

# STATS 782 Assignment 1

Chase Robertson - crob873

15 March 2022

I have read the declaration on the cover sheet and confirm my agreement with it.

## Question 1

(a)

A sequence of numbers between 1 and 3 using a step of 0.2

```
seq(1, 3, 0.2)
```

```
## [1] 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0
```

(b)

The sequence of integers between 1 and 4, repeated 3 times.

```
rep(1:4, times=3)
```

```
## [1] 1 2 3 4 1 2 3 4 1 2 3 4
```

(c)

A vector of strings where each of the first 7 letters is pasted to its index with a space.

```
paste(letters[1:7], 1:7)
```

```
## [1] "a 1" "b 2" "c 3" "d 4" "e 5" "f 6" "g 7"
```

(d)

A palindrome of inverse factors of 10, from 0.01 to 1 and back.

```
1 / 10**abs(-2:2)
```

```
## [1] 0.01 0.10 1.00 0.10 0.01
```

(e)

A string with all uppercase letters of the alphabet interlaced next to their lowercase.

```
paste0(LETTERS[1:26], letters[1:26], collapse="")
```

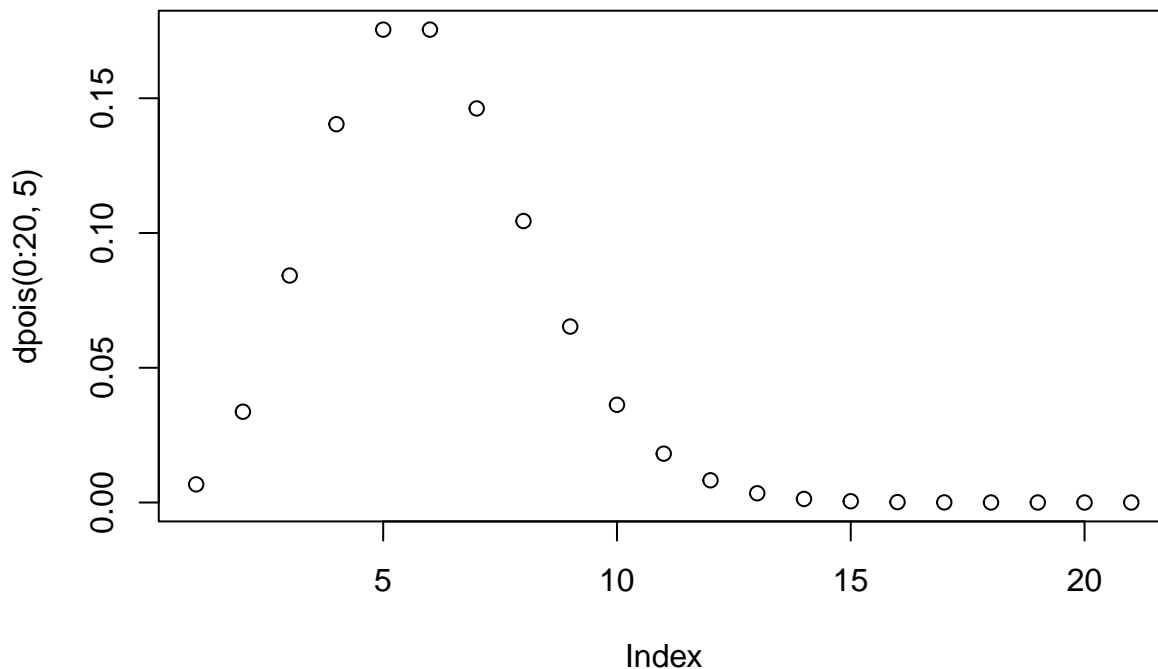
```
## [1] "AaBbCcDdEeFfGgHhIiJjKkLlMmNnOoPpQqRrSsTtUuVvWwXxYyZz"
```

## Question 2

(a)

The probability mass function from 0 to 20 of a Poisson distribution with  $\lambda = 5$ .

```
plot(dpois(0:20, 5))
```



(b)

Calculate  $P(x = 8) + P(x = 9) + P(x = 10)$  by calculating  $P(7 < x \leq 10)$ .

```
ppois(10, 5) - ppois(7, 5)
```

```
## [1] 0.1196764
```

(c)

Calculate  $P(x = 8) + P(x = 9) + P(x = 10)$  directly.

```
sum(dpois(8:10, 5))
```

```
## [1] 0.1196764
```

(d)

Calculate  $P(x = 8) + P(x = 9) + P(x = 10)$  by Monte Carlo simulation.

```
# simulate a single independent month of crashes  
sim_month <- function() {  
  rpois(1, 5)  
}
```

```
# repeat for 10,000 months  
results <- 0  
for (i in 1:10000) {
```

```

    results[i] = sim_month()
}

# separate months with between 8 and 10 crashes
eight_to_ten <- results[results >= 8 & results <= 10]

length(eight_to_ten) / length(results)

## [1] 0.1195

```

## Question 3

A function `analyse_text` which takes a single string and returns a named vector containing the number of whitespace-separated words, the length of the longest word, and the mean word length. Punctuation is ignored.

```

analyse_text <- function(x) {
  result <- vector('numeric', 3)
  names(result) <- c('num_words', 'max_word_length', 'mean_word_length')

  # remove special characters
  clean_x = gsub('[:punct:]', '', x)

  # split string into vector of individual words
  words = strsplit(clean_x, " ")[[1]]
  w_i = seq(along=words)

  result['num_words'] = length(words)
  result['max_word_length'] = max(nchar(words[w_i]))
  result['mean_word_length'] = mean(nchar(words[w_i]))

  result
}

```

Two tests of the above function.

```

analyse_text("Hello!")

##          num_words  max_word_length mean_word_length
##              1              5              5

analyse_text("The quick brown fox jumps over the lazy dog.")

##          num_words  max_word_length mean_word_length
##          9.000000      5.000000      3.888889

```

EOF