Lab 13 Answers

April 28, 2020

Example Questions

Example 1

A study examined 793 individuals who were in bike accidents. It was found that of 147 riders that wore helmets, 17 of them had a head injury. Of the rest of the bikers who did not wear helmets, 428 did not get a head injury.

1) Make a contingency table for the data
2) Without running any tests, does there appear to be a benefit to wearing a helmet? (hint: odds ratio)
3) Make a 95% CI for this odds ratio
4) What are the expected counts for the contingency table?
5) Calculate chi-squared statistic and interpret

Example 2

A study compared the miles per gallon of American cars (sample 1) to Japanese cars (sample 2). The sample size for American cars was 249 with a sample mean of 20.145 and sample standard deviation of 6.415. Japanese cars had a sample size of 79, sample mean of 30.481, and sample standard deviation of 6.108. (pooled standard deviation is 6.343)

1) If we assume data is normal what test do we run? What else might we consider to determine what test is most appropriate?
2) Conduct a t-test comparing the two group means and interpret the results
3) Further analysis shows that two of the American cars in the sample were getting less than 5 miles per gallon. How might this affect the test results? How might you remedy this issue?

Example 3

The Predators (a hockey team) just reached the second round of playoffs. I was curious if the Predators experienced any benefit to playing at home this season, so I gathered the data on how many goals they scored each game and whether they were home or away (regular season only). In home games they scored an average of 3.098 goals in 41 games with a sd of 1.841. In away games (41) they scored 2.237 with a sd of 1.43. The pooled sd is 1.650.

1) Which test would you use to examine the association between location and goals scored?
2) It seems they did better at home could this be a difference explained by chance alone?

#Answers

Example 1

1)

```r
table<- rbind(c(17,130), c(218,428))

colnames(table)<- c("Y", "N")

rownames(table)<- c("Helmet", "No Helmet")

print(table)
```

```
##             Y   N
## Helmet     17 130
## No Helmet 218 428
```

2) Yes, the odds ratio is $(130*218)/(17*428) = 3.89$. The odds of those **not wearing a helmet** experiencing a head injury is 3.89 times greater than those wearing a helmet experiencing a head injury.

Alternately, you could calculate the odds by ad/bc=0.26, but the group you will interpret the odds for is now switched. To interpret this, you can say: The odds of those who **wore a helmet** experiencing a head injury is 74% lower than the odds of those who did not wear a helmet.

3)

```r
log(3.89) + c(1.96, -1.96) * sqrt(1/130+1/428+1/17+1/218)
```

```
## [1] 1.8895635 0.8272548
```

```r
exp(c(0.8272548, 1.8895635))
```

```
## [1] 2.287032 6.616480
```

4) Expected probability of having a head injury (both groups): (17+218)/793= 0.296

```r
table2<- rbind(c(43.5,103.5), c(191.2,454.8))

colnames(table2)<- c("Y", "N")

rownames(table2)<- c("Helmet", "No Helmet")

print(table2)
```

```
##               Y     N
## Helmet     43.5 103.5
## No Helmet 191.2 454.8
```

5)

```r
(17-43.5)^2/43.5 + (218-191.2)^2/191.2 + (130-103.5)^2/103.5 + (428-454.8)^2/454.8
```

```
## [1] 28.26443
```

```r
pchisq(28.26443, df=1, lower.tail = FALSE)
```

```
## [1] 1.058228e-07
```

Example 2

1) Student's

2) $t = \frac{20.145 - 30.481}{6.34 * \sqrt{1/249 + 1/79}} = -12.62$
Using (249+79 - 2) = 326 degrees of freedom, the p-value is less than 0.05; there is a difference in MPG– American cars get fewer miles per gallon.

3) The t-test procedure can be affected by outlier, which can impact the mean and standard error.

Assuming the measurements are accurate you might consider a non-parametric (Wilcoxon Rank Sum) approach to remove the effect of these outliers.

You can also consider a log transformation to shrink the scale of the variables.

Note: A confidence interval log-transformed data will reflect the ratio group means after being exponentiated.

Example 3

1) Probably Student's

2) $t = \frac{3.098 - 2.237}{1.650 * \sqrt{1/41 + 1/41}} = 2.3626$; p-value = 0.02057