Simulation basics

Jeffrey Leek, Assistant Professor of Biostatistics
Johns Hopkins Bloomberg School of Public Health

Important simulation functions

Distributions

- rbeta, rbinom, rcauchy, rchisq, rexp, rf, rgamma, rgeom, rhyper, rlogis, rlnorm, rnbinom, rnorm, rpois, rt, runif, rweibull

Densities

- dbeta,dbinom, dcauchy, dchisq, dexp, df, dgamma, dgeom, dhyper, dlogis, dlnorm, dnbinom, dnorm, dpois, dt, dunif, dweibull

Sampling

- sample(,replace=TRUE),sample(replace=FALSE)

r*foo* functions generate data

Normal

```
args(rnorm)
```

```
function (n, mean = 0, sd = 1)
NULL
```

```
heights = rnorm(10,mean=188,sd=3)
heights
```

```
 [1] 186.0 191.2 187.6 187.9 186.6 187.2 187.2 189.5 190.8 186.4
```

r*foo* functions generate data

Binomial

```
args(rbinom)
```

```
function (n, size, prob)
NULL
```

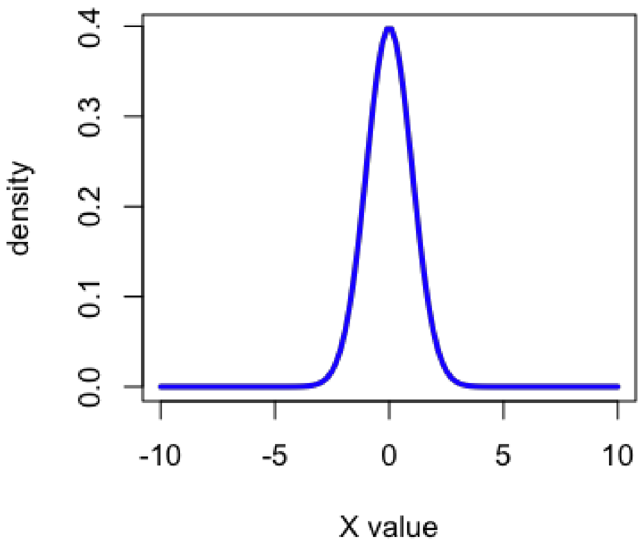
```
coinFlips = rbinom(10,size=10,prob=0.5)
coinFlips
```

```
 [1] 3 4 6 5 7 6 5 8 5 6
```

Example distribution: Normal

Normal Distribution: $N(\mu, \sigma)$

- $X \sim N(0, 1)$



5/13

d*foo* functions calculate the density

Normal

```
args(dnorm)
```

```
function (x, mean = 0, sd = 1, log = FALSE)
NULL
```

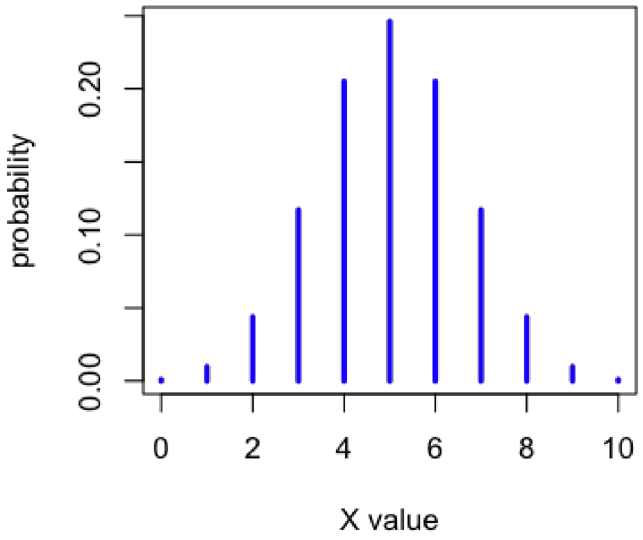
```
x = seq(from=-5,to=5,length=10)
normalDensity = dnorm(x,mean=0,sd=1)
round(normalDensity,2)
```

```
 [1] 0.00 0.00 0.01 0.10 0.34 0.34 0.10 0.01 0.00 0.00
```

Example distribution: Binomial

Binomial distribution: $Bin(n, p)$

- $X \sim Bin(10, 0.5)$



7/13

d*foo* functions calculate the density

Binomial

```
args(dbinom)
```

```
function (x, size, prob, log = FALSE) 
NULL
```

```
x = seq(0,10,by=1)
binomialDensity = dbinom(x,size=10,prob=0.5)
round(binomialDensity,2)
```

```
 [1] 0.00 0.01 0.04 0.12 0.21 0.25 0.21 0.12 0.04 0.01 0.00
```

Sample draws a random sample

```
args(sample)
```

```
function (x, size, replace = FALSE, prob = NULL)
NULL
```

```
heights = rnorm(10,mean=188,sd=3)
heights
```

```
 [1] 187.2 185.4 187.9 187.3 184.8 190.3 185.0 188.2 190.0 188.1
```

```
sample(heights,size=10,replace=TRUE)
```

```
 [1] 188.2 188.2 184.8 185.0 187.2 188.2 187.9 185.4 184.8 185.4
```

Sample draws a random sample

```
heights
```

```
 [1] 187.2 185.4 187.9 187.3 184.8 190.3 185.0 188.2 190.0 188.1
```

```
sample(heights,size=10,replace=FALSE)
```

```
 [1] 185.0 188.2 188.1 184.8 190.3 187.9 187.2 185.4 190.0 187.3
```

Sample can draw according to a set of probabilities

```
heights
```

```
 [1] 187.2 185.4 187.9 187.3 184.8 190.3 185.0 188.2 190.0 188.1
```

```
probs = c(0.4,0.3,0.2,0.1,0,0,0,0,0,0)
sum(probs)
```

```
[1] 1
```

```
sample(heights,size=10,replace=TRUE,prob=probs)
```

```
 [1] 187.2 185.4 187.9 187.3 185.4 187.2 185.4 187.2 187.2 185.4
```

Setting a seed

Setting a seed ensures reproducible results from random processes in R

```
set.seed(12345)
rnorm(5,mean=0,sd=1)
```

```
[1]  0.5855  0.7095 -0.1093 -0.4535  0.6059
```

```
set.seed(12345)
rnorm(5,mean=0,sd=1)
```

```
[1]  0.5855  0.7095 -0.1093 -0.4535  0.6059
```

For more information

More on distributions in R

http://cran.r-project.org/web/views/Distributions.html

Computing for Data Analysis

Simulation in R