Introduction to R

The purpose of this handout is to demonstrate how to conduct a χ^2 test in R.

1   The χ^2 goodness of fit test in R

The χ^2 goodness of fit test is one of the simplest statistical tests you can conduct. The basic equation is as such:

$$\chi^2 = \sum_{i=1}^{n} \frac{(Obs_i - Exp_i)^2}{Exp_i}$$

What that means is that for each category of your data (e.g., genotype, allele, etc) you separately subtract the Expected value from the Observed value, square it, and then divide it by the Expected value. The sum of those values (for each category) is your χ^2 statistic. If you were calculating a χ^2 by hand, you would then compare your calculated value to a table of critical values to estimate the p-value associated with our statistical test. With R, it is possible (and easy!) to calculate the exact p-value.

The function in R to conduct a χ^2 goodness of fit test is `chisq.test()`. This function requires two vectors:

1. A vector of your observed values

2. A vector of your expected frequencies

R does the rest of the math for you. Remember, to make a set of numbers a vector, we combine with the `c()` function.

2   Example: Sex ratio in recently metamorphosed red-eyed treefrogs

In the `tadpole` dataset, we have previously looked at things like the size of tadpoles in relation to different treatment variables. However, we also have data on the sex of individuals (male or female) after they metamorphosed.

```
> str(tadpole)
'data.frame':  300 obs. of  8 variables:
 $ Ind     : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Pond    : int  3 5 5 2 4 5 4 4 2 2 ...
 $ Food    : Factor w/ 2 levels "Cont","Prot": 2 1 2 2 2 2 1 1 2 2 ...
 $ Pred    : Factor w/ 3 levels "High","Low","Med": 3 1 2 2 2 3 2 3 3 1 ...
 $ Size    : num  15.2 12.7 18.9 19.6 19.5 ...
 $ Spots   : num  76 146 33 36 13 64 36 56 67 112 ...
 $ MetSex  : int  0 0 0 0 0 0 1 1 0 0 ...
 $ AgeAtMet: int  56 50 79 87 80 47 47 77 49 43 ...
```

The variable `MetSex` contains the data about frog sex: 0's are females and 1's are males. We would most likely assume that we would see a 50/50 sex ratio and we can check this with a χ^2 goodness of fit test. First, we need to tally up how many males and females we have. This is easily accomplished using the `table()` command.

```
> Sex.Table<-table(tadpole$MetSex)
> Sex.Table

  0   1
169 131
```

We have 169 females and 131 males. Let's go ahead and change the headings of our table to make it easy to remember which is which.

```
> names(Sex.Table)<-c("Female","Male")
> Sex.Table
Female    Male
   169     131
```

In order to run the χ^2 we just need to define our observed values (169 and 131) and our expected frequency (0.5 and 0.5). Remember, our observed values are called x and our expected frequencies are called p.

```
> chisq.test(x=c(169,131),p=c(.5,.5))

  Chi-squared test for given probabilities

data:  c(169, 131)
X-squared = 4.8133, df = 1, p-value = 0.02824
```

This output tells us several things. First, it tells us what we have done, a Chi-squared test. Secondly, it tells us our actual χ^2 test statistic, which is 4.8133. Next we have our degrees of freedom, and lastly the p-value. Since the p-value is below 0.05, we feel comfortable to reject the null hypothesis of a 50/50 sex ratio. Amphibian sex is not 100% genetically determined and can be affected by the environment, so perhaps something about our experiment has caused this result.

In the next tutorial, we will begin using slightly more complex statistics, such as a t-test or Analysis of Variance.