Extra Practice: Odds Ratios, Relative Risk, and GLMs
STAT 244NF: Infectious Disease Modeling

YOUR NAME HERE (AND PARTNER'S NAME, IF APPLICABLE)

06 October, 2021

Question 1

Suppose the following data are collected from a cohort study.

```
set.seed(1242356)
sim <- rmultinom(n = 1, size = 1000, prob = c(0.06, 0.03, 0.04, 0.87))
id_outbreak <- data.frame(case = c(rep(1, sim[1]), rep(1, sim[2]),
                                   rep(0, sim[3]), rep(0, sim[4])),
                          exposure = c(rep(1, sim[1]), rep(0, sim[2]),
                                       rep(1, sim[3]), rep(0, sim[4])))
```

For each of the following parts (a)-(c) lines of code, indicate the probability that is being calculated by selecting one of the choices. You should select this choice by putting two asterisks (*) on either side of your choice. This will make your choice bold when you knit your document.

Part (a)

```
id_outbreak %>%
  summarize(p = mean(case))
```

```
##       p
## 1 0.086
```

- $P(\text{disease}|\text{exposure})$
- $P(\text{disease}|\text{no exposure})$
- $P(\text{exposure}|\text{disease})$
- $P(\text{no exposure}|\text{disease})$
- $P(\text{disease})$
- $P(\text{exposure})$

Part (b)

```
id_outbreak %>%
  filter(exposure == 0) %>%
  summarize(p = mean(case))
```

```
##            p
## 1 0.03329634
```

- $P(\text{disease}|\text{exposure})$
- $P(\text{disease}|\text{no exposure})$
- $P(\text{exposure}|\text{disease})$
- $P(\text{no exposure}|\text{disease})$
- $P(\text{disease})$
- $P(\text{exposure})$

Part (c)

```
id_outbreak %>%
  filter(exposure == 1) %>%
  summarize(p = mean(case))
```

```
##           p
## 1 0.5656566
```

- $P(\text{disease}|\text{exposure})$
- $P(\text{disease}|\text{no exposure})$
- $P(\text{exposure}|\text{disease})$
- $P(\text{no exposure}|\text{disease})$
- $P(\text{disease})$
- $P(\text{exposure})$

Question 2

For each of the parts (a)-(f), use the relevant output from the two model fits below. For each line of code, indicate the quantity that is being calculated by selecting one of the choices. You should select this choice by putting two asterisks (*) on either side of your choice. This will make your choice bold when you knit your document.

```
mod1 <- glm(data=id_outbreak, case ~ exposure, family=binomial(link = "log"))
summary(mod1)
```

```
##
## Call:
## glm(formula = case ~ exposure, family = binomial(link = "log"),
##     data = id_outbreak)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2914  -0.2602  -0.2602  -0.2602   2.6086
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.4023     0.1795  -18.95   <2e-16 ***
## exposure      2.8325     0.1999   14.17   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 586.37  on 999  degrees of freedom
## Residual deviance: 398.66  on 998  degrees of freedom
## AIC: 402.66
##
## Number of Fisher Scoring iterations: 7
```

```
mod2 <- glm(data=id_outbreak, case ~ exposure, family=poisson(link = "log"))
summary(mod2)
```

```
##
## Call:
## glm(formula = case ~ exposure, family = poisson(link = "log"),
##     data = id_outbreak)
##
## Deviance Residuals:
##     Min        1Q    Median        3Q       Max
## -1.0636   -0.2581   -0.2581   -0.2581    2.2071
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.4023     0.1826  -18.64   <2e-16 ***
## exposure      2.8325     0.2263   12.52   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 421.99  on 999  degrees of freedom
## Residual deviance: 267.95  on 998  degrees of freedom
## AIC: 443.95
##
## Number of Fisher Scoring iterations: 6
```

Part (a)

```
exp(coefficients(mod1)[2])
```

```
## exposure
## 16.98855
```

- Relative risk
- Odds ratio
- $P(\text{disease}|\text{exposure})$
- $P(\text{disease}|\text{no exposure})$

Part (b)

```
exp(coefficients(mod2)[1])
```

```
## (Intercept)
##  0.03329634
```

- Relative risk
- Odds ratio
- $P(\text{disease}|\text{exposure})$
- $P(\text{disease}|\text{no exposure})$

Part (c)

```
exp(coefficients(mod2)[2])
```

```
## exposure
## 16.98855
```

- Relative risk
- Odds ratio
- $P(\text{disease}|\text{exposure})$
- $P(\text{disease}|\text{no exposure})$

Part (d). What do you notice about the relative risk and odds ratio in this fictional outbreak? Are they similar or different? Why?

Part (e). The 95% confidence interval for the measure (ratio) of risk (either odds ratio or relative risk - you decide!) is calculated for `mod2` below. Notice that it does not contain 1 (the lower end point is about 11, and the higher end point is about 27, so 1 is outside the interval, and specifically below it). Does this mean there is evidence that the probability (or odds) of getting disease for those that were exposed is significantly larger than the probability (or odds) of getting disease for those that were not exposed?

```
exp(confint(mod2, parm="exposure"))
```

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

```
##     2.5 %    97.5 %
## 10.99215 26.78486
```

Part (f). Another way of addressing the question posed in Part (e) would to carry out a hypothesis test, where we determine whether exposure has an effect on the probability (or odds) of developing disease. The relevant test would be $H_0 : \beta_{exposure} = 0$ versus $H_0 : \beta_{exposure} \neq 0$ (or in other words - the null hypothesis is that there is no effect of exposure on the probability (or odds) of developing disease versus the alternative that there is an effect on the probability (or odds) of exposure. Like in linear regression, the summary of our models automatically carries out this test for us, and the z-value and p-value for this test are reported on the line for `exposure` in each of the models. Using this output, is there any evidence that there is an effect of exposure on the probability (or odds) of developing disease?