

## Lab 2 Exercises

### Exercise 1. The use of the c and sum functions.

This exercise uses epidemiological data. Vicente et al. (2006) analysed data from observations of wild boar and red deer reared on a number of estates in Spain.

The dataset contains information on tuberculosis (Tb) in both species, and on the parasite *Elaphostrongylus cervi*, which only infects red deer.

In Zuur et al. (2009), Tb was modelled as a function of the continuous explanatory variable, length of the animal, denoted by LengthCT (CT is an abbreviation of cabeza-tronco, which is Spanish for head-body). Tb and Ecervi are shown as a vector of zeros and ones representing absence or presence of Tb and *E. cervi* larvae. Below, the first seven rows of the spreadsheet containing the deer data are given.

Farm	Month	Year	Sex	LengthClass	LengthCT	Ecervi	Tb
MO	11	00	1	1	75	0	0
MO	07	00	2	1	85	0	0
MO	07	01	2	1	91.6	0	1
MO	NA	NA	2	1	95	NA	NA
LN	09	03	1	1	NA	0	0
SE	09	03	2	1	105.5	0	0
QM	11	02	2	1	106	0	0

Using the c function, create a variable that contains the length values of the seven animals.

```
> LengthCT <- c(75, 85, 91.6, 95, NA, 105.5, 106)
```

Also create a variable that contains the Tb values. Include the NAs.

```
> Tb <- c(0, 0, 1, NA, 0, 0, 0)
```

What is the average length of the seven animals?

```
> mean(LengthCT, na.rm = TRUE)
```

## CIND 123 Data Analytics: Basic Methods

**Exercise 2.** The use of the cbind function using epidemiological data. We continue with the deer from Exercise 1. First create variables Farm and Month that contain the relevant information. Note that Farm is a string of characters.

```
Farm <- c("MO", "MO", "MO", "MO", "LN", "SE", "QM")
```

#Or:

```
#Farm <- rep(c("MO", "LN", "SE", "QM"), c(4, 1, 1, 1))
```

```
Month <- c(11, 7, 7, NA, 9, 9, 11)
```

Use the cbind command to combine month, length, and Tb data, and store the results in the variable, Boar.

```
LengthCT <- c(75, 85, 91.6, 95, NA, 105.5, 106)
```

```
Tb <- c(0, 0, 1, NA, 0, 0, 0)
```

```
Boar <- cbind(Month, LengthCT, Tb)
```

Make sure that you can extract rows, columns, and elements of Boar. Use the dim, nrow, and ncol functions to determine the number of animals and variables in Boar.

```
> Boar[,1]
```

```
> dim(Boar)
```

```
> nrow(Boar)
```

```
> ncol(Boar)
```

## CIND 123 Data Analytics: Basic Methods

**Exercise 3.** The use of the vector function using epidemiological data.

We continue with the deer from Exercise 1. Instead of the c function that you used in Exercise 2 to combine the Tb data, can you do the same with the Vector function? Give the vector a different name, for example, Tb2.

```
Tb2 <- vector(length = 7)
Tb2[] <- NA
Tb2[1] <- 0
Tb2[2] <- 0
Tb2[3] <- 1
Tb2[5] <- 0
Tb2[6] <- 0
Tb2[7] <- 0
Tb2
```

**Exercise 4.** Working with a matrix.

Create the following matrix in R and determine its transpose, its inverse, and multiply D by its inverse (the outcome should be the identity matrix).

$$D = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 2 & 1 \\ 2 & 3 & 0 \end{pmatrix}$$

```
D <- matrix(nrow = 3, ncol = 3)
```

```
D[1,] <- c(1, 2, 3)
```

```
D[2,] <- c(4, 2, 1)
```

```
D[3,] <- c(2, 2, 0)
```

```
D
```

```
t(D)           #Transpose
```

```
solve(D)       #Inverse
```

```
D %*% solve(D) #Should be the identity matrix
```