**Probability Distributions**

**Sampling:**

**From a given dataset:**

You often have a matrix where each row is a subject that has been asked several questions (the columns). If we want a random sample from our dataset, we use the sample ( ) function to generate a vector of random row numbers and then select that subset from the matrix.

> sample(1:20,5)

[1] 12 6 10 7 2

> random.sample<-sample(seq(1,64),10)

> random.sample

[1] 52 2 25 54 27 8 40 11 49 17

**From a distribution:**

Often we want to sample data from a specific distribution, also sometimes called simulating data. This data is usually used to test some algorithm or function that someone has written. Since the data is simulated, you know where it came from and so what the answer should be from your algorithm or function. Simulated data lets you double-check your work.

Each distribution has 4 functions associated with it: r--, d--, p--, and q--.

(For example, rnorm( ), dnorm( ), pnorm( ), and qnorm( )).

You must specify the parameters for the distribution.

The r-- function simulates random data from the distribution.

10 observations from a normal distribution with mean 0 and stdev 1:

> rnorm(10,0,1)

[1] 0.1872828 -0.1160541 1.8812873 1.0428904 -1.5879228

[6] 1.1440760 1.1122024 -0.3928728 0.7963980 1.2735349

6 observations from a normal distribution with mean 3 and stdev 2:

> rnorm(6,3,2)

[1] 3.618939 5.129774 2.553895 0.715022 7.506615 3.347618

The d-- function finds the density value of the number/vector you plug in.

The density value of 3.5 in a normal with mean 3 and st. dev 1:

> dnorm(3.5,3,1)

[1] 0.3520653

The density value of 0.5 in a normal with mean 3 and stdev 1:

> dnorm(0.5,3,1)

[1] 0.0175283

The density values of a vector in a normal with mean 0 and stdev 1:

> dnorm(c(-3,-2,-1,0,1,2,3),0,1)

[1] 0.004431848 0.053990967 0.241970725 0.398942280 0.241970725 [6] 0.053990967 0.004431848

The p-- function gives the distribution function; that is, it gives the probability of being at or below the number/vector you plug in for that distribution. P(X <= x).

The probability of being <=3 in a normal with mean 0 and stdev 1:

> pnorm(3,0,1)

[1] 0.9986501

The probability of being <=-3 in a normal with mean 0 and stdev 1:

> pnorm(-3,0,1)

[1] 0.001349898

Note: you can get the probability of being greater than a number by 1–p--

The probability of being >12 in a normal with mean 10 and stdev 3:

> 1-pnorm(12,10,3)

[1] 0.2524925

The q-- function gives the quantile function, or what number marks the x-th percentile of a specific distribution. P(X <= ?) = given percentile.

The 50th percentile of a normal with mean 0 and stdev 1:

> qnorm(0.50,0,1)

[1] 0

The 10th, 25th, 75th, and 90th percentile of a normal with mean -1 and stdev 0.5:

> qnorm(c(0.10,0.25,0.75,0.90),-1,0.5)

[1] -1.6407758 -1.3372449 -0.6627551 -0.3592242

There are R functions for the following distributions:

Distributions R Name Additional Arguments

|  |  |  |
| --- | --- | --- |
| beta | beta | shape1, shape2, ncp |
| binomial | binom | size, prob |
| Cauchy | cauchy | location, scale |
| chi-squared | chisq | df, ncp |
| exponential | exp | rate |
| F | f | df1, df2, ncp |
| gamma | gamma | shape, scale |
| geometric | geom. | prob |
| hypergeometric | hyper | m, n, k |
| log-normal | lnorm | meanlog, sdlog |
| logistic | logis | location, scale |
| negative binomial | nbinom | size, prob |
| normal | norm | mean, sd |
| Poisson | pois | lambda |
| Student’s t | t | df, ncp |
| uniform | unif | min,max |
| Weibull | weibull | shape, scale |
| Wilcoxon | wilcox | m, n |

**Examples of other distributions:**

**Binomial distribution**

If we use the rbinom( ), the n argument is how many trials, the size is how many coins, and the p is the probability of getting a heads:

Flip one coin once.

> rbinom(n=1,size=1,p=.5)

[1] 0

Flip three coins ten times; the results are the number of heads we saw each time.

> rbinom(n=10,size=3,p=.5)

[1] 2 1 1 2 0 3 2 1 3 3

We can change the coin to have a smaller chance of success.

> rbinom(n=10,size=3,p=.2)

[1] 2 1 0 1 1 0 0 1 1 1

What is the probability of seeing 7 heads if we flip 12 coins?

> dbinom(7,12,.5)

[1] 0.1933594

**Uniform distribution:**

Random samples from [0,1]:

> runif(10,0,1)

[1] 0.80193610 0.29145833 0.02581367 0.80001030 0.83624867

[6] 0.38835422 0.17701875 0.81880388 0.40883710 0.75325295

or from the unit square: [0,1] by [0,1]:

> cbind(runif(8,0,1),runif(8,0,1))

[,1] [,2]

[1,] 0.8091289 0.02334477

[2,] 0.2980009 0.95909988

[3,] 0.9597524 0.56358745

[4,] 0.6610231 0.22847434

[5,] 0.1445462 0.82469317

[6,] 0.6264433 0.71810215

[7,] 0.9222504 0.40311884

[8,] 0.4051854 0.97956278

**Poisson Distribution:**

What is the probability of getting 4 phone calls in the next hour if on average you receive 6 phone calls an hour?

> dpois(4,6)

[1] 0.1338526

**Exponential Distribution:**

What is the probability that a light bulb with an average lifetime of 200 hours burns out before 100 hours?

> pexp(100,1/200)

[1] 0.3934693