

Quantile Function, Quantile Regression

Quantile fonksiyonu $Q(\theta)$ ile veya QF ile gösterilir.

* Bir X rastgele değişkeninin dağılım fonksiyonu $F(X)$ ile gösterilir ve X 'in x 'e eşit ya da küçük olması olasılığıdır.

Yani;

$$F(X) = P(X \leq x) = \sum_{X \leq x} f(x)$$

$f(x) \Rightarrow$ Olasılık yoğunluk fonksiyonu.

$$f(x) dx = P(x \leq X \leq x + dx)$$

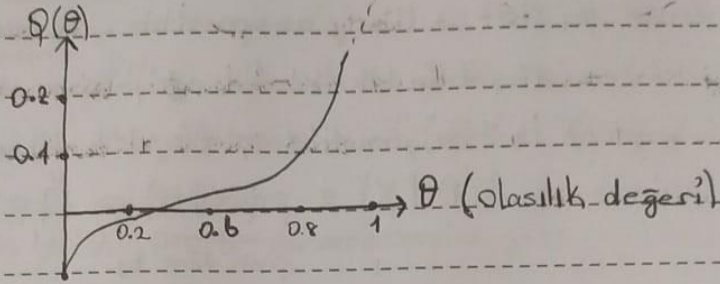
$f(x)$ eğrisinin altındaki alan 1 olmalıdır.

$f(x) =$ Olasılık yoğunluk fonksiyonu, *Kümülatif*

Dağılım fonksiyonunun Türevine Eşittir.

Quantile Fonksiyonu

Grafiki:



$x_\theta = Q(\theta) \Rightarrow$ Quantile Fonksiyonu

$x_\theta \Rightarrow \theta$ 'ı quantile olarak adlandırılır

$Q(\theta)$ ve CDF, herhangi (x, θ) çifti için $x = Q(\theta)$ ve $\theta = F(x)$ şeklinde yazılabilir. Bu fonksiyonlar birbirlerinin tersine eşittir.

Yani; $Q(\theta) = F^{-1}(\theta)$ ve $F(x) = Q^{-1}(x)$

$Q(\theta)$ quantile fonksiyonu θ 'nın tüm olasılıkları için $0 \leq \theta \leq 1$, quantile değerlerini verir.

CDF \rightarrow what area/probability corresponds to a known z score.

Quantile Fonk \rightarrow What z score corresponds to a known area/probability.

$Q(\theta)$; θ 'nin tüm olasılıkları için ($0 \leq \theta \leq 1$), quantile değerlerini verir.

\hookrightarrow Yan sayfada açıklayalım.

* Quantile Fonksiyonu Birikimli yoğunluk fonksiyonunun tersini verir. CDF ile bir değerin veya bu değerden daha düşük bir değerin elde edilme ihtimalini biliyoruz. Quantile Fonksiyonu ile bir ihtimalin hangi değere isabet ettiğini buluyoruz.

CDF \rightarrow Gives probability that a random variable has a value less than or equal to given value.

The Median value is where $CDF = 0.5$

The median (Md) of PDF is:

$$Md = \int_a^{Md} f(x) dx \Rightarrow \int_{Md}^b f(x) dx = 1/2$$

pdf($f(x)$) integrate \rightarrow cdf($F(x)$)

set $F(M) = 0.5$ where M represent Median.

$Q(\theta)$; θ 'nin tüm olasılıklarının ($0 \leq \theta \leq 1$) quantile değerlerini verir dedik.

ÖR $Q(0.5) \Rightarrow$ 0.5 olasılığa karşılık gelen değer? (Median)
(alan)

$Q(1/4) \Rightarrow$ ilk yüzde 25'lik olasılığa karşılık gelen değeri verir.
(0.25'lik alan)

$Q(3/4) \Rightarrow$ yüzde 75'lik alana (0.75) karşılık gelen değeri verir.

Median is value of x such that are equal probabilities of getting a larger value and getting a smaller value than x_{median} .

Quantile Regresyon:

Quantile Regresyon, quantillere bağlı olarak Regresyon katsayılarını belirler. Bu model uç değerlere OLS'den daha hassastır.

OLS'de Hataların karelerinin toplamı minimize ediliyor. Hataların kareleri yerine farklı değerlerin de minimizasyonu söz konusu olabilir.

Quantile Regresyon, lineer Regresyonun şartları (varsayımları) sağlanmadığı durumda kullanılır (linearity, homoscedasticity, independence or normality).

Linear Reg. Model Equation $\Rightarrow y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$

p = number of Features

n = number of data points

The best line found by minimizing the "Mean Square Error" $\Rightarrow MSE = \frac{1}{n} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}))^2$

Quantile Reg. Equation $\Rightarrow Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau) x_{i1} + \dots + \beta_p(\tau) x_{ip}$

Beta katsayıları sabit değil. They are functions of the quantile.

Quantile Reg. reduce the Median absolute Deviation.

$$MAD = \frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - (\beta_0(\tau) + \beta_1(\tau) x_{i1} + \dots + \beta_p(\tau) x_{ip}))$$

$p \Rightarrow$ check function

$$p_T(u) = T \max(u, 0) + (1-T) \max(-u, 0)$$

"u" is error of single data point.

max function returns the largest value in parentheses.

Yani; Error positive ise $T \max(u, 0) \rightarrow$ Function that multiplies error by T.
 \rightarrow check that function

Error negative ise $(1-T) \max(-u, 0) \rightarrow$ Function that multiplies error by $(1-T)$.
 \rightarrow check that func

(Not: Quantile level is the probability (or the proportion of the population) that is associated with a quantile)

$T = 0.5$ iken Median Regression.

$T = 0.10, 0.25 \dots$ olabilir

The quantile $q \in (0, 1) \rightarrow$ splits data into proportions

q below and $1-q$ above.

ÖR

Örneğin; 75th quantile regression fits a regression line through the data so that %75 percent of the observations are below the regression line and %25 percent are above.

* OLS'de line %30, %25 herhangi birinden geçebilir.

75th quantile regression durumunda, en küçük MAN değerini bulmak için (yani 75th durumundaki en iyi line'i bulmak) p_T (weights) have to be added to errors, 75th durumda a weight of 0.75 is added positive ones, while 0.25 is added to negative ones.