2023-2024 GÜZ DÖNEMİ İST369 R PROGRAMI İLE İSTATİSTİKSEL ANALİZ İST. DERSİ ÖDEVİ

Hazırlayan;

Yunus Emre BÜYÜKGÜLER - 21020513

25 Kasım 2023

Umut YAMAK

İST369 R PROGRAMI İLE İSTATİSTİKSEL ANALİZ

**Kesikli Dağılım Örneği**

**Problem: Online Alışveriş Sitesinde Ürün Satın Alma Olasılığı**

Diyelim ki bir online alışveriş sitesi, belirli bir ürün kategorisindeki kullanıcıların %20'sinin bir ürünü satın alma olasılığı olduğunu iddia ediyor. Ancak, gerçek durumu kontrol etmek için bir hipotez testi yapmak istiyoruz.

**Hipotezler:**

* H0: Kullanıcıların belirli bir ürünü satın alma olasılığı %20'dir.
* H1: Kullanıcıların belirli bir ürünü satın alma olasılığı %20'den farklıdır.

**Adımlar:**

* Rastgele bir kullanıcı grubu oluşturulur (100 kullanıcı).
* Hipotez testi için uygun test seçilir ve test istatistiği hesaplanır.
* P-değeri hesaplanır.
* Alfa düzeyi belirlenir (0.05).
* P-değeri alfa düzeyinden küçükse, H0 reddedilir; aksi takdirde, H0 kabul edilir.

R ortamında test edelim.

kullanici\_sayisi<-100

urun\_satin\_alma <- rbinom(kullanici\_sayisi, size = 1, prob = 0.2)

test\_sonucu <- prop.test(sum(urun\_satin\_alma), length(urun\_satin\_alma), p = 0.2, alternative = "two.sided")

#p-value = 0.4533 çıktı.

#p > α olduğundan H0 reddedilemez.

#**Kullanıcıların belirli bir ürünü satın alma olasılığı %20'dir denilebilir.**

**Sürekli Dağılım Örneği**

**Problem: Online Oyunlardaki Denge**

Bir oyun firması, kendi oyunlarındaki oyun dengesinin yüzde 50 olduğunu iddia etmektedir.

Yapılan araştırma sonucu ulaşılan 1000 oyuncudan 100 tanesinin maç başına kazanma oranı beklenenden uç değerlerde çıkmıştır. Buna göre oyun firmasının iddiasını test edelim.

**Adımlar:**

* **Hipotezlerin oluşturulması**
* **Anlam düzeyinin (α) belirlenmesi**
* **Örnekleme dağılımının belirlenmesi**
* **Ret bölgesinin belirlenmesi**
* **Test istatistiğinin hesaplanması ve yorum**

**Hipotezler**

**H0: Kazanma oranı 50%'dir**

**H1: Kazanma oranı 50%’den farklıdır.**

R ortamında test edelim:

prop.test(x=900,n=1000,p=0.5,alternative = "two.sided", correct = TRUE)  
  
 1-sample proportions test with continuity correction  
  
data: 900 out of 1000, null probability 0.5  
X-squared = 638.4, df = 1, p-value < 2.2e-16  
alternative hypothesis: true p is not equal to 0.5  
95 percent confidence interval:  
 0.8793091 0.9175476  
sample estimates:  
 p   
0.9

#P < α olduğundan H0 reddedilir.

#Oyunda bir dengesizlik olduğundan 95% güven düzeyinde söz edilebilir.

Bir kitle ortalaması kıyası testlerinden z testini uygulayalım:

**Problem: Ortalama Teslimat Süresi**

Bir şirketin belirli bir ürününün ortalama teslimat süresi konusunda bir iddia test edilecek olsun. Şirket, ortalama teslimat süresinin 3 gün olduğunu iddia ediyor. Ancak, bu iddiayı test etmek ve bu sürenin gerçekten doğru olup olmadığını belirlemek istiyoruz. Ayrıca, teslimat süresi varyansını biliyoruz gibi varsayalım.

populasyon\_ortalama <- 3 # Popülasyonun teorik ortalama teslimat süresi (gün)

populasyon\_varyans <- 1 # Popülasyonun teorik varyansı

orneklem\_buyuklugu <- 40 # Örneklem büyüklüğü

# Normal dağılıma dayalı rastgele örnekleme

ornek\_veri <- rnorm(orneklem\_buyuklugu, mean = populasyon\_ortalama, sd = sqrt(populasyon\_varyans))

# Oluşturulan örneklem veri setini kontrol et

print(ornek\_veri)

# Z-testi için parametreleri tanımla

h0\_ortalama <- 3 # Hipotez edilen popülasyon ortalaması

anlam\_düzeyi <- 0.05 # Anlamlılık düzeyi

# Z-testi gerçekleştir

z\_testi\_sonucu <- z.test(ornek\_veri, mu = h0\_ortalama, sigma = 1, conf.level = 0.95, alternative = "two.sided")

# Z-testi sonuçlarını göster

print(z\_testi\_sonucu)

Bir kitle ortalaması kıyası testlerinden t testini uygulayalım:

**Problem: Sınav Performansı**

Bir öğrenci kursu şirketi, iki farklı öğretim yöntemi arasında öğrenci sınav performansı üzerinde bir fark olup olmadığını değerlendirmek istiyor.

Varyans biliniyor ve öğrenci sınav notları normal bir dağılıma sahip olduğunu varsayalım.

yonetim1\_notlari <- rnorm(30, mean = 75, sd = 10) # Yöntem 1

yonetim2\_notlari <- rnorm(30, mean = 80, sd = 12) # Yöntem 2

# Oluşturulan örneklem veri setlerini kontrol et

print(yonetim1\_notlari)

print(yonetim2\_notlari)

# Bağımsız iki örneklem t-testi için parametreleri tanımla

anlam\_düzeyi <- 0.05 # Anlamlılık düzeyi

# Hipotezler:

# H0: İki öğretim yöntemi arasında bir fark yoktur. (Ortalama fark = 0)

# H1: İki öğretim yöntemi arasında bir fark vardır. (Ortalama fark ≠ 0)

# Bağımsız iki örneklem t-testi gerçekleştir

t\_testi\_sonucu <- t.test(yonetim1\_notlari, yonetim2\_notlari)

Welch Two Sample t-test  
  
data: yonetim1\_notlari and yonetim2\_notlari  
t = -0.96013, df = 49.053, p-value = 0.3417  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -9.012972 3.184926  
sample estimates:  
mean of x mean of y   
 78.19322 81.10724   
  
#p > α olduğundan H0 reddedilir.

#İki yönetim arasında bir fark olduğu 95% güvenle söylenebilir.

Bir kitle ortalaması kıyası testlerinden wilcoxon testini uygulayalım.

**Problem: Politikaları Optimize Etmek**

Diyelim ki bir şirket, çalışanlarının işe başlama sürelerini optimize etmeye çalışıyor ve yeni bir işe başlama süresi politikası uygulamaya koydu. Ancak, bu politikanın işe başlama süresi üzerinde bir etkisi olup olmadığını test etmek istiyorlar.

Elimizdeki veriler, eski işe başlama sürelerini ve yeni politika altındaki işe başlama sürelerini içeriyor.

# Eski ve yeni işe başlama süreleri veri seti oluştur

eski\_baslama\_sureleri <- c(8:45, 8:55, 9:05, 8:50, 9:10)

yeni\_politika\_baslama\_sureleri <- c(9:15, 9:25, 9:35, 9:20, 9:30)

# İki grup arasındaki veri noktalarını eşitle

yeni\_politika\_baslama\_sureleri <-yeni\_politika\_baslama\_sureleri[1:length(eski\_baslama\_sureleri)]

# Oluşturulan veri setlerini kontrol et

print(eski\_baslama\_sureleri)

print(yeni\_politika\_baslama\_sureleri)

# Wilcoxon işaretle testi için parametreleri tanımla

alpha <- 0.05 # Anlamlılık düzeyi

# Hipotezler:

# H0: Eski ve yeni işe başlama süreleri arasında bir fark yoktur.

# H1: Eski ve yeni işe başlama süreleri arasında bir fark vardır.

# Wilcoxon işaretle testi gerçekleştir

wilcoxon\_test\_sonucu <- wilcox.test(eski\_baslama\_sureleri, yeni\_politika\_baslama\_sureleri, paired = TRUE)

wilcoxon\_test\_sonucu

Wilcoxon signed rank test with continuity correction  
  
data: eski\_baslama\_sureleri and yeni\_politika\_baslama\_sureleri  
V = 3068, p-value = 4.816e-08  
alternative hypothesis: true location shift is not equal to 0

#p< α olduğundan H0 reddedilir

#95% güvenle eski ve yeni işe başlama süreleri arasında anlamlı bir farklılık vardır.

**Sürekli Olasılık Dağılımlarıyla Alakalı Ek Bir Soru**

**Problem: Müşterilere Ayırılan Hizmet Süresi**

Bir restoranın müşteri hizmet süresi, 0 ile 30 dakika arasında sürekli bir rastgele değişkenle modellendiğini düşünün. Bu süreklilik dağılımı, üniform dağılım ile modelensin. Restoranın bir müşteriye hizmet süresi 15 dakikadan fazla olma olasılığı nedir?

R ortamında problemi çözelim:

a <- 0 # Minimum süre

b <- 30 # Maksimum süre

# 15 dakikadan fazla hizmet süresi olma olasılığını hesapla

olasılık <- 1 - punif(15, min = a, max = b)

olasılık

0.5

#Restoranın bir müşteriye hizmet süresi 15 dakikadan fazla olma olasılığı 50%’dir