

# MA256 Reference Sheet

## R Guide and Distribution Characteristics

### Getting Help

**?mean**  
Get help for a particular function.

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

### 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 everytime you start R-Studio.

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	
<b>x[4]</b>	The fourth element
<b>x[-4]</b>	All but the fourth
<b>x[2:4]</b>	Elements two to four
<b>x[c(1, 5)]</b>	Elements one and five
By Value	
<b>x[x == 10]</b>	Elements equal to 10
<b>x[x &lt; 0]</b>	All elements less than zero

### Function

`function_name <- function(var){ Do something }`

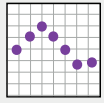
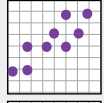
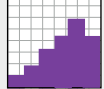
### Reading File

- Save file as .csv.
- Input command:
  - `data=read.csv("filename.csv", header=T)` (or F if no header)
  - `data=read.csv(file.choose(),header=T)` (if you want to select the file)
- The data is now read in as a data frame.
- You can index into sections of the data frame using the \$ operator (e.g., `data$column1`).

### 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 x against y.
hist(x)		Histogram of x

### Math Functions

log(x)	Natural log	sum(x)	Sum
exp(x)	Exponential.	mean(x)	Mean
max(x)	Largest element	median(x)	Median
min(x)	Smallest element	quantile(x)	Percentile or quantile
round(x,n)	Round to n decimal places	var(x)	The variance
sig.fig(x, n)	Round to n sig figs	sd(x)	The standard deviation
cor(x, y)	Correlation	length(x)	# of elements in vector
		rank(x)	Rank of elements

### Probability Functions in R

Distribution Names	Random Variates	Density Function	Cumulative Distribution	Quantile	Arguments
Binomial	rbinom	dbinom	pbinom	qbinom	x/q, size, prob
Poisson	rpois	dpois	ppois	qpois	x/q, lambda (same as Devore's μ)
Uniform	runif	dunif	punif	qunif	x/q, min, max
Normal	rnorm	dnorm	pnorm	qnorm	x/q, mean, sd
Exponential	rexp	dexp	pexp	qexp	x/q, rate (same as Devore's λ)
Gamma	rgamma	dgamma	pgamma	qgamma	x/q, shape, scale (same as Devore's α and β)
t	rt	dt	pt	qt	x/q, df
Chi-Square	rchisq	dchisq	pchisq	qchisq	x/q, df

### PMF Characteristics

- 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: Find the probability that X is between a and b using the PMF:  $P(a \leq X \leq b) = \sum_{x:a \leq x \leq b} p(x)$
- Characteristic 6: Find the probability that X is between a and b using the CDF:  $P(a \leq X \leq b) = F(b) - F(a-)$
- Characteristic 7: The expected value of X:  $E(X) = \mu_X = \sum_x x \cdot p(x)$
- Characteristic 8: The variance of X:  $V(X) = \sigma_X^2 = \sum_x (x - \mu_X)^2 \cdot p(x)$  or  $E[X^2] - (E[X])^2$
- Characteristic 9: The (100p)th percentile of the discrete random variable X is the min value of X such that  $F(x) \geq p$ .

### PDF Characteristics

- 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 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: Find the probability that X is between a and b using the pdf:  $P(a \leq X \leq b) = \int_a^b f(x)dx$
- Characteristic 6: Find the probability that X is between a and b 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:  $V(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f(x)dx$
- Characteristic 9: The (100p)th percentile of the continuous random variable X:  $x^* = F^{-1}(p)$

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