Quantile-Quantile Plots

Quantile-Quantile Plots

0:00 / 0:00 1.0x

for the normal distribution,

then it must be because the data is approximated by a normal distribution.

To obtain the quantiles for the data, we can use the quantile function

in r like this.

So we're going to define an object called observe quantiles, and calculate

the quantiles for x at the series of values of p,

which is stored in the object p.

To obtain the theoretical normal distribution quantiles

with the corresponding average and standard deviation,

we use the qnorm function, like this.

We're going to find an

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Textbook link

This video corresponds to the textbook section on quantile-quantile plots.

Key points

- Quantile-quantile plots, or QQ-plots, are used to check whether distributions are well-approximated by a normal distribution.
- Given a proportion p, the quantile q is the value such that the proportion of values in the data below q is p.
- In a QQ-plot, the sample quantiles in the observed data are compared to the theoretical quantiles expected from the normal distribution. If the data are well-approximated by the normal distribution, then the points on the QQ-plot will fall near the identity line (sample = theoretical).
- Calculate sample quantiles (observed quantiles) using the `quantile()` function.
- Calculate theoretical quantiles with the `qnorm()` function. `qnorm()` will calculate quantiles for the standard normal distribution ($\mu = 0, \sigma = 1$) by default, but it can calculate quantiles for any normal distribution given `mean()` and `sd()` arguments. We will learn more about `qnorm()` in the probability course.
- Note that we will learn alternate ways to make QQ-plots with less code later in the series.

Code

```
# define x and z
library(tidyverse)
library(dslabs)
data(heights)
index <- heights$sex=="Male"
x <- heights$height[index]
z <- scale(x)

# proportion of data below 69.5
mean(x <= 69.5)

# calculate observed and theoretical quantiles
p <- seq(0.05, 0.95, 0.05)
observed_quantiles <- quantile(x, p)
theoretical_quantiles <- qnorm(p, mean = mean(x), sd = sd(x))

# make QQ-plot
plot(theoretical_quantiles, observed_quantiles)
abline(0,1)

# make QQ-plot with scaled values
observed_quantiles <- quantile(z, p)
theoretical_quantiles <- qnorm(p)
plot(theoretical_quantiles, observed_quantiles)
abline(0,1)
```

Learn About Verified Certificates