

Sosyal Bilimlerde İstatistik

Bölüm 7: Korelasyon (İlişki) Analizi

Nihan Acar Denizli¹

¹Mimar Sinan Güzel Sanatlar Üniversitesi



1 Korelasyon Analizi

- Pearson Korelasyon Katsayısı (Parametrik Korelasyon Analizi)
- Spearman Korelasyon Katsayısı (Parametrik Olmayan Korelasyon Analizi)

2 İlginç Korelasyon Örnekleri



Korelasyon Analizi

Korelasyon: İki ya da daha fazla değişken arasındaki ilişkiyi ifade eder. Bir X değişkenine ait değerler değişirken Y değişkenine ait değerler de değişiyor ise bu iki değişken arasında bir ilişki/korelasyon olduğu söylenebilir. Bir değişkenin aldığı değerler artarken diğer değişkenin aldığı değerler artabilir ya da azalabilir.

Örnekler:

- *Öğrencilerin derse devamının artması ile derste başarı puanının artması.*
- *Bir işyerinde çalışanların işten memnuniyetinin azalması ile birlikte iş yerindeki performanslarının düşmesi.*



Korelasyon Analizi

Korelasyon Katsayısı: İki değişken arasındaki ilişkinin yönünü ve gücünü belirten değer korelasyon katsayısı olarak adlandırılır ve " r " ile gösterilir.

Korelasyon katsayısı, (r): $-1 < r < +1$ aralığında değerler alır.

- r , 0'a eşit ise iki değişken arasında anlamli bir ilişki yoktur.
- r , 0'a yakın ise iki değişken arasında **zayıf** bir ilişki vardır.
- r , +1'e yakın ise iki değişken arasında **kuvvetli ve pozitif yönde** bir ilişki vardır. (**Aynı yönlü**)
- r , -1'e yakın ise iki değişken arasında **kuvvetli ve negatif yönde** bir ilişki vardır. (**Ters yönlü**)



Korelasyon Analizi

Korelasyon Katsayısı (r) 'nın aldığı pozitif/negatif değerlere göre ilişkinin gücü:

0 - 0.3 ise **zayıf** ilişki

0.3 - 0.7 ise **orta** ilişki

0.7- 1 arası ise **kuvvetli** ilişki



Pearson Korelasyon Katsayısı (Parametrik)

Normal dağılıma sahip en az eşit aralıklı ölçüm düzeyinde iki nicel değişken arasındaki doğrusal ilişkinin yönü, kuvveti ve öneminin ölçülmesinde kullanılır.

$$r_{XY} = \frac{\sum XY - (\sum X \sum Y / n)}{\sqrt{(\sum X^2 - (\sum X)^2 / n) (\sum Y^2 - (\sum Y)^2 / n)}}$$



Pearson Korelasyon Katsayısı (Parametrik)

Normal dağılıma sahip en az eşit aralıklı ölçüm düzeyinde iki nicel değişken arasındaki doğrusal ilişkinin yönü, kuvveti ve öneminin ölçülmesinde kullanılır.

$$r_{XY} = \frac{\sum XY - (\sum X \sum Y / n)}{\sqrt{(\sum X^2 - (\sum X)^2 / n) (\sum Y^2 - (\sum Y)^2 / n)}}$$

- Değişkenler en az eşit aralıklı ölçülmüş ise,



Pearson Korelasyon Katsayısı (Parametrik)

Normal dağılıma sahip en az eşit aralıklı ölçüm düzeyinde iki nicel değişken arasındaki doğrusal ilişkinin yönü, kuvveti ve öneminin ölçülmesinde kullanılır.

$$r_{XY} = \frac{\sum XY - (\sum X \sum Y / n)}{\sqrt{(\sum X^2 - (\sum X)^2 / n) (\sum Y^2 - (\sum Y)^2 / n)}}$$

- Değişkenler en az eşit aralıklı ölçülmüş ise,
- Normalliğin sağlanması durumunda,



Pearson Korelasyon Katsayısı (Parametrik)

Normal dağılıma sahip en az eşit aralıklı ölçüm düzeyinde iki nicel değişken arasındaki doğrusal ilişkinin yönü, kuvveti ve öneminin ölçülmesinde kullanılır.

$$r_{XY} = \frac{\sum XY - (\sum X \sum Y / n)}{\sqrt{(\sum X^2 - (\sum X)^2 / n) (\sum Y^2 - (\sum Y)^2 / n)}}$$

- Değişkenler en az eşit aralıklı ölçülmüş ise,
- Normalliğin sağlanması durumunda,
- $n > 30$ olması durumunda,



Pearson Korelasyon Katsayısı (Parametrik)

Normal dağılıma sahip en az eşit aralıklı ölçüm düzeyinde iki nicel değişken arasındaki doğrusal ilişkinin yönü, kuvveti ve öneminin ölçülmesinde kullanılır.

$$r_{XY} = \frac{\sum XY - (\sum X \sum Y / n)}{\sqrt{(\sum X^2 - (\sum X)^2 / n) (\sum Y^2 - (\sum Y)^2 / n)}}$$

- Değişkenler en az eşit aralıklı ölçülmüş ise,
- Normallığın sağlanması durumunda,
- $n > 30$ olması durumunda,
- İki değişken arasındaki doğrusal ilişkiyi ölçer.



Spearman Korelasyon Katsayısı (Parametrik Olmayan)

Değişkenlerin normal dağılmadığı, sıralı ya da aralıklı ölçülen iki değişken arasındaki ilişkinin/ birlikteliğin ölçülmesinde kullanılır.

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

D_i : X değişkeninin i. gözleminin rankı ile Y değişkeninin i. gözleminin rankı arasındaki fark ($\text{Rank}(X_i) - \text{Rank}(Y_i)$) .



Spearman Korelasyon Katsayısı (Parametrik Olmayan)

Değişkenlerin normal dağılmadığı, sıralı ya da aralıklı ölçülen iki değişken arasındaki ilişkinin/ birlikteliğin ölçülmesinde kullanılır.

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

D_i : X değişkeninin i. gözleminin rankı ile Y değişkeninin i. gözleminin rankı arasındaki fark ($\text{Rank}(X_i) - \text{Rank}(Y_i)$) .

- Değişkenlerden en az biri sıralı (ordinal) ya da likert ölçeği ile ölçülmüş ise,



Spearman Korelasyon Katsayısı (Parametrik Olmayan)

Değişkenlerin normal dağılmadığı, sıralı ya da aralıklı ölçülen iki değişken arasındaki ilişkinin/ birlikteliğin ölçülmesinde kullanılır.

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

D_i : X değişkeninin i. gözleminin rankı ile Y değişkeninin i. gözleminin rankı arasındaki fark ($\text{Rank}(X_i) - \text{Rank}(Y_i)$) .

- Değişkenlerden en az biri sıralı (ordinal) ya da likert ölçeği ile ölçülmüş ise,
- Normalliğin sağlanmaması durumunda,



Spearman Korelasyon Katsayısı (Parametrik Olmayan)

Değişkenlerin normal dağılmadığı, sıralı ya da aralıklı ölçülen iki değişken arasındaki ilişkinin/ birlikteliğin ölçülmesinde kullanılır.

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

D_i : X değişkeninin i. gözleminin rankı ile Y değişkeninin i. gözleminin rankı arasındaki fark ($\text{Rank}(X_i) - \text{Rank}(Y_i)$) .

- Değişkenlerden en az biri sıralı (ordinal) ya da likert ölçeği ile ölçülmüş ise,
- Normalliğin sağlanmaması durumunda,
- $n < 30$ olması durumunda,



Spearman Korelasyon Katsayısı (Parametrik Olmayan)

Değişkenlerin normal dağılmadığı, sıralı ya da aralıklı ölçülen iki değişken arasındaki ilişkinin/ birlikteliğin ölçülmesinde kullanılır.

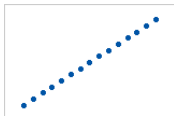
$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

D_i : X değişkeninin i. gözleminin rankı ile Y değişkeninin i. gözleminin rankı arasındaki fark ($\text{Rank}(X_i) - \text{Rank}(Y_i)$) .

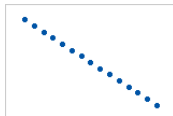
- Değişkenlerden en az biri sıralı (ordinal) ya da likert ölçeği ile ölçülmüş ise,
- Normalliğin sağlanmaması durumunda,
- $n < 30$ olması durumunda,
- İki değişken arasındaki monoton ilişkiyi ölçer.



İki nicel değişken arasındaki örnek ilişki grafikleri



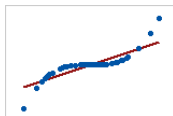
(a) Pearson $r = +1$, Spearman $r = +1$



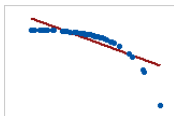
(b) Pearson $r = -1$, Spearman $r = -1$



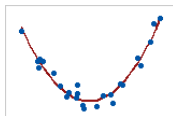
(c) Pearson $r = -0.093$, Spearman $r = -0.093$



(d) Pearson $r = +0.851$, Spearman $r = +1$



(d) Pearson $r = -0.799$, Spearman $r = -1$



(e) Pearson $r = 0$, Spearman $r = 0$



Korelasyon Analizinde Test Edilen Hipotez

H_0 : X ve Y değişkenleri arasında ilişki yoktur.

H_1 : X ve Y değişkenleri birbiri ile ilişkilidir.

H_0 : X ve Y değişkenleri arasındaki ilişki katsayısı sıfıra eşittir ($\rho_{XY} = 0$)

H_1 : X ve Y değişkenleri arasındaki ilişki katsayısı sıfırdan farklıdır ($\rho_{XY} \neq 0$).



Örnek araştırma problemleri (Gürbüz ve Şahin,2016)

H_0 : Bir işyerinde çalışanların iletişim kalitesi ile iş tatminleri arasında anlamlı bir ilişki yoktur.

H_1 : Bir işyerinde çalışanların iletişim kalitesi ile iş tatminleri arasında anlamlı bir ilişki vardır.



Örnek araştırma problemleri (Nuran Bayram,2009)

- Örnek 1: Sistolik Kan Basıncı (SKB) ile günlük içilen sigara sayısı arasında doğrusal bir ilişki var mıdır?



Örnek araştırma problemleri (Nuran Bayram,2009)

- Örnek 1: Sistolik Kan Basıncı (SKB) ile günlük içilen sigara sayısı arasında doğrusal bir ilişki var mıdır?

Pearson Korelasyon Katsayısı



Örnek araştırma problemleri (Nuran Bayram,2009)

- Örnek 1: Sistolik Kan Basıncı (SKB) ile günlük içilen sigara sayısı arasında doğrusal bir ilişki var mıdır?

Pearson Korelasyon Katsayısı

- Örnek 2: Öğrencilerin kağıtları teslim etme sırası ile sınavdan aldıkları puanlar arasında bir ilişki var mıdır?



Örnek araştırma problemleri (Nuran Bayram,2009)

- Örnek 1: Sistolik Kan Basıncı (SKB) ile günlük içilen sigara sayısı arasında doğrusal bir ilişki var mıdır?

Pearson Korelasyon Katsayısı

- Örnek 2: Öğrencilerin kağıtları teslim etme sırası ile sınavdan aldıkları puanlar arasında bir ilişki var mıdır?

Spearman Korelasyon Katsayısı



Önemli Not

- Korelasyon iki değişken arasındaki ilişkinin bir ölçüsüdür.
- Nedensellik ifade etmez. Nedensellik için korelasyon gereklidir ancak her korelasyon nedensellik içermemektedir.
- İki ya da daha fazla sayıda değişken arasındaki ilişkilerin açıklanması için "Regresyon Analizi" uygulanır.



1 Korelasyon Analizi

- Pearson Korelasyon Katsayısı (Parametrik Korelasyon Analizi)
- Spearman Korelasyon Katsayısı (Parametrik Olmayan Korelasyon Analizi)

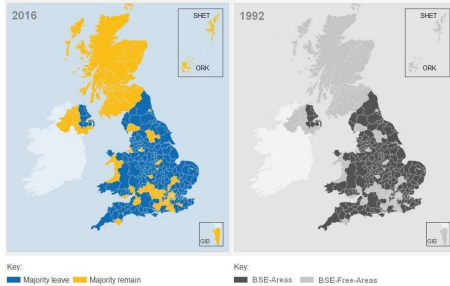
2 İlginç Korelasyon Örnekleri



İlginç Korelasyonlar

İki değişken arasında neden-sonuç ilişkisi olmadığı halde ilişki(korelasyon) bulunması durumuna **sahte korelasyon** denir.

Literatürde sahte korelasyona dair pek çok ilginç örnek bulunmaktadır.



Şekil: 1992 yılında deli dana hastalığından etkilenen yerler vs. 2016 Brexit referandumunda AB'den çıkmaya yönelik oy kullanan yerler



İlginç Korelasyonlar

■ Bebekleri leylek mi getiriyor?

Ülkelere gelen leylek sayısı ile yıllık çocuk doğumları arasında pozitif ilişki bulunmuştur.

Country	Area (km ²)	Storks (pairs)	Humans (10 ³)	Birth rate (10 ³ /yr)
Albania	28,750	100	3.2	83
Austria	83,860	300	7.6	87
Belgium	30,520	1	9.9	118
Bulgaria	111,000	5000	9.0	117
Denmark	43,100	9	5.1	59
France	544,000	140	56	774
Germany	357,000	3300	78	901
Greece	132,000	2500	10	106
Holland	41,900	4	15	188
Hungary	93,000	5000	11	124
Italy	301,280	5	57	551
Poland	312,680	30,000	38	610
Portugal	92,390	1500	10	120
Romania	237,500	5000	23	367
Spain	504,750	8000	39	439
Switzerland	41,290	150	6.7	82
Turkey	779,450	25,000	56	1576

Table 1. Geographic, human and stork data for 17 European countries



İlginç Korelasyonlar

■ Bebekleri leylek mi getiriyor?

Ülkelere gelen leylek sayısı ile yıllık çocuk doğumları arasında pozitif ilişki bulunmuştur.

Country	Area (km ²)	Storks (pairs)	Humans (10 ³)	Birth rate (10 ³ /yr)
Albania	28,750	100	3.2	83
Austria	83,860	300	7.6	87
Belgium	30,520	1	9.9	118
Bulgaria	111,000	5000	9.0	117
Denmark	43,100	9	5.1	59
France	544,000	140	56	774
Germany	357,000	3300	78	901
Greece	132,000	2500	10	106
Holland	41,900	4	15	188
Hungary	93,000	5000	11	124
Italy	301,280	5	57	551
Poland	312,680	30,000	38	610
Portugal	92,390	1500	10	120
Romania	237,500	5000	23	367
Spain	504,750	8000	39	439
Switzerland	41,290	150	6.7	82
Turkey	779,450	25,000	56	1576

Table 1. Geographic, human and stork data for 17 European countries

Ekonomik gelişme ve şehirleşme nedeniyle çocuk doğum sayısı da leylek sayısı da azalmıştır. Ancak bu iki değişkenin birlikte azalması birinin diğerine neden olduğu anlamına gelmemektedir.



İlginç Korelasyonlar

- Şehre gelen turistler suç olaylarının artmasına mı neden olmaktadır?

Bir sahil şehrine gelen turist sayısı ile vuku bulan suç oranları dönemsel olarak birlikte artmakta ve azalmaktadır.



İlginç Korelasyonlar

- Şehre gelen turistler suç olaylarının artmasına mı neden olmaktadır?

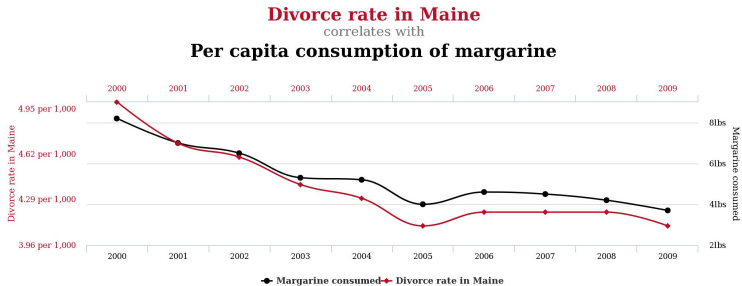
Bir sahil şehrine gelen turist sayısı ile vuku bulan suç oranları dönemsel olarak birlikte artmakta ve azalmaktadır.

Suç olaylarının artışı bahar ve yaz mevsiminden ötürü ya da şehrin büyüklüğü gibi farklı nedenlerden etkilenmiş olabilir. Bu örnekte sadece turist sayısı ve suç olayları aynı yönde değişmektedir.



İlginç Korelasyonlar

- Maine'de margarin tüketiminin azalması mı boşanmaya sebep olur yoksa boşananlar mı daha az margarin tüketir? ($r = 0.99$)

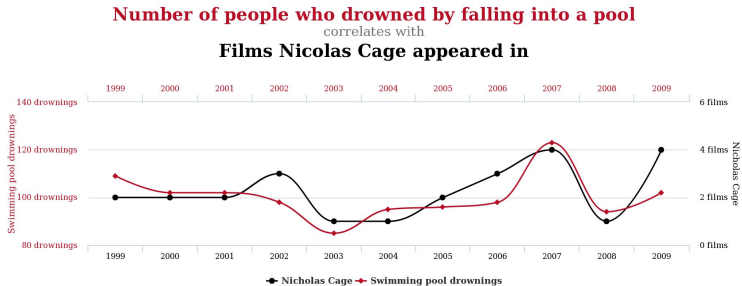


tylervigen.com



İlginç Korelasyonlar

- Nicolas Cage'in film çekmesi o yıla ait sudaki boğulma oranlarını etkiliyor mu? ($r = 0.66$)



tylervigen.com



KAYNAKLAR



Nuran Bayram (2009). Sosyal Bilimlerde SPSS ile Veri Analizi, Ezgi Kitabevi.



Nuran Bayram (2016). Veri Analizi (Excel ve SPSS Uygulamalarıyla Birlikte) , Ezgi Kitabevi.



Abdullah Can (2014). SPSS ile Bilimsel Araştırma Sürecinde Nicel Veri Analizi, Pegem Akademi Yayınları, 2. Baskı.



Sait Gürbüz ve Faruk Şahin (2016). Sosyal Bilimlerde Araştırma Yöntemleri Felsefe-Yöntem-Analiz, 3.Baskı, Seçkin Yayıncılık.



Kazım Özdamar (2013). Paket Programlar İle İstatistiksel Veri Analizi, Nisan Kitabevi.



R. Mark Sirkin (2005). Statistics for the Social Sciences, SAGE.



Yahşi Yazıcıoğlu ve Samiye Erdoğan (2014). SPSS Uygulamalı Bilimsel Araştırma Yöntemleri, Detay Yayıncılık.



Robert Mathews(2000). Storks Deliver Babies, *Teaching Statistics*,22(2), 36-38.



The Minitab Express Support.

<http://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/a-comparison-of-the-pearson-and-spearman-correlation-methods/>



Spurious Correlations

<http://www.tylervigen.com/spurious-correlations>

