DERS 1

Değişken = Yaş ve Boy da "YAŞ" bir değişkendir

Bağımlı Değişken = Birbirine bağımlı olan yani şu olursa şunu da etkiler tarzı

Bağımsız Değişken = Birbirlerini etkilemeyen

BAĞIMLI BAĞIMSIZ ÖRNEK = Sigara içen biri kanser olursa | Sigara içen biri bağımsız | kanser olma olayı ise bağımlı olmuş oluyor

DERS 2

METRIC

Değer hem pozitif hem de negatif alıyorsa = Aralık ölçeği

Başlangıç sayısı 0’dır = Oran ölçeği

NOMINAL

Herhangi bir şey söylemiyor bir grup içinde kız erkek olabilir cinsiyet grubu içinde

ORDINAL

Bir sıralama olduğu durumunda kullanılır

DERS 3

Merkezi eğilim ölçüsü = Veriler hakkında basit bilgi verir

Merkezi dağılım ölçüsü = Verinin nasıl bir alan oluşturduğunu gösterir. Çan eğrisi gibi

DERS 4

Aritmetik ortalama = biliyorsun zaten

DERS 5

Uç değer = Bazen verilerde bulunan hatalardan kaynaklı olarak çok düşük veya çok yüksek bir değer ortaya çıkıyor. Bu duruma Uç Değer deniyor

Medyan = Ortanca demek | Kullanılma amacı, bazen verilerde uç değer olabiliyor bu yüzden aritmetik ortalama yapmak çok sağlıklı sonuçlar ortaya çıkarmıyor. Bu sebepten dolayı Medyan kullanmak daha mantıklı hale geliyor.

DERS6

Mod(TEPE DEĞERİ) = En fazla tekrar eden

DERS 7

Kartiller = Bir pastayı 4’e bölmek gibi düşün

Pastanın 1. Kısmı (1. Kartiller) %25 | Pastanın 2. Kısmı (2. Kartiller) Medyan oluyor çünkü tam yarısına geliyor | Pastanın 3. Kısmı (3. Kartiller) de Medyandan sonra %75 oluyor

DERS 8

Ranj = En büyük ölçü ile en küçük ölçü arasındaki farktır. Fazla verilerde kullanılmaz yanıltır. Genel olarak çok fazla da kullanılmaz.

DERS 9

Standart Sapma = Her bir gözlemin ortalamadan sapmanın ortalamasıdır. Bir ortalama değer bulunur, bu ortalama değer rastgele bir değerle karşılaştırıldığında aradaki değer standart sapma olarak adlandırılır. Bunu tüm veriyle karşılaştırdığında, o verinin standart sapması oluyor.

DERS 10

Varyans, standart sapma olayı = Eğer Standart Sapma yüksekse daha yayılmış bir grafik olur. Standart Sapması düşükse daha Medyana yakınlaşmış bir dağılım gözükür.

Varyans = Standart Sapmayla aynı, tek fark karekökünü almıyoruz.

DERS 11

Skewness = Eğer veri de işleri bozan varsa, çarpıklıklar ortaya çıkıyor. Eğer negatif bir çarpıklıksa negatife doğru uzuyor pozitifse pozitife doğru uzuyor.

DERS 12

Kurtosis = Çarpıklık değeri demek | Kurtosis 3 ise normal dağılım | 3’ten büyükse normal dağılımın biraz daha sivri hali oluyor | 3’ten küçükse daha çok yere basık gibi gözükür.

DERS 13

PYTHON DERSİ (Betimsel İstatistik Borsa Örneği)

DERS 14

PYTHON DERSİ(Betimsel İstatistik Satış Örneği)

DERS 15

Veri seti tanıma

DERS 16

Çıkarımsal İstatistik = Genel Konu Başlığı

DERS 17

Merkezi Limit Teoremi = Elimde 30 farklı kaynak bulunuyor ve bu kaynaklardan her biri 1000 öğeyi içeriyor. Bu verileri birleştirip ortalamasını aldığımda sonuç karmaşık bir şekil oluşturabilir, ancak her bir kaynağın ayrı ayrı ortalamasını alarak daha anlaşılır ve tahmin edilebilir bir grafik elde edebilirim. | Aldığım verileri rassal(rastgele) olmalı | Ne kadar çok rassal veri alırsam o kadar çok normal dağılıma yaklaşırım.

DERS 18

PYTHON DERSİ (Merkezi Limit Teoremi Örnek)

DERS 19

Standart Hata = Merkezi Limit Teoremi ile örneklemleri çekiyorduk ya işte o çektiğimiz örneklemlerin Standart Sapması.

DERS 20

PYTHON DERSİ (SEABORN KÜTÜPHANESİ)

DERS 21

Z-Score = Dağılımdaki verilerin ortalamadan kaç standart sapma uzaklıkta olduğunu verir.

DERS 22

PYTHON DERSİ (Z-Score Örnek)

DERS 23

Z-Table = Z-tablosu, bir standart normal dağılımın altındaki alanları hesaplamak için kullanılan bir tablodur | Z = (X(Değerimiz) – μ(Ortalama)) / σ(S.S.) | Z değeri elde ettiğinizde, örneğin 1.33 çıktı. 1.30'u alıp tablonun sol tarafından aradınız ve buldunuz. Geriye kalan 0.03'lük kısmı ise üstten arayarak eşleştireceksiniz. Elde ettiğiniz değer, sizin değerin yüzde oranıdır.

DERS 24

Z-Table Örneği

DERS 25

PYTHON DERSİ (Z Table Python Uygulama)

DERS 26

Serbestlik Derecesi = Serbestlik Derecesi'nin açıklaması şöyledir: Birisi bize 3 tane sayı seçmemizi söylesin. Bu 3 sayı toplamı 15'i asla geçemez. Şimdi ben ilk sayıyı 5, ikinci sayıyı 6 seçtim. Geriye 15'e ulaşabilmem için zorunlu olarak 4'ü seçmem gerekiyor. İşte bu, ilk seçtiğim 2 sayıya Serbestlik Derecesi denir.

DERS 27

PYTHON DERSİ (Student'in t Dağılımı)

DERS 28

Tahmin Teorisi = Aralık ve nokta olarak ikiye ayrılıyor. Aralık, şu aralıkta olduğunu belirtirken, nokta ise direkt olarak söyler.

DERS 29

Normal Dağılım Ortalama Güven Aralığı = %99 düzeyinde olsun. Çünkü simetrik olduğu için geriye %0.5 kalır | İki tarafta alfa / 2 oluyor simetri olduğu için. Alfa / 2'ler birbirini götürür ve geriye sadece alfa kalır. Ortalama güven aralığı bu nedenle 1-alfa'ya eşittir. |

DERS 30

PYTHON DERSİ (Normal Dağılım Ortalama Güven Aralığı Uygulama)

DERS 31

PYTHON + NORMAL DERS

T Dağılımı Ortalama Güven Aralığı = n=30 diyelim | sd(Serbestlik Dağılımı) = 29 oluyor çünkü alfa – 1 den dolayı | Z ile nerdeyse aynı şeyler zaten sadece tek farkı baktığın tablo farklı eğer unuttuysan ders31 den tekrar et

DERS 32

PYTHON + NORMAL DERS (İki Popülasyon Ortalama Farkı Güven Aralığı)

DERS 33

Rassal(Rastgele) Değişken = Olasılık konusunu anlatıyor hoca | Kesikli ve sürekli bir olasılık süreci vardır. Bir paranın yazı tura gelmesi kesikli bir olasılık. Örnek(P(X=x)) | Normal dağılıma sahip olup, bir sayıya ihtiyaç duymayanlar ise sürekli olasılıktır. Örnek(P(X < 1.45)) | Eğer bir şey garantiyse kesikli değilse süreklidir. Örnek(Cinsiyet kesin olduğu için süreklidir. Boy ise kesiklidir çünkü hassas terazide farklı çıkarsın)

DERS 34

Beklenen Değer = E(x) | Her bir random değişkenin onun olasığı ile çarpılıp toplanma olayı. Kesikli değişkenler için geçerli | Sürekli de örnek var

DERS 35

PYTHON DERSİ (Beklenen Değer Örnek)

DERS 36

PYTHON + NORMAL DERS

Büyük Sayılar Yasası = Yapılan deney sayısı aynı olasılık geçerliyse mesela 1 para attın 0.50 çıktı 2 para attın 1 çıktı 3 para attın 1.50 çıktı olasılıklar aynı. Hep bir 0.50 ye yakınsama var burada ki olay da o. Yani 100 para atsan 100\*0.50 = 50 çıkacak

DERS 37

Olasılık Dağılımı = Kesikli ve sürekli olarak 2 tane şey var. Kesikli de çarp topla | Sürekli de bir fonksiyon verir onun üzerinden integral alırız.

DERS 38

Olasılık Kütle ve Fonksiyon = DERS 38’E SS ATTIM | Kesikli ve sürekli grafiklerinde en önemli farklardan biri sol tarafta kesikli de olasılık varken, sürekli de f(x) yani bir fonksiyon vardır

DERS 39

PYTHON + NORMAL DERS

Birikimli Kümülatif Dağılım = CDF olarak kısaltılır | F(x) olarak gösterilir | Bu konuyla ilgili özet, bir PDF (Olasılık Yoğunluk Fonksiyonu) grafiği üzerinde belirli bir değere karşılık gelen kısmı hesaplar. Hesaplanan bu değer, CDF (Birikimli Dağılım Fonksiyonu) olduğu için arkasından toplandığı için, bize bir olasılık değeri sunar.

DERS 40

PYTHON + NORMAL DERS

Bernoulli Dağılımı = 2 tane çizgisi vardır. Bunlardan biri başarılı, diğeri de başarısızı temsil eder | Pythondan devam et

DERS 41

Binom Dağılımı = Birden fazla deneme yapıldığında Binom Dağılımı kullanılır. | Uygulama - Ders 41 – Örnek çözümleri var. | Bu deneme sayısına “n” denir. Eğer birden fazla deneme yapıyorsak bunu kullanılırız.

DERS 42

Poisson Dağılımı = Nadir gerçekleşen olaylar için kullanılır (örneğin, bir saat içinde kaza olma oranı gibi)| Lambda, yaptığımız gözlemlerin ortalama değerini temsil eder | Deneme sayımız 50'den büyük olmalıdır ve deneme sayısı ile lambda değerini çarptığımızda 5'ten küçük olmalıdır.

DERS 43

PYTHON + NORMAL DERS (Poisson Dağılımı Örnekleri)

DERS 44

PYTHON + NORMAL DERS

Kesikli Uniform Dağılımı = Bir sayı grubunda hepsinin aynı değeri vermesi gerekir. Yani elimde 6 tane zar olsun, bu zarların rastgele birinin gelme olasılığı 1/6'dır. Bu durum, kesikli uniform dağılımı olarak adlandırılır.

DERS 45

PYTHON + NORMAL DERS

Sürekli Uniform Dağılımı = Bildiğimiz sürekli konusunda ne geçiyorsa bunda da aynı olay

DERS 46

Normal Dağılım = Tarihsel Normal Dağılımı anlattı

DERS 47

PYTHON + NORMAL DERS (Normal Dağılım Özellikleri)

DERS 48

PYTHON + NORMAL DERS (Normal Dağılım Örnekleri)

DERS 49

Hipotez Testi = Bir varsayım ortaya attık ve bu varsayımı kanıtlayabilmek için test etmemiz gerekiyor. Bu teste Hipotez Testi diyoruz | Hipotez testi üç adımda gerçekleşir | İlk adım, varsayım yapmaktır| İkinci adım, veri toplamaktır | Üçüncü adım ise topladığımız verileri ve varsayımları kabul veya reddetmektir | Örneğin, elinde bir COVID aşısı yaptın ve bu aşının COVID'e karşı etkili olduğunu düşünüyorsun. Bu, ilk adım yani varsayım aşamasıdır. Ardından, tüm popülasyona ulaşamayacağın için birkaç örneklem alıp üzerinde test ettin ve verileri kaydettin, bu ikinci adımı oluşturur. Son olarak, topladığın verilerin işe yarayıp yaramadığını kontrol ediyorsun, bu da üçüncü aşamadır.| Bir şeyi %100 kesinlikle tahmin edemeyeceğimiz için yapılan sapmalara ise örneklem hatası denir. Birisi geldi ve dedi ki, 'Ben bir aşı yaptım ve kan değerleri 20'den 30'a çıktı.' Bu sunumu yaptıktan sonra biri gelip, 'Ya, bu şans eseri mi ortaya çıktı?' gibi bir soru yöneltirse, bu şans eseri olmadığını kanıtlayabilmem için Hipotez testine ihtiyaç vardır

DERS 50

Boş ve Alternatif Hipotezler = “H0 ve H1“ Hipotez kurarken bunları kullanıyoruz. | Birisi 'x eşittir y' olduğunu savunuyor, biz de ortaya çıkıp 'hayır kardeşim, x eşit değildir y' dedik. Bu durumda H0: x = y ve H1: x ≠ y olur

DERS 51

PYTHON + NORMAL DERS

Hipotez Testinde Anlamlılık Düzeyi = Alfa: Hata olasılığımızın max sınırı | 1 – alfa = Güven düzeyi.

DERS 52

Hipotez Testinde Tek ve Çift Kuyruk Yapısı = Eğer hipotezimiz H değerleri için 'Eşittir' ve 'Eşit değildir' diye ifade ediliyorsa, çift kuyruk yapısı kullanılır. Eğer herhangi bir değerden büyük veya küçük olduğu yani '>' veya '<' gibi işaretler kullanılıyorsa, tek kuyruk yapısı kullanılır.

DERS 53

Hipotez Testinde Tip I ve Tip II Hatalar = Bizim Tip 1 hatayı yapma olasılığımız %5dir.

DERS 54

Hipotez Testinde P Değeri Nedir = P(Probability/Olasılık) | Eğer P < Alfadan H0 Reddedilir | Eğer P => Alfadan H0 Reddedilemez | Z hesap = P | Z tablo değeri = alfa

DERS 55

Hipotez Testi Örnekleri (Hipotez Testi Örnekleri)

DERS 56

Test İstatistiği Nedir = 2’ye ayrılır | Parametrik Test | Parametrik olmayan test |

Parametrik Test = Veriler normal dağılır (basık, çarpık, bazı grafik yapıları, bazı istatistiksel testler) | Merkezi Limit Teoremi, n > 30 | Veri yapısının sürekli olması gerekiyor | Gruplar arası varyans (2 veya daha fazlaysa)

DERS 57

PYTHON + NORMAL DERS (Normallik Varsayımı Testleri)

DERS 58

PYTHON + NORMAL DERS (Tek Örneklem Z Testi)

DERS 59

PYTHON + NORMAL DERS

Tek Örneklem T Testi = Bunun olayı, Z Testi ile aynıdır, tek fark olarak örneklemin standart sapmasını alıyoruz. Z testi ile T testi arasındaki farkı şu şekilde açıklanır: Z testi için örneklem büyüklüğü > 30 olmalıdır, T testi için ise örneklem büyüklüğü < 30 olmalıdır. Ayrıca, T testi uç değerler içermemelidir

DERS 60

PYTHON + NORMAL DERS

Bağımsız İki Örneklem T Testi = Değişken sayısının 2 olduğu zaman kullanılır (Kadın ve Erkek) | Bağımlı değişken sürekli olmalı (Ders notları için bir yapı kuruyorsan, notlar sürekli olmalı) | Bağımsız değişkenler kategorik olmalı (Erkek-Kadın, EVLİ-BEKAR) | Her bir grup için bağımlı değişken normal dağılım göstermelidir | Normal bir dağılım göstermiyorsa: n1,n2 > 25 veya 15 | Aykırı değer olmamalı | Varyansların homojenliği (varyansların eşitliği varsayımı) | Varyans homojenliği geçerliyse T testi | Değilse Welch’in t-testi

DERS 61

PYTHON (Varyansların Homojenliği Nedir?)

DERS 62

PYTHON (Bartlett ve Levene Testleri Nedir?)

DERS 63

PYTHON (Bağımsız İki Örneklem T Testi Örnekleri)

DERS 64

PYTHON + NORMAL DERS

Bağımlı Örneklem T Testi = Bir X aşısı olduğunu düşünün ve bu X aşısı bir hastalığı daha hafif geçirmenizi sağlıyor. Kişinin X aşısı vurulmadan önceki durumlarına bakılıyor ve sonra aşının uygulanmasından sonra tekrar aynı durumlara bakılıyor. İşte bu duruma bağımlı örneklem t-testi kullanılıyor. Burada iki grup bulunsa da, aynı grup üzerinde aynı işlemler yapılıyor

DERS 65

PYTHON (Bağımlı Örneklem T Testi Örnekleri)

DERS 66

PYTHON (Z Oran Testi Nedir?)

DERS 67

PYTHON (A/B Testi Örneği)

DERS 68

FamilyWise Error Rate-Alpha Inflation (Aile bazında hata kavramı) = Elinde bir zar var, bu zarı 1 defa havaya attığında yazı gelme olasılığı 0.5 olsun. Bu durumda, 0.5 gelme olasılığını bildiğin için rahatça işlemler yapabilirsin. Ancak, aynı zarı 3 defa havaya attığında 1 defa yazı gelme olasılığını sorduğunda, bu olasılık 0.5'ten farklı bir rakam olabilir ve bu durum bizi hata yapmaya yönlendirebilir.

DERS 69

ANOVA = Varyansların analizi | Tek Yönlü Anova = Üçten fazla grup olmalı ve bu gruplar bağımsız bir değişkene bağlı olmalıdır (örneğin, Ürün fiyatı = 2x \* Medeni Durum). Burada medeni durum bağımsız değişkenin aldığı değerlerden geçerek işlemler sonucu bir ürün fiyatını yani bağımlı bir değişkeni oluşturur| Sigara içme alışkanlığını eğitim seviyesiyle karşılaştırsak, bağımsız değişken eğitim durumu olurken, bağımlı değişken sigara içme alışkanlığı olur. Post-hoc gruplar arası testidir

DERS 70

One-Way ANOVA (Tek yönlü anova) = Bağımlı değişken Sürekli olmalı – Bağımsız değişken kategorik olmalı – grup sayısı min 3 olmalı – Varyanslar homojen olmalı – gözlemler rassal seçilmeli (Bağımsızlık, 100 farklı evden ayrı su örneği almayı düşün, ancak sadece 1 evden su örneği alıp bu örneği 100 farklı örnekleme çevirmeyi düşünürsen, bu yanlış sonuçlara yol açabilir.)

DERS 71

PYTHON (One-Way ANOVA Örnek)

DERS 72

PYTHON + NORMAL DERS

Post-Hoc Analizi = Gruplar arasında kimlerin uyuşmadığını görmek istediğimizde, ANOVA bize doğrudan uyuşup uyuşmadığını söyler, ancak Post-Hoc hangi grupların uyuşmadığını belirtir | Önce normal dağılım mı değil mi? Eğer normal dağılımsa, yolumuza devam ediyoruz; aksi takdirde merkezi limit teoremi ile rassal örneklem çekip normal dağılıma uyduruyoruz | 2. ANOVA testi için varyanslar homojen mi değil mi diye test ediyoruz. Ardından, eğer varyanslar homojen ise, Post-Hoc testine geçiyoruz; değilse Welch's ANOVA testine geçiyoruz. Welch's ANOVA testinde anlamlılık çıkıyorsa, tekrar Post-Hoc analizine geçiyoruz. Varyanslarımız eşit olmadığı için Tamhane kullanıyoruz. Eğer varyanslar homojen ise ve doğrudan Post-Hoc analizine gitmişsek, Tukey HSD kullanıyoruz

DERS 73

PYTHON + NORMAL DERS

Sürekli birleşik getiri = Uç değerleri en aza indirebildiğimiz bir yapıdır (=LN(B3/B2))

DERS 74

PYTHON + NORMAL DERS

Two-Way ANOVA = Tek yönlü ANOVA, yalnızca bir bağımsız değişkeni içerirken, 2 yönlü ANOVA iki farklı bağımsız değişken alır. Örneğin, eğitim seviyesi ve medeni durum gibi iki bağımsız değişkenimiz olsun. Bu değişkenler üzerinden bir ürün satın alma oranını inceleyeceğiz. İki yönlü ANOVA, önce eğitim ile ilgili farklılık olup olmadığını, sonra medeni durum ile ilgili farklılık olup olmadığını ve son olarak bu iki faktör arasında farklılık olup olmadığını değerlendirir|

DERS 75

PYTHON (Two-Way ANOVA Örnek)

DERS 76

PYTHON (Pandas Melt Fonksiyonu)

DERS 77

PYTHON + NORMAL DERS

Repeated Measures ANOVA = Tekrarlı T testinin birden fazla gruplu versiyonu | Küresellik = Varyansların, homojenli varsayımı önce ve sonrayı çıkarıp hepsini kontrol ediyor | Bonferroni düzeltmesi, alfa şişmelerini biraz olsun önlemek için kullanılan bir yöntem

DERS 78

PYTHON + NORMAL DERS

Varyans-Kovaryans Matrisi = Kovaryans, iki değişkenin birlikte ne kadar değiştiklerinin ölçüsüdür. Örneğin, pozitif mi yoksa negatif mi bir ilişki içinde olduklarını belirlemek için kullanılabilir. Diyelim ki iki borsa var, bu borsaların fiyatları birlikte pozitif mi ilerliyor yoksa negatif mi ilerliyor, kovaryans bu ilişkiyi ölçer|

DERS 79

One-Way MANOVA = Kovaryans kullanarak çok değişkenli bir varyans analizidir | Bunu bağımlı değişken sayısı arttığında kullanıyoruz | Varyans kovaryans matrixi eşitliği | BUNLARI EN İYİ AÇIKLAYAN UYGULAMA – DERS79 – ANOVA MANOVA FARK GÖRSELİ

DERS 80

PYTHON (Two-Way MANOVA)

DERS 81

Non-Parametrik Testler Nedir = Hem Merkezi Limit Teoremi testini geçemedi hem de yeterli miktarda örneklem yoksa Non-Parametrik testlere yöneliyoruz | Aykırı değerler fazla olduğunda

DERS 82

PYTHON + NORMAL DERS

Ki Kare Uyum Testi = Kategorik testlerde kullanıyoruz | Testleri frekanslara ayırmadan ki kare testi yapılamaz | Bir müşteri sayısı olsun. Bu müşteri sayılarını toplayıp ortalamasını aldığınızda bu beklenen değer oluyor. Bu beklenen değeri her bir değişkenden çıkardığın zaman ise X^2 (ki kare) testi olmuş oluyor.

DERS 83

PYTHON + NORMAL DERS

Ki Kare Bağımsızlık Testi = Kategorik değişken ama 2 farklı kategorik değişken | X – Y yapısı 2x2 bir tablo olucak. R-C yapısı 2x3 veya daha fazla olacak | 2x2 için En küçük beklenen frekans > 25 büyükse “pearson ki kare testi” | En küçük beklenen 5 < frekans < 25 “yates ki kare testi” | frekans < 5 “fisher ki kare testi” | R X C beklenen frekansların < 5 sayısı toplam frekans içinde %20 den küçük ise “pearson ki kare testi” | %20 den büyükse “fisher ki kare testi”

DERS 84

PYTHON (İşaret Testi)

DERS 85

PYTHON + NORMAL DERS

Mann-Whitney U Testi = Bağımsız 2 örneklem t testin nonparametric karşılığı | Kategorik bir değişken ve ya Sıralı ya da Sürekli bir değişken olacak.

DERS 86

PYTHON + NORMAL DERS

Wilcoxon T Testi = Bağımlı Örneklem T Testi nonparametric karşılığı

DERS 87

PYTHON + NORMAL DERS

Kruskal Wallis H Testi = Tek yönlü anova nonparametric karşılığı

DERS 88

PYTHON + NORMAL DERS

Friedman Testi = Tek yönlü tekrarlı anova nonparametric karşılığı

DERS 89

Korelasyon Nedir = Neden sonuç ilişkisini baz alır | Koralasyon bir nedensellik değildir, bir nedensellik aramak için en iyi sebeptir | r ile gösterilir |

DERS 90

PYTHON + NORMAL DERS

Kovaryans ve Korelasyon İlişkisi = Cov pozitifse pozitif ilişki, negatifse negatif ilişki | İkisi de aynı işlemleri yapıyor, ancak korelasyon daha fazla sonuç verirken sabitleştirme üzerine odaklanıyor |

DERS 91

PYTHON + NORMAL DERS

Doğrusallık Nedir = Bir doğrunun bir denklemi olması gerekir, yani genellikle y = 2x gibi bir eğimi olur | orijin= y = mx | orijinden geçmiyorsa = y = b(sabit sayı) + m(eğim katsayısı)x |

DERS 92

PYTHON + NORMAL DERS

Pearson Korelasyon Testi = Parametrik bir testtir | 2 değişken sürekli olacak | Normallik testi | Doğrusallık olması gerekiyo

DERS 93

PYTHON + NORMAL DERS

Spearman Korelasyon Testi = Nonparametrik bir testtir |

DERS 94

PYTHON + NORMAL DERS

Korelasyon Matrisi Nedir = Varyans kovaryans matrisiyle aynı mantıkta

DERS 95

PYTHON + NORMAL DERS

Kısmi Korelasyon Nedir = Kilo ile öğün arasında bir ilişki var mı, bunu test ederken biri yanımıza gelip, 'Belki adamın yaşı büyüdüğü için kemikleri gelişti ve kilo aldı' diyor. Bu ifadenin üzerine, yeni bir değişken ekliyoruz, bu da Yaş değişkeni oluyor ve bu üçü için kısmi korelasyon hesaplanıyor

DERS 96

PYTHON (Bist Korelasyon Matrisi)

DERS 97

Veri Görselleştirme Nedir = Ham veriler hakkında bilgi sahibi olmak| Analiz sonuçlarını sunmak (reklam, manipülasyon veya kişileri etkileme amacıyla) | Veriye göre grafik türünün seçilmesi | grafik üzerinde kişilerin dikkatini çekebilecek noktaların iyi hazırlanması gerekmektedir |

DERS 98

Sütun Grafik = Dikey sütun | Yatay cubuk | Birbirinden farklı nicel yapılar(x y z w markaları olsun bunlar birbirinden farklı) için kullanılır | Aynı nicel yapıların süreç içindeki değişimlerini gösterir | Sıfırdan başlaması önemlidir

DERS 99-100

PYTHON (Sütun Grafik Örnek-1-2)

DERS 101

Çizgi Grafik = Bir veya birden fazla yapının zaman içindeki değişimini gösterir | Sıfırdan başlaması önemlidir

DERS 102

PYTHON (Çizgi Grafik Örnek)

DERS 103

Saçılım Grafiği = İki değişken arasındaki ilişki durumunu bir grafikle gösteriyor

DERS 104

PYTHON (Saçılım Grafiği Örnek)

DERS 105

Kutu Grafiği = Minimum ve maksimum değerler, altındakiler ve üstündekiler aykırı değerlerdir

DERS 106

PYTHON (Kutu Grafiği Örnek)

DERS 107

Sohbet

DERS 108

Eksik Değer Nedir = Bazı şeyleri kendi yorumlarımızla doldurabiliyoruz, yani tahmin edebiliyoruz; ancak tahmin edemediğimiz durumlarda arkasında istatistiksel bir süreç bulunur | Rassal eksik değerler (insan ve makine hataları sonucu ortaya çıkar) ve rassal olmayan eksik değerler olarak iki kategoriye ayrılırlar |SÜREÇ = Eksik veriler tespit edilir | rassal mı değil mi | ardından çözüm

DERS 109

PYTHON (Eksik Gözlem Nasıl Tespit Edilir)

DERS 110

Eksik Gözlem Rassallık Testi =

Rassal(diğer değişkenle bir ilişkisi yok) olarak ortaya çıkmışsa: Göz ardı edilebilir; eksik veri oranı %5'i geçmemelidir | Satır olarak silinebilir | Ortalama, medyan, veya mod gibi değerler kullanılabilir | Regresyon modeli tahmini| Makine Öğrenmesi | Trend oluşturuyorsa, bu durum tahmin olarak yerleştirilebilir |

Rassal olarak ortaya çıkmamışsa: Satır olarak silinemez |

DERS 111

PYTHON(Eksik Gözlem Çözümü)

Eksik Gözlem Çözümü = Eğer Sütün bazlı %75 80 bir kayıp varsa silme işlemi yapılabilir

DERS 112

Aykırı Değer Nedir = Normal bir sıralama içinde sırayı bozan kişi aykırı bir gözlemdir | Tek değişkenle çalışılıyorsa, aykırı gözlem değerini kutu grafiği üzerinde gözlemleyebiliriz | Kutu grafiğinde Q3 – Q1 > %50 büyük | Kutu grafiğinde aykırı değerleri tespit etmek için IQR(Q3 - Q1) + 1,5

DERS 113

PYTHON + NORMAL DERS

Tek ve Çok Değişkenli Aykırı Değer = Tek değişkende doğrudan yaş verisi bulunuyor ve içinde aykırı değerler mevcut | Çok değişkenli bir analizde yaş ve kaç yıl sigara içtiği gibi değişkenlere tek tek bakmak yerine bu değişkenlere grup olarak bakmamız gerekiyor. İlk olarak saçılım diyagramı üzerinden kontrol ederiz, ardından Local Outlier Factor (LOF) analizi yaparız

DERS 114

PYTHON (Aykırı Değer Tespit Uygulama)

DERS 115

PYTHON (Aykırı Değer Silme Yöntemi)

DERS 116

PYTHON (Aykırı Değer Ortalama Atama Yöntemi)

DERS 117

PYTHON (Aykırı Değer Baskılama Yöntemi)

DERS 118

PYTHON (Değişken Ölçeklendirme)

DERS 119

PYTHON (Kategorik Değişken Kodlama)

DERS 120

PYTHON + NORMAL DERS

One-Hot Dönüşümü = Kategorik değişken yapısında bir sorun ortaya çıkıyor. Bu sorun, 5 farklı kategori olduğunu düşünün ve bu kategorilere 0'dan 4'e kadar değerler atandığını varsayalım. Program, 4'ün 1'den büyük olduğunu tespit ediyor ve bu nedenle bu yapıyı bir hiyerarşiye dönüştürmek istiyor. Ancak biz hiyerarşik bir yapı istemiyoruz, bu yüzden bu sorunu bu ders çözüyoruz

DERS 121

PYTHON + NORMAL DERS

Kukla Değişken Tuzağı = One-Hot üzerinden başka bir sorun ortaya çıkıyor. Bu sorun şöyle: İlkbahar, Yaz, Sonbahar ve Kış olmak üzere 4 mevsim olduğunu düşünelim, ve 4 ayrı gözlemimiz olsun. Gözlemlerimiz "Hangi aylarda tatil yapıyorsunuz?" sorusunun cevapları olsun ve ben ilkbahar kuklasını sildim, bu yüzden şu an elimde 3 kukla değişkeni var. İlkbaharı sildim ve işlemlere başladım. İlk gözlemde "Yaz" denildi, Yaz hariç diğer 2 gözlemi 0 yaparak Yaz'ı 1 yaptım. Ardından 2. gözlemde "Kış" denildi, Kış hariç diğer 2 gözlemi 0 yaparak Kış'ı 1 yaptım. Daha sonra 3. gözlemde "Sonbahar" denildi, Sonbahar hariç diğer 2 gözlemi 0 yaparak Sonbahar'ı 1 yaptım. En son gözleme geldiğimde tüm değişkenlerim 0 oldu ve bu sayede program ilkbahar olduğunu anladı. Bu duruma "M-1 tane kukla değeri" denir. Eğer M-1 tane kukla değeri yapmazsanız, kukla değişken tuzağına düşersin

DERS 122

PYTHON (Ordinal Değişken Dönüşümü)

DERS 123

PYTHON (Veri Önişleme Örnek 1.Bölüm)

DERS 124

PYTHON (Veri Önişleme Örnek 2.Bölüm)

DERS 125

PYTHON (Veri Önişleme Örnek 3.Bölüm)

DERS 126

PYTHON (Veri Önişleme Örnek 4.Bölüm)

DERS 127

Regresyon Nedir = Doğrusal regresyon tahmin modelinde bir doğru olsun ve bu doğrunun etrafında noktalar bulunsun. Bu noktalar ile doğru arasındaki mesafe hata terimi olarak adlandırılır. Amaç ise sabit sayı ile eğim katsayısını tahmin etmektir| Regresyonun amacı ise hava sıcaklığı arttığında dondurma satışının arttığını bilmekle birlikte, hava 1 derece arttığında dondurma satışının kaç birim arttığını öğrenmek istiyorsak regresyon kullanmaktır

DERS 128

En Küçük Kareler Yöntemi = Amacımız hata terimini minimize etmektir. Bu, basit doğrusal regresyon için kullanılan bir yöntemdir

DERS 129

PYTHON + NORMAL DERS

Basit Doğrusal Regresyon = Bağımsız değişkenin sadece bir tane olduğu için bununa basit denir | Yorumlamasını Python'a yazdım

DERS 130

Parametre Anlamlılığı = Python’da açıkladım

DERS 131

PYTHON + NORMAL DERS (Model Anlamlılığı)

DERS 132

PYTHON + NORMAL DERS

R Kare Nedir = Bağımsız değişkenler, bağımlı değişkenleri ne kadar iyi tahmin edebildiğini gösteren bir ölçüdür | 0 ile 1 arasında bir değer alır | Değer 1'e ne kadar yakınsa, modelin bağımsız değişken ile bağımlı değişken arasında o kadar anlamlı bir uyum gösterdiği anlamına gelir | Eğer tek bir bağımsız değişken varsa, bu ölçümü kullanabilirsiniz; ancak birden fazla bağımsız değişken varsa, düzeltilmiş R-kare (Adj. R-squared) kullanılır

DERS 133

PYTHON + NORMAL DERS

Düzeltilmiş R Kare Nedir = Birden fazla bağımsız değişken varsa düzeltilmiş R-kare kullanılır | Çünkü 2 tane bağımsız değişken kullanıldığında R-kare değeri büyük çıkabilir ve bizi yanıltabilir; bu nedenle düzeltilmiş R-kare kullanılır.

DERS 134

PYTHON + NORMAL DERS (Çoklu Doğrusal Regresyon)

DERS 135

EKK Varsayımları = Doğrusallık varsayımı (Modelin katsayılarının doğrusal olup olmaması) | 2- Normallik varsayımı (Model hatalarının normal dağılım göstermesi) | 3- Otokorelasyon varsayımı (Hatalar arasında otokorelasyon olmamalı) | 4- Sabit varyans (Her bir gözlem için hata varyansının sabit olması) | 5- Çoklu doğrusal bağıntı (Bağımsız değişkenler arasında çoklu doğrusal problem olmamalı) | 6- Teorik olarak söylenir (Gözlem sayısı bağımsız değişken sayısından fazla olmalı)

DERS 136

PYTHON + NORMAL DERS

1- Doğrusallık Varsayımı = X ve Y arasında bir doğrusallığın olması gerekir.

DERS 137

PYTHON + NORMAL DERS

2- Normallik Varsayımı = sm.qqplot(model.resid,line="s") Normal bir dağılım göstermesi gerekiyor

DERS 138

PYTHON + NORMAL DERS

3- Otokorelasyon Varsayımı = Korelasyon, başka bir değişkenle olan ilişkiyi ifade ederken, otokorelasyon ise aynı değişkenin kendi geçmişiyle olan ilişkisini ifade eder | Otokorelasyonu kullanmamızın nedeni hata terimlerinin rastgele (rassal) olarak oluşmasını istememizdir, yani bu hata terimleri birbiriyle ilişkili olmamalıdır.

DERS 139

PYTHON + NORMAL DERS

Newey–West Yöntemi = Eğer otokorelasyon mevcutsa, bu yöntemi kullanırız | Newey-West otokorelasyonu çözmez, sadece otokorelasyonun neden olduğu tahribatı giderir.

DERS 140

PYTHON + NORMAL DERS

4- Sabit Varyans Varsayımı = Hataların rassallığına bakılır. Eğer rassal değillerse, bize sunulan p-değerleri yanıltabilir | Eğer hatalar normal dağılım gösteriyorsa, Pagan testi yapılabilir, ancak normal dağılım varsayımı geçerli değilse ve merkezi limit teoremi ile ilgiliysek, White test daha yaygın olarak kullanılır

DERS 141

PYTHON + NORMAL DERS

Log Model Dönüşümü = Eğer sabit varyans varsayımı geçerli değilse, bunu kullanırız

DERS 142

PYTHON + NORMAL DERS

5- Çoklu Doğrusal Bağlantı = Modelimiz içinde bulunan bağımsız değişkenlerden herhangi biri ile model içinde bir ilişki olduğunda böyle bir sorun ortaya çıkabilir | Bu nedenle korelasyon matrisi oluşturulmalı ve VIF testi yapılmalıdır

DERS 143

PYTHON + NORMAL DERS

Değişken Dışlama = Çoklu doğrusal bağlantının bir çözüm yolu, maksimumdan (max) başlanır.