

Name: Greendee Roper B. Panogalon BSIT 3B

## *Handling Missing Values*

Questions 1-3: Write the correct R syntax

```
# vector with missing data
```

```
x <- c(15, 20, NA, 30, NA, 45)
```

1. Remove NAs from the vector and display the results.

```
> x <- c(15, 20, NA, 30, NA, 45)
> x <- na.omit(x)
> print(x)
[1] 15 20 30 45
attr(,"na.action")
[1] 3 5
attr(,"class")
[1] "omit"
> |
```

2. Identify the location of NAs in the vector

```
> x <- c(15, 20, NA, 30, NA, 45)
> naLocations <- which(is.na(x))
> print(naLocations)
[1] 3 5
> |
```

3. Count the number of NAs in the vector

```
> x <- c(15, 20, NA, 30, NA, 45)
> naCount <- sum(is.na(x))
> print(naCount)
[1] 2
> |
```

## Questions 4-6: Tracing and recoding

```
x <- c(2:5, NA, NA, 7:9, NA)
```

4. What are the values of x when displayed?

```
> x <- c(2:5, NA, NA, 7:9, NA)
> print(x)
[1] 2 3 4 5 NA NA 7 8 9 NA
> |
```

5. Replace all missing values with mean.

```
> x <- c(2:5, NA, NA, 7:9, NA)
> x[is.na(x)] <- mean(x, na.rm = TRUE)
> print(x)
[1] 2.000000 3.000000 4.000000 5.000000 5.428571 5.428571 7.000000 8.000000
[9] 9.000000 5.428571
```

6. Display the new vector rounding to 2 decimal places.

```
> x <- c(2:5, NA, NA, 7:9, NA)
> x[is.na(x)] <- mean(x, na.rm = TRUE)
> print(round(x, 2))
[1] 2.00 3.00 4.00 5.00 5.43 5.43 7.00 8.00 9.00 5.43
> |
```

Question 7-15: Fill the NA values with the median of the corresponding and the display the result.

```
> dataframe <- data.frame(
+   Name = c("Jon", "Anna", "Alex", "Mona"),
+   Physics = c(98, 87, 91, 94),
+   Chemistry = c(NA, 84, 93, 87),
+   Mathematics = c(91, 86, NA, NA)
+ )
> dataframe$Chemistry[is.na(dataframe$Chemistry)] <- median(dataframe$Chemistry)
> dataframe$Mathematics[is.na(dataframe$Mathematics)] <- median(dataframe$Mathematics)
> print(dataframe)
  Name Physics Chemistry Mathematics
1  Jon      98         87          91.0
2 Anna      87         84          86.0
3 Alex      91         93          88.5
4 Mona      94         87          88.5
> |
```