Otokorelasyon

* Ekonometride otokorelasyon denildiğinde hata terimlerinin diğer hata terimleri ile ilişkisi akla gelir.
* Otokorelasyona serisel korelasyon da denilmektedir.
* Literatürde bazen farklılıklar olsa da genelde otokorelasyon ve serisel korelasyon aynı şeyleri ifade eder.
  - Fakat bazı yazarlar aynı serinin farklı devreleri arasındaki ilişkiyi ifade etmek için otokorelasyon terimini kullanır. ÖR: $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_{10}$ ve $\varepsilon_2, \varepsilon_3, \varepsilon_4, \ldots, \varepsilon_{11}$
  - İki farklı seri arasındaki ilişkiyi ifade etmek için ise serisel korelasyon terimini kullanırlar: ÖR: $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_{10}$ ve $u_1, u_2, u_3, \ldots, u_{10}$
  - fakat biz bunları aynı kavramlarmış gibi kullanacağız.

---

* Hata terimleri arasında otokorelasyon olmaması regresyonun temel varsayımlarından biridir.
  - Çoklu Doğrusal Regresyonda

$$y = B_0 + B_1 X_1 + B_2 X_2 + \varepsilon$$

$$y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + \varepsilon_i \Rightarrow \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

ÇDR 6

$$① \; Corr(\varepsilon_i, \varepsilon_J \mid X) = 0 \quad \forall\, i, J = 1, \ldots, n$$

$$② \; Cov(\varepsilon_i, \varepsilon_J \mid X) = 0 \quad i \neq J$$

Not: $① \; Cov(\varepsilon_i, \varepsilon_J) = E\left[\varepsilon_i - E(\varepsilon_i)\right]\left[\varepsilon_J - E(\varepsilon_J)\right] = 0$

$$= E\left[\varepsilon_i \varepsilon_J\right] = 0 \quad i \neq J \text{ için}$$

ÇDR5: $E(\varepsilon_i|X) = 0$, $E(\varepsilon_J|X) = 0$

sonuç hata terimleri arasında lineer bir bağıntı yok.

② $Corr(\varepsilon_i, \varepsilon_J) = \frac{\overbrace{Cov(\varepsilon_i, \varepsilon_J)}^{=0}}{\sqrt{Var(\varepsilon_i)}\sqrt{Var(\varepsilon_J)}} = 0$ → GAR 7

$Var(\varepsilon_i | X) = \sigma_\varepsilon^2$

$Var(\varepsilon_J | X) = \sigma_\varepsilon^2$

$$\rho = \frac{E(\varepsilon_i \varepsilon_J)}{\sigma_\varepsilon^2}$$

$\rho = 0$ $i \neq J$ için

↳ yani hata terimleri arasında otokorelasyon yok. her hangi bir i, J ikilisi için

* Otokorelasyon, model doğru tanımlanmışsa ortadan kaldırılabilir. ama sistematik olmayan yani rassal kısmın sınanması gerekir
* Artıklarda rassal olmayan bir görünüm varsa otokorelasyon (OK) olabilir.
* OK zaman serilerinde daha sık görülür.

Otokorelasyonun Mantığı

* örneğin 3 aylık zaman serisi datasıyla ilgilendiğimizi düşünelim ve çıktıları sermaye ve emek üzerine regres edelim. Eğer işçiler bu 3 aylık ilk dönemde greve giderlerse, bu grevin ikinci 3 aylık dönemi etkilemeyeceğini varsaymamız otokorelasyon olmaması durumudur. Yani emekteki bir değişiklik bu dönemdeki çıktıyı etkileyecek fakat bir dönem sonraki çıktıyı etkilemeyecektir

Zaman serisi

$$E(\varepsilon_1 \varepsilon_2)=0 \begin{cases} \text{Çıktı}_1 = B_0 + B_1 \text{Sermaye}_1 + B_2 \text{Emek}_1 + \varepsilon_1 \\ \text{Çıktı}_2 = B_0 + B_1 \text{Sermaye}_2 + B_2 \text{Emek}_2 + \varepsilon_2 \end{cases} \text{OK YOK}$$

$$E(\varepsilon_1 \varepsilon_2)\neq 0 \begin{cases} \text{Çıktı}_1 = B_0 + B_1 \text{Sermaye}_1 + B_2 \text{Emek}_1 + \varepsilon_1 \\ \text{Çıktı}_2 = B_0 + B_1 \text{Sermaye}_2 + B_2 \text{Emek}_2 + \varepsilon_2 \end{cases} \text{OK VAR}$$

Yatay kesit verisi

$$\begin{cases} \text{Tüketim}_1 = B_0 + B_1 \text{Gelir}_1 + \varepsilon_1 \Rightarrow \text{aile } 1 \\ \text{Tüketim}_2 = B_0 + B_1 \text{Gelir}_2 + \varepsilon_{12} \Rightarrow \text{aile } 2 \end{cases}$$

$E(\varepsilon_1 \varepsilon_2) = 0 \Rightarrow$ OK YOK

$$\begin{cases} \text{Tüketim}_1 = B_0 + B_1 \text{Gelir}_1 + \varepsilon_1 \Rightarrow \text{aile } 1 \\ \text{Tüketim}_2 = B_0 + B_1 \text{Gelir}_2 + \varepsilon_2 \Rightarrow \text{aile } 2 \end{cases}$$

$E(\varepsilon_i \varepsilon_j) \neq 0$ <u>OK VAR</u>

---

* Otokorelasyon denilince regresyon modellerinin hata terimleri için OK hatırlanır fakat her değişken için korelasyon hesaplanabilir.
* Değişkenler arasında bir ilişki olup olmadığını kovaryans ve onun bir türevi olan korelasyon katsayısı ifade eder.

$$Cov(X,Y) = E\left[X - E(X)\right]\left[Y - E(Y)\right]$$

→ bazen $\sigma_{x,y}$ olarak da ifade edilir.

arakütle için

örneklem ?

hesaplama

$$Cov(X,Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$Cov(X,X) = E\left[X - E(X)\right]\left[X - E(X)\right]$$

hesaplama

$$= E\left[X - E(X)\right]^2$$

$$= \frac{\sum (x_i - \bar{x})^2}{n} = \sigma_x^2$$

* Eğer kovaryans sıfırdan farklı ise değişkenler arasında lineer ilişki vardır.

* Kovaryans "−" ve "+" olması ilişkinin yönünü belirtir ama standard bir ölçüt olmadığından ilişkinin gücü hakkında bilgi vermez $-\infty < Cov < +\infty$

* Bu nedenle korelasyon katsayısı hesaplanır.

* Korelasyon katsayısı ± 1 arasında değer alan standart bir ölçüdür $-1 \leq Corr \leq +1$

r = korelasyon katsayısı

ρ = otokorelasyon katsayısı

$$r = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

ara kütle ✗ örneklem ?

$$r = \frac{E\left[X - E(X)\right]\left[Y - E(Y)\right]}{\sqrt{E\left[X - E(X)\right]^2}\sqrt{E\left[Y - E(Y)\right]^2}} \Rightarrow r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y}) / n}{\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}\sqrt{\frac{\sum (y_i - \bar{y})^2}{n}}}$$

* Otokorelasyon katsoyısı ise herhangi bir değişkenin kendinden önceki ya da sonraki gözlem değerleri ile ilişkisi olup olmadığını ve varsa gücünü gösterir.

ÖR: X değişkeni için X_i ve X_{i-s} arasındaki otokorelasyon katsoyısına bakalım.

Not: S = devre kaybı.

$X = \{1, 3, 4, 8, 20, 30\}$

s = 2 olsun

#	X_i	X_{i-2}
1	1	–
2	3	–
3	4	1
4	8	3
5	20	4
6	30	8

2 devre kaybı çünkü s=2

kullanılacak data

n−s gözlem kaldı

s = 1 olsun

#	X_i	X_{i-1}
1	1	–
2	3	1
3	4	3
4	8	4
5	20	8
6	30	20

anakütle

$$Cov(X_i, X_{i-s}) = E\left[X_i - E(X_i)\right]\left[X_{i-s} - E(X_{i-s})\right]$$

$$= \frac{\sum_{i=s+1}^{n} (X_i - \bar{X}_i)(X_{i-s} - \bar{X}_{i-s})}{n}$$

$$\rho_s = \frac{Cov(X_i, X_{i-s})}{\sqrt{\sigma^2_{X_i}} \cdot \sqrt{\sigma^2_{X_{i-s}}}}$$

Aşağıdakileri kullanılarak $\hat{\rho}_s$ yazılabilir

tahmin

$$\hat{Cov}(X_i, X_{i-s}) = \frac{\sum_{i=s+1}^{n} (X_i - \bar{X}_i)(X_{i-s} - \bar{X}_{i-s})}{n-1}$$

örneklem

$$\hat{\sigma}^2_{X_i} = \frac{\sum_{i=s+1}^{n} (X_i - \bar{X}_i)^2}{n-1} \qquad \hat{\sigma}^2_{X_{i-s}} = \frac{\sum_{i=s+1}^{n} (X_{i-s} - \bar{X}_{i-s})^2}{n-1}$$

* ρ_s => otokorelasyon fonksiyonu
↳ belli devre kayıpları için
$s = 1, 2, 3, \dots, M \Rightarrow \rho_1, \rho_2, \rho_3, \dots, \rho_M$

} tekrar döneceğiz.

Hata Terimleri için Otokorelasyon Katsayısı

Basit Regresyonda

$$y_t = \beta_0 + \beta_1 X_t + u_t \quad \forall, \ t = 1, \dots, n$$

n = gözlem sayısı
t = zaman ifade eden indeks.

OK YOK a) $Cov(u_t, u_{t-s}) = 0 \ ; \ s > 0$

OK VAR b) $Cov(u_t, u_{t-s}) \neq 0 \ ; \ s > 0$

* zaman serilerinde zaman t indeksi ile ifade edilir.
* s ile gecikmeden (devre kaybı) bahsedilir.

u_t	u_{t-s}	
u_1	–	kayıp
u_2	–	
u_3	–	
⋮	–	
u_s	–	
u_{s+1}	u_1	$n-s$ kalan gözlem sayısı
⋮	⋮	
u_t	u_{t-s}	
⋮		
u_n	u_{n-s}	

* s teorik olarak $1 \leq s \leq n-1$ olabilir. ama uygulamada çok büyük değildir. Eğer çok büyük olursa serbestlik derecesi düşer.

serbestlik derecesi → bağımsız bilgi sayısı

* iki önemli varsayımı hatırlayın

$$\left.\begin{array}{l} \text{GDR 5} \Rightarrow E(u_t \mid X) = 0 \\ \text{GDR 7} \Rightarrow Var(u_t \mid X) = \sigma_u^2 \end{array}\right\} \Rightarrow t\text{-ye göre değişmiyor}$$

* Korelasyon katsayısının formülünü kullanarak ve basit regresyondaki eğim parametresinin formülünü kullanarak ρ_s ve $\hat{\rho}_s$ değerlerini hesaplayalım

(1) Korelasyon katsayısının formülü ile

GDRS kullanıldı

$$\rho_s = \frac{Cov(u_t, u_{t-s})}{\sqrt{\sigma^2_{u_t}}\sqrt{\sigma^2_{u_{t-s}}}} = \frac{E[u_t - \underbrace{E(u_t)}_{=0}][u_{t-s} - \underbrace{E(u_{t-s})}_{=0}]}{\sigma^2_{u_t}}$$

GDR7 kullanıldı

$$\rho_s = \frac{E[u_t \cdot u_{t-s}]}{E[u_t - \underbrace{E(u_t)}_{=0}]^2} = \frac{E[u_t \cdot u_{t-s}]}{E[u_t^2]}$$

$$\hat{\rho}_s = \frac{\sum_{t=s+1}^{n}(u_t \cdot u_{t-s})}{\sum_{t=s+1}^{n} u_t^2}$$

(2) Basit regresyon eğim parametresi ile.

$$y_t = \beta_0 + \beta_1 x_t + u_t \qquad \forall\ t = 1, 2, 3, \dots, n$$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

u_t-leri regresyon ile ifade edelim

yeni regresyondaki hata terimi

$$u_t = \rho_s u_{t-s} + \varepsilon_t$$

AR(s) otoregresif model s. dereceden

$\varepsilon_t \sim N(0, \sigma^2_\varepsilon)$

$Cov(\varepsilon_t, \varepsilon_s) = 0$

$Cov(\varepsilon_t, u_{t-s}) = 0$

$$\hat{\rho}_s = \frac{\sum_{t=s+1}^{n}(u_{t-s} - \underbrace{\bar{u}_{t-s}}_{=0})(u_t - \underbrace{\bar{u}_t}_{=0})}{\sum_{t=s+1}^{n}(u_{t-s} - \underbrace{\bar{u}_{t-s}}_{=0})^2}$$

$$= \frac{\sum_{t=s+1}^{n} u_t \cdot u_{t-s}}{\sum (u_{t-s})^2} \cong \frac{\sum_{t=s+1} u_t u_{t-s}}{\sum_{t=s+1} u_t^2}$$

Sonuç: Korelasyon katsayısı $\hat{=}$ Regresyon modeli SEKK tahmincisi

---

Otokorelasyon Fonksiyonu

$$U_t = \rho_1 U_{t-1} + \varepsilon_{1t}$$
$$U_t = \rho_2 U_{t-2} + \varepsilon_{2t}$$
$$U_t = \rho_3 U_{t-3} + \varepsilon_{3t}$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$U_t = \rho_m U_{t-m} + \varepsilon_{mt}$$

$\hat{\rho}_i$-ler otokorelasyon fonksiyonunu oluşturur.

$i = 1, 2, \ldots, m$

$\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_m$ → otokorelasyon katsayıları

Otoregresif süreç

* AR(1) ⇒ birinci dereceden otoregresif süreç ⇒ (s = 1)
  AR(2) ⇒ ikinci 〃 〃 〃 ⇒ (s = 2)
  AR(q) ⇒ q. 〃 〃 〃 ⇒ (s = q)

* Eğer otokorelasyon mevcut ise herhangi bir hata terimi s devre önceki / sonraki hata teriminden etkileniyor demektir. Bu ilişki otokorelasyon katsayısı kadar olacaktır.

* Ana model daha önce verilmişti.
* Hata terimi için AR modeli...

AR(1) hata terimi için

$$U_t = \rho U_{t-1} + \varepsilon_t$$

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$

$Cov(\varepsilon_t, \varepsilon_s) = 0$

$Cov(\varepsilon_t, U_{t-1}) = 0$

kovaryans formülünden ve GDR5-den

$$Cov(\varepsilon_t, U_{t-1}) = E\Big[\varepsilon_t - \underbrace{E(\varepsilon_t)}_{=0}\Big]\Big[U_{t-1} - \underbrace{E(U_{t-1})}_{=0}\Big]$$

$$= E[\varepsilon_t . U_{t-1}] = 0$$

teorik olarak

$$-1 \leq \rho \leq 1$$

peki AR(1) modelde doğru mu?

* Hata terimi için AR modelinin her iki tarafının da varyansını alalım.

$$u_t = \rho u_{t-1} + \varepsilon_t$$

$$Var(u_t) = Var(\rho u_{t-1} + \varepsilon_t) \quad Cov(u_{t-1}, \varepsilon_t) = 0$$

$$Var(u_t) = Var(\rho u_{t-1}) + Var(\varepsilon_t) + 2\rho\, Cov(u_{t-1}, \varepsilon_t)$$

$$Var(u_t) = \rho^2 Var(u_{t-1}) + Var(\varepsilon_t)$$

$$\sigma_{u_t}^2 = \rho^2 \sigma_{u_{t-1}}^2 + \sigma_{\varepsilon_t}^2 \quad Var(u_t) = \sigma_{u_t}^2 = \sigma_{u_{t-1}}^2 \text{ her } t \text{ için}$$

$$\sigma_{u_t}^2 = \rho^2 \sigma_{u_t}^2 + \sigma_{\varepsilon_t}^2$$

$$\sigma_{u_t}^2 = \frac{\sigma_\varepsilon^2}{(1-\rho^2)} \Rightarrow \sigma_{u_t}^2 \text{-nin tanımlanabilmesi için}$$

$$\underbrace{1-\rho^2 > 0 \Rightarrow 1 > \rho^2}$$

$$\Downarrow$$

$$-1 < \rho < 1$$

* Sonuç olarak AR(1) model için denge koşulu $-1 < \rho < 1$ olarak verilir. Eğer $\rho = -1$ ya da $\rho = 1$ olur ise patlayan bir sistem oluşur.

* Şimdi de u_t ve u_{t-s} arasındaki kovaryansı bulalım.

$$u_t = \rho u_{t-1} + \varepsilon_t$$

$$u_{t-1} = \rho u_{t-2} + \varepsilon_{t-1}$$

$$u_{t-2} = \rho u_{t-3} + \varepsilon_{t-2}$$

$$\vdots$$

① $\Rightarrow u_t = \rho[\rho u_{t-2} + \varepsilon_{t-1}] + \varepsilon_t$

$$u_t = \rho^2 u_{t-2} + \rho \varepsilon_{t-1} + \varepsilon_t$$

② $u_t = \rho^2[\rho u_{t-3} + \varepsilon_{t-2}] + \rho \varepsilon_{t-1} + \varepsilon_t$

$$u_t = \rho^3 u_{t-3} + \rho^2 \varepsilon_{t-2} + \rho \varepsilon_{t-1} + \varepsilon_t$$

genel form

$$u_t = \rho^s u_{t-s} + \rho^{s-1}\varepsilon_{t-s+1} + \rho^{s-2}\varepsilon_{t-s+2} + \dots + \rho \varepsilon_{t-1} + \varepsilon_t$$

* Her iki tarafı da U_{t-s} ile çarpıp beklenen değerini alalım.

$$U_t = \rho^s U_{t-s} + \rho^{s-1} \varepsilon_{t-s+1} + \dots + \rho \varepsilon_{t-1} + \varepsilon_t$$

$$E[U_t . U_{t-s}] = E[\rho^s U_{t-s} U_{t-s} + \rho^{s-1} \varepsilon_{t-s+1} . U_{t-s} + \dots + \rho \varepsilon_{t-1} U_{t-s} + \varepsilon_t U_{t-s}]$$

$$\underbrace{E[U_t . U_{t-s}]}_{= Cov(U_t, U_{t-s})} = \rho^s \underbrace{E[U_{t-s}^2]}_{= \sigma^2_{ut}} + \rho^{s-1} \underbrace{E[\varepsilon_{t-s+1} U_{t-s}]}_{=0} + \dots + \rho \underbrace{E[\varepsilon_{t-1} U_{t-s}]}_{=0} + \underbrace{E[\varepsilon_t U_{t-s}]}_{=0}$$

GDR 5   GDR 7   GDR 5   GDR 5   GDR 5

$$Var(U_{t-s}) = E[U_{t-s} - \underbrace{E(U_{t-s})}_{=0}]^2$$

$$= E[U_{t-s}^2]$$

$$= \sigma^2_{ut-s}$$

$$= \sigma^2_{ut}$$

$$Cov(U_t, U_{t-s}) = E[U_t - \underbrace{E(U_t)}_{=0}][U_{t-s} - \underbrace{E(U_{t-s})}_{=0}]$$

$$= E[U_t . U_{t-s}]$$

$$Cov(U_t, U_{t-s}) = \rho^s \sigma^2_{ut}$$

$$Corr(U_t, U_{t-s}) = \frac{Cov(U_t, U_{t-s})}{\sigma^2_{ut}} = \rho^s$$

yani s devre kaybı kadar üssel alınır.

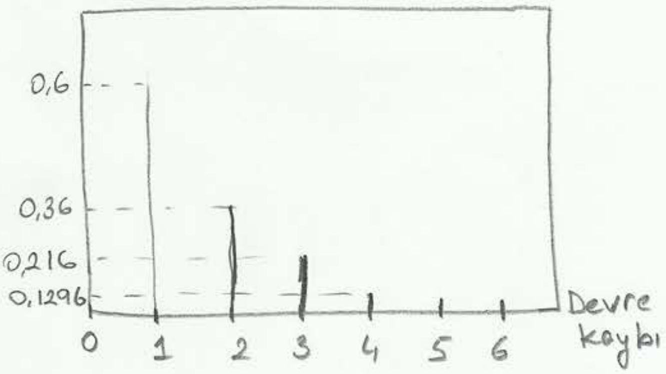
sonuç $\Rightarrow \rho_s = \rho^s \Rightarrow$ işte bu otokorelasyon fonksiyonudur. Farklı s devre kayıplarına göre değişir.

* Otokorelasyon katsayısı U_t, U_{t-1} ; U_t, U_{t-2} ; U_t, U_{t-3} ; ... için hesaplanırsa gecikmelere göre hesaplanan katsayıların oluşturduğu fonksiyona otokorelasyon fonksiyonu denir.

* Bu fonksiyonun grafiği çizilebilir. Bu grafiğe korelogram denir. Korelogram gecikme uzunluğuna göre otokorelasyon katsayılarını gösteren bir grafiktir.

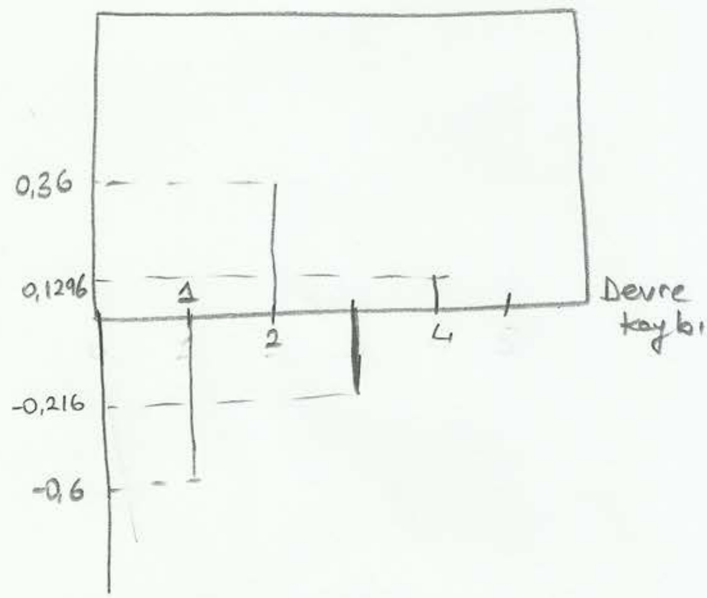
ÖR:

$\rho = 0,6$ varsayalım

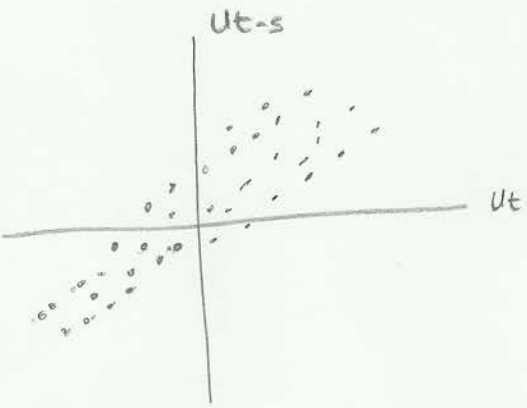


★ Azalarak giden lineer ilişki var yön hep pozitif

$\rho = -0,6$ varsayalım

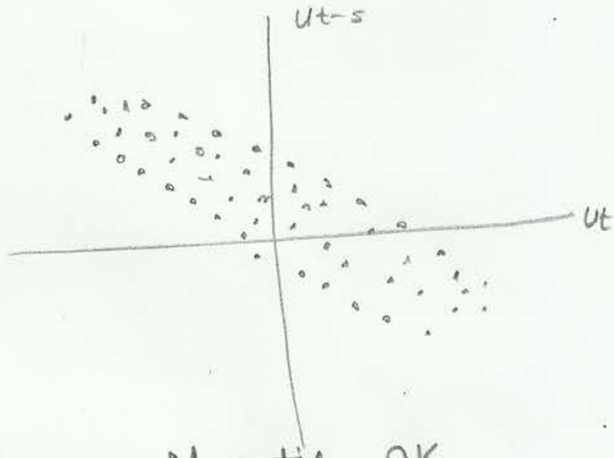


★ Azalarak giden lineer ilişki var fakat yönü negatif ve pozitif olarak değişiyor



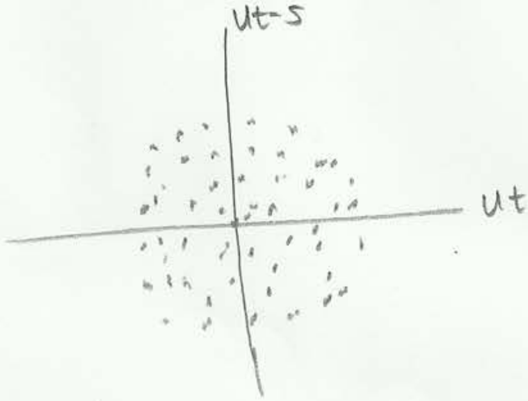
Pozitif OK

$0 < \rho < 1$



Negatif OK

$-1 < \rho < 0$



OK YOK

$\rho = 0$