Moodle Questions Chapter 11

Homer White

Friday, July 5, 2014

# Seating at GC I

Thinking back again to the m111survey data frame, recall that the variable **seat** records where each subject prefers to sit in a classroom: Front, Middle, or Back.

Here is a table of the results:

gcSeat <- xtabs(~seat,data=m111survey)  
gcSeat

## seat  
## 1\_front 2\_middle 3\_back   
## 27 32 12

Now at Georgetown most classrooms are fairly small, with at about four rows actually in use:

* the first row is obviously the Front;
* most people would think of the second and third rows as the Middle;
* the fourth row would count as the Back.

If preferences for the four rows are exactly the same out there in the GC population, then one would expect that:

* 25% of the population prefers the Front;
* 50% prefer the middle (twice as many rows in the Middle, after all);
* 25% prefer the back.

We wonder if the available data provide strong evidence against the idea of equal preference among all four rows.

Since **seat** is a factor variable, it makes sense to investigate our Research Question with chisqtestGC(). We define parameters and hypotheses as follows:

Let

pf = the proportion of all GC students who prefer the Front.

pm = the proportion of all GC students who prefer the Middle.

pb = the proportion of all GC students who prefer the Back.

The Hypotheses are:

H0: pf = 0.25, pm = 0.50, pb = 0.25.

Ha: At least one of the values in H0 is not correct.

Which of the following bits of R-code would do the job for us? (There are three correct answers.)

chisqtestGC(~seat,data=m111survey,  
 p=c(0.25,0.50,0.25))

chisqtestGC(gcSeat,p=c(0.25,0.50,0.25))

Feedback: chisqtestGC() will take pre-made tables, so this is also correct!

chisqtestGC(~seat,data=m111survey,  
 p=c(0.25,0.50,0.25),  
 simulate.p.value=TRUE,B=5000)

Feedback: It never hurts to d osimulation, with a few thousand resamples, and it only takes a fraction of a second longer than the routines that are based on approximating P-values iwth chi-square curves!

chisqtestGC(~seat,data=m111survey,  
 p=c(0.33,0.33,0.33))

Feedback: The p argument does not represent the Null Hypothesis accurately.

chisqtestGC(gcSeat,  
 p=c(0.33,0.33,0.33))

Feedback: The p argument does not represent the Null Hypothesis accurately.

chisqtestGC(~Seat,data=m111survey,  
 p=c(0.25,0.50,0.25))

Feedback: Watch out: R is case-sensitive!

# Seating at GC: II

In the Seating study, the results of the chi-sqaure test are as follows:

## Chi-squared test for given probabilities   
##   
## observed counts Expected by Null contribution to chisq statistic  
## 1\_front 27 17.75 4.82  
## 2\_middle 32 35.50 0.35  
## 3\_back 12 17.75 1.86  
##   
##   
## Chi-Square Statistic = 7.028   
## Degrees of Freedom of the table = 2   
## P-Value = 0.0298

Which of the following statements are true? (Select all correct statements.)

If the Null is right, then the chi-square statistic shold turn out to be about 2, give or take 2 or so.

If the Null is right, then the chi-square statistic shold turn out to be about 7.0282, give or take 2 or so.

There were 9.25 more front-sitters than the Null would have expected to see in a study of this size.

The Null expects to see about 17.75 back-sitters, give or take a few for chance error in the sampling process.

The Null expects to see about 12 back-sitters, give or take a few for chance error in the sampling process.

The Null expects to see about 27 back-sitters, give or take a few for chance error in the sampling process.

# Seating at GC: III

In the Seating study, what is the best interpretation of the P-value?

If the Null is correct, then there is about a 2.98% chance of getting a chi-square statistic that is at least as big as the one one we got in our study.

If the Alternative is correct, then there is about a 2.98% chance of getting a chi-square statistic that is at least as big as the one one we got in our study.

If the Null is correct, then there is about a 2.98% chance of getting a chi-square statistic that is less than the one one we got in our study.

If the Alternative is correct, then there is about a 2.98% chance of getting a chi-square statistic that is less than the one one we got in our study.

# Seating at GC: IV

In the Seating study, what's the best way to write a conclusion? (There are tow equally good ways: find and select both.)

The data provide strong evidence that students don't have equal preference for all class rows. In particular, they seem to prefer the Front somewhat.

Observed counts differed so much from what the Null expected that it's not reasonable to ascribe the difference to chance variation in the sampling process. In particular, people seemed to prefer the much more often than the Null expected.

The data provide strong evidence that students don't have equal preference for all class rows. In particular, they seem to prefer the Middle somewhat.

Observed counts differed from the counts the Null expected, but it's reasonable to ascribe the difference to chance variation in the sampling process.

# Good Chi-Square Code

A certain species of lizard comes in three colors: Red, Green and Purple. Here are the observed colors for 10 randomly-selected lizards:

* 1 Red
* 7 Green
* 2 Purple

We would like to test the following two hypotheses:

H0: All three colors are equally common in the population.

Ha: The three colors are not all equally common in the population.

What's the correct R-code to perform the test?

lizardColors <- c(red=1,green=7,purple=2)  
chisqtestGC(lizardColors,p=c(0.33,0.33,0.33),  
 simulate.p.value=TRUE,B=5000)

lizardColors <- c(red=1,green=7,purple=2)  
chisqtestGC(lizardColors,p=c(0.33,0.33,0.33),  
 simulate.p.value=TRUE,B=50)

Feedback: That's not enough resamples! Try setting `B~ to several thousand!

lizardColors <- c(red=1,green=7,purple=2)  
chisqtestGC(lizardColors,p=c(0.33,0.33,0.33))

Feedback: If you try this code, R will warn you about expected cell counts are too low to rely on a chi-square curve to approximate the P-value.

lizardColors <- c(red=1,green=7,purple=2)  
chisqtestGC(lizardColors,p=c(0.1,0.7,0.2),  
 simulate.p.value=TRUE,B=5000)

Feedback: The p argument reflects the data, but it needs to reflect what the Null Hypothesis thinks!