

## R PRACTICAL 1

DAVID STEINSALTZ

(First two based in large part on problems in the MSc course by Robin Evans.)

(1) Create the following vectors in R using `seq()` and `rep()`.

(a)  $1, 1.5, 2, \dots, 12$ .

(b)  $1, 8, 27, \dots, 1000$ .

(c)  $1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \dots, -\frac{1}{100}$ .

(d)  $1, 0, 3, 0, 5, 0, \dots, 49$ .

(e)  $1, 9, 36, 100, \dots, \sum_{i=1}^k i^3, \dots, 44100$ . [look up `?cumsum`]

(f)  $1, 3, 3, 5, 5, 5, \dots, \underbrace{19, \dots, 19}_{10 \text{ times}}$ .

(2) The  $i$ th term in the Taylor series expansion of  $\log(1+x)$  is  $(-1)^{i+1}x^i/i$ .

(a) Create a vector containing the first 100 terms for  $x = 0.5$ .

(b) Let

$$r_n(x) = \log(1+x) - \sum_{i=1}^n \frac{(-1)^{i+1}x^i}{i}.$$

Evaluate  $r_n(1)$  for  $n = 10, 100, 1000, \dots, 10^6$ .

(3) Type `load(url('http://steinsaltz.me.uk/DTC/abdata.RData'))` to load two vectors, named `a` and `b`. Think of `a[i]` and `b[i]` as being measurements of two different quantities for a given patient  $i$ .

(a) Use the functions `length`, `summary`, and others you know to figure out what you can about these vectors.

(b) Look up the function `unique`. Use it to determine how many individuals have identical values of `a`. And of `b`.

(c) Make new vectors `x` and `y`, containing in place  $i$  the larger of  $b$  and  $e^a$ , and the smaller of  $b$  and  $e^a$ , respectively

(d) `summary` gives you the quartiles of the vector; that is, the values that split up the range of the vector into equal quarters. Make new vectors `a1` through `a4`, consisting of the values of `a` corresponding to the patients whose `b` value is in the first, second, third, and fourth quartiles respectively.

(e) Investigate how spread out these four vectors are, and their means and medians. Does it seem that there is a connection between an individual's `a` and `b` values? Why or why not?

(f) Look up the `sort` and `order` functions. Use them to create a new vector `aSb` which contains the values of `b`, ordered by `a`. That is, the first element is the value `b[i]`, where  $i$  is the index of the patient with the smallest value of `a`. Next the `b[i]` corresponding to the second smallest value of `a`. And so on.