

Statistical models and probability distributions

Gerardo Ferrara

Master in Economics and Complexity, Collegio Carlo Alberto

Fall 2013

A full range of summary statistics are available, including:

- `mean(x)`
- `mean(x,trim=0.95)`
- `median(x)`
- `sqrt(var(x))`
- `sd(x)`
- `quantile(x,probs=0.25)`

Some functions have functionality for dealing with missing values **NA**. The function *mean*, e.g., has argument prototype **na.rm=FALSE**.

Summary Function

The function *summary* is particularly useful with many of the statistical functions. When applied to a numeric vector:

```
> summary(x)
```

the function produces a 6 point summary, comprising the minimum, maximum, lower and upper quartiles, and the median and mean. Applied to a matrix (or a data frame), the summary function generates this summaries for each column.

R has an impressive graphics functionality. A common data analysis tool is the histogram:

```
> hist(x, prob = TRUE)
```

Histograms display the frequency distribution of a data set. The area of the rectangles (bins) is proportional to the number of observations. They are generated using a default binning strategy. The appearance can depend on the choice of breakpoints. To specify the number of bins or the actual breakpoints, modify the argument **breaks=**.

A histogram is a crude form of density estimate. A better alternative is, often, a smooth density estimate:

```
> dens <- density(x); lines(dens)
```

Density curve are preferable to histograms for drawing attention to particular form of non-normality like skewness.

Probability Distributions

R includes functions for computing quantities associated with a variety of distributions, including:

- **pdist** # probability
- **qdist** # quantile
- **ddist** # density
- **rdist** # random number

where *dist* is the nickname of the distribution (like normal, Poisson, binomial, uniform, negative exponential, gamma, chi-squared and so on).

Probability Distributions (cont.)

For example, to generate 1000 random variates from a standard normal distribution (with mean 0 and standard deviation 1):

```
> rnorm(1000, mean = 0, sd = 1)
```

while, to calculate the probability masses at 0, 1, 2 of a Poisson distribution of rate 2:

```
> dpois(0 : 2, lambda = 2)
```

Other examples of distribution are:

- binom
- exp
- gamma
- chisq
- unif
- beta