Quantile-Quantile Plot (QQ Plot)

A QQ plot investigates whether it's **plausible** that an observed sample is from a specific distribution by plotting theoretical quantiles of the distribution of interest against the sample quantiles, also called empirical quantiles. For example, you may ask whether an observed set appears to be generated from a normal distribution.

During this activity, you will learn the steps involved in constructing a QQ plot, apply the steps to investigate the possible distribution of a sample, and explore how to interpret a QQ plot. Throughout the activity, refer to Table 1 for theoretical quantiles, q, of $N(0,1)$ and Exp(1) for selected values of p such that $P(X \leq q) = p$. Their probability density functions are also provided for reference in Figure 1.

Recall *that p^{th} quantiles are the value of the **random variable** that corresponds to a left-tail probability of p. e.g., $Z = 0$ is the value of a $N(0,1)$ random variable that corresponds with a left-tail probability of 0.5. i.e., $P(Z \leq 0 = 0.5)$.*

Table 1: Theoretical quantiles, q, of $N(0,1)$ and Exp(1) selected values of $p = P(X \leq q)$.

	q			q			q	
p	$N(0,1)$	Exp(1)	p	$N(0,1)$	Exp(1)	p	$N(0,1)$	Exp(1)
1/16	-1.53	0.06	3/8	-0.32	0.47	3/4	0.67	1.39
1/8	-1.15	0.13	7/16	-0.16	0.58	13/16	0.89	1.67
1/6	-0.97	0.18	1/2	0	0.69	5/6	0.97	1.79
3/16	-0.89	0.21	9/16	0.16	0.83	7/8	1.15	2.08
1/4	-0.67	0.29	5/8	0.32	0.98	15/16	1.53	2.77
5/16	-0.49	0.37	2/3	0.43	1.1			
1/3	-0.43	0.41	11/16	0.49	1.16			

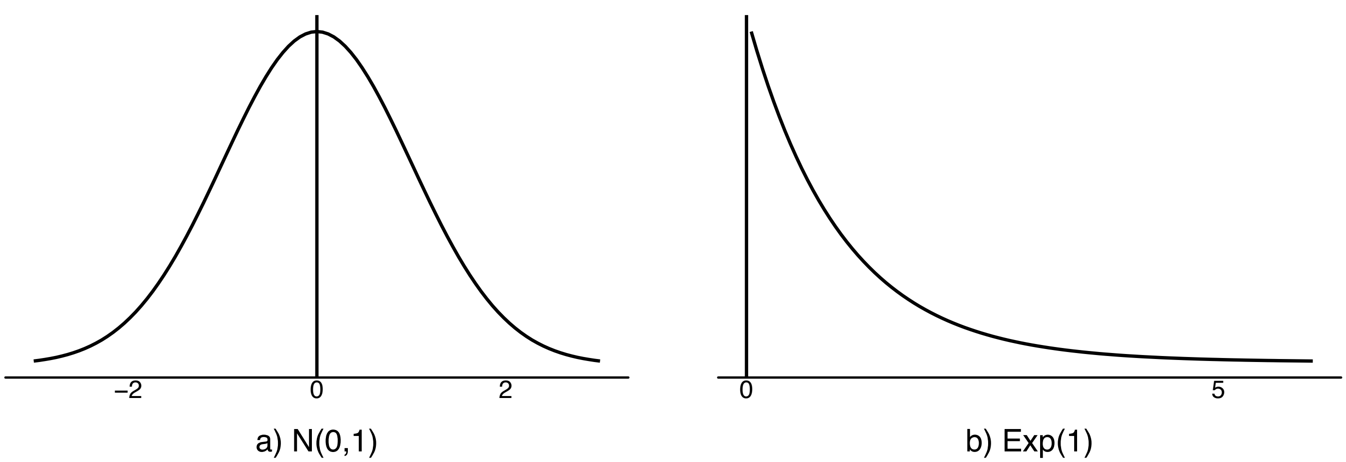


Figure 1: Probability density functions

Learn [2 points]

In this section, you will work through the steps to learn how to build a QQ plot for the following sample:

$$0 \qquad -0.98 \qquad 0.48 \qquad 0.95 \qquad -0.68$$

The sample consists of independent observations from a distribution with a mean of 0 and standard deviation of 1. We will investigate the **plausibility** of the sample coming from the standard normal distribution, $N(0, 1)$.

1. **Sort** the data points in the sample from the smallest to the largest. The sorted values are the *Empirical quantile,* q_n values in order. Write the values on the *Empirical quantile,* q_n row of the table on the next page.

   We placed the smallest value, -0.98, as the first q_n value.

2. **Compute** the corresponding p for each data point. For this tutorial, use the following method to find the p in our data:

   $$p = \frac{i}{n+1}$$

   for $i = 1, 2, ..., n$ where n is the number of observed values in the sample. Write them in order in the p row of the table.

   For the first column, $i = 1$, the corresponding p is $\dfrac{1}{5+1} = \dfrac{1}{6}$.

__Note__: Intuitively, you maybe tempted to use $1/n$ which leads to $p = 1$. The corresponding theoretical quantile of an unbounded distribution such as $N(0, 1)$ is positive infinity, which isn't useful for plotting. Using $(n + 1)$ in the denominator instead is one way to avoid infinities.

---

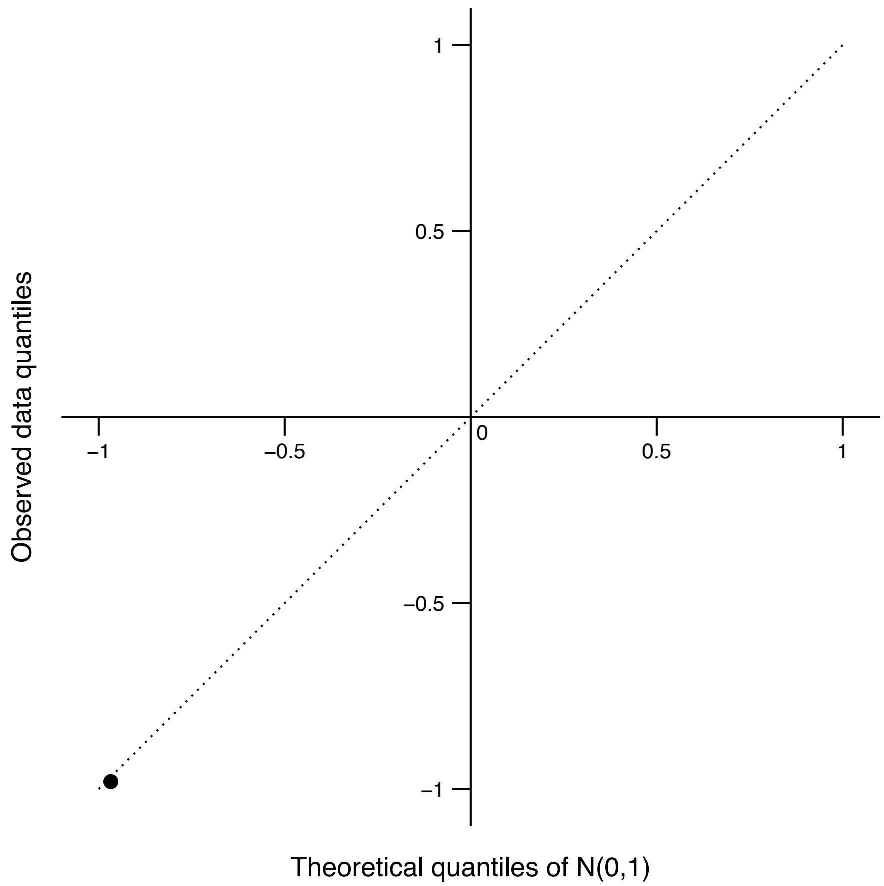
3. **Find** the quantiles of the standard normal distribution that correspond to the values of p. Write them in the *Theoretical* $N(0, 1)$ *quantile,* q row.

   The corresponding $N(0, 1)$ quantile for $p = \frac{1}{6}$ is -0.97 based on Table 1.

4. **Plot** points that correspond to the coordinates $(x, y) = (q, q_n)$. You can use the 45 degree line, $y = x$, as a guide.

   We plotted the first point at (-0.97, -0.98).

Empirical quantile, q_n	-0.98				
p	$1/6$				
Theoretical $N(0,1)$ quantile q	-0.97				



How to read a QQ plot

- QQ plots that align closely along a line suggest that it is plausible the two distributions match in shape - e.g., normally distributed, exponentially distributed, etc. For simplicity, we provided the $y = x$ line around which points may align when the two distributions match in means and variances as well.
- We can interpret this as having no clear signs that the underlying distribution of the sample does **NOT** match the theoretical distribution.
- One common use of QQ plots is to investigate normality of observed data as illustrated above. We call QQ plots that plot sample data points against a theoretical normal distribution, a *normal QQ plot*.