

Quantile-quantile plots

Quantiles

1. If $p = 3/4$ where is the p -th quantile for:

2. Where is the $1/10$ -th quantile for:

Where is the $1/4$ -th quantile?

3. (Def) A value q is a p -th quantile for a probability distribution P if

$$\begin{aligned}P(X \leq q) &\geq p \\P(X \geq q) &\geq 1 - p\end{aligned}$$

Look at the examples.

4. (Note) If the distribution has a continuous strictly increasing CDF, it has at most one p -th quantile.

$$q = F^{-1}(p)$$

5. (Picture)

6. (Van der Weerden quantiles) The values

$$q_1, \dots, q_n$$

are called **Van der Weerden quantiles** for P if

$$q_i = \frac{i}{n+1} \text{ - th quantile}$$

7. (Continuous distribution) $n = 4$

8. (discrete distribution) $n = 4$

Note the VW quantiles are the mass points.

Q-Q and normal probability plots

1. (Def) A plot of the quantiles of one distribution on the corresponding quantiles of another is called a **Q-Q plot**.
Let

$$x_p = \text{a } p\text{-th quantile for } \mathcal{D}_1$$

$$y_p = \text{a } p\text{-th quantile for } \mathcal{D}_2$$

The Q-Q plot is the parametric plot:

2. (Th) Let

$$z_p = \text{ } p\text{-th quantile for } N(0, 1)$$

$$x_p = \text{ } p\text{-th quantile for } N(\mu, \sigma^2)$$

Then

$$x_p = \mu + \sigma z_p$$

Pf: Let $Z \sim N(0, 1)$ and $X \sim N(\mu, \sigma^2)$.

$$\begin{aligned} p &= P(Z \leq z_p) = P(\mu + \sigma Z \leq \mu + \sigma z_p) \\ &= P(X \leq \mu + \sigma z_p) \end{aligned}$$

Thus $x_p = \mu + \sigma z_p$.

3. (Def) Let \mathcal{D}_n be the sample distribution defined by the sample values x_1, \dots, x_n . A plot of the VW quantiles of \mathcal{D}_n on those of $N(0, 1)$ is a **normal probability plot** for the sample values.
4. (Note) If the sample distribution is approximately normal we expect the normal probability plot to be approximately linear by Th 2.
5. (The plot) Let

$$\begin{aligned} q_1, \dots, q_n &\text{ be the VW quantiles of } \mathcal{D}_n \\ z_1, \dots, z_n &\text{ be the VW quantiles of } N(0, 1) \end{aligned}$$

The normal probability plot is a plot of q_i on z_i . Recall, however, that

$$q_i = x_{(i)} = i\text{-h order statistic}$$

Thus the normal probability plot is:

This is the usual text book plot, but not the usual computer plot.

6. (SAS plot) Let

$$r_i = \text{rank of } x_i \text{ in } x_1, \dots, x_n$$

Then

$$\begin{aligned} x_{(r_i)} &= x_i \\ z_{r_i} &= \Phi^{-1}\left(\frac{r_i}{n+1}\right) \stackrel{\text{df}}{=} s_i \end{aligned}$$

The value s_i is called the normal score for x_i . It may be obtained using PROC RANK. The SAS plot is:

3.4 Weighted regression

1. (Def) The weighted LS criterion is

$$Q_w(\alpha, \beta) = \sum w_i (y_i - \alpha - \beta x_i)^2$$

2. (Gauss Markov Th) If the model $y_i = \alpha + \beta x_i + e_i$ is unbiased, has uncorrelated errors, and $w_i = c/\text{var } y_i$, then the weighted LS estimates of α and β are minimum variance linear unbiased (BLUE) estimates.

Pf: Ch 6

3. (Terminology) The WLS estimates from Th 2 are called
Optimal or GM estimates

The weights

$$w_i = c/\text{var } y_i$$

are called optimal or GM weights.

4. (Note) Weights w_i are optimal iff

$$\text{var}(\sqrt{w_i} y_i) = \text{const}$$

Pf: $\text{var}(\sqrt{w_i} y_i) = w_i \text{var } y_i = c$

5. (Ex) $y_i \sim \mathcal{P}(\mu_i)$

$$\text{var } y_i = \mu_i$$

$$w_i = 1/\mu_i \approx 1/\hat{y}_i \quad , \quad \hat{y}_i \text{ from LS}$$

6. (Ex) $y_i = \mu_i e_i =$ multiplicative errors

$$\text{var } y_i = \mu_i^2 \sigma^2$$

$$w_i = 1/\mu_i^2 \approx 1/\hat{y}_i^2$$