MA206 Reference Sheet

Getting Help in R

?mean
Get help for a particular function.

help(mean)
Search the help files for a word or phrase.

Reading File

1. Save file as .csv.
2. Input command:
   a. data=read.csv("filename.csv", header=T)
      (or F if no header)
   b. data=read.csv(file.choose(),header=T)
      (uses a dialog box to select the file)
3. The data is now read in as a data frame.
4. You can index into sections of the data frame using the $ operator (e.g., data$column1).

Vectors

c(2, 4, 6)	2 4 6	Join elements into a vector
2:6	2 3 4 5 6	An integer sequence
seq(2,3,by=0.5)	2.0 2.5 3.0	A complex sequence
rep(1:2, times=3)	1 2 1 2 1 2	Repeat a vector
rep(1:2, each=3)	1 1 1 2 2 2	Repeat elements of a vector

Selecting Vector Elements

By Position

x[4]	The fourth element
x[-4]	All but the fourth
x[2:4]	2nd through 4th elements
x[c(1, 5)]	1st and 5th element

By Value

x[x == 10]	Elements equal to 10
x[x < 0]	All elements less than zero

Working Directory

getwd()
Find the current working directory.

setwd('C://file/path') Change the current working directory.

Tools-GlobalOptions-*C:\\yourRdirectory*-**Apply**
Make your folder the working directory every time you start R-Studio.

Statistics and Regression

Function Name	Arguments
t.test()	x, y(if needed), alternative, mu, paired, conf.level
lm()	y~x (simple) y~x1+x2+... (multiple)
summary()	object or model (This command summarizes the model or data set.)

Plotting

plot(x)		Values of x in order.
plot(x, y)		Values of y against x.
hist(x)		Histogram of x

Math Functions

log(x)	Natural Log
exp(x)	Exponential
max(x)	Largest Element
min(x)	Smallest Element
round(x,n)	Round to n Decimal Places
sig.fig(x, n)	Round to n Sig Figs
cor(x, y)	Correlation

sum(x)	Sum
mean(x)	Mean
median(x)	Median
quantile(x)	Percentile or Quantile
var(x)	Variance
sd(x)	Standard Deviation
length(x)	# of Elements in Vector
rank(x)	Rank of Elements

Conditions

a == b	Are equal	a > b	Greater than	a >= b	Greater than or equal to	is.na(a)	Is missing	a & b	a and b
a != b	Not equal	a < b	Less than	a <= b	Less than or equal to	is.null(a)	Is null	a|b	a or b

The Nine Characteristics of Discrete Distributions

Characteristic 1: The pmf must be greater than or equal to zero for all x. $p(x) \geq 0 \quad \forall\, x$

Characteristic 2: The sum of probabilities of x, $p(x)$, over all possible values of x must be one. $\sum_x p(x) = 1$

Characteristic 3: For a discrete random variable X, the probability that X is equal to a specific value, c, is: $P(X = c) = p(c)$

Characteristic 4: Given the pmf $p\,(x)$ for random variable X, we define the cumulative distribution function (CDF) $F(x)$ as follows: $F(x) = P(X \leq x) = \sum_{y:y\leq x} p(y)$

Characteristic 5: Using the pmf: $P(a \leq X \leq b) = \sum_{x:a\leq x\leq b} p(x)$

Characteristic 6: Using the CDF: $P(a \leq X \leq b) = F(b) - F(a^-)$

Characteristic 7: The expected value of X: $E(X) = \mu_X = \sum x \cdot p(x)$

Characteristic 8: The variance of X: $V(X) = E(X^2) - (E(X)^2)$ or $V(X) = \sigma_X^2 = E[(X - \mu_X)^2] = \sum (x - \mu_X)^2 p(x)$

Characteristic 9: The $(100p)^{th}$ percentile of the discrete random variable X is the minimum value of x such that: $F(x) \geq p$

The Nine Characteristics of Continuous Distributions

Characteristic 1: The pdf must be greater than or equal to zero for all x. $f(x) \geq 0 \quad \forall\, x$

Characteristic 2: The total area under $f(x)$ must be equal one. $\int_{-\infty}^{\infty} f(x)dx = 1$

Characteristic 3: For a continuous random variable X, the probability that X is equal to a specific value, c, is: $P(X = c) = 0$

Characteristic 4: Given the pdf $f(x)$ for random variable X, we define the cumulative distribution function (CDF) $F(x)$ as follows: $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$

Characteristic 5: Using the pdf: $P(a \leq X \leq b) = \int_a^b f(x)dx$

Characteristic 6: Using the CDF: $P(a \leq X \leq b) = F(b) - F(a)$

Characteristic 7: The expected value of X: $E(X) = \mu_X = \int_{-\infty}^{\infty} x \cdot f(x)dx$

Characteristic 8: The variance of X: $V(X) = E(X^2) - (E(X)^2)$ or $V(X) = \sigma_X^2 = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f(x)dx$

Characteristic 9: The $(100p)^{th}$ percentile of the continuous random variable X where F^{-1} denotes the inverse CDF of X is: $x^* = F^{-1}(p)$ or $F(x^*) = p$

Named Probability Distributions

	Binomial	Uniform	Normal	t
PDF	$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x \in \{0,1,2,\dots,n\} \\ 0 & otherwise \end{cases}$	$f(x) = \begin{cases} \frac{1}{B-A} & A \le X \le B \\ 0 & otherwise \end{cases}$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-((x-\mu)^2)/2\sigma^2} \quad \forall\, x$	$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad \forall\, x > 2$
Parameters	$n \text{ and } p$	$A \text{ and } B$	$\mu \text{ and } \sigma$	ν $(degrees\ of\ freedom)$
$E[X]$	np	$\frac{A+B}{2}$	μ	0
$V[X]$	$np(1-p)$	$\frac{1}{12}(B-A)^2$	σ^2	$\frac{\nu}{\nu-2}$
R Help				
PMF	`dbinom(x,size,prob)`			
CDF	`pbinom(q,size,prob)`	`punif(q,min,max)`	`pnorm(q,mean,sd)`	`pt(q,df)`
Percentile	`qbinom(p,size,prob)`	`qunif(p,min,max)`	`qnorm(p,mean,sd)`	`qt(p,df)`
Random Variate	`rbinom(n,size,prob)`	`runif(n,min,max)`	`rnorm(n,mean,sd)`	`rt(n,df)`