Q-Q Plot

Q stands for quantiles (or percentiles).

Quantiles represent the percentage of data points in a frequency distribution that are less than or equal to a given value.

The intention here is to compare the quantile of the given data set against the quantiles of a theoretical normal distribution.

If quantiles from the 2 distributions, i.e., given data set and theoretical normal distribution are *exactly* same then Q-Q plot results in a straight line. This indicates that the given data likely follows a normal distribution.

---

p-Value

Given a random sample with mean $\overline{X}$.

We want to determine if the sample comes from a population with population mean

H_0: = μ, or
H_a: > μ

μ can be any number.

Assume:
- Population standard Deviation **σ**
- α = given level of significance based on which we will reject or not reject H_0.

Process:
To test the H_0 versus H_a,
- Assume H_0 is true.
- Normalize the sample mean, $\overline{X}$, and
- Compute p-value = P(z > $\frac{(\overline{X}-\mu)}{\sigma/\sqrt{n}}$ >) using standard normal tables.

If: p-value > α, Accept H_0 Else: Cannot accept H_0.

Note: p value provides a "probability of the sample coming from the population as specified in the null hypothesis (H_0)". A low probability value rejects the null hypothesis; if so this means that we cannot conclude that the population comes from a normal distribution.

Confidence Interval

With a level of confidence (C), we can say that

- a population parameter (say, population mean μ) lies within the given confidence interval (CI).
- The confidence interval being derived from a random sample of the population.