Upload solutions to *Problems* to Moodle by 5pm, Friday, Feb 19

Submit your solutions to the assessed questions in the form of a Word document — see instructions in Section 4 of the Course Moodle page.

Note the policy on late submissions stated on the Moodle page.

This lab sheet provides an introduction to R.

Before starting, read "Brief Introduction to the Statistics Package R".

Questions are labelled:

Exercises To give practice with concepts and methods,

Problems To be submitted for marking.

Work through the **Exercises during Lab 1** and on your own, as necessary. Your **homework is** to complete the **Problems**. Tutors will discuss the solutions to the Exercises but they will not provide hints for the Problems.

The arrangements will be the same for all five Computer Lab Sheets. This is a dry-run: Sheets 2 to 5 will form the **assessed coursework** for this Unit.

Exercise 1

Use R to do the following:

(a) Create a vector named `x` with elements −1, 2, 3, 0, 2, −1, 1, 4.

(b) Extract the 4th element of the vector `x`.

(c) Extract as a single vector the 2nd, 3rd, 4th and 5th elements of `x`.

(d) Try the effect of `x[1 <= x & x <= 3]` and `x[x < 1 | x > 3]`.

The first command in (d) extracts those elements of `x` that are at least 1 and at most 3, and the second one those that are less than 1 or greater than 3.

(e) Try setting `x[0]=-7`. What happens?

(f) Create a vector named `y` with elements $\frac{1}{101}, \frac{2}{101}, \ldots, \frac{100}{101}$.

(g) Create a vector called `z` that has 100 elements (you can make these anything you like — try to find a simple command to do this).

(h) The command `sum(v)` returns the sum of the elements of the vector `v`. Find the sum of squares of the elements of `y` above. (Answer: 33.16832.) Type `sum(c(1,12)^3)-sum(c(9,10)^3)` — a simple answer, but why?

(i) Create vector `a` with elements (1,2,3,4,5) and vector `b` with elements (0,6,0,4,6). Try the commands `a>3`, `a >= 3`, `a>b`, `max(a,b)` and `pmax(a,b)`. What would be a neat way to sum those elements of `a` that are strictly greater than the corresponding elements of `b`?

(j) The natural exponential and logarithm functions are called `exp` and `log`. Use these to evaluate $e^{(\ln 2)^2}$. (Answer: 1.616807.)

(k) **Recycling.** When `x` is a vector and `h` is a number, `x+h` will add `h` to each component of `x`. If `h` is a vector but `x` is longer than `h`, R *recycles* the value of `h` as many times as necessary to make a vector of the same length as `x`. Try the results of the following: `c(1,1,1,1,1,1)+c(2,3)` and `c(1,1,1,1,1)+c(2,3)`.

Exercise 2

The R command `dbinom(x,n,prob)` computes $P[X = x]$, where X is a Binomial RV with number of trials `n` and probability of success `prob`. The R command `pbinom(q,n,prob)` computes $P[X \leq q]$ for the same X. Thus, if X is Binom(n, p), with $n = 5$ and $p = 0.3$, then $P[X = 3]$ can be computed using `dbinom(3,5,0.3)`, and $P[X \leq 3]$ is computed using `pbinom(3,5,0.3)`.

A balanced die is rolled 100 times. Let X denote the number of times it shows a 6. Find two different ways of computing the probability that X is at least 25, one based on `pbinom`, and one based on `dbinom`. (Answer: 0.02170338.)

Information for Exercise 3

The R command `rbinom(1,1,prob)` returns the value `1` with probability `prob`, and `0` with probability `1-prob`. Use the command `rbinom(1,1,0.6)` to simulate the toss of a coin that lands H with probability 0.6.

The command `rbinom(nreps,1,prob)` simulates `nreps` *independent* trials, each with probability of success `prob`. So, tossing the same coin as above 10 times can be simulated with the command `rbinom(10,1,0.6)`. Try it.

Recall that the total number of successes in n independent trials, each with probability of success p, is a Binom(n, p) RV. This total can be simulated directly without generating results of the individual trials. The R command `rbinom(1,n,prob)` simulates a Binomial RV with `n` trials and probability of success `prob`. For example, the total number of Hs in 10 tosses of the coin above can be simulated with `rbinom(1,10,0.6)`. The result of this is an integer in the set $\{0, 1, 2, \ldots, 10\}$. Try it. Note that above we had `rbinom(1,1,prob)` as the special case of a single trial.

Finally, we may sometimes be interested in simulating `nreps` independent Binomial RVs, each with the same number of trials `n` and probability of success `prob`. This can be done using `rbinom(nreps,n,prob)`: the result is a vector with `nreps` integer elements, each between `0` and `n`.

Exercise 3

One hundred children rolled a die 10 times and counted the number of sixes.

(a) Simulate the outcome of this experiment.

(b) If `x` and `y` are vectors of length n, the R command `plot(x,y)` plots n points with the specified x and y coordinates. The command `plot(x,y,pch=20)` uses a prettier plotting character. Plot your results for the 100 children, with the x coordinates being $1, \ldots, 100$, and the y coordinates the number of sixes obtained by each child.

(c) Plot the probability mass function of a $\mathrm{Binom}(10, 1/6)$ RV.

(d) Does the plot in (c) match with the result of the experiment?

For homework

Problem 1

(a) (2 marks) Let X be the number of Heads in 100 tosses of a fair coin. Plot the probability mass function of X over the range $20 - 70$.

(b) (4 marks) Find the probabilities of the events $E_1 = \{X \leq 40\}$ and $E_2 = \{X \geq 70\}$ when X is the number of Heads in 100 tosses of a fair coin.

Find the probabilities of the same events, E_1 and E_2, when X is the number of Heads in 100 tosses of a coin which lands Heads with probability $2/3$.

(c) (2 marks) In 100 tosses, Coin 1 lands Heads 40 times, and Coin 2 lands Heads 70 times. If it is known that each coin is either fair or has $P(\text{Heads}) = 2/3$, draw your conclusions about the nature of each coin.

(d) (2 marks) Find the probability of the event $E_3 = \{X \text{ is even}\}$ when X is the number of Heads in 10 tosses of (i) a fair coin and (ii) a coin which lands Heads with probability $2/3$.

Problem 2

(a) (2 marks) In R, define a vector with elements $1, \dots, 6$, use this to create a vector with elements 10, 10^2, 10^3, 10^4, 10^5 and 10^6, and hence compute

$$\left(1 - \frac{1}{n}\right)^n \tag{1}$$

for $n = 10,\ 10^2, \dots, 10^6$.

(b) (2 marks) It is possible to represent the limit of (1) as $n \to \infty$ by a simple mathematical expression. Do this, guessing the answer if necessary. (Hint: It may help to examine the reciprocals of your answers in (a).)

(c) (4 marks) Each student attending a lecture course is asked to roll a die and choose a card at random from a standard pack of playing cards. If the student rolls a 6 and chooses the Ace of Spades, they win a chocolate egg. Suppose there are 312 students in the class, what is the probability that nobody wins a chocolate egg? Explain how your answer relates to part (b) of this question.

(d) (4 marks) Over a period of five years, the above experiment is conducted with a new cohort of 312 students each year.

(i) What is the probability that not a single chocolate egg is won over the five year period?

(ii) State the distribution of the total number of chocolate eggs awarded as prizes over the five years. Draw a sample of 20 realisations from this distribution and present an appropriate graphical display of this sample.

Presentation

(3 marks) Your submission should be well organized and easy to follow.

Make sure you tell the reader what each plot is meant to show.

Be sure to answer the question: so, for example, in Problem 1 (b), write

$\mathbf{P}(\boldsymbol{E_1}) = \mathbf{0.}\boldsymbol{xxxx}$ (giving as many decimal places as appropriate)

and in Problem 1 (c), state clearly

your opinion as to whether each coin is fair or not.

Using colour to highlight these answers will make your work easy to mark.