Swetha Adike

Venu Goud Raparti

Problems from Chapter 18

MSIS 545

9.

This problem is about cardio vascular deaths in men, which came out in an argument between two schools about the relationship of obesity and heart disease. As a part of this study, researchers categorized subjects as obese or not according to their weight and height ratio. The data consisted of number of deaths in men in each category who died and did not die of CVD in a period of time. As shown in the data, there are a total of 1201 obese men out of which 22 were dead because of cvd and 1431 men who are not obese, of which 22 were dead and 1409 were not dead on cvd. In this problem, let's find out few computations to relating to the deaths and obesity in men

a. i.

Let's find the sample proportion cvd deaths for obese and nonobese groups in general. Below is code in R

```
> obese <- 22/(1179+22)

> nonobese <- 22/(1409+22)

> obese

[1] 0.01831807

> nonobese

[1] 0.01537386
```

The above results show that 0.0183 proportion is for cvd deaths in obese men and 0.0154 for nonobese men

ii.

For the above resulted sample proportions, let's find the standard error for the difference

```
> sdofdiff <- sqrt(obese*(1-obese)/1201 + nonobese*(1-nonobese)/1431)

> sdofdiff

[1] 0.005098606
```

From the above formula, 0.0051 is standard error of difference

iii.

Below is 95% confidence interval calculation for the population proportions. Though a prop test can be used for this, since the confidence interval is calculated for population, let's use conventional formula along with prop test results

```
> ll <- (obese - nonobese) - 1.96 * sdofdiff

> ul <- (obese - nonobese) + 1.96 * sdofdiff

> c(ll, ul)

[1] -0.007049065  0.012937473

> stats::prop.test(cvd, correct = FALSE)

     2-sample test for equality of proportions without continuity correction

data:  cvd

X-squared = 0.34434, df = 1, p-value = 0.5573

alternative hypothesis: two.sided

95 percent confidence interval:

 -0.006963065  0.012851473

sample estimates:

    prop 1     prop 2

0.01831807 0.01537386
```

More or less, both the methods give similar result of confidence interval between -0.00696 to 0.01285

b.

Let's find the one-sided p value for the test of equal population proportions, using the standard error of difference, which is already computed in the previous section

```
> nstat <- (obese - nonobese)/sdofdiff

> nstat

[1] 0.5774527

> 1 - pnorm(nstat)

[1] 0.2801169
```

From the above obtained result, 0.2801 is the one-sided p value for the test

c. i.

Let's compute odds of cvd deaths for obese and nonobese men

In general, odds for obese men = 22/1179 = 0.01866 and odds for nonobese men = 22/1409 = 0.01561.

Computing the same in R by below command, we get

```
> oddsRatio(cvd, verbose = TRUE )

Odds Ratio

Proportions

        Prop. 1:     0.01832

        Prop. 2:     0.01537

      Rel. Risk:     0.8393

Odds

         Odds 1:     0.01866

         Odds 2:     0.01561

     Odds Ratio:     0.8368

95 percent confidence interval:

      0.4671 < RR < 1.508

      0.4611 < OR < 1.519

NULL

[1] 0.8367637
```

ii.

Let's compute estimated odds ratio for obese deaths to nonobese deaths = 0.01866/0.01561 = 1.195388 or

```
> cvd <- matrix(c(1179, 1409, 22, 22), ncol = 2)

> oddsRatio(cvd)

[1] 1.195081
```

iii.

Now let's compute standard error of estimated log odds ratio, by the below formula

```
> logstd <- sqrt((1/1179)+(1/1409)+(1/22)+(1/22))

> logstd
```

```
[1] 0.3040839
```

The standard error for log transformed odds ratio for cvd deaths in obese and nonobese men is 0.3041

iv.

Let's calculate 95% confidence interval for odds ratio

```
> cvd <- matrix(c(1179, 1409, 22, 22), ncol = 2)

> oddsRatio(cvd, verbose = TRUE )

Odds Ratio

Proportions

        Prop. 1:     0.9817

        Prop. 2:     0.9846

      Rel. Risk:     1.003

Odds

         Odds 1:     53.59

         Odds 2:     64.05

     Odds Ratio:     1.195

95 percent confidence interval:

      0.9929 < RR < 1.013

      0.6585 < OR < 2.169
```

The confidence interval for the odds ratio is between 0.6585 to 2.169

d.

The above RR and OR in the above section indicate the risk and odds comparison between obese and nonobese groups. It says that the risk of cvd death in obese group can be estimated to be between 0.9929 and 1.013 times the risk of cvd deaths in nonobese men and similarly, the odds cvd death in obese men can be between 0.6585 to 2.169 times the odds of cvd deaths in nonobese men.

12.

This problem is regarding the probability of disease in population of unvaccinated and vaccinated groups. It is given that the probability of disease is 0.00369 in a population of unvaccinated subjects and the probability of disease is 0.001 in population of vaccinated subjects. Let's find the odds ration and compute the other findings that can be drawn from this data.

a.

Let's calculate the odds of disease in unvaccinated group relative to the odds of disease with vaccinated group

odds of disease without vaccination = prob(disease without vaccine) / 1 – prob(disease without vaccine) = 0.00369/(1-0.00369) = 0.00370

odds of disease with vaccination = prob(disease with vaccine) / 1 – prob(disease with vaccine) = 0.001/(1-0.001) = 0.0009

Now, odds of unvaccinated group = odds of disease without vaccination / odds of disease with vaccination

= prob(unvaccinated population) / prob(vaccinated population) = 0.00369 / 0.0009 = **3.6999**

b.

Out of a population of 100,000, the chance of people getting the disease if they were not treated with vaccine is n* prob(disease without vaccine) = 100,000 * 0.00369 = 369

369 members out of 100,000 is prone to get the disease if they are not vaccinated

c.

Similarly, Out of a population of 100,000, the chance of people getting the disease if they were treated with vaccine is n* prob(disease with vaccine) = 100,000 * 0.001 = 100

100 members out of 100,000 is prone to get the disease if they are vaccinated

d.

Let's find out the proportion of the people out of 100,000 who had the disease would be spared if all 100,000 were vaccinated. In simple, if a person had chance of getting the disease but spared because of vaccination.

number of unvaccinated people getting the disease out of 100,000 is 369. Out of 369 the chance of a person being saved if they are vaccinated = 369-100 = 269 and the protection rate can be found by taking the ratio between the spared people and the unvaccinated people = 269/369 = 0.729

e.

Given the protection rate if unvaccinated probability if disease = 0.48052 and

vaccinated probability = 0.2

the odds of disease in unvaccinated group relative to the odds of disease with vaccinated group

odds of disease without vaccination = prob(disease without vaccine) / 1 – prob(disease without vaccine) = 0.48052/(1-0.48052) = 0.925

odds of disease with vaccination = prob(disease with vaccine) / 1 – prob(disease with vaccine) = 0.2/(1-0.2) = 0.25

Now, odds of unvaccinated group = odds of disease without vaccination / odds of disease with vaccination

= prob(unvaccinated population) / prob(vaccinated population) = 0.925/0.25 = 3.70

Out of a population of 100,000, the chance of people getting the disease if they were not treated with vaccine is n* prob(disease without vaccine) = 100,000 * 0.48052 = 48052

48052 members out of 100,000 is prone to get the disease if they are not vaccinated

Similarly, Out of a population of 100,000, the chance of people getting the disease if they were treated with vaccine is n* prob(disease with vaccine) = 100,000 * 0.2 = 20000

20000 members out of 100,000 is prone to get the disease if they are vaccinated

Let's find out the proportion of the people out of 100,000 who had the disease would be spared if all 100,000 were vaccinated. In simple, if a person had chance of getting the disease but spared because of vaccination.

number of unvaccinated people getting the disease out of 100,000 is 48052. Out of 48052 the chance of a person being saved if they are vaccinated = 48052-20000 = 28052 and the protection rate can be found by taking the ratio between the spared people and the unvaccinated people = 28052/48052 = 0.584

13.

Given that πu = probability of disease for unvaccinated subjects

and πv = probability of disease for vaccinated subjects

the relative risk due to not vaccinating is $\beta = \pi u / \pi v$

a.

Let's consider small sample probability for disease for unvaccinated and vaccinated subjects and see if there will be any significance can be found between relative risk and odds ratio

Before the sample probability considerations,

Relative risk due to no vaccine, RR = $\pi u / \pi v$ and odds ratio due to no vaccination, OR = odds of no vaccination with respect to vaccination and odds of no vaccination in absence of vaccination

Let $\pi u = 0.004$ and $\pi v = 0.0004$

Then RR = 10

Odds ratio = (0.004/(1-0.004)) / (0.0004/(1-0.0004)) = 10.036

And so, by the above, we can say that for small probabilities, relative risk and odds ratio are nearly same

b.

For the probability of disease, 0.00369 in a population of unvaccinated subjects and the probability of disease, 0.001 in population of vaccinated subjects, let’s find the relative risk

RR = probability of disease for unvaccinated subjects / probability of disease for vaccinated subjects = 0.00369/0.001 = 3.69

c.

For the probability of disease, 0.48052 in a population of unvaccinated subjects and the probability of disease, 0.2 in population of vaccinated subjects, let’s find the relative risk

RR = probability of disease for unvaccinated subjects / probability of disease for vaccinated subjects = 0.48052/0.2 = 2.4026

d.

Given that relative risk = 50

πu1, unvaccinated probability of disease = 0.0050

πv1, vaccinated probability of disease = 0.0001

Relative risk = 0.0050/0.0001 = 50

in another situation, πv2 = 0.05, given RR2 = 50, πu2 = 50 * 0.05 = 2.5, which cannot be considered as probability as it greater than 1

19.

This problem consists of 123 low income, African-American children who were at high risk of school failure. 58 of them were randomly assigned to participate in high quality preschool program at age of 3 and 4 and remaining 65 received no preschool education. At the age of 40, they were again interviewed and the results were as follows:

35 subjects who received no preschool education were arrested 5 or more time by the age of 40 while 28 of these subjects were not; totaling to 63

20 subjects who received preschool education were arrested 5 or more time by the age of 40 while 36 of these subjects were not; totaling to 56

And 38 subjects with no preschool education made and annual income of 20K or more by the age of 40 while 25 of them could not; totaling to 63

22 subjects with preschool education made and annual income of 20K or more by the age of 40 while 34 of them could not; totaling to 56

Now Let's find the probability of the subjects with no preschool education and have arrested 5 or more times

```
> nopresarr.y <- 35/(35+28)

> nopresarr.y

[1] 0.5555556
```

And the probability of the subjects with preschool education and have arrested 5 or more times

```
> presarr.y <- 20/(36+20)

> presarr.y

[1] 0.3571429
```

Below is the standard error of difference of in the above probabilities

```
>    arr.sd.diff    <-    sqrt(nopresarr.y*(1-nopresarr.y)/63    +
presarr.y*(1-presarr.y)/56)

> arr.sd.diff

[1] 0.08954952
```

Let's compute prop.test to see if there is any evidence of effect of preschool on the subjects who were arrested and not arrested 5 or more time

```
> stats::prop.test(arr, correct = FALSE, alternative = "less")

     2-sample test for equality of proportions without continuity
correction

data:  arr

X-squared = 4.6954, df = 1, p-value = 0.01512

alternative hypothesis: less

95 percent confidence interval:

 -1.00000000 -0.05126079

sample estimates:

   prop 1    prop 2

0.4375000 0.6363636
```

Let's find the one-sided p value

```
> arr.stat <- (nopresarr.y - presarr.y)/arr.sd.diff

> arr.stat
```

```
[1] 2.215676
> 1 - pnorm(arr.stat)
[1] 0.01335686
```

The p value of 0.0133 shows a clear evidence of effect of preschool education in the subjects who were arrested 5 or more times

Let’s find the odds ratio to analyze further

```
> arr <- matrix( c(28, 35, 36, 20), ncol = 2 )
> oddsRatio(arr, verbose = TRUE )
Odds Ratio
Proportions
        Prop. 1:     0.4375
        Prop. 2:     0.6364
      Rel. Risk:     1.455
Odds
         Odds 1:     0.7778
         Odds 2:     1.75
     Odds Ratio:     2.25
95 percent confidence interval:
      1.033 < RR < 2.048
      1.075 < OR < 4.71
NULL
[1] 2.25
```

The above results show that the odds for the subjects without preschool education who were arrested 5 or more times by age of 40 is estimated to be 2.25 times the odds of the subjects with preschool education with a 95% confidence interval of 1.075 to 4.71

Similarly, let’s find the probability of the subjects with no preschool education and had made to an income of 20k or more

```
> nopresinc.y <- 25/(38+25)
> nopresinc.y
[1] 0.3968254
```

And the probability of the subjects with preschool education and had made to an income of 20k or more

```
> presinc.y <- 34/(22+34)

> presinc.y

[1] 0.6071429
```

Below is the standard error of difference of in the above probabilities

```
> inc.sd.diff <- sqrt(nopresinc.y*(1-nopresinc.y)/63 + presinc.y*(1-presinc.y)/56)

> inc.sd.diff

[1] 0.08976959
```

Let's compute prop.test to see if there is any evidence of effect of preschool on the subjects who made and income of 20k or more

```
> stats::prop.test(income, correct = FALSE, alternative = "less")

     2-sample test for equality of proportions without continuity correction

data:  income

X-squared = 5.2459, df = 1, p-value = 0.011

alternative hypothesis: less

95 percent confidence interval:

 -1.00000000 -0.06265963

sample estimates:

   prop 1    prop 2

0.3928571 0.6031746
```

Let's find the one-sided p value

```
> inc.stat <- (presinc.y - nopresinc.y)/inc.sd.diff

> inc.stat

[1] 2.342859

> 1 - pnorm(inc.stat)

[1] 0.009568314
```

The one-sided p value of 0.009 shows a clear evidence of effect of preschool education in the subjects whose income was 20k or more and the other group

Let's find the odds ratio to analyze further

```
> income <- matrix( c(22, 38, 34, 25), ncol = 2 )

> oddsRatio(income, verbose = TRUE )

Odds Ratio

Proportions

        Prop. 1:     0.3929

        Prop. 2:     0.6032

      Rel. Risk:     1.535

Odds

         Odds 1:     0.6471

         Odds 2:     1.52

     Odds Ratio:     2.349

95 percent confidence interval:

      1.048 < RR < 2.25

      1.125 < OR < 4.906

NULL

[1] 2.349091
```

The above results show that the odds for the subjects without preschool education who earned 20k or more income by age of 40 is estimated to be 2.35 times the odds of the subjects with preschool education with a 95% confidence interval of 1.125 to 4.906.

The above results are just for reference as the author established an uncertainty on the random assignment of the children in preschool.