Probability

Objectives

- Connect random variables, probabilities, and parameters
- define prob. functions
  - discrete and continuous random variables
- use/plot prob. functions
- notation

Probability/Statistics

Probability and statistics are the opposite sides of the same coin.

. . .

To understand statistics, we need to understand probability and probability functions.

. . .

The two key things to understand this connection is the random variable (RV) and parameters (e.g., θ, σ, ϵ, μ).

Motivation

Why learn about RVs and probability math?

. . .

Foundations of:

- linear regression
- generalized linear models

- mixed models / hierarchical models

. . .

Our Goal:

- conceptual framework to think about *data*, *probabilities*, and *parameters*
- mathematical connections and notation

Not Random Variables

$$
\begin{aligned}
a =& 10 \\
b =& \log(a) \times 12 \\
c =& \frac{a}{b} \\
y =& \beta_0 + \beta_1 * c
\end{aligned}
$$

. . .

All variables here are scalars. They are what they are and that is it. β variables and y are currently unknown, but still scalars.

. . .

Scalars are quantities that are fully described by a magnitude (or numerical value) alone.

Random Variables

$$y \sim f(y)$$

y is a random variable which may change values each observation; it changes based on a probability function, $f(y)$.

. . .

The tilde ($\sim$) denotes "has the probability distribution of".

. . .

Which value (y) is observed is predictable. Need to know *parameters* (θ) of the probability function $f(y)$.

Random Variables

$$y \sim f(y|\theta)$$

Specifically, $f(y|\theta)$, where '|' is read as 'given'.

. . .

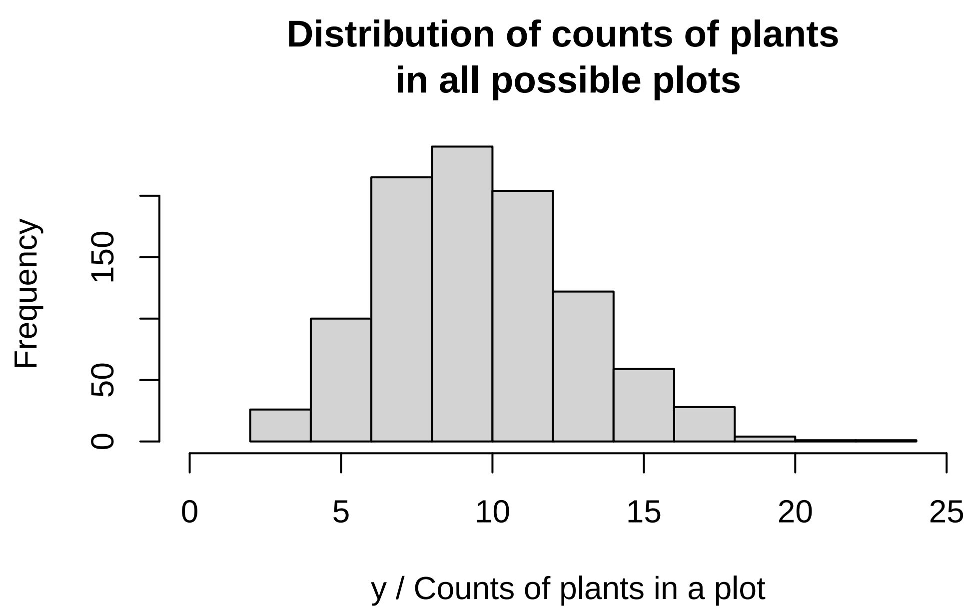
Toss of a coin Roll of a die Weight of a captured elk Count of plants in a sampled plot

. . .

The values observed can be understand based on the frequency within the population or presumed super-population. These frequencies can be described by probabilities.

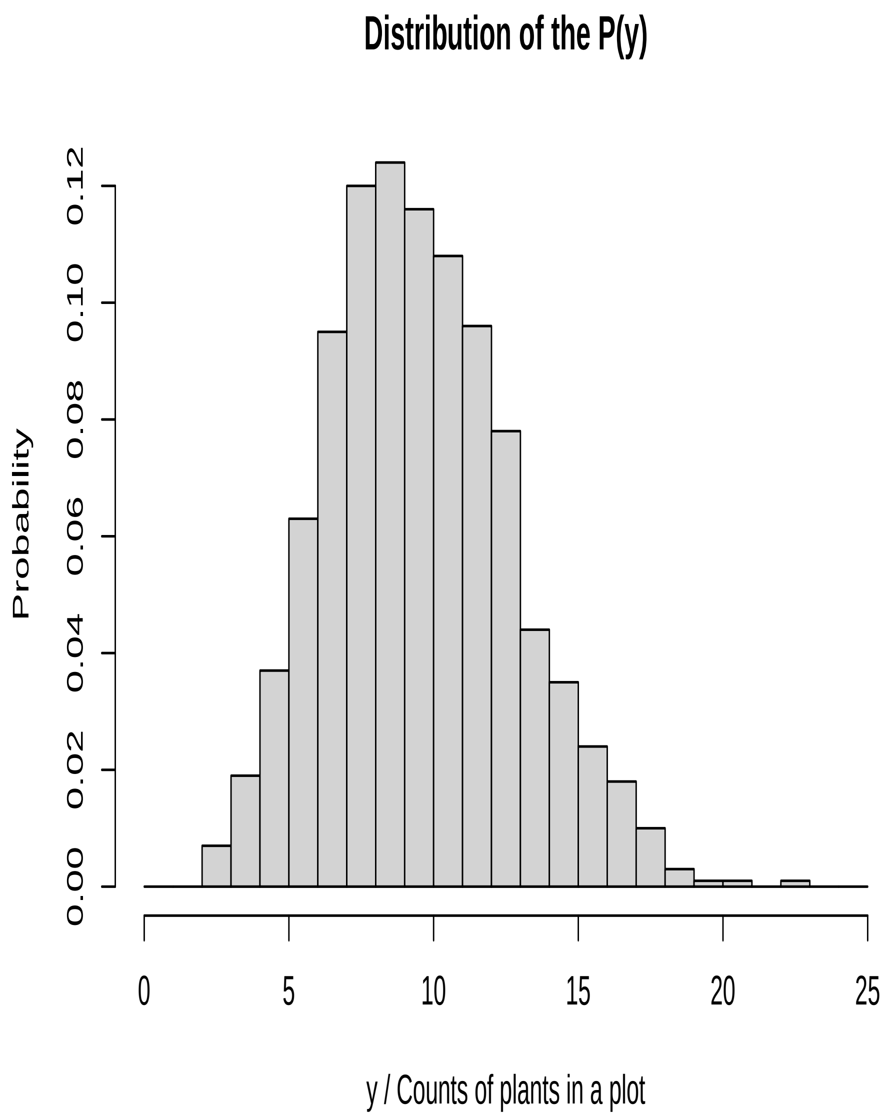
Frequency / Probabilities

```
hist(y, xlim=c(0,25),main=main,xlab="y / Counts of plants in a plot")
```



Frequency / Probabilities

```
hist.values=hist(y,breaks = seq(from=0, to=25, by = 1),freq = FALSE,
                 ylab="Probability",
                 main="Distribution of the P(y)",
                 xlab="y / Counts of plants in a plot")
```

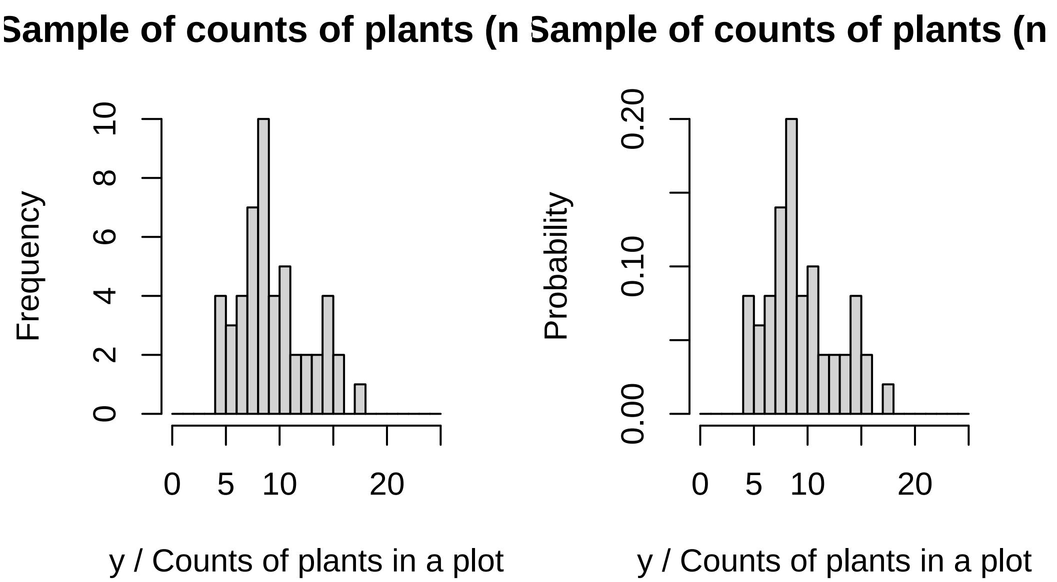


```r
sum(hist.values$density)
```

```
[1] 1
```

Sampling

We often only get to see **ONE** sample from this distribution.



Random Variables

We are often interested in the characteristics of the whole population of frequencies,

- central tendency (mean, mode, median)
- variability (var, sd)
- proportion of the population that meets some condition $\mathrm{P}(8 \leq y \leq 12) = 0.68$

. . .

We infer what these are based on our sample (i.e., statistical inference).

Philosophy

Frequentist Paradigm:

Data (e.g., y) are random variables that can be described by probability distributions with unknown parameters that (e.g., θ) are *fixed* (scalars).

. . .

Bayesian Paradigm:

Data (e.g., y) are random variables that can be described by probability functions where the unknown parameters (e.g., θ) are also random variables that have probability functions that describe them.

Random Variables

$$y \sim f(y|\theta)$$

$$\begin{aligned} y &= \text{event/outcome} \\ f(y|\theta) = [y|\theta] &= \text{ process governing the value of } y \\ \theta &= \text{parameters} \end{aligned}$$

. . .

$f()$ or [ ] is conveying a function (math).

. . .

It is called a PDF when y is continuous and a PMF when y is discrete.

- PDF: **probability density function**
- PMF: **probability mass function**

Functions

We commonly use *deterministic* functions (indicated by non-italic letter); e.g., log(), exp(). Output is always the same with the same input.

$$\mathrm{g}x \Longrightarrow \boxed{\text{DO STUFF}} \Longrightarrow \mathrm{g}(x)$$

. . .

$$\mathrm{g}x \Longrightarrow \boxed{+7} \Longrightarrow \mathrm{g}(x)$$

. . .

$$\mathrm{g}(x) = x + 7$$

. . .

Random Variables

Probability: Interested in y, the data, and the probability function that "generates" the data.

$$y \leftarrow f(y|\theta)$$

. . .

Statistics: Interested in population characteristics of y; i.e., the parameters,

$$y \rightarrow f(y|\theta)$$

Probability Functions

Special functions with rules to guarantee our logic of probabilities are maintained.

. . .

Discrete RVs

y can only be a certain set of values.

1. $y \in \{0, 1\}$
    - 0 = dead,
    - 1 = alive

Probability Functions

Special functions with rules to guarantee our logic of probabilities are maintained.

Discrete RVs

y can only be a certain set of values.

2. $y \in \{0, 1, 2\}$
    - 0 = site unoccupied
    - 1 = site occupied w/o young
    - 2 = site occupied with young

Probability Functions

Special functions with rules to guarantee our logic of probabilities are maintained.

Discrete RVs

y can only be a certain set of values.

3. $y \in \{0, 1, 2, ..., 15\}$
    - count of pups in a litter; max could by physiological constraint

These sets are called the sample space (Ω) or the support of the RV.

PMF

$$f(y) = P(Y = y)$$

. . .

Data has two outcomes (0 = dead, 1 = alive)

$y \in \{0, 1\}$

. . .

There are two probabilities

- $f(0) = P(Y = 0)$
- $f(1) = P(Y = 1)$

. . .

Axiom 1: The probability of an event is greater than or equal to zero and less than or equal to 1.

$$0 \leq f(y) \leq 1$$

Example,

- $f(0) = 0.1$
- $f(1) = 0.9$

. . .

Axiom 2: The sum of the probabilities of all possible values (sample space) is one.

. . .

$$\sum_i f(y_i) = f(y_1) + f(y_2) + ... = P(\Omega) = 1$$

Example,

- $f(0) + f(1) = 0.1 + 0.9 = 1$

PMF

Still need to define $f()$, our PMF for $y \in \{0, 1\}$

. . .

The Bernoulli distribution

$$ f(y|) = [y|]=

```
\theta^{y}\times(1-\theta)^{1-y}
```

$$

$\theta = \mathrm{P(Y = 1)} = 0.2$

$$ f(y|) = [y|]

```
= 0.2^{1}\times(1-0.2)^{1-1}
```

$$

$$ f(y|) = [y|]=

```
= 0.2 \times (0.8)^{0} = 0.2
```

$$

Bernoulli PMF

$$ f(y|) = [y|]=

```
\theta^{y}\times(1-\theta)^{1-y}
```

$$

Sample space support (Ω):

- $y \in \{0, 1\}$

Parameter space support (Θ):

- $\theta \in [0, 1]$
- General: $\theta \in \Theta$

Bernoulli PMF (Code)

What would our data look like for 10 ducks that had a probability of survival (Y = 1) of 0.20?

. . .

```
# Define inputs
  theta = 0.2
  N = 1

# Random sample - 1 duck
  rbinom(n = 1,
         size = N,
         prob = theta
         )
```

```
[1] 0
```

. . .

```
# Random sample - 10 ducks
  rbinom(n = 10,
         size = N,
         prob = theta
         )
```

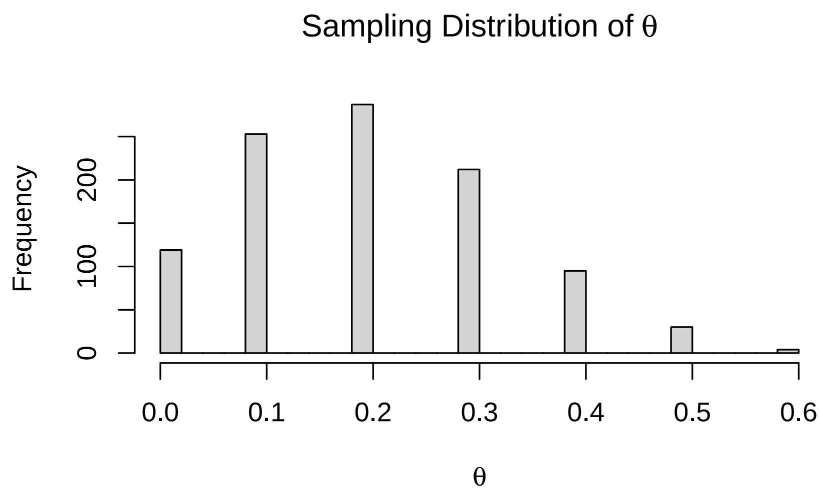
```
 [1] 1 0 0 0 1 0 1 0 1 0
```

Why is this useful to us?

How about to evaluate the sample size of ducks needed to estimate θ?

. . .

```
y.mat = replicate(1000, rbinom(n = 10,size = N, prob = theta))
theta.hat = apply(y.mat, 2, mean)
```



Sampling Distribution of θ
Frequency
0
100
200
0.0
0.1
0.2
0.3
0.4
0.5
0.6
θ