

Homework Assignment

Team Member1 and Team Member2

October 12, 2021

1. Write an R function that returns the following dispersion measures:

- Estimator of standard deviation (SD):

$$SD = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

```
s.d <- function(x){  
  n <- length(x) # Sample size  
  s2 <- sum((x - mean(x))^2)/(n-1) # sample variance  
  s.d <- sqrt(s2) # sample standard deviation  
  return(s.d)  
}
```

- Estimator of mean absolute deviation (MAD):

$$MAD = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

```
mean.abs.d <- function(x){  
  n <- length(x) # Sample size  
  m <- sum(abs(x - mean(x)))/n # mean average deviation  
  return(m)  
}
```

2. Construct box-plots, histograms, QQ-plots and kernel density estimates for these variables. Comment on features such as the distribution and outliers in these plots.

When asked to construct a graph, you should always precede your graph by the R command/function that generated it properly annotated.

```
library(MASS)  
pima2 = rbind(Pima.tr, Pima.tr2, Pima.te)  
  
x = pima2$age  
var.name = 'age'  
  
library(ggplot2)  
  
ggplot(pima2, aes(x=age)) + geom_histogram()  
  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
ggplot(pima2, aes(x = factor(0), y = age)) + geom_boxplot() + xlab("") +  
  scale_x_discrete(breaks = NULL) + coord_flip()
```

```

y    <- quantile(pima2$age, c(0.25, 0.75)) # Find the 1st and 3rd quartiles
x    <- qnorm( c(0.25, 0.75))              # Find the matching normal values on the x-axis
slope <- diff(y) / diff(x)                 # Compute the line slope
int  <- y[1] - slope * x[1]                # Compute the line intercept

ggplot(pima2, aes(sample=age)) + stat_qq() +
  geom_abline(intercept=int, slope=slope, color='red')

```

