

hw1

danie

2025-01-21

```
knitr::opts_chunk$set(fig.width=5, fig.height=3.5, fig.align = "center")
```

1

i

Using:

```
x <- c(2.7, 4.0, 2.3, 5.4, -5.3, 1.8, -1.3, -2.9, 2.1, 3.9,  
      -1.8, 0.4, -4.2, 0.5, -0.1, 1.5, -0.7)  
y <- c(1.4, 2.5, 2.6, 5.6, -2.2, 0.4, 0.1, -3.0, 2.2, 0.9 ,  
      -2.4, 1.6, -2.5, 0.1, -9.9, 1.1, -1.7)
```

Five-number summary of x :

```
summary(x)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
## -5.3000 -1.3000  0.5000  0.4882  2.3000  5.4000
```

Five-number summary of y :

```
summary(y)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
## -9.9000 -2.2000  0.4000 -0.1882  1.6000  5.6000
```

Sample variance of x :

```
var(x)
```

```
## [1] 8.673603
```

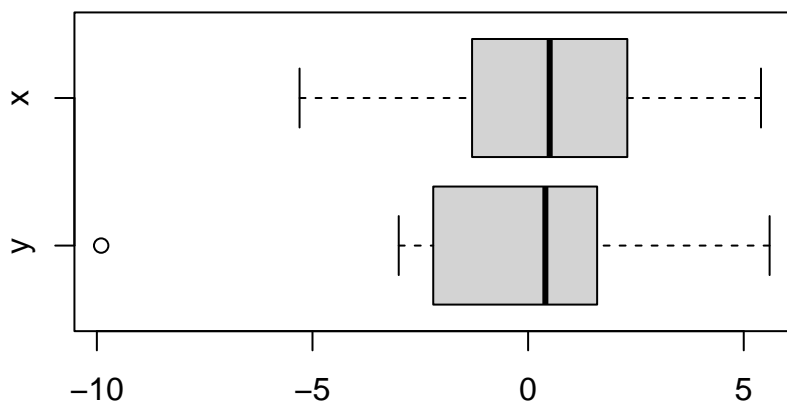
Sample variance of y :

```
var(y)
```

```
## [1] 11.37985
```

```
boxplot(y,x,  
  main = "Distribution of xs and ys",  
  names = c("y", "x"),  
  horizontal = TRUE  
)
```

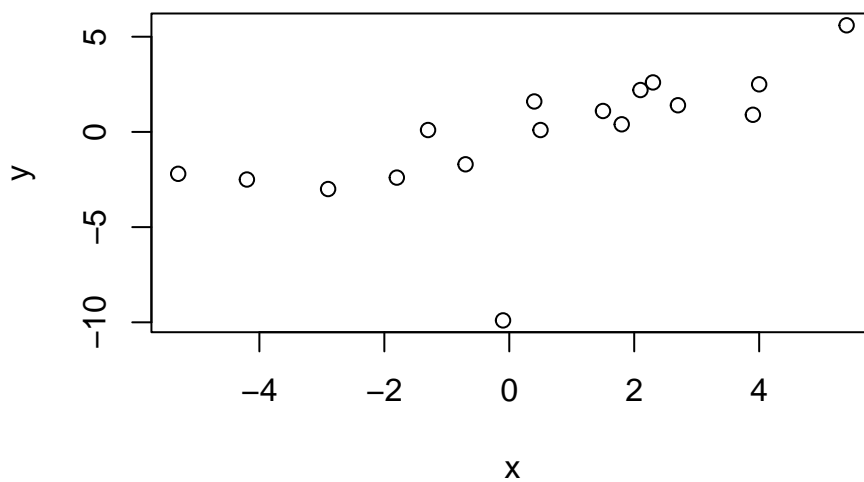
Distribution of x s and y s



The x s are skewed negative/left while the y s have no skew. The outlier is the y -value -9.9 from the point $(-0.1, -9.9)$.

ii

```
plot(x, y)
```



Correlation coefficient:

```
cor(x, y)
```

```
## [1] 0.6289777
```

Which means x and y are moderately linearly correlated.

iii

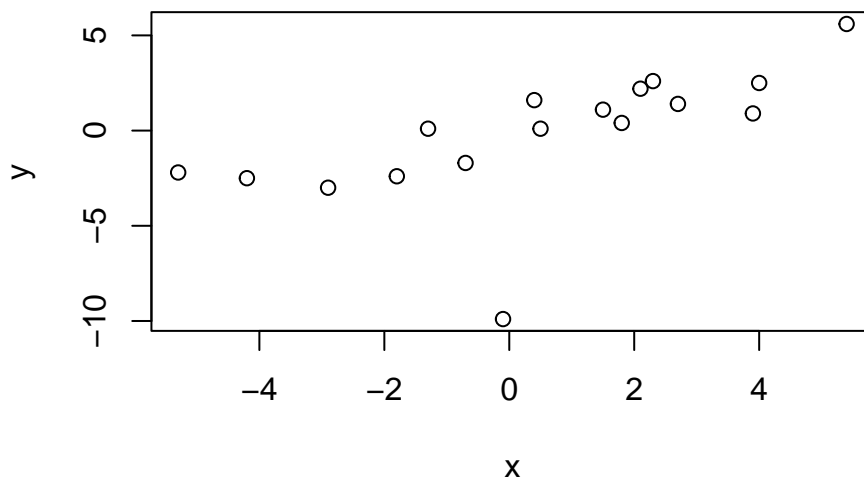
Yes; $(-0.1, -9.9)$ is an outlier.

```
x2 <- c(2.7, 4.0, 2.3, 5.4, -5.3, 1.8, -1.3, -2.9, 2.1, 3.9,  
        -1.8, 0.4, -4.2, 0.5, 1.5, -0.7)  
y2 <- c(1.4, 2.5, 2.6, 5.6, -2.2, 0.4, 0.1, -3.0, 2.2, 0.9,  
        -2.4, 1.6, -2.5, 0.1, 1.1, -1.7)  
cor(x2, y2)
```

```
## [1] 0.8822511
```

iv

```
plot(x, y)
```



I can see the outlier $(-0.1, -9.9)$ at the bottom-center of the graph.

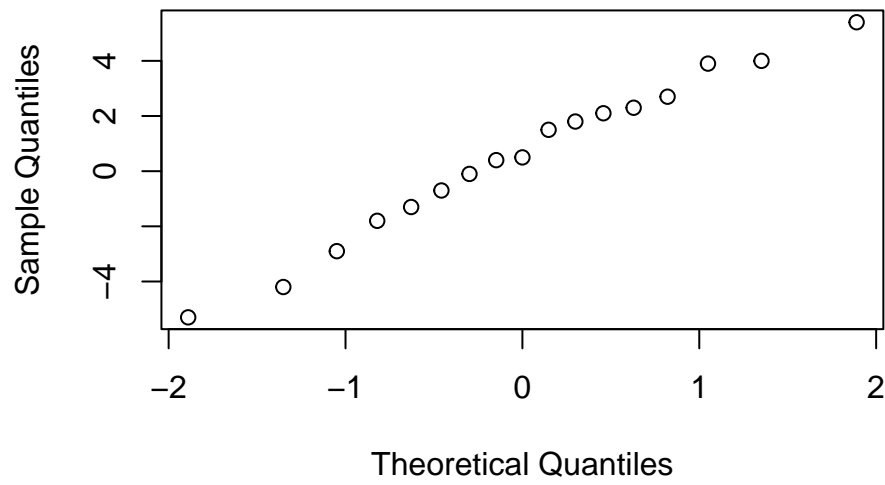
v

The sample correlation coefficient in iii is much higher than the one in ii.

vi

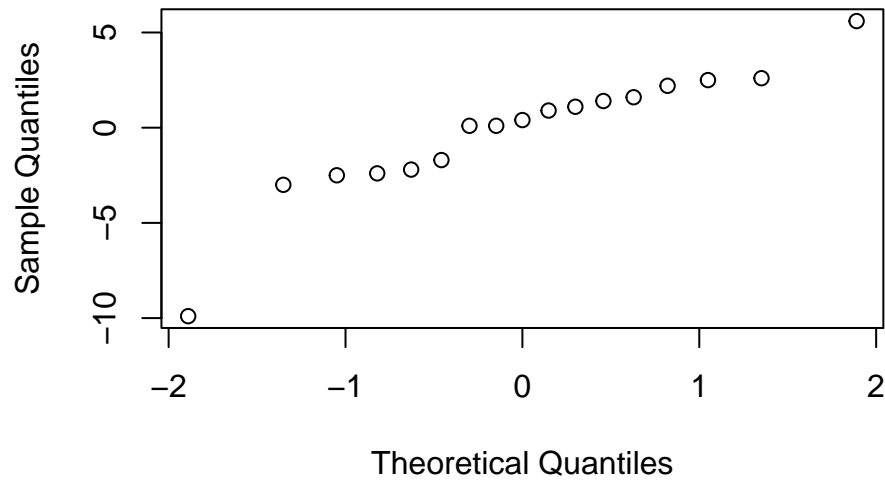
```
qqnorm(x, main="xs with outlier")
```

xs with outlier



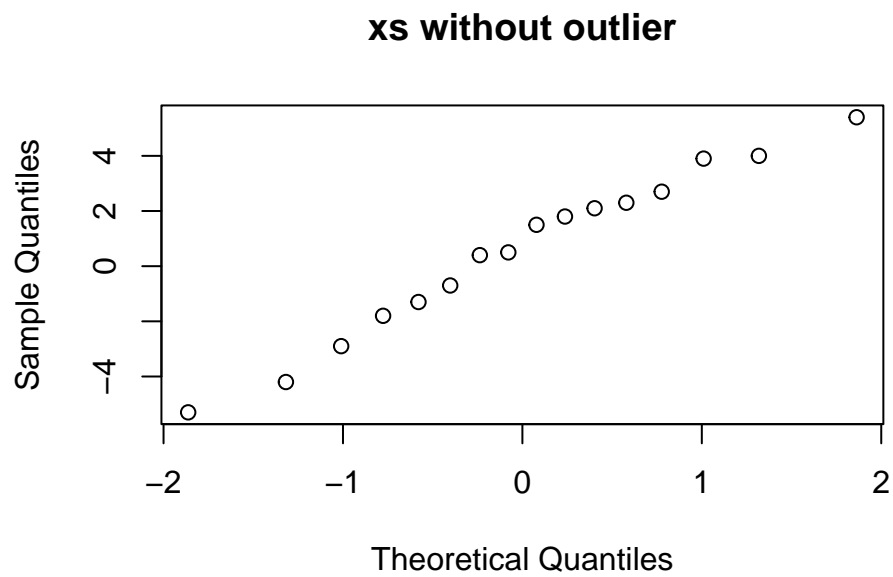
```
qqnorm(y, main="ys with outlier")
```

ys with outlier

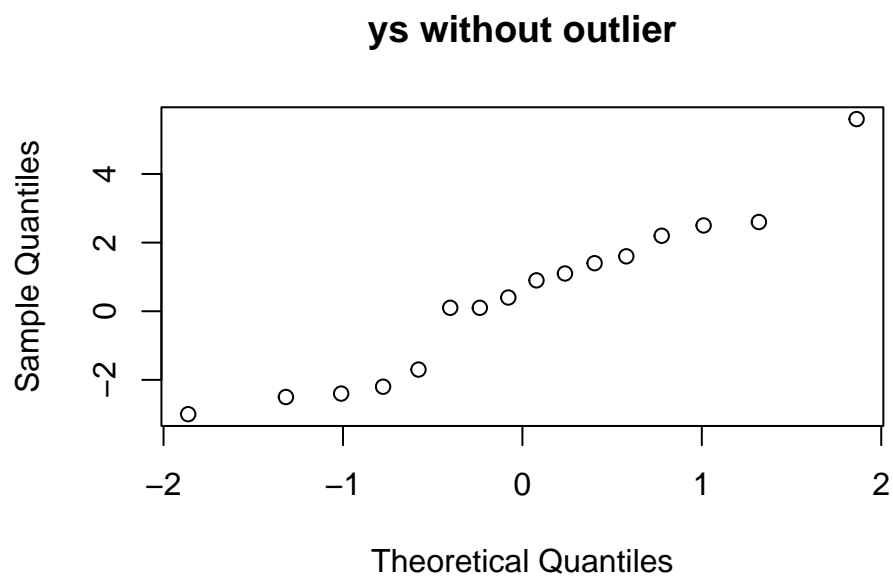


The *xs* look much closer to normal distribution than the *ys*.

```
qqnorm(x2, main="xs without outlier")
```



```
qqnorm(y2, main="ys without outlier")
```



The *xs* still look much closer to normal distribution than the *ys*.