Minnesota University

Study Resource Link: <https://www.stat.umn.edu/geyer/old/5101/rlook.html>

**R Functions for Probability Distributions**

**General Command**

Every distribution that R handles has four functions. There is a root name, for example, the root name for the normal distribution is norm. This root is prefixed by one of the letters

• p for "probability", the cumulative distribution function (c. d. f.)

• q for "quantile", the inverse c. d. f.

• d for "density", the density function (p. f. or p. d. f.)

• r for "random", a random variable having the specified distribution

For the normal distribution, these functions are pnorm, qnorm, dnorm, and rnorm.

For the binomial distribution, these functions arepbinom, qbinom, dbinom, and rbinom. And so forth

**R Function Quick Recap**

1. rnorm: to generate a series of random value. For example, rnorm(20) gives you 20 random variables

2. PNORM DIRECT LOOK-UP PNORM IS THE R FUNCTION THAT CALCULATES THE C. D. F.

F(x) = P(X <= x),where X is normal. Optional arguments described on the

on-line documentation specify the parameters of the particular normal

distribution.

3. Both of the R commands in the box below do exactly the same thing.

Question: Suppose widgit weights produced at Acme Widgit Works have weights that are normally distributed with mean 17.46 grams and variance 375.67 grams. What is the probability that a randomly chosen widgit weighs more then 19 grams?

Question Rephrased: What is P(X > 19) when X has the N(17.46, 375.67) distribution?

Caution: R wants the s. d. as the parameter, not the variance. We'll need to take a square root!

**INVERSE LOOK-UP**

qnorm is the R function that calculates the inverse c. d. f. F-1 of the normal distribution The c. d. f. and the inverse c. d. f. are related by

p = F(x)

x = F-1(p)

So given a number p between zero and one, qnorm looks up the p-th quantile of the normal distribution. As with pnorm, optional arguments specify the mean and standard deviation of the distribution.

Question: Suppose widgits produced at Acme Widgit Works have probability 0.005 of being defective. Suppose widgits are shipped in cartons containing 25 widgits. What is the probability that a randomly chosen carton contains exactly one defective widgit?

Question Rephrased: What is P(X = 1) when X has the Bin(25, 0.005) distribution?

**THE BINOMIAL DISTRIBTION**

DIRECT LOOK-UP, POINTS

dbinom is the R function that calculates the p. f. of the binomial distribution. Optional arguments described on the on-line documentation specify the parameters of the particular binomial distribution.

Both of the R commands in the box below do exactly the same thing.

They look up P(X = 27) when X is has the Bin(100, 0.25) distribution.

EXAMPLE

Question: Suppose widgits produced at Acme Widgit Works have probability 0.005 of being defective. Suppose widgits are shipped in cartons containing 25 widgits. What is the probability that a randomly chosen carton contains exactly one defective widgit?

Question Rephrased: What is P(X = 1) when X has the Bin(25, 0.005) distribution?

Answer:DIRECT LOOK-UP, INTERVALS

pbinom is the R function that calculates the c. d. f. of the binomial distribution. Optional arguments described on the on-line documentation specify the parameters of the particular binomial distribution.

Both of the R commands in the box below do exactly the same thing.

They look up P(X <= 27) when X is has the Bin(100, 0.25) distribution. (Note the LESS THAN OR EQUAL TO sign. It's important when working with a discrete distribution!)

EXAMPLE

Question: Suppose widgits produced at Acme Widgit Works have probability 0.005 of being defective. Suppose widgits are shipped in cartons containing 25 widgits. What is the probability that a randomly chosen carton contains no more than one defective widgit?

Question Rephrased: What is P(X <= 1) when X has the Bin(25, 0.005) distribution?

Answer:

INVERSE LOOK-UP

qbinom is the R function that calculates the "inverse c. d. f." of the binomial distribution. How does it do that when the c. d. f. is a step function and hence not invertible? The on-line documentation for the binomial probability functions explains.

The quantile is defined as the smallest value x such that F(x) >= p, where F is the distribution function.

When the P-th quantile is nonunique, there is a whole interval of values each of which is a P-th quantile. The documentation says that qbinom (and other "q" functions, for that matter) returns the smallest of these values. That is one sensible definition of an "inverse c. d. f." In the terminology of Section of the course notes, the function defined by qbinom is a RIGHT INVERSE of the function defined by pbinom, that is,

q == pbinom(qbinom(q, n, p)), 0 < q < 1, 0 < p < 1, n a positive integer

is always true, but the analogous formula with pnorm and qnorm reversed does not necessarily hold. Question: What are the 10th, 20th, and so forth quantiles of the Bin(10, 1/3) distribution?

Answer:Note the nonuniqueness.

**Normal Distribution**

**Graphical user interface, text, application, email

Description automatically generated**

**Histogram

Description automatically generated**

The normal distribution density function f(z) is called the Bell Curve because it has the shape that resembles a bell.

The area under the bell curve indicates the probability of a value range and the total area is 1. The center line of the bell curve intercepts the mean value and mean value does not necessary to be 0. It can be any value based on a given population.

Standard normal distribution table is used to find the area under the f(z) function in order to find the probability of a specified range of distribution. For example, you can utilize the table to find probability between Z-Score 0.9 and 1.6. Z-score is the independent variable of PDF, but essentially the Z-score is a statistical measurement that describes a value's relationship to the mean of a group of values.

The feature of normal distribution

1. The mean, median and mode are exactly the same number, or 50th percentile

2. 68% of the data falls within 1 standard deviation.

95.4% of the data falls within 2 standard deviations.

99.7% of the data falls within 3 standard deviations.

3. Normal distribution graph is a symmetrical Bell-Shape Curve

The chi-square test of independence is an inferential statistical test, meaning that it allows you to draw conclusions about a population based on a sample. Specifically, it allows you to conclude whether two variables are related in the population.