

## DATA202/472 - Practice Test questions

### Solutions

**Q1. (4 Marks)** Classify the following variables as boolean, nominal categorical, ordinal categorical, binary, integer or continuous. Give the R data type that would be used to store each type of variable.

- Name of a species of bird.
- Number of rooms in an apartment
- Level of agreement with the statement “I am satisfied with my home circumstances” (agree, neutral, disagree).
- Whether a person was born in New Zealand or not.

### Q2. (6 Marks)

- Write R code that creates a vector called `x` with the 4 elements 5, 3, 1, 10 and a vector called `y` with the 4 elements 3, 3, 1 and -1.
- Write R code that multiplies the each of the elements of `x` with the corresponding elements of `y`, and give the value of the vector that will result.
- Write R code that extracts the first and last elements of `x` and stores them in a vector `z`, and give the value of `z` that will result.
- The code `a <- cbind(x, y)` creates a two column array `a`. Write out its values.
- Write code that extracts second and third rows of `a` and saves the result in an object `b`.
- What will be the output of the expression `dim(b)`?

**Q3. (4 Marks)**

The vectors  $u$ ,  $v$  and  $w$  have the following values:

$$u = \begin{bmatrix} 5 \\ 1 \\ 4 \end{bmatrix} \quad v = \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} \quad w = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

- What is the value of the expression  $u >= v$ ?
- What is the value of the expression  $u[u >= v]$ ?
- What is the value of the expression  $w[u >= v]$ ?
- What is the value of the expression  $w[u == v]$ ?

**Q4. (3 Marks)** A survey of 1200 patients who have completed treatment in a hospital asks each patient about their experience. A subset of the data is available in a file called `patientsurvey.csv` with 1200 rows and 5 columns:

- `PatientID` = the unique identification number of the patient, issued by the hospital
- `Age` = the patient's age in years
- `Length` = the number of days the patient was in hospital
- `Reason` = a coded value with the patient's primary diagnosis
- `Satisfied` = level of satisfaction (scale 1-5) with the standard of care provided

- Write code that will read the file into R and create a data frame called `pdata`
- Some of the patients have missing age information: after reading the data file into R their ages appear as `NA` in the `Age` column of `pdata`. Write code that eliminates these rows from the data frame, and stores the result in the same object `pdata`.
- Write code to find out how many rows remain in `pdata`.

**Q5. (6 Marks)** Assume you have the data frame `pdata` from the previous question already read into R.

In the spaces provided below explain what each of the following lines of code produces. If the code won't execute, and will produce an error message, say what is wrong with the code.

- `pdata`
- `pdata[1:5,]`
- `pdata[,4:10]`
- `pdata[,c(3,1,1)]`
- `pdata$Age[1:5]`
- `pdata$Reason[100,]`

**Q6. (5 Marks)** Consider the following code which defines a function called `gfunc()`. (There is some code missing in the first line, indicated by `*****`.)

```
gfunc <- function( ***** ) {  
  out <- NA  
  if(x>y) {  
    out <- x  
  } else if(y>x) {  
    out <- y  
  }  
  return(out)  
}
```

- (a) What does NA mean here?
- (b) What code should replace the asterisks `*****` in the first line? (The number of asterisks is not related to the amount of code that is missing.)
- (c) If the function is called using the code

```
gfunc(2,4)
```

what will happen?

- (d) If the function is called using the code

```
gfunc(7,7)
```

what will happen?

**(Quiz total: 28 Marks)**