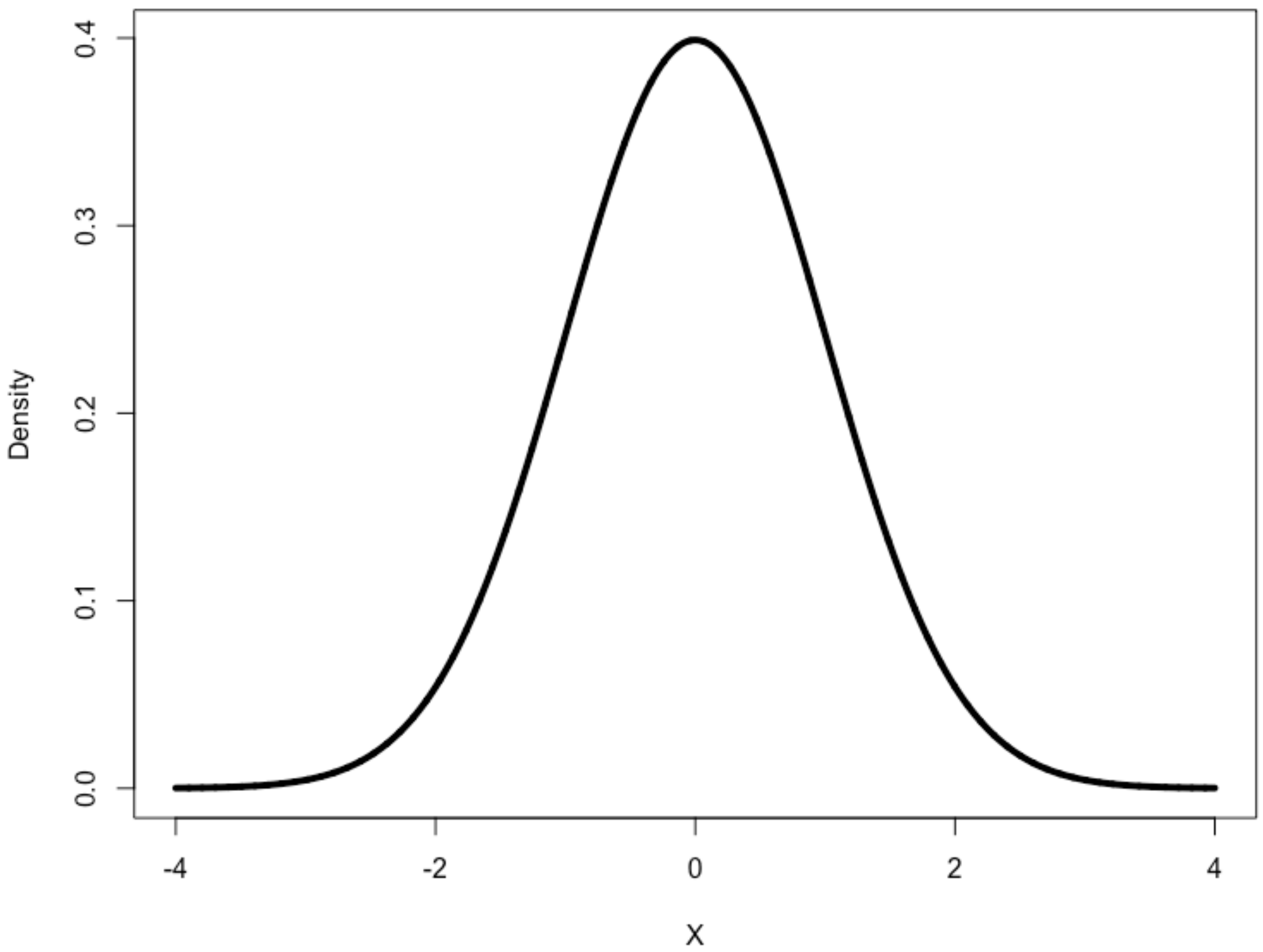
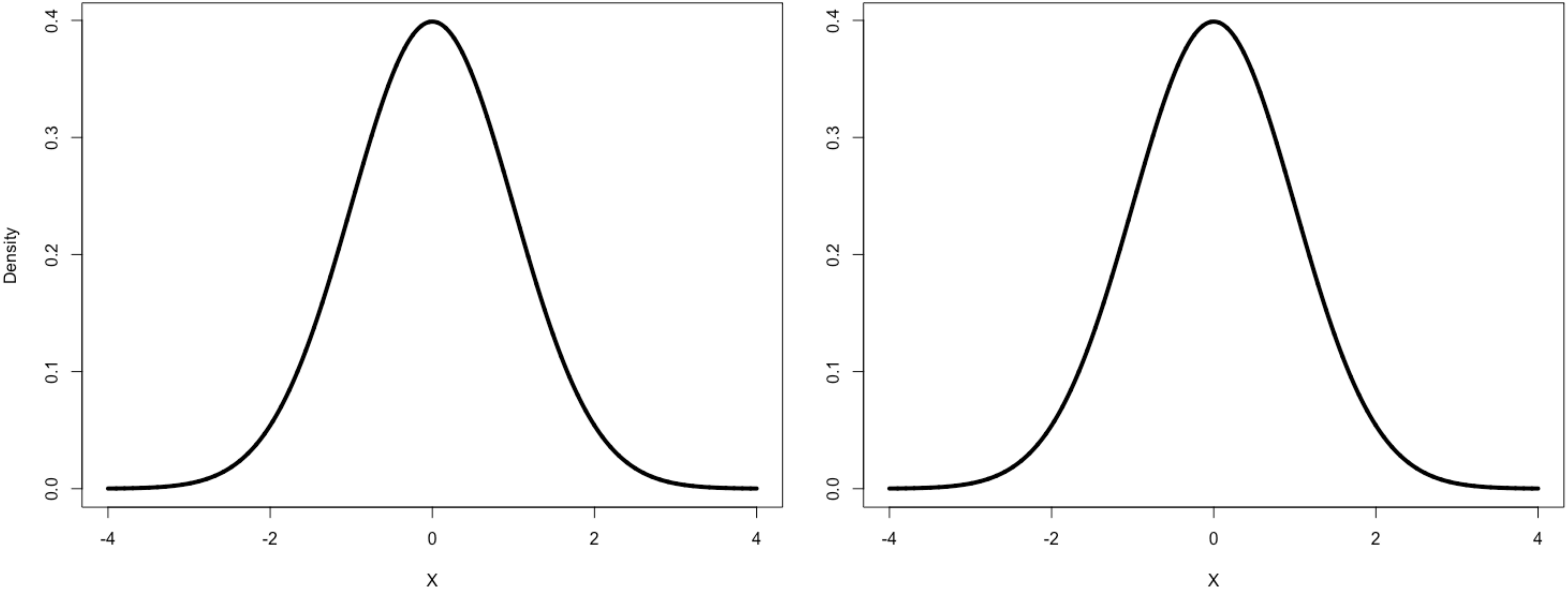
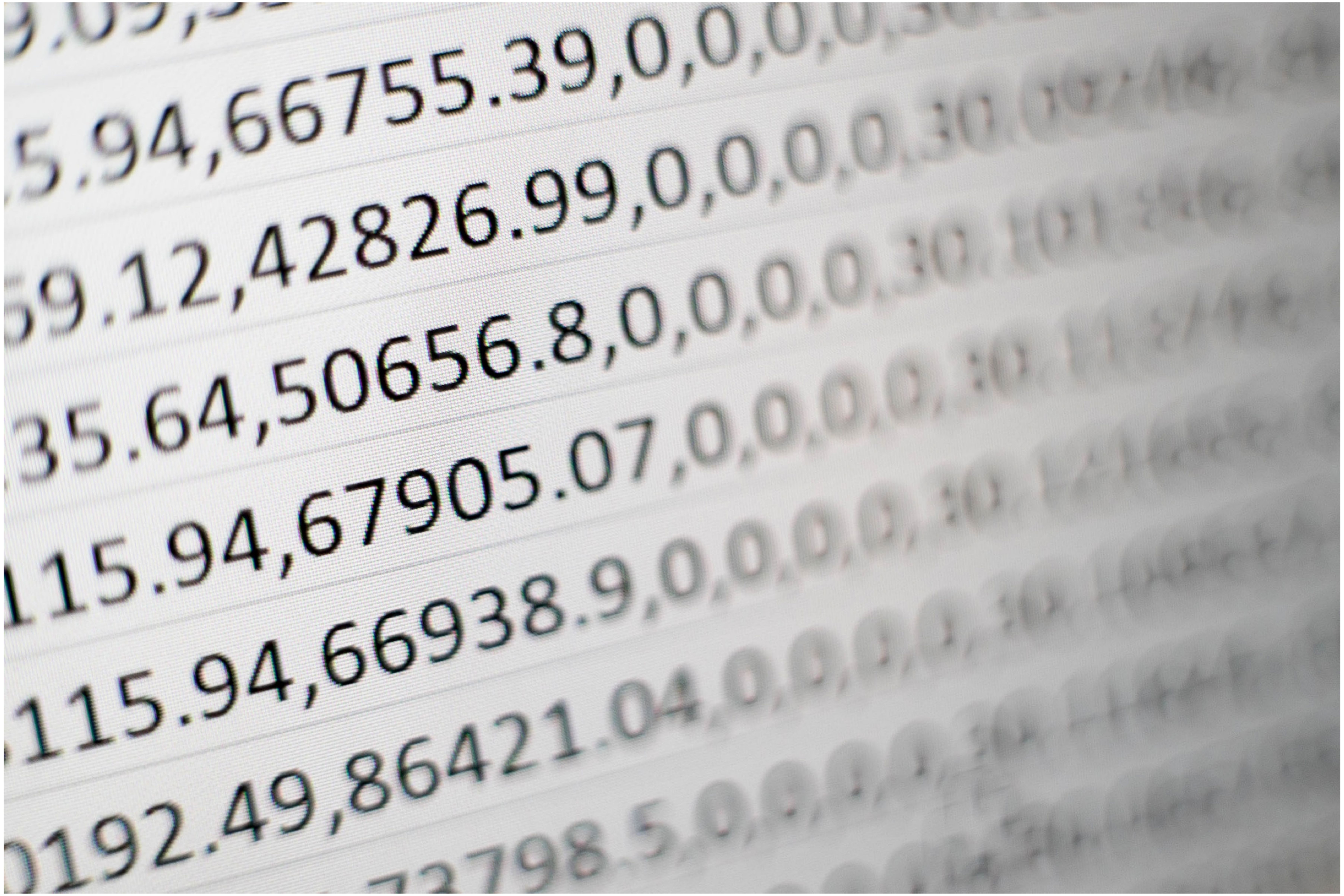
Lesson 020
Probability Plots

Friday, October 27


0.4
0.3
0.2
0.1
0.0
Density
-4
-2
0
2
4
X


Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
X
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
X


5.94,66755.39,0,0,0,0
59.12,42826.99,0,0,0,0
35.64,50656.8,0,0,0,0
115.94,67905.07,0,0,0
115.94,66938.9,0,0
192.49,86421.04

Probability Plots

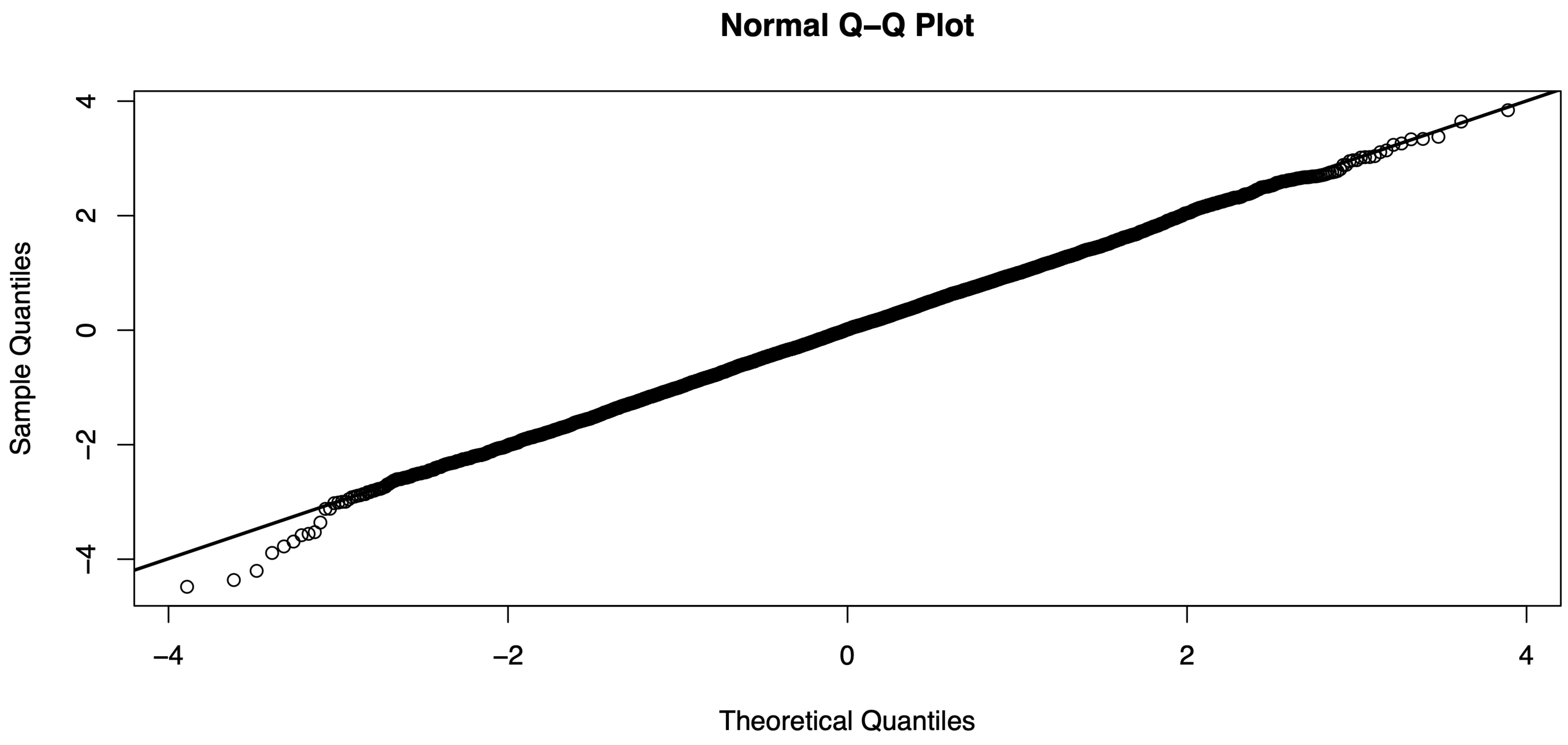
- To test whether a particular distribution is followed, we should plot our data.
- Histograms are likely to be biased or challenging to draw accurate conclusions with.
- Instead, we rely on **probability plots.**
    - **Note:** these are more commonly called QQ-plots outside of this course.

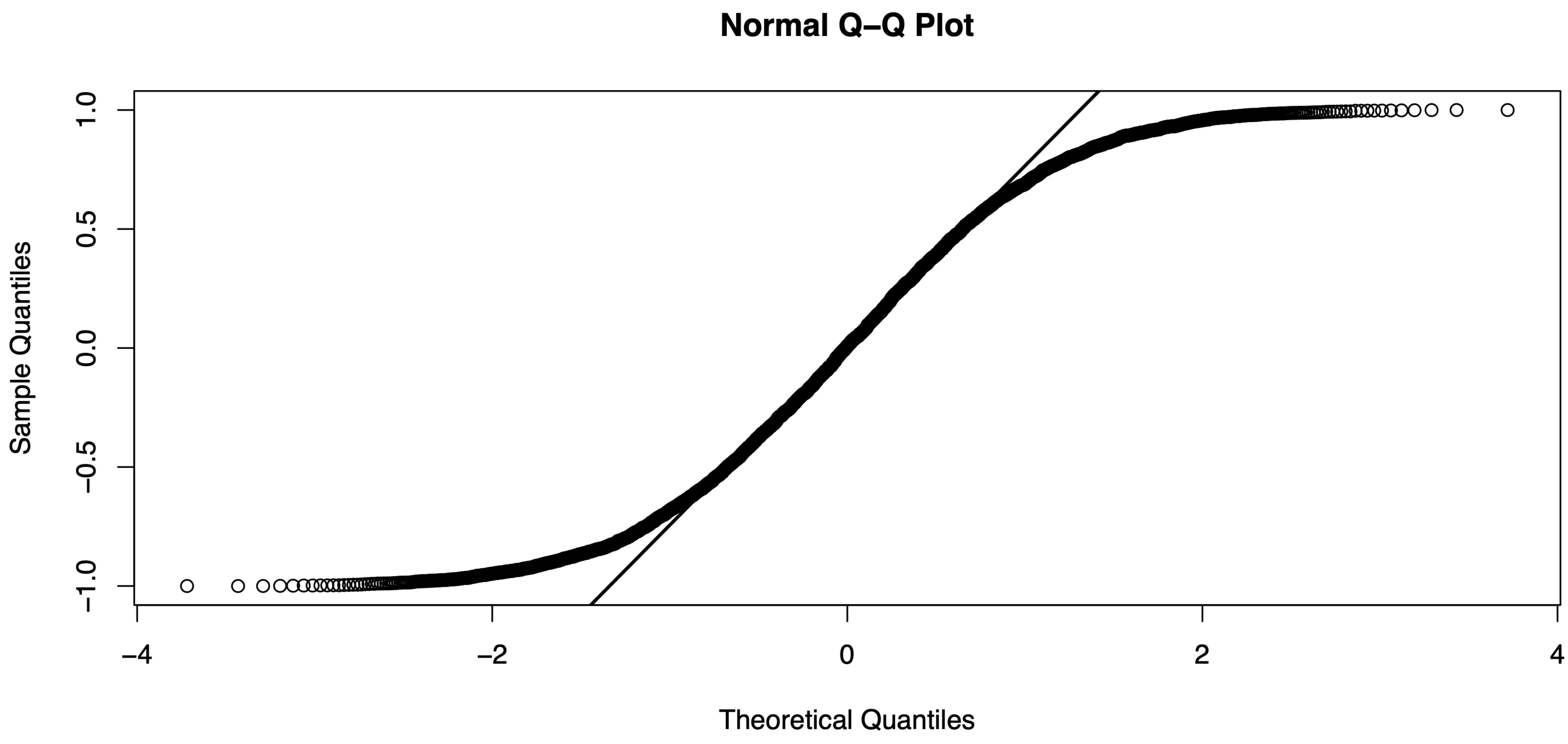
Using Quantiles to Define a Distribution

- Suppose that data are drawn from a particular distribution, with quantiles $\eta(p)$.
  - In the sample we expect that:
    - Median $\approx \eta(0.5)$
    - $Q1 \approx \eta(0.25)$
    - $Q3 \approx \eta(0.75)$
  - In general, all sample quantiles should approximately equal the corresponding theoretical values.

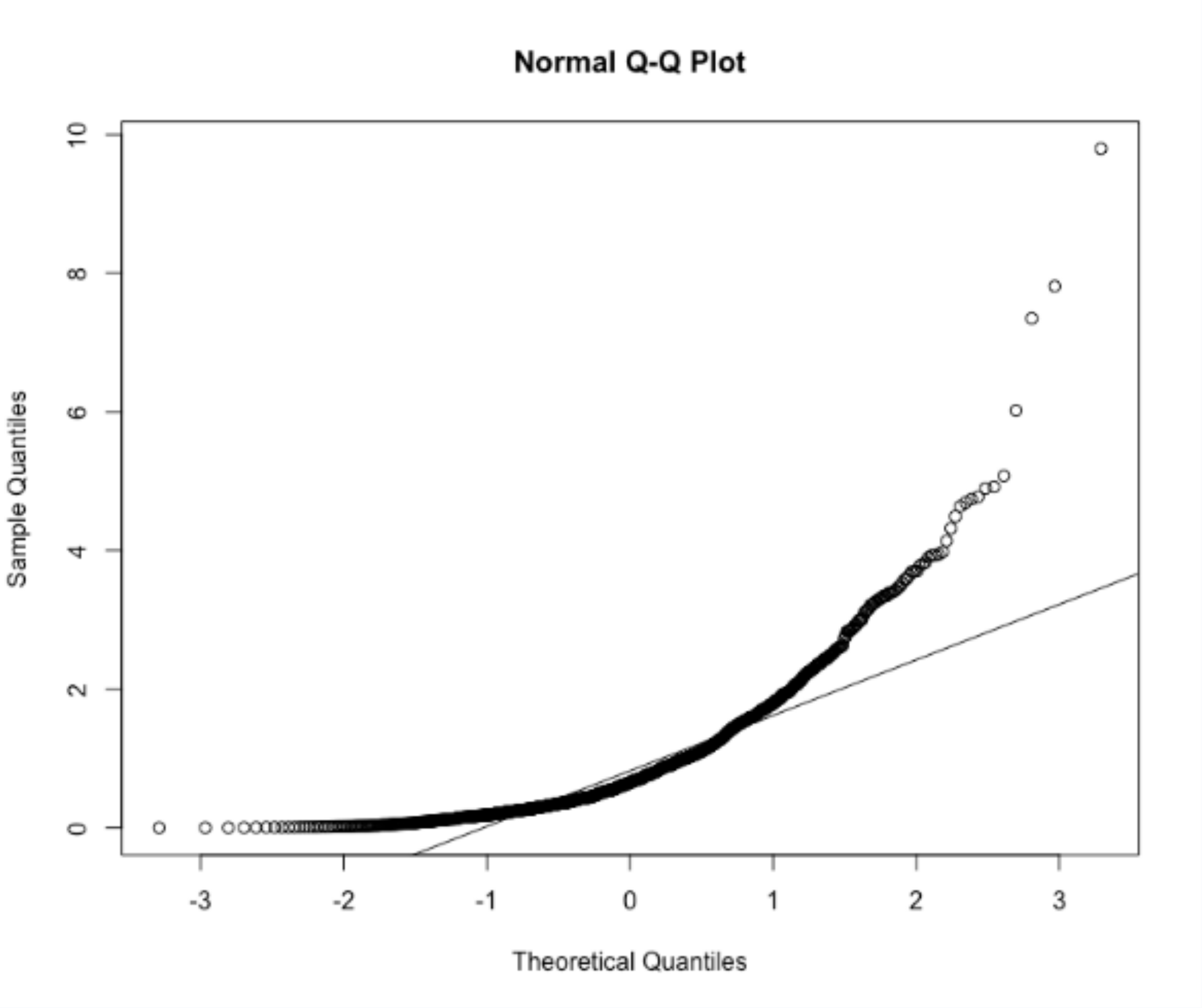
Plotting Sample versus Theoretical Quantiles

- If we call the sample quantiles $\tilde{\eta}(p)$ then imagine we make a plot of $\tilde{\eta}(p)$ versus $\eta(p)$.
- If the distribution is correct, we expect $\tilde{\eta}(p) \approx \eta(p)$
- In graphical terms, this corresponds to $y \approx x$.
- If the plot ends up with an approximately straight line, evidence of a good fit.


Normal Q–Q Plot
Sample Quantiles
Theoretical Quantiles
4
2
0
−2
−4
−4
−2
0
2
4


Normal Q–Q Plot
Sample Quantiles
Theoretical Quantiles
1.0
0.5
0.0
−0.5
−1.0
−4
−2
0
2
4

True or False: The following probability plot displays evidence of normality.



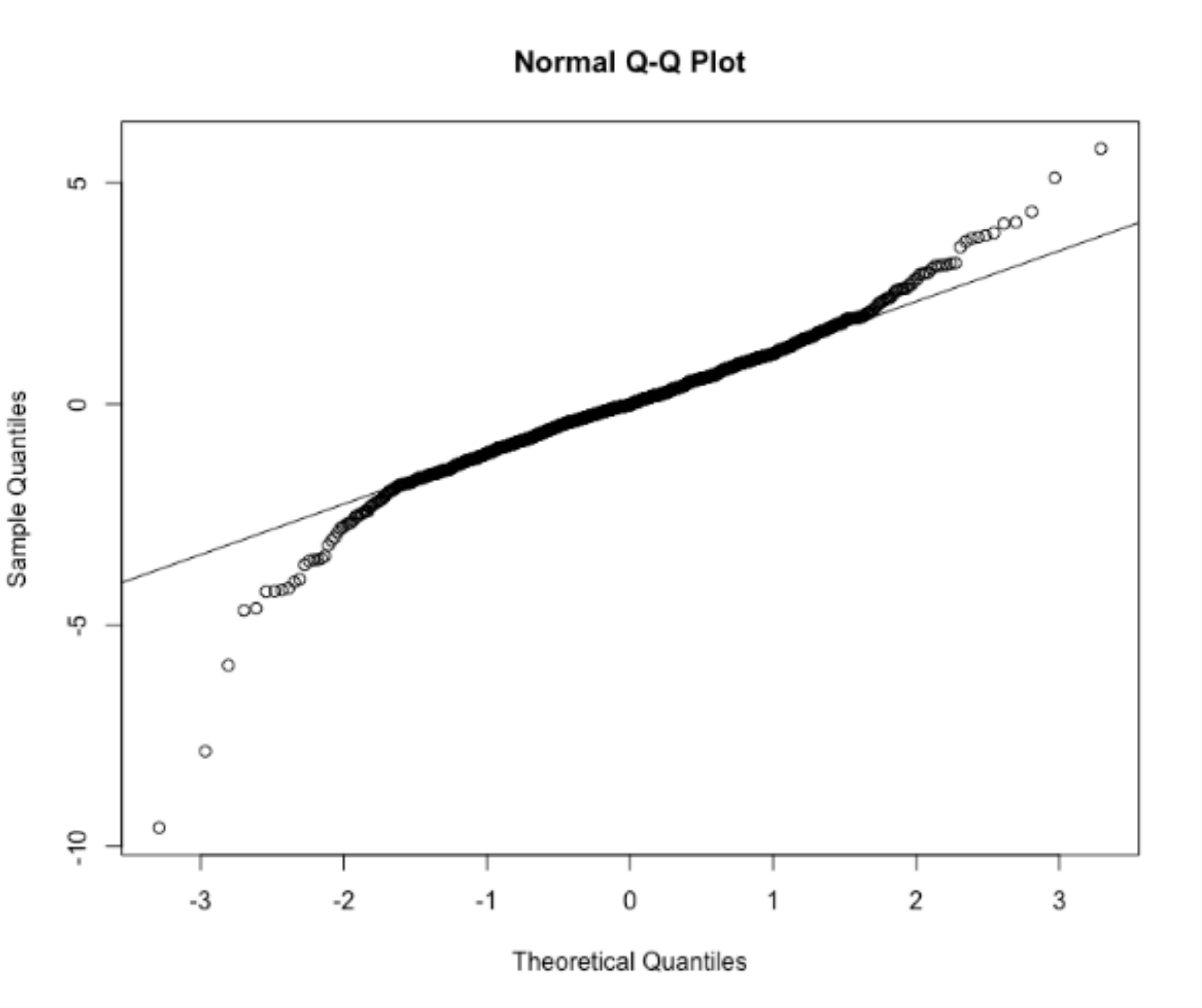
0%

True

0%

False

True or False: The following probability plot displays evidence of normality.



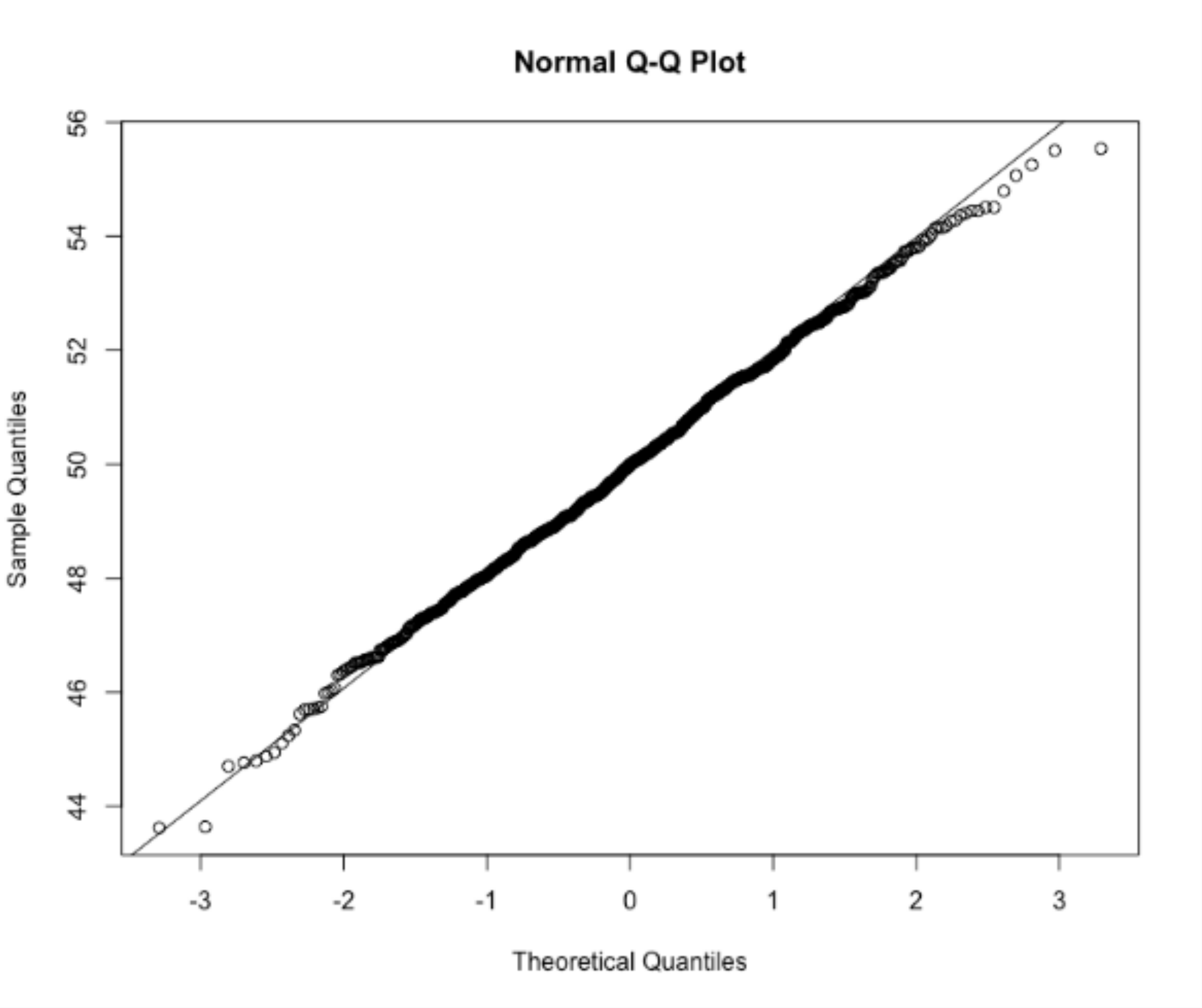
0%

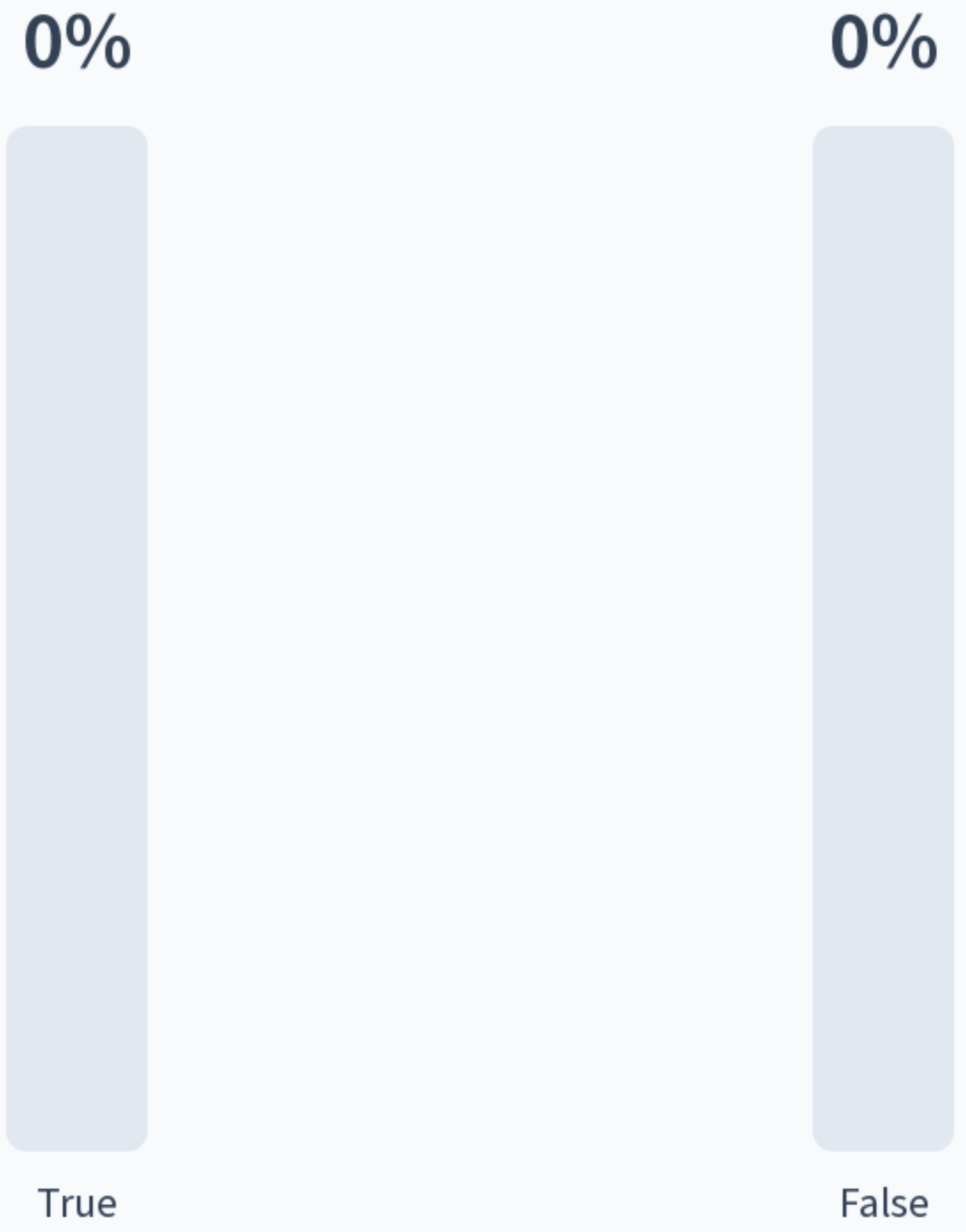
True

0%

False

True or False: The following probability plot displays evidence of normality.





Computing Quantiles

- We saw how to compute theoretical quantiles for a given distribution

$$p = \int_{-\infty}^{\eta(p)} f(x)dx$$

- For sample quantiles, we need a different procedure.
- If data are ordered smallest to largest, this roughly corresponds to the sample quantiles.

Sample Quantiles

- Suppose we observe n data points, and order them $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$.
- $x_{(j)}$ corresponds to the jth smallest value
- In **this course** we will take

$$x_{(j)} = \tilde{\eta}(p) \text{ with } p = \frac{j - 0.5}{n}$$

Example

- If we observe $n = 10$ values what sample quantile is $x_{(1)}$?

$$\frac{1 - 0.5}{10} = \frac{0.5}{10} = 0.05, \text{ or the } 5\text{th percentile.}$$

- Which value corresponds to the 87th percentile?

$$0.87 = \frac{x - 0.5}{10} \implies x = 8.75$$

- **We only consider the theoretical percentiles observed in the sample.**

Example

- Suppose we observe $\{-1.17, -2.30, -1.25\}$, what will our points for the probability plot be?
  - $x_{(1)}: \dfrac{1-0.5}{3} = 0.1\dot{6}$
  - $x_{(2)}: \dfrac{2-0.5}{3} = 0.5$
  - $x_{(3)}: \dfrac{3-0.5}{3} = 0.8\dot{3}$

Example

- Suppose we observe $\{-1.17, -2.30, -1.25\}$, what will our points for the probability plot be?

$$x_{(1)} = \widetilde{\eta}(0.1\dot{6}) = -2.30 \leftrightarrow \eta(0.1\dot{6}) = Z_{0.16} = -0.97$$

$$x_{(2)} = \widetilde{\eta}(0.5) = -1.25 \leftrightarrow \eta(0.5) = Z_{0.5} = 0$$

$$x_{(3)} = \widetilde{\eta}(0.8\dot{3}) = -1.17 \leftrightarrow \eta(0.8\dot{3}) = -Z_{0.16} = 0.97$$

In a sample of 10 observations, $x_{(4)}$ corresponds to which percentile?

$\eta(0.4)$

0%

$\eta(0.35)$

0%

$\eta(0.04)$

0%

$\eta(0.035)$

0%

If there are $n = 15$ observations, which of the following theoretical quantities will not correspond to a value on the probability plot?

$\eta(0.9\dot{6})$

0%

$\eta(0.0\dot{3})$

0%

$\eta(0.5)$

0%

$\eta(0.9\dot{2})$

0%

Suppose that we observe $\{-5, -2, 0, 1, 4, 8, 15\}$. What is $\tilde{\eta}(0.5)$?

0

0%

1

0%

4

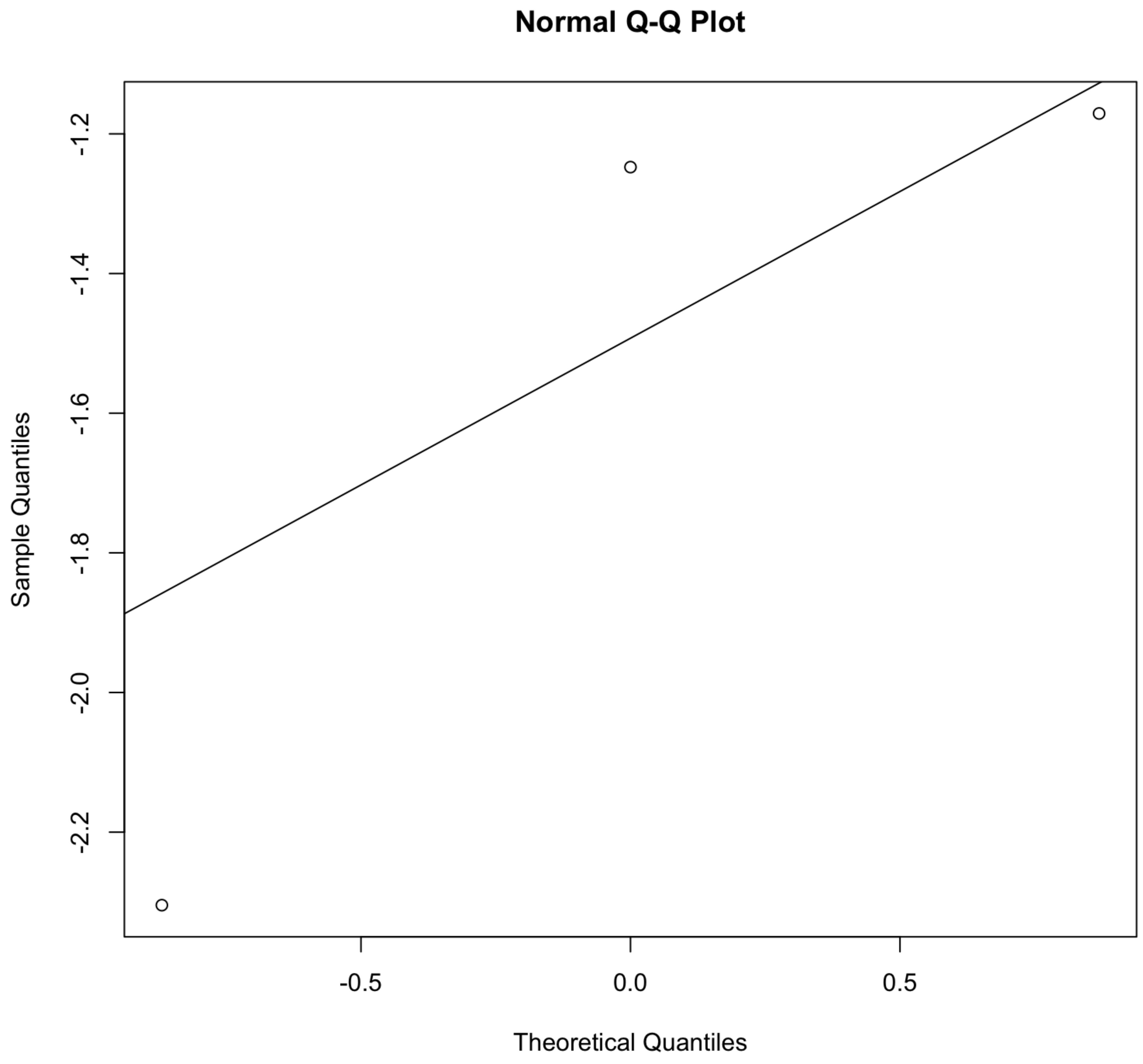
0%

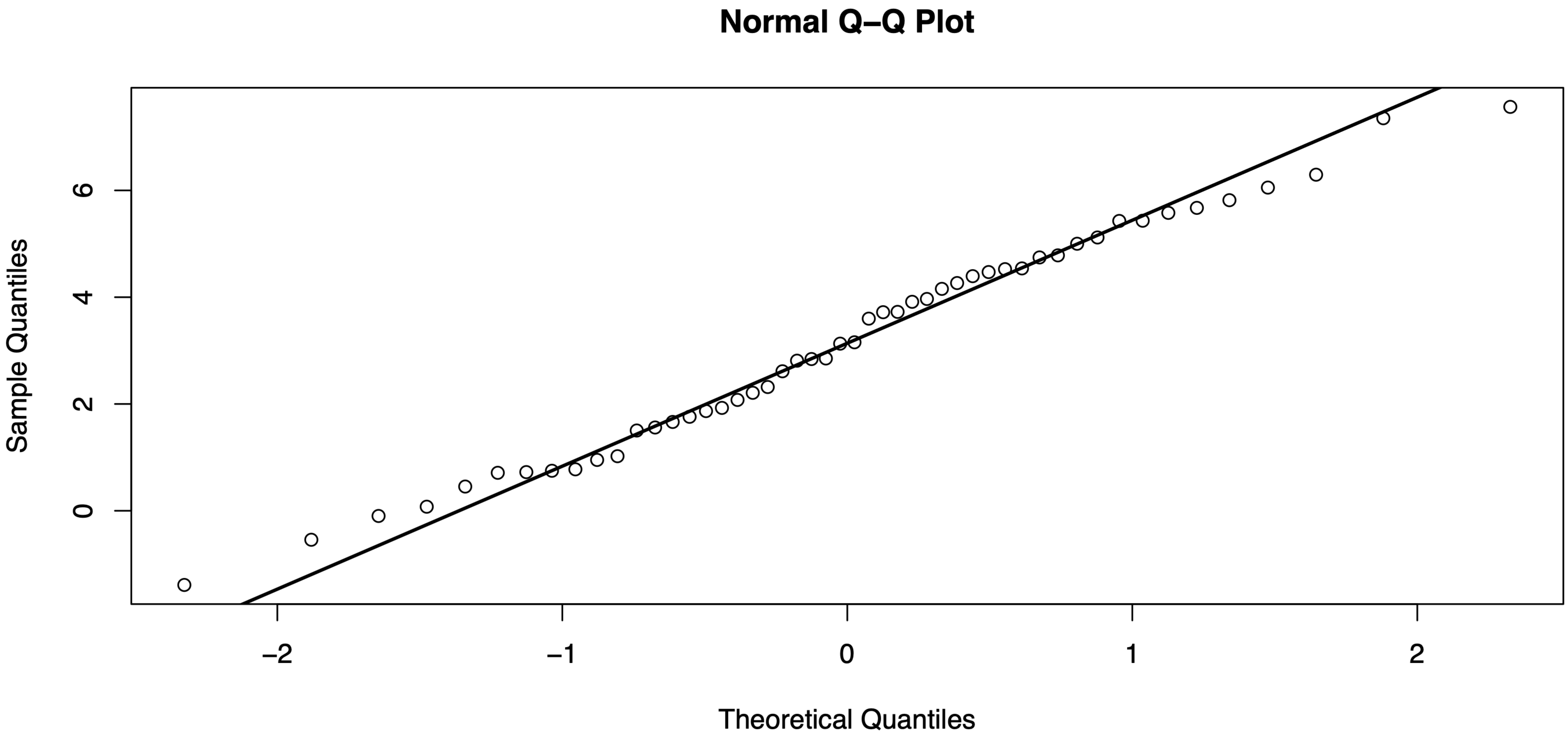
Does not exist in the sample.

0%

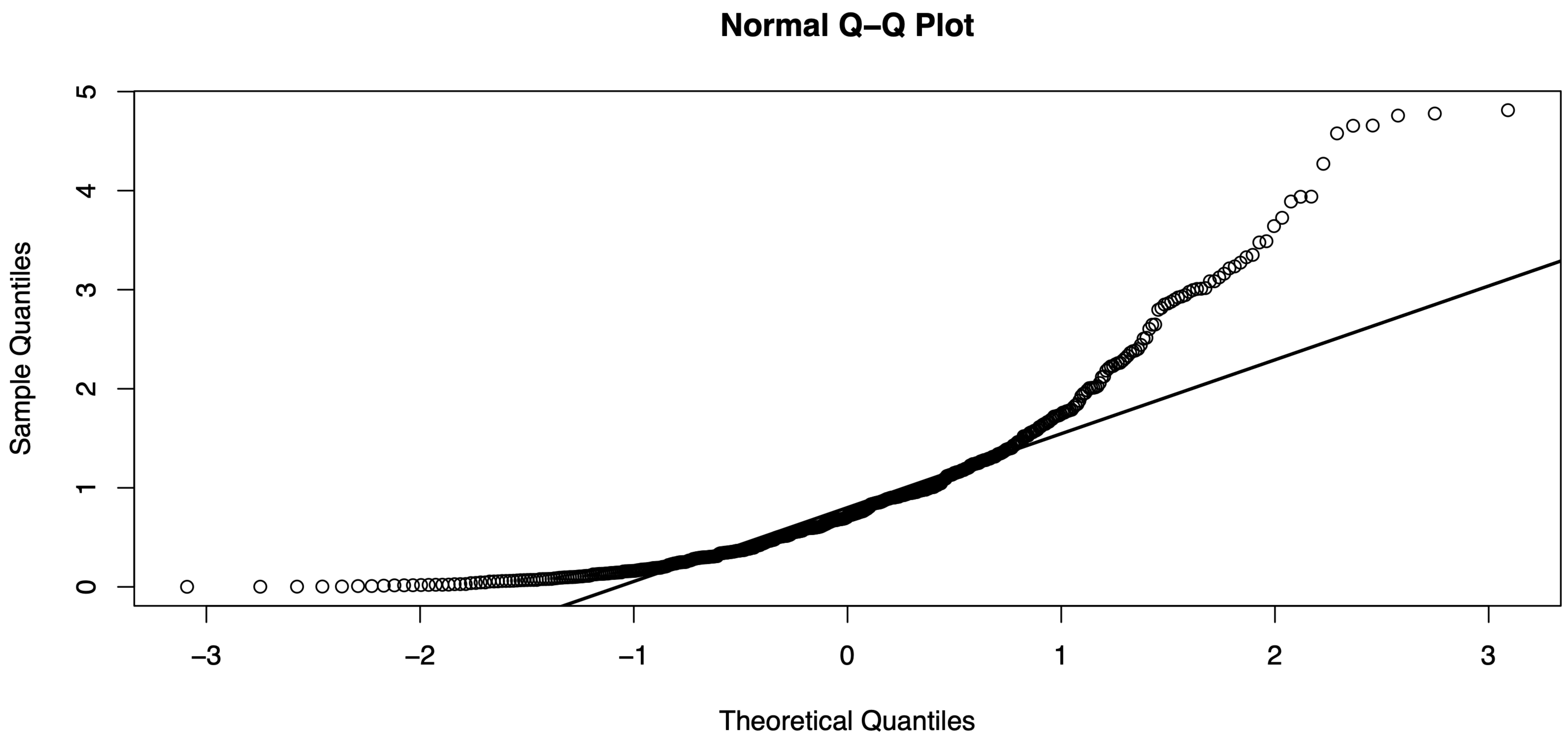
Considerations for Probability Plots

- If the sample size is small, the plots can be challenging to accurately read
  - The last example *was* from a normal distribution but would have had a very messy plot.
- More of "an art" than "a science".

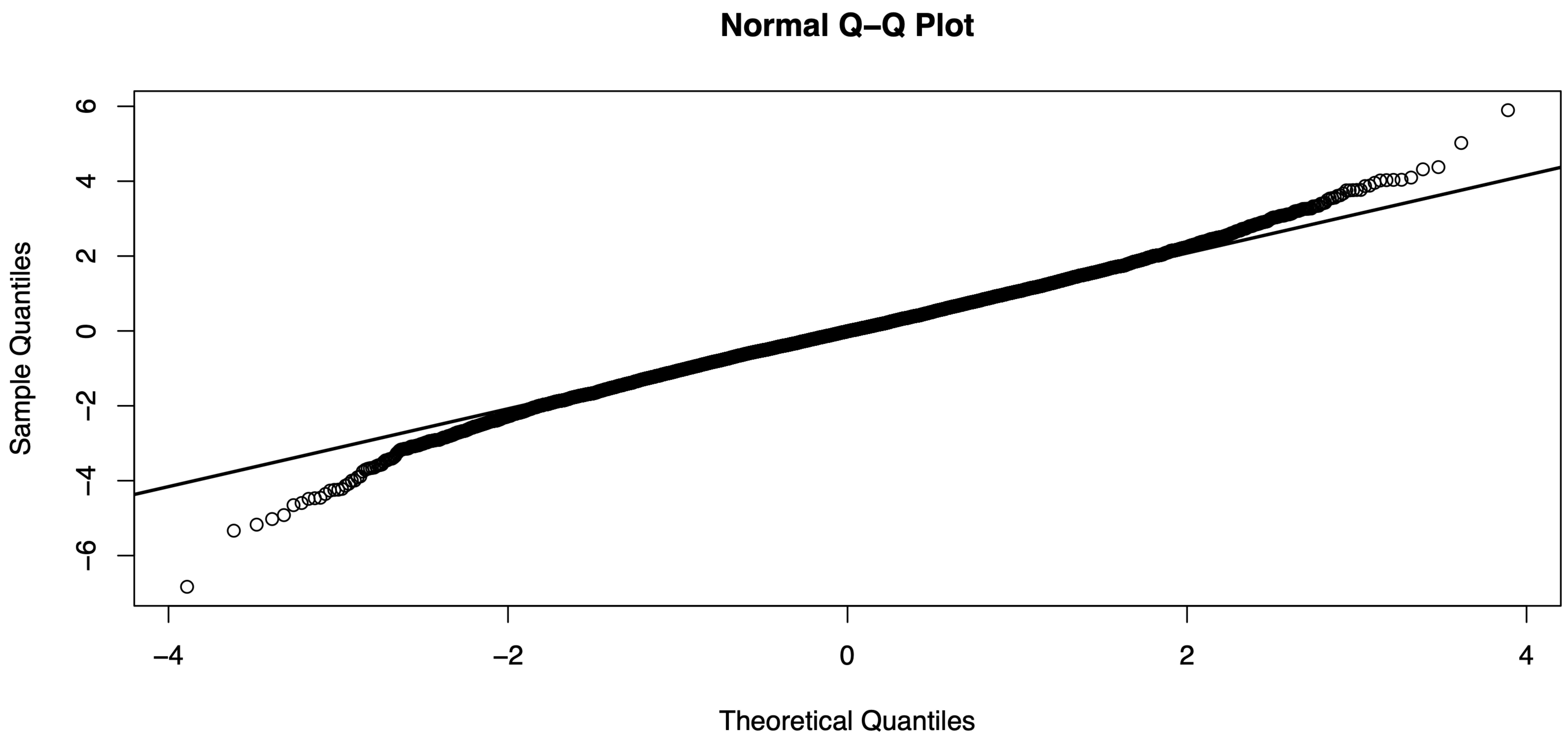




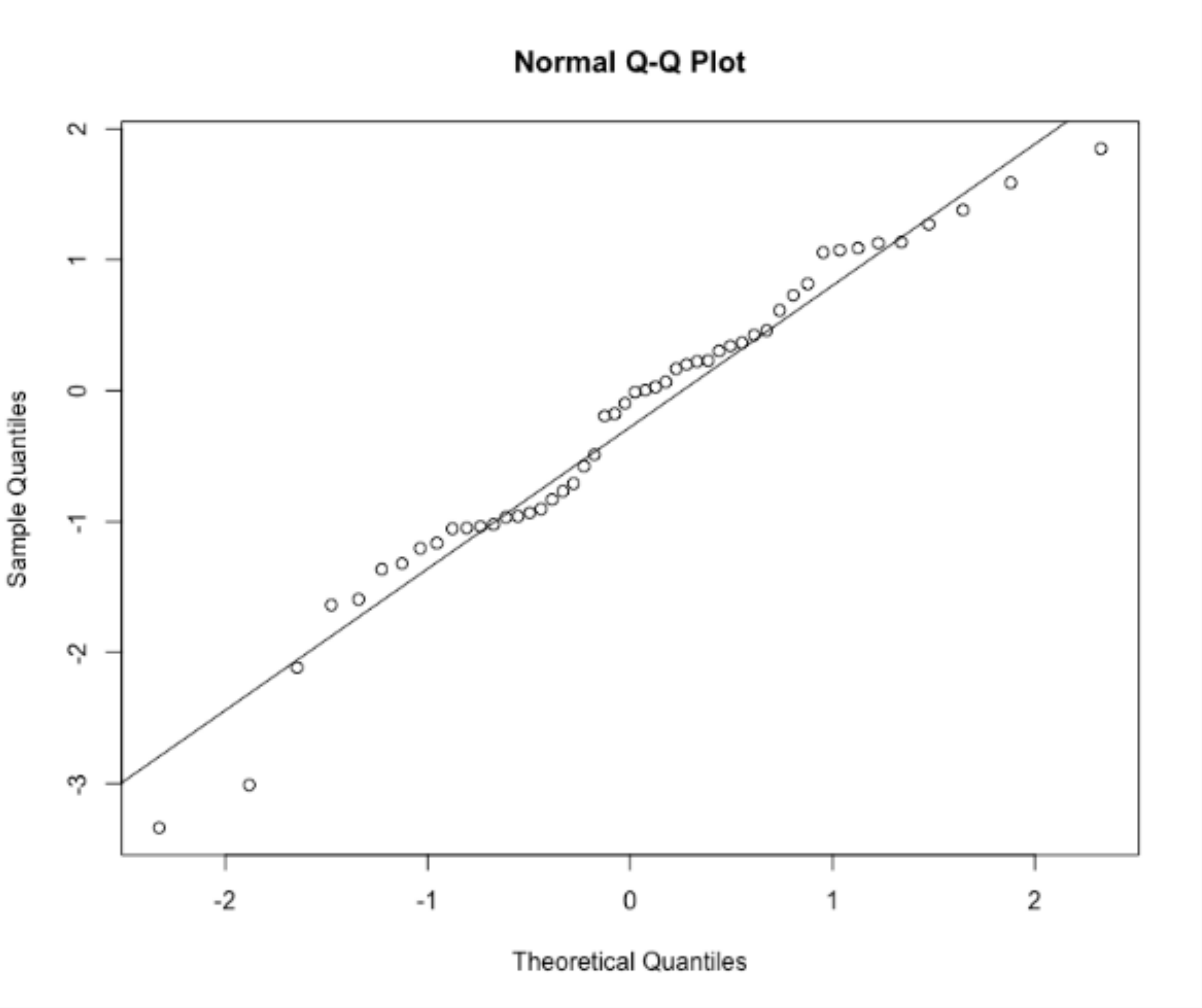
Normal Q–Q Plot
Sample Quantiles
Theoretical Quantiles
0
2
4
6
−2
−1
0
1
2



Normal Q–Q Plot
Sample Quantiles
0
1
2
3
4
5
−3
−2
−1
0
1
2
3
Theoretical Quantiles


Normal Q–Q Plot
Sample Quantiles
Theoretical Quantiles
6
4
2
0
−2
−4
−6
−4
−2
0
2
4

True or False: The following probability plot displays evidence of normality.



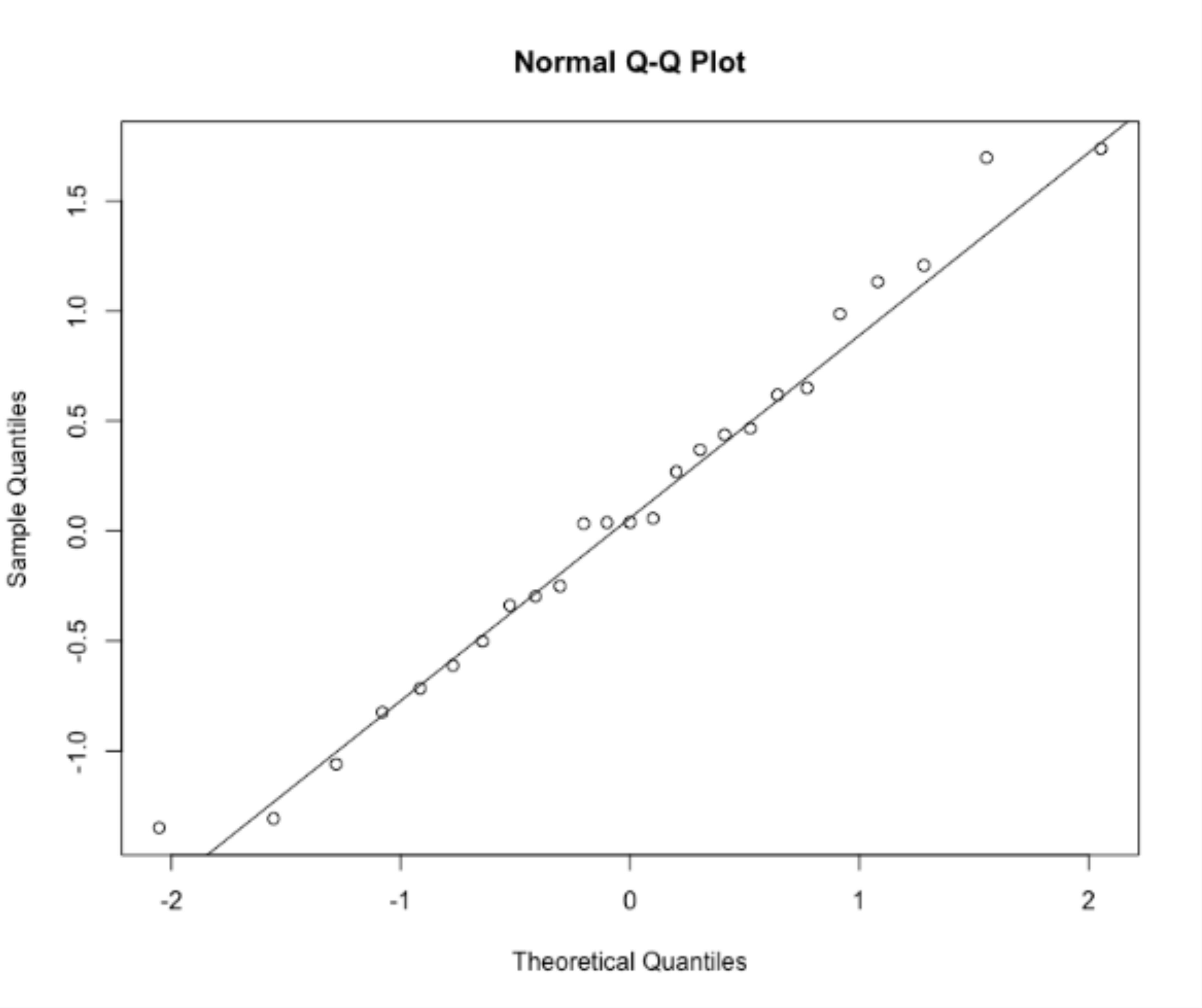
0%

True

0%

False

True or False: The following probability plot displays evidence of normality.



0%

True

0%

False

Non-Standard Normal

- What if we have $X \sim N(\mu, \sigma^2)$ and want to construct a probability plot, but μ and σ^2 are unknown?
- Recall that $\eta(p) = \sigma Z_p + \mu$.
- If $\tilde{\eta}(p) \approx \eta(p) = \sigma Z_p + \mu$ then this provides the ability to consider *any* linear relationship as evidence.
  - Also provides a way to estimate σ and μ.

A probability plot of a dataset versus the standard normal displays a fairly straight linear pattern, with a slope of 3 and a y-intercept of -2. What can we say about the distribution of the data?

It is likely to be approximately $N(0, 1)$.

0%

It is likely to be approximately $N(-2, 3)$.

0%

It is likely to be approximately $N(-2, 9)$.

0%

It is likely to be approximately $N(3, 4)$.

0%

None of the above

0%