P-P and Q-Q plots

Graphical approaches to model checking

P-P plots

The P-P plot plots

observed cumulative probabilities

at the observed data, against

theoretical cumulative probabilities

computed for the proposed probability model (eg Normal).

The resulting plot for a sample of size n should be a straight line if the proposed model is correct.

P-P plots

For data $y_1, \ldots, y_n$, we first sort the data into ascending order to get

$$y_{(1)} < y_{(2)} < \ldots < y_{(n)}$$

then plot

- for x: $F(y_{(i)})$ for $i = 1, \ldots, n$, where F is the theoretical cumulative distribution function (**cdf**) for the proposed probability model.

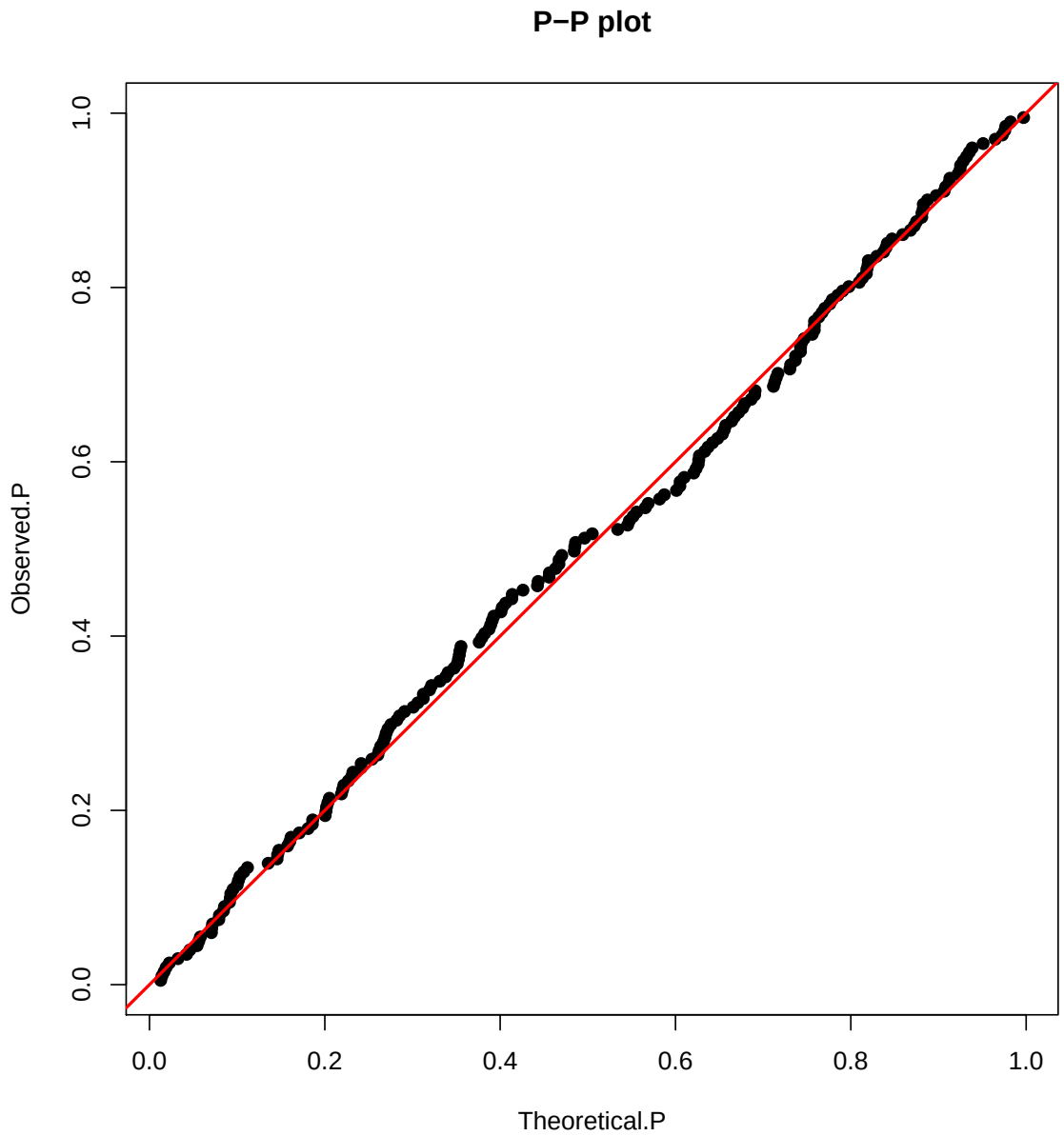
- for y: $i/(n+1)$ for $i = 1, \ldots, n$.

Example: P-P plot for normality

```
set.seed(23412)
n<-200
x<-rnorm(n)

Theoretical.P<-pnorm(sort(x))
Observed.P<-c(1:n)/(n+1)
plot(Theoretical.P,Observed.P,pch=19)
abline(0,1,col="red",lwd=2)
title("P-P plot")
```

Example: P-P plot for normality



P-P plots

The Q-Q plot plots

observed quantiles

against

theoretical quantiles

computed for the proposed probability model (eg Normal), with both evaluated for the probability grid

$$\frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1}$$

The resulting plot for a sample of size n should be a straight line if the proposed model is correct.

Q-Q plots

For data $y_1, \ldots, y_n$, we first sort the data into ascending order to get

$$y_{(1)} < y_{(2)} < \ldots < y_{(n)}$$

then plot

- for x: $F^{-1}(i/(n+1))$ for $i = 1, \ldots, n$, where F^{-1} is the theoretical inverse cumulative distribution function (**inverse cdf**) for the proposed probability model.

- for y: $y_{(i)}$ for $i = 1, \ldots, n$.

Example: Q-Q plot for normality

```
set.seed(23412)
n<-200
x<-rnorm(n)

Theoretical.Q<-qnorm((1:n)/(n+1))
Observed.Q<-sort(x)
plot(Theoretical.Q,Observed.Q,pch=19)
abline(0,1,col="red",lwd=2)
title("Q-Q plot")
```

Example: Q-Q plot for normality

