Q-Q Grafiği (Q-Q Plot)

Bir veri setinin belli bir dağılımdan gelip gelmediğine dair istatistiksel testlerin yanı sıra grafiksel yöntemlerden de faydalanılır. Bunlardan en yaygın olarak kullanılanı Q-Q grafiği (Q-Q plot) dir.

<u>Normal dağılım için Q-Q grafiği</u>

Q-Q grafiği $x_1, x_2, \dots, x_n$ örnekleminin normal dağılımdan gelip gelmediğini belirlemek için kantilleri (quantile) esas alan ve uygulamada yaygın olarak kullanılan grafiksel bir yöntemdir.

Q-Q grafik yöntemin aşamaları aşağıda gösterildiği gibidir.

1. Öncelikle veriler küçükten büyüğe doğru sıralanır

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}.$$

2. Daha sonra sıralanmış veriler $(x_{(j)})$ ile standart normal dağılımın tablo değerleri (z_i) koordinat sisteminde gösterilir. Burada,

$$P(Z \leq z_i) = \Phi(z_i) = \frac{i - 0.5}{n}$$

şeklindedir.

3. Eğer koordinat sisteminde gösterilen sıralanmış veriler ile standart normal dağılımın tablo değerleri doğrusal bir çizgi üzerinde yayılım gösteriyorsa verilerin normal dağılıma sahip olduğu söylenir.

<u>Örnek:</u> Aşağıdaki veri setinin normal dağılımdan geldiği varsayılmaktadır.

x_i: 1172, 1156, 1201, 1232, 1114, 1250, 1303, 1271, 1325, 1131

Varsayımın doğru olup olmadığını Q-Q grafik yöntemini kullanarak sınayınız.

Çözüm:

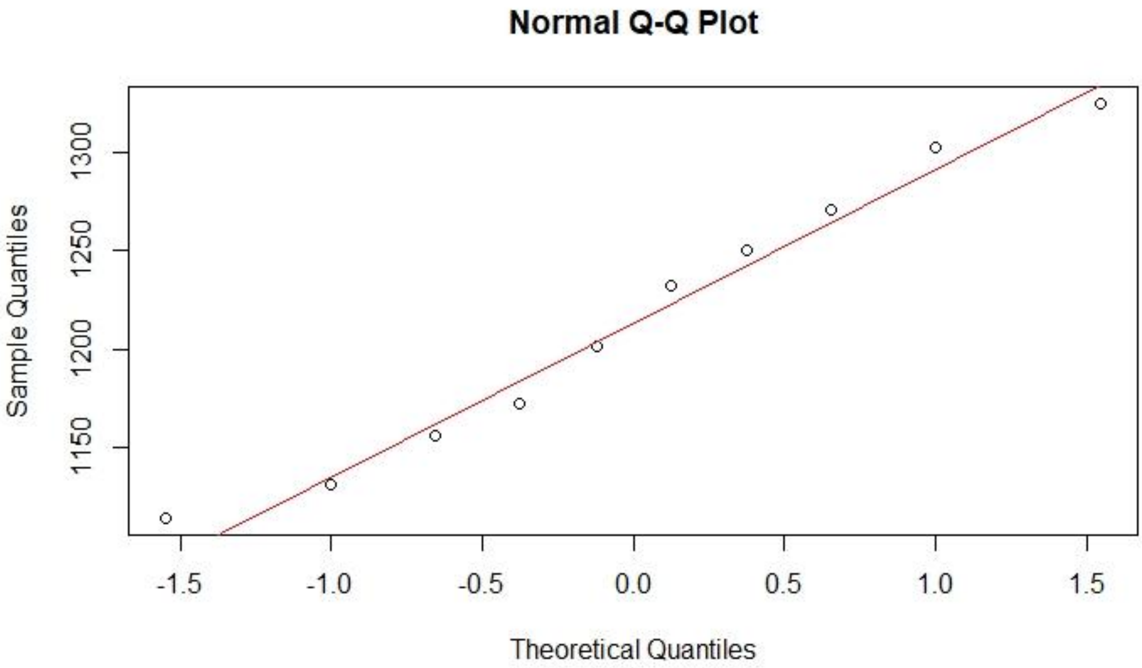
1. Öncelikle veriler küçükten büyüğe doğru sıralanır.

$x_{(i)}$: 1114, 1131, 1156, 1172, 1201, 1232, 1250, 1271, 1303, 1325

2. Daha sonra standart normal dağılımın tablo değerleri $P(Z \leq z_i) = \Phi(z_i) = \frac{i-0.5}{n}$ eşitliği yardımıyla bulunur.

$\boldsymbol{i}$	$\boldsymbol{x_{(i)}}$	$\frac{i-0.5}{n}$	z_i
1	1114	0.05	-1.64
2	1131	0.15	-1.04
3	1156	0.25	-0.67
4	1172	0.35	-0.39
5	1201	0.45	-0.13
6	1232	0.55	0.13
7	1250	0.65	0.39
8	1271	0.75	0.67
9	1303	0.85	1.04
10	1325	0.95	1.64

3. Bu veri seti için Q-Q grafiği aşağıdaki gibi elde edilir.



Gözlemler düz bir doğru etrafında yayılım gösterdiğinden, veri setinin dağılımının normal olduğu sonucuna varılır.

Diğer dağılımlar için Q-Q grafiği

Q-Q grafiği $x_1, x_2, \dots, x_n$ örnekleminin dağılımının normal olup olmadığını belirlemenin yanı sıra belirlenen başka bir dağılımdan geldiği varsayımını test etmek için de kullanılır. Diğer dağılımlar için Q-Q grafiği aşağıdaki adımlar izlenerek bulunur.

1. Öncelikle veriler küçükten büyüğe doğru sıralanır.

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

2. Daha sonra standartlaştırılmış sıra istatistiklerinin beklenen değeri olarak tanımlanan

$$t_{(i)} = E\left(z_{(i)}\right) \text{ ve } z_{(i)} = \frac{x_{(i)}-\mu}{\sigma}$$

değerleri aşağıdaki eşitlik yardımıyla

$$F\left(t_{(i)}\right) = \int_{-\infty}^{t_{(i)}} f(z)dz = \frac{i}{n+1} \Rightarrow t_{(i)} = F^{-1}(\frac{i}{n+1})$$

şeklinde hesaplanır.

3. Daha sonra sıralanmış $x_{(i)}$ değerleri ile $t_{(i)}$ değerleri koordinat sisteminde gösterilir. Gözlemlerin düz bir doğru etrafında yayılım gösterip göstermediğine bakılarak verinin belirlenen dağılımdan gelip gelmediği hakkında sonuca varılır.

<u>Örnek:</u> Örneklem çapı n=15 olan aşağıdaki veri setinin şekil parametresi p=2.0 olan Weibull dağılımından gelip gelmediğini Q-Q grafik yardımıyla gösteriniz.

x_i: 0.2701, 0.5768, 1.4174, 1.4345, 1.5772, 0.4074, 0.8643, 0.2879, 1.5019, 0.9539, 0.5689, 0.7615, 1.1153, 0.6541, 1.4064

Weibull dağılımının olasılık yoğunluk fonksiyonu aşağıda verildiği gibidir

$$f(x) = px^{p-1}e^{-x^p}, \quad x \geq 0.$$

<u>Çözüm:</u>

1. Veriler küçükten büyüğe doğru sıralanır.

$\boldsymbol{i}$	$\boldsymbol{x_{(i)}}$
1	0.2701
2	0.2879
3	0.4074
4	0.5689
5	0.5768
6	0.6541
7	0.7615
8	0.8643
9	0.9539
10	1.1153
11	1.4064
12	1.4174
13	1.4345
14	1.5019
15	1.5772

2. t(i) değerleri Weibull dağılımının birikimli dağılım fonksiyonu kullanılarak

$$F(z) = 1 - e^{-z^p}$$

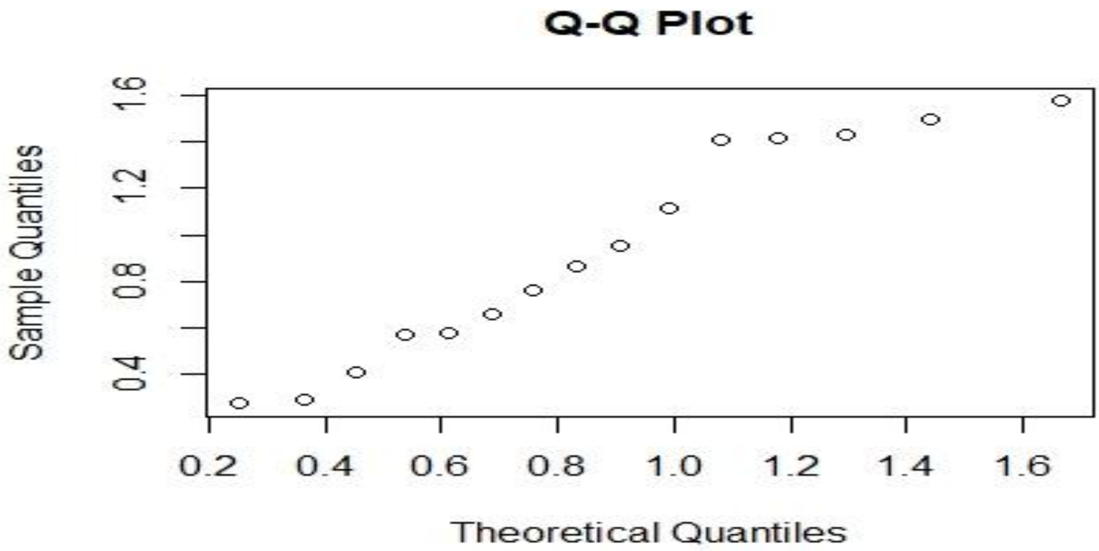
$$F\big(t(i)\big) = \int_{-\propto}^{t(i)} f(z)dz = \frac{i}{n+i}$$

$$t(i) = F^{-1}\left(\frac{i}{n+i}\right) = \left(-ln\left(1 - \frac{i}{n+1}\right)\right)^{1/p}$$

eşitliği yardımıyla hesaplanır ve aşağıdaki tabloda gösterildiği gibi elde edilir.

$\boldsymbol{i}$	$\boldsymbol{t_{(i)}}$
1	0.2540
2	0.3654
3	0.4557
4	0.5364
5	0.6121
6	0.6156
7	0.7585
8	0.8326
9	0.9092
10	0.9904
11	1.0785
12	1.1774
13	1.2938
14	1.4420
15	1.6651

3. Bu veri seti için Q-Q grafiği aşağıdaki gibi elde edilir.



Gözlemlerin düz bir doğru etrafında yayılım gösterip göstermediğine bakarak, veri setinin dağılımının şekil parametresi 2 olan Weibull olduğu sonucuna varılabilir. Ancak ortada yer alan bazı değerlerin doğrudan sapma gösterdiği düşünülerek veri setinin dağılımının şekil parametresi 2 olan Weibull olmadığı sonucuna da varılabilir. Bu yöntemin en büyük dezavantajı kişiden kişiye değişiklik göstermesi bir başka deyişle subjektif olmasıdır.

Örnek: Aşağıdaki veri setinin dağılımının Lojistik olup olmadığını Q-Q grafik yardımıyla belirleyiniz.

x_i: 1.5355, 2.8508, 0.2865, -0.3421, -0.8526, -0.1192, 1.0987, -1.3796, 1.6854, 0.2758, 0.4584, 1.1094, -0.9482, -3.0990, 1.8967

Lojistik dağılımın olasılık yoğunluk fonksiyonu aşağıda gösterildiği gibidir

$$f(z) = \frac{e^{-z}}{(1+e^{-z})^2}.$$

1. Veriler küçükten büyüğe doğru sıralanır.

$\boldsymbol{i}$	$\boldsymbol{x_{(i)}}$
1	-3.0990
2	-1.3796
3	-0.9482
4	-0.8526
5	-0.3421
6	-0.1192
7	0.2758
8	0.2865
9	0.4584
10	1.0987
11	1.1094
12	1.5355
13	1.6854
14	1.8967
15	2.8508

2. t(i) değerleri Lojistik dağılımının birikimli dağılım fonksiyonu kullanılarak

$$F(z) = \frac{1}{(1+e^{-z})}$$

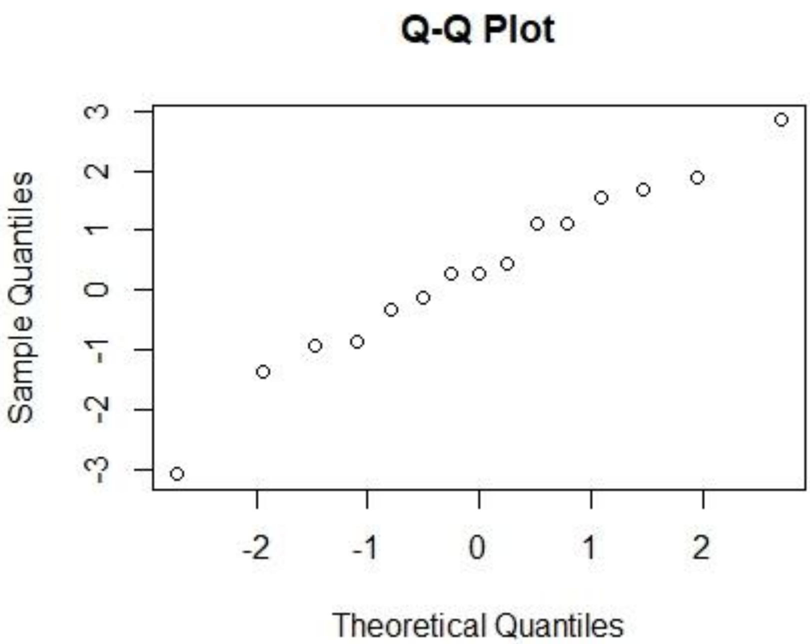
$$F\big(t(i)\big) = \int_{-\propto}^{t(i)} f(z)dz = \frac{i}{n+i}$$

$$t(i) = F^{-1}\left(\frac{i}{n+i}\right) = -ln\left(\frac{n+1}{i} - 1\right)$$

eşitliği yardımıyla hesaplanır ve aşağıdaki tabloda gösterildiği gibi elde edilir.

i	$t_{(i)}$
1	-2.7081
2	-1.9459
3	-1.4663
4	-1.0986
5	-0.7885
6	-0.5108
7	-0.2513
8	0.0000
9	0.2513
10	0.5108
11	0.7885
12	1.0986
13	1.4663
14	1.9459
15	2.7081

3. Bu veri seti için Q-Q grafiği aşağıdaki gibi elde edilir.



Gözlemler düz bir doğru etrafında yayılım gösterdiğinden, veri setinin dağılımının Lojistik olduğu sonucuna varılabilir.