Assignment 1

The Old Republic

2023-06-21

```
library(Pareto)
set.seed(100)
Data = data.frame(x.n = rnorm(50000), x.p = rPareto(50000, t=1, alpha=2))
summary(Data)
```

```
##       x.n                 x.p         
##  Min.   :-4.087893   Min.   :  1.000  
##  1st Qu.:-0.671144   1st Qu.:  1.154  
##  Median :-0.005919   Median :  1.412  
##  Mean   :-0.000208   Mean   :  1.994  
##  3rd Qu.: 0.672466   3rd Qu.:  1.992  
##  Max.   : 4.363243   Max.   :159.275
```

Question 1

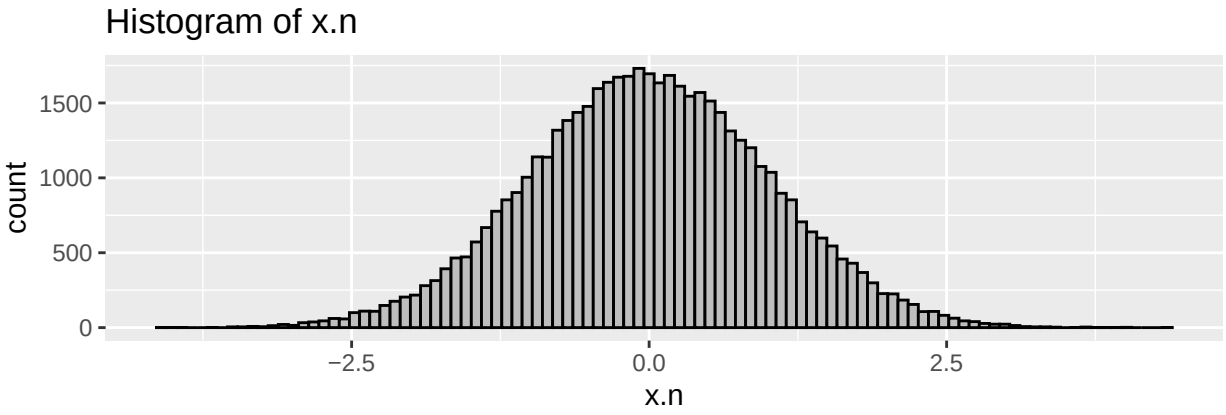
1. Histogram and Box Plot of the Variable x.n

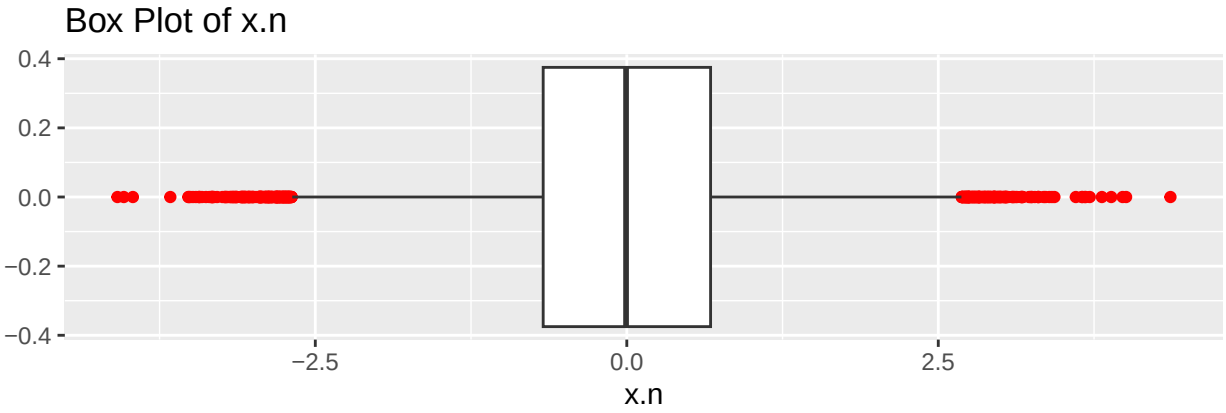
```
library(ggplot2)
library(grid)
library(gridExtra)

x.n.hist = ggplot(data=Data, aes(x=x.n)) +
  geom_histogram(bins=100, fill="grey", color="black") +
  ggtitle("Histogram of x.n")

x.n.box = ggplot(data=Data, mapping=aes(x=x.n)) +
  geom_boxplot(outlier.color="red") +
  ggtitle("Box Plot of x.n")

grid.arrange(x.n.hist,x.n.box)
```





2. The sample mean, and standard deviation of x.n are $-2.0849558 \times 10^{-4}$ and 0.9989658 respectively rounded to 3 decimal places. It is evident that these parameters are very close to the parameters of a variable Z following a standard normal distribution, namely $Z \sim N(0,1)$. This result is not surprising since the observations were taken from a normal distribution with $\mu = 0$ and $\sigma^2 = 1$.

3. Judging by the shape of the distribution of the observations of x.n and the values for the mean and standard deviation of these observations, we may conclude that $x.n \sim N(0,1)$ approximately. This means that it is highly likely that the new observations from the variable x.n will have values close to 0 (μ). More specifically, we know that approximately 68% of the observations will be between -1 and 1 ($\mu \pm \sigma$), 95% of the observations will be between -2 and 2 ($\mu \pm 2\sigma$) and 99.7% of the observations will be between -3 and 3 ($\mu \pm 3\sigma$).

4. Mean and standard deviation of the variable x.p

```
x.p.mean = mean(Data$x.p)
x.p.sd = sd(Data$x.p)
x.p.IQR = quantile(Data$x.p, 0.75)-quantile(Data$x.p, 0.25)
```

Histogram and Box Plot of the variable x.p

```
x.p.hist = ggplot(data=Data, mapping=aes(x=x.p)) +
  geom_histogram(bins=50, fill="grey", color="black",) +
  ggtitle("x.p Histogram") +
  geom_vline(aes(xintercept=x.p.mean), color="blue", linetype="dashed") +
  geom_vline(aes(xintercept= (quantile(x.p, 0.75) + 1.5*x.p.IQR)), color="red", linetype="dashed")+
  geom_vline(aes(xintercept= (quantile(x.p, 0.25) - 1.5*x.p.IQR)), color="red", linetype="dashed")

x.p.box = ggplot(data=Data, mapping=aes(x=x.p)) +
  geom_boxplot() +
```

```
  ggtitle("x.p box plot")

grid.arrange(x.p.hist, x.p.box)
```