DSCI/MATH 530 RLab Seven

Overview

In this assignment we will use R to

-examine the distribution of the Chi square test statistic using Monte Carlo sampling.

Submission

Please submit through gradescope. Complete this assignment using R Markdown and converting to pdf. *Make sure each problem is on a separate page.* Use `\newpage` before each problem to insert a page break.

You can also refer to the .Rmd file for this assignment to have an example of this markdown document. Note however, that this version is setup to also use Latex for some math and conversion directly to pdf.

1 Checking test statistic distribution for homework 9.30

Homework problem 9.30 is a clincal study that has the following results

	airsick	not airsick	Total
Placebo	60	48	108
Treatment	31	77	108
Totals	91	125	216

To test for a treatment effect one can form the statistic

```
n11<- 60
n12<- 48
n21<- 31
n22<- 77

n1<- n11 + n21
n2<- n12 + n22

e11<- (108*n1)/216
e12<- (108*n2)/216
e21<- (108*n1)/216
e22<- (108*n2)/216

W =  (e11- n11)^2 /e11 + (e12- n12)^2 /e12 +
     (e21- n21)^2 /e21 + (e22- n22)^2 /e22
W
```

```
## [1] 15.96976
```

```
# to test at .05 compare to
qchisq(.95, 4-3)
```

```
## [1] 3.841459
```

```
# p-value for test
pchisq( W, 4 - 3, lower.tail = FALSE)
```

```
## [1] 6.436247e-05
```

Here we are using the fact that equal numbers of subjects are allocated to the placebo and treatment group (108) and of course the total number of subject is $108 + 108 = 216$. Under the null hypothesis that there is no treatment effect the expected counts are the same for both groups because they are equal in size.

1(a) Scrutinzing the Chi Square business

To be honest using the chi square statistic W and the approximate distribution of a chi squared are mysterious to me. It is easy to check this, however, using simulation. Here is how to simulate one case under H_0 and based on the observed values from the study. Here **pH0** is the estimated probability of airsickness under no treatment effect. We need to assume this as the "true" value to do the simulation.

```r
WSim=c()
for (i in 0:10000){

NTotal<- 216
pH0<- 91/216
NGroup<- NTotal/2
#simulated counts

n11<- rbinom(1, NGroup, pH0)
n12 <- NGroup - n11
n21<- rbinom(1, NGroup, pH0)
n22 <- NGroup - n21

n1<- n11 + n21
n2<- n12 + n22

# expected counts under H0 and using "pHat = n1/216"
e11<- (NGroup*n1)/NTotal
e12<- (NGroup*n2)/NTotal
e21<- (NGroup*n1)/NTotal
e22<- (NGroup*n2)/NTotal

WSim[i] =  (e11- n11)^2 /e11 + (e12- n12)^2 /e12 +
     (e21- n21)^2 /e21 + (e22- n22)^2 /e22

}
quantile(WSim,.95)
```

```
##      95%
## 3.758523
```

Use this code inside a **for** loop to generate 10000 samples of **WSim** saving these in an array. Compare the .95 percentile from these exact samples to the chi squared value to test at $\alpha = .05$. Does your Monte Carlo give you an estimate for the p-value?

This is similar to the estimate at the alpha .05 and gives an estimate for the p value.

2 Small sample size

Now suppose that you want to redo this study but with a smaller group size. Assume **pH0 = 91/216** is still a reasonable estimate under the null hypothesis but your budget only allows for 50 subjects – 25 in each group. For these smaller groups is the chi squared distribution still a good approximation for testing at $\alpha = .05$?

Hint: If you have followed the code above the only change you need to make is for `NTotal` .

```
for (i in 0:10000){

NTotal<- 50
pH0<- 91/216
NGroup<- NTotal/2
#simulated counts

n11<- rbinom(1, NGroup, pH0)
n12 <- NGroup - n11
n21<- rbinom(1, NGroup, pH0)
n22 <- NGroup - n21

n1<- n11 + n21
n2<- n12 + n22

# expected counts under H0 and using "pHat = n1/216"
e11<- (NGroup*n1)/NTotal
e12<- (NGroup*n2)/NTotal
e21<- (NGroup*n1)/NTotal
e22<- (NGroup*n2)/NTotal

WSim[i] =  (e11- n11)^2 /e11 + (e12- n12)^2 /e12 +
     (e21- n21)^2 /e21 + (e22- n22)^2 /e22

}
quantile(WSim,.95)
```

```
##     95%
## 3.94525
```

This approximation still works but its more inaccurate.

3 EXTRA CREDIT

Suggest one improvement in the course that would increase your data analysis skills using R and R Studio.

I feel as though a dedicated video on hypothesis testing would have helped me.