Midterm1

Adam Kaderbhai

```
library(readxl)
library(tidyverse)
```

a.)

```
my_data <- read_excel("body_temps.xlsx")
glimpse(my_data)
```

```
Rows: 106
Columns: 3
$ Smoke <chr> "Y", "Y", "N", "N", "Y", "Y", "Y", "Y", "N", "Y", "Y", "N", "Y",~
$ Sex   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0~
$ Temp  <dbl> 98.6, 98.6, 98.0, 98.0, 99.0, 98.4, 98.4, 98.4, 98.4, 98.6, 98.6~
```

- Sex - Sex is a categorical data because it does not involve numerical measurements 0s and 1s here are used as labels. The level of measurement is nominal because there is no specific level of order.

- Smoke - This is Categorical and the level of measurement is nominal as there is no specific level of order

- Temp - This is Quantitative data as it includes numerical measurements and the level of measurement is Interval as 0 here means something and we can't discuss about temperatures in terms of ratios

b.)

```
mean(my_data$Temp)
```

```
[1] 98.2
```

```
quantile(my_data$Temp)
```

```
  0%  25%  50%  75% 100%
96.5 97.8 98.4 98.6 99.6
```

The mean is 98.2 and the Median is 98.4

c.)

```
sd(my_data$Temp)
```

```
[1] 0.6228965
```

Standard deviation shows the average spread of the values in the dataset from the mean. In our case, individual data points on average are approximately 0.623 units away from the mean of our data which is 98.2.

d.)

```
quantile(my_data$Temp)
```

```
  0%  25%  50%  75% 100%
96.5 97.8 98.4 98.6 99.6
```

```
98.6 - 97.8
```

```
[1] 0.8
```

IQR is calculated by taking the difference between Q3(75%) and Q1(25%). This shows how the values are spread in the middle 50% of our data. This tells us that the difference between the highest and the lowest observation in the middle 50% of our data is 0.8.

e.)

```
quantile(my_data$Temp, 0.20)
```

```
 20%
97.6
```

97.6 falls in the 20th percentile of temperatures in the survey

f.)

```
ecdf(my_data$Temp)(98.6)
```

```
[1] 0.7830189
```

```
0.78 * 100
```

```
[1] 78
```

98.6 falls in the 78th percentile. This means that 78% of our data is less or equal to 98.6