Basic Concepts

1 Start

1.1 Simple calculations

Run and comment

```
> 2+2
> # blabla
> blah
> 2+2 # this is an addition
> pi
> exp(2)
> log(10)
> sin(5*pi)
> (1+3/5)*5
```

1.2 More complex calculations

Run and comment

```
> mean(c(4,10,16))
> sqrt(c(5,6))
> var(c(4,10,16))
```

Practical

1. Calculate the average of 1, 3, 5, 4, and 8.
2. Calculate the sum of 4, 10, and 16.
3. Calculate the median of 4, 10, and 16.

1.3 Assignment

Run and comment

```
> x <- pi
> print(x)
> x
> objects()
> y=pi
> objects()
> y
> x <- c(4,10,16)
> print(x)
> x
```

Practical

1. Calculate the max of `x`.
2. Calculate the min of `x`.
3. Calculate the mean of `x`.
4. Calculate the length of `x`.
5. Calculate the numeric summary of `x`.

2 Vectors

2.1 Calculation

Run and comment

```
> y=c(-1,5,0)
> x
> y
> x+y
> -y
> x+2
> abs(y)
> x*y
> x/y
> x^2
> 1:3
> 1:10
> -1:5
> -(1:5)
```

2.2 Logical

Logics are either `TRUE` or `FALSE` (which can be abbreviated as `T` and `F`).
Run and comment

```
> w <- c(TRUE,FALSE,FALSE)
> sum(w)
> any(w)
> all(w)
> !w
> (TRUE)&(FALSE)
> (TRUE)|(FALSE)
> (TRUE)|(TRUE)
```

2.3 Special values

The value `NA` is the missing value, `NaN` is the Not a Number value (indeterminate form), `Inf` is infinity.

Run and comment

```
> log(0)
> log(Inf)
> 1/0
> 0/0
> max(c( 0/0,1,10))
> max(c(NA,1,10))
> max(c(-Inf,1,10))
> is.finite(c(-Inf,1,10))
> is.na(c(NA,1,10))
> is.nan(c(NaN,1,10))
```

Practical

1. Create the vector of integers from 5 to 23.
2. Create the vector from 6 to 24 going from 2 to 2.
3. Create the vector of 100 regularly spaced values between 0 and 1.
4. Create the following vector

   ```
   [1] 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
   ```

5. Create the following vector

   ```
   [1] 1 1 1 2 2 2 3 3 3 4 4 4 5 5 5
   ```

6. Create the following vector

   ```
   [1] 1 1 2 2 2 3 3 3 3
   ```

2.4 Selection within a vector

Run and comment

```
> x[1]
> x[2]
> x[c(1,2,3)]]
> x[1:3]
> x[c(2,2,1,3)]
> x[c(1:3,2,1)]
> x[-1]
> x[-c(1,2)]
> x[-(1:2)]
```

Run and comment

```
> objects()
> vec1<-c(3,NA,4)
> objects()
> vec2<-c(FALSE,TRUE,FALSE)
> objects(pattern='vec*')
> vec2
> vec1
> vec1[vec2]
> is.na(vec1)
```

2.5 Characters

Run and comment

```
> z=c('iut','fds')
> z[1]
> paste('m',1:3)
> paste('m',1:3,sep='')
> c(paste('m',1:3,sep=''),paste('p',1:4,sep='.'),z)
```

3 Datasets

Practical

1. Import the data from the table contained in the file `tab1.ods` into the object R which you will call `don1`. The result of the screen display should be

   ```
   > don1
   V1 V2
   1 1 2
   2 0 2
   3 3 1
   ```

2. Import the data from the table contained in the file `tab2.xls` into the object R which you will call `don2`. The result of the screen display should be

   ```
   > don2
   variable1 variable2
   1 -1.0  0
   2  2.0 -2
   3  3.1  4
   ```

3. Import the data from the table contained in the file `tab3.xls` into the object R which you will call `don3`. The result of the screen display should be

   ```
   > don3
   sex    size
   gege   male   180.6
   simone female 175.2
   albert male   172.9
   ```

4. Import the data from the table contained in the file `tournesols.csv` into the object R which you will call `tournesols`. This table contains the measurements, on different statistical individuals (wild sunflower plants), of the variables described in the table 1.

   This table will be assigned (after import into R) in the object called `tournesols` and the summary of this dataset is obtained using

   ```
   > summary(tournesols)
   ```

3.1 Names of the variables (columns) and individuals (rows)

Run and comment

```
> rownames(don3) ## it's a vector
> names(don3) ## it's a vector
> colnames(don3) ## it's a vector
> colnames(don1)[2]
> colnames(don1)[2] <- 'var2'
> colnames(don1)
> colnames(don1) <- colnames(don2)
> colnames(don1) <- c('variable1', 'variable2')
```

3.2 Selection in datasets

Run and comment

```
> don1[1,]
> don3[, 'sex']
> don3$sex
> don3[,2]
> don3[,c(FALSE,TRUE)]
> don3[,c('size', 'gender')]]
> don1[1,2]
> don1[,1:2]
> don1[-1,]
> don1[c(2,3),c(2,1)]
> don1[c(TRUE,FALSE,TRUE),]
> don1[don1[,1]>0,]
> don1[-(1),]
> don1[,c(2,1)]
```

3.3 Operation on columns

Run and comment

```
> don1[,1]+don1[,2]
> exp(don1[,1])
> don3[1,]+don3[2,]
```

Variable code	Variable description
ecotype	plant code
plt	plant number of a given ecotype
state	state of origin of the plant
longitude	longitude of the place of collection
latitude	latitude of place of collection
height	plant height
semflo	day of flowering
rambas	basal branching score
longfeu	cumulated length of leaf blade and petiole
grlon	maximum length of the seed
oil	percentage of oil

Table 1: Variables measured on sunflowers

.

4 Eliminate missing values

Import the table `tournesols_brut.csv`.
Knowing the missing values (`NA`) of a vector

```
> which(is.na(tbrut[,'oil'])
```

When you want to know the missing values (`NA`) of an entire table:

```
> which(is.na(tbrut),arr.ind=TRUE)
```