

Example:

In 2019, the average health care annual premium for a family of 4 in the United States, was reported to be \$6,015.

[The Kaiser Family Foundation, “Employer Health Benefits 2019 Annual Survey”]

In a more recent survey, 100 randomly sampled families of 4 reported an average annual health care premium of \$6,537.

Can we say that the true average is currently greater than \$6,015 for all families of 4?

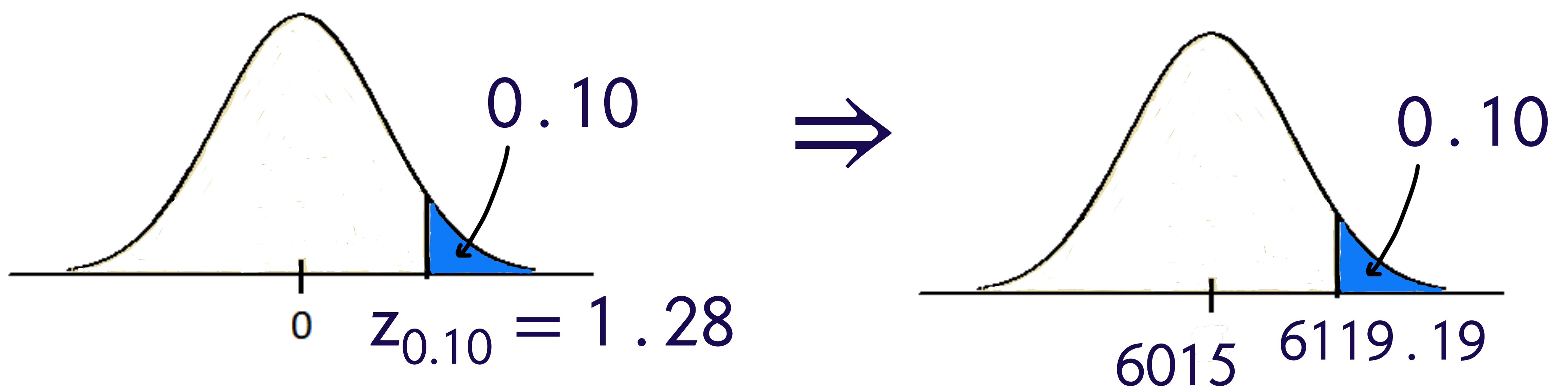
$(\sigma = 814)$

Our size 0.10 test said to

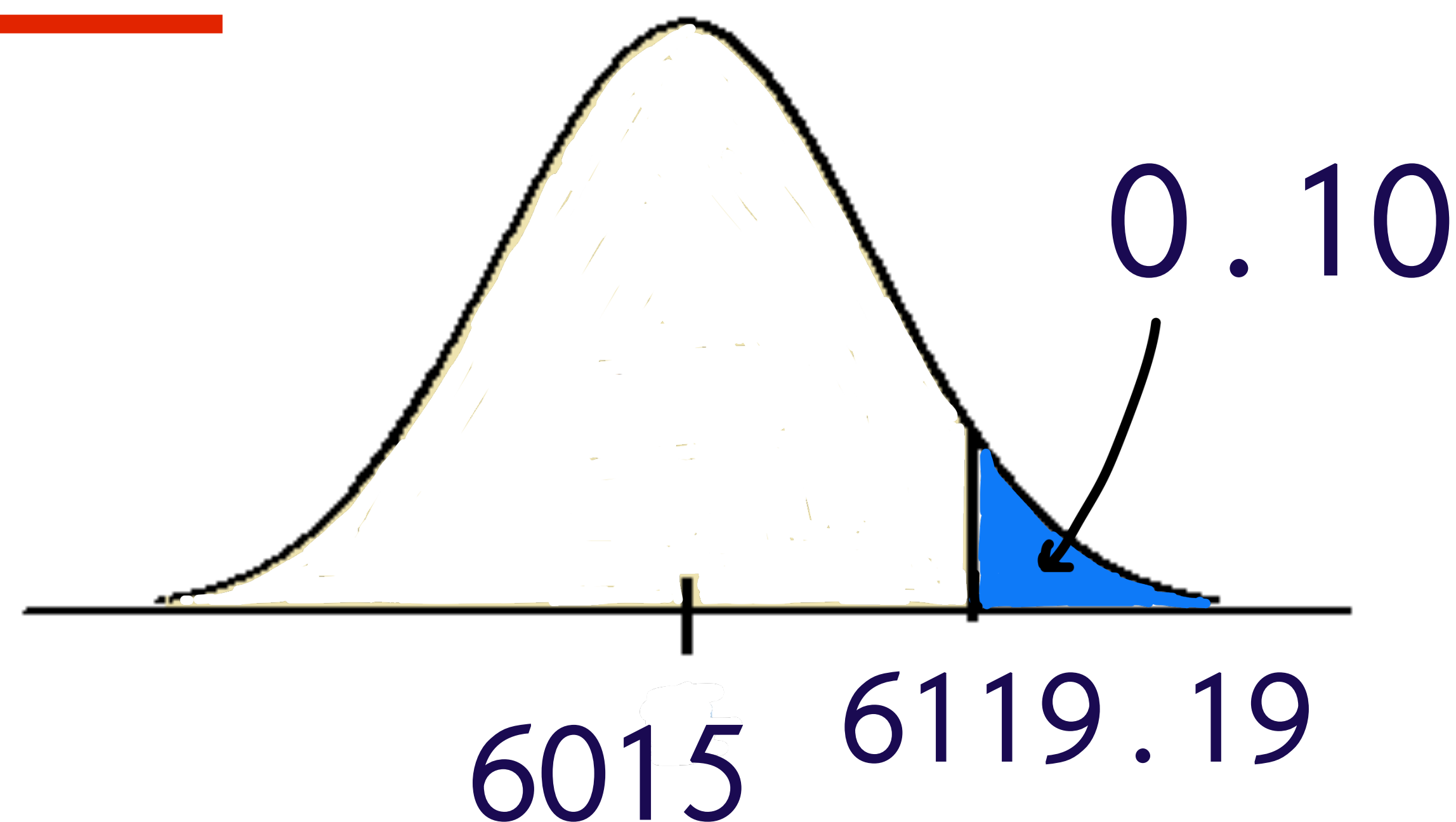
Reject H_0 if $\bar{X} > 6119.19$

Our observed sample mean was

$$\bar{x} = 6537$$

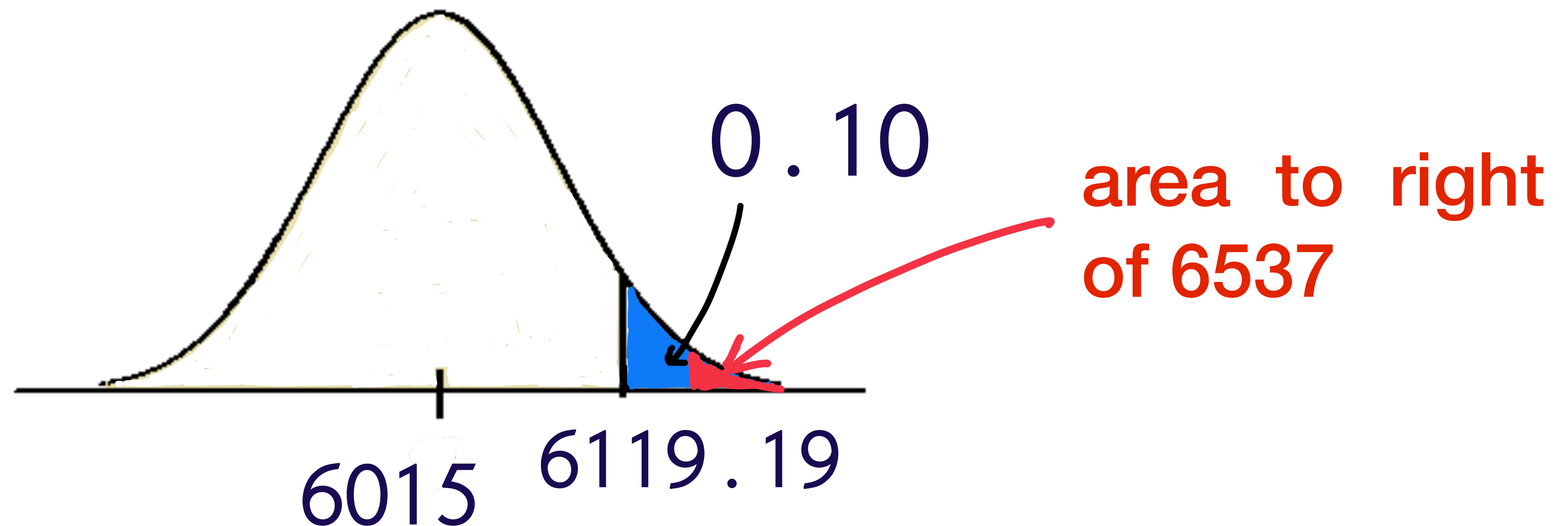


The P-Value



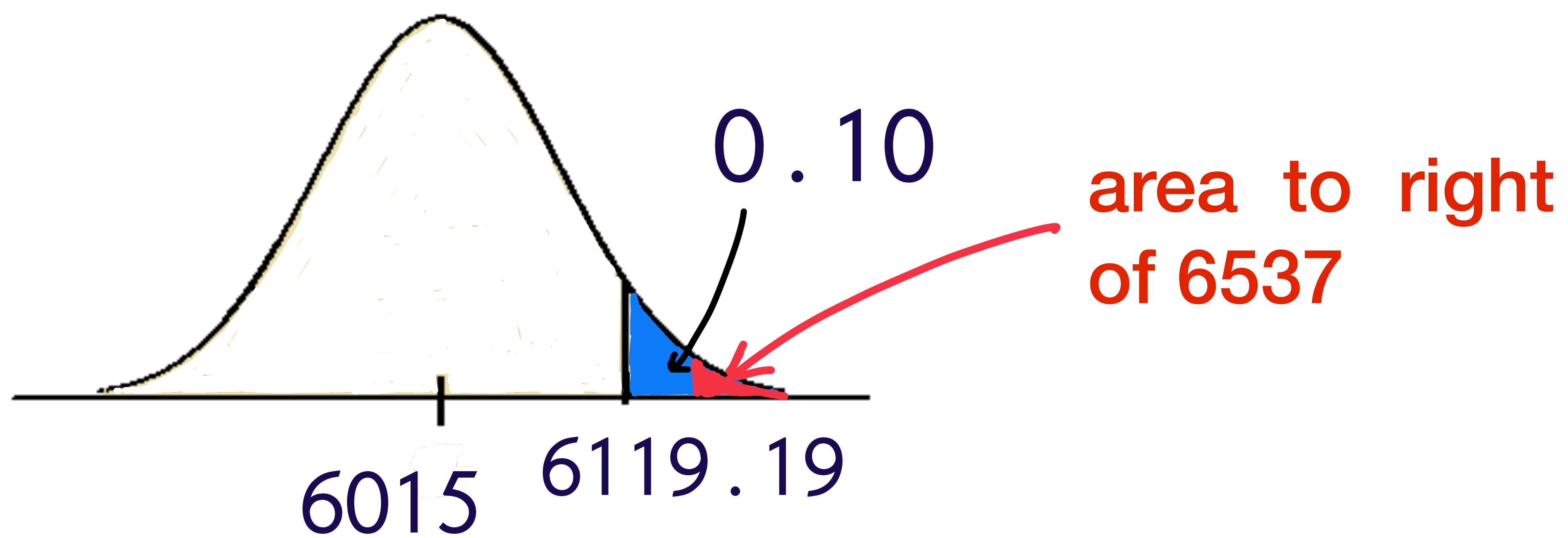
- Our sample mean (6537) fell into the **rejection region**, so we rejected H_0 .
- Note then that the area to the right of our sample mean of 6537 must be less than 0.10.

The P-Value



- Our sample mean (6537) fell into the **rejection region**, so we rejected H_0 .
- Note then that the area to the right of our sample mean of 6537 must be less than 0.10.

The 0.10 is a probability that is relevant when H_0 is true.



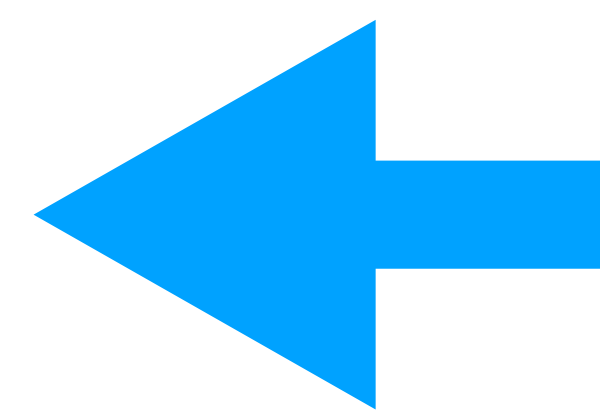
- This is the $N(6015, 814^2/100)$ pdf.
- The red area is $P(\bar{X} > 6537)$.

$$P(\bar{X} > 6537)$$

$$= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{6537 - 6015}{814/\sqrt{100}}\right)$$

$$= P(Z > 6.4127)$$

$$\approx 0.000000001$$



Super small
and way out
“in the tail”.

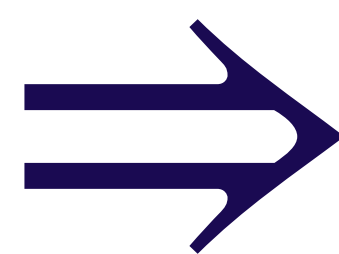
The P-Value

The P-Value is the area to the right (in this case) of the test statistic \bar{X} .

- The P-value being less than 0.10 puts \bar{X} in the rejection region.
- The P-value is also less than 0.05 and 0.01.
- It looks like we will reject H_0 for the most typical values of α .

The P-Value

small
P-Value



reject
 H_0



Checking for Normality

Consider the following “data”:

1.678, 2.024, 2.168, 3.018, 1.689,
1.727, 1.743, 3.234, 2.008, 1.309

Now in order:

1.309, 1.678, 1.689, 1.727, 1.743,
2.008, 2.024, 2.168, 3.018, 3.234

Checking for Normality

1.309, 1.678, 1.689, 1.727, 1.743,
2.008, 2.024, 2.168, 3.018, 3.234

- Note that 10% of the data are at or below 1.309.
- 20% of the data are at or below 1.678.

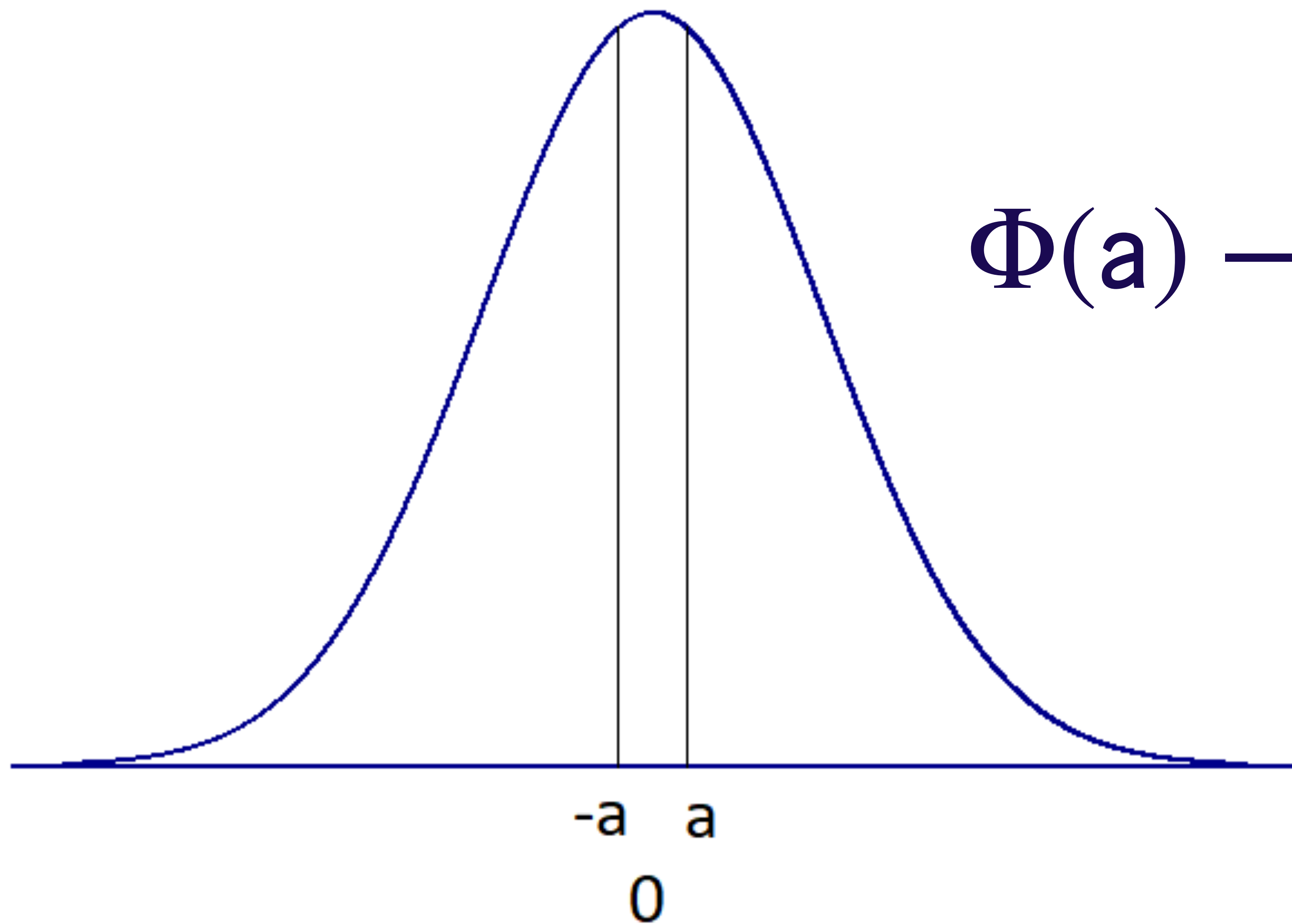
Numbers that divide data into equal groups like this are called **quantiles**.

Checking for Normality

These numbers taken from the sample are **sample quantiles**.

We compare them with numbers that divide the area under the normal curve into 10 equal parts.

Checking for Normality



$$\Phi(a) - \Phi(-a) = 0.10$$

$$\Rightarrow \Phi(a) = 0.55$$

$$\Rightarrow a = \Phi^{-1}(0.55)$$

$$\text{qnorm}(0.55)=0.1257$$

Checking for Normality

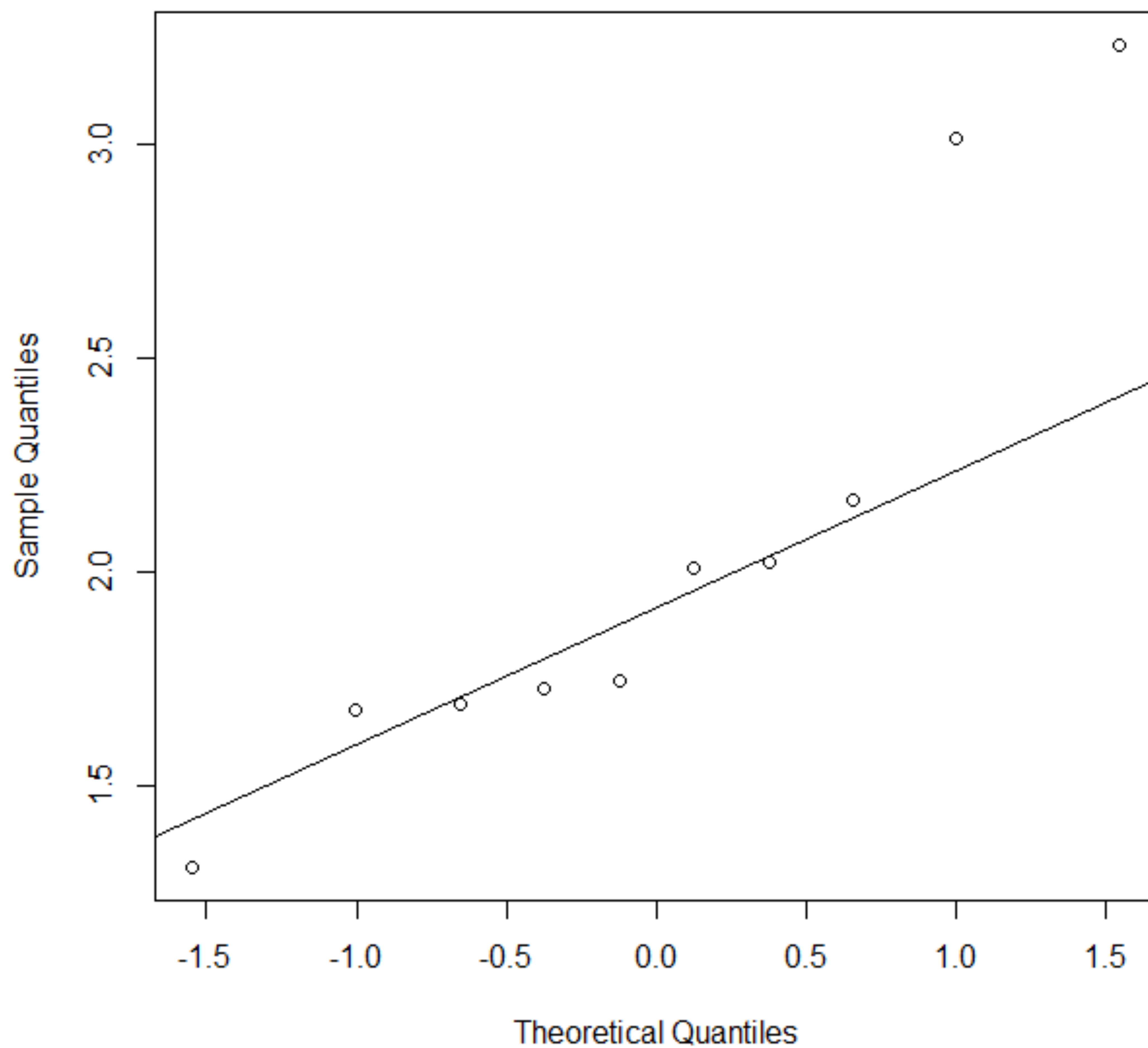
Let's have R compare all the quantiles for us.

```
mysample<-c(1.678, 2.024, 2.168, 3.018,  
1.689, 1.727, 1.743, 3.234, 2.008, 1.309)
```

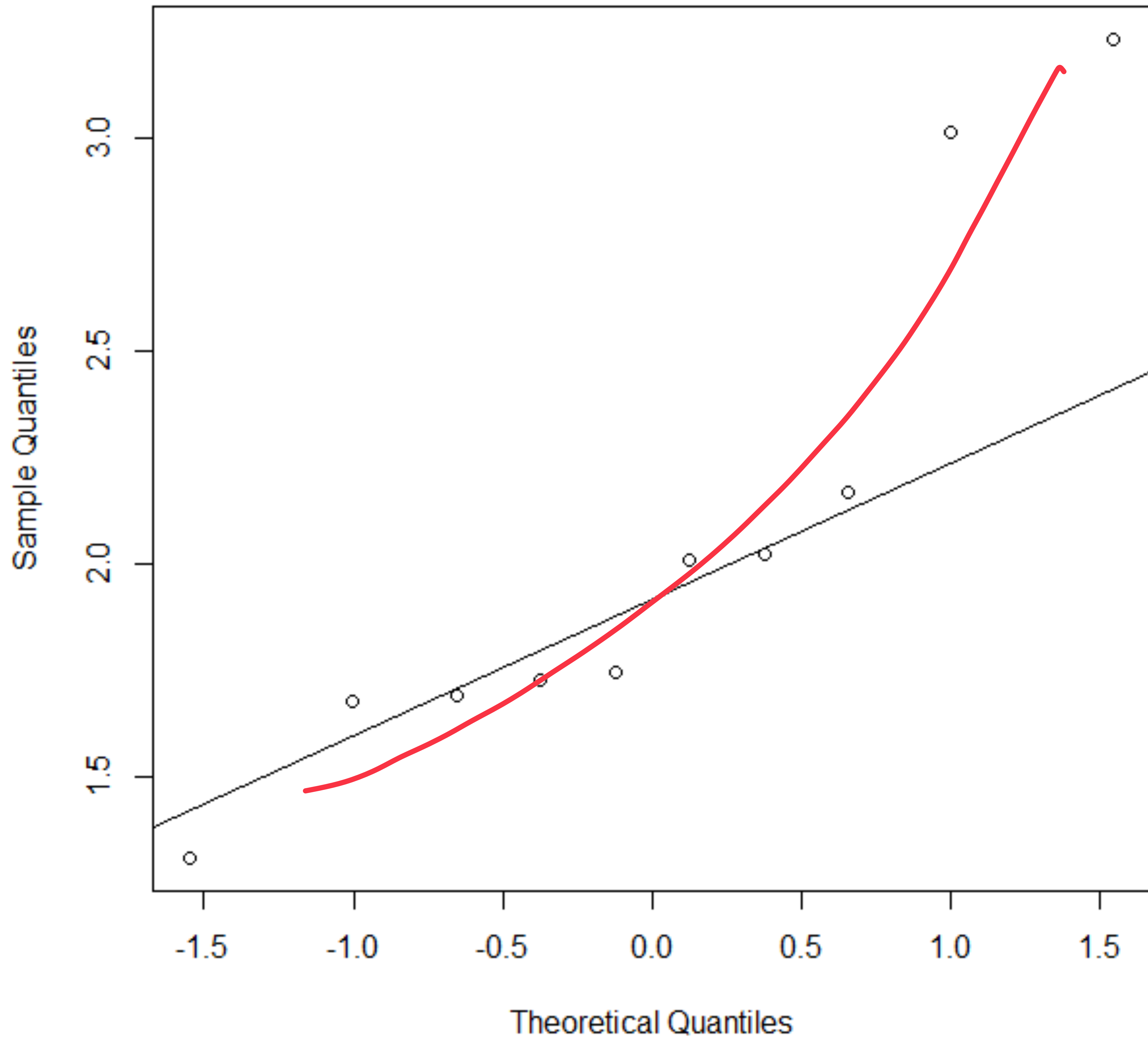
```
qqnorm(mysample)
```

```
qqline(mysample)
```

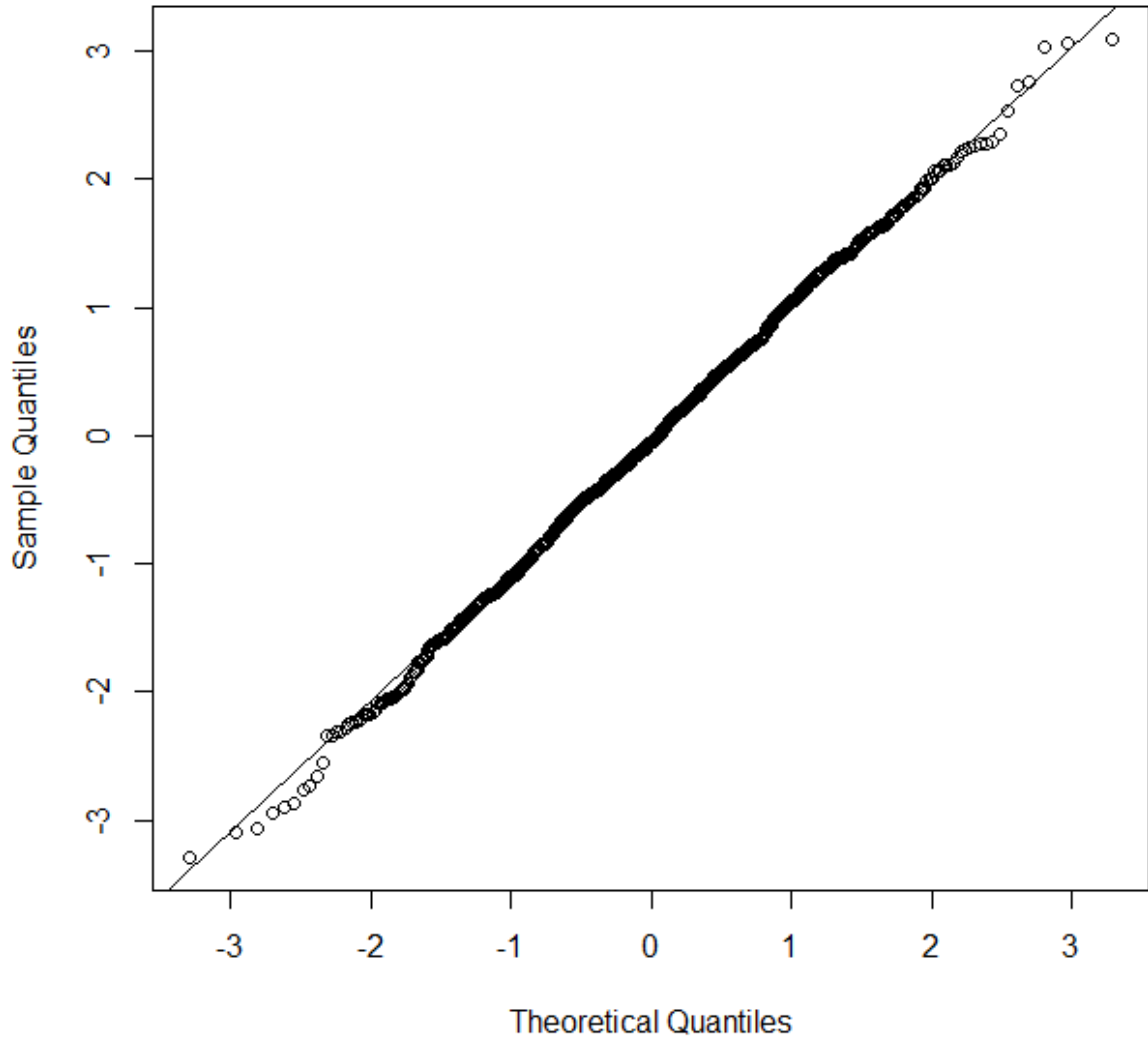
Normal Q-Q Plot



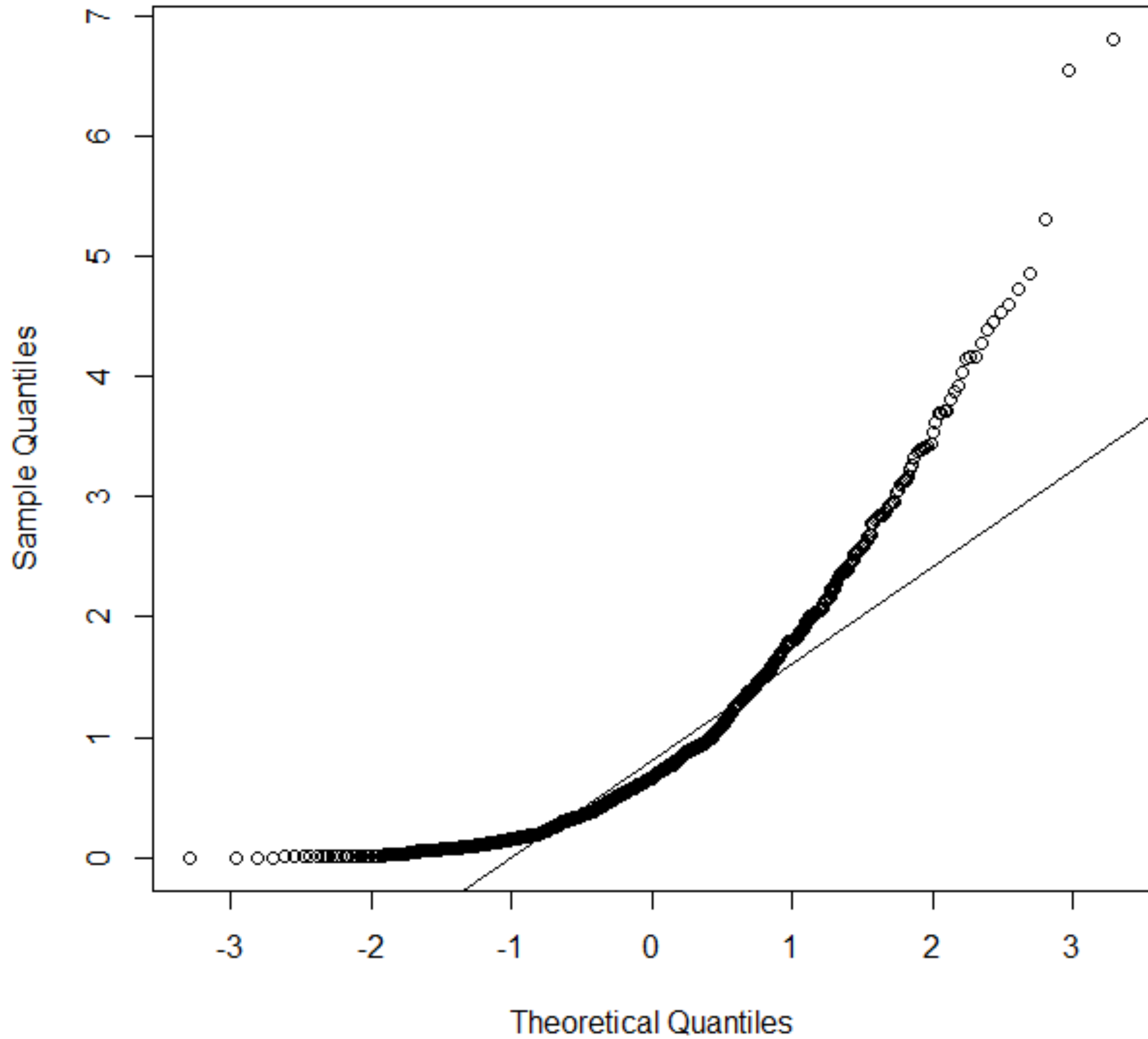
Normal Q-Q Plot

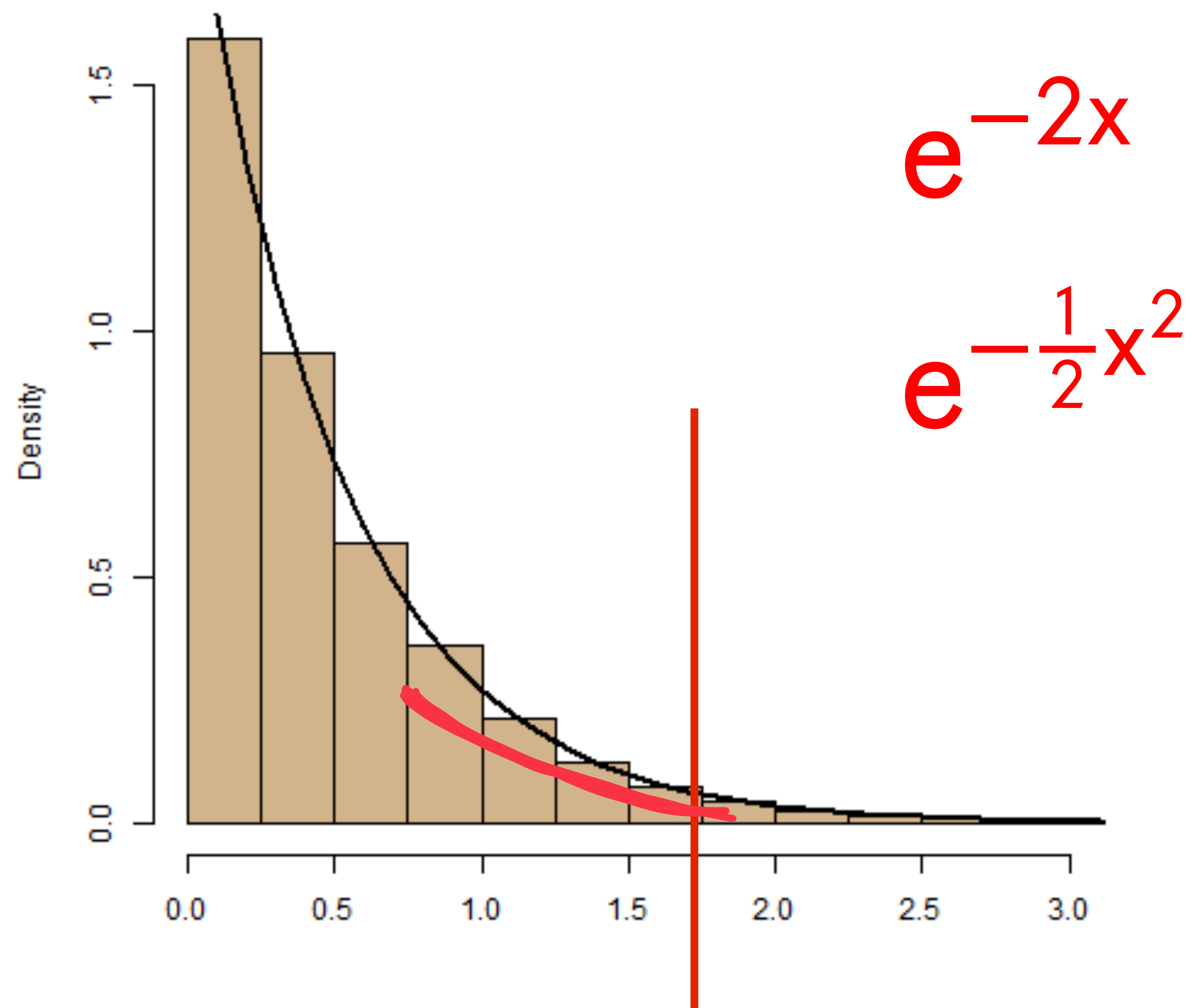


Normal Q-Q Plot

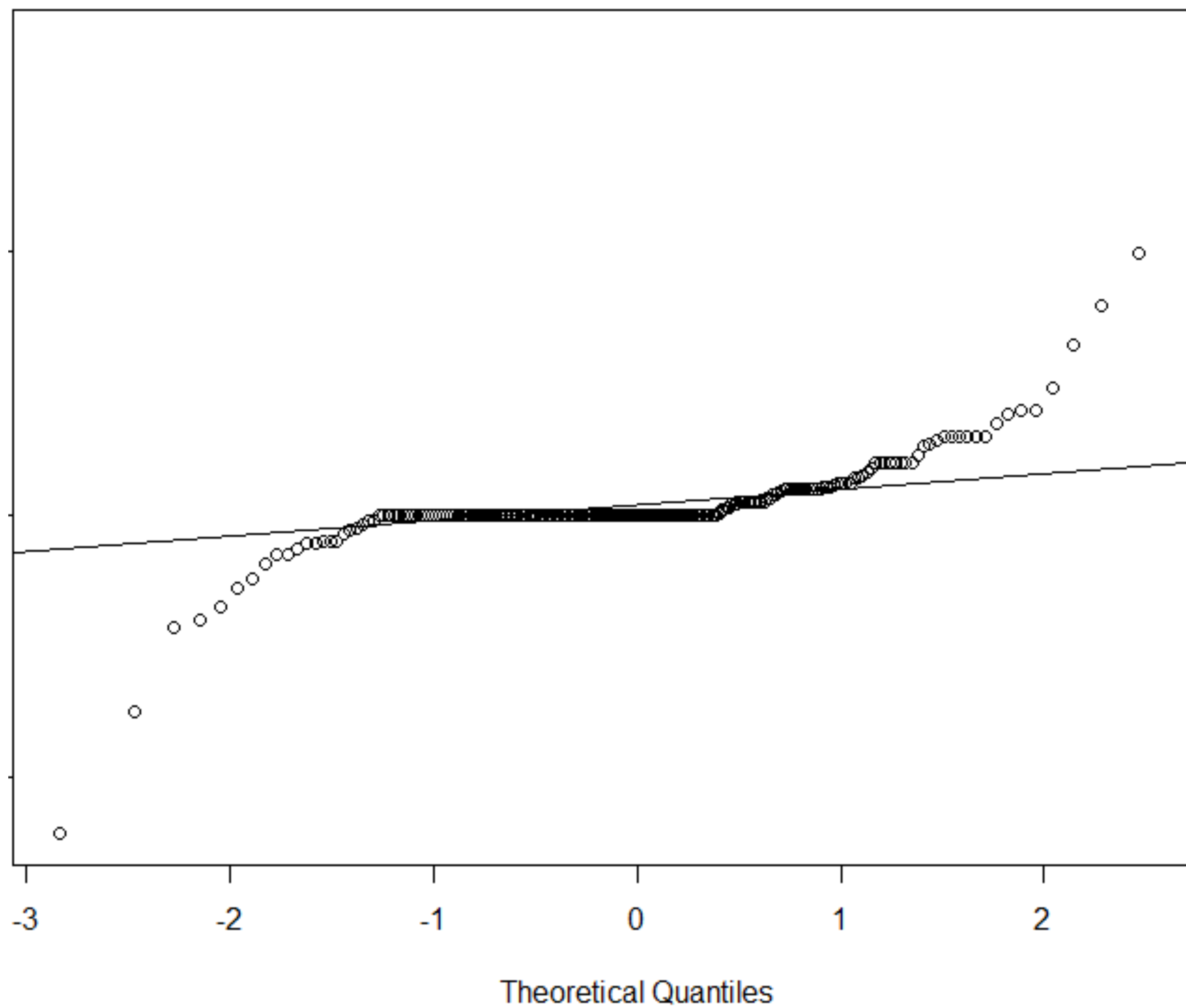


Normal Q-Q Plot





Q-Q Plot for Dose Changes Post Op to Discharge



It takes practice to read **qq-plots**.

You can get some practice by going in to R, simulating some normal samples and some not normal samples of different sizes and looking at the qq-plot.

```
mysample<-rnorm(100,2,1)
```

```
qqnorm(mysample)
```

```
qqline(mysample)
```

After doing this many times, try again with a random sample from the exponential distribution.

What do you see?