

BACS-hw03-107070004

Question 1)

Let's reexamine how to standardize data: subtract the mean of a vector from all its values, and divide this difference by the standard deviation to get a vector of standardized values.

a) Create a normal distribution (mean=940, sd=190) and standardize it (let's call it `rnorm_std`)

```
# Standardized dataset
standardized <- function(x) {
  x <- (x-mean(x))/sd(x)
  return(x)
}

# create a normally distributed data set
rnorm <- rnorm(n=500, mean=940, sd=190)
rnorm_std <- standardized(rnorm)
```

i) What should we expect the mean and standard deviation of `rnorm_std` to be, and why?

We expect the mean is 0 and the standard deviation is 1.

$$E[Y] = E\left[\frac{X-E[x]}{sd(X)}\right] = \frac{E[X]-E[X]}{sd(X)} = 0$$

$$\begin{aligned} sd[Y] &= sd\left[\frac{X-E[x]}{sd(X)}\right] = \sqrt{E\left[\left(\frac{X-E[x]}{sd(X)} - E\left[\frac{X-E[x]}{sd(X)}\right]\right)^2\right]} = \sqrt{E\left[\left(\frac{X-E[x]}{sd(X)}\right)^2 - 2\left(\frac{X-E[x]}{sd(X)}\right)\left(E\left[\frac{X-E[x]}{sd(X)}\right]\right) + \left(E\left[\frac{X-E[x]}{sd(X)}\right]\right)^2\right]} \\ &= \sqrt{E\left[\left(\frac{X-E[x]}{sd(X)}\right)^2 - 2\left(\frac{X-E[x]}{sd(X)}\right)0 + 0^2\right]} \\ &= \frac{\sqrt{E[(X-E[X])^2]}}{sd[X]} = \frac{sd[X]}{sd[X]} = 1 \end{aligned}$$

Normalization typically means rescales the values into a range of [0,1]. Standardization typically means rescales data to have a mean of 0 and a standard deviation of 1 (unit variance).

```
# find its mean and std
mean(rnorm)
```

```
## [1] 958.8315
```

```
sd(rnorm)
```

```
## [1] 189.6496
```

```
mean(rnorm_std)
```

```
## [1] -5.234962e-17
```

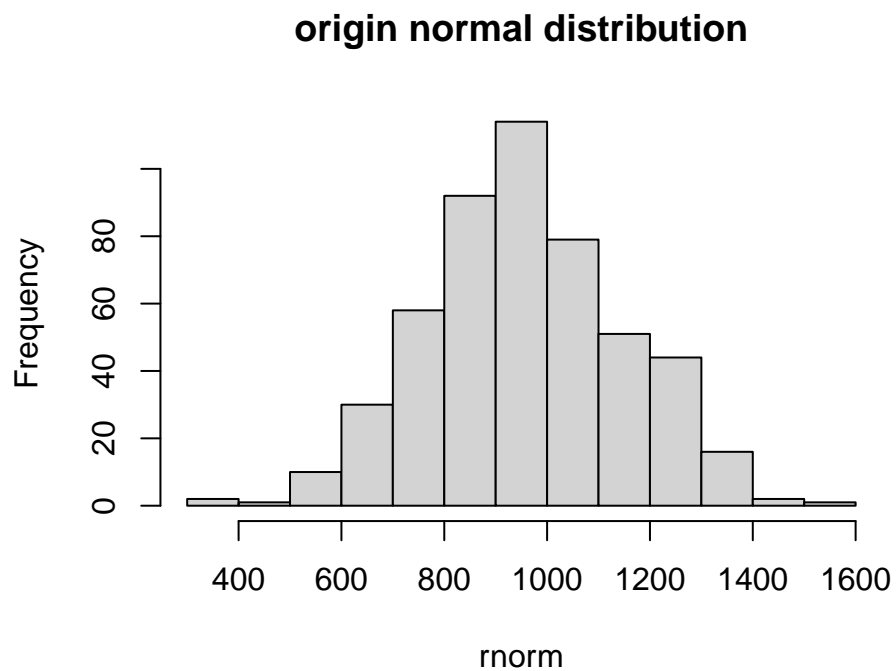
```
sd(rnorm_std)
```

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

ii) What should the distribution (shape) of `rnorm_std` look like, and why?

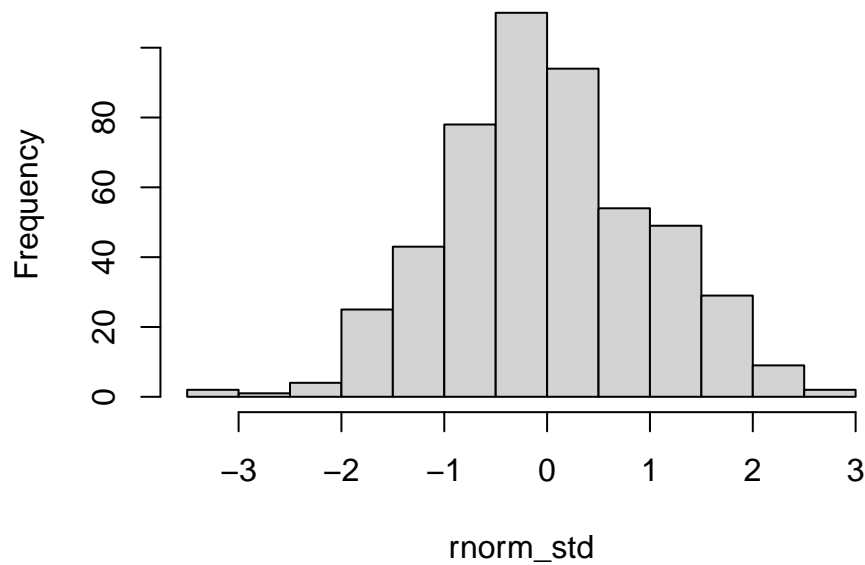
the value is rescaled and the mean and std will change. However, the curve will look the same.(explanation is beyond i))

```
# show the shape  
hist(rnorm, main = "origin normal distribution")
```



```
hist(rnorm_std, main = "standardized normal distribution")
```

standardized normal distribution



iii) What do we generally call distributions that are normal and standardized?

Standard Normal Distribution

b) Create a standardized version of minday discussed in question 3 (let's call it minday_std)

```
minday_std <- standardized(minday)
```

i) What should we expect the mean and standard deviation of minday_std to be, and why?

same as a)i)

```
# find its mean and std  
mean(minday_std)
```

```
## [1] -4.25589e-17
```

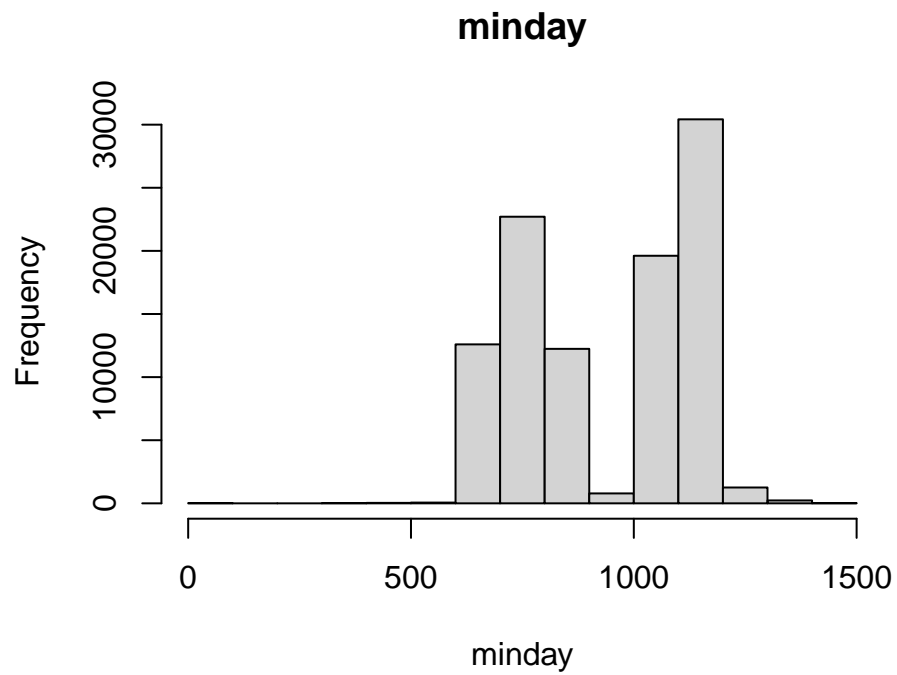
```
sd(minday_std)
```

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

ii) What should the distribution of `minday_std` look like compared to `minday`, and why?

same as a)ii)

```
# Show the hist  
hist(minday,main = "minday")
```



```
hist(minday_std,main = "minday_std")
```

