WO1D

Benjamin Huang

2024-09-05

Standard Deviation is a measure of spread

$$S_Y = \sqrt{\frac{1}{n-1}\sum(y_i - \bar{y})^2}$$

```
#assigns values to A
A <- c(4, 5, 6)
#sample deviation of A
sd(A)
```

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

Q1. Use the code above to find the standard deviation of B=(3, 5, 7)

```
B <- c(3, 5, 7)
sd(B)
```

```
## [1] 2
```

Correlation describes the linear relationship between A and B.

$$r_{X,Y} = \frac{1}{(n-1)s_X s_Y}\sum(x_i - \bar{x})(y_i - \bar{y})$$

Q2. Use the cor() function to find the correlation between A and B

```
A <- c(4, 5, 6)
B <- c(3, 5, 7)
cor(A, B)
```

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

Sometimes you need to use a function that isn't in base R. Fortunately, R is a open resource and many packages have been created to solve these problems.

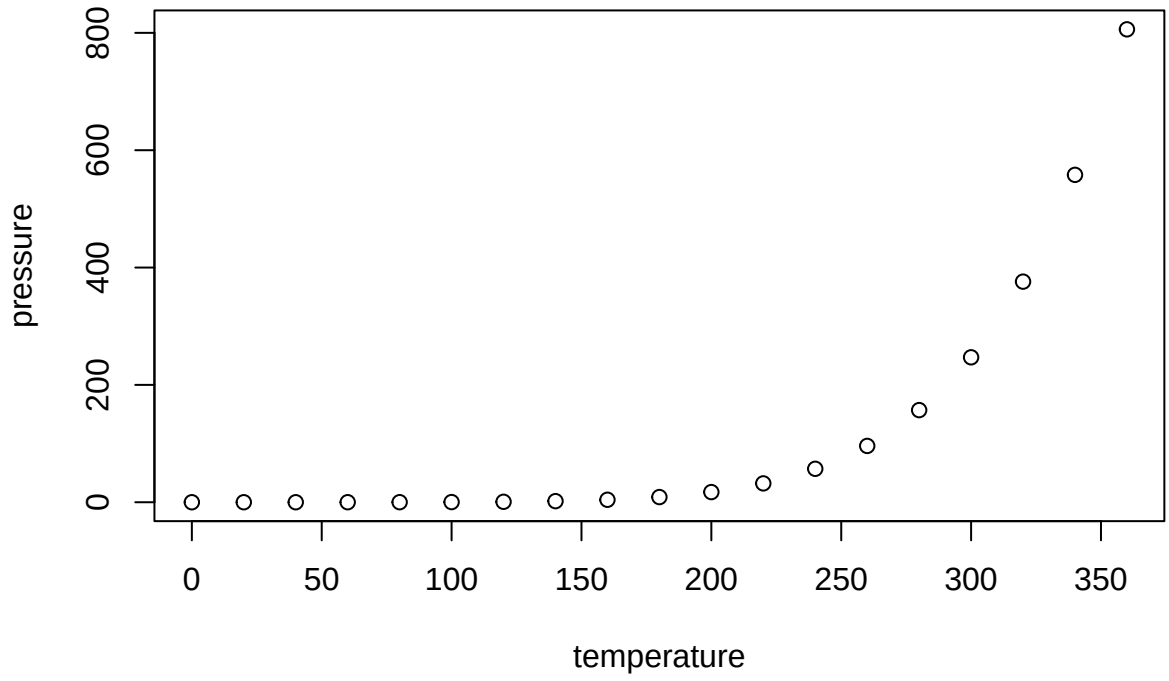
```
#install.packages("moments")
library(moments)
A <- c(4, 5, 6)
skewness(A)
```

```
## [1] 0
```

A is symmetrical around its mean, 5.

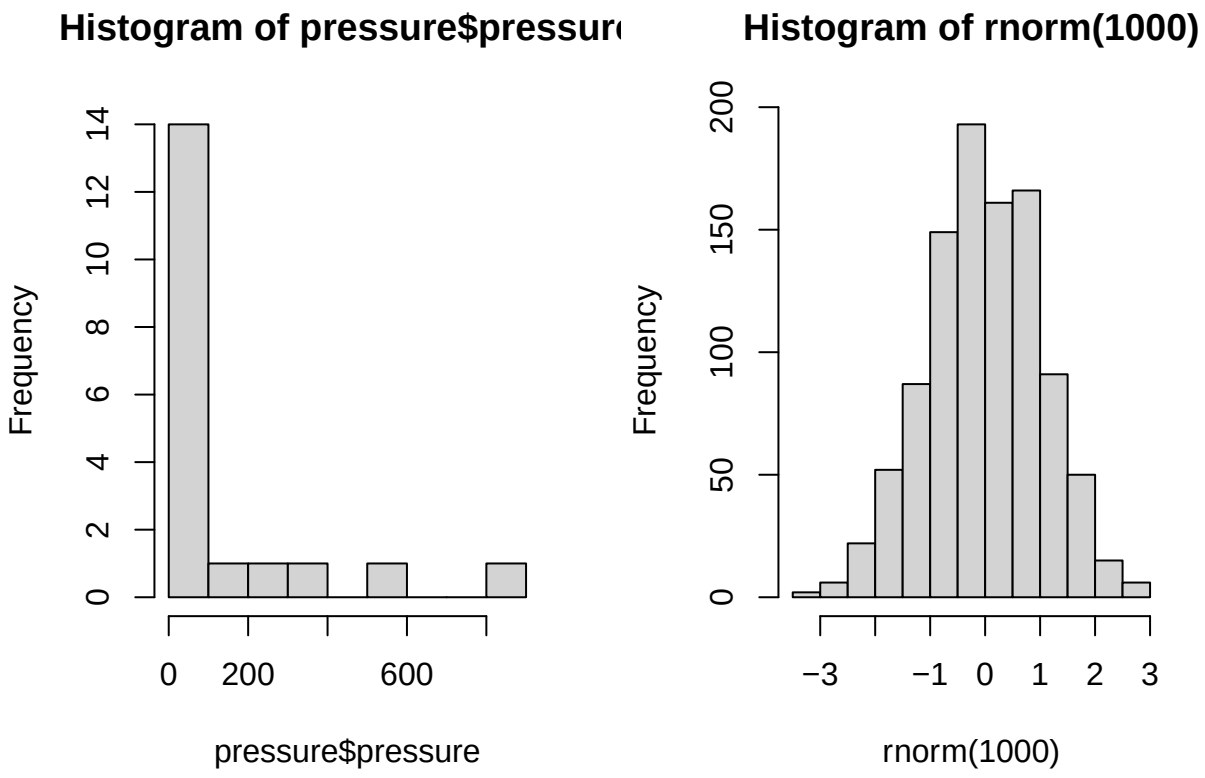
Q3. Plot and find the skewness of the data set pressure. Is this data normally distributed?

```
data("pressure")
plot(pressure)
```



A histogram can make it clear if a data set is normally distributed.

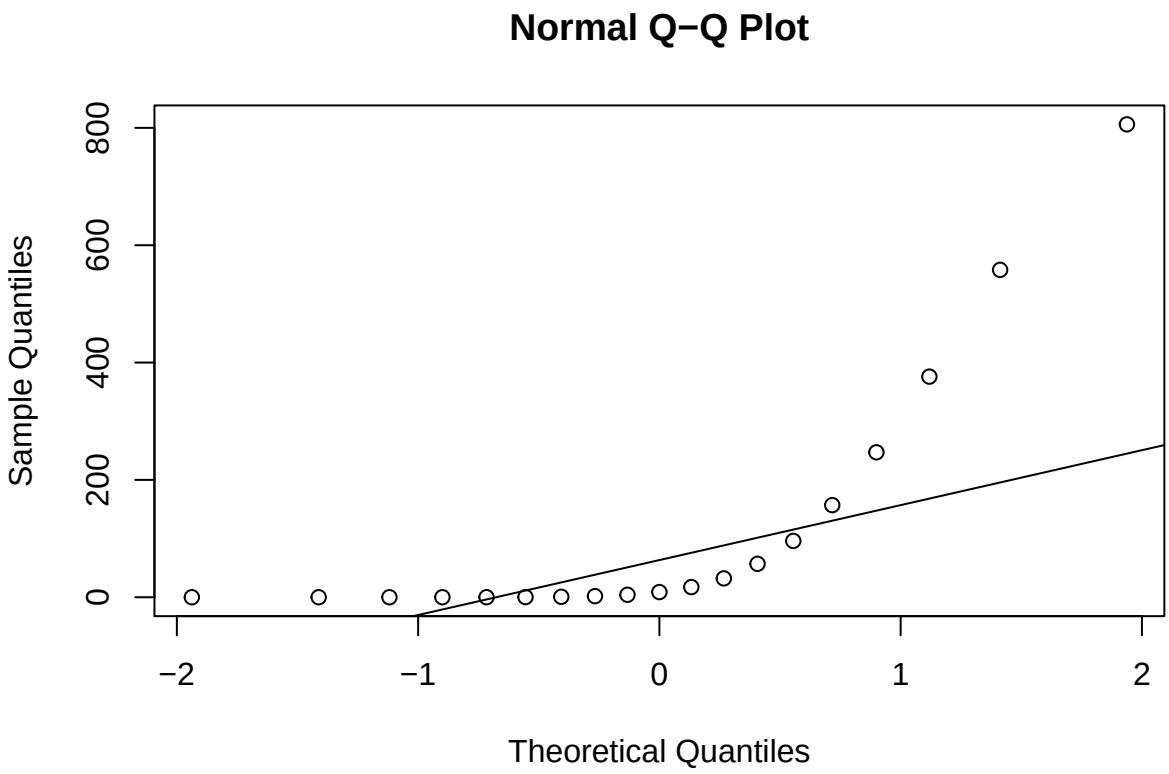
```
set.seed(441)
par(mfrow = c(1,2))
hist(pressure$pressure)
hist(rnorm(1000))
```




```
#they do not look similar
```

Another graphical way of checking if data is normally distributed is with a qq plot (Quantile-Quantile plot). The idea is to plot the percentiles, or quantiles, of one distribution against the other. If they come from the same distribution, then the points should lie on a line

```
qqnorm(pressure$pressure); qqline(pressure$pressure)
```



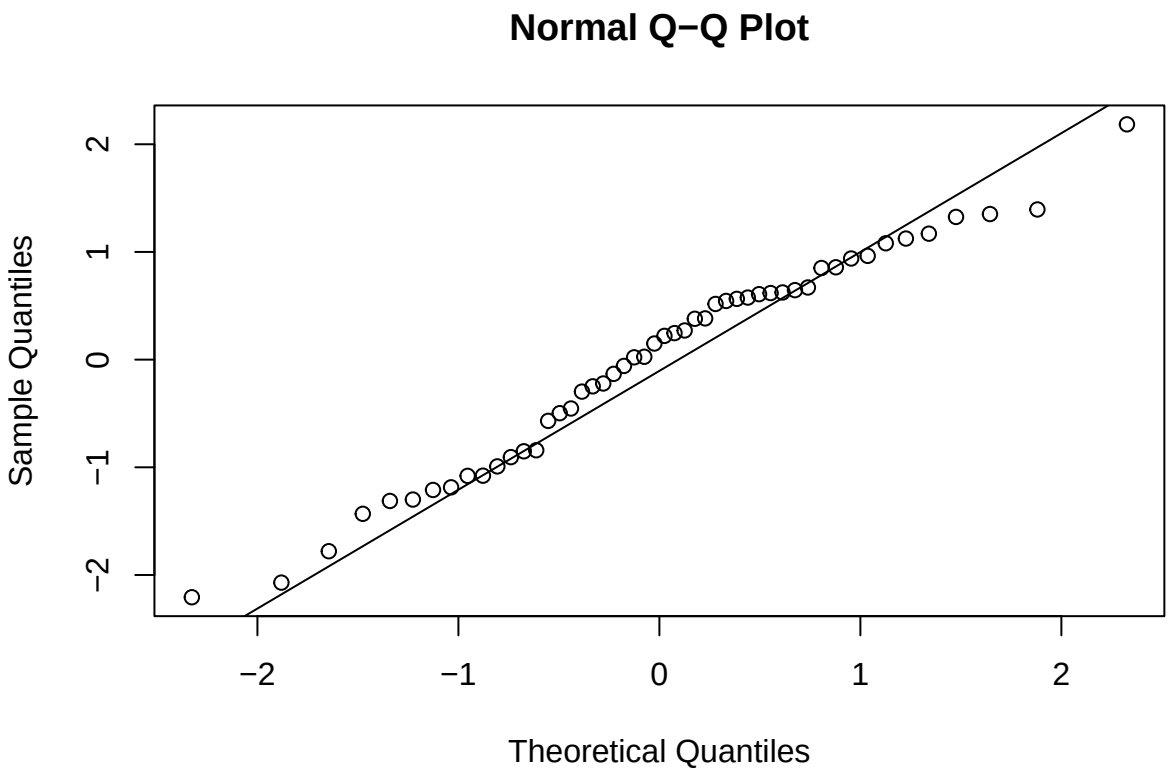

```
#not normally distributed
```

Q4. Below is a block of code that shows what a qq plot will look like for normally distributed data. Describbe what each line does in front of the comments, #.

```
#
set.seed(441)
n <- 50
(x <- rnorm(n, 0, 1))
```

```
##  [1]  1.32518470  0.54485211  2.18499532 -1.07788555 -1.07965590 -0.24804462
##  [7] -2.20678533  0.85087062 -1.77931074 -2.07085785  0.02138647 -1.43195727
## [13]  0.66979866  0.38296818 -1.18614319  0.24621160 -0.90565357  0.51690611
## [19]  1.08054090 -0.05858168 -0.49741949 -1.29949181  0.37924140  0.02641973
## [25]  0.27079952  0.57635343 -0.29708763  1.35203154  0.96351499  0.61813940
## [31]  0.93923917  1.39453759 -0.22158733  1.12445554  0.64517594  0.62395387
## [37] -1.21052784 -0.99170887 -1.31251128  1.16976839  0.14952998  0.85755173
## [43] -0.56887278  0.56427675 -0.84173713  0.22036298 -0.13225248 -0.45391433
## [49]  0.60774629 -0.85143628
```

```
qqnorm(x); qqline(x)
```



The pnorm(x) function gives $P(Z < x)$ for a normal distribution. The qnorm(p) function gives the x-value for a given percentile.

```
pnorm(1.96)
```

```
## [1] 0.9750021
```

```
qnorm(0.975)
```

```
## [1] 1.959964
```

Q5. Find the probability that a draw from a standard normal distribution is between -1.5 and 0.5.

```
pnorm(0.5) - pnorm(-1,5)
```

```
## [1] 0.6914625
```

Q6. Determine the z-values that contain the middle 60% of the standard normal distribution.

```
qnorm(0.2)
```

```
## [1] -0.8416212
```

```r
qnorm(0.8)
```

```
## [1] 0.8416212
```