

STAT2009 Introduction to Programming Language for Statistics
Midterm Examination
25th October, 2017 14:30 – 16:15

Answer ALL questions.

1. (a) Create each of the following R objects in a single line command. (5 marks)

```
> A
      [,1] [,2] [,3] [,4]
[1,]    1    2    3    4
[2,]    2    3    4    5
[3,]    3    4    5    6
[4,]    4    5    6    7
> b
[1] 1 2 3 4
```

(b) Given A and b as in part (a). Write down the results of the following expressions. Indicate ERROR if you think that the expression is not valid. (20 marks, 2 marks each)

- i. `A[3,]`
- ii. `dim(A)`
- iii. `b(c(3,1,1,2))`
- iv. `(A^2)[,1]`
- v. `A[c(-1,-2),]`
- vi. `b+10%%3`
- vii. `b[rep(2:3,2:3)]`
- viii. `(b<=3)&(A[,3]==3)`
- ix. `!(b>1)|(b<4)`
- x. `trunc(b/2-2)`

2. (a) With the use of `sample()` function, write down a command to generate a sample for rolling three fair dice. (4 marks)

Note: A die is a cube, with each of its six faces showing a different number of dots from 1 to 6.

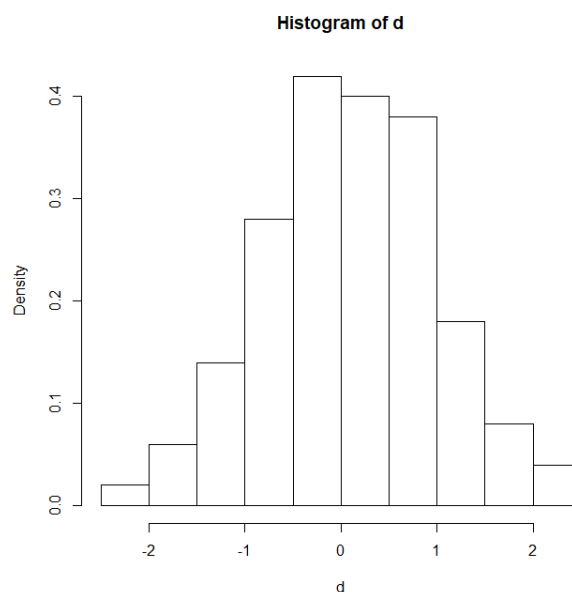
(b) Generate 3,000 random sample of rolling a fair die and save them as `r`. Transform `r` into a 1,000-by-3 matrix and save them again as `r`, such that each row in `r` represents the results of rolling three fair dice. (7 marks)

(c) Write down the command(s) to estimate the probability that the sum of three dice being an even number using the sample obtained in part (b). (7 marks)

(d) Suppose you have \$1,000 initially. For each roll of three dice, you win \$1 if the sum of three is even, and loss \$1 otherwise. Generate a vector of length 1,001 with the first entry being 1,000 and the remaining 1,000 entries contain your worth level after each roll. (7 marks)

3. (a) Given a vector of i.i.d. sample `d`, a histogram is generated with the following command.

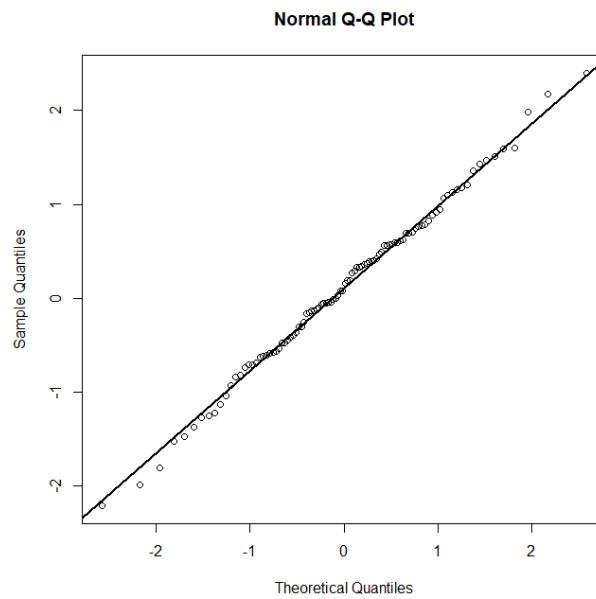
```
> hist(d, freq=F)
```



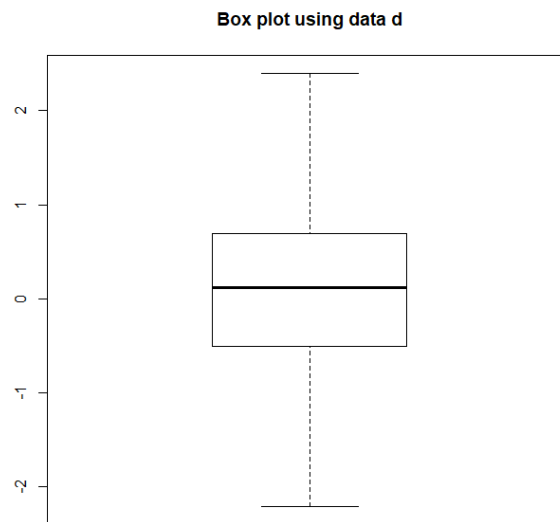
Add a normal density curve in red colour which fits the mean and standard deviation of the data in `d`. (6 marks)

(b) Generate a normal Q-Q plot using the data in `d` and add a reference line with width = 3. (5 marks)

(c) Explain the meanings and the implications of the following normal Q-Q plot. (6 marks)



(d) Create a box plot using the data in d with titles as below and explain the meanings of the plot. (8 marks)



Created on 25th Oct 2017

4. (a) Write down the matrix A generated by the following for-loop. (10 marks)

```
A <- matrix(rep(0,7^2), ncol=7)
for (i in 1:7) {
  for (j in 1:7) {
    if (i>j) {
      A[i,j] <- i
    } else if (i<j) {
      A[i,j] <- 1
    } else {
      A[i,j] <- 0
    }
  }
}
```

(b) Rewrite the following statements in for-loop using the while-loop to produce the same results in the numeric vector Fib. (8 marks)

```
Fib <- numeric(12)
Fib[1] <- Fib [2] <- 1
for (i in 3:12) {
  Fib[i] <- Fib [i - 2] + Fib[i - 1]
}
```

(c) Do you think that we can rewrite every for-loop into while-loop? Explain your answer. (7 marks)

End of Questions

Appendix

Useful R syntax given in the mid-term question paper.

(A) List of operators

Priority	Operator	Meaning
1	\$	component selection
2	[] [[]]	subscripts, elements
3	^	exponentiation
4	-	unary minus
5	:	sequence operator
6	% % / % % * %	modulus, integer division, matrix multiplication
7	* /	multiplication, division
8	+ -	addition, subtraction
9	< > <= >= == !=	comparison
10	!	logical negation
11	& &&	vectorized and or, control and or
12	<- ->	assignments

(B) Control flow

Syntax	Description
if (cond) expr	evaluate cond; if TRUE, evaluate expr
if (cond) expr1 else expr2	evaluate cond; if TRUE, expr1, if FALSE, expr2
ifelse(cond, expr1, expr2)	a vectorized version of if-else
switch(expr, ...)	evaluate expr and compare it to arguments
break	terminate current loop and jump out
next	terminate current iteration and immediately start next iteration of the loop
return(expr)	terminate current function and immediately return the value of expr
stop(message)	terminate evaluation of the current function and display message
while (cond) expr	evaluate cond; if TRUE evaluate expr, until cond is FALSE
repeat expr	repeat expr indefinitely, some breaks should be included inside expr
for (name in expr1) expr2	evaluate expr2 once for each name in expr1

(C) Statistical Distributions

Distribution	R name	Additional arguments
Binomial	binom	size, prob
Normal	norm	mean, sd
Uniform	unif	min, max

Adding the letter d, p, q or r in front of the R's name represents the density, cdf, quantile and pseudo-random number respectively.

(D) Build-in Functions

Name	Operations
ceiling(x)	smallest integer greater than or equal to x
floor(x)	largest integer less than or equal to x
trunc(x)	ignore the decimal part
round(x)	round up for positive x and round down for negative x
sqrt(x), abs(x)	return square root and absolute value of x
sample(v,n,p,replace=F)	generate n discrete random numbers which take value v with prob. p
sort(x,ascending=T)	sort the vector in ascending or descending order
order(x), rank(x)	return the order and rank of numbers in x
sum(x), prod(x)	return the sum and product of all the elements in a vector
cumsum(x), cumprod(x)	cumulative sum and product of elements of a vector
min(x), max(x), mean(x)	return the smallest, largest values and sample mean of x
var(x), sd(x)	return the sample variance and standard deviation of a vector
seq(a,b,by=c)	generate a sequence of number from a to b with step c
rep(x,n)	replicate elements in a vector n times
apply(d,n,f)	apply a function f to a matrix d, n=1 row, n=2 column
length(x)	return the length of a vector
list(x,y,...)	construct list of items x, y, ... etc
sample(x, size, replace, prob)	takes a sample of the specified size from the elements of x using either with or without replacement
plot(x,y,pch=,bg=)	plot (x,y), pch is print character, bg is background color
lines(x,y)	add lines or curves with coordinates (x,y) in the plot
abline(h=, v=)	add a horizontal or vertical line in the plot
matrix(data, nrow, ncol, byrow)	create a matrix from the given set of values in data
cbind(a,b), rbind(a,b)	binding two matrices column-wise and row-wise
dim(A), nrow(A), ncol(A)	return the row and column dimension of a matrix A