ORIGIN ≡ 1

Assessing Data Normality using Q-Q Plots

W. Stein

Assessing Normality of sample data is an essential part of statistical analysis. Q-Q Plots are one easy way to do this. They are also interesting at this point in our course since they demonstrate the use of the inverse cumulative probability function for the Normal Distribution.

Reading Anderson's Iris data:

iris := READPRN("c:/2008LinearModelsData/iris.txt")

$SL := iris^{\langle 2 \rangle}$ **< assigning variable SL**

n := length(SL) n = 150 **< n = number of observations X**

i := 1 .. n **< constructing index variable i**

$Xbar_{SL} := mean(SL)$ $Xbar_{SL} = 5.8433$ **< mean of X**

$SD_{SL} := \sqrt{Var(SL)}$ $SD_{SL} = 0.8281$ **< sample standard deviation of X**

$SE_{SL} := \frac{SD_{SL}}{\sqrt{n}}$ $SE_{SL} = 0.0676$ **< standard error of the sample *mean* of X**

Calculating Cumultive Probability levels $\Phi_N(X)$:

We will look at variable SL here:

SL =

	1
1	5.1
2	4.9
3	4.7
4	4.6
5	5
6	5.4
7	4.6
8	5
9	4.4
10	4.9
11	5.4
12	4.8
13	4.8
14	4.3
15	5.8
16	5.7

First we sort SL:

$SL_{sort} := sort(SL)$

SL_{sort} =

	1
12	4.8
13	4.8
14	4.8
15	4.8
16	4.8
17	4.9
18	4.9
19	4.9
20	4.9
21	4.9
22	4.9
23	5
24	5
25	5
26	5
27	5

Now we treat each index of SL as a quantile:

$$P_i := \frac{\left(i - \frac{1}{2}\right)}{n}$$

^ the 1/2 here is a correction factor

P =

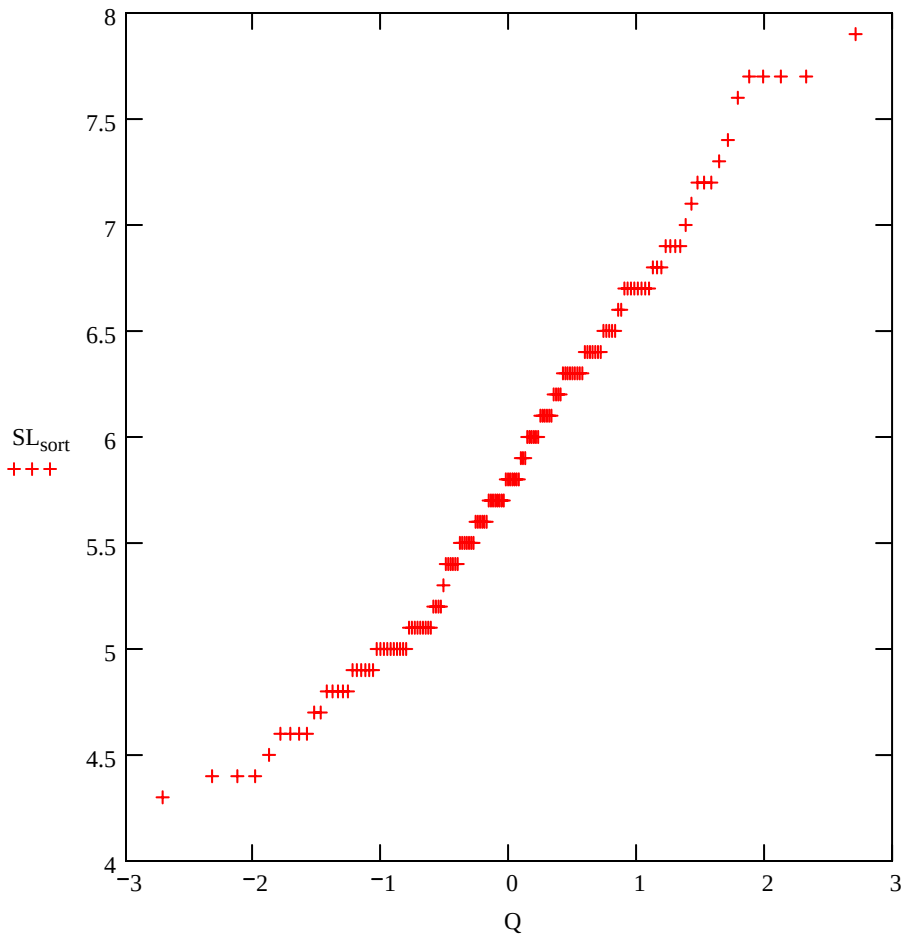
	1
1	0.0033
2	0.01
3	0.0167
4	0.0233
5	0.03
6	0.0367
7	0.0433
8	0.05
9	0.0567
10	0.0633
11	0.07
12	0.0767
13	0.0833
14	0.09
15	0.0967
16	0.1033

From the values of $P = \Phi_N(X)$, we now convert back to X

$Q_i := qnorm(P_i, 0, 1)$

Q =

	1
135	1.2628
136	1.3008
137	1.3408
138	1.383
139	1.4279
140	1.4758
141	1.5274
142	1.5834
143	1.6449
144	1.7132
145	1.7908
146	1.8808
147	1.9893
148	2.128
149	2.3263
150	2.7131



If the sample data are distributed close to the Normal distribution, the Q-Q plot should be mostly a straight line in the center with an overall S-shaped curve towards each end.

Prototype in R:

COMMANDS:

```
#READ IRIS TABLE AND ASSIGN VARIABLE SL
K=read.table("c:/2007BiostatsData/iris.txt")
attach(K)
SL=Sepal.Length

#LOAD PACKAGE  choose "lattice" from pop-up list
local({pkg <- select.list(sort(.packages(all.available = TRUE)))
if(nchar(pkg)) library(pkg, character.only=TRUE)})
#DOCUMENTATION:
? qqmath()
#QQ PLOT IN lattice:
qqmath(SL)
```

RETURN: