Use c() only when necessary

```
> worms( , 2:4)

# not
worms( , c(2:4))

# not
worms( , c(2,3,4))
```

Include the comma when accessing a data frame

```
> worms[ , 2:3]

# is more explicit than
> worms[2:3]
```

Use a text editor, not a word processor

and, beware of smart quotes

Calling by name is safer and self-commenting

```
> rlnorm(n=1000, meanlog=7.1, sdlog=1.1)

# is better than
> rlnorm(1000, 7.1, 1.1)
```

But take advantage of common conventions

```
> mean(myData)
# instead of
> mean(x=myData)

> read.table('myFile.txt')
# instead of
> read.table(file='myFile.txt')
```

Statistics Fundamentals

Kinds of data

Nominal Data

Mutually exclusive, unranked categories

rock type

vegetation type

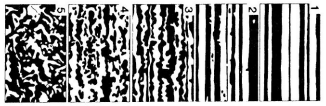
color

Ordinal Data

Like nominal, but ordered or ranked

Moh's hardness scale

Droser burrowing scale



Interval Data

Like ordinal, but steps of equal size

Fahrenheit & Celsius

isotopic ratios

Ratio Data

Like interval, but with natural zero point

Kelvin and Rankine

length

Open vs. Closed

Open: not constrained by measurement system
length

Closed: constrained by measurement system
percentages: must sum to 100%

Continuous vs. Discrete

Continuous: all intermediate values possible
length

Discontinuous: only certain values possible, such as integers
number of clams

Populations vs. Samples

Population

Well-defined set of elements

Finite or infinite

Total collection of objects

Sample

Subset of elements from population

Distinct from a geologic sample, isotopic sample, tissue sample, etc.

The Problem

We collect a sample,
but we're interested in the population

Statistics vs. Parameters

Parameter

A measurement that describes a population

standard deviation: σ

Statistic

A measurement that describes a sample

standard deviation: s

The Problem, Restated

We measure statistics,
but we're interested in parameters

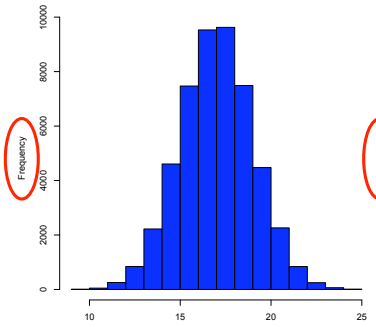
Replication improves our
inferences about parameters

Randomization is required
to make these inferences

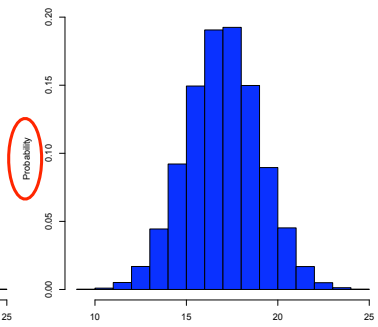
Distributions

Frequency Distribution

in terms of frequency...



or probability

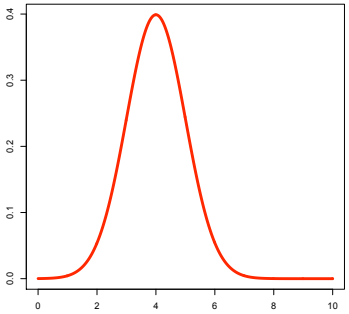


Normal Distribution

Symmetrical

Continuous

Many independent, additive factors

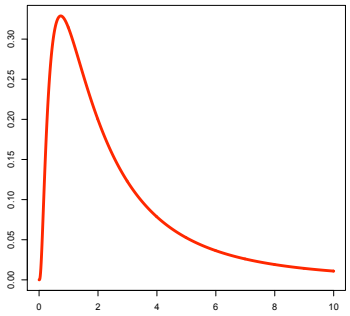


Lognormal Distribution

Asymmetrical, right-tailed

Continuous

Many independent, multiplicative factors

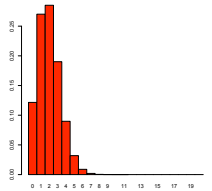


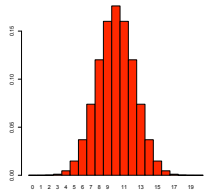
Binomial Distribution

Symmetrical or asymmetrical

Discrete

Number of successes in n trials



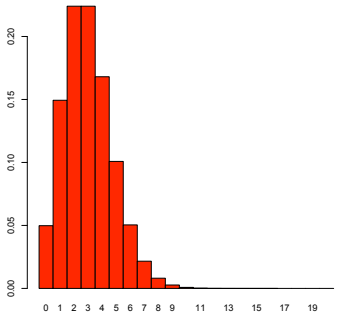


Poisson Distribution

Right-tailed or symmetrical

Discrete

Number of events in fixed window of space or time

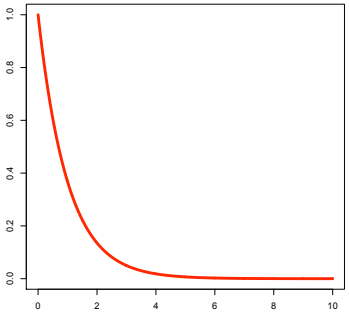


Exponential Distribution

Right-tailed

Continuous

Waiting time between events in fixed window of space or time

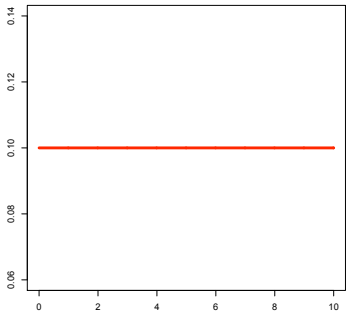


Uniform Distribution

Flat

Discrete or Continuous

Rare in nature



Descriptive Statistics

Central Tendency

Mean

Median

Mode

Geometric Mean

Variation

Range

Interquartile Deviation

Variance

Standard Deviation

Coefficient of Variation

Shape

Unimodal, bimodal, multimodal

Skewness: right-skewed, left-skewed

Kurtosis: leptokurtic, platykurtic