

Quartiles, Quantile, Percentile, Decile

Quartiles: (4'te 1'lik)

Quartiles are values that divide your data into quarters.

2, 5, 6 | 7, 10, 12 | 13, 14, 15 | 22, 45, 65 g'bi.

(Lower Quart) Q_1

Q_2

Q_3 (Upper Quartile)

1st Quartile

2nd Quartile

3rd Quartile

4th Quartile

-3

-2

-1

0

1

2

3

Second Quartile (Q_2) is known as "Median".

.../.../...
IQR (Interquartile Range) : Çeyrekler arası aralık

$$IQR = Q_3 - Q_1$$

* Değerlerin büyük kısmının nerede bulunduğu bir ölçüsüdür.

Percentile : (Yüzdelik Dilim)

Q_1 is also called as 25th Percentile.

Q_2 is also called as 50th Percentile.

Q_3 is also called as 75th Percentile.

Percentile divide the dataset to 100 equally pieces.

Deciles : (Onlar Birlik)

Divide the dataset to 10 equally pieces.

Quantile :

* Quantile is where a sample is divided into equal-sized, subgroups.

* It can also refer to dividing a probability distribution into areas of equal probability.

ÖR

Quantiles are also quantiles; they divide distribution into 4 equal parts.

Percentiles are also quantiles that divide distribution into 100 equal parts.

Median (Q_2) is 0.5 quantile (Çünkü veriyi 2'ye bölüyor).

Quantiles, q -quantiles have special names

→ 2-quantile is called Median.

4-quantile is called Quartiles.

10-quantile is called Deciles.

100-quantile is called Percentiles.

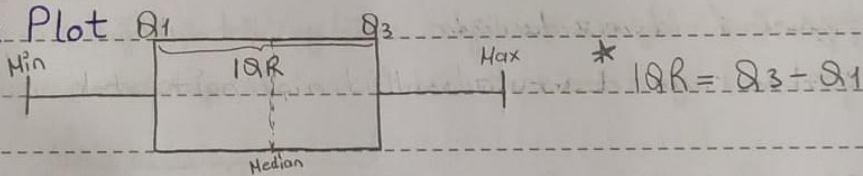
1000-quantile is called Permilles

Sonuç olarak;

q -quantiles, sonlu bir değerler kümesini eşit büyüklükteki q alt kümelerine bölen değerlerdir.



Box Plot



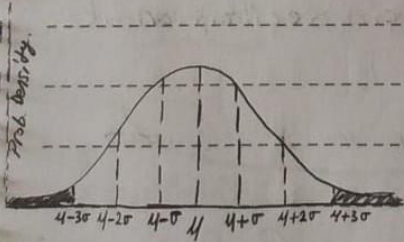
Box plot kullanarak outlier detect etmek için Min (Lower Bound);

ve Max (Upper Bound) dışında kalan gözlemlere odaklanılır.

$$\text{Min} = \text{Lower Bound} : (Q1 - 1.5 \times \text{IQR})$$

$$\text{Max} = \text{Upper Bound} : (Q3 + 1.5 \times \text{IQR})$$

Peki neden 1.5 IQR?



(Arka sayfada 1'şpat olacak.)

$$\%68,26 < \sigma$$

$$\%95,44 < 2\sigma$$

$$\%99,72 < 3\sigma$$

Rest $> 3\sigma$

$\%0,28$ (Taraflı Alan)

Dipnot: $Q1$ and $Q3$ ortalamadan

-0.675σ ve $+0.675\sigma$

uzaktadırlar.

Ne dedik?

%68,26 ortalamadan $\pm \sigma$ uzaklıktadır.

%95,44 ortalamadan $\pm 2\sigma$ uzaklıktadır.

%99,72 " $\pm 3\sigma$ "

ve

%0,28 $\pm 3\sigma$ 'dan daha uzak.

1QR çarpanını 1 aldığımız Durum:

Lower Bound $\Rightarrow Q1 - 1 \times IQR$

$$= Q1 - 1 \times (Q3 - Q1) \Rightarrow -0,675\sigma - 1 \times (0,675 - [-0,675])\sigma = -2,025\sigma$$

Upper Bound $\Rightarrow Q3 + 1 \times IQR \Rightarrow +2,025\sigma$

Yani 1 aldığımız durumda 3σ 'ya çok uzaktır ve çok gözlem kaybederiz.

1QR çarpanı 2 olduğu Durum:

Bu durumda ise $\pm 3,375\sigma$ uzaklıkta oluruz mean'den

Buda $\pm 3\sigma$ 'dan çok fazla.

Kısacası, box plot ile outlier detect ederken 1,5 IQR

3σ 'ya en yakın sonucu verdiği için seçilmiş ancak probleme göre değiştirebiliriz.

Buradaki tüm çıkartmalar, data'mızın Gaussian Dağılımına sahip olduğu farz edilerek yapılmıştır. Yani 1,5 değeri değiştirilebilir.