Midterm

Adam Orr

October 12, 2017

Section: 91973 Friday 9am

Question 1

Script

```
# a. What is the mean number of yellow-coated mice the geneticist expects to see
# in the litter?
n <- 11
p <- 2/3
(mu <- n * p)

# b. What is the standard deviation of the number of yellow-coated mice in the
# litter?
(sigma <- sqrt(n * p * (1 - p)))

# c. What is the variance of the number of yellow-coated mice in the litter?
(sigma2 <- n * p * (1 - p))

# d. What is the probability that he will see exactly 3 yellow-coated mice in
# the litter?
dbinom(3, n, p)

# e. What is the probability that he will see 5 or more yellow-coated mice in
# the litter?
1 - pbinom(5,n,p)
```

Output

```
# a. What is the mean number of yellow-coated mice the geneticist expects to see
# in the litter?
```

```
## [1] 7.333333
```

```
# b. What is the standard deviation of the number of yellow-coated mice in the
# litter?
```

```
## [1] 1.563472
```

```
# c. What is the variance of the number of yellow-coated mice in the litter?
```

```
## [1] 2.444444
```

```
# d. What is the probability that he will see exactly 3 yellow-coated mice in
# the litter?
```

```
## [1] 0.007451439
```

```r
# e. What is the probability that he will see 5 or more yellow-coated mice in
# the litter?
```

```
## [1] 0.877915
```

Answers

Question 2

Script

```r
#input data
x <- c(38.51, 38.36, 38.23, 38.49, 38.51, 38.39, 38.31, 38.35, 38.59, 38.48,
       38.44, 38.40, 38.30, 38.29, 38.51, 38.41, 38.37, 38.43, 38.50, 38.38,
       38.50, 38.33, 38.46, 38.36, 38.41)

#estimate mean of x
(xbar <- sum(x) / length(x))

#estimate variance of x
(s2 <- sum((x - xbar)^2)/(length(x) - 1))

#estimate standard deviation of x
(s <- sqrt(s2))

#estimate standard error of the mean
(sem <- s / sqrt(length(x)))

#approximate 95% confidence interval of the mean
c(xbar - 2 * sem, xbar + 2 * sem)

#make a plot that includes both a histogram of the data and a plot of the
#estimated probability density function of X.

#plot histogram
hist(x, xlab = "Ostrich Temperature (degrees C)",
     ylab = "Probability Density",
     main = "Distribution of Ostrich Temperatures",
     xlim = c(38.1,38.7),
     ylim = c(0,5),
     freq = FALSE)

#add a curve showing estimate of the pdf
xvals<-seq(xbar-4*s, xbar+4*s, by = .01)
pdf <- dnorm(xvals,xbar,s)
lines(pdf~xvals, col='red')
legend("topright",
       legend = c("Histogram", "PDF"),
       col = c('black','red'),
       lty = 1,
       bty = "n")
```

Output

```r
#estimate mean of x
```

```
## [1] 38.4124
```

```r
#estimate variance of x
```

```
## [1] 0.007519
```

```r
#estimate standard deviation of x
```

```
## [1] 0.08671217
```

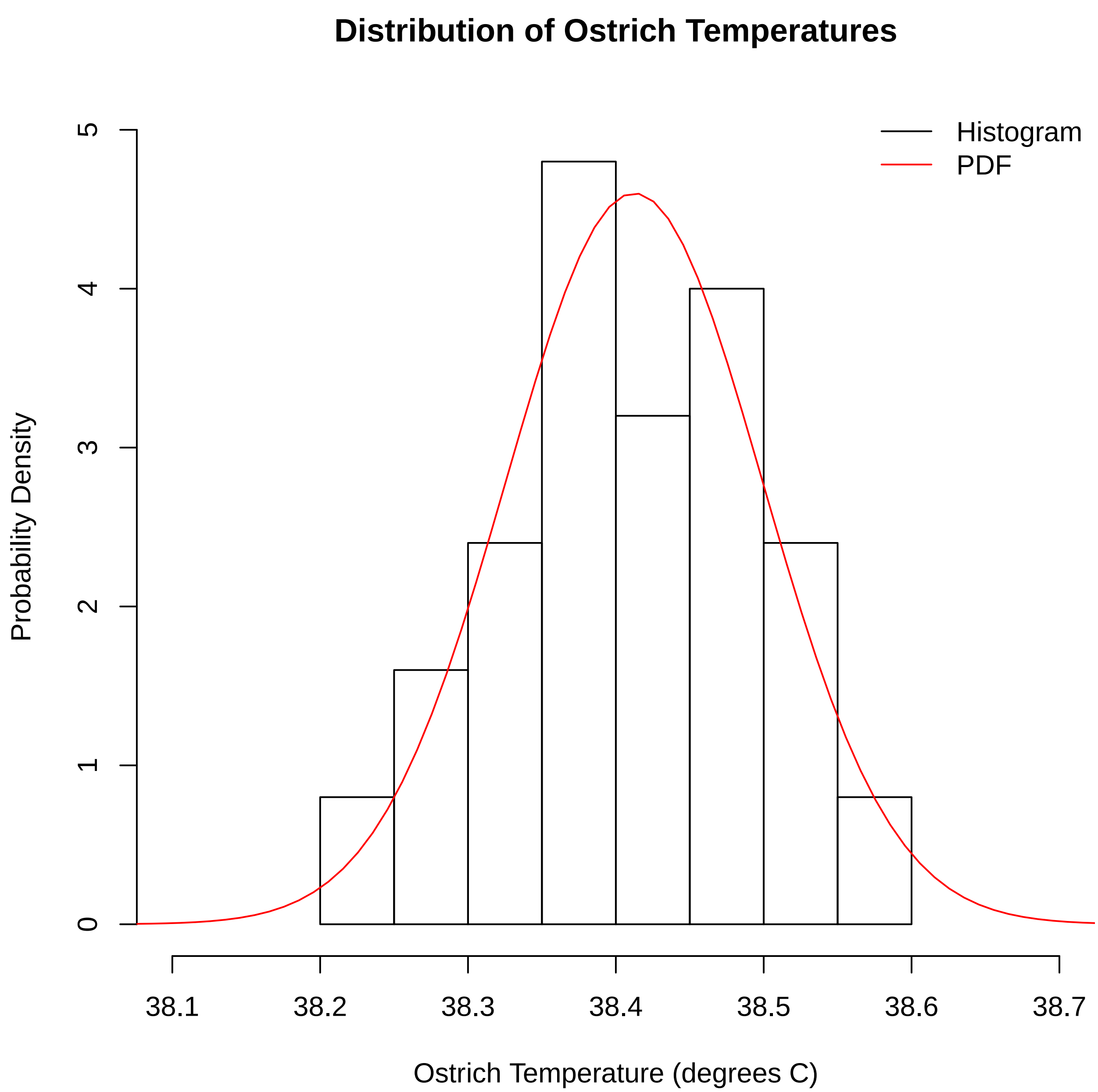
```r
#estimate standard error of the mean
```

```
## [1] 0.01734243
```

```r
#approximate 95% confidence interval of the mean
```

```
## [1] 38.37772 38.44708
```

```r
#make a plot that includes both a histogram of the data and a plot of the
#estimated probability density function of X.
```



Answers

Question 3

Script

```
#birth weight parameters
mu <- 3.8 #kg
sigma <- 0.7 #kg

# a. What proportion of birth weights are expected to exceed 4.5 kg?
1 - pnorm(4.5, mu, sigma)
```

```r
# b. 80% of birth weights are expected to be below what value?
qnorm(.8, mu, sigma)

# c. What proportion of birth weights are expected to be between 2.5 and 3.0 kg?
pnorm(3, mu, sigma) - pnorm(2.5, mu, sigma)
```

Output

```r
# a. What proportion of birth weights are expected to exceed 4.5 kg?
```

```
## [1] 0.1586553
```

```r
# b. 80% of birth weights are expected to be below what value?
```

```
## [1] 4.389135
```

```r
# c. What proportion of birth weights are expected to be between 2.5 and 3.0 kg?
```

```
## [1] 0.09490354
```

Answers

Question 4

Script

```r
#input data
x <- 43
n <- 300

# a. Use these data to estimate the proportion of shoppers in the population who
# have injured themselves with their food and drink packaging.
(phat <- x / n)

# b. Calculate the 95% confidence interval of the proportion you estimated
# above. Use the Agresti-Coull method.
pgrave <- (x + 2)/(n + 4)
se <- 2 * sqrt(pgrave * (1 - pgrave) / (n + 4))
c(pgrave - 2 * se, pgrave + 2 * se)

# c. If the study had included only 30 shoppers, would you expect the confidence
# interval to be larger, smaller, or about the same?

#If the study had included only 30 shoppers, I would expect the confidence
#interval to be much larger. A smaller sample size means the estimate of p is
#less precise. This is apparent in the equation for the magnitude of the
#interval in each direction; it is divided by n, so a smaller n produces a wider
#interval.
```

Output

```
# a. Use these data to estimate the proportion of shoppers in the population who
# have injured themselves with their food and drink packaging.
```

```
## [1] 0.1433333
```

```
# b. Calculate the 95% confidence interval of the proportion you estimated
# above. Use the Agresti-Coull method.
```

```
## [1] 0.06655481 0.22949782
```

```
# c. If the study had included only 30 shoppers, would you expect the confidence
# interval to be larger, smaller, or about the same?

#If the study had included only 30 shoppers, I would expect the confidence
#interval to be much larger. A smaller sample size means the estimate of p is
#less precise. This is apparent in the equation for the magnitude of the
#interval in each direction; it is divided by n, so a smaller n produces a wider
#interval.
```

Answers

Question 5

Script

```
#input data
categories <- c('Vertical', '20 Degrees', '40 Degrees')
obs <- c(30,15,8)

#state null and alternative hypothesis
#H0: The number of bird deaths is randomly distributed between 3 window angles
#Ha: The number of bird deaths is not randomly distributed between 3 window angles.

#state significance level
#alpha = 0.05

#state conclusion of test including name, value of test statistic, p-value, and
#sample size or degrees of freedom
chisq.test(obs)

#We reject the null hypothesis that the number of bird deaths is randomly
#distributed across the 3 window angles with alpha of 0.05 (Chi-squared: X2 =
#14.302, df = 2, p = 0.00078)

#plot observed and expected distribution
exp <- sum(obs) / length(obs)
mat <- t(data.frame(obs = obs, exp = exp))
barplot(mat, beside = T,
        main= "Bird Deaths By Windows of Various Angles",
        xlab = "Window Angle",
        ylab = "Number of Deaths",
```

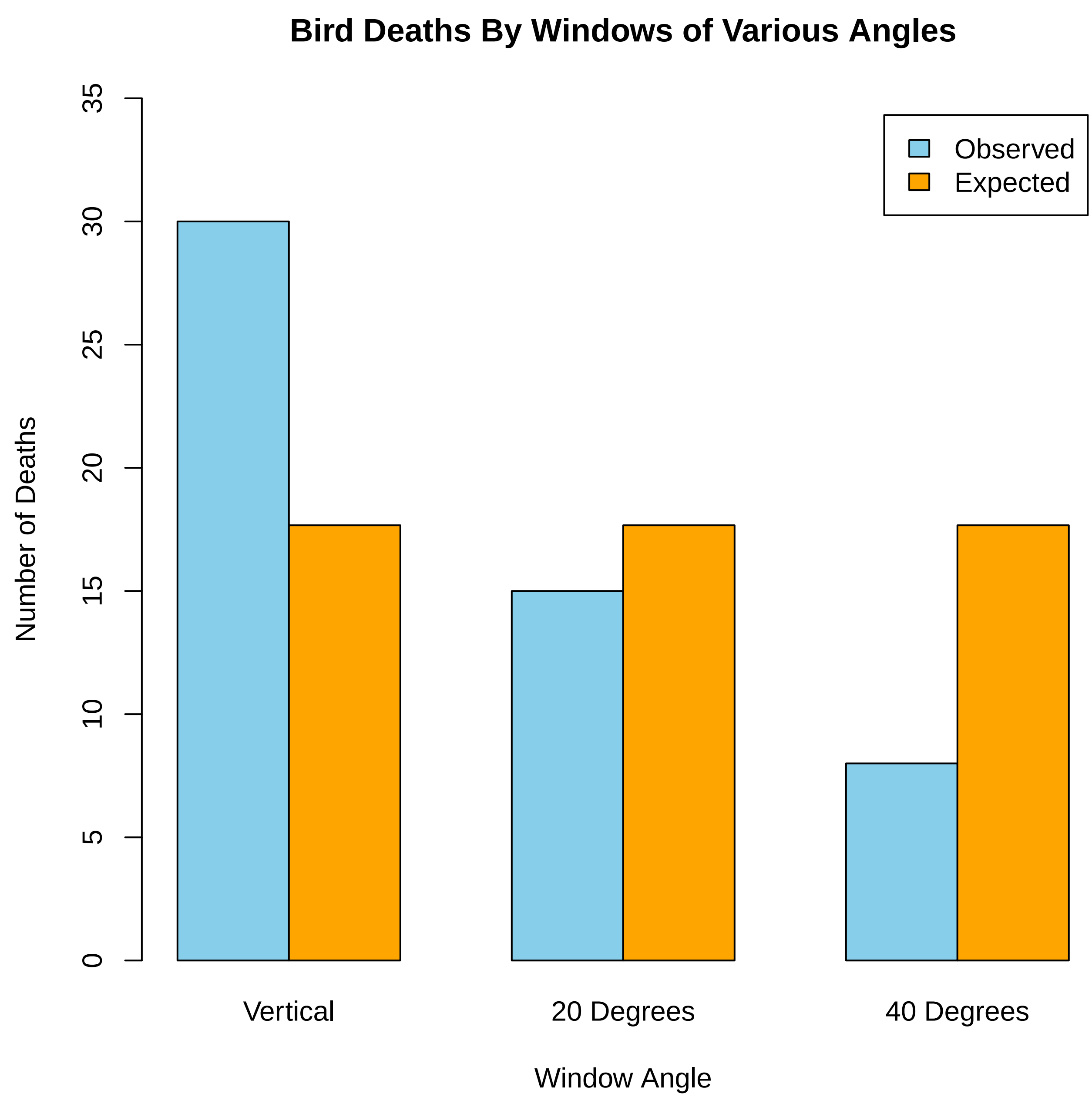
```
        names.arg=categories,
        col = c("skyblue","orange"),
        legend = c("Observed","Expected"),
        ylim = c(0,max(c(obs, exp)) + 5))
```

Output

```
#state null and alternative hypothesis
#H0: The number of bird deaths is randomly distributed between 3 window angles
#Ha: The number of bird deaths is not randomly distributed between 3 window angles.
#state significance level
#alpha = 0.05
#state conclusion of test including name, value of test statistic, p-value, and
#sample size or degrees of freedom
```

```
##
##  Chi-squared test for given probabilities
##
## data:  obs
## X-squared = 14.302, df = 2, p-value = 0.0007841
```

```
#We reject the null hypothesis that the number of bird deaths is randomly
#distributed across the 3 window angles with alpha of 0.05 (Chi-squared: X2 =
#14.302, df = 2, p = 0.00078)
#plot observed and expected distribution
```



Answers

Question 6

Script

```
#input data
numfemales <- c(0:3)
numterritories <- c(60,130,10,0)

#state null and alternative hypothesis
```

```r
#HO: The number of females in territories having three fish is binomially
#distributed.
#Ha: The number of females in territories having three fish is not
#binomially distributed.

#state significance level
#alpha = 0.05

#state conclusion of test including name, value of test statistic, p-value, and
#sample size or degrees of freedom

#estimate parameters
phat <- sum(numfemales * numterritories)/sum(numterritories * 4)
pnull <- dbinom(numfemales, size = 3, p = phat)
exp <- pnull * sum(numterritories)

#calculate statistic
(chisq <- sum((numterritories - exp)^2/exp))

#degrees of freedom
(df <- 4 - 1 - 1)

#p-value
(p <- 1 - pchisq(chisq, df = df))

#We reject the null hypothesis that the number of females in territories is
#binomially distributed (Chi-square: X2 = 66.949, df = 2, p = 2.8e-15).

#plot observed and expected distribution
mat <- t(data.frame(obs = numterritories, exp = exp))
barplot(mat, beside = T,
        main = "Number of Female Anemonefish\nin Territories of Three Fish",
        xlab = "Number of Females",
        ylab = "Number of Territories",
        names.arg=numfemales,
        col = c("skyblue","orange"),
        legend = c("Observed","Expected"),
        ylim = c(0,max(c(numterritories, exp)) + 20))
```

Output

```r
#state null and alternative hypothesis
#HO: The number of females in territories having three fish is binomially
#distributed.
#Ha: The number of females in territories having three fish is not
#binomially distributed.

#state significance level
#alpha = 0.05

#state conclusion of test including name, value of test statistic, p-value, and
#sample size or degrees of freedom
```

```
#calculate statistic
```

```
## [1] 66.9488
```

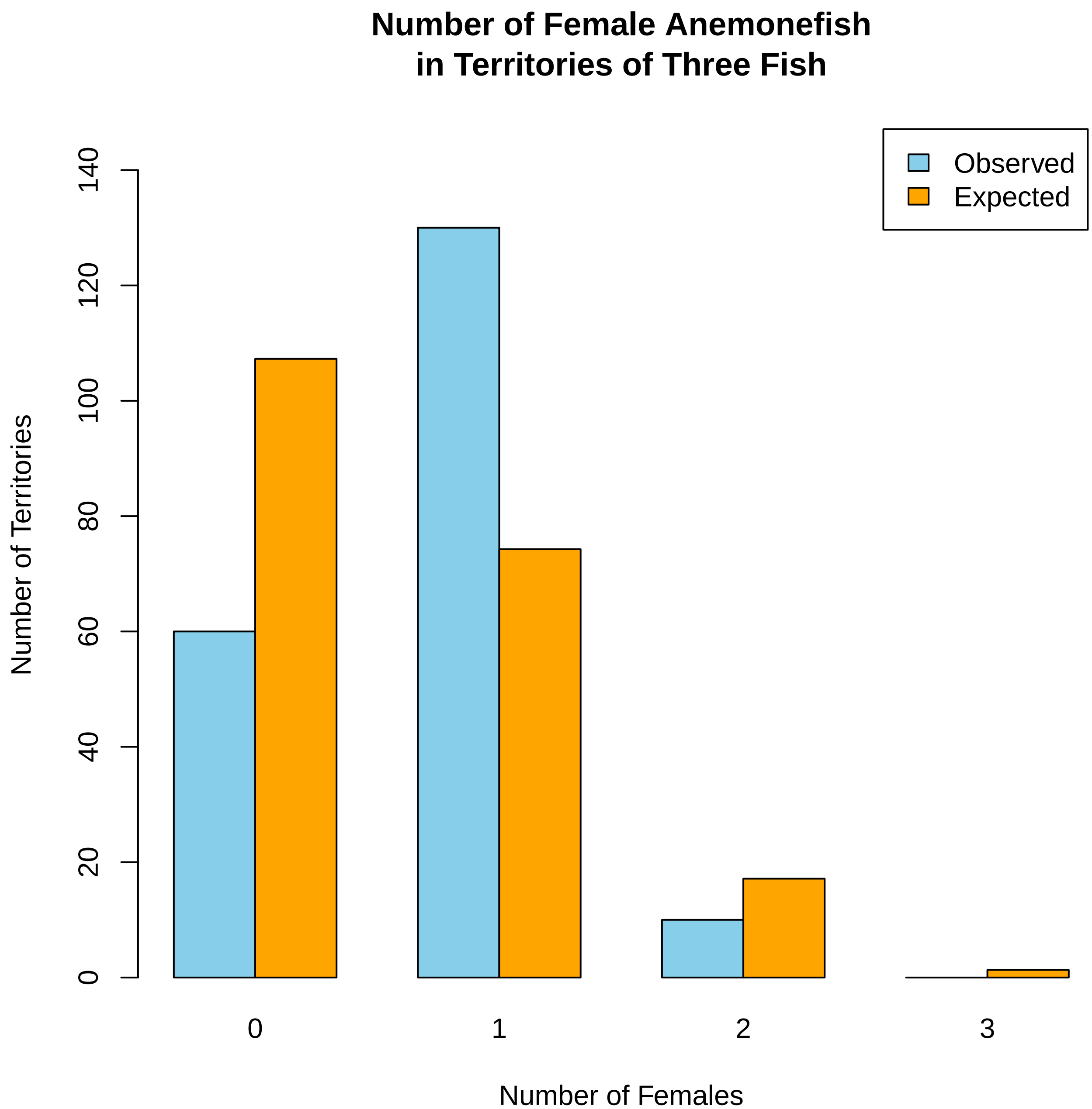
```
#degrees of freedom
```

```
## [1] 2
```

```
#p-value
```

```
## [1] 2.88658e-15
```

```
#We reject the null hypothesis that the number of females in territories is
#binomially distributed (Chi-square: X2 = 66.949, df = 2, p = 2.8e-15).

#plot observed and expected distribution
```



Answers

Question 7

Script

```
#input data
x <- 6101 #num successes
n <- 9821 #num trials

#state null and alternative hypothesis
#H0: The probability of toast landing butter-side down is .5
#Ha: The probability of toast landing butter-side down is greater than .5

#state significance level
#alpha = 0.05

#state conclusion of test including name, value of test statistic, p-value, and
#sample size or degrees of freedom
binom.test(x,n, alternative = 'greater')

#We reject the null hypothesis that the probability of landing butter-side down
#is .5 (Binomial test: X = 6101, n = 9821, p = 2.2e-16).

#plot results
obs <- c(x, n-x)
exp <- rep(n * 0.5,2)
mat <- t(data.frame(obs = obs, exp = exp))
barplot(mat, beside = T,
        main = "Butter Orientation of 9821 Toast Flips",
        xlab = "Trial Outcome",
        ylab = "Number of Landings",
        names.arg=c("Butter Side Down","Butter Side Up"),
        col = c("skyblue","orange"),
        legend = c("Observed","Expected"),
        ylim = c(0,max(c(x, exp)) + 1000))
```

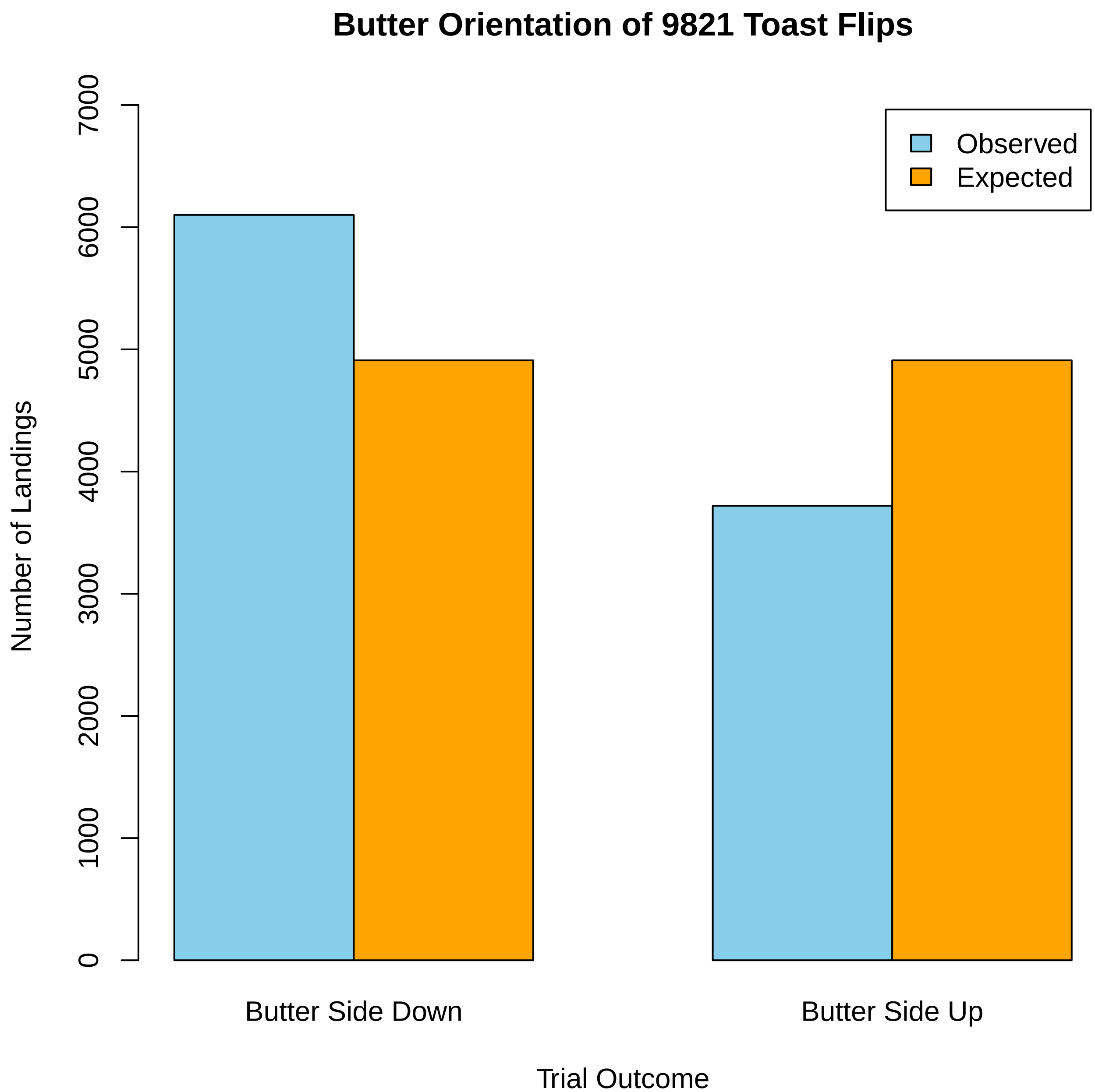
Output

```
#state null and alternative hypothesis
#H0: The probability of toast landing butter-side down is .5
#Ha: The probability of toast landing butter-side down is greater than .5

#state significance level
#alpha = 0.05

#state conclusion of test including name, value of test statistic, p-value, and
#sample size or degrees of freedom
```

```
##
##  Exact binomial test
```

```
##
## data:  x and n
## number of successes = 6101, number of trials = 9821, p-value <
## 2.2e-16
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.6130918 1.0000000
## sample estimates:
## probability of success
##              0.6212198
```

```
#We reject the null hypothesis that the probability of landing butter-side down
#is .5 (Binomial test: X = 6101, n = 9821, p = 2.2e-16).

#plot results
```



Answers