Lab 4: Heart of the (Tiny) Tiger

July 10, 2020

Agenda: Writing functions to automate repetitive tasks; fitting statistical models.

The ***gamma*** distributions are a family of probability distributions defined by the density functions,

$$f(x) = \frac{x^{a-1}e^{-x/s}}{s^a\Gamma(a)}$$

where the ***gamma function*** $\Gamma(a) = \int_0^\infty u^{a-1}e^{-u}du$ is chosen so that the total probability of all non-negative x is 1. The parameter a is called the ***shape***, and s is the ***scale***. When $a = 1$, this becomes the exponential distributions we saw in the first lab. The gamma probability density function is called `dgamma()` in R. You can prove (as a calculus exercise) that the expectation value of this distribution is as, and the variance as^2. If the mean and variance are known, μ and σ^2, then we can solve for the parameters,

$$a = \frac{a^2s^2}{as^2} = \frac{\mu^2}{\sigma^2}$$

$$s = \frac{as^2}{as} = \frac{\sigma^2}{\mu}$$

In this lab, you will fit a gamma distribution to data, and estimate the uncertainty in the fit.

Our data today are measurements of the weight of the hearts of 144 cats.

Part I

1. The data is contained in a data frame called `cats`, in the R package `MASS`. (This package is part of the standard R installation.) This records the sex of each cat, its weight in kilograms, and the weight of its heart in grams. Load the data as follows:

```
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```
data(cats)
summary(cats)
```

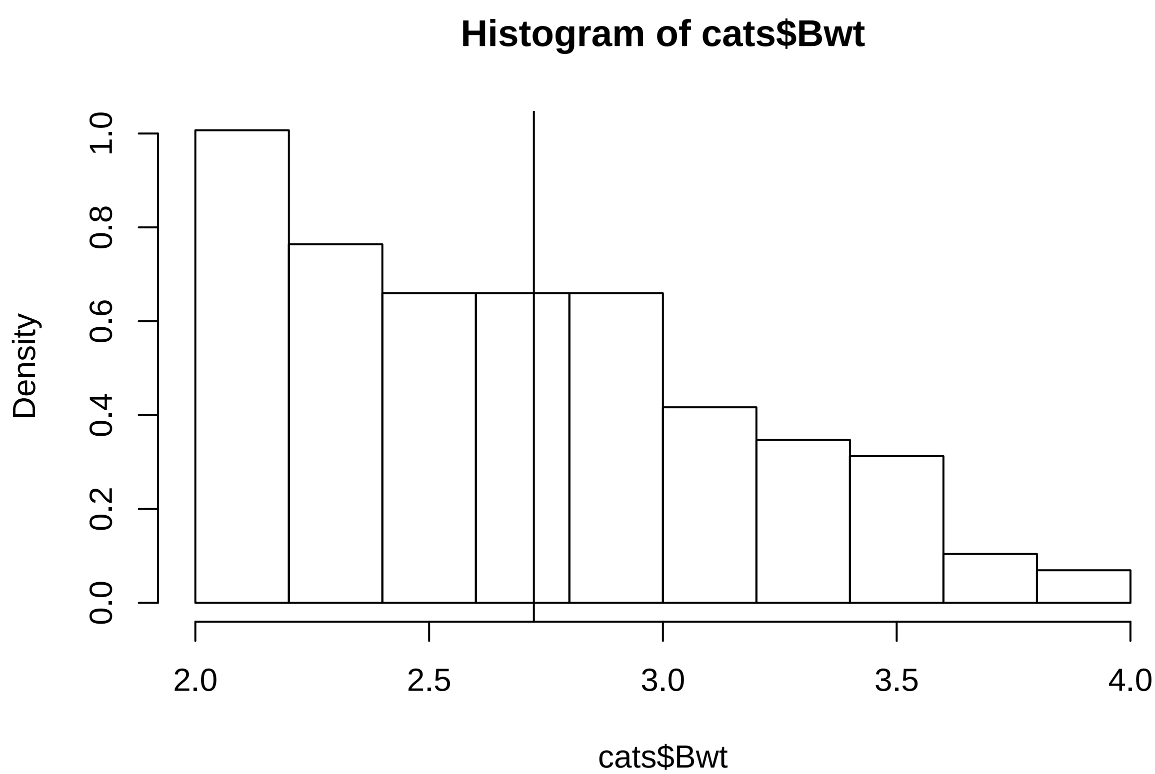
```
##  Sex         Bwt             Hwt
##  F:47   Min.   :2.000   Min.   : 6.30
##  M:97   1st Qu.:2.300   1st Qu.: 8.95
##         Median :2.700   Median :10.10
##         Mean   :2.724   Mean   :10.63
```

```
##         3rd Qu.:3.025   3rd Qu.:12.12
##         Max.   :3.900   Max.   :20.50
```

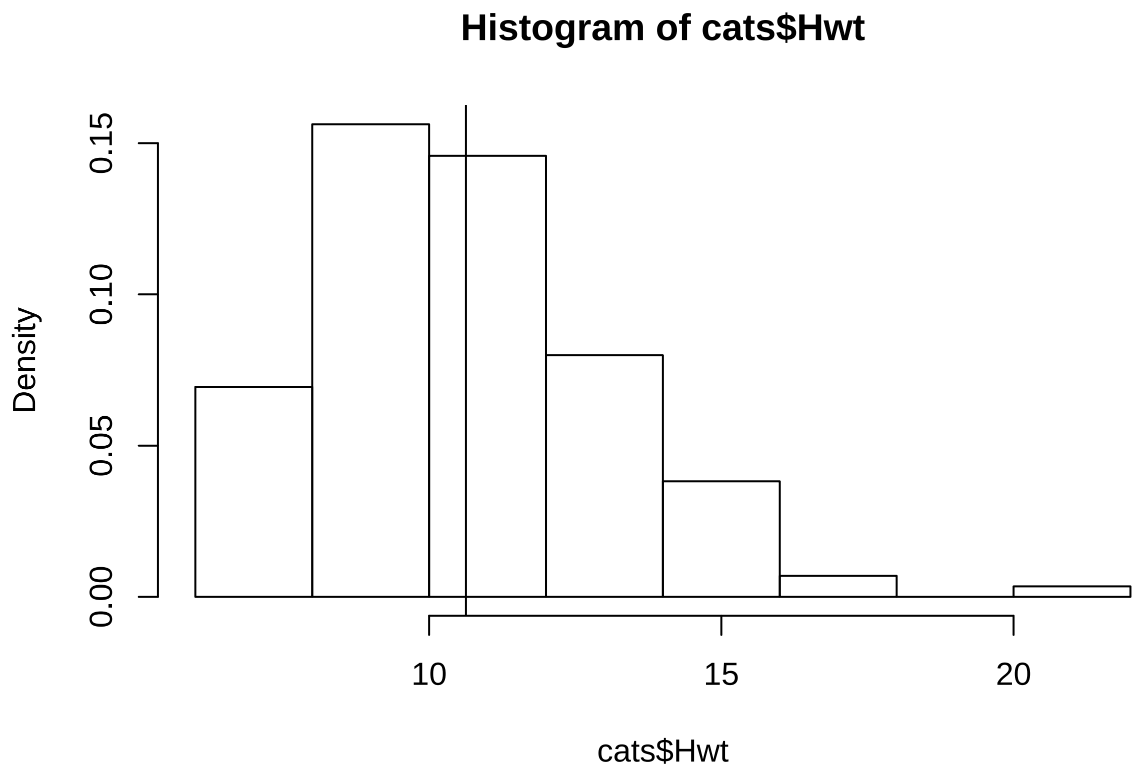
Run `summary(cats)` and explain the results.

2. Plot a histogram of these weights using the `probability=TRUE` option. Add a vertical line with your calculated mean using `abline(v=yourmeanvaluehere)`. Does this calculated mean look correct?

```
hist(cats$Bwt,probability=TRUE)
abline(v=2.724)
```




```
hist(cats$Hwt,probability=TRUE)
abline(v=10.63)
```



3. Define two variables, `fake.mean <- 10` and `fake.var <- 8`. Write an expression for a using these placeholder values. Does it equal what you expected given the solutions above? Once it does, write another such expression for s and confirm.

```
fake.mean <- 10
fake.var <- 8
a = (fake.mean/fake.var)^2
s = fake.var^2/fake.mean
```

4. Calculate the mean, standard deviation, and variance of the heart weights using R's existing functions for these tasks. Plug the mean and variance of the cats' hearts into your formulas from the previous question and get estimates of a and s. What are they? Do not report them to more significant digits than is reasonable.

```
m = mean(cats$Hwt)
std = sd(cats$Hwt)
var = var(cats$Hwt)
a = (m/var)^2
s = var^2/m
```

5. Write a function, `cat.stats()`, which takes as input a vector of numbers and returns the mean and variances of these cat hearts. (You can use the existing mean and variance functions within this function.) Confirm that you are returning the values from above.

```
cat.stats<-function(vec){
  m = mean(cats$Hwt[vec])
  var = var(cats$Hwt[vec])
  return(c(m,var))
```

```
}
cat.stats(seq(from=1,to=144,by=1))
```

```
## [1] 10.630556  5.927451
```

Part II

6. Now, use your existing function as a template for a new function, `gamma.cat()`, that calculates the mean and variances and returns the estimate of a and s. What estimates does it give on the cats' hearts weight? Should it agree with your previous calculation?

```
gamma.cat<-function(vec){
  m_var = cat.stats(vec)
  a = (m_var[1]/m_var[2])^2
  s = m_var[2]^2/m_var[1]
  return(c(a,s))
}
gamma.cat(seq(from=1,to=144,by=1))
```

```
## [1] 3.216443 3.305065
```

7. Estimate the a and s separately for all the male cats and all the female cats, using `gamma.cat()`. Give the commands you used and the results.

```
gamma.cat(which(cats$Sex=="M"))
```

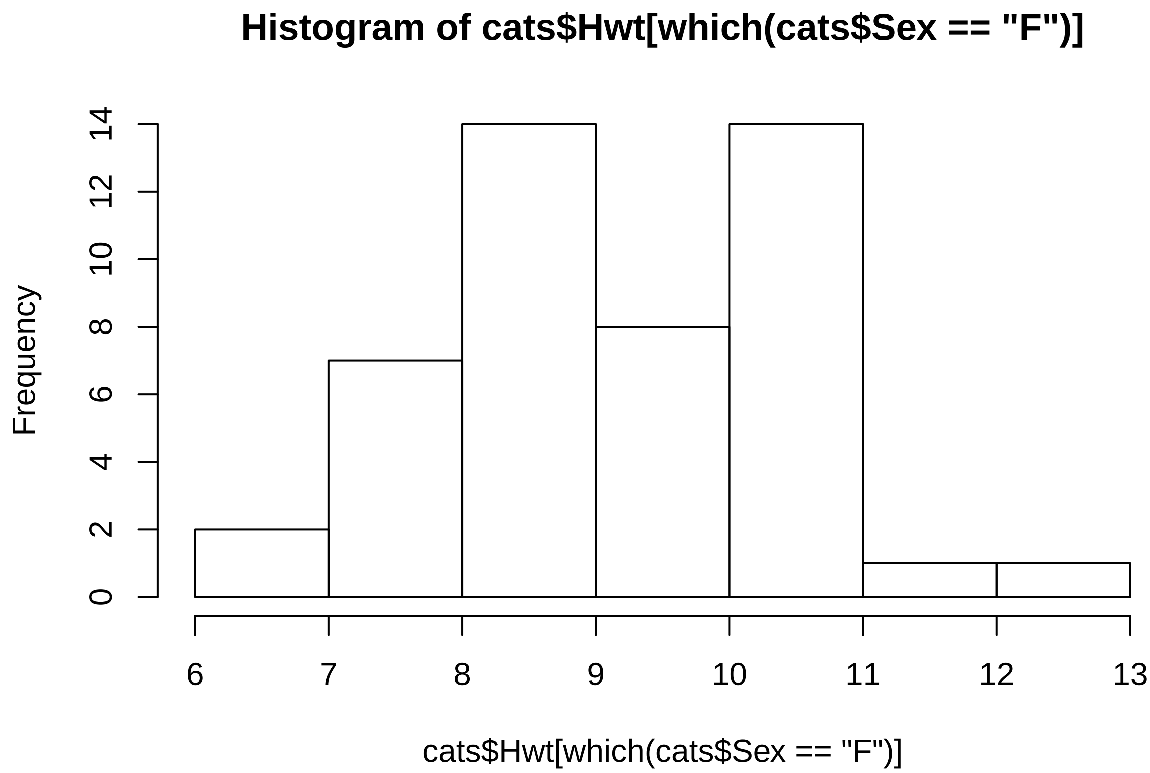
```
## [1] 3.069017 3.689351
```

```
gamma.cat(which(cats$Sex=="F"))
```

```
## [1] 24.9232733  0.3692183
```

8. Now, produce a histogram for the female cats. On top of this, add the shape of the gamma PDF using `curve()` with its first argument as `dgamma()`, the known PDF for the Gamma distribution. Is this distribution consistent with the empirical probability density of the histogram?

```
hist(cats$Hwt[which(cats$Sex=="F")])
```




```
#curve(expr = dgamma(x = seq(0,10,0.1),shape = a, scale = s))
```

9. Repeat the previous step for male cats. How do the distributions compare?