

**İSTATİSTİK II**

**ORTAK DERS**

**YRD. DOÇ. DR. LEYLA İŞBİLEN YÜCEL (Bölüm 1-2)**

**DOÇ. DR.ÖZLEM YORULMAZ (Bölüm 3-14)**

**İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ**

**İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ**

**ORTAK DERS**

**İSTATİSTİK II**

**YRD. DOÇ. DR. LEYLA İŞBİLEN YÜCEL (Bölüm 1-2)**

**DOÇ. DR.ÖZLEM YORULMAZ (Bölüm 3-14)**

# ÖNSÖZ

Bu ders günlük hayatta, özellikle ekonomik alanda karşılaşılan çeşitli problemleri istatistiksel metotlarla ele alma, çeşitli iddiaları sınama, çıkarsama yapma, ilişkileri modelleme ve değerlendirme konularında bilgi vermeyi amaçlar. İlk iki bölümde sürekli dağılım, örnekleme kavramı gibi temel konulara değinilmiş, sonraki bölümlerde ise istatistiğin çıkarımsal konularına (aralık tahmini, hipotez testleri, regresyon analizi, ki-kare analizi, endeksler) yer verilmiştir. Bu kitap fazla sayıda örnek uygulamalarıyla, istatistiksel kavramları ve teknikleri anlamayı kolaylaştırmak hedefiyle derlenmiştir.

DOÇ. DR.ÖZLEM YORULMAZ

# 8. REGRESYON ANALİZİ I

**Bu Bölümde Neler Öğreneceğiz?**

Bu bölümde iki değişken arasındaki ilişkiyi modellemeyi öğreneceğiz. Regresyon analizi iki değişken arasındaki nedensel ilişkinin fonksiyonel yapısını verir.

**Bölüm Hakkında İlgi Oluşturan Sorular**

1. Bir topluluktaki kişilerin gelir ve tüketimleri üzerine araştırılma yapılmak istenmiştir. Aynı gelire sahip kişilerin tüketimlerinin farklı olabildiği görülmüştür. Topluluktaki tüm bireyler gözetilerek gelir ve tüketim arasındaki ilişki nasıl modellenebilir?
2. Suç oranı ve işsizlik arasındaki beklenen ilişki nasıl modellenebilir?

**Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri**

|  |  |  |
| --- | --- | --- |
| **Konu** | **Kazanım** | **Kazanımın nasıl elde edileceği veya geliştirileceği** |
| Regresyon analizi, sabit ve eğim parametrelerinin belirlenmesi | İki değişken arasındaki ilişki, değişkenlerinin bağımlı ve bağımsız yapıda olmaları dikkate alınarak modellenir. | Metinler, çözümlü problemler. |

**Anahtar Kavramlar**

* Regresyon Analizi
* Bağımlı değişken
* Bağımsız değişken
* Sabit ve eğim katsayısı

**Giriş**

Bu bölümde regresyon analizini ele alarak iki değişken arasındaki ilişkiyi modelleyeceğiz. Değişkenlerden hangisinin etkileyen, hangisinin etkilenen değişken olduğunu belirleyerek bir modelleme yapacağız.

## 8.1. Regresyon Analizi

Regresyon analizi değişkenler arasındaki ilişkinin araştırılmasında kullanılan istatistiksel bir araçtır. Bu araçla, bir değişkenin diğer değişken üzerindeki nedensel ilişkisi araştırılır. İncelenen ilişkideki değişkenler aralarındaki ilişki göz önüne alınarak değişkenler bağımlı ve bağımsız olarak isimlendirilir. Önceki derslerde de değinildiği gibi istatistiğin öncelikli ilgi alanını rastlantı değişkeninin davranışını bir modelle tahmin etmek oluşturur. Davranışı tahmin edilecek olan rastlantı değişkeni bir diğer değişken(ler)in fonksiyonu olarak gösterilebilir ve bu değişken bağımlı olarak isimlendirilir ve Y ile gösterilir. Bağımlı değişkeni etkileyen değişken ise X ile gösterilir ve bağımsız değişken olarak isimlendirilir. Yani, bağımlı değişken, bağımsız değişken(ler) tarafından açıklanmaya çalışılır ve açıklayıcı değişkenlerin modelde bilinen sabitler olduğu varsayılır.

Model şu şekilde gösterilir; 

burada b0 sabit katsayıyı gösterir buna başlangıç parametresi de denir, b1 ise eğim parametresidir. X’deki 1 birimlik değişmenin Y üzerinde nasıl bir değişim yaptığını gösterir. Denklemdeki  ise daha sonra da açıklanacağı gibi hata terimine karşılık gelir.

Örneğin, tüketim ve gelir üzerine yapılan bir çalışmada bağımsız değişken gelir, bağımlı değişkense tüketimdir ya da bir hastaya uygulanan ilacın dozu ve hastanın iyileşme süreci çalışmasında bağımsız değişken ilacın dozu ve bağımlı değişkense hastanın iyileşme süreci olur.

Regresyon analizi, bilinen gerçekleşen olaylar sonucunda elde edilen bulgulardan yola çıkarak gelecekteki olaylarla ilgili tahmin yani öngörü yapılmasını sağlar. Regresyon modelinde amaç, koşullar değiştiğinde bağımlı değişkenin ortalamasının E(Yi) ‘nin nasıl değiştiğini tanımlamaktır.

Değişkenler arasındaki ilişki deterministik (kesin) ya da olasılıksal (istatistiksel) olarak isimlendirilir. Deterministik ilişkileri açıklamada kullanılan matematiksel fonksiyondan farklı olarak, regresyon analizi olasılıksal ilişkileri açıklar. Arz-talep, gelir-tüketim gibi ilişkilerin modellenmesinde deterministik ilişkiye sahip değişkenler yerine istatistiksel ilişkiye sahip değişkenler kullanılır.

Söz konusu ilişkide bağımsız değişken sayısının bir tane olması basit regresyonla, birden fazla olması ise çoklu regresyonla açıklanır. Regresyon modeli, doğrusal yapıda olabileceği gibi parabolik, logaritmik, üstel biçimli de olabilir. Modelde bir bağımlı ve bir bağımsız değişken söz konusu olduğunda, yani basit regresyon söz konusu iken serpilme diyagramı kullanılarak uygun model seçimi yapılabilir. Serpilme diyagramı, i. gözlemin bağımlı değeri yi ve bağımsız değeri xi olmak üzere tüm gözlem çiftleri üzerinden, her ikili yani Y ve X değişkenlerinin aldığı tüm değerler birer nokta ile temsil edilecek şekilde çizilir. Diyagramdaki dağılıma bakılarak uygun model belirlenir.









Yukarıdaki serpilme diyagramlarında noktaların ortasından geçecek olan eğri dikkate alınır ve bu eğri incelenen ilişki biçimi hakkında bilgi verir. Buna göre, ilk çizimde noktaların bir doğru etrafında toplandığı söylenebilir ve değişkenler arasında aynı yönlü doğrusal bir ilişkinin varlığından söz edilebilir. İkinci çizimse ters yönlü doğrusal bir ilişkinin varlığını gösterir. Üçüncü çizimde doğrusal olmayan bir ilişkinin varlığı söz konusudur. Son çizim dikkate alındığında ise bir eğri oluşturmak mümkün görünmemektedir, değişkenler arasında bir ilişkinin olmadığı sonucuna varılır.

Serpilme diyagramı çizimi sonrasında uygun modele karar verilir ve modeldeki parametreler tahmin edilir. Tahmin sürecinde çeşitli kriterler doğrultusunda kullanılan yöntemler olmakla beraber, burada bu yöntemlerden sadece En Küçük Kareler Yöntemine (EKK) değinilecektir.

## 8.2. Regresyon Katsayılarının Tahmini

EKK ile bulunacak eğrinin her (xi,yi) gözlem çiftine karşılık gelen nokta ile bu noktanın EKK ile elde edilecek eğri üzerindeki dik izdüşümü arasındaki farklar toplamı sıfır olmalıdır. Bu farklar, yani Yi değerlerinin regresyon doğrusuna olan uzaklığı, daha sonrada bahsedileceği gibi ‘hata’ olarak isimlendirilir. Yi değerlerinin regresyon doğrusu üzerindeki görüntüsü  (tahmini Yi) ile arasındaki fark hataya karşılık gelir. İdeal regresyon doğrusu, bu farkların karelerinin toplamını , minimum verenle elde edilir.



Hata kareler toplamının minimum olabilmesi için, sabit ve eğim parametrelerine göre türevleri alınarak sıfıra eşitlenir:



Bu işlem sonrasında elde edilen denklemler Normal Denklemler olarak isimlendirilir:





Cramer yöntemine göre bu denklemler çözüldüğünde;





elde edilir. Ancak eğim ve sabit parametre tahminlerini ortalamadan sapmalar üzerinden giderekte belirlemek mümkündür:





Eğim parametresi bu x ve y üzerinden gidilerek belirlenir.





olduğundan eşitlikler düzenlenirse;



b1=

Sabit parametre için,

 olduğundan, eşitliğin her iki tarafı n’e bölünürse,



elde edilir.

**Örnek:** Aşağıda bir sınıftaki öğrencilerin muhasebe ve matematik derslerine ait veri bulunmaktadır. Muhasebe dersinden başarının matematik dersinden başarıya bağımlı olup olmadığını sınamak için regresyon denklemini oluşturunuz.

|  |  |
| --- | --- |
| **Muhasebe** | **Matematik** |
| 1 | 2 |
| 2 | 3 |
| 3 | 5 |
| 5 | 6 |
| 6 | 7 |
| 7 | 10 |
| 8 | 7 |
| 8 | 8 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Y** | **X** | **Y^2** | **X^2** | **YX** | **X-Xort=x** | **Y-Yort=y** | **xy** | **x^2** |
| 1 | 2 | 1 | 4 | 2 | -4 | -4 | 16 | 16 |
| 2 | 3 | 4 | 9 | 6 | -3 | -3 | 9 | 9 |
| 3 | 5 | 9 | 25 | 15 | -1 | -2 | 2 | 1 |
| 5 | 6 | 25 | 36 | 30 | 0 | 0 | 0 | 0 |
| 6 | 7 | 36 | 49 | 42 | 1 | 1 | 1 | 1 |
| 7 | 10 | 49 | 100 | 70 | 4 | 2 | 8 | 16 |
| 8 | 7 | 64 | 49 | 56 | 1 | 3 | 3 | 1 |
| 8 | 8 | 64 | 64 | 64 | 2 | 3 | 6 | 4 |
| **40** | **48** | **252** | **336** | **285** | **0** | **0** | **45** | **48** |

|  |
| --- |
| **Yort=5** |
| **Xort=6** |

Normal denklemlerden elde edilen eşitlikler kullanılarak katsayılar şöyle tahmin edilmiştir:









Ortalamadan sapmalarla,





yine aynı donuca ulaşılmıştır. Tahmin edilen regresyon denklemi şöyledir:



**Örnek:**

Aşağıda bir eyaletteki suç ve işsizlik oranlarına ilişkin veri mevcuttur.

|  |  |
| --- | --- |
| **işşizlik oranı** | **suç oranı** |
| 0,8 | 3 |
| 1,4 | 6 |
| 2,3 | 7 |
| 3,5 | 15 |
| 4,5 | 19 |

İşsizlik ve suç işleme oranları arasındaki ilişkiyi gösteren regresyon denklemini oluşturunuz.

Sorunun çözümünün ilk aşamasında bağımlı ve bağımsız değişkenleri belirleyelim. İşssizlik oranı bağımsız değişkendir (x), suç oranını etkiler ki bu da bağımlı değişken (Y) olarak isimlendirilir. İlgili kolonların toplam ve çarpımlarının toplamına ait bilgiler aşağıdaki gibi özetlenmiştir.





İşsizlik oranı 1 birim arttığında suç oranı 4.365 birim artar.

Ortalamadan sapmalar serisi yerine orijinal seriden hareket edilirse eğim katsayısı şöyle bulunacaktır:



**Örnek:**

Bir firmanın reklam harcamaları ve satış rakamlarına ilişkin veri mevcuttur. Regresyon denklemini oluşturunuz

|  |  |
| --- | --- |
| **Reklam**  **Harca.** | **Satışlar** |
| 1.6 | 6 |
| 2.8 | 12 |
| 4.6 | 14 |
| 7 | 30 |
| 9 | 38 |

Bağımlı değişken satışlardır., Reklam harcamaları ise bağımsız değişkendir.





Reklam harcamaları 1 birim artarsa satış 4.37 birim artar.

Diğer yaklaşımla:



**Örnek:** Aşağıdaki x,y serilerinden hareketle regresyon denklemini bularak yorumlayınız.

|  |  |
| --- | --- |
| **x** | **y** |
| 10.2 | 7 |
| 8.4 | 5 |
| 6.2 | 4 |
| 4.2 | 1 |
| 11 | 8 |

**Örnek:**Aşağıda verilen x,y ikilisinden hareketle regresyon denklemini bulunuz.

|  |  |
| --- | --- |
| **x** | **y** |
| 20 | 12 |
| 19 | 10 |
| 17 | 9 |
| 16 | 8 |
| 13 | 6 |

**Bölüm Soruları**

1. X={2,3,4,5}, Y={6,8,8,10} X ve Y sırasıyla bağımsız ve bağımlı değişkenler ise regresyon denklemi nasıl ifade edilir?,
2. Uygulanan ilacın dozu ve ilacın etki süresi aşağıdaki gibidir. Regresyon denklemini bularak yorumlayınız.

|  |  |
| --- | --- |
| **Doz** | **İyileşme süresi** |
| 1 | 8 |
| 2 | 8 |
| 3 | 6 |
| 4 | 4 |
| 5 | 2 |

**BÖLÜM 8 ÇOKTAN SEÇMELİ SORULARI**

1- Aşağıdaki grafiklerden hangisi kullanılarak değişkenler arasındaki ilişki hakkında fikir sahibi olunabilir?

1. Pasta Grafiği
2. Frekans Poligonu
3. Serpilme diyagramı
4. Histogram
5. Dal-yaprak grafiği

2-Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

I)EKK Yöntemi, artıkların karelerinin toplamını en küçük yapabilen regresyon denklemindeki sabit ve eğim parametrelerinin elde edilmesini sağlayan bir yöntemdir.

II)Korelasyon analizi ile değişkenler arasındaki ilişkinin yönü, büyüklüğü ve matematiksel modeli belirlenir.

III) Regresyon analizinde bağımlı değişken ve gözlem değerlerinden elde edilen tahmini Y değerleri arasındaki farka hata terimi adı verilir.

a)I ve III b) Yalnız I c) Yalnız II d) Yalnız III e) I ve II

3- Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

1. Regresyon analizinde değişkenlerden biri artarken diğeri azalıyorsa eğim katsayısı negatif değerlidir.
2. Basit regresyon analizinde serpilme diyagramına bakılarak ilişkinin yönünü belirlemek mümkündür.
3. Regresyon analizinde hata terimlerinin toplamı sıfırdır.
4. Yalnız I b) Yalnız II c)Yalnız III d)I ve II e) I, II ve III

4-X bağımsız ve Y ise bağımlı değişkeni göstermek üzere,  regresyon denkleminin eğim parametresi aşağıdakilerden hangisine eşit olur?

a)1.58 b) 1.64 c) 2.11 d)-1.58 e)-1.64

5-Yukarıdaki bilgiden hareketle regresyon denkleminin sabit parametresi aşağıdakilerden hangisine eşit olur?

a)8.66 b) 7.66 c) -12.15 d) -8.66 e)-2.23

6-Y=8.66+1.58X regresyon denklemi için aşağıdakilerden hangisi söylenemez?

a)X artarken Y artar

b)Y artarken X artar

c)X ve Y arasında aynı yönlü ilişki vardır

d)X de 1 birimlik artıs Y’de ortalama 1.58 birimlik artmaya neden olur.

e) X ile Y arasında doğrusal artan bir ilişki vardır

7-Regresyon modelinde gözlenen değerler ve tahmin arasındaki farkın kareli ortalamasına ne denir?

1. Değişim katsayısı b)Standart sapma c)Korelasyon katsayısı d)Eğim katsayısı e)Tahminin standart hatası

8- X bağımsız ve Y ise bağımlı değişkeni göstermek üzere,

 regresyon denkleminin eğim parametresi aşağıdakilerden hangisine eşit olur?

1. -2.3 b) -1.5 c)0.8 d)1.5 e)2.3

9) Yukarıdaki bilgiden hareketle regresyon denkleminin sabit parametresi aşağıdakilerden hangisine eşit olur?

a) 60 b) 40 c) 20 d) -20 e)-40

10-Y=-40+1.5X regresyon denklemi için aşağıdakilerden hangisi söylenemez?

a)X artarken Y artar

b)Y artarken X azalır

c)X ve Y arasında aynı yönlü ilişki vardır

d)X de 1 birimlik artıs Y’de ortalama 1.5 birimlik artmaya neden olur.

e) X ile Y arasında doğrusal artan bir ilişki vardır

Yanıtlar

1)C 2)A 3)E 4)A 5)A 6)B 7)E 8)D 9)E 10) B

**Bu Bölümde Ne Öğrendik Özeti**

Çeşitli olayları, olguları açıklamakta kullandığımız değişkenler ve bunlar arasındaki ilişkinin anlaşılmasında kullanılan regresyon analizini, basit doğrusal regreyon denkleminde yer alan sabit ve eğim parametrelerinin elde edilmesini, yorumlanmasını öğrendik.

# 9. REGRESYON ANALİZİ II

**Bu Bölümde Neler Öğreneceğiz?**

Bu bölümde tahmin edilen anakütle regresyon parametrelerinin istatistiksel olarak anlamlılığını sınayacağız.

**Bölüm Hakkında İlgi Oluşturan Sorular**

1. Tahmin edilmiş olan gelir tüketim denkleminde, gelir gerçekten istatistiksel olarak anlamlı mıdır?
2. Tahmin edilen regresyon denkleminde gerçekleşen hata nedir?

**Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri**

|  |  |  |
| --- | --- | --- |
| **Konu** | **Kazanım** | **Kazanımın nasıl elde edileceği veya geliştirileceği** |
| Regresyon parametrelerinin anlamlılığının sınanması | Regresyon denkleminde tahminin standart hatasını bulmak, parametrelerin anlamlılığını sınamak | Metinler, çözümlü problemler. |

**Anahtar Kavramlar**

* Standart hata
* Eğim katsayısının/sabit katsayının standart hatası
* Eğim katsayısının/standart hatanın anlamlılığı

**Giriş**

Önceki bölümde basit doğrusal regresyon modelinde sabit ve eğim parametrelerinin nasıl belirlendiğini öğrendik bu bölümde ise söz konusu parametre tahminlerinin istatistiksel olarak anlamlılığını sınayacağız.

## 9.1. Regresyon Doğrusunun Özellikleri ve Hata (Artık) Kavramı

Regresyon doğrusu denklemi  olmak üzere burada , başlangıç terimidir, modelin sabit değeri olarak da bilinir. X=0 iken

’nin değeridir.  doğrunun eğimine karşılık gelir. X’deki bir birim değişmenin ’deki değişimine karşılık gelir. Burada ve  parametrelerdir. Bu parametlere veri üzerinden b0 ve b1 tahminleriyle ulaşılır.

Y, bağımlı değişkenin elde edilen her bir gözlemi (Yi) ana kütle ortalaması  olan bir ana kütleden gelen rastlantı değişkeni olduğu varsayılır. Yi gözleminin ’ den sapması hata terimi olarak isimlendirilir ve  ile gösterilir.

Belirlenen X değerine karşılık gelen tahmini Y değeri ya da diğer bir ifadeyle X’in belirli değeri için, Y’nin anakütle ortalamasının tahmini  şöyle bulunur:





Gözlenen  değeri ile tahmini değer yani  değeri karşılaştırıldığında model ile veri arasındaki uyum için bir büyüklük elde edilir buna **artık** denir.



Artık, tahmin edilen modelle veri arasındaki farka karşılık gelir eğer modelde sabit terim varsa artıkların toplamı sıfırdır.

 eşitliği dikkate alındığında,   gözlemini açıklayabilen fakat   gözlemini açıklayamayan kısımdır.Belirli varsayımlar sağlandığında artıklar tahmini hatalar olarak kabul edilir. Hatırlanacağı üzere hata,

 şeklindedir.

### Regresyon Doğrusunun Özellikleri

-Artıkların toplamı sıfırdır



-Artık Kareler toplamı minimumdur.

-Gözlenen Y ve tahmini Y değerlerinin toplamı birbirine eşittir.



Dolayısıyla ortalamaları da eşittir.

- Bağımsız değişkenle artıkların çarpım toplamları sıfırdır.



-Tahmini Y değerleri ile artıkların çarpımlar toplamı sıfırdır.



- Regresyon doğrusu daima  noktasından geçer.

Gözlenen Yi, tahmini Yi ve hata arasındaki ilişki aşağıdaki temsili çizimde de görülebilmektedir.



**Standart Hata**

Regresyon denklemini standart hatasının tahmini hata kareler toplamının n-2’ye bölünerek karekökünün alınmasıyla bulunur. Buna tahminin standart hatası da denilir.

 şeklinde tanımlanır.

Regresyon Denkleminin Varsayımları

* Bağımlı değişken tesadüfi değişkendir, normal dağılır.
* Hatalar tesadüfidir ve normal dağılır.
* Hatalar birbirinden bağımsızdır (otokorelasyon yoktur)
* Her bağımsız değişkenin değerlerine ait olan bağımlı değişken değerlerinin alt kümelerinin varyansları birbirine eşittir (Homoskedastisite)
* Bağımsız değişkenler arasında basit doğrusal ilişki olmamalıdır.

## 9.2. Regresyon Denklemi Parametrelenin Hipotez Testi ve Güven Aralığı

EKK yöntemi ile  regresyon denkleminin  ve parametrelerinin tahmini değerleri ve  örneklemden örnekleme değişen birer rastlantı değişkenidir. Bu rastlantı değişkenlerinin ortalama ve varyansları sırayla aşağıdaki gibidir.

Regresyon analizindeki varsayımlardan ilki hatırlanacağı üzere bağımlı değişkenin Y’nin normal dağıldığı varsayımı idi. Gerek  ve gerekse  bağımlı değişkenle doğrusal ilişki içinde olduğundan bu iki rastlantı değişkenin de normal dağıldığı sonucuna varılır. Bu sonuca dayanarak  ve için  ve  üzerinden yola çıkarak hipotez testi ve güven aralıklarını tanımlamak mümkündür.  ve  rastlantı değişkenlerinin normal dağılması halinde önceki konulardan hatırlanacağı üzere

 ve 

rastlantı değişkenleri de normal dağılır.

 ve  tahminlerinin standart hataları şöyledir:





Varyans bilinmediği zaman z yerine t eşitlikleri kullanılır:

Bu istatistikler n-2 serbestlik dereceli t dağılır. Bu eşitliklerden yola çıkarak sırasıyla sabit ve eğim paramereleri için güven aralıkları şöyledir:



Sabit parametre için yukarıdaki eşitsizliklerin düzenlenmesiyle bulunan yaklaşım eğim parametresi için de yapıldığında bulunan güven aralığı şöyledir:



İlgilenilen hipotez testi sabit katsayı için aşağıdaki gibi düzenlendiğinde, aşağıdaki test istatistiği kullanılır:





Alternatif hipotez yukarıda çift yönlü düzenlenmiş olmakla beraber aşağıdaki gibi tek yönlü de düzenlenebilir:





Alternatif hipotezin çift yönlü olması halinde eğer bulunan kritik değer ve tablo değeri arasında aşağıdaki gibi bir ilişki varsa H0 reddedilir.

 veya 

Alternatif hipotez eğer aşağıdaki gibi tek yönlü ise,



 halinde H0 hipotezi reddedilir.

Ve son olarak da alternatif hipotez

 şeklindeyse

 olmalı halinde H0 hipotezi reddedilir.

Eğim parametresinin hipotez testinde genellikle aşağıdaki şekilde düzenlenir. Katsayının genellikle sıfıra eşitliği sınanır. Çünkü katsayının sıfıra eşitliği bağımlı değişkenin bağımsız değişkene doğrusal bağımlı olmadığı anlamına gelir. Sıfır hipotezi ve olası alternatif hipotezler aşağıdaki gibi düzenlenir;







Burada kullanılan istatistik aşağıdaki gibidir.



Tablodan bulunan kritik değerle yukarıdaki istatistikten hesaplanan değerin kıyaslanması sabit parametre için hipotez testinde değinildiği gibidir, aynıdır.

**Örnek:** Önceki örnekte bir sınıftaki öğrencilerin matematik ve muhasebe derslerindeki başarıları arasındaki ilişki incelenmişti. Regresyon denkleminin paremetrelerinin %99 ve %95 güvenle bulundukları aralıkları belirleyiniz ve katsayların anlamlılığını aynı güven düzeyinde sınayınız.

Öncelikle tahminin standart hatasını bulalım;

(2)=1.25

(3)=2.188

(5)=4.063

(6)=5

(7)=5.938

(10)=8.75

(7)=5.938

(8)=6.875

|  |  |  |
| --- | --- | --- |
| **Y** |  |  |
| 1 | 1,25 | 0,0625 |
| 2 | 2,188 | 0,035344 |
| 3 | 4,063 | 1,129969 |
| 5 | 5 | 0 |
| 6 | 5,938 | 0,003844 |
| 7 | 8,75 | 3,0625 |
| 8 | 5,938 | 4,251844 |
| 8 | 6,875 | 1,265625 |
| **40** | **40,002** | **9,811626** |





 olacağından %99 ve %95 güvenle parametrenin yer alacağı aralıklar şöyledir:





(t0.01;6 = 3.707)





(t0.05;6 = 2.447)

Sabit parametrenin standart hatası, güven aralığı aşağıdaki gibidir:











Eğim parametresinin anlamlılığı için kurulan hipotezler ve testi şöyledir:





elde edilen değer her iki anlamlılık düzeyindeki kritik değerlerle sınandığında sıfır hipotezi reddedilir. Yani, Y bağımlı değişkeni X’e bağımlıdır.

Sabit katsayının testi benzeri yaklaşımla incelendiğinde,





Her iki anlamlılık düzeyinde sıfır hipotezi kabul edilir.

-Bağımlı değişkenin ortalama değeri için güven aralığı bulunmak istendiğinde, belirli bir x değerinde bağımlı değişkenin ortalamasının standart hatası aşağıdaki biçimde bulunur:



- Eğer belirli bir x değerine karşılık gelen tek bir nokta için güven aralığı oluşturmak istenirse kullanılacak standart hata şöyledir:



**Örnek**: Yukarıdaki örneğe ait verileri kullanarak matematik dersinden 9 alan öğrencilerin muhasebe dersinden alacakları ortalama not için %95 güven aralığı oluşturunuz.





Aralık tahmini için:





**Örnek:** Yukarıdaki örneğe ait verileri kullanarak matematik dersinden 9 alan bir öğrencinin muhasebe dersinden alacağı not için %95 güven aralığı oluşturunuz.





Aralık tahmini:





**Örnek:**

Önceki bölümde ele alınan suç ve işsizlik oranlarına ilişkin örnekteki parametrelerin anlamlılığını % 5 önem düzeyinde sınamak istersek;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | Y |  |  |  |
| **işşizlik oranı** | **suç oranı** | Ytahmin | (Y-Ytah) | (Y-Ytah)^2 |
| 0,8 | 3 | 2,592 | 0,408 | 0,166464 |
| 1,4 | 6 | 5,211 | 0,789 | 0,622521 |
| 2,3 | 7 | 9,1395 | -2,1395 | 4,57746 |
| 3,5 | 15 | 14,3775 | 0,6225 | 0,387506 |
| 4,5 | 19 | 18,7425 | 0,2575 | 0,066306 |
| 12,5 | 50 | 50,0625 | -0,0625 | 5,820258 |







Eğim parametresi anlamlıdır, güven aralığı sıfırı kapsamadığı gibi hesaplanan t istatistik değeri de tablo değerinden büyüktür (9.6> 3.82)



Sabit parametresi ise anlamsızdır. Aralık sıfırı kapsar ve test istatistiği sonucu kritik değerden küçük çıkmıştır.

**Örnek:** Aşağıdaki veriden hareketle regresyon denklemi eğim katsayısının anlamlılığını değerlendiriniz.

|  |  |
| --- | --- |
| **x** | **y** |
| 20 | 6 |
| 19 | 8 |
| 17 | 9 |
| 16 | 10 |
| 13 | 12 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **x** | **y** | **ytah** | **e** | **e^2** |
| 20 | 6 | 6.6 | -0.6 | 0.36 |
| 19 | 8 | 7.4 | 0.6 | 0.36 |
| 17 | 9 | 9 | 0 | 0 |
| 16 | 10 | 9.8 | 0.2 | 0.04 |
| 13 | 12 | 12.2 | -0.2 | 0.04 |
| 85 | 45 | 45 | -3.6E-15 | 0.8 |

****

3 sd li t tablo değeri -3.82 olduğundan eğim parametresinin anlamsız olduğunu ifade eden sıfır hipotezi reddedilir.

**Bölüm Soruları**

**1-**Aşağıda matematik dersine çalışma süresi ve alınan notlara ilişkin bilgi mevcuttur. Regresyon denklemini kurarak eğim ve sabit parametrelerin anlamlılığını sınayınız.

|  |  |
| --- | --- |
| **çalışma saati** | **alınan not** |
| 3 | 30 |
| 5 | 45 |
| 6 | 55 |
| 8 | 65 |
| 9 | 80 |

-Tahmin edilen notların toplamı yaklaşık olarak alınan notlara eşit mi?

-Hataların toplamı yaklaşık olarak sıfır mı?

**2-** Uygulanan ilaç dozu ve iyileşme sürelerine ilişkin regresyon denkleminin parametrelerinin anlamlılığını sınayınız.

|  |  |
| --- | --- |
| **Doz** | **İyileşme süresi** |
| 1 | 8 |
| 2 | 8 |
| 3 | 6 |
| 4 | 4 |
| 5 | 2 |

**BÖLÜM 9 ÇOKTAN SEÇMELİ SORULARI**

1. Aşağıdaki 1-2-3 ve 4 numaralı soruları şu soru metninden hareketle çözünüz:

X bağımsız ve Y bağımlı değişkene karşılık gelmek üzere,



1. Regresyon denkleminin eğim katsayısı aşağıdakilerden hangisine eşittir?
2. 1.45 b)0.69 c) -0.80 d)-1.29 e)-1.45
3. Regresyon denkleminde tahminin standart hatası nedir?
4. 0.896 b)0.799 c)0.639 d)0.489 e)0.398
5. Kitle için eğim katsayısının yer alacağı sınırlar % 95 olasılıkla aşağıdakilerden hangisine eşittir?

a)-0.72- 0.80 b)-0.52-1.90 c)0.40-1.96 d)0.58-2.16 e)0.70-3.28

1. Bir önceki soruda elde edilen güven aralığından hareketle aşağıdakilerden hangisi söylenemez?
2. X değişkeninin Y değişkeni üzerinde etkisi yoktur.
3. Eğim katsayısı %5 önem düzeyinde anlamlı değildir.
4. Y değişkenin X üzerinde etkisi vardır
5. Yalnız I b) Yalnız II c) I ve II d) Yalnız III e) II ve III
6. 5-6-7-8-9-10 numaralı soruları aşağıdaki metinden hareketle yanıtlayınız. X bağımsız ve Y bağımlı değişkene karşılık gelmek üzere;



5- Sabit ve eğim katsayıları sırasıyla aşağıdakilerden hangisine eşit olur?

a)(2, 6) b) (1, 3) c) (6, 2) d) (4,5) e)(2,5)

6- Regresyon denkleminde hata karelerinin toplamı aşağıdakilerden hangisine eşittir?

1. 22 b) 19 c) 17 d)12 e)11

7-Regresyon denkleminde eğim katsayısının standart hatası aşağıdakilerden hangisine eşittir?

a)0.91 b)0.78 c) 0.67 d)0.17 e)0.14

8- Regresyon denkleminde sabit katsayının standart hatası aşağıdakilerden hangisine eşit olur?

a)0.91 b)0.78 c) 0.67 d)0.57 e)0.4

9- Eğim katsayısının anlamlılığını sınamak için yapılan t testi skoru aşağıdakilerden hangisine eşittir?

1. 16.23 b)14.78 c)15.47 d)14.06 e) 11.98

10- Eğim katsayısının anlamlılığı için kullanılan t testi skoru aşağıdakilerden hangisine eşittir (alfa=0.05)?

1. 2.899 b) 2.306 c) 1.95 d) 1.65 e)1.28

Yanıtlar

1)A 2)B 3)B 4)D 5)C 6)D 7)D 8)A 9)E 10)B

**Bu Bölümde Ne Öğrendik Özeti**

Bu bölümde regresyon parametrelerinin anlamlı olup olmadığını sınadık.

# 10. KORELASYON KATSAYISI-DETERMİNASYON KATSAYISI

**Bu Bölümde Neler Öğreneceğiz?**

Bu bölümde regresyon denkleminde bağımlı değişkenin ne oranda bağımsız değişken tarafından açıklandığını determinasyon (Belirlilik) katsayısı ile değerlendireceğiz. Ayrıca iki değişken arasındaki ilişkinin yönünü ve büyüklüğünü korelasyon katsayısı ile inceleyeceğiz.

**Bölüm Hakkında İlgi Oluşturan Sorular**

1. Enflasyon ve faiz oranı arasında nasıl bir ilişki var? Aynı yönlü mü? Ters yönlü mü? Güçlü mü? Zayıf mı?
2. Gelir ve tüketim arasındaki regresyon denkleminde, tüketimin gelir tarafından açıklanma yüzdesi nedir?

**Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri**

|  |  |  |
| --- | --- | --- |
| **Konu** | **Kazanım** | **Kazanımın nasıl elde edileceği veya geliştirileceği** |
| Determinasyon katsayısı, Korelasyon katsayısı | İki değişken arasındaki ilişkinin yönünün ve büyüklüğünün belirlenmesi, regresyon denkleminin gücünün belirlenmesi. | Metinler, çözümlü problemler |

**Anahtar Kavramlar**

* Determinasyon katsayısı
* Korelasyon katsayısı
* Regresyon denkleminin gücü

**Giriş**

Bu bölümde regresyon denkleminde yer alan bağımlı değişkenin bağımsız değişken tarafından açıklanma yüzdesini determinasyon katsayısı ile değerlendireceğiz. Determinasyon katsayısı, regresyon denkleminin gücünü veren bir büyüklüktür. Konularımıza dahil ettiğimiz korelasyon katsayısı ise değişkenler arasındaki ilişkinin nedenselliği ile değil de ilişkinin sadece yönü ve büyüklüğü hakkında bilgi veren bir istatistiktir.

## 10.1. Regresyon Denkleminin Gücü

Regresyon analiziyle bağımsız değişken(ler)le bağımlı değişkendeki değişme açıklanır. Değişimi açıklamanın tam olması beklenemez, örnekleme ve regresyon denklemin yapısına bağlı olarak farklı oranlarda açıklanır ancak istenen yüksek bir açıklanma oranıdır. Bu orana regresyon denkleminin gücü- determinasyon katsayısı- belirlilik katsayısı denilir ve  ile gösterilir.

Determinasyon katsayısı bağımlı değişkendeki değişimin ne kadarının bağımsız değişken(ler) tarafından açıklandığını gösterir ve  arasındadır. Katsayının 1’e yaklaşması açıklama gücünün yüksekliğini gösterir.

Aşağıdaki çizimden de görülebileceği gibi bir gözlem için şu eşitlik yazılabilir:





bu eşitlikte şöyle bir düzenleme yapılır.



ve her iki tarafın karesi alınırsa



elde edilir. Bu eşitlikte bir takım değişiklikler yapmak için normal denklemler hatırlansın.



denkleminde her iki taraf n’e bölünürek yukarıdaki karesel eşitliğin sağ tarafındaki ifade yeniden yazılır:



Karesel eşitlik yeniden düzenlenerek şu hale getirilir:



Burada, sol taraf regresyon denklemindeki toplam değişmeye karşılık gelir, eşitliğin sağ tarafının ilk kısmı yani regresyon parametresinin yer aldığı kısım açıklanabilen değişimdir. Kalan diğer kısımsa açıklanamayan değişim olarak isimlendirilir.

Toplam Değişme= Açıklanabilen Değişme + Açıklanamayan Değişme

Kısaca şu notasyonla da gösterilebilir:

SST=SSR+SSE

İşte determinasyon katsayısı burada, toplam değişim içerisindeki açıklanabilen değişim yüzdesidir:



**Örnek:** Önceki örnekten yola çıkarak regresyon denkleminin gücünü belirleyiniz.

|  |  |  |  |
| --- | --- | --- | --- |
| **Y** |  |  |  |
| 1 | 1,25 | 0,0625 | 16 |
| 2 | 2,188 | 0,035344 | 9 |
| 3 | 4,063 | 1,129969 | 4 |
| 5 | 5 | 0 | 0 |
| 6 | 5,938 | 0,003844 | 1 |
| 7 | 8,75 | 3,0625 | 4 |
| 8 | 5,938 | 4,251844 | 9 |
| 8 | 6,875 | 1,265625 | 9 |
| **40** | **40** | **9,811626** | **52** |



**Örnek:**

Önceki iki bölümde ele alınan işsizlik ve suç oranı istatistikleri verisinden hareketle determinasyon katsayısını bulunuz.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | Y |  |  |  |
| **işşizlik oranı** | **suç oranı** | **Ytahmin** | **(Y-Ytah)** | **(Y-Ytah)^2** |
| 0,8 | 3 | 2,592 | 0,408 | 0,166464 |
| 1,4 | 6 | 5,211 | 0,789 | 0,622521 |
| 2,3 | 7 | 9,1395 | -2,1395 | 4,5774603 |
| 3,5 | 15 | 14,3775 | 0,6225 | 0,3875063 |
| 4,5 | 19 | 18,7425 | 0,2575 | 0,0663062 |
| 12,5 | 50 | 50,0625 | -0,0625 | 5,8202578 |





Suç oranı istatistiklerinin açıklanmasında işsizlik oranı oldukça önemli bir yer tutar, %97 oranında açıklar.

## 10.2. Korelasyon Katsayısı

Regresyon analizi ile bağımlı ve bağımsız değişken(ler) arasında ilişkinin varlığı, bu ilişkinin gücü sorgulandı ve değişkenler arasındaki ilişki matematiksel bir modelle açıklandı. Bu yeni konuda, iki değişken arasındaki ilişkinin ölçüsünün, büyüklüğü ve yönü incelenecek. İki değişken arasındaki ilişkinin ölçüsü değerlendirilirken Korelasyon Katsayısı (r) kullanılır. Korelasyon analizinde regresyon analizindeki gibi serpilme diyagramı çizilerek değişkenler arasındaki ilişki için genel bir bilgi edinilir. Korelasyon katsayısı –1 ve +1 arasında değiişm gösterir.

Aşağıda farklı korelasyon ilişkisini gösteren çizimlere yer verilmiştir (Gujarati, Temel Ekonometri 1998).



0-0.5 arası ilişki zayıf ilişki olarak değerlendirilir. Korelasyon katsayısının “-“ olması değişkenler arasındaki ters yönlü ilişkiye işaret ederken, bu katsayının “+” olması aynı yönlü ilişkiye işaret eder.

Çizim (a) da aynı yönlü tam doğrusal bir ilişki söz konsudur.

Çizim (b)’de ise ters yönlü tam doğrusal bir ilişki söz konusudur.

Çizim (c) ve (d) tama yakın sırayla aynı yönlü ve ters yönlü ilişki söz konusudur.

(e) ve (f) çizimlerinde ise korelasyon katsayısı sıfıra yakındır. Çizim (g)’de ilişki yoktur, korelasyon katsayısı sıfırdır. Son çizim (h) de de korelasyon katsayısı sıfırdır ama bu durum değişkenler arasında ilişki olmamasından değil ilişkinin doğrusal olmamasından kaynaklanır.

Regresyon analizinde değişkenler arasındaki ilişkinin matematiksel ifadesi öne çıkar, değişkenler bağımlı ve bağımsız değişkenler olarak ayrılır, bağımlı değişken, rastlantı değişkenidir bağımsız değişkense kontrol edilebilen değişkendir. Korelasyon analizinde ise değişkenlerde bağımlı ve bağımsız ayırımı yoktur tüm değişkenler rastlantı değişkenidir.

Regresyon analizinin korelasyon analizinden bir diğer farklı yaklaşımıysa, regresyon analizinin değişkenler arasındaki neden sonuç önbilgisini sınaması fakat korelasyon katsayısının her zaman nedensel bir ilişkinin göstergesi olmamasıdır.

Korelasyon katsayılar ilgilenilen değişkenlerin türlerine göre sınıflanabilir. Buna göre eğer ilgilenilen değişkenler nitelse kullanılan korelasyon katsayılarından bazıları şöyledir:

-Phi Katsayısı

-Cramer Katsayısı

-Lambda Katsayısı

-Gamma Katsayısı

-Kendall’ın tau Katsayısı

İlgilenilen değişkenler nicelse kullanılan korelasyon katsayıları şöyle sıralanabilir:

-Pearson Korelasyon Katsayısı

-Spearman Korelasyon Katsayısı

Burada sıralanan korelasyon katsayılarına daha sonra üzerinde detaylı bir biçimde durulacak olan Kısmi Korelasyon katsayısını da eklemek mümkündür. Kısmi korelasyon katsayısı iki değişken arasındaki ilişkiyi gösterirken diğer değişkenlerin etkilerini dikkate almaz.

### 10.2.1. Pearson Korelasyon Katsayısı

İki değişken arasındaki doğrusal ilişkiyi gösteren bir diğer ölçütte hatırlanacağı gibi Kovaryans katsayısıdır. Ancak kovaryans katsayısı ölçü birimlerine bağlı bir büyüklüktür.



Korelasyon katsayısı ise kovaryans katsayısının standart sapma değerlerine oranlanmasıyla elde edilir ve ölçü birimlerinden bağımsızdır. Korelasyon katsayısı şöyle ifade edilir:



Anakütleye ait korelasyon katsayısı yukarıdaki gibi olmakla beraber örneğe ait olan korelasyon katsayısı ise şöyle gösterilir:







Korelasyon katsayısının regresyon denkleminin gücünü ifade eden determinasyon katsayısıyla ve eğim katsayısı ile arasında ilişki vardır. Şöyle ki;

Eğim katsayısı, ortalamadan sapmalarla  şeklinde gösterilmek üzere,



eğim katsayısı ve korelasyon katsayısı birbirine oranlanırsa,



ve sonrasında eşitliğin sağ tarafı ’e bölünürse,

 elde edilir.

Eğim katsayısı ve korelasyon arasındaki ilişki şöyle gösterilebilir:

Anlaşılacağı üzere korelasyon katsayısı ve regresyon katsayısının işaretleri aynı olacaktır.

Hatırlanacağı üzere regresyonun temel eşitliğinde, toplam değişme açıklanabilen ve açıklanamayan değişmeye eşittir.









Buradan açıkça basit regresyondaki determinasyon katsayısının korelasyon katsayısının karesine yani  eşit olduğu görülebilir.

**Örnek:**

Muhasebe dersinden başarının matematik dersinden başarıya bağlı olup olmadığını araştırmak için seçilen öğrencilere ait notlar aşağıdaki gibidir. Korelasyon katsayısını hesaplayınız.

|  |  |
| --- | --- |
| **Muhasebe** | **Matematik** |
| 1 | 2 |
| 2 | 3 |
| 3 | 5 |
| 5 | 6 |
| 6 | 7 |
| 7 | 10 |
| 8 | 7 |
| 8 | 8 |

Korelasyon katsayısı ve regresyon denklemi arasındaki ilişkiyi değerlendirmek için öncelikle regresyon denklemi bulunmak istenirse;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **y** | **x** | **y^2** | **x^2** | **xy** |
| 1 | 2 | 1 | 4 | 2 |
| 2 | 3 | 4 | 9 | 6 |
| 3 | 5 | 9 | 25 | 15 |
| 5 | 6 | 25 | 36 | 30 |
| 6 | 7 | 36 | 49 | 42 |
| 7 | 10 | 49 | 100 | 70 |
| 8 | 7 | 64 | 49 | 56 |
| 8 | 8 | 64 | 64 | 64 |
| **40** | **48** | **252** | **336** | **285** |

Normal denklemler:

40=8b0+48b1

285=48b0+336b1

b1=0.9375

b0=-0.625





Regresyon denkleminin eğimini veren katsayı 0.9375’tir. Bu katsayının pozitif olması aynı yönlü ilişki olduğunu gösterir. Nitekim bulunun korelasyon katsayısı da pozitifit. Korelasyon katsayısı, r, oldukça yüksektir. Buradan muhasebe ve matemetik dersleri arasında güçlü doğrusal bir ilişki olduğu sonucuna varılır.

Tahmini y değerleri üzerinden bulunacak determinasyon katsayısının karekökü de aynı sonucu verecektir. Şöyle ki;

|  |  |  |  |
| --- | --- | --- | --- |
| **y** | **ytah** | **(y-ytah)^2** | **(y-yort)^2** |
| 1 | 1,25 | 0,0625 | 16 |
| 2 | 2,188 | 0,035344 | 9 |
| 3 | 4,063 | 1,129969 | 4 |
| 5 | 5 | 0 | 0 |
| 6 | 5,938 | 0,003844 | 1 |
| 7 | 8,75 | 3,0625 | 4 |
| 8 | 5,938 | 4,251844 | 9 |
| 8 | 6,875 | 1,265625 | 9 |
| **40** | **40** | **9,811626** | **52** |





**Örnek:**

Önceki iki bölümde ele alınan işsizlik ve suç oranı istatistikleri verisinden hareketle korelasyon katsayısını bulunuz.

|  |  |
| --- | --- |
| X | Y |
| **işşizlik oranı** | **suç oranı** |
| 0,8 | 3 |
| 1,4 | 6 |
| 2,3 | 7 |
| 3,5 | 15 |
| 4,5 | 19 |
| 12,5 | 50 |





**Örnek:**

Aşağıda bir firmanın reklam harcamalarına ve satışlarına ait bilgi mevcuttur. Satış rakamlarının reklam harcamalarına göre regresyon denklemini bularak eğim katsayısının anlamlılığını sınayın ve regresyon denkleminin gücünü değerlendirin.

|  |  |
| --- | --- |
| **Reklam Harcamaları(X)** | **Satışlar(Y)** |
| 10 | 3 |
| 20 | 4 |
| 30 | 6 |
| 40 | 7 |
| 50 | 10 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Y** | **X** | **Y^2** | **X^2** | **YX** | **Y-Yort=y** | **X-Xort=x** | **xy** | **x^2** |
| 3 | 10 | 9 | 100 | 30 | -3 | -20 | 60 | 400 |
| 4 | 20 | 16 | 400 | 80 | -2 | -10 | 20 | 100 |
| 6 | 30 | 36 | 900 | 180 | 0 | 0 | 0 | 0 |
| 7 | 40 | 49 | 1600 | 280 | 1 | 10 | 10 | 100 |
| 10 | 50 | 100 | 2500 | 500 | 4 | 20 | 80 | 400 |
| **30** | **150** | **210** | **5500** | **1070** | **0** | **0** | **170** | **1000** |









Reklam harcamalarındaki 1 birimlik değişme satıişlarda 0.17 birimlik değişime neden   
olur.

|  |  |  |  |
| --- | --- | --- | --- |
| **Y** |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| 3 | 2,6 | 0,16 | 9 |
| 4 | 4,3 | 0,09 | 4 |
| 6 | 6 | 0 | 0 |
| 7 | 7,7 | 0,49 | 1 |
| 10 | 9,4 | 0,36 | 16 |
| 30 | 30 | 1,1 | 30 |



Reklam harcamaları satışları %96 oranında açıklamaktadır.

Eğim parametresi için güven aralığı şöyledir:







Eğim parametresinin testinde bulunan değer tablo kritik değeri 3.18’den büyük olduğu için bağımsız değişkenin ağımlı değişkeni açıklama da yetersiz olduğunu savunan sıfır hipotezi reddedilir.





**Örnek:** Bir ilacın dozu ve iyileşme süresine ait seri aşağıdaki gibidir. Korelasyon katsayısını ve regresyon denklemini bulunuz.

|  |  |
| --- | --- |
| **İlaç Dozu(x)** | **İyileşme süresi** |
| 0 | 90 |
| 1 | 70 |
| 4 | 70 |
| 6 | 40 |
| 7 | 50 |
| 12 | 10 |





**Örnek:** Aşağıda yer alan x,y ikilisinden hareketle regresyon denklemini oluşturunuz, eğim katsayısının anlamlılığını değerlendiriniz, determinasyon ve korelasyon katsayısını bulunuz.

|  |  |
| --- | --- |
| **X** | **Y** |
| 2 | 18 |
| 3 | 22 |
| 5 | 25 |
| 6 | 26 |
| 8 | 30 |
| 9 | 32 |
| 12 | 40 |





****

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **X** | **Y** | ytah | y-ytah=e | hata kare |
| 2 | 18 | 18.543 | -0.543 | 0.294849 |
| 3 | 22 | 20.582 | 1.418 | 2.010724 |
| 5 | 25 | 24.66 | 0.34 | 0.1156 |
| 6 | 26 | 26.699 | -0.699 | 0.488601 |
| 8 | 30 | 30.777 | -0.777 | 0.603729 |
| 9 | 32 | 32.816 | -0.816 | 0.665856 |
| 12 | 40 | 38.933 | 1.067 | 1.138489 |
| 45 | 193 |  | -0.01 | 5.317848 |

****

****

**Bölüm Soruları**

**1-**Aşağıda matematik dersine çalışma süresi ve alınan notlara ilişkin bilgi mevcuttur. Determinasyon ve korelasyon katsayılarını hesaplayarak yorumlayınız.

|  |  |
| --- | --- |
| **Çalışma saati** | **Alınan not** |
| 3 | 30 |
| 5 | 45 |
| 6 | 55 |
| 8 | 65 |
| 9 | 80 |

**2-** Uygulanan ilaç dozu ve iyileşme sürelerine ilişkin bilgi mevcuttur. Determinasyon ve korelasyon katsayılarını bularak yorumlayınız.

|  |  |
| --- | --- |
| **Doz** | **İyileşme süresi** |
| 1 | 8 |
| 2 | 8 |
| 3 | 6 |
| 4 | 4 |
| 5 | 2 |

**BÖLÜM 10 ÇOKTAN SEÇMELİ SORULARI**

1-Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

1. Korelasyon katsayısı 2 değişken arasındaki ilişkinin yönünün ve büyüklüğünün belirlenmesinde kullanılır.
2. Korelasyon katsayısının mutlak değerce 1’e yaklaşması güçlü bir ilişki göstergesidir.
3. Determinasyon katsayısı korelasyon katsayısının karesine eşittir.

a)Yalnız I b) Yalnız II c)Yalnız III d)I ve II e)Hepsi

2- X ve Y iki değişkene karşılık gelmek üzere, aşağıdaki bilgiden hareketle elde edilecek olan pearson korelasyon katsayısı kaçtır?



a)0.93 b) 0.98 c)-0.87 d)-0.92 e)-0.98

3- 8 ‘er gözlemden oluşan X ve Y değişkenleri arasındaki korelasyon katsayısı aşağıdakilerden hangisine eşittir?



a)0.72 b)0.82 c)0.92 d)-0.72 e)-0.92

1. Aşağıdaki soruları burada verilen bilgiden hareketle yanıtlayınız:

6 adet gözlemden oluşan X bağımsız ve Y bağımlı değişkenlerine ait değerler şöyledir,



4- Regresyon denkleminin eğim katsayısı aşağıdakilerden hangisine eşittir?

1. -12.01 b) -9.76 c) -6.26 d)9.66 e)12.23

5- Regresyon denkleminde bağımlı değişken ve tahmin edilen bağımlı değişken arasındaki farkı gösteren hata terimlerinin toplamı aşağıdakilerden hangisine eşittir?

a)0 b)2 c)12 d)14 e)20

6)Regresyon denkleminde hata kareler toplamı aşağıdakilerden hangisine eşittir?

a) 0 b) 200 c)890 c)1280 d)1571 e)1600

7) Regresyon denkleminde determinasyon katsayısı kaçtır?

a) 0.78 b) 0.86 c) 0.91 d)0.93 e)0.98

8)Regresyon denkleminde X değişkeni Y değişkenin % kaçını açıklar?

a) 0.78 b) 0.86 c) 0.91 d)0.93 e)0.98

9-X ve Y değişkenleri arasındaki korelasyon katsayısı kaçtır?

a)0.95 b)0.90 c)0.88 d) 0.83 e)0.80

10- Regresyon denkleminde sabit katsayı kaçtır?

a)-3.18 b)-1.81 c)-0.78 d) 0.79 e)1.81

Yanıtlar

1. E 2) A 3)C 4)D 5)A 6)D 7)C 8)C 9)A 10)A

**Bu Bölümde Ne Öğrendik Özeti**

Determinasyon katsayısının ve korelasyon katsayısının hangi amaçla kullanıldığı, nasıl hesaplandığı ve yorumlandığı konuları ele alındı.

# 11. Kİ-KARE TESTLERİ

**Bu Bölümde Neler Öğreneceğiz?**

Ki kare testleri ile sayısal olmayan değişkenler arasındaki ilişkinin varlığı, iki ya da daha fazla sayıda örneklemin aynı anakütleden gelip gelmediği, bir seriye ait frekans dağılımının belli bir dağılıma uygun olup olmadığı sınamalarını yapmayı öğreneceğiz.

**Bölüm Hakkında İlgi Oluşturan Sorular**

1. Eğitim düzeyi ile izlenen televizyon programları arasında ilişki var mıdır?
2. Kadınların makyaj yapma sıklığı ile yaşları arasında bir ilişki var mıdır?
3. Annenin eğitim düzeyi çocuğun okuldaki başarısını etkiler mi?

**Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri**

|  |  |  |
| --- | --- | --- |
| **Konu** | **Kazanım** | **Kazanımın nasıl elde edileceği veya geliştirileceği** |
| Kikare Testleri | İki Nitel değişken arasında ilişkinin varlığını değerlendirmek | Metinler, çözümlü problemler |

**Anahtar Kavramlar**

* Ki kare Uyum testi
* Ki kare Bağımsızlık testi

## 11.1. Ki-Kare Testlerini Neden Kullanırız?

İstatistik I derslerinden hatırlanacağı üzere istatistikte değişkenler sadece nicel(sayısal) olanlarla sınırlı değildir, nitel değişkenler de istatistiğin konusunu oluşturur. Önceki derslerde nicel değişkenler tahmin teknikleri ve hipotez testlerinin konusunu oluşturdu. Zaman zaman analizlerde sayısal olmayan yani nitel değişkenler de dikkate alınabilir. Örneğin insanların eğitim düzeyi ve izledikleri programlar arasındaki ilişkinin varlığı araştırılmak istenebilir. Böyle bir durumda söz konusu 2 değişken de niteldir ve şıkları sırasıyla şu şekilde ifade edilebilir; “ilkokul-lise-üniversite”, “dizi-eğlence-yarışma”.

Ki kare testleri ile sayısal olmayan değişkenler arasındaki ilişkinin varlığı (ilişkinin varolmadığı sıfır hipotezi varsayımı ile), iki ya da daha fazla sayıda örneklemin aynı anakütleden gelip gelmediği, bir seriye ait frekans dağılımının belli bir dağılıma uygun olup olmadığı sınamaları yapılabilir. Ki-kare testleri üç başlık altında incelenir:

-Ki kare bağımsızlık

-Ki-kare homojenlik

-Ki-kare uyum

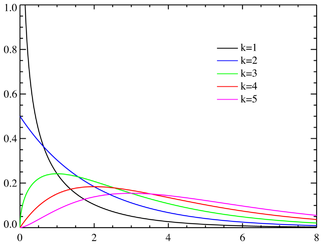
Ki kare testlerinde  dağılımı kullanılır. Kısaca bahsetmek gerekirse ki-kare dağılımı, X tesadüfi değişken dağılımından bir değer seçilip standart hale dönüştürülüp, karesinin alınmasıyla bulunur.

Standardize edilen X tesadüfi değişkeni, standart normal değişken Zi’ninaşağıdaki şekilde karesi alındığında ki-kare dağılımına dönüşür.





Ki-kare dağılımlarının serbestlik derecesi n ise şu şekilde gösterilir:



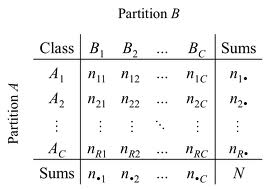
Ki-kare dağılımı sağa çarpıktır ve normalden daha diktir. N büyüdükçe merkezi limit teoreminin bir sonucu olarak diklik ve asimetri azalır ve dağılım normale yaklaşır.

Ki-kare değerleri 0 ile arasında değişir. Yukarıdaki tabloda yer alan k değerleri serbestlik derecesine karşılık gelmektedir.

## 11.2. Ki-Kare Bağımsızlık Testi

İncelenen birimlerin iki ya da daha fazla özelliğinin birbirlerinden bağımsız olup olmadıklarını sınamak amacıyla kullanılan bir testtir. Örneğin sigara içme alışkanlığı ve cinsiyet gibi, eğitim düzeyi ve alınan ücret düzeyi (düşük/orta/yüksek) gibi…

Bu tür araştırmalarda test yapılırken kontenjans tablosundan faydalanılır. Tabloda incelenen iki değişkene ilişkin şıklar yatay ve dikey gözelerde yer alır.



Hipotezler şu şekilde kurulur:

H0: A ve B özellikleri bağımsızdır

HA: A ve B özellikleri bağımlıdır

Yukarıdaki kontenjans tablosunda satır sayısı R ve kolon sayısı C ile gösterilmiştir buna göre serbestlik derecesi, sd= (c-1)(n-1)’dir.

Bu hipotez testini uygulamak için öncelikle her bir “gözlenen” göze için “beklenen” değerlerin bulunması gerekir. Beklenen değerler, satır ve sütün toplamlarının çarpımlarının toplam gözlem sayısına oranlanması ile bulunur. Örneğin n11 gözlenen değerdir. Bu gözeye ilişkin beklenen değerse ((n,1)\*(n1.))/N ile bulunur. Gözlenen ve beklenen arasındaki farkların büyüklüğüne bakmak gerekir. Aradaki fark mutlak değerce ne kadar büyükse, sıfır hipotezi hakkında o ölçüde şüphe duyulur. Her bir göze için beklenen ve gözlenen değerler gözönüne alınarak aşağıdaki test istatistiği hesaplanır.

****



**Örnek**

Türkiye’de (Doğu ve Güneydoğu Anadolu B.) yaşayan kadınların evlenme şekilleri ve yaşları arasında bir ilişki var mıdır?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Görücü** | **İsteyerek** | **Zorla** | **Akraba** |
| **13-16yaş** | 15 | 20 | 10 | 5 |
| **17-22 yaş** | 12 | 19 | 5 | 6 |
| **22+** | 6 | 5 | 5 | 6 |

H0:Evlenme usulü yaşa bağımlı değildir.

H1: bağımlıdır.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gözlenen** | **Görücü** | **İsteyerek** | **Zorla** | **Akraba** | **Satır toplamı** |
| **13-16yaş** | 15 | 20 | 10 | 5 | 50 |
| **17-22 yaş** | 12 | 19 | 5 | 6 | 42 |
| **22+** | 6 | 5 | 5 | 6 | 22 |
| **Sütun toplamı** | 33 | 44 | 20 | 17 | 114 |

B11=(29\*46)/114= 13.3 … B34=(17\*22)/114=..

Tüm beklenen değerler yukarıdaki gibi hesaplanırsa şu tabloya ulaşılır;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Görücü** | **İsteyerek** | **Zorla** | **Akraba** |
| **13-16yaş** | 13,3 | 20,2 | 9,2 | 5,2 |
| **17-22 yaş** | 11,3 | 17,2 | 7,8 | 5,7 |
| **22+** | 5,4 | 6,6 | 6 | 5,1 |

 hesaplanan bu değer tablo değeri ile kıyalanırsa,



(Ki-kare tablosu bölüm sonunda yer almaktadır)

H0 hipotezi kabul edilir yani, evlenme usulu yaşa bağımlı değildir.

**Örnek**

Sigara kullanma alışkanlığına ilişkin bir araştırma için 120 kadın ve 140 erkekle görüşülüyor. Sigara kullanımının her iki cinsiyet için de aynı olduğu iddiası desteklenmeli midir (

|  |  |  |
| --- | --- | --- |
|  | Kullanıyor | Kullanmıyor |
| Erkek | 80 | 60 |
| Kadın | 60 | 60 |

Tabloda gözlenen değerler yer almaktadır. Her gözeye ilişkin beklenen değerlerse şu şekilde bulunur:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Kullanıyor | Kullanmıyor |  |
| Erkek | 80 | 60 | 140 |
| Kadın | 60 | 60 | 120 |
|  | 140 | 120 | 260 |

H0:Sigara içme alışkanlığı cinsiyete bağımlı değildir.

H1:Sigara içme alışkanlığı cinsiyete bağımlıdır.

B11= (140x180)/260=75.38 B12=(120x140)/260=64.62

B21= (140x120)/260=64.62 B22=(120x120)/260=55.38



Tablo değeri  olduğundan sigara içme alışkanlığı ve cinsiyet arasında ilişki yoktur hipotezi kabul edilir.

## 11.3. Ki-Kare Homojenlik Testi

Bu test ile iki ya da daha fazla bağımsız örneklemin aynı anakütleden seçilip seçilmediğinin araştırılmasında kullanılır. Bağımsızlık testinden farklı olarak, sıfır hipotezi örneklemlerin aynı ana kütleden seçildiği şeklinde kurulur.

H0: Örneklemler aynı ana kütleden seçilmiştir.

H1: Örneklemler farklı ana kütlelerden seçilmiştir.

**Örnek**: Bir süpermarket aynı işletmenin piyasaya sürdüğü 2 farklı şampuan markasına ilişkin tüketici ile memnuniyet anketi yapmıştır. Bu amaçla ilk grupta 50 ikinci grupta ise 75 kişiyle görüşülmüştür. Buna göre seçilen örneklemler aynı anakütleden mi gelmektedir?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Şampuan** | **Begenmeyen** | **Kararsız** | **Begenen** | Toplam |
| **A** | 30 | 16 | 5 | 50 |
| **B** | 40 | 25 | 10 | 75 |
| Toplam | 70 | 40 | 15 | 125 |

Yukarıdaki örneklerde hesaplanıldığı gibi burada da her bir göze için tek tek beklenen frekanslar bulunur. Buna göre elde edilen değerler aşağıdaki gibi olacaktır.

B11= (50x70)/125=28 . . . B23= (15x75)/125=9

|  |  |  |  |
| --- | --- | --- | --- |
| **Şampuan** | **Beğenmeyen** | **Kararsız** | **Beğenen** |
| **A** | 28 | 16 | 6 |
| **B** | 42 | 24 | 9 |

H0: İki örneklem aynı ana kütleden seçilmiştir.

H1: Örneklemler farklı ana kütlelerden seçilmiştir.



Tablo değeri  olduğundan örneklemler aynı anakütleden seçilmiştir.

## 11.4. Ki-Kare Uyum Testi

Bir örneklemin ilgili olduğu ana kütleyi temsil edip etmediğini belirlemek amacıyla kullanılan bir testtir. Testte beklenen frekanslar ilgili olasılık dağılımından elde edilmektedir.

**Örnek:**

Bir fotokopi makinesinin verdiği arıza sayılarının Poisson dağıldığı iddia edilmektedir. Aşağıda arıza haftaları ve sayılarının sıklığına ilişkin veri mevcuttur. Buna göre arıza sayılarının poisson dağılımına uyduğu iddiası desteklenmeli midir?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Arıza sayısı | 0 | 1 | 2 | 3 |
| Hafta | 32 | 16 | 8 | 4 |

H0: Arıza sayıları poisson dağılımına uyar

H1: Arıza sayıları poisson dağılımına uymaz

Yukarıdaki tablodan görülebileceği gibi 60 haftanın 32 sinde hiç arıza oluşmazken, 4’ün de 3 arıza oluşmaktadır.

Öncelikle varsayılan poisson dağılımına ilişkin olasılıklar ve beklenen frekanslar bulunur sonrasında ki-kare sınaması yapılır.



Ortalama arızalanma sayısı 









|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gözlenen frekans | 32 | 16 | 8 | 4 |
| Beklenen frekans | 27 | 21 | 8,4 | 2,4 |

Beklenen frekans değerleri olasılık değerleri ve toplam frekans değeriyle çarpılarak bulunmuştur.

0.45x60=27…



 (Burada serbestlik derecesi k-m-1 den, 4-1-1 şeklinde bulunmuştur.

Buna göre dağılım poisson’a uymaktadır.

## 11.5. Kontenjans Katsayısı

İki değişken arasındaki ilişkinin büyüklüğü hakkında bilgi verir. Kontenjans tablolarından (r>2, c>2) hesaplanır, “c” notasyonu kontenjans katsayısını göstermek üzere c=0 ise ilişki yoktur eğer c=1 ise oldukça güçlü ilişki varlığını gösterir.



İle hesaplanır.

**Örnek:**

Eğitim Düzeyi ve Gelir arasındaki ilişki var mıdır? Bu ilişkinin derecesi nedir?

|  |  |  |  |
| --- | --- | --- | --- |
| Gelir Düzeyi/ Eğitim Düzeyi | Lise | Üniversite | Yüksek Lisans |
| Düşük | 50 | 20 | 10 |
| Orta | 20 | 30 | 10 |
| Yüksek | 20 | 10 | 30 |

Soruda verilen gözlenen frekanslarla beraber beklenen frekans değerleri de dikkate alınmalıdır. Her bir gözeye ilişkin hesaplanan beklenen frekanslardan sonra ki-kare değeri hesaplanır.



Tablo değeri  olduğundan, gelir düzeyi ve eğitim düzeyi arasındaki ilişki varlığı kabul edilir.

Kontenjans katsayısı ise,



Orta düzeyde bir ilişki varlığına işaret eder.

**Bölüm Soruları**

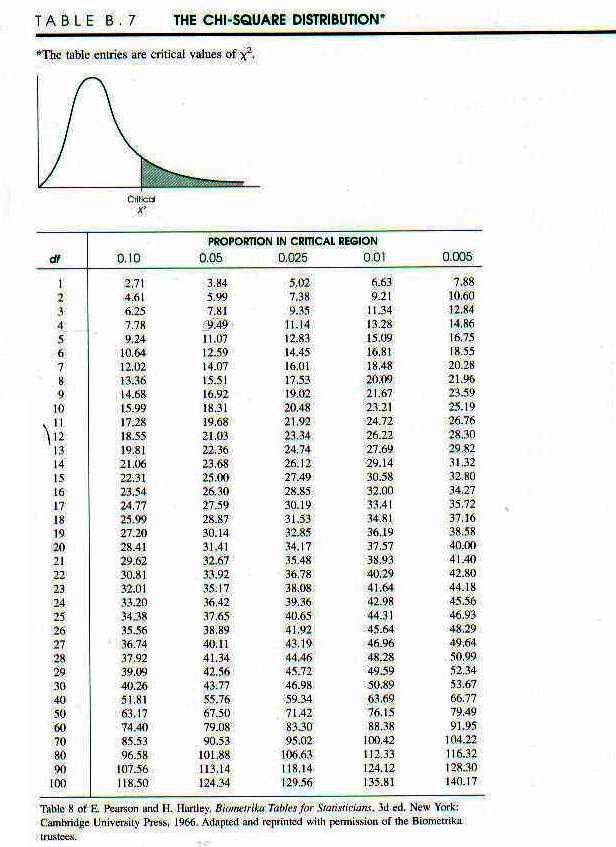
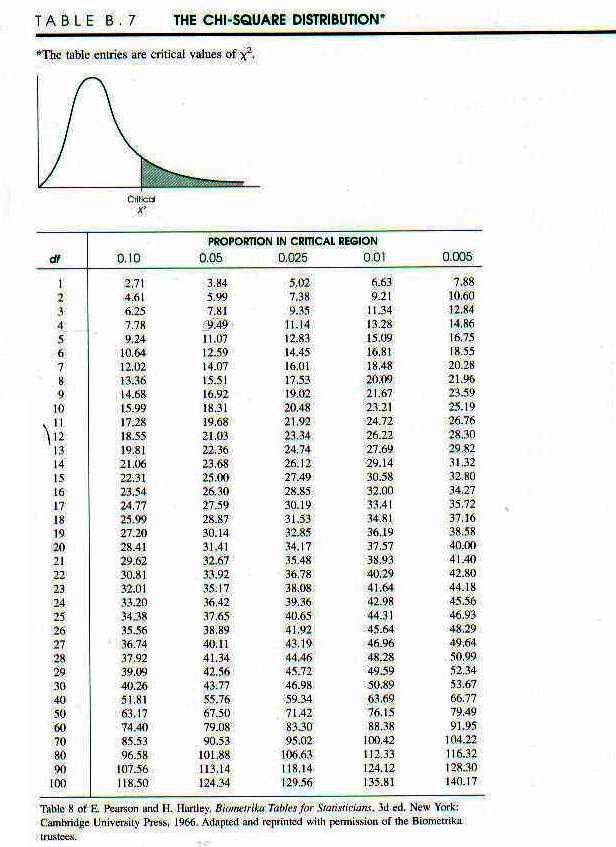
1. Evli kadınların maruz kaldığı şiddet ve eşlerinin eğitim düzeyleri arasında ilişkiyi değerlendirmek için 150 evli kadınla görüşülmüştür, çalışma sonuçları aşağıdaki gibidir. Erkeğin eğitim düzeyi kadına yönelik şiddet eğilimi üzerinde etkili midir? (=0,01) (15 puan)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Sözlü şiddet** | **Fiziksel şiddet** | **Şiddet yok** |
| **İlk+ Orta** | 30 | 10 | 20 |
| **Yüksek** | 40 | 10 | 40 |

1. Sigara içme ve akciğer kanseri arasında bir ilişki varlığının sınanması için hangi test seçilmelidir?

|  |  |  |
| --- | --- | --- |
|  | **Akciğer Kanseri** |  |
| **Sigara içme** | Var | Yok |
| Kullanan | 20 | 80 |
| Kullanmayan | 5 | 95 |

1. Tablo değerlerini  ve  alarak değerlendirin
2. Kontenjans katsayısını bularak yorumlayın



BÖLÜM 11 SORULARI

1-Aşağıdaki ifadelerden hangisi ya da hangileri doğrudur?

1. İki değişken arasında ilişki yoksa bu iki değişkenin bağımsız olduğu söylenir
2. Sınıflama düzeyinde ölçülmüş iki değişkenin bağımsızlığı kikare analizi ile araştırılır.
3. Kikare bağımsızlık testinde serbestlik derecesi (satır sayısı-1)x(sütun sayısı-1) ile bulunur

a)Yalnız I b)Yalnız II c)Yalnız III d)I veII e) Hepsi

A-Aşağıdaki soruları tablodan hareketle çözünüz.

|  |  |  |
| --- | --- | --- |
|  | sigara kullanan | sigara kullanmayan |
| Kadın | 5 | 115 |
| Erkek | 20 | 160 |

2-Yukarıdaki tablodan hareketle uygun sıfır hipotezi aşağıdakilerden hangisidir?

1. Kadınlarda sigara içme alışkanlığı daha azdır
2. Erkeklerde sigara içme alışkanlığı daha fazladır
3. Sigara içme alışkanlığı ile cinsiyet arasında ilişki yoktur
4. Sigara içme alışkanlığı ile cinsiyet arasında ilişki vardır
5. Hiçbiri

3-Tablodaki serbestlik derecesi kaçtır?

1. 5 b) 4 c) 3 d)2 e)1

4-% 5 önem düzeyinde tablodan bulunacak kritik değer nedir?

a)1.96 b) 3.84 c)5.52 d)9.22 e)13.82

5-Tablodan hesaplanacak kikare değeri aşağıdakilerden hangisine eşittir?

a)1.04 b)2.21 c)3.18 d)4.57 e)8.41

B)Tablodan hareketle aşağıdaki soruları yanıtlayınız

Dört farklı yabancı dil eğitimine katılan öğrencilerin başarı ve başarısızlık durumları tabloda özetlenmiştir.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | B | C | D |
| Başarılı | 23 | 30 | 40 | 25 |
| Başarısız | 27 | 20 | 10 | 25 |

6)Tablodaki bilgi ile uyumlu olan alternatif hipotez aşağıdakilerden hangisidir?

a) Eğitim türü ve öğrencinin başarısı arasında ilişki vardır

b)Eğitim türü ve öğrencinin başarısı arasında bir ilişki yoktur

c)C türü eğitim en iyi sonucu verir

d)C türü eğitim diğer tekniklerden daha üstündür

e)Hiçbiri

7) Tablodaki serbestlik derecesi kaçtır?

1. 5 b) 4 c) 3 d)2 e)1

8) % 5 önem düzeyinde tablodan bulunacak kritik değer nedir?

a)1.96 b) 3.84 c)5.52 d)7.81 e)13.82

9) Tablodan hesaplanacak kikare değeri aşağıdakilerden hangisine eşittir?

a)11.04 b)12.21 c)13.18 d)14.37 e)18.41

10) Kikare bağımsızlık testinin uygulanacağı 60 gözlemden oluşan 3x3 kontenjans tablosu için uygun serbestlik derecesi aşağıdakilerden hangisidir?

a)59 b) 57 c)9 d) 4 e)3

Yanıtlar:

1)E 2) C 3) E 4) B 5) A 6) A 7) C 8) D 9)D 10) D