Chapter 6 Population Proportions and the Binomial Distribution

Angelo LaCommare-Soto

2025-03-05

Avian Alarm Calls in the Tropics

Research Question 1: Does the presence of a predator increases the proportion of alarm calls produced by sentinel birds?

Section 1 - Importing Data

```
# set working directory for all chunks in this file (default working directory is wherever Rmd file is)
getwd()
```

```
## [1] "C:/Users/Angelo L/Documents/GitHub/BIOL710/RCode710/RCode/working_directory"
```

```
library(tidyverse)
```

```
## Warning: package 'purrr' was built under R version 4.4.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# Importing data
alarm <- read.csv("alarm.csv",header=TRUE)

# viewing the structure of the data
str(alarm)
```

```
## 'data.frame':    369 obs. of  28 variables:
##  $ SEQ                    : int  360 361 284 266 267 281 282 264 265 279 ...
##  $ DATE                   : chr  "18-Jun-16" "18-Jun-16" "19-Jul-16" "18-Jul-16" ...
##  $ TIME                   : chr  "15:54" "15:54" "14:33" "13:52" ...
##  $ FLOCKNAME              : chr  "GY" "GY" "YO" "TNT" ...
```

```
##  $ FLOCK                    : int  6 6 16 14 14 1 1 13 13 4 ...
##  $ TREATMENT                : chr  "PAUN FAR" "PAUN FAR" "ACBI FAR" "PAUN NEAR" ...
##  $ RAPTOR                   : chr  "PAUN" "PAUN" "ACBI" "PAUN" ...
##  $ FOCALBIRD                : chr  "THAR" "THAR" "THSY" "THAR" ...
##  $ SEX                      : chr  "M" "M" "M" "M" ...
##  $ VEG                      : chr  "OPEN" "OPEN" "OPEN" "OPEN" ...
##  $ HEIGHT                   : num  4 4 2 1 1 1 1 8 8 6 ...
##  $ PREDDIST                 : num  16 16 11 1.5 1.5 7 7 4 4 7 ...
##  $ PRED.ANGLE               : int  45 45 80 0 0 90 90 0 0 80 ...
##  $ OBS.DIST                 : int  17 17 12 11 11 7 7 8 8 13 ...
##  $ PRED.PERCH.HEIGHT        : chr  "7" "7" "8" "5" ...
##  $ FB.FLIGHT.ANGLE          : chr  "8" "8" "8" "9" ...
##  $ SPALARMCALL              : chr  "THAR" "THAR" "THSY" "THAR" ...
##  $ ALARM.CALLER.DISTANCE    : chr  NA NA NA NA ...
##  $ FLIGHT.DISTANCE          : num  3 3 14 13 13 15 15 4 4 8 ...
##  $ Alarm_Call               : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ NoAlarmCall              : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Alarm.Latency.Focal.Bird: chr  NA "1.172" NA NA ...
##  $ Focal.Bird               : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ NonFocal.Bird            : int  0 0 0 NA 0 0 0 0 0 0 ...
##  $ TRIAL                    : chr  "BEFORE" "POST" "BEFORE" "BEFORE" ...
##  $ Alarm.Total              : int  0 2 0 0 0 0 0 0 1 11 ...
##  $ Urgent.Total             : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ SIZE                     : chr  "large" "large" "small" "large" ...
```

```r
head(alarm)
```

```
##   SEQ      DATE  TIME FLOCKNAME FLOCK TREATMENT RAPTOR FOCALBIRD SEX    VEG
## 1 360 18-Jun-16 15:54        GY     6  PAUN FAR   PAUN      THAR   M   OPEN
## 2 361 18-Jun-16 15:54        GY     6  PAUN FAR   PAUN      THAR   M   OPEN
## 3 284 19-Jul-16 14:33        YO    16  ACBI FAR   ACBI      THSY   M   OPEN
## 4 266 18-Jul-16 13:52       TNT    14 PAUN NEAR   PAUN      THAR   M   OPEN
## 5 267 18-Jul-16 13:52       TNT    14 PAUN NEAR   PAUN      THAR   M   OPEN
## 6 281 19-Jul-16 13:40        AH     1  ACBI MED   ACBI      THSY   F COVERD
##   HEIGHT PREDDIST PRED.ANGLE OBS.DIST PRED.PERCH.HEIGHT FB.FLIGHT.ANGLE
## 1      4     16.0         45       17                 7               8
## 2      4     16.0         45       17                 7               8
## 3      2     11.0         80       12                 8               8
## 4      1      1.5          0       11                 5               9
## 5      1      1.5          0       11                 5               9
## 6      1      7.0         90        7                 6               2
##   SPALARMCALL ALARM.CALLER.DISTANCE FLIGHT.DISTANCE Alarm_Call NoAlarmCall
## 1        THAR                  <NA>               3          0           1
## 2        THAR                  <NA>               3          0           1
## 3        THSY                  <NA>              14          0           1
## 4        THAR                  <NA>              13          0           1
## 5        THAR                  <NA>              13          0           1
## 6        THSY                  <NA>              15          0           1
##   Alarm.Latency.Focal.Bird Focal.Bird NonFocal.Bird  TRIAL Alarm.Total
## 1                     <NA>          1             0 BEFORE           0
## 2                    1.172          1             0   POST           2
## 3                     <NA>          1             0 BEFORE           0
## 4                     <NA>          1            NA BEFORE           0
## 5                     <NA>          1             0   POST           0
```

```
## 6                     <NA>          1             0 BEFORE           0
##   Urgent.Total  SIZE
## 1            0 large
## 2            0 large
## 3            0 small
## 4            0 large
## 5            0 large
## 6            0 small
```

Question Answers

a. The 'alarm' dataset has 369 observations of 28 variables.
b. Given the question 'Does the presence of a predator increases the proportion of alarm calls produced by sentinel birds?', we are interested in the 'RAPTOR' and 'Alarm_Call' variables.
c. Considering the response variable of 'Alarm_Call', a success is represented by a 1, which is an alarm call.

Section 2 - Estimating the Binomial Formula

```r
# ensuring the levels of the variable Raptor read as a factor
levels(as.factor(alarm$RAPTOR))
```

```
## [1] "ACBI"    "CONTROL" "FAFE"    "PAUN"
```

```r
# filtering "alarm" for control trials only
control <- filter(alarm,RAPTOR=="CONTROL")

# filtering "alarm" for raptor trials only, syntax of != gives the values that are everything but what
raptor <- filter(alarm,RAPTOR!="CONTROL")
```

```r
# Estimating the binomial formula using raptor dataframe
# number of trials n
n <- length(raptor$RAPTOR)

# number of successes X (alarms)
X <-sum(raptor$Alarm_Call)

# observed proportion of successes in sample population p_hat
p_hat <- X/n
```

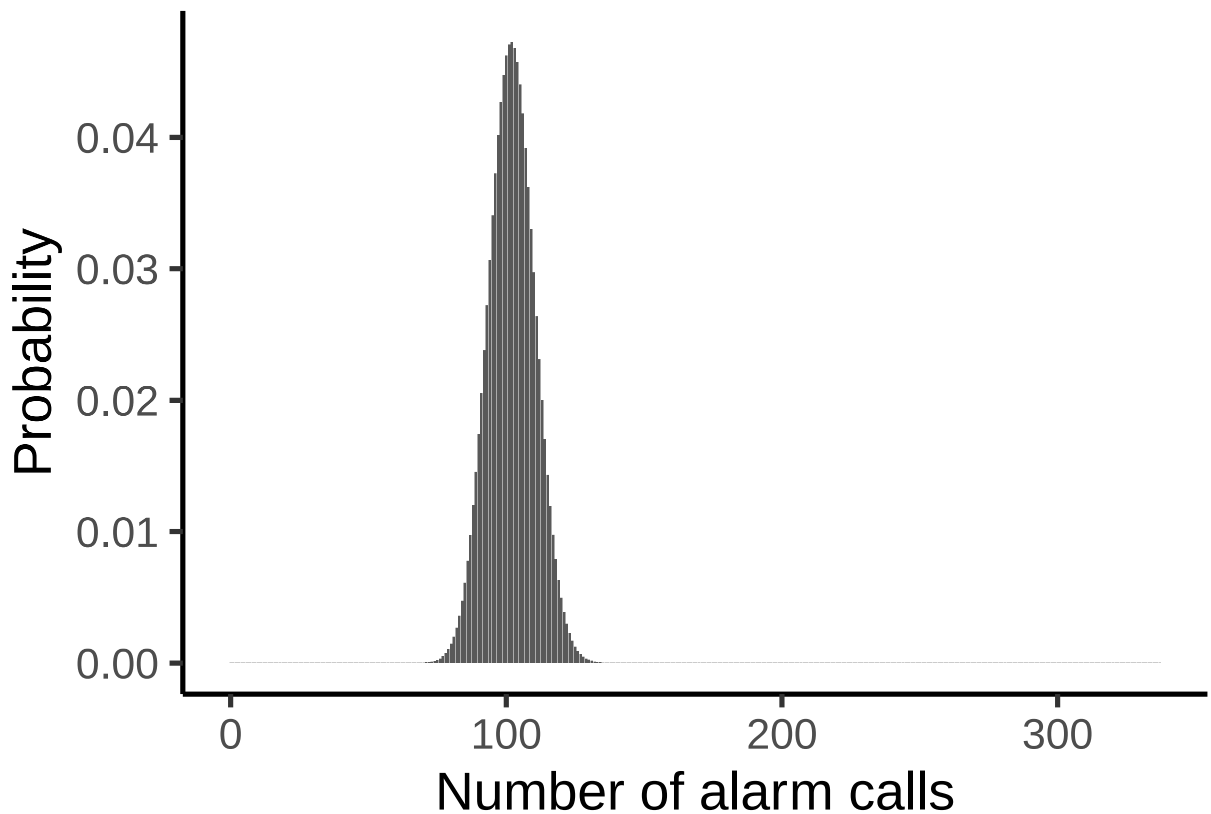
```r
# Estimating the binomial function formula using dbinom()
# example: the probability of getting exactly 100 alarm calls with n=337 and p_hat=0.303.
dbinom(100,n,p_hat)
```

```
## [1] 0.0462047
```

Section 3 - Estimating the Binomial Probability Distribution

```r
# all possible values of X out of n trials
X_success <- 0:n
# X_success

# binomial probabilities (same function as above)
prob_X <- dbinom(X_success, n, p_hat)
# prob_X
```

```r
# generating a dataframe with X_success and binomial probabilities for the plot
probTable <- data.frame(X_success, prob_X)
# probTable

# binomial probability distribution plot
p1 <- ggplot(probTable,aes(x=X_success,y=prob_X)) +
  geom_bar(stat = "identity") +
  ylab("Probability") +
  xlab("Number of alarm calls") +
  theme_classic(20)
p1
```



Question Answers

a. The probability of 120 sentinel birds out of 337 trials producing an alarm call is 0.00497.
b. The number of alarm calls with the highest probability out of the 337 trials is 102.

Section 4 - Estimating the Standard Error of the Proportion

```
# standard error of p as in probability
SE <- sqrt((p_hat*(1-p_hat))/n)
SE
```

```
## [1] 0.02502587
```

Section 5 - Testing the Hypothesis Using the Binomial Test

```
# proportion of birds producing alarm calls in a control trial; note use of two dataframes
p_null <- sum(control$Alarm_Call)/length(control$RAPTOR)
p_null
```

```
## [1] 0.03125
```

Considering our research question, our null and alternative hypotheses are:

H0: In the presence of raptors, 3.125% of sentinel birds produce an alarm call.

HA: In the presence of raptors, the proportion of sentinel birds producing an alarm call is not 3.125%.

```
# binomial test
binom.test(X, n, p_null)
```

```
##
##  Exact binomial test
##
## data:  X and n
## number of successes = 102, number of trials = 337, p-value < 2.2e-16
## alternative hypothesis: true probability of success is not equal to 0.03125
## 95 percent confidence interval:
##  0.2540623 0.3547869
## sample estimates:
## probability of success
##              0.3026706
```

Question Answers

a. The presence of a predator does increase the proportion of alarm calls produced by sentinel birds; the sample probability of alarm calls in the presence of raptors of 0.303 is much higher than the probability of alarm calls in the control treatment of raptor absence of 0.031.
b. Therefore, we can make an argument that raptors are a major driver in shaping the landscape of fear of Amazonian birds.

Research Question 2: Is raptor size a visual cue driving the proportion of sentinel birds producing alarm calls?

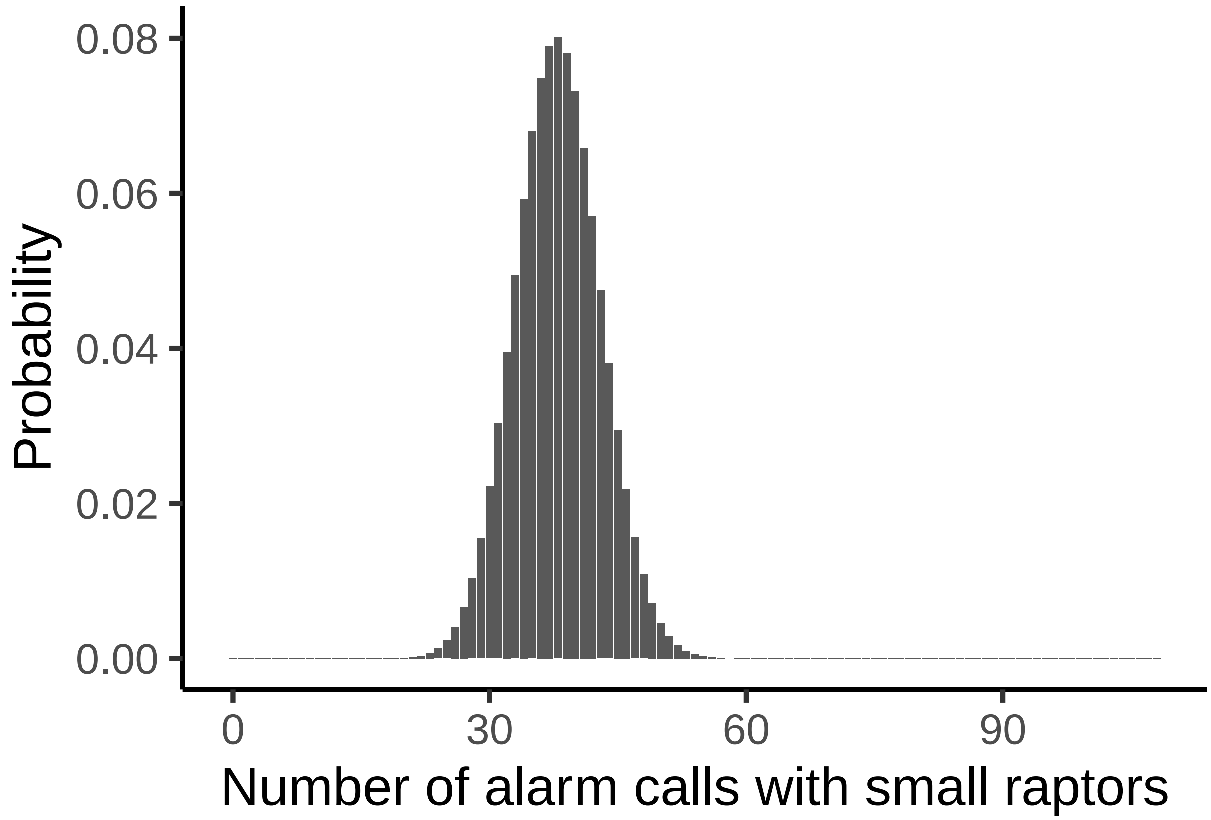
Stop, Think, Do: Answer Research Question 2.

```r
# Creating a new data table for size of raptor and eliminating NA values
size <- filter(raptor, SIZE != 'NA')

# Creating a new data table for small size of raptor
smsize <- filter(size, SIZE == 'small')

# Creating a new data table for large size of raptor
lgsize <- filter(size, SIZE == 'large')
```

```r
# For small raptor size
# Estimating the binomial formula using smsize dataframe
# number of trials n
nsmall <- length(smsize$SIZE)

# number of successes X (alarms)
Xsmall <-sum(smsize$Alarm_Call)

# observed proportion of successes in sample population p_hat
p_hatsmall <- Xsmall/nsmall
```

```r
# For small raptor size
# all possible values of X out of n trials
X_success_small <- 0:nsmall
# X_success_small

# binomial probabilities
prob_Xsmall <- dbinom(X_success_small, nsmall, p_hatsmall)
# prob_Xsmall
```

```r
# For small raptor size
# generating a dataframe with X_success_small and binomial probabilities for the plot
probTable1 <- data.frame(X_success_small, prob_Xsmall)
# probTable1

# binomial probability distribution plot
p2 <- ggplot(probTable1,aes(x=X_success_small,y=prob_Xsmall)) +
  geom_bar(stat = "identity") +
  ylab("Probability") +
  xlab("Number of alarm calls with small raptors") +
  theme_classic(20)
p2
```




```r
# standard error of p as in probability
SE_small <- sqrt((p_hatsmall*(1-p_hatsmall))/nsmall)
SE_small
```

```
## [1] 0.04595208
```

```r
# binomial test
binom.test(Xsmall, nsmall, p_null)
```

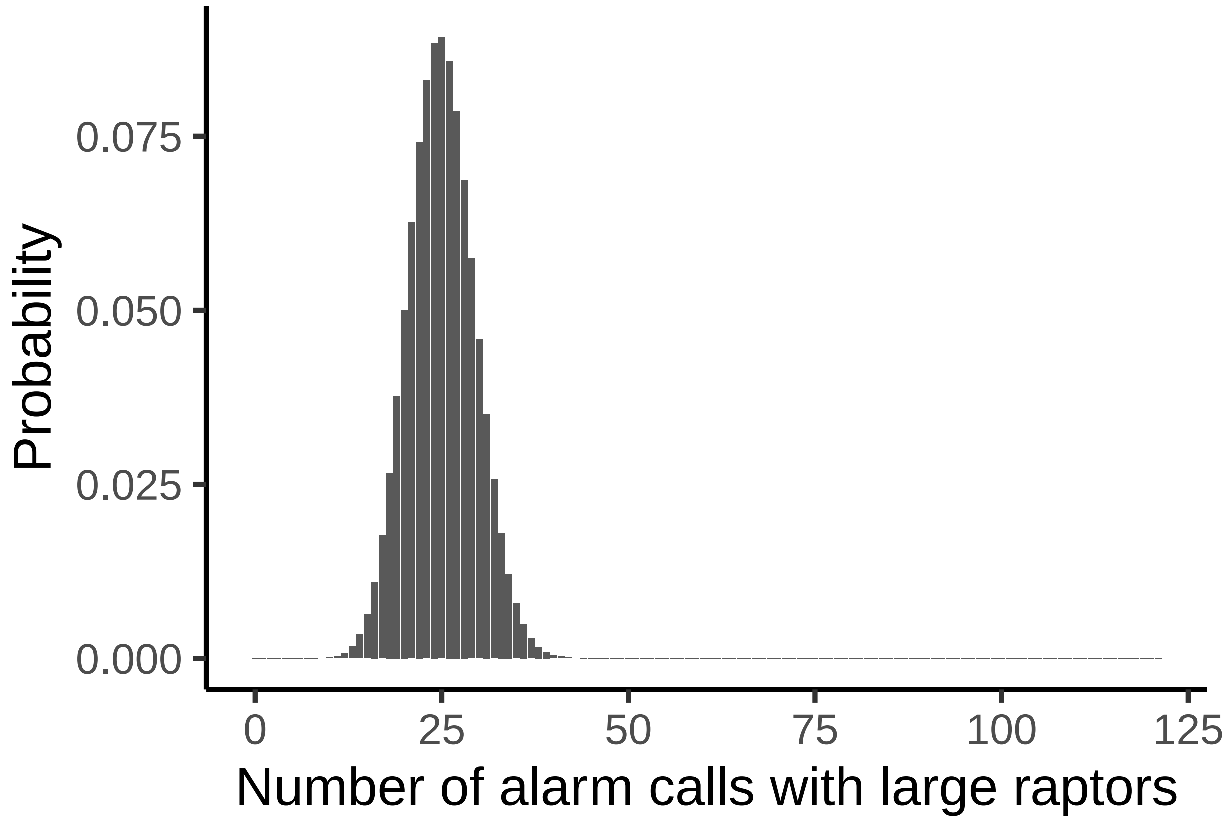
```
##
##  Exact binomial test
##
## data:  Xsmall and nsmall
## number of successes = 38, number of trials = 108, p-value < 2.2e-16
## alternative hypothesis: true probability of success is not equal to 0.03125
## 95 percent confidence interval:
##  0.2624064 0.4496477
## sample estimates:
## probability of success
##              0.3518519
```

```r
# For large raptor size
# Estimating the binomial formula using lgsize dataframe
# number of trials n
```

```r
nlarge <- length(lgsize$SIZE)

# number of successes X (alarms)
Xlarge <-sum(lgsize$Alarm_Call)

# observed proportion of successes in sample population p_hat
p_hatlarge <- Xlarge/nlarge
```

```r
# For large raptor size
# all possible values of X out of n trials
X_success_large <- 0:nlarge
# X_success_large

# binomial probabilities
prob_Xlarge <- dbinom(X_success_large, nlarge, p_hatlarge)
# prob_Xlarge
```

```r
# For large raptor size
# generating a dataframe with X_success_large and binomial probabilities for the plot
probTable2 <- data.frame(X_success_large, prob_Xlarge)
# probTable2

# binomial probability distribution plot
p3 <- ggplot(probTable2,aes(x=X_success_large,y=prob_Xlarge)) +
  geom_bar(stat = "identity") +
  ylab("Probability") +
  xlab("Number of alarm calls with large raptors") +
  theme_classic(20)
p3
```




```
# standard error of p as in probability
SE_large <- sqrt((p_hatlarge*(1-p_hatlarge))/nlarge)
SE_large
```

```
## [1] 0.03680676
```

```
# binomial test
binom.test(Xlarge, nlarge, p_null)
```

```
##
##  Exact binomial test
##
## data:  Xlarge and nlarge
## number of successes = 25, number of trials = 121, p-value = 6.66e-14
## alternative hypothesis: true probability of success is not equal to 0.03125
## 95 percent confidence interval:
##  0.1383954 0.2897084
## sample estimates:
## probability of success
##              0.2066116
```

Discussion Question Answers

a. Successes and failures in binomial trials represent the answer to a yes or no question posed by the researchers.

b. H0: Raptor size is not a visual cue driving the proportion of sentinel birds producing alarm calls. The estimated p-value for the binomial test of small raptors driving the proportion of alarm calls among sentinel birds was <2.2x10^-16. Additionally, the estimated p-value for the binomial test of large raptors driving the proportion of alarm calls among sentinel birds was 6.66x10^-14. Given the statistical significance of both binomial tests, raptor size does not drive the proportion of alarm calls in sentinel birds.
c. For the small raptors, the standard error (precision) of the distribution of alarm calls among sentinel birds was 0.046 alarm calls. For the large raptors, the standard error (precision) of the distribution of alarm calls among sentinel birds was 0.0368 alarm calls.