Introduction to R

1. Install R and RStudio on your laptop.
2. Introduction to R. Work through the following R instructions and find out what they do.

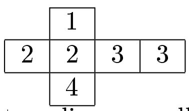
   (a)
   ```
   print("Hello World!")
   ?rnorm
   #I am a comment
   cat("Hello", "World!",123)
   ```
   (b)
   ```
   1+1
   3/3==1
   3.3/3==1.1
   c(1,2,3)+4
   c(1,2,3)==c(3,2,1)
   ```
   (c)
   ```
   x <- 1:10
   x+1
   x**2
   sin(x)
   sapply(x,sin)
   ```
   (d)
   ```
   x=1:20
   x[4]
   x[-4]
   x[seq(2,13,2)]
   x[-seq(2,13,2)]
   ```
   (e)
   ```
   #Plotting in R
   x <- seq(0,10,0.1)
   y <- x**2
   plot(x,y)
   plot(x,y,type="l",col=2)
   ```
   (f)
   ```
   #Random numbers, and summary statistics
   x <- rnorm(10**6)
   mean(x)
   summary(x)
   hist(x,40)
   library(MASS)
   truehist(x)
   ```
   (g)
   ```
   #Distributions
   x=seq(-3,3,0.01);plot(x,dnorm(x))
   x=seq(-3,3,0.01);plot(x,pnorm(x))
   x=seq(0,1,0.01);plot(x,qnorm(x))
   ```
   (h)
   ```
   #Plot random points in [0,1]^2. Why are there clusters?
   x <- runif(1000)
   y <- runif(1000)
   plot(x,y,pch=15,cex=0.3)
   ```
   (i)
   ```
   #Functions,
   parity <- function(x) {
     if (x %% 2 == 0) {
       cat("Number", x, "is even.\n")
     } else {
       cat("Number", x, "is odd.\n")
     }
   }
   ```

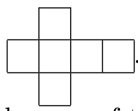
(j)
```
#Loops
for (x in 1:10) {
  parity(x)
}

x<-1
while(x<=10) {
  cat(x)
  x <- x+1
}

replicate(10,"Hello")
replicate(10,sample(4)) #Why is replicate not behaving like a normal function?
sapply(1:10,function(x) 2*x)
```

(k)
```
#Matrices in R
rbind(1:3,4:6,7:9)
cbind(1:3,4:6,7:9)

n <- 3
a <- matrix(rnorm(n**2),n,n)  # A random matrix
a * a                         # Entry-wise multiplication
a %*% a                       # Matrix multiplication
a[2,]                         # Second row of a
a[,2]                         # Second column of a
a[1:2,1:2]                    # The top left corner of a

n <- 1000
a <- matrix(rnorm(n**2),n,n)  # A bigger random matrix
s <- (a+t(a))/n               # Form a symmetric random matrix:
                              # t(.) means transpose
hist(eigen(s)$values)         # What shape do you get?
```

(l)
```
#Recursion
factorial<-function(n) {
  if (n==0) {
    1
  } else {
    n*factorial(n-1)
 }
}
factorial(10)

#Caution - simple recursion can be inefficient
fibonacci<-function(n) {
  if (n<=1) {
    1
  } else {
    fibonacci(n-1)+fibonacci(n-2)
  }
}
fibonacci(30)
```

Lab 1

1. What is a probability space $(S, \mathcal{B}, \mathbb{P})$?
2. Generate one thousand samples from the normal N(0,1) distribution.
   (a) How many lie between -1 and +1?
   (b) How many lie between -2 and +2?
   (c) Plot a histogram. Compare your answer to a plot of the `dnorm` function.
   (d) Sort the data and plot it. Compare your answer to the `pnorm` function.
   (e) What is the maximum value? How does the maximum value change if you resample your 1000 data points repeatedly?
3. You have a sequence of hospitals 1,2,...,1000. The n-th hospital does n operations per year. The outcome of each operation might be modelled by a N(0,1) random variable; higher is better. Each hospital is judged by the mean value M_n of all the operations there.
   (a) Plot M_n against n. What do you see?
   (b) Sort the hospitals by M_n. Which sort of hospitals rank highest? Or lowest
   (c) Suppose that just one hospital is providing poor service: the outcome of operations there is really better modelled as N(-0.5,1)? For which values of n might you be able to find it?
4. Sample X and Y from the N(0,1) distribution. Repeat this process 1000 times and plot the resulting (X,Y) points as a scatter plot (with aspect ratio 1).
   (a) Do some regions have surprisingly few points in?
   (b) Now rotate your laptop screen by 45°. Does the distribution of points change?
5. Sample X from the Uniform(0,1) distribution. What can you say about the distribution of $-\log(X)$?
6. Uncertainty:
   (a) What is the probability of Labour (or the Conservatives, Greens, etc) winning an overall majority in the next general election? How can you give a probability to this uncertain situation? [Perhaps look at http://www.paddypower.com/bet/politics/other-politics/uk-politics?ev_oc_grp_ids=600667]
   (b) How might you use that number? When might it be rational to gamble on such an event?
7. On any given day, your bicycle gets a flat tire with probability 0.01. Consider a period of 100 consecutive days. What is the probability that you get at least one flat tire during the period? Can you give an exact answer or just upper and lower bounds?
8. You have a die labelled as 

and a blank die, i.e. . Label the blank die so that when the two dice are rolled at the same time, the sum of the numbers shown has the same distribution as for two regular dice.
9. From Wikipedia, partially redacted: "Sally Clark (August 1964 – 15 March 2007) was a British solicitor ... was found guilty of the murder of two of her sons. ... Clark's first son died suddenly within a few weeks of his birth in September 1996, and in December 1998 her second died in a similar manner. A month later, she was arrested and subsequently tried for the murder of both children. The prosecution case relied on statistical evidence presented by pediatrician Professor Sir Roy Meadow, who testified that the chance of two children from an affluent family suffering sudden infant death syndrome was 1 in 73 million. He had arrived at this figure by squaring 1 in 8500, as being the likelihood of a cot death in similar circumstances." Is squaring the odds a reasonable thing to do?

10. Which of the following are generally true for events A, B, and C? Give a reason or a counter-example.

   (a) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
   (b) $A \cap (B \cap C) = (A \cap B) \cap C$
   (c) $(A \cup B) \cap C = A \cup (B \cap C)$
   (d) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$

11. Inclusion/Exclusion principle:

   (a) Show that $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
   (b) Show that

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) - \mathbb{P}(A \cap C) + \mathbb{P}(A \cap B \cap C).$$