

Appendix A

R Basic Reference Guide

This appendix is intended to provide a brief but broad collection of functions commonly used in R. See references and Sect. 2.8 in Chap. 2 to study R in depth. Type `help(foo)` in the R Console to see the documentation and the complete list of arguments for the function `foo`.

Remember that R is case sensitive, and a is not equal to A.

Table A.1 Help functions

Function	Action
<code>help()</code> or <code>?</code>	Opens help about a topic
<code>help.search()</code> or <code>??</code>	Returns list of topics containing some text
<code>apropos()</code>	Returns list of functions containing some text
<code>find()</code>	Tells what package something is in
<code>find()</code>	Shows which libraries and dataframes are attached
<code>example()</code>	Runs example for a topic
<code>demo()</code>	Runs demo about a topic
<code>args()</code>	Returns syntax of a function
<code>attach()</code>	Makes variables in a dataframe accessible by name
<code>detach()</code>	For a dataframe, stops accessibility by name; for a package, unloads it

Table A.2 Functions for packages and workspace

Function	Action
<code>library()</code>	Loads package or shows available package
<code>save.image()</code>	Saves all objects of R session
<code>save()</code>	Saves list of objects to file
<code>load()</code>	Loads objects from file system to R session
<code>rm()</code>	Removes objects
<code>source()</code>	Runs script in file
<code>sink()</code>	Sends output to text file instead of R Console
<code>savehistory()</code>	Saves command history of session
<code>loadhistory()</code>	Loads previously saved history
<code>q()</code>	Exits R

Table A.3 Constants and special values

Expression	Value
letters	26 lowercase letters
LETTERS	26 uppercase letters
month.abb	Three-letter abbreviation of months
month.name	English names of months
pi	π constant
NA	Not available
NaN	Not a number (e.g. $\frac{0}{0}$)
Inf	Infinite
NULL	No value
1i	Imaginary constant $\sqrt{-1}$

Table A.4 Operators for objects

Operator	Operation
<-	Assigns value
[]	Subscript vector
[[]]	Subscript list
:	Generates sequences
~	Model formula
\$	List indexing

Table A.5 Functions for data types

Create	Verify	Convert	Description
character()	is.character()	as.character()	String type
numeric()	is.numeric()	as.numeric()	Numeric type
logical()	is.logical()	as.logical()	Logical data
vector()	is.vector()	as.vector()	One-dimensional data
matrix()	is.matrix()	as.matrix()	Two-dimensional data
array()	is.array()	as.array()	Any dimensional data
factor()	is.factor()	as.factor()	Categorical variable in a vector
list()	is.list()	as.list()	Collection of objects of different types
data.frame()	is.data.frame()	as.data.frame()	Data set as a list of vectors
integer()	is.integer()	as.integer()	Integers
ts()	is.ts()	as.ts()	Time-series objects
complex()	is.complex()	as.complex()	Complex vector (real and imaginary parts)
	class()		Returns the class of any object

Table A.6 Functions for importing and exporting data

Function	Action
<code>read.table()</code>	Imports text files
<code>read.csv()</code> , <code>read.csv2()</code>	Imports CSV files
<code>read.delim()</code>	Imports delimited text files
<code>read.spss()</code>	Imports SPSS files
<code>read.mtp()</code>	Imports Minitab files
<code>read.arff()</code>	Imports Weka files (Data mining)
<code>read.dta()</code>	Imports Stata binary files
<code>write.table()</code>	Writes a text file
<code>write.csv()</code>	Writes a CSV file
<code>writeClipboard()</code>	Saves data in Clipboard to paste in other applications

Table A.7 Functions for managing data

Function	Action
<code>c()</code>	Concatenates elements
<code>seq()</code>	Generates regular sequences
<code>rep()</code>	Replicates values
<code>length()</code>	Gets or sets length of an object
<code>sort()</code>	Sorts vectors
<code>rev()</code>	Reverses vectors
<code>order()</code>	Returns permutation of vector indices
<code>rank()</code>	Returns position of each vector element
<code>unique()</code>	Returns unique values of a vector
<code>duplicated()</code>	Returns indices that contain duplicated values
<code>which()</code>	Gives TRUE indices of logical object
<code>levels()</code>	Returns levels of a factor
<code>unlist()</code>	Converts a list into a vector
<code>cut()</code>	Converts a vector into a factor
<code>transform()</code>	Transforms a dataframe
<code>aggregate()</code>	Gets subtotals of variables in data set
<code>subset()</code>	Gets subset of data set

Table A.8 String functions

Function	Action
<code>nchar()</code>	Counts number of characters in a string
<code>substr()</code>	Returns a substring
<code>paste()</code>	Builds a string with substrings

Table A.9 Functions for exploring data

Function	Action
<code>data()</code>	Loads a data set
<code>head()</code>	Returns first rows of a data set
<code>tail()</code>	Returns last rows of a data set
<code>str()</code>	Returns structure of data set
<code>attributes()</code>	Accesses object attributes

Table A.10 Operators

Operator	Meaning
<code>+</code> , <code>-</code> , <code>*</code> , <code>/</code>	Arithmetic operators
<code>/%</code>	Integer division
<code>%%</code>	Remainder of a division (modulo)
<code>%*%</code>	Matrix multiplication
<code>^</code>	Powers
<code>></code> , <code>>=</code> , <code><</code> , <code><=</code> , <code>==</code> , <code>!=</code>	Relational operators
<code>!</code> , <code>&</code> , <code> </code>	Logical operators

Table A.11 Set operations

Function	Action
<code>union()</code>	Set union
<code>intersection()</code>	Set intersection
<code>setdiff()</code>	Set difference
<code>setequal()</code>	Set equality
<code>is.element()</code>	Set membership

Table A.12 Mathematical functions

Function	Meaning
<code>sqrt()</code>	Square root
<code>log()</code> , <code>exp()</code>	Logarithmic and exponential functions
<code>sin()</code> , <code>cos()</code> , <code>tan()</code> , <code>asin()</code> , <code>acos()</code> , <code>atan()</code>	Trigonometry functions
<code>cosh()</code> , <code>sinh()</code> , <code>tanh()</code> , <code>acosh()</code> , <code>asinh()</code> , <code>atanh()</code>	Hyperbolic functions
<code>round()</code> , <code>trunc()</code> , <code>ceiling()</code> , <code>floor()</code>	Different ways to round a number
<code>abs()</code>	Returns absolute value of a number

Table A.13 Vector functions

Function	Meaning
<code>max()</code> , <code>min()</code>	Maximum and minimum values in vector
<code>sum()</code> , <code>prod()</code>	Sum and product
<code>mean()</code> , <code>var()</code> , <code>sd()</code>	Sample mean, variance, and standard deviation
<code>median()</code>	Median value in vector
<code>range()</code>	Vector with a maximum and minimum
<code>cor()</code>	Correlation between vectors
<code>cumsum()</code> , <code>cumprod()</code> , <code>cummax()</code> , <code>cummin()</code>	Cumulative functions
<code>colMeans()</code> , <code>colSums()</code> , <code>rowMeans()</code> , <code>rowSums()</code>	Totals by rows or columns for dataframes and matrices
<code>ifelse()</code>	Returns a value for each component depending on a logical expression

Table A.14 Summarize functions

Function	Action
<code>apply</code>	Applies function to margins in matrix or vector
<code>lapply</code>	Applies function to all elements of list or vector
<code>tapply</code>	Applies function to ragged array
<code>sapply</code>	Friendly version of <code>lapply</code>
<code>summary</code>	Returns summary of object; for data objects, some statistics
<code>table</code>	Returns data frequency table

Table A.15 Probability distributions

Name	Probability	Random	Quantile	Density	Parameters	Default
Binomial	<code>pbinom()</code>	<code>rbinom()</code>	<code>qbinom()</code>	<code>dbinom()</code>	size, prob	
Normal	<code>pnorm()</code>	<code>rnorm()</code>	<code>qnorm()</code>	<code>dnorm()</code>	mean, sd	0,1
Uniform	<code>punif()</code>	<code>runif()</code>	<code>qunif()</code>	<code>dunif()</code>	min, max	0,1
Poisson	<code>ppois()</code>	<code>rpois()</code>	<code>qpois()</code>	<code>dpois()</code>	lambda	
Exponential	<code>pexp()</code>	<code>rexp()</code>	<code>qexp()</code>	<code>dexp()</code>	rate	1
<i>t</i> -Student	<code>pt()</code>	<code>rt()</code>	<code>qt()</code>	<code>dt()</code>	df	
Chi-square	<code>pchisq()</code>	<code>rchisq()</code>	<code>qchisq()</code>	<code>dchisq()</code>	df	
F	<code>pf()</code>	<code>rf()</code>	<code>qf()</code>	<code>df()</code>	df1, df2	
Beta	<code>pbeta()</code>	<code>rbeta()</code>	<code>qbeta()</code>	<code>dbeta()</code>	shape1, shape2	
Gamma	<code>pgamma()</code>	<code>rgamma()</code>	<code>qgamma()</code>	<code>dgamma()</code>	shape, scale	NULL, 1
Cauchy	<code>pcauchy()</code>	<code>rcauchy()</code>	<code>qcauchy()</code>	<code>dcauchy()</code>	location, scale	0,1
Weibull	<code>pweibull()</code>	<code>rweibull()</code>	<code>qweibull()</code>	<code>dweibull()</code>	shape, scale	NULL, 1
Tukey	<code>ptukey()</code>	—	<code>qtukey()</code>	—	nmeans, nranges	NULL, 1

Table A.16 Sampling and combinatorial

Function	Action
<code>set.seed()</code>	Sets seed for replicable randomization
<code>sample()</code>	Extracts sample from a set
<code>choose()</code>	Computes binomial coefficient

Table A.17 Graphic functions

Function	Action
<code>plot()</code>	Produces different types of plots
<code>boxplot()</code>	Produces boxplot
<code>hist()</code>	Produces histogram
<code>coplot()</code>	Conditional plots
<code>dotchart()</code>	Produces dot chart
<code>stripchart()</code>	Produces 1-D scatterplot
<code>mosaicplot()</code>	Produces mosaic plot
<code>matplot()</code>	Plots matrix columns
<code>pairs()</code>	Matrix of scatterplots
<code>persp()</code>	Produces perspective plot (three dimensions)
<code>spineplot()</code>	Produces spine plot and spinogram
<code>text()</code>	Superposes text on a plot
<code>abline()</code>	Prints line in plot
<code>lines()</code>	Superposes line on a plot
<code>points()</code>	Superposes points on a plot
<code>polygon()</code>	Draws polygon
<code>xspline()</code>	Draws an X-spline
<code>identify()</code>	Identifies point in a plot with mouse

Table A.18 Plot arguments

Argument	Meaning
<code>type</code>	Type: p(points), l(lines), b(both), h(histogram), s(steps)
<code>main</code>	Main title
<code>sub</code>	Subtitle
<code>xlab</code>	Label for x axis
<code>ylab</code>	Label for y axis
<code>asp</code>	yx aspect ratio

Table A.19 Graphic parameters

Parameter	Description	Values
<code>col</code>	Color	Names, numbers, hex codes
<code>lwd</code>	Line width	Number
<code>lty</code>	Line type	Number
<code>pch</code>	Point symbol	Number
<code>las</code>	Style of axis labels	Number
<code>cex</code>	Magnifying ratio	Number (1 default)

Table A.20 Hypothesis tests and confidence intervals

Function	Task
<code>binom.test</code>	Exact test for binomial distributions
<code>bartlett.test</code>	Tests equal variances
<code>chisq.test</code>	Test for contingency tables and goodness-of-fit tests
<code>lillie.test</code>	Normality test
<code>pairwise.t.test</code>	Pairwise comparison with correction for multiple testing
<code>poisson.test</code>	Test for Poisson distribution parameter
<code>prop.test</code>	Test for proportions
<code>shapiro.test</code>	Normality test
<code>t.test</code>	Test for means
<code>var.test</code>	Test for comparing variances

Table A.21 Model fitting

Package	Function	Models
<code>lm</code>	<code>stats</code>	Linear model
<code>anova</code>	<code>stats</code>	Gets ANOVA table for a model
<code>manova</code>	<code>stats</code>	Performs multivariate ANOVA
<code>glm</code>	<code>stats</code>	Generalized linear model
<code>gam</code>	<code>gam</code>	Generalized additive model
<code>gam</code>	<code>mgcv</code>	Generalized additive model
<code>lm.ridge</code>	<code>MASS</code>	Linear models by ridge regression
<code>coef</code>	<code>stats</code>	Extracts coefficients from a model
<code>predict</code>	<code>stats</code>	Computes predictions for data input based on a model
<code>confint</code>	<code>stats</code>	Gets a confidence interval for model parameters
<code>residuals</code>	<code>stats</code>	Gets residuals of a model
<code>step</code>	<code>stats</code>	Selects variables for a model
<code>update</code>	<code>stats</code>	Updates and refits formula of a model

Table A.22 SixSigma package data sets

Function	Task
<code>ss.data.bolts</code>	Bolts example, Chap. 4
<code>ss.data.ca</code>	Winery example, Chap. 7
<code>ss.data.doe1</code> , <code>ss.data.doe2</code>	Pizza dough example, Chap. 11
<code>ss.data.pastries</code>	Pastries exercise, Chap. 5
<code>ss.data.pc</code> , <code>ss.data.pc.big</code> , <code>ss.data.pc.r</code>	Print cartridges example, Chap. 8
<code>ss.data.strings</code>	Guitar strings example, Chaps. 9 and 10
<code>ss.data.rr</code>	Data for a paper helicopter experiment

Table A.23 SixSigma package functions

Function	Task
<code>ss.ceDiag</code>	Cause-and-effect diagram
<code>ss.pMap</code>	Process map
<code>ss.lfa</code>	Loss function analysis
<code>ss.lfa</code>	Computes loss function value
<code>ss.rr</code>	Gage R&R study
<code>ss.study.ca</code>	Capability analysis study
<code>ss.ca.cp</code>	Capability index
<code>ss.ca.cpk</code>	Corrected capability index
<code>ss.ca.yield</code>	Computes yield of a process
<code>ss.ci</code>	Confidence interval for mean and normality test

Table A.24 Cited packages

Package	Use	Reference
base, stats, utils, grid, graphics, tools	Base packages	[84]
AcceptanceSampling	Acceptance sampling	[52]
AMORE	Neural network modeling	[62]
e1071	SVM, data mining, classification	[21]
kernlab	Data mining, classification, SVM	[48]
foreign	Data format conversion	[83]
gam	General additive model (inference)	[37]
ggplot2	Elegant graphics using grammar of graphics	[105]
Hmisc	Harrell miscellaneous (useful functions for many tasks by Frank E. Harrell)	[35]
knitr	Report generation	[111]
lattice	Lattice graphics	[91]
lpridge	Ridge regression	[95]
MASS	Support functions and data sets for Venables and Ripley's MASS	[102]
mgcv	Generalized additive and other models	[109]
monmlp	Neural network modeling	[14]
neuralnet	Neural network modeling	[27]
nnet	Neural network and multinomial log-linear models	[102]
nortest	Normality test	[33]
odfWeave	Reporting	[57]
plan	Gantt chart	[50]
plotrix	Gantt chart and other plots	[61]
pls	Partial least squares (PLS) regression	[67]
plsdoef	Partial least squares (PLS) regression	[56]
plspm	Partial least squares (PLS) regression	[90]
plsRbeta	Partial least squares (PLS) regression	[26]
plsRcox	Partial least squares (PLS) regression	[6]
plsRglm	Partial least squares (PLS) regression	[7]
qAnalyst	Control chart and capability analysis	[98]
qcc	Quality control chart	[93]
qualityTools	Methods associated with DMAIC cycle	[88]
RODBC	Data source connection	[87]
RWeka	SVM and other data mining tools	[39, 108]
SixSigma	Six Sigma tools and data sets for examples in this book	[15]
sp	Spatial data analysis	[78]
spc	Statistical process control computations	[54]
spcadjust	Calibration of control charts	
survival	Survival analysis (including ridge regression); reliability studies	[100]
XLConnect	Import/export, manipulate Microsoft Excel files	[29]
xtable	Elegant tables for reports and publications	[19]

References

1. Agresti, A., & Coull, B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *American Statistician*, 52, 119–126.
2. Allen, T. T. (2010). *Introduction to engineering statistics and lean Six Sigma—Statistical quality control and design of experiments and systems*. New York: Springer.
3. Arthur, J. (2006). *Lean six sigma demystified. Demystified series*. New York: McGraw-Hill.
4. Bentley, W., & Davis, P. (2009). *Lean Six Sigma secrets for the CIO: ITIL, COBIT, and beyond*. USA: CRC Press.
5. Berger, P., & Maurer, R. (2002). *Experimental design: With applications in management, engineering, and the sciences. Duxbury titles of related interest*. CA: Duxbury/Thomson Learning.
6. Bertrand, F., Maumy-Bertrand, M., & Meyer, N. (2011) *Partial least squares regression for Cox models and related techniques*. <http://CRAN.R-project.org/package=plsRcox>, r package version 0.2.0.
7. Bertrand, F., Meyer, N., & Maumy-Bertrand, M. (2011). *Partial least squares regression for generalized linear models*. <http://CRAN.R-project.org/package=plsRglm>, r package version 0.7.6.
8. Bivand, R., Pebesma, E., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R. Use R!*. New York: Springer.
9. Box G (1992). Teaching engineers experimental design with a paper helicopter. *Quality Engineering*, 4(3), 453–459.
10. Box, G., & Jones, S. (1992). Designing products that are robust to the environment. *Total Quality Management*, 3(3), 265–285.
11. Box, G., Hunter, J., & Hunter, W. (2005). *Statistics for experimenters: Design, innovation, and discovery. Wiley series in probability and statistics*. New York: Wiley.
12. Burdick, R., Borror, C., & Montgomery, D. (2005). *Design and analysis of gauge R&R studies: Making decisions with confidence intervals in random and mixed ANOVA models. ASA-SIAM series on statistics and applied probability*. Philadelphia: Society for Industrial Applied Mathematics.
13. Burton, T. T., & Sams, J. L. (2005). *Six Sigma for small and mid-sized organizations: Success through scaleable deployment*. Boca Rotan: J. Ross Publishing.
14. Cannon, A. J. (2011). *monmlp: Monotone multi-layer perceptron neural network*. <http://CRAN.R-project.org/package=monmlp>, r package version 1.1.
15. Cano, E. L., Moguerza, J. M., & Redchuk, A. (2011). *SixSigma: Six Sigma tools for quality improvement*. R package version 0.6.0.
16. Chambers, J. M. (2008). *Software for data analysis. Programming with R. Statistics and computing*. Berlin: Springer.

17. Cook, D., & Swayne, D. (2007). *Interactive and dynamic graphics for data analysis: With R and GGobi. Use R!*. New York: Springer.
18. Crawley, M. J. (2007). *The R Book*. UK: Wiley.
19. Dahl, D. B. (2009). *xtable: Export tables to LaTeX or HTML*. <http://CRAN.R-project.org/package=xtable>, r package version 1.5-6.
20. Dalgaard, P. (2008). *Introductory statistics with R. Statistics and computing*. New York: Springer.
21. Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., & Weingessel, A. (2011). *e1071: Misc functions of the department of statistics (e1071), TU Wien*. <http://CRAN.R-project.org/package=e1071>, r package version 1.6.
22. Eckes, G. (2003). *Six sigma for everyone*. New Jersey: Wiley.
23. El-Haik, B., & Roy, D. (2005). *Service design for six sigma: A roadmap for excellence*. New York: Wiley.
24. Faraway, J. (2006). *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models. Texts in statistical science*. Boca Raton: Chapman & Hall/CRC. <http://books.google.es/books?id=ODcRsWpGji4C>.
25. Faraway, J. J. (2002). *Practical regression and anova using r*. <http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>. Accessed 01.08.2011.
26. Bertrand, F., & Myriam Maumy-Bertrand N. M. (2011). *plsRbeta: Partial least squares regression for beta regression models*. <http://CRAN.R-project.org/package=plsRbeta>, r package version 0.1.1.
27. Fritsch, S., Guenther, F., & following earlier work by Marc Suling (2010). *neuralnet: Training of neural networks*. <http://CRAN.R-project.org/package=neuralnet>, r package version 1.31.
28. Gandy, A., & Kvaloy, J. T. (2011). *spcadjust: Functions for calibrating control charts*. <http://CRAN.R-project.org/package=spcadjust>, r package version 0.1-1.
29. GmbH MS (2011). *XLConnect: Excel Connector for R*. <http://CRAN.R-project.org/package=XLConnect>, r package version 0.1-7.
30. Graham, B. (2004). *Detail process charting: Speaking the language of process*. New Jersey: Wiley.
31. Grömping, U. (2011). *Cran task view: Design of experiments (doe) & analysis of experimental data*. <http://cran.r-project.org/web/views/ExperimentalDesign.html>. Retrieved 24.01.2012.
32. Grömping, U. (2012). *Project: (industrial) doe in r*. <http://prof.beuth-hochschule.de/groemping/software/design-of-experiments/project-industrial-doe-in-r/>. Retrieved 24.01.2012.
33. Gross, J. (2006). *nortest: Tests for normality*. R package version 1.0.
34. Gygi, C., DeCarlo, N., & Williams, B. (2005). *Six sigma for dummies*. Hoboken: Wiley.
35. Harrell, F. E. (2011). *Hmisc: Harrell miscellaneous*. <http://CRAN.R-project.org/package=Hmisc>, r package version 3.9-0.
36. Harry, M., & Schroeder, R. (1999). *Six sigma: The breakthrough management strategy revolutionizing the world's top corporations*. New York: Doubleday Business.
37. Hastie, T. (2011). *gam: Generalized additive models*. <http://CRAN.R-project.org/package=gam>, r package version 1.06.2.
38. Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The elements of statistical learning: Data mining, inference, and prediction. Springer series in statistics*. New York: Springer. <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>.
39. Hornik, K., Buchta, C., & Zeileis, A. (2009) Open-source machine learning: R meets weka. *Computational Statistics*, 24(2), 225–232.
40. Hsu, H. (2010). *Schaum's outline of probability, random variables, and random processes. Schaum's Outline Series* (2nd ed.). New York: McGraw-Hill.
41. Ishikawa, K. (1990). *Introduction to quality control*. Japan: 3A Corporation
42. ISO (1994). *ISO 5725-1: Accuracy (trueness and precision) of measurement methods and results—Part 1: General principles and definitions*. Geneva: International Organization for Standardization.

43. ISO (2009). *ISO/TS 16949: Quality management systems—Particular requirements for the application of ISO 9001:2008 for automotive production and relevant service part organizations*. Geneva: International Organization for Standardization.
44. ISO (2011). *Iso 13053-1:2011—Quantitative methods in process improvement—six sigma —part 1: DMAIC methodology*. Geneva: International Organization for Standardization.
45. Jakir, B. (2011). *Introduction to statistical thinking (with r, without calculus)*. <http://pluto.huji.ac.il/~msby/StatThink/index.html>. Accessed 01.08.2011.
46. John, A., Meran, R., Roenpage, O., Lunau, S., Staudter, C., & Schmitz, A. (2008). *Six Sigma+Lean Toolset*. Berlin: Springer.
47. Juran, J., & Defeo, J. (2010). *Juran's quality handbook: The complete guide to performance excellence*. New York: McGraw-Hill.
48. Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab—An S4 package for kernel methods in R. *Journal of Statistical Software*, 11(9), 1–20. <http://www.jstatsoft.org/v11/i09/>.
49. Keller, P., & Keller, P. (2011). *Six Sigma demystified. Demystified series*. New York: McGraw-Hill.
50. Kelley, D. (2009). *plan: Tools for project planning*. <http://CRAN.R-project.org/package=plan>, r package version 0.3-1.
51. Kerns, G. J. (2010). *Introduction to probability and statistics using r*. <http://ipsur.r-forge.r-project.org/book/index.html>. Accessed 01.08.2011.
52. Kiermeier, A. (2008). Visualizing and assessing acceptance sampling plans: The R package AcceptanceSampling. *Journal of Statistical Software*, 26(6), 1–20. <http://www.jstatsoft.org/v26/i06/>.
53. Knight, E., Russell, M., Sawalka, D., & Yendell, S. (2007). *Taguchi quality loss function and specification tolerance design*. https://controls.engin.umich.edu/wiki/index.php/Taguchi_quality_loss_function_and_specification_tolerance_design in Michigan chemical process dynamics and controls open textbook. Accessed 20.09.2011.
54. Knoth, S. (2011). *spc: Statistical process control*. <http://CRAN.R-project.org/package=spc>, r package version 0.4.1.
55. Koch, R. (2005). *Living the 80/20 way: Work less, worry less, succeed more, enjoy more*. London: Nicholas Brealey Publishing.
56. Kraemer, N., & Braun, M. L. (2011). *plsdof: Degrees of freedom and statistical inference for partial least squares regression*. <http://CRAN.R-project.org/package=plsdof>, r package version 0.2-2.
57. Kuhn, M. (2011). *odfWeave: Sweave processing of open document format (ODF) files*. <http://CRAN.R-project.org/package=odfWeave>, r package version 0.7.17.
58. Lalanne, C. (2006). *R companion to montgomery's design and analysis of experiments*. <http://www.aliquote.org/articles/tech/dae/>. Retrieved 19.01.2012.
59. Larsson, A. (2008). *Dia software*. <http://live.gnome.org/Dia>.
60. Leisch, F. (2002). Sweave: Dynamic generation of statistical reports using literate data analysis. In W. Härdle & B. Rönz (Eds.), *Compstat 2002—Proceedings In Computational Statistics* (pp. 575–580). Heidelberg: Physica Verlag. <http://www.stat.uni-muenchen.de/~leisch/Sweave>, ISBN 3-7908-1517-9.
61. Lemon, J. (2006). Plotrix: A package in the red light district of r. *R-News*, 6(4), 8–12.
62. Limas, M. C., Meré, J. B. O., Marcos, A. G., de Pisón Ascacibar, F. J. M., Espinoza, A. V. P., & Elías, F. A. (2010). *AMORE: A MORE flexible neural network package*. <http://CRAN.R-project.org/package=AMORE>, r package version 0.2-12.
63. Lopez-Fidalgo, J. (2009). A critical overview on optimal experimental designs. *Boletín de Estadística e Investigación Operativa*, 25(1), 14–21. http://www.seio.es/BEIO/files/BEIOv25n1_ES_J.Lopez-Fidalgo.pdf. Retrieved 19.01.2012.
64. Maindonald, J. H. (2008). *Using r for data analysis and graphics. Introduction, code and commentary*. <http://cran.r-project.org/doc/contrib/usingR.pdf>. Accessed 01.08.2011

65. Martin, J. (2007). *Operational excellence: Using lean six sigma to translate customer value through global supply chains. Series on resource management*. PA: Auerbach Publications.
66. Mee, R. (2009). *A comprehensive guide to factorial two-level experimentation*. New York: Springer.
67. Mevik, B. H., Wehrens, R., & Liland, K. H. (2011). *pls: Partial least squares and principal component regression*. <http://CRAN.R-project.org/package=pls>, r package version 2.3-0.
68. Moguerza, J. M., Muñoz, A. (2006). Support vector machines with applications. *Statistical Science*, 21(3), 322–336.
69. Montgomery, D. (2005). *Introduction to statistical quality control* (6th ed.). New York: Wiley.
70. Montgomery, D. (2008). *Design and analysis of experiments. Student solutions manual*. New York: Wiley.
71. Muir, A. (2005). *Lean Six sigma statistics: Calculating process efficiencies in transactional projects. Six sigma operational methods series*. New York: McGraw-Hill.
72. Murrell, P. (2009). Drawing Diagrams with R. *The R Journal*, 1(1), 15–21. http://journal.r-project.org/archive/2009-1/RJournal_2009-1_Murrell.pdf.
73. Murrell, P. (2009). *Introduction to data technologies. Series in computer science and data analysis*. Boca Raton: CRC Press. <http://www.stat.auckland.ac.nz/~paul/ItDT/>.
74. Murrell, P. (2011). *R graphics* (2nd ed.). Boca Raton: Chapman & Hall/CRC.
75. Myers, R., Montgomery, D., & Anderson-Cook, C. (2009). *Response surface methodology: Process and product optimization using designed experiments. Wiley series in probability and statistics*. New York: Wiley.
76. Pande, P., Neuman, R., & Cavanagh, R. (2000). *The Six Sigma way: How GE, Motorola, and other top companies are honing their performance*. New York: McGraw-Hill.
77. Pearn, W., & Kotz, S. (2006). *Encyclopedia and handbook of process capability indices: A comprehensive exposition of quality control measures. Series on quality, reliability & engineering statistics*. New York: World Scientific.
78. Pebesma, E., & Bivand, R. (2005). Classes and methods for spatial data in r. *R News*, 5(2), 9–13.
79. Poincaré, H. (1914). *Science and method*. London: Nelson. Originally published in French as *Science et Méthode* in 1908.
80. Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson. Originally published in German as *Logik der Forschung* in 1934.
81. Prieto, M. (2005). 6 sigma Qué es y cómo aplicarlo a la empresa española. Asociación Española para la Calidad.
82. Pyzdek, T., & Keller, P. (2009). *The Six Sigma handbook: A complete guide for green belts, black belts, and managers at all levels*. New York: McGraw-Hill.
83. R-core (2011). *foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, dBase,* <http://CRAN.R-project.org/package=foreign>, r package version 0.8-48.
84. 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/>, ISBN 3-900051-07-0.
85. Rasch, D., Pilz, J., & Simecek, P. (2010). *Optimal experimental design with R*. London: Taylor & Francis.
86. Recchia, D. R., Barbosa, E. P., & de Jesus Goncalves, E. (2010). *IQCC: Improved quality control charts*. <http://CRAN.R-project.org/package=IQCC>, r package version 0.5.
87. Ripley, B. (2012). *RODBC: ODBC Database Access*. R package version 1.3-4.
88. Roth, T. (2011). *qualityTools: Statistics in quality science*. R package version 1.50.
89. Roy, R. (2010). *A primer on the Taguchi Method* (2nd ed.). Michigan: Society of Manufacturing Engineers.
90. Sanchez, G., & Trinchera, L. (2012). *plsplm: Partial least squares data analysis methods*. <http://CRAN.R-project.org/package=plsplm>, r package version 0.2-2.
91. Sarkar, D. (2008). *Lattice: Multivariate data visualization with R*. New York: Springer. <http://lmdvr.r-forge.r-project.org>, ISBN 978-0-387-75968-5.

92. Schonberger, R. (2008). *Best practices in lean six sigma process improvement: A deeper look*. Safari Books Online. Chichester: Wiley.
93. Scrucca, L. (2004). qcc: An r package for quality control charting and statistical process control. *R News*, 4(1), 11–17, <http://CRAN.R-project.org/doc/Rnews/>.
94. Searle, S., Casella, G., & McCulloch, C. (2006). Variance components. Wiley series in probability and statistics. New York: Wiley.
95. Seifert, B. (2011). lpridge: Local polynomial (ridge) regression. <http://CRAN.R-project.org/package=lpridge>, r package version 1.0-6; Packaged for R by Martin Maechler.
96. Shaffer, L., Young, T., Guess, F., Bensmail, H., & León, R. (2008). Using r software for reliability data analysis. *International Journal of Reliability and Application*, 91(1), 53–70.
97. Snee, R. D. (2004). Six-sigma: The evolution of 100 years of business improvement methodology. *International Journal of Six Sigma and Competitive Advantage*, 1(1), 4–20.
98. Spano, A. (2011). *qAnalyst: Control charts, capability and distribution identification*. <http://CRAN.R-project.org/package=qAnalyst>, r package version 0.6.4.
99. Taguchi, G., Chowdhury, S., & Wu, Y. (2005). *Taguchi's quality engineering handbook*. USA: Wiley.
100. Therneau, T. (2011). *survival: Survival analysis, including penalised likelihood*. <http://CRAN.R-project.org/package=survival>, r package version 2.36-10. Original Splus->R port by Thomas Lumley.
101. Vardeman, S. B., & VanValkenburg, E. S. (1999). Two-way random-effects analyses and gauge r&r studies. *Technometrics*, 41(3), 202–211.
102. Venables, W., & Ripley, B. (2002). *Modern applied statistics with S. Statistics and computing*. New York: Springer.
103. Vikneswaran (2005). *An r companion to "experimental design"*. http://cran.r-project.org/doc/contrib/Vikneswaran-ED_companion.pdf. Retrieved 19.01.2012.
104. Wahlbrink, S. (2012). *Statet: Eclipse based ide for r*. <http://www.walware.de/goto/statet>. Retrieved 16.01.2012.
105. Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis. Use R!*. New York: Springer.
106. Wilkinson, L., & Wills, G. (2005). *The grammar of graphics. Statistics and computing*. New York: Springer.
107. Wilson, G. (2005). Six sigma and the product development cycle. Butterworth-Heinemann: Elsevier.
108. Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco: Morgan Kaufmann.
109. Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society*, 73(1), 3–36.
110. Workgroup, M. (2010). *MSA manual* (4th ed.). Automotive Industry Action Group (AIAG).
111. Xie, Y. (2012). *knitr: A general-purpose package for dynamic report generation in R*. <http://CRAN.R-project.org/package=knitr>, r package version 0.1.
112. Zuur, A., Ieno, E., & Meesters, E. (2009). *A beginner's guide to R. Use R!*. New York: Springer.

Solutions

Practice for Chap. 2

2.1

```
> install.packages("SixSigma", dependencies = TRUE)
```

Select a mirror close to your location when prompted.

2.2

```
> FailureTime <- c(0.29, 0.32, 1.21, 0.95, 0.14, 2, 0.81,
0.88)
> Temp <- c(63.89, 63.38, 65.05, 62.31, 68.04, 59.12, 62.80,
61.89)
> Factory <- c("A", "B", "C", "C", "B", "B", "A", "B")
> sol1a <- data.frame(FailureTime,
Temp, Factory = as.factor(Factory))
> summary(sol1a)
> write.csv(sol1a, file = "sol1a.csv")
```

2.3

```
> plot(sol1a$Temp, sol1a$FailureTime)
> boxplot(sol1a$FailureTime ~ sol1a$Factory)
> hist(sol1a$Temp)
```

2.4

```
> table(sol1a$Factory)
> sol3 <- sol1a$Temp[sol1a$FailureTime < 1]
```


Practice for Chap. 3

3.1

```

> grid.roundrect(width = .25,
  height = unit(1.8, "inches"),
  x = 0.25)
> grid.text("INPUTS\n\nData\nComputer\nOperator",
  x = 0.25,
  y = 0.66,
  just = "top")
> grid.roundrect(width = .25,
  height = unit(1.8, "inches"),
  x = 0.75)
> grid.text("OUTPUTS (Y)\n\nInvoice\n(pages,\nsize,\ncolor)",
  x = 0.75,
  y = 0.66,
  just = "top")
> grid.lines(x = c(0.375, 0.625),
  y = c(0.5, 0.5),
  arrow = arrow())

```

3.2

```

> inputs <- c("Clothes", "Machine", "Powder")
> outputs <- c("dryness", "cleanness", "time", "creases")
> steps <- c("PREPARE", "WASH", "HANG OUT")
> io <- list()
> io[[1]] <- list("Clothes", "Machine", "Powder")
> io[[2]] <- list("Clothes", "MachineState")
> io[[3]] <- list("dryness", "cleanness", "time")
> param <- list()
> param[[1]] <- list(c("type", "P"), c("amount", "N"))
> param[[2]] <- list(c("time", "Cr"), c("Powder.Brand", "Cr"),
  c("weight", "P"), c("MachineAge", "Cr"))
> param[[3]] <- list(c("WindSpeed", "N"), c("temperature", "N"),
  c("location", "C"))
> feat <- list()
> feat[[1]] <- list("cleanness")
> feat[[2]] <- list("cleanness", "time", "dryness")
> feat[[3]] <- list("cleanness", "time", "creases")
> ss.pMap(steps, inputs, outputs,
  io, param, feat,
  sub = "Laundry Process")

```

Practice for Chap. 4

4.1

$$L(Y) = 175(Y - 15)^2$$

4.2

Formula:

$$L(Y) = 0.125 \cdot (Y - 750)^2$$

Graphic of function:

```
> curve(0.125*(x-750)^2,
        from = 735, to = 765)
```

4.3 Loss function analysis:

```
> ss.lfa(ss.data.ca, "Volume", 10, 750, 1.25, 850)
```

The average cost is \$0.53 and the total loss is \$452.9.

Practice for Chap. 5

5.1

The variable `lab` (laboratory) is the appraiser factor (three levels), and the variable `batch` is the part factor, also with three levels. Two measures of each batch were taken in each laboratory.

5.2

The command to run the MSA is as follows:

```
> ss.rr(var = comp, part = batch, appr = lab,
        data = ss.data.pastries)
```

This is a good measurement system. The number of distinct categories is greater than 4 (6) and the %Study Var for R&R is lower than 30%. However, this value (20.27%) is not lower than 10%, and the measurement system may be improved. The plots confirm these results.

Practice for Chap. 6

6.1

First we create the data from the example:

```
> b.effect <- "Delay"
> b.groups <- c("Personnel", "Weather", "Suppliers", "
               Planning")
```

```

> b.causes <- vector(mode="list", length=length(b.groups))
> b.causes[1] <- list(c("Training", "Inadequate"))
> b.causes[2] <- list(c("Rain", "Temperature", "Wind"))
> b.causes[3] <- list(c("Materials", "Delays", "Rework"))
> b.causes[4] <- list(c("Customer", "Permissions", "Errors"))
>

```

We can plot a cause-and-effect diagram:

```

> ss.ceDiag(b.effect, b.groups, b.causes, sub="Construction
  Example")
> b.data <- data.frame(cause = factor(unlist(b.causes)),
  count = c(5, 1, 3, 1, 2, 18, 20, 4, 15, 2, 4),
  cost = c(50, 150, 50, 10, 20, 180, 200, 10, 5, 20, 150))

```

Next we plot the Pareto chart:

```

> require(qcc)
> b.vector <- b.data$cost
> names(b.vector) <- b.data$cause
> pareto.chart(b.vector, cumperc = c(80))

```

6.2 We need the two charts in order to choose the adequate causes to focus on.

If we only use the Pareto chart for the count of errors, then we will select “Customer” as one of the “vital few.” However, in the cost chart, it is apparent that delays due to customer needs do not generate a high cost. It is due to the fact that a new specification usually involves a revision of the contract and, therefore, an extended deadline without extra costs.

Using the cost chart, the factors to focus on are the delay in the receipt of materials, quality of materials, inadequateness of the personnel, and planning errors.

Practice for Chap. 7

7.1

With $\sigma = 2$:

$$Z = \min \left\{ \frac{(14 - 10)}{2}, \frac{(10 - 4)}{2} \right\} = \min 2, 3 = 2,$$

$$Z_{LT} = Z_{ST} - 1.5 = 2 \times 1.5 = 0.5.$$

With $\sigma = 1$:

$$Z = \min \left\{ \frac{(14 - 10)}{1}, \frac{(10 - 4)}{1} \right\} = \min 4, 6 = 4,$$

$$Z_{LT} = Z_{ST} - 1.5 = 4 \times 1.5 = 2.5,$$

7.2 Now we have less than 308,000 DPMO (previously we had more than 690,000 DPMO).

Practice for Chap. 8

8.1

```
> newData <- aggregate(pc.volume ~ pc.filler,
  sum, data = ss.data.pc)
> barplot(newData$pc.volume,
  names.arg = newData$pc.filler,
  main = "Total volume by filler")
```

8.2

```
> newData <- ss.data.pc.big[ss.data.pc.big$batch == 1,5]
> plot(newData, type = "b")
```

There seems to be a descendant trend over time.

Practice for Chap. 9

9.1 To save the subset:

```
> data.stats.probl <- ss.data.strings[ss.data.strings$type ==
  "E1",]
```

```
> summary(data.stats.probl$res)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.00	5.75	6.50	6.20	8.00	9.00

```
> sd(data.stats.probl$res)
```

```
[1] 2.067289
```

```
> IQR(data.stats.probl$res)
```

```
[1] 2.25
```

The mean and median are not equal, so the distribution might not be symmetric. We can generate a boxplot to see the data from the summary.

```
> boxplot(data.stats.probl$res)
```

9.2

```
> pbinom(4, 100, 0.01)
```

```
[1] 0.9965677
```

Practice for Chap. 10

10.1

Using the `shapiro.test` function:

```
> shapiro.test(ss.data.strings$res)
```

Using the `ss.ci` function from the `SixSigma` package:

```
> ss.ci(res, data = ss.data.strings)
```

These data are nonnormal. We need other methods for inference.

10.2

```
> prob2.model <- lm(res ~ len, data = ss.data.strings)
> summary(prob2.model)
```

The linear model is not good for these data. The p -value for the goodness-of-fit hypothesis test is very high (0.7149). Furthermore, the R-squared statistic is very low.

10.3 Solution using `lm` function:

```
> prob3.model1 <- lm(len ~ type, data = ss.data.strings)
> summary(prob3.model1)
```

Solution using `aov` function:

```
> prob3.model2 <- aov(len ~ type, data = ss.data.strings)
> summary(prob3.model2)
```

The length of the strings is independent of the type of string (p -value > 0.05).

Plotting the effects:

```
> qplot(type, len, data = ss.data.strings) +
  stat_summary(fun.y = mean, geom = "line",
    aes(group = 1), col = "orangered") +
  stat_summary(fun.y = mean, geom = "point",
    shape = 17, size = 3, col = "red") +
  opts(title = "Effects of factor Type of string on Length"
    )
```

Practice for Chap. 11

11.1

```
> myDesign <- expand.grid(factor1 = gl(2, 1, labels = c("-",
  "+")),
  factor2 = gl(2, 1, labels = c("-", "+")),
  factor3 = gl(2, 1, labels = c("-", "+")),
  factor4 = gl(2, 1, labels = c("-", "+")),
  response = NA)
```

```
> myDesign$order <- sample(1:16, 16)
> myDesign[order(myDesign$order),]
```

11.2

The formula in the following code allows one to check up to three-way interactions.

```
> model.probl <- lm(score ~ (.-repl)^3 , data = ss.data.doe2
)
> summary(model.probl)
```

There are some important interactions; therefore, we keep them in the model jointly with the significant main effects.

```
> selectionvar <- step(model.probl, method="backwards")
> summary(selectionvar)
> coef(selectionvar)
```

Practice for Chap. 12

12.1

```
> pb.groups.s <- with(ss.data.pb2,
  qcc.groups(pb.humidity, pb.group))
> pb.s <- qcc(pb.groups.s, type = "S")
> summary(pb.s)
> plot(pb.s)
```

12.2

As the number of errors is registered by surface unit, the sample size for each group is 1, and we will have straight control limit lines. The u chart is plotted with the following command:

```
> with(ss.data.pb4,
  plot(qcc(defects, sizes = rep(1, 80), type = "u")))
```

R Packages and Functions Used in the Book

A

AMORE, 186

B

base

~, 40, 182

<- (assign), 25

[], 28

[[]], 29

\$, 29

apply, 38

as.character, 32

as.data.frame, 32

c, 25, 81

choose, 44

class, 29, 32

colnames, 26

cos, 36

cut, 143

data.frame, 27, 81, 95, 178, 204

diff, 148, 170

dim, 28

exp, 36

expand.grid, 201, 213

factor, 26, 81

factorial, 44

for, 245

function, 246

gamma, 169

getwd, 22

gl, 201

if, 245

is.element, 37

is.na, 145

lapply, 37

length, 28

list, 26, 95

load, 33

ls, 32

matrix, 26, 34

mean, 37, 146, 169, 173

names, 97, 183, 189

nrow, 151, 170–174

objects, 32, 34

order, 97

q, 23

range, 170

rbind, 204

rev, 97

rm, 33

round, 36

rownames, 26

sample, 43, 203

sapply, 37, 246

save, 33

save.image, 33

scan, 25

set.seed, 245

setwd, 22

sort, 97

source, 23

sqrt, 36, 171, 173

sum, 151, 170, 171

summary, 38, 148, 183, 189, 206

table, 37, 143, 146, 151

union, 37

vector, 95

E

e1071, 186

F

foreign, 31
 read.mtp, 31
 read.spss, 31

G

gam, 186
 ggplot2, 39, 100
 facet_grid, 209
 geom_line, 209
 geom_point, 209
 geom_point, 181
 ggplot, 209
 opts, 181, 191, 209
 qplot, 181, 191
 scale_x_continuous, 209
 stat_abline, 209
 stat_summary, 191
 graphics, 38, 225
 abline, 40, 41, 68, 128, 130, 181, 208
 axis, 128
 barplot, 97, 120
 box, 210
 boxplot, 131, 132, 188
 curve, 39, 68, 123
 grid, 123, 125
 hist, 41, 121, 123, 155
 lines, 41, 123, 130, 133, 155
 par, 41, 184, 210
 pie, 42
 plot, 39–41, 117, 125, 128, 181, 184, 208, 210
 plot.design, 191
 points, 133
 stripchart, 130, 133
 text, 39, 68, 97, 123, 132, 133
 grid, 39
 grid.lines, 54
 grid.roundrect, 54
 grid.text, 54

H

Hmisc, 171
 binconf, 171, 172

I

IQCC, 225

K

kernlab, 186

L

lattice, 39, 100, 225
 xyplot, 135
 lpridge, 186
 lpridge, 186

M

MASS, 186
 lm.ridge, 186
 mgcv, 186
 monmlp, 186

N

neuralnet, 186
 nnet, 186
 nortest
 lillie.test, 211

P

plan, 244
 plotrix, 244
 gantt.chart, 244
 pls, 186
 plsdf, 186
 plsgenomics, 186
 plsppm, 186
 plsRbeta, 186
 plsRcox, 186
 plsRglm, 186

Q

qAnalyst, 225
 plot.spc, 231
 print.spc, 229
 spc, 229
 summary.spc, 229
 qcc, 96, 225
 pareto.chart, 97
 plot.qcc, 228
 qcc, 227, 232, 235
 qcc.groups, 227, 232
 qualityTools
 mvPlot, 135
 paretoChart, 97

R

RODBC, 31
 RWeka, 186

S

SixSigma

ss.ca.cp, 109
 ss.ca.cpk, 109
 ss.ca.yield, 106
 ss.ca.z, 107
 ss.cc.getc4, 169, 170
 ss.cc.getd2, 169
 ss.ceDiag, 95
 ss.ci, 174
 ss.data.bolts, 69
 ss.data.doe1, 204
 ss.getd2, 170
 ss.heli, 12
 ss.lf, 71
 ss.lfa, 70
 ss.pMap, 58
 ss.rr, 85
 ss.study, 110

sp, 39

stats

aggregate, 178, 205
 anova, 82, 184, 189
 aov, 189
 bartlett.test, 176
 binom.test, 171, 176
 chisq.test, 176
 coef, 207
 confint, 184, 189, 208
 cor, 180
 dbinom, 153
 dnorm, 44
 glm, 186
 IQR, 148
 lm, 82, 181, 182, 189, 206
 manova, 192
 median, 146
 na.omit, 180
 pairwise.t.test, 190
 pbeta, 162
 pbinom, 153
 pchisq, 164
 pexp, 160
 pf, 165
 pgamma, 161
 pgeom, 158
 phyper, 157
 plnorm, 160
 plot.lm, 184
 pnbinom, 159
 pnorm, 44, 155
 poisson.test, 176
 power.prop.test, 179
 power.t.test, 179
 ppois, 159

predict, 184, 208
 prop.test, 171, 176, 178
 pt, 163
 ptukey, 169
 punif, 161
 pweibull, 163
 qchisq, 164, 174
 qf, 165
 qnorm, 44, 156, 171, 173
 qqline, 155
 qqnorm, 155
 qt, 163, 172
 quantile, 148
 residuals, 184
 rnorm, 44
 runif, 161
 sd, 148, 170, 173
 shapiro.test, 176, 178, 184, 211
 step, 185
 t.test, 174, 176, 177
 update.formula, 185
 var, 169, 174
 var.test, 176, 177
 survival, 186
 ridge, 186

U

utils

?, 24
 ??, 24
 apropos, 24
 browseVignettes, 24
 data, 30
 demo, 24
 example, 24
 head, 27
 help.start, 24
 install.packages, 23
 read.csv, 30
 read.table, 31
 str, 27, 81, 135
 Sweave, 247, 248
 texi2svi, 248
 vignette, 12, 24
 write.csv, 31

X

XLConnect

loadWorkbook, 31
 readWorksheet, 31
 xtable, 247
 xtable, 247

Subject Index

Symbols

2^k factorial design, 203
NA, 145
6 M's, 53
80/20 distribution, 92
80/20 rule, 92

A

acceptance sampling, 70, 237
accuracy, 80
affinity diagram, 94
alternative hypothesis, 176
analysis of covariance, 191
analysis of variance, 82, 187, 203
ANCOVA, 191
ANOVA, 82, 187, 203
 model fitting, 189
 multi-way, 187
 one-way, 187
 two-way, 187
area under the curve, 152
argument, 34
assignable causes, 149
attributes control charts, 234
atypical data, 145
average, 146
average loss function, 69
average range, 89, 149
axis, 116
 x, 116
 y, 116
a priori probability, 150

B

bar chart, 96

base

 range, 148
Bayesian inference, 141
Bayesian probability, 150
Bernoulli experiment, 152, 168
binomial distribution, 152, 157, 171, 234
blind trial, 204
blocking, 198
box plot, 41, 130, 188
box-whisker chart, 130
brainstorming, 7, 52, 94
bubbles, 136

C

c chart, 237
capability
 adjusted index, 108
 indices, 108
 of a process, 108
capability analysis, 237
cause-and-effect, 200
cause-and-effect diagram, 93
cause-and-effect relationship, 125, 193
center control line, 221
center line, 225
central limit theorem, 154, 171, 172
central tendency, 146
characterizing experiments, 202
chart, 115
 limits, 116
chart patterns, 223
class, 32, 143
classical probability, 150
clipboard, 31
coefficients, 207
columns, 38, 142

combinatorial, 44
 Command, 19
 common causes, 149, 222
 conditions, 244
 confidence coefficient, 170
 confidence interval, 109, 167, 176
 mean, 172
 proportion, 171
 variance, 173
 confidence level, 170, 179
 confusion of effects, 198
 continuous variable, 142
 control, 219
 control chart, 88, 221
 control limits, 88, 89, 198, 221, 226, 227, 231,
 232, 234, 236
 Cook's distance, 185
 coordinates, 54
 correlation, 124, 180, 198
 negative, 180
 positive, 180
 correlation coefficient, 180
 cost of poor quality, 63, 64
 cost of quality, 63
 covariance, 180
 CRAN, 16
 CTQ, 6, 53, 63, 64, 179, 212, 219
 cumulative distribution function, 150
 cycle, 126

D

data frame, 27
 data science, 141
 data scientist, 141
 data types, 25, 32
 dataframe, 142
 default value, 34
 Defect, 105
 per Unit, 105
 defective, 65
 Defects
 per Opportunity, 105
 defects, 95
 per million opportunities, 105
 degrees of freedom, 172
 density, 41, 44
 density probability function, 152
 dependent variable, 124, 182, 198
 descriptive analysis, 115
 descriptive statistics, 141
 design matrix, 201
 design of experiments, 99, 191, 197
 deterministic event, 152

device, 21
 dfss, 242
 DFSS cycles, 242
 dimension, 28, 116
 discrete variable, 142
 DMAIC, 5
 analyze, 6
 control, 7
 define, 6
 improve, 7
 measure, 6
 DoE, 197
 DPMO, 107

E

ECLIPSE, 247
 effect, 100
 effective limits, 108
 efficient process, 243
 empirical probability, 150
 error type I, 179
 error type II, 179
 estimator, 147, 168
 EWMA, 237
 Excel, 30
 expectation, 152
 experiment, 198
 experiment replication, 201, 202
 experimental conditions, 201, 202
 experimentation strategy, 200
 explanatory variable, 182
 exploratory data analysis, 141
 export data, 31

F

factor, 26, 41, 142, 187, 200, 203
 factor levels, 187, 200
 factorial design, 203
 features, 55
 fishbone diagram, 94
 flow chart, 52
 FMECA, 241
 formula, 40
 fraction of nonconforming, 171
 freely available software, 16
 frequency, 143
 frequency table, 143
 function, 21
 arguments, 246
 body, 246
 function value, 36
 functions, 34
 future state process map, 7

G

Gage R&R, 79
 gantt chart, 243
 generalized additive model, 186
 generalized linear model, 186
 geometric mean, 146
 ggplot2, 225
 group chart, 133

H

harmonic mean, 146
 hat, 168
 histogram, 41, 110, 120
 homoscedasticity, 185
 hypergeometric distribution, 157
 hypothesis testing, 174, 175, 203

I

I/MR chart, 226
 IDE, 247
 import data, 31
 improve phase, 70
 Improvement, 197
 inconsistent data, 198
 independent variable, 124, 182, 198
 indices, 28
 individual control chart, 226
 inference, 144, 152, 167
 inferential statistics, 141
 inputs, 53
 inspection, 219
 interaction, 83, 187, 191, 202
 interaction effect, 203
 interactions plot, 209
 interquartile range, 148
 interval estimation, 170
 IQR, 148
 Ishikawa diagram, 94
 Ishikawa, Kaoru, 94
 ISO, 64, 80

J

Juran, Joseph M., 92

K

key characteristic, 106

L

larger-the-better, 72

latex, 247
 law of large numbers, 150
 LCL, 221, 225
 lean manufacturing, 243
 length, 28
 library, 23
 likelihood, 151
 linear relation, 180
 list, 26
 literate programming, 247
 location chart, 134
 long term variability, 106
 long term variation, 129
 long-term variability, 149
 loops, 244
 loss function, 63, 65, 102
 lower control limit, 221
 LSL, 102

M

magnificent seven, 93
 main effect, 202, 203
 main effects plot, 209
 MANOVA, 192
 mathematical expression, 39
 mathematical function, 39
 matrix, 26, 142
 maximum, 147
 mean, 106, 146, 152
 mean of squares, 82
 mean squared deviation, 69
 mean-range chart, 231
 mean-sd chart, 234
 measure phase, 69
 measurement system analysis, 79
 median, 146
 median absolute deviation,
 147
 minimum, 147
 Minitab, 31
 missing data, 145
 mistake-proofing, 219
 mistake-proofing design, 219
 mixed models, 186
 mode, 146
 model, 204, 207
 model diagnostics, 210
 moving range, 226
 moving range chart, 226
 multiple regression, 185
 multivariate analysis of variance,
 192
 multivariate chart, 134

N

natural limits, 108
 neural networks, 186
 noise factor, 198, 212
 nominal-the-better, 73
 normal distribution, 154, 157, 171, 221
 normality, 211
 normality test, 173, 178, 184, 211
 not available, 145
 np chart, 237
 null hypothesis, 176, 179
 number of distinct categories, 84

O

object, 21, 32
 observation, 142
 observational study, 197
 OC curve, 237
 one-at-a-time, 202
 one-sided test, 176
 operational structure, 212
 optimization experiments, 202
 outlier, 130, 145, 146
 outputs, 53

P

p chart, 234
 p-value, 176, 206, 211
 package, 23
 parallel computing, 246
 parameter, 167, 175, 180
 parameters, 55, 93
 classification, 56
 Pareto chart, 93, 120
 Pareto Principle, 91
 Pareto, Vilfredo, 91
 perspective, 136
 phase I (control charts), 221
 phase II (control charts), 221
 pie plot, 42
 pilot program, 7
 planning, 243
 plot, 40
 PLS, 186
 point estimation, 167
 Poisson distribution, 234
 poka-yoke, 219, 220
 population, 144, 152, 167
 population mean, 168
 population variance, 168
 posterior probability, 150
 precision, 80

probability, 150
 probability density function, 150
 probability distribution, 44, 142, 152, 167
 probability function, 150
 probability theory, 141
 problem solving, 92
 process, 4
 breakdown, 54
 steps, 54
 process improvement, 4, 147, 198, 212
 process map, 6, 51
 final, 52
 ideal, 52
 top level, 52
 process mapping, 51
 process owner, 212
 program block, 246
 program statement, 246
 progressive experimentation, 202
 project charter, 6, 52
 prompt symbol, 19

Q

QC, 219
 qualitative variable, 142, 187
 quality, 64
 quality control, 219
 quantile–quantile chart, 110
 quantitative variable, 142
 quartile, 148

R

R
 Console, 19
 Editor, 22
 formulas, 181
 graphics, 21, 38
 GUI, 19
 Help, 24
 Installation, 16
 Licence, 16
 package installation, 23
 session, 32
 workspace, 32
 workspace image, 33
 R functions, 246
 R language, 244
 random sample, 43, 144
 random variable, 142, 152
 randomization, 202, 223
 randomness, 222
 range, 147, 169

- rational groups, 223
- raw data, 142, 143
- redesign, 242
- reference level, 189
- refutability, 176
- regression, 180
 - model fitting, 181
 - model validation, 184
- regression coefficients, 183
- regression line, 40
- rejection region, 176
- relative frequency, 143, 150
- relative range, 150, 169, 226
- reliability, 237
- reliability analysis, 237
- repeatability, 79, 80
- replication, 201
- reproducibility, 79, 80
- reproducible research, 58, 247
- requirements, 64, 101
- residuals, 180, 183
- response, 182, 198, 202
- restore data, 33
- ridge regression, 186
- risk, 200, 241
- risk matrix, 242
- RNG, 144
- robust design, 212
- robust process, 197
- roles
 - BB, 9
 - champion, 9
 - green belt, 9
 - MBB, 9
 - process owner, 9
- root cause, 6, 51, 92
- row, 142
- rows, 38
- RUMBA, 102
- run chart, 126, 128

- S**
- sample, 144, 152, 167
 - order, 203
- sample mean, 168
- sample proportion, 168
- sample size, 144, 150, 223
- sample space, 150
- sample standard deviation, 169
- sample variance, 147, 168
- sampling, 223
- sampling distribution, 168
- save data, 33
- scale, 116
- scatterplot, 40, 124, 180
- scheduling, 243
- scientific method, 4, 198
- screening experiments, 202
- Script, 21
- seasonality pattern, 223
- segmented implementation, 7
- select project, 99
- sequential experimentation, 200
- Shapiro–Wilk, 178
- Shewart’s constants, 89
- shift, 126
- shift pattern, 223
- short term variability, 106
- short-term variability, 149
- short-term variation, 129
- Sigma
 - Score, 106
- significance level, 176, 179
- significant effect, 202
- simulation, 7
- SIPOC, 52
- six Ms, 94
- Six Sigma
 - methodology, 3, 4
 - project, 5
 - roles, 9
- SixSigma package
 - Installation, 46
- smaller-the-better, 71
- solution criteria list, 7
- SPC, 219, 237
- special causes, 222
- special characters, 23
 - ;, 23
 - #, 23
- Special Values
 - Inf, 25
 - NA, 25
 - NaN, 25
 - NULL, 25
- specification limits, 64, 102, 110, 221
- specifications, 101
- split-plot design, 214
- spreadsheet, 30
- SPSS, 31
- stakeholder, 52
- standard deviation, 106, 147
- standard normal distribution, 156
- standardization, 156
- statistic, 168, 173, 176
- statistical inference, 167
- statistical power, 179

statistician, 141
 stochastic event, 152
 structure, 27
 Student t, 172
 study var, 86
 subgroup, 223
 subscript, 28, 30
 subset, 28
 sum of squares, 82
 summarize, 37
 symmetric distribution, 147

T

t-test, 187
 Taguchi, 63, 66, 215
 Taguchi loss function, 63
 target, 64, 66, 102, 110
 Task View, 24
 tier chart, 129
 time series, 126
 tolerance, 64, 66
 tolerance limits, 70
 trend, 126, 145
 trend pattern, 223
 trial, 200, 202
 trial-and-error, 200
 trimmed mean, 146
 trivial many, 92
 trueness, 80
 two-dimensional chart, 221
 two-dimensional charts, 116
 two-sided test, 176

U

u chart, 236
 UCL, 221, 225
 unbiasedness, 168
 uncertainty, 170

upper control limit, 221
 useful many, 92
 USL, 102

V

value stream, 243
 variability, 147
 variable, 27, 142
 variables control charts, 226
 variables selection, 185
 variance, 147, 152, 168
 vector, 25, 142
 vignette, 24, 248
 vital few, 92
 Voice of the Customer, 101
 Voice of the Process, 102
 VSM, 52

W

waste, 243
 weighted mean, 146
 Working Directory, 22

X

Xs of the process, 220

Y

Yield, 103
 yield
 first-time yield, 103
 rolled throughput yield, 104
 Ys, 220

Z

zero defects, 220