Lab 6

Adam Orr

September 29, 2017

Section: 91973 Friday 9am

Question 1

Script

```r
# Read in these data and make a bar plot of the results.
setwd("~/code/biometry-lab/lab6")
cats <- read.csv("cats.csv")
barplot(cats$falls,
        main = "Monthly Number of Cat Falls",
        xlab = "Month",
        ylab = "Number of Falls",
        names.arg = cats$month,
        las = 2,
        ylim = c(0,max(cats$falls) + 5))

# Carry out a test of the hypothesis that falls are randomly distributed across
# months of the year. Explicitly perform the five basic steps of hypothesis
# testing. Clearly state the conclusion of the test. Do this test "by hand" in
# R.

# Step 1: State hypotheses
# h0: Falls occur randomly across months.
# ha: Falls are not randomly distributed across months.

# Step 2: State alpha
(alpha <- 0.05)

# Step 3: Define test statistic
# chisq = sum((obs - exp)^2/exp)

# Step 4: Calculate the test statistic
obs <- cats$falls
p_null <- rep(1 / length(obs),length(obs))
exp <- p_null * sum(obs)
(chisq <- sum((obs - exp)^2/exp))

# Step 5: Calculate the p-value and make a conclusion
# degrees of freedom
(df <- length(obs) - 0 - 1)

# p-value
1 - pchisq(chisq, df)
```
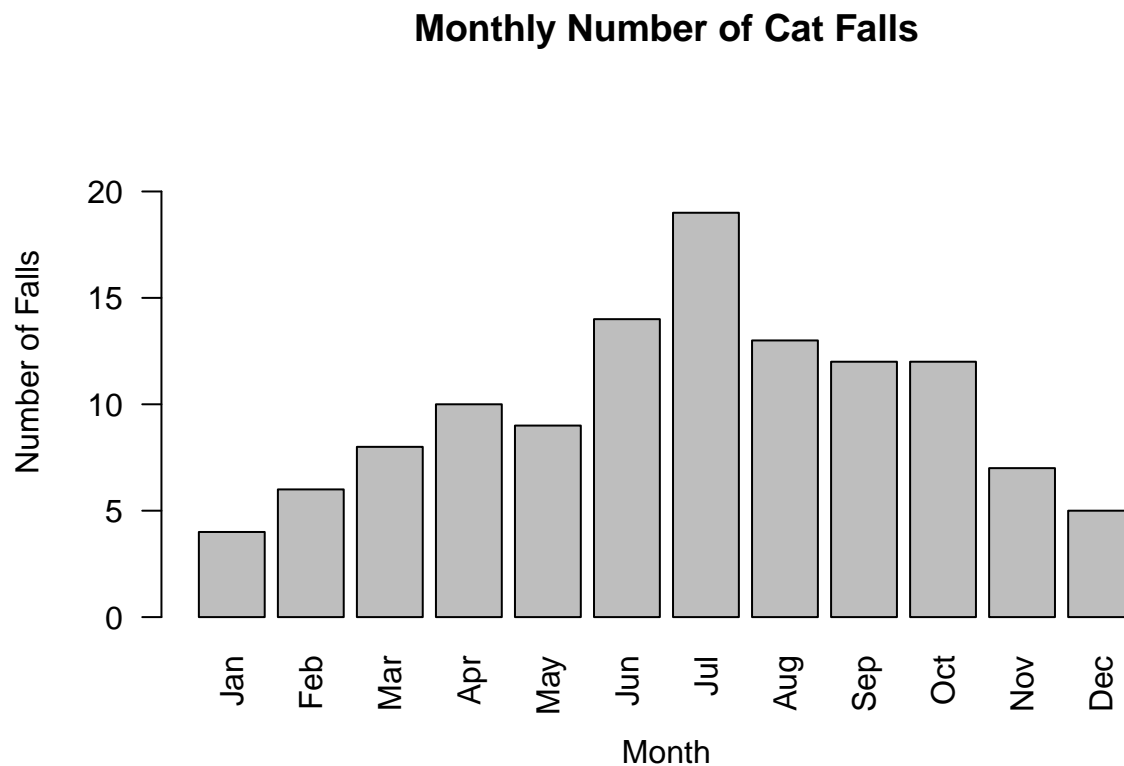
```
# conclusion
# We reject the null hypothesis that the falls occur randomly distributed across
# months of the year with an alpha of 0.05 (Chi-square: X2 = 20.6, df = 11, p =
# 0.00015)

# Repeat the test using R's chisq.test function.
chisq.test(obs)
```

Output

```
# Read in these data and make a bar plot of the results.
```

Monthly Number of Cat Falls


```
# Step 1: State hypotheses
# h0: Falls occur randomly across months.
# ha: Falls are not randomly distributed across months.

# Step 2: State alpha
```

```
## [1] 0.05
```

```
# Step 3: Define test statistic
# chisq = sum((obs - exp)^2/exp)

# Step 4: Calculate the test statistic
```

```
## [1] 20.66387
```

```
# Step 5: Calculate the p-value and make a conclusion
# degrees of freedom
```

```
## [1] 11
```
```
# p-value
```
```
## [1] 0.0370255
```
```
# conclusion
# We reject the null hypothesis that the falls occur randomly distributed across
# months of the year with an alpha of 0.05 (Chi-square: X2 = 20.6, df = 11, p =
# 0.00015)

# Repeat the test using R's chisq.test function.
chisq.test(obs)
```
```
##
##  Chi-squared test for given probabilities
##
## data:  obs
## X-squared = 20.664, df = 11, p-value = 0.03703
```

Answers

Question 2

Script

```r
# The results of such a cross were 10 red-, 21 pink-, and 9
# white-flowered offspring. Make a bar plot that shows both these observed
# values and the corresponding expected values.
obs <- c(10,21,9)
p_null <- c(1/4, 2/4, 1/4)
exp <- p_null * sum(obs)
mat <- t(data.frame(obs = obs, exp = exp))
barplot(mat, beside = T,
        main= "Flower Offspring Colors",
        xlab = "Flower Color",
        ylab = "Number of Flowers",
        names.arg=c("Red","Pink","White"),
        col = c("skyblue","orange"),
        legend = c("Observed","Expected"),
        ylim = c(0,max(c(obs, exp)) + 5))

# Test the hypothesis that the true ratio is 1:2:1. Do this test "by hand" in R.
# You do not need to explicitly perform the five basic steps of hypothesis
# testing, but you do need to clearly state the conclusion of the test.
chisq <- sum((obs - exp)^2/exp)
df <- length(obs) - 0 - 1
(p <- 1 - pchisq(chisq, df))

# We fail to reject the null hypothesis that the phenotypes of the offspring are
# in a 1:2:1 ratio (Chi-square: X2 = 0.15, df = 2, p = 0.928)

# Repeat the test using R's chisq.test function.
chisq.test(obs, p = p_null)
```

```r
# In another, larger experiment, you count 100 times as many flowers as in the
# experiment above, and get 1000 red, 2100 pink, and 900 white offspring. Repeat
# the hypothesis test with these data, using chisq.test.
obs <- c(1000,2100,900)
p_null <- c(1/4, 2/4, 1/4)
exp <- p_null * sum(obs)
chisq.test(obs,p = p_null)

# We reject the null hypothesis that the phenotypes of the offspring are
# in a 1:2:1 ratio (Chi-square: X2 = 15, df = 2, p = 0.00055)

# Do the proportions observed in the two experiments differ? Did the results of
# the two hypothesis tests differ? Briefly explain why or why not.

# The proportions observed do not differ. However, the results of the hypothesis
# tests did differ. They differ because the higher number of observations makes
# it more apparent that the true proportions of the data are not 1:2:1. While
# it's somewhat probable to observe the proportions observed in the first
# experiment even if the true ratio is 1:2:1, it's less probable that the same
# proportion would be observed with so many samples.
```
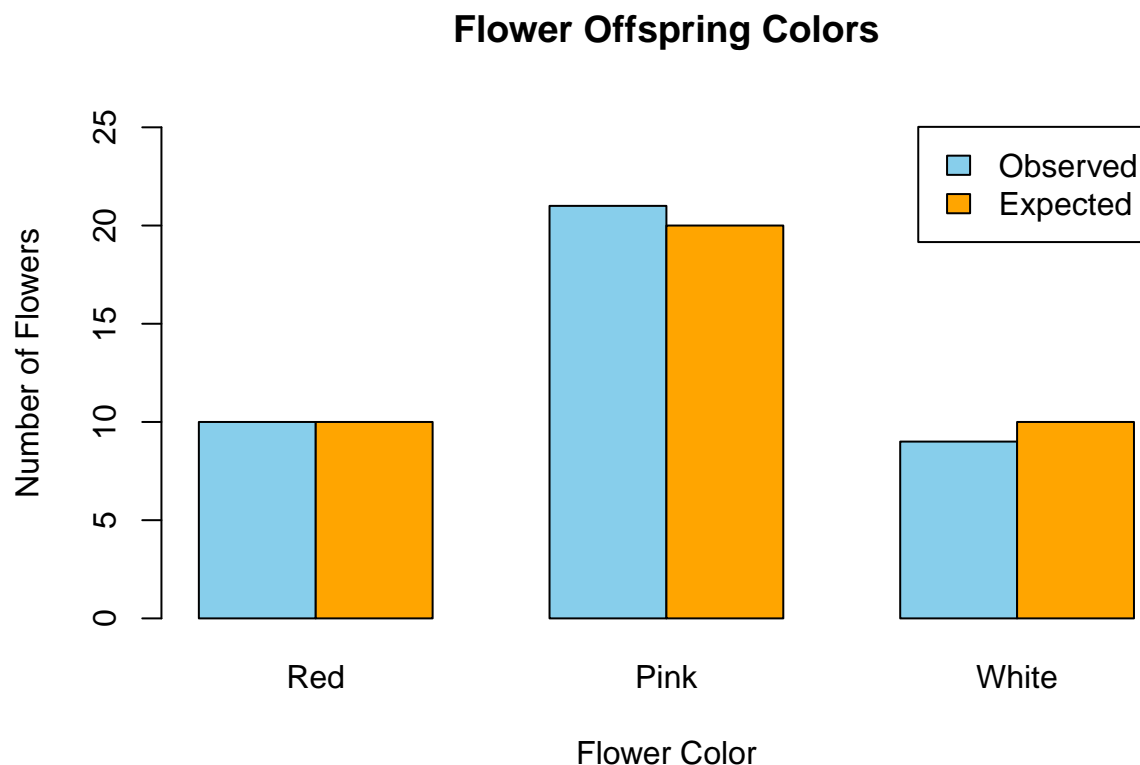
Output

```r
# The results of such a cross were 10 red-, 21 pink-, and 9
# white-flowered offspring. Make a bar plot that shows both these observed
# values and the corresponding expected values.
```


Flower Offspring Colors

```
# Test the hypothesis that the true ratio is 1:2:1. Do this test "by hand" in R.
# You do not need to explicitly perform the five basic steps of hypothesis
# testing, but you do need to clearly state the conclusion of the test.
```

```
## [1] 0.9277435
```

```
# We fail to reject the null hypothesis that the phenotypes of the offspring are
# in a 1:2:1 ratio (Chi-square: X2 = 0.15, df = 2, p = 0.928)

# Repeat the test using R's chisq.test function.
```

```
##
##  Chi-squared test for given probabilities
##
## data:  obs
## X-squared = 0.15, df = 2, p-value = 0.9277
```

```
# In another, larger experiment, you count 100 times as many flowers as in the
# experiment above, and get 1000 red, 2100 pink, and 900 white offspring. Repeat
# the hypothesis test with these data, using chisq.test.
```

```
##
##  Chi-squared test for given probabilities
##
## data:  obs
## X-squared = 15, df = 2, p-value = 0.0005531
```

```
# We reject the null hypothesis that the phenotypes of the offspring are
# in a 1:2:1 ratio (Chi-square: X2 = 15, df = 2, p = 0.00055)

# Do the proportions observed in the two experiments differ? Did the results of
# the two hypothesis tests differ? Briefly explain why or why not.

# The proportions observed do not differ. However, the results of the hypothesis
# tests did differ. They differ because the higher number of observations makes
# it more apparent that the true proportions of the data are not 1:2:1. While
# it's somewhat probable to observe the proportions observed in the first
# experiment even if the true ratio is 1:2:1, it's less probable that the same
# proportion would be observed with so many samples.
```

Answers

Question 3

Script

```
# Estimate the proportion of parasitized pups in this population of armadillos.
parasitized <- c(0,1,2,3,4)
numdens <- c(12,25,23,19,17)
(phat <- sum(parasitized * numdens) / sum(numdens * 4))

# Consider the hypothesis that the number of parasitized pups in each burrow
# follows a binomial distribution. Calculate how many burrows of each type are
# expected if this hypothesis is true.
```

```r
p_null <- dbinom(parasitized, size = max(parasitized), p = phat)
(exp <- p_null * sum(numdens))

# Make a bar plot that shows both the observed values and the corresponding
# expected values that you just calculated.
mat <- t(data.frame(obs = numdens, exp = exp))
barplot(mat, beside = T,
        main= "Number of Parasitized Pups in Armadillo Dens",
        xlab = "Number of Parasitized Pups",
        ylab = "Number of Dens",
        names.arg=parasitized,
        col = c("skyblue","orange"),
        legend = c("Observed","Expected"),
        ylim = c(0,max(c(numdens, exp)) + 5))

# Use the expected values you calculated to test the hypothesis that the number
# of parasitized pups follows a binomial distribution. Carry out an appropriate
# hypothesis test "by hand" in R. You do not need to explicitly perform the five
# basic steps of hypothesis testing, but you do need to clearly state the
# conclusion of the test.
chisq <- sum((numdens - exp)^2/exp)
df <- length(parasitized) - 1 - 1
1 - pchisq(chisq, df = df)

#We reject the null hypothesis that the number of parasitized pups in
#armadillo dens is binomially distributed (Chi-square: X2 = 30.78297, df = 3,
#p = .0000009).

# If the number of parasitized pups is not binomially distributed, suggest a
# possible explanation (i.e., what assumption of the binomial distribution is
# violated and why)?

#The binomial distribution assumes each trial is independent. However, when one
#armadillo in a den is parasitized, it's in close proximity to the other pups in
#the den, meaning they are more likely to all become parasitized than if the
#parasitization was independent. We also see this in the bar plot above; our
#data is enriched for dens with no parasitized pups and dens with all 4
#parasitized pups.
```

Output

```r
# Estimate the proportion of parasitized pups in this population of armadillos.
```

```
## [1] 0.5104167
```

```r
# Consider the hypothesis that the number of parasitized pups in each burrow
# follows a binomial distribution. Calculate how many burrows of each type are
# expected if this hypothesis is true.
```

```
## [1]  5.515409 23.000430 35.968757 24.999561  6.515843
```

```r
# Make a bar plot that shows both the observed values and the corresponding
# expected values that you just calculated.
```

Number of Parasitized Pups in Armadillo Dens
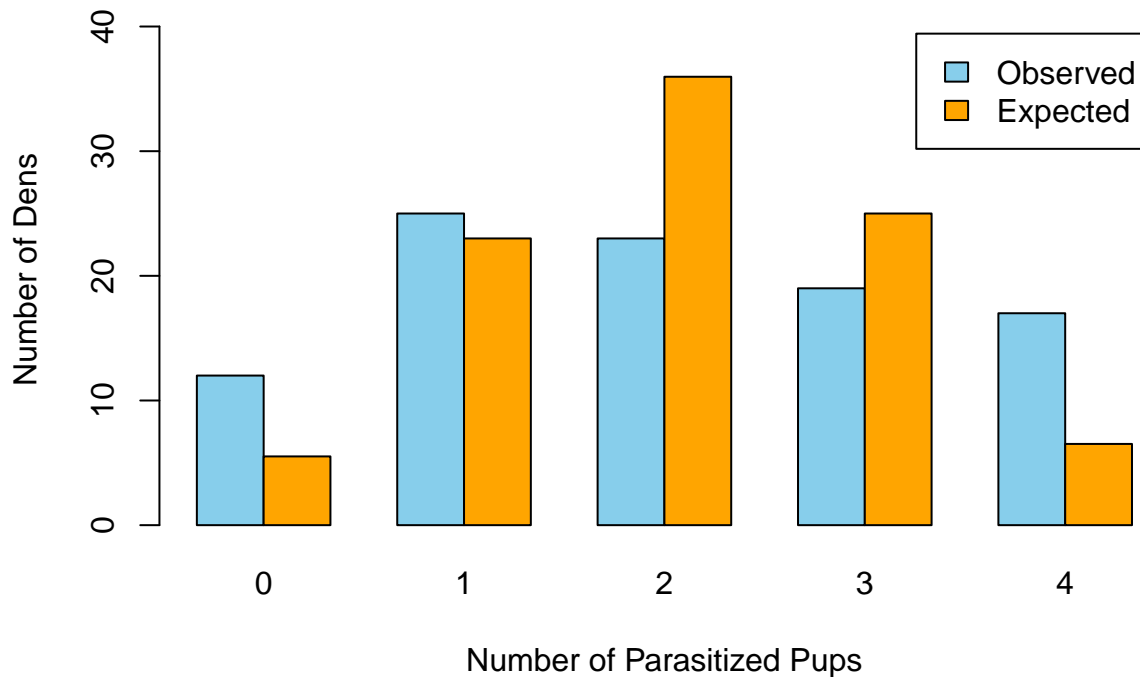

```r
# Use the expected values you calculated to test the hypothesis that the number
# of parasitized pups follows a binomial distribution. Carry out an appropriate
# hypothesis test "by hand" in R. You do not need to explicitly perform the five
# basic steps of hypothesis testing, but you do need to clearly state the
# conclusion of the test.
```

```
## [1] 9.443563e-07
```

```r
#We reject the null hypothesis that the number of parasitized pups in
#armadillo dens is binomially distributed (Chi-square: X2 = 30.78297, df = 3,
#p = .0000009).

# If the number of parasitized pups is not binomially distributed, suggest a
# possible explanation (i.e., what assumption of the binomial distribution is
# violated and why)?

#The binomial distribution assumes each trial is independent. However, when one
#armadillo in a den is parasitized, it's in close proximity to the other pups in
#the den, meaning they are more likely to all become parasitized than if the
#parasitization was independent. We also see this in the bar plot above; our
#data is enriched for dens with no parasitized pups and dens with all 4
#parasitized pups.
```

Answers