)

(a) Suppose we observe 1 observation from a Poisson model with mean μ, obtaining $y = 3$. Using the `dpois` function in `R`, plot the likelihood function over the range from $\mu = 0$ to $\mu = 10$. Hint: see an analogous Binomial example in the the class notes "1.4 Likelihood Based Inferences - Part II", and also check the R help documentation for `dpois`.

Answer:

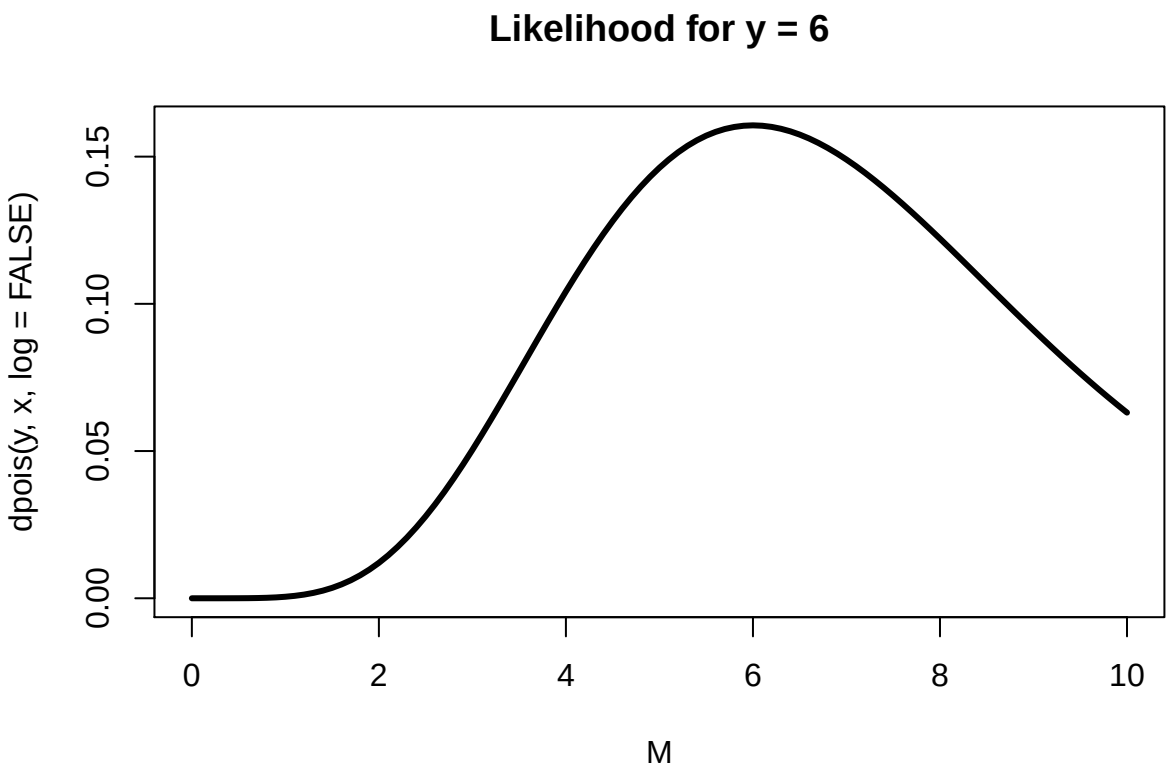
```
y = 3
curve(dpois(y, x, log = FALSE), xlim=c(0,10), xlab="M",
main=paste("Likelihood for y =" , y), lwd=3)
```



(b) Redo the plot in (a) if we observe $y = 6$ instead.

Answer:

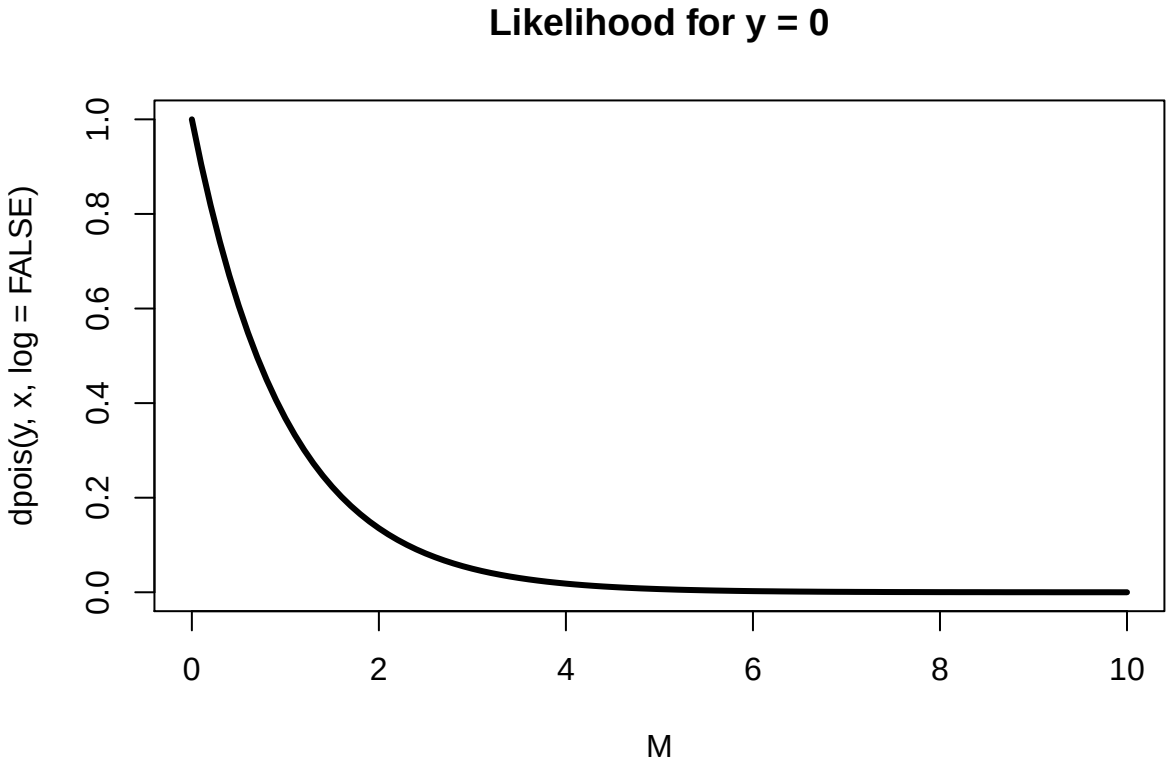
```
y = 6
curve(dpois(y, x, log = FALSE), xlim=c(0,10), xlab="M",
main=paste("Likelihood for y =" , y), lwd=3)
```



(c) Redo the plot in (a) if we observe $y = 0$.

Answer:

```
y = 0
curve(dpois(y, x, log = FALSE), xlim=c(0,10), xlab="M",
main=paste("Likelihood for y =" , y), lwd=3)
```

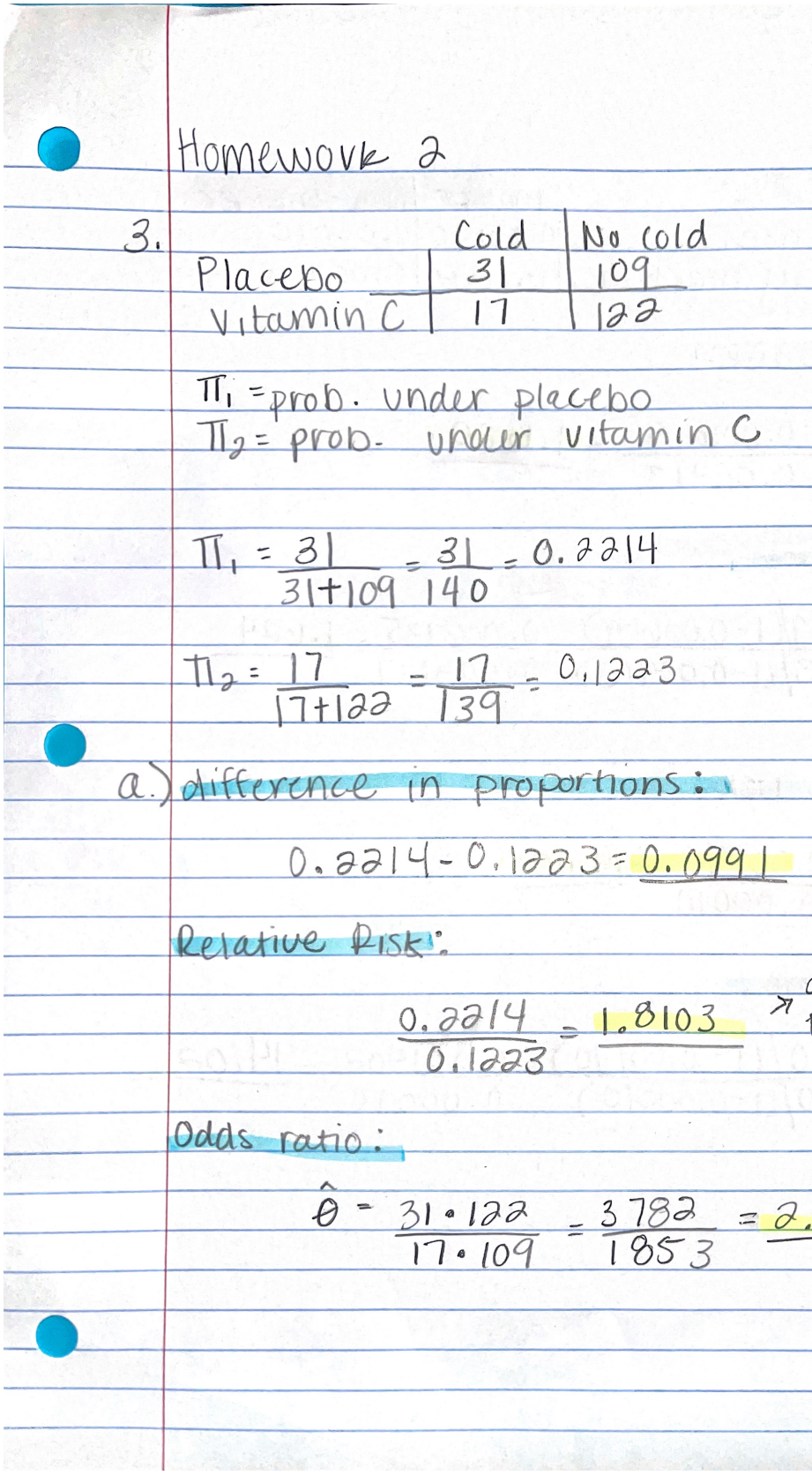


Problem 3 (4 pts)

The table below summarizes data from a double blind study involving 270 French skiers to compare incidence of Common Colds for skiers given placebo or Vitamin C supplements.

	Cold	No Cold
Placebo	31	109
Vitamin C	17	122

(a) Compute the estimated difference in proportions, relative risk and odds ratio based on these data.



Homework 2

3.

	Cold	No cold
Placebo	31	109
Vitamin C	17	122

π_1 = prob. under placebo
π_2 = prob. under vitamin C

$\pi_1 = \frac{31}{31+109} = \frac{31}{140} = 0.2214$

$\pi_2 = \frac{17}{17+122} = \frac{17}{139} = 0.1223$

a.) difference in proportions:

$0.2214 - 0.1223 = 0.0991$

Relative Risk:

$\frac{0.2214}{0.1223} = 1.8103$

Odds ratio:

$\hat{\theta} = \frac{31 \cdot 122}{17 \cdot 109} = \frac{3782}{1853} = 2.$

Answer: Work done on separate sheet.

(b) Interpret the results in (a): according to each measure, which treatment appears better

for preventing the Common Cold, and how large is the estimated effect (difference, ratio etc)?

Answer: Vitamin C is appears to be better at preventing the Common Cold because it has a probability of 0.122 compared to placebo with probability if 0.221. Based off of the relative risks, there is a 1.81 more of a chance that the population has a cold with the placebo.

Problem 4 (4 pts)

A 20-year cohort study of British male physicians found that the proportion per year who died from lung cancer was 0.00140 for cigarette smokers and 0.00010 for nonsmokers. The proportion who died from coronary heart disease was 0.00669 for smokers and 0.00413 for nonsmokers.

(a) Describe the association of smoking with coronary heart disease using the relative risk and odds ratio. Interpret the results in terms of how the coronary heart disease risks compare for the two groups.

4.

	Smoker	Non-Smoker
lung cancer	0.00140	0.00010
Coronary heart dis.	0.00669	0.00413

a) relative risk:

$$\frac{0.006669}{0.00413} = 1.620$$

Odd ratio:

$$\frac{0.006669/(1-0.006669)}{0.00413/(1-0.00413)} = \frac{0.006735}{0.004147} = 1.624$$

b) relative risk:

$$\frac{0.00140}{0.00010} = 14.00$$

Odd ratio:

$$\frac{0.00140/(1-0.00140)}{0.00010/(1-0.00010)} = \frac{0.001402}{0.00010} = 14.02$$

Answer: Work done on separate sheet. Based off of the results, smokers are about 1.62 times more likely to die form Coronary heart disease.

(b) Describe the association of smoking with lung cancer using the relative risk and odds ratio. Interpret the results in terms of how the lung cancer risks compare for the two groups.

Answer: Work done on separate sheet. Based off of the results, a smoker is about 14.0 time more likely to die lung cancer than a non-smoker.