

Appendix A

Vectors, Matrices, and Lists

A.1 Vectors

A.1.1 *Creating a vector*

Vectors can be created in several ways. We have already seen two methods, the combine function `c` and the colon operator `:` for creating vectors with given values. For example, to create a vector of integers 1 through 9 and assign it to `i`, we can use either of the following.

```
i = c(1,2,3,4,5,6,7,8,9)
i = 1:9
```

To create a vector without specifying any of its elements, we only need to specify the type of vector and its length. For example,

```
y = numeric(10)
```

creates a numeric vector `y` of length 10, and

```
a = character(4)
```

creates a vector of 4 empty character strings. The vectors created are

```
> y
[1] 0 0 0 0 0 0 0 0 0 0
> a
[1] "" "" "" ""
```

A.1.2 *Sequences*

There are several functions in R that generate sequences of special types and patterns. In addition to `:`, there is a sequence function `seq` and a repeat `rep` function. We often use these functions to create vectors.

The `seq` function generates ‘regular’ sequences that are not necessarily integers. The basic syntax is `seq(from, to, by)`, where `by` is the size of the increment. Instead of `by`, we could specify the `length` of the sequence. To create a sequence of numbers $0, 0.1, 0.2, \dots, 1$, which has 11 regularly spaced elements, either command below works.

```
> seq(0, 1, .1)
[1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
> seq(0, 1, length=11)
[1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
```

The `rep` function generates vectors by repeating a given pattern a given number of times. The `rep` function is easily explained by the results of a few examples:

```
> rep(1, 5)
[1] 1 1 1 1 1

> rep(1:2, 5)
[1] 1 2 1 2 1 2 1 2 1 2

> rep(1:3, times=1:3)
[1] 1 2 2 3 3 3

> rep(1:2, each=2)
[1] 1 1 2 2

> rep(c("a", "b", "c"), 2)
[1] "a" "b" "c" "a" "b" "c"
```

A.1.3 Extracting and replacing elements of vectors

If `x` is a vector, `x[i]` is the i^{th} element of `x`. This syntax is used both for assignment and extracting values. If `i` happens to be a vector of positive integers (i_1, \dots, i_k) , then `x[i]` is the vector containing x_{i_1}, \dots, x_{i_k} , provided these are valid indices for `x`.

A few examples to illustrate extracting elements from a vector are:

```
> x = letters[1:8] #letters of the alphabet a to h
> x[4]           #fourth element
[1] "d"

> x[4:5]        #elements 4 to 5
[1] "d" "e"

> i = c(1, 5, 7)
> x[i]         #elements 1, 5, 7
[1] "a" "e" "g"
```

Examples illustrating assignment are:

```

> x = seq(0, 20, 2)
> x[4] = NA           #assigns a missing value
> x
[1] 0 2 4 NA 8 10 12 14 16 18 20

> x[4:5] = c(6, 7)   #assigns two values
> x
[1] 0 2 4 6 7 10 12 14 16 18 20

> i = c(3, 5, 7)
> x[i] = 0           #assigns 3 zeros, at positions i
> x
[1] 0 2 0 6 0 10 0 14 16 18 20

```

Sometimes it is easier to specify what to “leave out” rather than what to include. In this case, we can use negative indices. An expression `x[-2]`, for example, is evaluated as all elements of `x` *except* the second one. The negative sign can also be used with a vector of indices, as shown below.

```

> x = seq(0, 20, 5)
> x
[1] 0 5 10 15 20

> i = c(1, 5)
> y = x[-i]         #leave out the first and last element
> y
[1] 5 10 15

```

A.2 The sort and order functions

The `sort` function sorts the values of a vector in ascending order (by default) or descending order. At times, one wants to sort several variables according to the order of one of the variables. This can be accomplished with the `order` function.

Example A.1 (The `order` function). Suppose that we have the following data for temperatures and ozone levels

```

> temps = c(67, 72, 74, 62)
> ozone = c(41, 36, 12, 18)

```

and wish to sort the pairs (`ozone`, `temps`) in increasing order of `ozone`. The expression `order(ozone)` is a vector of indices that can be used to rearrange `ozone` into ascending order. This is the same order required for `temps`, so this order is what we need as an index vector for `temps`.

```

> oo = order(ozone)
> oo
[1] 3 4 2 1

> Ozone = sort(ozone)  #same as ozone[oo]

```

```
> Temps = temps[oo]

> Ozone
[1] 12 18 36 41
> Temps
[1] 74 62 72 67
```

R_x A.1 *In Example A.1 we used `sort` to sort the values of `ozone`; however, it is not really necessary to sort (again) because `order(ozone)` contains the information for sorting `ozone`. In this example, `sort(ozone)` is equivalent to `ozone[oo]`.*

See Section 2.7.3 for an example of sorting a data frame.

A.3 Matrices

A.3.1 Creating a matrix

A matrix can be created using the `matrix` function. The basic syntax is `matrix(x, nrow, ncol)`, where `x` is a constant or a vector of values, `nrow` is the number of rows and `ncol` is the number of columns. For example, to create a 4 by 4 matrix of 0's, we use

```
> matrix(0, 4, 4)
      [,1] [,2] [,3] [,4]
[1,]  0   0   0   0
[2,]  0   0   0   0
[3,]  0   0   0   0
[4,]  0   0   0   0
```

To create a matrix with specified elements, supply those elements in a vector as the first argument to `matrix`.

```
> X = matrix(1:16, 4, 4)
> X
      [,1] [,2] [,3] [,4]
[1,]   1   5   9  13
[2,]   2   6  10  14
[3,]   3   7  11  15
[4,]   4   8  12  16
```

The number of rows, number of columns, and dimension of a matrix that is already in the R workspace is returned by the functions `nrow` (or `NROW`), `ncol`, and `dim`, respectively.

```
> nrow(X)
[1] 4
> NROW(X)
[1] 4
> ncol(X)
```

```
[1] 4
> dim(X)
[1] 4 4
```

R_x A.2 (`NROW`, `nrow`, `length`) *It is helpful to know when to use `NROW`, `nrow`, or `length`. `length` gives the length of a vector, and `nrow` gives the number of rows of a matrix. But `length` applied to a matrix does not return the number of rows, but rather, the number of entries in the matrix. An advantage of `NROW` is that it computes `length` for vectors and `nrow` for matrices. This is helpful when the object could be either a vector or a matrix.*

Notice that when we supply the vector `x=1:16` as the entries of the matrix `X`, the matrix was filled in column by column. We can change this pattern with the optional argument `byrow=TRUE`.

```
> matrix(1:16, 4, 4)
      [,1] [,2] [,3] [,4]
[1,]    1    5    9   13
[2,]    2    6   10   14
[3,]    3    7   11   15
[4,]    4    8   12   16

> A = matrix(1:16, 4, 4, byrow=TRUE)
> A
      [,1] [,2] [,3] [,4]
[1,]    1    2    3    4
[2,]    5    6    7    8
[3,]    9   10   11   12
[4,]   13   14   15   16
```

The row and column labels of the matrix also indicate how to extract each row or column from the matrix. To extract all of row 2 we use `A[2,]`. To extract all of column 4 we use `A[,4]`. To extract the element in row 2, column 4, we use `A[2, 4]`.

```
> A[2,]
[1] 5 6 7 8
> A[,4]
[1] 4 8 12 16
> A[2,4]
[1] 8
```

A submatrix can be extracted by specifying a vector of row indices and/or a vector of column indices. For example, to extract the submatrix with the last two rows and columns of `A` we use

```
> A[3:4, 3:4]
      [,1] [,2]
[1,]   11   12
[2,]   15   16
```

To omit a few rows or columns we can use negative indices, in the same way that we did for vectors. To omit just the third row, we would use

```
> A[-3, ]
      [,1] [,2] [,3] [,4]
[1,]    1    2    3    4
[2,]    5    6    7    8
[3,]   13   14   15   16
```

The rows and columns of matrices can be named using the `rownames` or `colnames`. For example, we can create names for `A` as follows.

```
> rownames(A) = letters[1:4]
> colnames(A) = c("FR", "SO", "JR", "SR")
> A
  FR SO JR SR
a  1  2  3  4
b  5  6  7  8
c  9 10 11 12
d 13 14 15 16
```

Now one can optionally extract elements by name. To extract the column labeled "JR" we could use `A[, 3]` or

```
> A[, "JR"]
  a b c d
  3 7 11 15
```

A.3.2 Arithmetic on matrices

The basic arithmetic operators (`+` `-` `*` `/` `^`) on matrices apply the operations elementwise, analogous to vectorized operations. This means that if $A = (a_{ij})$ and $B = (b_{ij})$ are matrices with the same dimension, then $A*B$ is a matrix of the products $a_{ij}b_{ij}$. Multiplying a matrix `A` above with itself using the `*` operator squares every element of the matrix.

```
> A = matrix(1:16, 4, 4, byrow=TRUE)
> A
      [,1] [,2] [,3] [,4]
[1,]    1    2    3    4
[2,]    5    6    7    8
[3,]    9   10   11   12
[4,]   13   14   15   16
> A * A
      [,1] [,2] [,3] [,4]
[1,]    1    4    9   16
[2,]   25   36   49   64
[3,]   81  100  121  144
[4,]  169  196  225  256
```

The exponentiation operator is also applied to each entry of a matrix. The R expression `A^2` is evaluated as the matrix of squared elements a_{ij}^2 , not the matrix product AA .

```
> A^2      #not the matrix product
      [,1] [,2] [,3] [,4]
[1,]    1    4    9   16
[2,]   25   36   49   64
[3,]   81  100  121  144
[4,]  169  196  225  256
```

Matrix multiplication is obtained by the operator `%*%`. To obtain the square of matrix A (using matrix multiplication) we need `A %*% A`.

```
> A %*% A  #the matrix product
      [,1] [,2] [,3] [,4]
[1,]   90  100  110  120
[2,]  202  228  254  280
[3,]  314  356  398  440
[4,]  426  484  542  600
```

Many of the one-variable functions in R will be applied to individual elements of a matrix, also. For example, `log(A)` returns a matrix with the natural logarithm $\log(a_{ij})$ as its entries.

```
> log(A)
      [,1]      [,2]      [,3]      [,4]
[1,] 0.000000 0.6931472 1.098612 1.386294
[2,] 1.609438 1.7917595 1.945910 2.079442
[3,] 2.197225 2.3025851 2.397895 2.484907
[4,] 2.564949 2.6390573 2.708050 2.772589
```

The `apply` function can be used to *apply* a function to rows or columns of a matrix. For example, we obtain the vector of column minimums and the vector of column maximums of A by

```
> apply(A, MARGIN=1, FUN="min")
[1] 1 5 9 13

> apply(A, MARGIN=2, FUN="max")
[1] 13 14 15 16
```

Row means and column means can be computed using `apply` or by

```
> rowMeans(A)
[1] 2.5 6.5 10.5 14.5

> colMeans(A)
[1] 7 8 9 10
```

The `sweep` function can be used to *sweep* out a statistic from a matrix. For example, to subtract the minimum of the matrix and divide the result by its maximum we can use

```
> m = min(A)
> A1 = sweep(A, MARGIN=1:2, STATS=m, FUN="-") #subtract min

> M = max(A1)
> A2 = sweep(A1, 1:2, M, "/") #divide by max
```

```
> A2
      [,1]      [,2]      [,3]      [,4]
[1,] 0.0000000 0.06666667 0.1333333 0.2000000
[2,] 0.2666667 0.33333333 0.4000000 0.4666667
[3,] 0.5333333 0.60000000 0.6666667 0.7333333
[4,] 0.8000000 0.86666667 0.9333333 1.0000000
```

Here we specified `MARGIN=1:2` indicating all entries of the matrix. The default function is subtraction, so the `"-"` argument could have been omitted in the first application of `sweep`.

The column mean can be subtracted from each column by

```
> sweep(A, 2, colMeans(A))
      [,1] [,2] [,3] [,4]
[1,]  -6  -6  -6  -6
[2,]  -2  -2  -2  -2
[3,]   2   2   2   2
[4,]   6   6   6   6
```

Example 1.8 in Section 1.3 illustrates matrix operations.

A.4 Lists

One limitation of vectors, matrices, and arrays is that each of these types of objects may only contain one type of data. For example, a vector may contain all numeric data or all character data. A list is a special type of object that can contain data of multiple types.

Objects in a list can have names; if they do, they can be referenced using the `$` operator. Objects can also be referenced using double square brackets `[[]]` by name or by position.

Example A.2 (Creating a list). If the results of Example 1.3 (fatalities due to horsekicks) are not in the current workspace, run the script “horsekicks.R” discussed in Section 1.1.3. Now suppose that we would like to store the data (`k` and `x`), sample mean `r` and sample variance `v`, all in one object. The data are vectors `x` and `k`, both of length 5, but the mean and variance are each length 1. This data cannot be combined in a matrix, but it can be combined in a list. This type of list can be created as

```
> mylist = list(k=k, count=x, mean=r, var=v)
```

The contents of the list can be displayed like any other object, by typing the name of the object or by the `print` function.

```
> mylist
$k
[1] 0 1 2 3 4

$count
[1] 109 65 22 3 1
```



```
$mean
[1] 0.61
```

```
$var
[1] 0.6109548
```

Names of list components can be displayed by the `names` function.

```
> names(mylist)
[1] "k"      "count"  "mean"   "var"
```

All of the components of this list have names, so they can be referenced by either name or position. Examples follow.

```
> mylist$count      #by name
[1] 109 65 22 3 1
```

```
> mylist[[2]]      #by position
[1] 109 65 22 3 1
```

```
> mylist["mean"]  #by name
$mean
[1] 0.61
```

A compact way to describe the contents of a list is to use the `str` function. It is a general way to describe any object (`str` is an abbreviation for structure).

```
> str(mylist)
List of 4
 $ k      : num [1:5] 0 1 2 3 4
 $ count: num [1:5] 109 65 22 3 1
 $ mean  : num 0.61
 $ var   : num 0.611
```

The `str` function is particularly convenient when the list is too large to read on one screen or the object is a large data set.

Many functions in R return values that are lists. An interesting example is the `hist` function, which displays a histogram; its return value (a list) is discussed in the next example.

Example A.3 (The histogram object (a list)). One of the many data sets included in the R installation is `faithful`. This data records waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA. We are interested in the second variable `waiting`, the waiting time in minutes to the next eruption.

We construct a frequency histogram of the times between eruptions using the `hist` function, displayed in [Figure A.1](#). The heights of the bars correspond to the counts for each of the bins. The histogram has an interesting shape. It is clearly not close to a normal distribution, and in fact it has two modes; one near 50-55, and the other near 80.

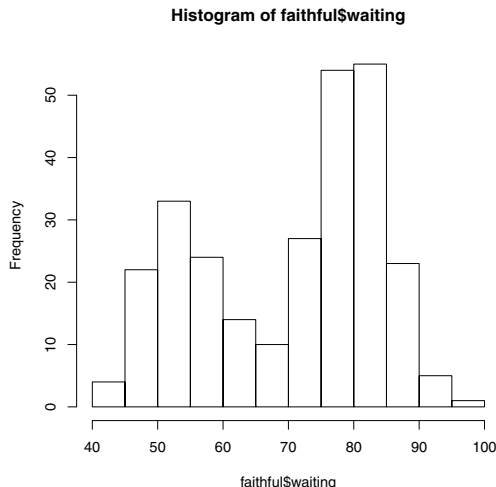


Fig. A.1 Histogram of waiting times between Old Faithful eruptions in Example A.3.

Typically one is only interested in the graphical output of the `hist` function, but some useful information is returned by the function. That information can be saved if we assign the value of the function to an object. Here we saved the result of `hist` in an object `H`. The object `H` is actually a list; it stores a variety of information about the histogram such as the bin counts, breaks (endpoints), midpoints of intervals, etc.

```
> H = hist(faithful$waiting)

> names(H)
[1] "breaks"      "counts"      "intensities" "density"
[5] "mids"        "xname"       "equidist"
```

The endpoints of the intervals (the *breaks*) are

```
> H$breaks
[1] 40 45 50 55 60 65 70 75 80 85 90 95 100
```

The frequencies in each interval (the bin *counts*) are

```
H$counts
[1] 4 22 33 24 14 10 27 54 55 23 5 1
```

The help topic for `hist` describes each of the components of the list `H` in the *Value* section.

To create a list, use the `list` function; a simple example is at the beginning of this section (Example A.2). There are several other examples throughout this book that illustrate how to create lists.

A.5 Sampling from a data frame

To draw a random sample of observations from a data frame, use the `sample` function to sample the row indices or the row labels, and extract these rows. Refer to the `USArrests` data introduced in Example 1.9 on page 20. To obtain a random sample of five states,

```
> i = sample(1:50, size=5, replace=FALSE)
> i
[1] 30  1  6 50  3
> USArrests[i, ]
      Murder Assault UrbanPop Rape
New Jersey   7.4    159      89 18.8
Alabama     13.2    236      58 21.2
Colorado     7.9    204      78 38.7
Wyoming      6.8    161      60 15.6
Arizona      8.1    294      80 31.0
```

Alternately, we could have sampled the row labels:

```
> samplerows = sample(rownames(USArrests), size=5, replace=FALSE)
> samplerows
[1] "North Dakota" "West Virginia" "Montana"      "Idaho"
[5] "Virginia"
> USArrests[samplerows, ]
      Murder Assault UrbanPop Rape
North Dakota  0.8     45      44  7.3
West Virginia 5.7     81      39  9.3
Montana       6.0    109      53 16.4
Idaho        2.6    120      54 14.2
Virginia     8.5    156      63 20.7
```

References

1. Albert, J. (2009), *Bayesian Computation with R*, Springer, New York.
2. Agresti, A. and Coull, B. (1998), "Approximate is better than 'exact' for interval estimation of binomial proportions," *The American Statistician*, 52, 119-126.
3. Ashenfelter, O. and Krueger, A. (1994), "Estimates of the economic return to schooling from a new sample of twins," *The American Economic Review*, 84, 1157-1173.
4. Bolstad, W. (2007), *Introduction to Bayesian Statistics*, Wiley-Interscience.
5. Box, G. E. P., Hunter W. G. and Hunter J. S. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Wiley, New York.
6. Cameron E. and Pauling L. (1978), "Experimental studies designed to evaluate the management of patients with incurable cancer," *Proc. Natl. Acad. Sci. U. S. A.*, 75, 4538-4542.
7. Canty, A. and Ripley. B. (2010), *boot: Bootstrap R Functions*, R package version 1.2-43.
8. Casella, G., and George, E. (1992), "Explaining the Gibbs sampler," *The American Statistician*, 46, 167-174.
9. Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, 49, 327-335.
10. Cleveland, W. (1979), "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, 74, 829-83.
11. Carmer, S. G. and Swanson, M. R. (1973), "Evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods," *Journal of the American Statistical Association*, 68, 66-74.
12. DASL: The Data and Story Library, <http://lib.stat.cmu.edu/DASL/>.
13. Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge.
14. Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall/CRC, Boca Raton, FL.
15. Everitt, B. S., Landau, S., And Leese, M. (2001), *Cluster Analysis*, 4th edition, Oxford University Press, Inc. New York.
16. Faraway, J. (2002), "Practical regression and ANOVA Using R," Contributed documentation in PDF format available at <http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>.
17. Friendly, M. (2002), "A brief history of the mosaic display," *Journal of Computational and Graphical Statistics*, 11, 89-107.
18. Fienberg, S. E. (1971), "Randomization and social affairs: The 1970 draft lottery," *Science*, 171, 255-261.

19. Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2003), *Bayesian Data Analysis*, second edition, Chapman and Hall, New York.
20. Gentle, J. E. (2003), *Random Number Generation and Monte Carlo Methods*, second edition, Springer, New York.
21. Hand D. J., Daly F., Lunn A. D., McConway K. J., Ostrowski E. (1994), *A Handbook of Small Data Sets*, Chapman and Hall, London.
22. Hoaglin, D., Mosteller, F., and Tukey, J. (2000), *Understanding Robust and Exploratory Data Analysis*, Wiley-Interscience.
23. Hoff, P. (2009), *A First Course in Bayesian Statistics*, Springer, New York.
24. Hogg, R. and Klugman, S. (1984), *Loss Distributions*, Wiley, New York.
25. Hollander. M. and Wolfe, D. (1999), *Nonparametric Statistical Methods*, second edition, Wiley, New York.
26. Hornik, K (2009). *R FAQ: Frequently Asked Questions on R*, Version 2.13.2011-04-07, <http://cran.r-project.org/doc/FAQ/R-FAQ.html>.
27. Hothorn, T., Hornik, K., van de Wiel, M. and Zeileis, A. (2008), "Implementing a class of permutation tests: The coin package," *Journal of Statistical Software*, 28, 1-23.
28. Koopmans, L. (1987), *Introduction to Contemporary Statistical Methods*, Duxbury Press.
29. Labby, Z. (2009), "Weldon's dice, automated" *Chance*, 22, 258-264.
30. Larsen, R. J. and Marx, M. L. (2006), *An Introduction to Mathematical Statistics and Its Applications*, 4th edition, Pearson Prentice Hall, Saddle River, New Jersey.
31. Leisch, F. (2007), *bootstrap: Functions for the Book "An Introduction to the Bootstrap"*, R package version 1.0-22.
32. Meyer, D., Zeileis, A., and Hornik, K. (2010), *vcd: Visualizing Categorical Data*, R package version 1.2-9.
33. Montgomery, D. G. (2001), *Design and Analysis of Experiments*, 5th edition., Wiley, New York.
34. Moore, G. (1965), "Cramming more components onto integrated circuits," *Electronics Magazine*, April, 114-117.
35. Mosteller, F. (2010), *The Pleasures of Statistics*, Springer, New York.
36. Mosteller, F. and Wallace, D. L. (1984), *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*, Springer-Verlag, New York.
37. Murrell, P. (2006), *R Graphics*, Chapman and Hall, Boca Raton, Florida.
38. *OzDASL, Australian Data and Story Library*, <http://www.statsci.org/data/index.html>.
39. Pearson, K. (1900), "On the criterion that a given system of derivations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Philosophical Magazine*, 5, 157-175.
40. R Development Core Team (2011), "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>.
41. The R Development Core Team (2011), "R: A Language and Environment for Statistical Computing, Reference Index." Version 2.13.0 (2011-04-13)
42. Ross, S. (2007), *Introduction to Probability Models*, 9th edition, Academic Press, Burlington, MA.
43. Sarkar, D. (2008), *Lattice: Multivariate Data Visualization with R*, Springer, New York.
44. Smith, J. M. and Misiak, H. (1973), "The effect of iris color on critical flicker frequency (CFF)," *Journal of General Psychology* 89, 91-95.
45. Snell, L. (1988), *Introduction to Probability*, Random House, New York.

46. Starr, N. (1997), “Nonrandom risk: the 1970 draft lottery. *Journal of Statistics Education*,” 5, <http://www.amstat.org/publications/jse/v5n2/datasets.starr.html>
47. Tukey, J. (1977), *Exploratory Data Analysis*, Addison-Wesley.
48. “Campaign 2004: Time-tested formulas suggest both Bush and Kerry will win on Nov. 2,” USA Today, http://www.usatoday.com/news/politicselections/nation/president/2004-06-23-bush-kerry-cover_x.htm.
49. Venables, W. N., Smith, D. M. and the R Development Core Team (2011). “An Introduction to R”, Version 2.13.0 (2011-04-13).
50. Venables, W. N. and Ripley, B. D. (2002), *Modern Applied Statistics with S*, fourth edition. Springer, New York.
51. Verzani, J. (2010), *UsingR: Data sets for the text Using R for Introductory Statistics*,” R package version 0.1-13.
52. Wickham, H. (2009), *ggplot2: Elegant Graphics for Data Analysis*, Springer, New York.
53. *The Washington Post* (2007) “Heads and Shoulders above the Rest,”, http://blog.washingtonpost.com/44/2007/10/11/head_and_shoulders_above.html.
54. Wikipedia, “Heights of Presidents of the United States and presidential candidates,” retrieved June 28, 2011.
55. Wilkinson, L. (2005), *The Grammar of Graphics*, second edition, Springer, New York.
56. Willerman, L., Schultz, R., Rutledge, J. N., and Bigler, E. (1991), “In vivo brain size and intelligence,” *Intelligence*, 15, 223–228.

Index

`:`, 5, 337
`??`, 13
`?`, 13
NROW, 15, 20, 340, 341
TukeyHSD, 213, 214, 231, 238
abbreviate, 189
abline, 32, 59, 60, 85, 106, 110, 111,
128, 140, 141, 143, 149, 167, 175,
178, 189, 193, 209, 320
add4ci, 158
anova method for aov, 206–209, 236
anova method for lm, 184, 185, 206–209,
213
any, 22
aov, 206, 207, 213, 214, 217, 230, 236
apply, 18, 66, 317, 318, 343
arrows, 115
as.data.frame, 225
as.factor, 216
as.integer, 49
as.matrix, 21, 72
as.numeric, 271
attach, 55, 101, 187, 202, 230, 245
axis, 117
barchart, 121
barplot, 6, 8, 32, 80, 83, 87, 92, 121
binom.test, 157, 170
boxplot, 49, 64, 121, 137, 138, 163, 202,
216, 233, 245, 317
box, 120, 129
bwplot, 121, 130
by, 54, 203, 211, 216, 245
cbind, 10, 83, 169
chisq.test, 85, 94, 95, 97–99
cloud, 125
colMeans, 343
colnames, 17, 21, 342
complete.cases, 134, 168
coplot, 50
cor.test, 58, 250
cor, 25, 56, 65, 179, 188
cumsum, 256, 273
curve, 16, 67, 114, 129, 130, 189, 192,
280, 282, 303, 304, 312, 320
cutree, 77
cut, 87, 90, 98
c, 3, 4, 7, 220, 337
data.frame, 5, 26, 68, 83, 180, 186, 195,
229
data, 19
dbinom, 82, 97, 113, 273
dcauchy, 335
densityplot, 124, 130
density, 67
detach, 191
dev.off, 121
dexp, 325, 326
dgamma, 282, 285, 286, 290
diag, 65, 89
diff, 60
dimnames, 21
dim, 20, 260, 340, 341
dist, 72, 74
dnbinom, 299
dnorm, 114, 290, 303
example, 13
expression, 114, 115
factor, 81, 201, 229
floor, 10
function, 14
getwd, 11
grid, 269
hclust, 73, 74
head, 44, 68, 169, 215, 221, 228, 229, 244

help.search, 13
 help, 13
 histogram, 121
 hist, 23, 66, 108, 121, 150, 159, 167,
 169, 305, 312
 identify, 111, 112, 129, 136, 141, 149,
 189
 ifelse, 68, 154, 163, 168, 267
 integrate, 15
 interaction.plot, 236
 is.data.frame, 44
 is.factor, 201, 216, 221
 is.matrix, 44
 is.na, 22, 58
 jpeg, 121
 layout, 116
 legend, 50, 55, 106
 length, 80, 267, 341
 levels, 201
 library, 122, 228
 lines, 59, 60, 84, 105, 107, 119, 128
 line, 129, 139, 149
 list, 225, 344
 lm, 143, 144, 150, 174–175, 177, 180,
 183, 188, 207, 213, 246
 locator, 115, 119
 lowess, 59, 60, 105
 matrix, 16–19, 323, 340
 max, 70, 271, 343
 mean, 5, 59, 65, 66, 269
 median, 62, 63, 68, 312, 313
 mfcoll, 37
 mfrow, 37
 min, 343
 model.tables, 207, 208, 231, 237
 mosaicplot, 88, 96, 98
 mtext, 117
 na.omit, 220, 244
 names, 21, 61, 82, 95, 155, 162, 220, 230,
 245, 345
 ncol, 340, 341
 nrow, 20, 340, 341
 oneway.test, 204, 247, 248
 options, 36
 order, 71, 339
 outer, 211
 pairs, 25, 55, 179
 pairwise.t.test, 212, 239
 par, 37, 113, 117, 129, 217
 paste, 115
 pbinom, 157
 pbirthday, 274
 pchisq, 95
 pdf, 121
 plot.new, 119, 129
 plot.window, 119, 129
 plot, 25, 55, 59, 60, 63, 72, 85, 89,
 102–104, 106–109, 111, 113, 114,
 117, 121, 127, 139–141, 143, 149,
 160, 175, 178, 188, 189, 193, 258,
 260, 261, 272, 332
 plot method for TukeyHSD, 214, 217,
 232, 238
 plot method for aov, 217, 232
 plot method for density, 67
 plot method for hclust, 74
 plot method for lm, 176, 181, 193, 209,
 246
 pmax, 99
 pnorm, 303
 points, 104, 119
 polygon, 72, 309
 postscript, 121
 predict, 180
 predict method for lm, 185, 186, 207
 print, 11
 prop.table, 87, 91, 98, 260
 prop.test, 155, 156, 170
 qbeta, 304, 305
 qgamma, 283, 285
 qnorm, 303, 319
 qqline, 67, 159, 209
 qqnorm, 67, 159, 209
 qtkey, 212, 213
 qt, 211
 quantile, 301, 313
 rbeta, 305, 331, 332
 rbind, 10, 31
 rbinom, 259, 319, 332
 read.csv, 30, 222, 234
 read.table, 29–31, 53, 61, 82, 134, 163,
 187, 201, 215, 219, 220, 244
 replicate, 167, 255, 258, 260, 264,
 267–269, 272–276, 312, 315
 rep, 287, 326, 337
 return, 15
 rev, 6
 rexp, 312, 326
 rgamma, 284, 291
 rgam, 331
 rnorm, 331
 round, 56
 rowMeans, 66, 343
 rowSums, 18
 rownames, 17, 21
 rpois, 291
 runif, 309, 326

- sample, 99, 167, 247, 255, 256, 262, 266, 267, 271, 273, 275, 276, 315, 323, 347
- sapply, 62, 63, 228, 265, 269, 320
- scan, 80, 96
- seq, 5, 337
- setwd, 11
- smoothScatter, 300
- solve, 19
- sort, 339
- source, 11
- sqrt, 31
- stack, 220, 229, 244
- stem, 145
- stripchart, 135, 136, 138, 149, 151, 202, 216, 245
- str, 21, 28, 29, 230, 234, 345
- subset, 69, 73, 134, 145, 148, 168
- summary, 22, 44, 49, 53, 221, 244, 245, 268
- summary method for aov, 230
- summary method for lm, 184, 194, 214
- sum, 89, 263, 309
- sweep, 343
- t.test, 158–162, 164–167, 169–172
- table, 6, 31, 80, 87, 88, 90, 91, 96–99, 154, 166, 211, 258, 260, 261, 264, 267, 276, 291, 324, 332
- tail, 71, 221, 229
- text, 73, 104, 110, 111, 114, 117, 119, 128, 130
- title, 104, 115
- truehist, 23, 67
- t, 92
- unclass, 201
- unique, 267
- var, 14, 65
- which.max, 70
- which.min, 70
- which, 58, 69
- wilcox.test, 161, 162, 165, 166, 170–172
- windows, 116
- with, 24, 236, 245
- xypLOT, 51, 121–123, 292

- abline, 178, 180
- acceptance rate
 - of Metropolis-Hastings algorithm, 299
- aesthetics: ggplot2 package, 126
- Agresti-Coull interval, 157
- Analysis of Variance, 227
- analysis of variance, 199
- ANOVA, 199, 217, 227
 - calculations, 224
 - coefficients, 207
 - comparison of means, 231
 - data entry, 200
 - F test, 230, 236
 - fitted values, 207
 - one-way F test, 204
 - one-way model, 205
 - randomized block model, 229
 - residual plot, 209
 - table, 208, 214
 - table of effects, 208
 - table of means, 207, 237
 - two-way factorial design, 234
- ANOVA table for regression, 185
- arrow
 - drawing, 115
- assignment operator, 2

- bar graph, 80, 87
- barplot, 6
- Bayesian
 - interval estimate, 283, 285
 - point estimate, 282
- beta density, 332
- bias
 - of estimator, 316
- binomial cumulative distribution, 157
- binomial distribution, 82, 85, 113, 156, 332
- block, 229
- block design, 227
- bootstrap package, 228
- boxplot, 64, 216, 233
- boxplots
 - parallel, 163

- checking function, 293
- chi-square distribution, 94
- chi-square statistic, 155
- Clopper-Pearson confidence interval, 157
- cluster analysis, 73
- coefficient of determination, 194
- coin package, 248
- colors in graphing, 108
- column proportions, 91
- comparison of means, 237
- conditional graph
 - lattice, 123
- conditional plot, 50, 51
- confidence interval
 - difference in means, 170
 - for a mean, 160
 - for a median, 161, 311, 313
 - for a proportion, 156

- for difference of two locations, 166
 - Wald, 318
- contingency table, 88, 91
- continuity-adjusted statistic
 - for testing a proportion, 156
- contributed documentation, 38
- coordinate system
 - setting up, 119
- coplot, 51
- correlation, 25, 56, 58, 65, 179
- data frame, 19, 26
 - attaching, 24, 25
 - detaching, 24
 - extracting elements, 23
 - extracting variables, 23
 - importing, 26
 - missing values in, 22
 - sampling from, 347
 - structure, 21
 - summary statistics, 22
- dendrogram, 74
- density estimate, 67
- density estimate plot, 288
- differences, 60
- distance, 71, 72
- distance matrix, 72
- draft lottery, 61
- Euclidean distance, 72
- Example
 - order function, 339
 - ANOVA calculations, 224
 - Brain and body size of mammals, 44–47
 - cancer survival times, 217
 - car gas consumption, 122
 - cars, 174
 - cars, cont., 177
 - Cherry Trees Model 2, 181
 - Cherry Trees Model 3, 184
 - Cherry Trees, interval estimates, 186
 - Class mobility, 16
 - coin flipping, 79
 - coin-tossing game, 255
 - collecting baseball cards, 265
 - college ratings, 134
 - confidence interval for a proportion, 318
 - Creating a list, 344
 - deciding authorship, 277
 - die rolling, 81
 - Digits of π , 30
 - Exam scores, 227, 229
 - flicker, 199, 204, 207, 213, 224
 - flipping random coin, 331
 - function definition, 14
 - functions as arguments, 15
 - graphing a function using `curve`, 16
 - hat problem, 262
 - histogram object (a list), 345
 - home run hitting, 101
 - horsekicks, 7
 - IQ of twins, 48
 - LSD method, 210
 - lunatics data, 187
 - mammals, 44
 - mammals (cluster analysis), 73
 - mammals (distances), 71
 - mammals, cont., 68
 - Massachusetts Lunatics, 29
 - mile run records, 126
 - Moore’s Law, 191
 - New Haven temperatures, 59
 - pattern in website counts, 249
 - Poisons, 234
 - President’s heights, 4
 - random walk, 321
 - Randomization test, 246
 - RANDU, 65
 - Sample means, 65
 - sample means, 125
 - sampling distribution of a median, 311, 325, 329
 - SiRstv from NIST, 215
 - sleeping patterns, 153
 - Sleepless in Seattle, 308
 - sorting mammals, 70
 - streaky hitting in baseball, 270
 - taxi problem, 314
 - Temperatures, 3
 - Tukey’s multiple comparison procedure, 213
 - twins dataset, 85, 163
 - Two-way ANOVA, 236, 237
 - USArrests, 20
 - USArrests, cont., 22
 - Using the bootstrap package, 33
 - Volume of Black Cherry Trees, 178
 - Waste Run-up data, 243
 - Weldon’s dice, 82
 - expected counts, 84, 94
 - expected outcomes, 82
 - exponential density, 311, 325
 - exponential fitting, 144
 - exporting graph, 120
 - faceting: `ggplot2` package, 126

- factor, 81, 199, 201–202, 229
- factor variable, 81
- factorial design, 235
- figure region, 116
- Fisher Least Significant Difference, 210–212
- five-number summary, 137
- folded fraction, 146
- folded log, 146
- folded root, 146
- formula, 206
- frequency table, 80, 82, 87, 90, 258, 260, 264, 278, 324
- functions, 14

- gamma density, 279, 282, 286
- gamma quantile, 283
- geoms: `ggplot2` package, 126
- grammar of graphics, 126
- graph history, 37
- graphical parameters, 37
- graphics device, 121
- graphics frame
 - opening, 119
- group comparison
 - lattice, 124

- help system, 13
- histogram, 23, 66, 67, 159, 284, 292, 312
- Holm's p -value adjustment, 239
- hypothesis test
 - about a median, 161

- importing data, 26
- inline, 216
- inter-quartile range, 245
- interaction, 234, 235
- interaction plot, 236
- interval estimate
 - Bayesian, 301

- Jacobian of a transformation, 298

- labeling outliers, 141
- least squares, 173
- least-squares line, 143
- legend
 - in a plot, 107
- levels, 199
- likelihood function, 279, 296
- line graph, 80, 260
- line styles, 106
- linear model, 173
- lists, 344–346

- log-logistic density, 296
- lowess, 60
- lowess smoother, 105
- LSD, 210

- Mann-Whitney-Wilcoxon test, 165
- marginal posterior density, 300
- marginal probability distribution, 332
- Markov chain simulation
 - discrete case, 321
- mathematics
 - displaying, 114
- matrix, 16, 340
 - arithmetic, 342
 - converting to, 21
 - creating, 17, 19, 340
 - diagonal, 65
 - extracting elements, 18, 341
 - multiplication, 18, 343
 - negative indices, 341
 - operations, 19
 - row and column names, 17, 342
- maximum, 70
- mean absolute error
 - of estimator, 316
- mean squared error, 204
- median, 62
- median split, 68
- Metropolis-Hastings algorithm, 285, 298
 - acceptance rate, 288, 327
 - burn-in period, 327
 - good mixing, 327
 - standard deviation parameter, 328
 - trace plot, 327
- missing values, 54, 56, 58
- model formula, 174
- Monte Carlo estimate
 - of a mean, 307, 310
 - of a probability, 309
 - standard error, 308, 309
- Monte Carlo experiment, 258
 - coverage probability of an interval estimate, 319
- Moore's Law, 191–195
- mosaic plot, 88, 89, 96
- MSE, 208
- multiple comparisons, 210–214
- multiple graphs in window, 256

- negative binomial density, 299
- negative binomial distribution, 295
- normal density
 - as a prior, 289
- normal distribution, 113

- normal probability plot, 159
- normal-QQ plot, 209
- one-dimensional scatterplot, 135
 - options, 36
- outliers, 139
 - identifying, 112, 136
- overdispersion parameter, 300
- p-value, 58, 85, 94, 155, 161, 162, 165, 167
- p-value adjustment, 212
- pairs plot, 55–57
- Pearson residuals, 85, 95, 96
- Pearson’s chi-square test, 94
- permutation
 - random, 262
- permutation test, 243
 - for comparing two samples, 166
- plot, 178, 180
- plot region, 116
- plotting character, 103
- Poisson
 - density, 9
 - distribution, 9
 - probability functions, 10
- Poisson density, 291
- Poisson distribution, 278
- position adjustments: `ggplot2` package, 126
- posterior density, 281, 297
- posterior distribution
 - by simulation, 284
- posterior median, 282
- posterior predictive distribution, 292
- prediction interval, 186
- predictive distribution, 291
- predictor, 173
- prior density, 279, 296
 - construction, 279
- prompt character, 3
- PropCIs package, 158
- QQ plot, 67, 194, 232, 246
- quantiles
 - posterior, 285
- quartile spread, 139
- R Graph Gallery, 38
- R Graphical Manual, 38
- R Wiki, 38
- Randomization test, 246
- randomization test, 243
- randomized block design, 227
- regression, 173, 174, 188
 - adjusted R^2 , 184
 - ANOVA, 185
 - confidence interval, 185, 186
 - fitted line plot, 176
 - multiple linear regression, 181
 - plot fitted line, 175, 178
 - prediction, 180
 - prediction intervals, 185
 - QQ plot of residuals, 181
 - residual plot, 181, 189
 - residual plots, 194
 - residual standard error, 184
 - through the origin, 177, 178
 - relative frequencies, 80
 - relative frequency, 87
 - replicate, 247
 - reproducible research, 39
 - residual analysis, 193
 - residual plot, 85, 140, 143, 176
 - residual plots, 232, 246
 - residuals, 176–177, 193
 - resistant line, 139
 - response, 173, 199, 229
 - row proportions, 91
- sampling
 - with replacement, 256, 266
- sampling density
 - of a median, 311
- sampling distribution, 315
 - empirical, 315
- scale parameter
 - for Metropolis-Hastings algorithm, 299
- scales: `ggplot2` package, 126
- scatterplot, 25, 50, 55, 72, 332
 - lattice, 122
- scripts, 10–12
- segmented bar chart, 92
- sequence, 5, 337
- side-by-side bar plots, 83
- side-by-side barplots, 92
- simple linear regression, 173
- smoother, 194
- sort, 71
- Spearman’s rank test, 250
- stacked format, 201
- stacking data, 220
- standard error
 - of a Monte Carlo estimate, 316, 318
- stationary distribution, 323
- statistic: `ggplot2` package, 126
- stemplot, 145, 147

- stripchart, 135
- stripcharts
 - parallel, 135
- studentized range statistic, 212, 213
- subjective probability, 278
- sum of squared error, 204
- sum of squares, 204

- t statistic
 - for a mean, 158
- t-test
 - for paired differences, 169
 - traditional for two samples, 165
 - Welch for two samples, 165
- t-test statistic
 - in permutation test, 167
- tab character, 30, 134, 220
- Task Views, 38
- test
 - of a normal mean, 160
- test of independence, 93
- testing hypothesis
 - about proportion, 154
- textttNROW, 14
- textttrownames, 342
- time plot, 60
- time series, 59
- time series plot, 102
- trace plot, 288

- transformation
 - parameter, 300
- transforming
 - logarithm, 144
 - parameters, 297
- transition probability matrix, 322
- treatment, 199, 229
- treatment effect, 206
- Tukey Honest Significant Difference, 231
- Tukey's Honest Significant Difference,
 - 213, 217, 238
- Tukey's multiple comparison method,
 - 212–214
- Tukey-Kramer method, 212

- unbiased estimator, 315
- uniform distribution, 308
 - discrete, 314
- unstacked format, 201

- variance, 14
- variance-covariance matrix, 65
- vectorized operations, 3
- vectors, 16
- vertical line plot, 84

- Wilcoxon signed rank method, 161
- Wilson score interval, 156
- working directory, 11