Homework #3

Dustin Leatherman

April 27, 2019

`p. 570-571, #19`

9 (p568)

| Yes | No |
Obese | 22 | 1179 |
Non-Obese | 22 | 1409 |

(a.i)

Compute the sample proportions of CVD deaths for obese and non-obese groups

Obese CVD death Proportion: 22/ (1179 + 22) = 0.0183 Non-obese CVD death Proportion: 22/ (1409 + 22) = 0.0154

(a.ii)

Compute the standard error for the difference in sample proportions

$$\sqrt{\frac{\pi_1(1-\pi_2)}{n_1} + \frac{\pi_2(1-\pi_1)}{n_2}}$$

$\pi_1 = 0.0183$

$\pi_2 = 0.0154$

$n_1 = 22 + 1179 = 1201$

$n_2 = 22 + 1409 = 1431$

0.0051

(a.iii)

Compute a 95% C.I for the difference in population proportions

$$\pi_1 - \pi_2 \pm z_{0.975}\sqrt{\frac{\pi_1(1-\pi_2)}{n_1} + \frac{\pi_2(1-\pi_1)}{n_2}}$$

(-0.007, 0.0129)

(b)

Find a one-sided p-value for the test of equal populations proportions (using the std error already computed)

$Z = \frac{\pi_1 - \pi_2}{\sqrt{\frac{\pi_1(1-\pi_2)}{n_1} + \frac{\pi_2(1-\pi_1)}{n_2}}} = 0.5823$

1 - pnorm(0.5823) = **0.2802**

(c.i)

Compute the sample odds ratio for CVD death for Obese vs Non-Obese groups

$$\frac{22 * 1409}{22 * 1179} = 1.1951$$

(c.ii)

Compute the estimated odds ratio

$$log(1.1951) = 0.1782$$

(c.iii)

Compute the standard error of the estimated log odds ratio

$$\sqrt{\frac{1}{n_1\pi_1(1-\pi_1)} + \frac{1}{n_2\pi_2(1-\pi_2)}} = 0.3041$$

(c.iv)

Compute a C.I. for the odds ratio

(-2.4783, 2.8347)

(d)

There is no evidence that there is a difference in odds from samoans who died from CardioVascular Disease when obese compared to somoan patients who died from CardioVascular Disease when non-obese (Approximate 95% C.I: -2.4783 to 2.8347).

10 (p568)

Infantile paralysis victim YES NO Placebo 142 199,858 Vaccine 56 199,944

a

Find a one-sided p-value for testing the hypothesis that the odds of infantile paralysis are the same for the placebo and the vaccine treatments.

$\pi_1 = 142/(142 + 199858) = 0.0007$ $\pi_2 = 56/(56 + 199944) = 0.0003$ $n_1 = 142 + 199858 = 200000$ $n_2 = 56 + 199944 = 200000$

```r
n1 <- 142 + 199858
n2 <- 56 + 199944

pi1 <- 142 / n1
pi2 <- 56 / n2

std.err.log <- sqrt((1 / (n1 * pi1 *(1 - pi1))) + (1 / (n2 * pi2 * (1 - pi2))))

1 - pnorm(log(pi1 / pi2) / std.err.log)
```

```
## [1] 1.867345e-09
```

The one-sided p-value is approximately 1.867345e-142/1989

(b)

```r
pi <- 142/198
z <- (pi - 0.5) / sqrt(pi * (1 - pi) * (1 / 198))
1 - pnorm(z)
```

```
## [1] 5.796141e-12
```

There is convincing evidence that the proportion of Infantile Paralysis victims in the placebo group is greater than the proportion of Infantile Paralysis Victims in the Vaccine group (p-value = 5.796141e-12).

11

```
In the hypothetical example of Display 18.11, calculate each of the following quantities:

i. in the population,
ii. in the expected results from prospective sampling
iii. in the expected results from retrospective sampling (with equal numbers of lung cancer patients and
```

	Cancer	No Cancer
Smokers	1000	2,000,000
Non-Smokers	4000	16,000,000

(a)

The proportion of lung cancer patients among smokers.

i. 1000 / (2,000,000 + 1000) = 0.0005
ii. 1000 / (2,000,000 + 1000) = 0.0005
iii. 1000/5000 = 1/5 => 18000000 / 5 = 3600000 => 3600000 / (3600000 + 2000000) = 0.6429

(b)

The proportion of lung cancer patients among nonsmokers.

i. 4000 / (4000 + 16000000) = 0.0002
ii. 4000 / (4000 + 16000000) = 0.0002
iii. 4000 / 5000 = 4/5 => 18000000 * .8 = 14400000 / (14400000 + 2000000) = 0.878

(c)

The difference between parts (a) and (b).

i. (4000 - 1000) / (4000 + 16000000 - (2000000 + 1000)) = 0.0002
ii. (4000 - 1000) / (4000 + 16000000 - (2000000 + 1000)) = 0.0002
iii. 4000 - 1000 / 5000 = 3/5 => 18000000 * .6 = 10800000 => 10800000/(10800000 + 2000000) = 0.8434

(d)

Verify that retrospective sampling results cannot be used to estimate these population properties.

Retrospective sampling can only be used to estimate the odds ratio. It cannot be used to estimate population proportions or differences in population proportions.

12

Suppose that the probability of a disease is 0.00369 in a population of unvaccinated subjects and that the probability of the disease is 0.001 in a population of vaccinated subjects.

(a)

What are the odds of disease without vaccine relative to the odds of disease with vaccine?

0.00369 / 0.001 = 3.69

(b)

How many people out of 100,000 would get the disease if they were not treated?

0.00369 * 100000 = 369

(c)

How many people out of 100,000 would get the disease if they were vaccinated?

0.001 * 100000 = 100

(d)

What proportion of people out of 100,000 who would have gotten the disease would be spared from it if all 100,000 were vaccinated? (This is called the protection rate.)

(369 - 100) / 369 = 0.729

(e)

Follow the steps in parts (a)–(d) to derive the odds ratio and the protection rate if the unvaccinated probability of disease is 0.48052 and the vaccinated probability is 0.2. (The point is that the odds ratio is the same in the two situations, but the total benefit of vaccination also depends on the probabilities.)

0.48052 / 0.2 = 2.4026.

0.48052 * 100000 = 48052

0.2 * 100000 = 20000

(48052 - 20000) / 48052 = 0.5838

13

An alternative to the odds ratio and the difference in probabilities is the relative risk, which is a common statistic for comparing disease rates between risk groups. If u and v are the probabilities of disease for unvaccinated and vaccinated subjects, the relative risk (due to not vaccinating) is pi_u = pi_v. (So if the relative risk is 2, the probability of disease is twice as large for unvaccinated as for vaccinated individuals.)

(a)

Show that the odds ratio is very close to the relative risk when very small proportions are involved. (You may do this by showing that the two quantities are quite similar for a few illustrative choices of small values of v and u.)

Let u = .001 and v = .005. Then $\rho = .001/.005 = 0.2$ For a population of 1000, $\phi = \frac{0.001 \times 1000}{0.005 \times 1000} = 0.2$

(b)

Calculate the relative risk of not vaccinating for the situation introduced at the beginning of Exercise 12.

0.00369 / 0.001 = 3.69

(c)

Calculate the relative risk of not vaccinating for the situation described in Exercise 12(e).

0.48052 / 0.2 = 2.4026

(d)

Here is one situation with a relative risk of 50: The unvaccinated probability of disease is 0.0050 and the vaccinated probability of disease is 0.0001 (the probability of disease is 50 times greater for unvaccinated subjects than for vaccinated ones). Suppose in another situation the vaccinated probability of disease is 0.05. What unvaccinated probability of disease would imply a relative risk of 50? (Answer: None. This shows that the range of relative risk possibilities depends on the baseline disease rate. This is an undesirable feature of relative risk.)

None as it said. 0.05 * 50 > 1 so its not possible.

14

The Pique Technique was developed because a target for solicitation is more likely to comply if mindless refusal is disrupted by a strange or unusual request. Researchers had young people ask 144 targets for money. The targets were asked either for a standard amount—a quarter—or an unusual amount—17 cents. Forty-three point one percent of those asked for 17 cents responded, compared to 30.6% of those asked for a quarter. Each group contained 72 targets. What do you make of that?

u = .431 v = .306

| Resp | No Resp |
17 | 31.032 | 40.968 |
25 | 22.032 | 49.968 |

$\phi = \frac{31.032 \times 49.968}{40.968 \times 22.032} = 1.718$

Young people who asked for 17 cents were 1.718 times more likely to get a response over young people who asked for 25 cents.

18

It is known that left-handed people tend to recall orientations of human heads or figures differently than right-handed people. To investigate whether there is a similar systematic difference in recollection of inanimate object orientation, researchers quizzed University of Oxford undergraduates on the orientation of the tail of the Hale-Bopp comet. The students were shown eight photographic pictures with the comet in different orientations (head of the comet facing left down, left level, left up, center up, right up, right level, right down, or center down) six months after the comet was visible in 1997. The students were asked to select the correct orientation. (The comet faced to the left and downward.) Shown below are the responses categorized as correct or not, shown separately for left- and right-handed students.

Is there evidence that left or right-handedness is associated with correct recollection of the orientation? If so, quantify the association. Write a brief statistical report of the findings. (Data from M. Martin, and G. V. Jones, "Hale-Bopp and Handed- ness: Individual Difference in Memory for Orientation American Psychological Science 10 (1999): 267–70.)

There is moderate evidence that the odds ratios between the two populations are not equal to 1 (Fisher Exact Test. p-value = 0.03572). It is estimated that Right Handed individuals are 1.636 times more likely to select orientation correctly than Left handed students. With 95% confidence, Right Handed Students select orientation between 1.0557 and 2.5363 times more correctly than left handed students.

```
library(epitools)
df <- data.frame(Incorrect=c(149,129), Correct=c(48,68), row.names = c("Left-Handed","Right-Handed"))
oddsratio.wald(as.matrix(df))
```

```
## $data
##              Incorrect Correct Total
## Left-Handed        149      48   197
## Right-Handed       129      68   197
## Total              278     116   394
##
## $measure
##                           NA
## odds ratio with 95% C.I. estimate    lower    upper
##             Left-Handed  1.000000       NA       NA
##             Right-Handed 1.636305 1.055662 2.536317
##
## $p.value
##               NA
## two-sided      midp.exact fisher.exact chi.square
##   Left-Handed          NA           NA         NA
##   Right-Handed 0.02776869   0.03547372 0.02705771
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

19

As discussed in Exercise 4, researchers in the mid-1960s identified 123 low-income, African-American children who were thought to be at high risk of school failure and randomly assigned 58 of them to participate in a high-quality preschool program at ages 3 and 4; the remaining 65 received no preschool education. One hundred nineteen of the study subjects were available for re-interview at age 40. The following two tables show results on two of the variables considered:

Assess the evidence for an effect of preschool on these two variables and report the estimated odds ratios for 5 or more arrests and for annual income of $20K or more. (Note: Although the authors reported that random assignment was used, there was, in fact, some switching of groups after the initial assignment. For example, children placed in the preschool group were switched to the control group if their mothers could not transport them to preschool. Although the authors defend the results as just as good as if random assignment was made, the failure to follow through with exact randomization—even though it may have been unavoidable—causes a necessary extra debate about the causal conclusions from this study.)

$\phi_{arrest} = \frac{36\times35}{28\times20} = 2.25$

It is estimated that African American Children who did not attend preschool are 2.25 times more likely to get arrested at least 5 times by the age of 40 than African American Children who did attend preschool.

$\phi_{20k} = \frac{38\times34}{22\times25} = 2.3491$

It is estimated that African American Children who attend preschool are 2.3491 times more likely to recieve an annual income of 20K or more by age 40 than African American Children who did not attend preschool.