Week 03 Questions

Scott Graham 10131323

September 29, 2017

2.29

Let $\alpha = 0.05$, and:

$$H_0 : OR = \frac{Odds_{Drug}}{Odds_{Control}} \leq 1$$

$$H_1 : OR \frac{Odds_{Drug}}{Odds_{Control}} > 1$$

```
results_mat <- matrix(nrow = 2, ncol = 2, c(7, 0, 8, 15))
rownames(results_mat) <- c("Drug", "Control")
colnames(results_mat) <- c("Normalized", "Not Normalized")
results_mat
```

```
##         Normalized Not Normalized
## Drug             7              8
## Control          0             15
```

```
fisher.test(results_mat, alternative = "greater")
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  results_mat
## p-value = 0.003161
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  2.645931      Inf
## sample estimates:
## odds ratio
##        Inf
```

Since our p-value (0.003161) is $< \alpha$, we reject our null hypothesis that the odd's ratios for the drug and the control $= 1$, and assume that the odds ratio is greater than 1.

2.31

a.

Let $\alpha = 0.05$, and:

$$H_0 : OR = \frac{Odds_{Surgery}}{Odds_{Radiation}} = 1$$

$$H_1 : OR \frac{Odds_{Surgery}}{Odds_{Radiation}} \neq 1$$

```
results_mat <- matrix(nrow = 2, ncol = 2, c(21, 15, 2, 3))
rownames(results_mat) <- c("Surgey", "Radiation")
colnames(results_mat) <- c(" Cancer Controlled", "Cancer Not Controlled")
results_mat
```

```
##             Cancer Controlled Cancer Not Controlled
## Surgey                     21                      2
## Radiation                  15                      3
```

```r
fisher.test(results_mat, alternative = "two.sided")
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  results_mat
## p-value = 0.6384
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##   0.2089115 27.5538747
## sample estimates:
## odds ratio
##   2.061731
```

Because our p-value (0.6384) $> \alpha$, we fail to reject our null hypothesis, and assume the odds of controlling cancer are the same if you have surgery or radiation based on the sample.

b.

```r
1 -
  phyper(
    q = results_mat[1, 1]
    ,m = sum(results_mat[, 1])
    ,n = sum(results_mat[, 2])
    ,k = sum(results_mat[1, ])) + 0.5*
  dhyper(
    x = results_mat[1, 1]
    ,m = sum(results_mat[, 1])
    ,n = sum(results_mat[, 2])
    ,k = sum(results_mat[1, ]))
```

```
## [1] 0.2430911
```

Because our p-value (0.2431) $> \alpha$, we fail to reject our null hypothesis, and assume the odds of controlling cancer are the same if you have surgery or radiation based on the sample.

2.39

a.

True

b.

True

c.

False

d.

True

e.

False

3.3

a.

$$\mathrm{E}\left[\mathrm{P}\left(Y=1|X=x\right)\right]=0.00255+0.00109x, x=\{0,0.5,1.5,4.0,7.0\}$$

That is, on average with no alcohol consumption, there is a 0.255% probability of a sex organ malformation. And on average, for each unit increase in x (alcohol consumption), there is a 0.108% increase in the probability of sex organ malformation, based on the sample.

b.

$$\mathrm{P}\left(Y=1|X=(0,7.0)\right)=0.00255{+}0.00109x=(0.00255{+}0.00109(0)), 0.00255{+}0.00109(7.0))=(0.00255,0.01018) \implies$$

$$RR=\frac{\mathrm{P}\left(Y=1|X=7.0\right)}{\mathrm{P}\left(Y=1|X=0\right)}=\frac{0.01018}{0.00255}=3.992157$$

3.9

```r
income.credit <- read_csv("Week_03_Data_3.09.csv")
```

```
## Parsed with column specification:
## cols(
##   Inc = col_integer(),
##   Card = col_integer()
## )
```

```r
income.credit
```

```
## # A tibble: 100 x 2
##       Inc  Card
##     <int> <int>
##  1    24     0
##  2    27     0
##  3    28     1
##  4    28     1
##  5    28     0
##  6    28     0
##  7    28     0
##  8    29     0
##  9    29     0
## 10    29     0
## # ... with 90 more rows
```

```r
income.credit.glm <- glm(Card ~ Inc, data = income.credit, family = "binomial")
summary(income.credit.glm)
```

```
##
## Call:
## glm(formula = Card ~ Inc, family = "binomial", data = income.credit)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8192  -0.6593  -0.5261   0.3394   2.0911
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.55611    0.71685  -4.961 7.02e-07 ***
## Inc          0.05318    0.01314   4.047 5.19e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 123.820  on 99  degrees of freedom
## Residual deviance:  96.963  on 98  degrees of freedom
## AIC: 100.96
##
## Number of Fisher Scoring iterations: 4
```

a.

$$\ln\left(\frac{\mathrm{P}\left(Y=1|X=x\right)}{1-\mathrm{P}\left(Y=1|X=x\right)}\right) = -3.5561 + 0.0532x \implies$$
$$\mathrm{P}\left(Y=1|X=x\right) = \frac{e^{-3.5561+0.0532x}}{1+e^{-3.5561+0.0532x}}$$

b.

On average, as the annual income in Lira increase, the probability of owning at least one travel credit card increases as well, based on the sample.

c.

$$\ln\left(\frac{0.5}{1-0.5}\right) = -3.5561 + 0.0532x \implies 0 = -3.5561 + 0.0532x \implies x = 66.84398 \approx 66.86$$

When our LHS $= 0$, we can solve for x, which gives $x \approx 66.86$ million Lira. That is, on average there is a 50% chance of possessing at least one travel credit card when one's annual income is 66.86 million Lira, based on the sample.