

Araştırma sorusu: Malatya ilinde yetişkin insanların kilo ortalaması 75 kg mıdır?



$$H_0 : \mu = 75kg$$

Bunun yerine populasyondan bir örneklem seçeriz



$$H_0: \bar{x} = 75kg$$

$$H_A: \bar{x} \neq 75kg$$

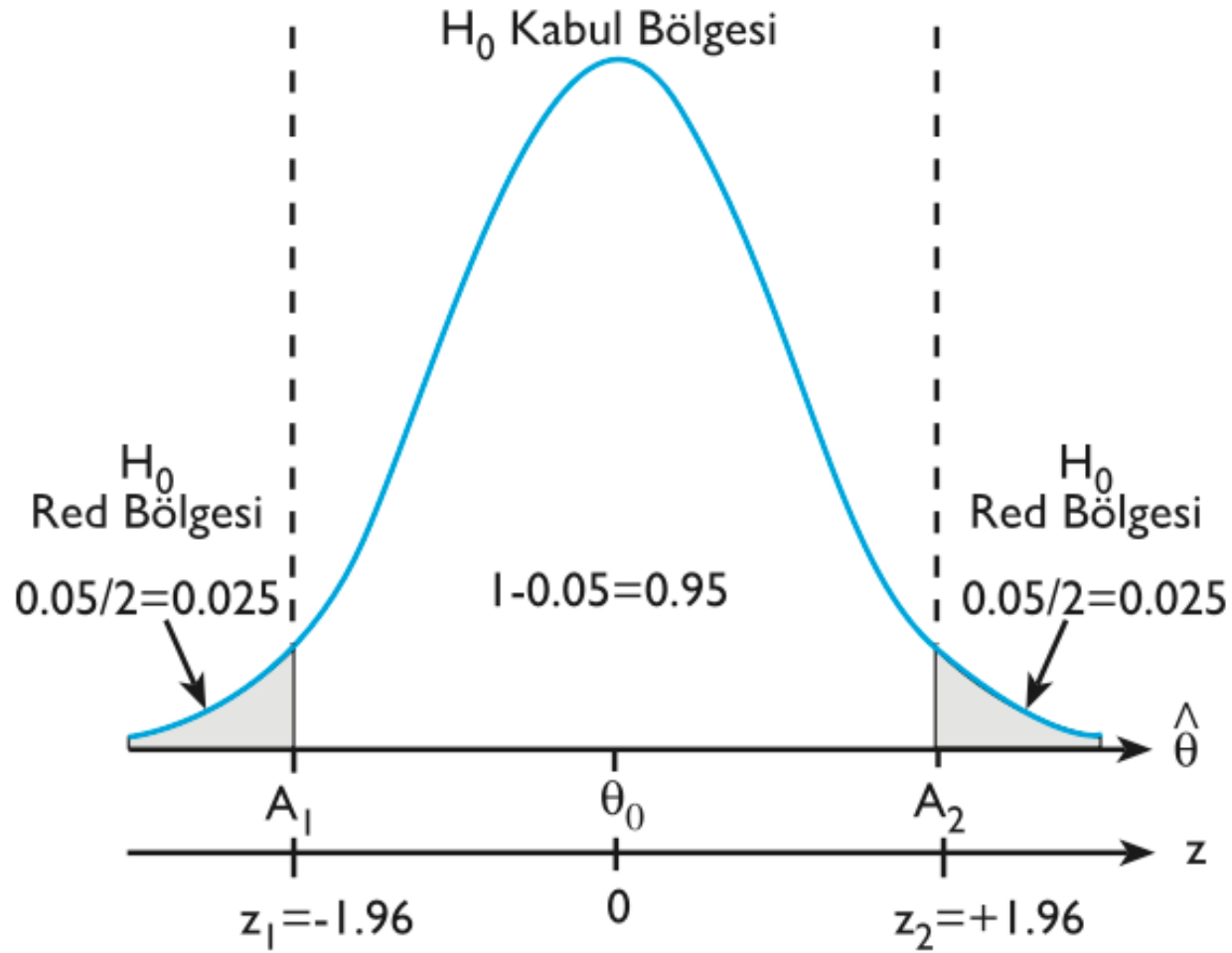
Dikkat! Bunu doğrulamak için Malatya ilinde herkese ulaşmamız gerekir. Bu zaman ve maliyet açısından zordur.

İŞTE BU SORUNUN CEVABINI ÖĞRENMEK İÇİN KULLANMAMIZ GEREKEN TEST “tek örneklem t-testi” dir.

KRİTİK BÖLGENİN BELİRLENMESİ

$\alpha=0.05$ anlamlılık düzeyi için

$1-\alpha=0.95$ testin güven düzeyini gösterir



Tek Örneklem t -testi

- Tek Örneklem t Testi, bir popülasyonun ortalamasının bilinen veya varsayılan bir değerden istatistiksel olarak farklı olup olmadığını inceler. Tek Örneklem t testi parametrik bir testtir.
- Tek Örneklem t - testinde, test değişkeninin ortalaması, popülasyondaki ortalamanın bilinen veya varsayılan bir değeri olan bir "test değeri" ile karşılaştırılır. Test değerleri bir literatür taramasından, güvenilir bir araştırma kuruluşundan, yasal gerekliliklerden veya endüstri standartlarından gelebilir

Varsayımları

1. Sürekli olan test değişkeni (yani, aralık veya oran düzeyi)
2. Test değişkenine ilişkin puanlar bağımsızdır (yani, gözlemlerin bağımsızlığı)
 - Test değişkenindeki puanlar arasında bir ilişki yoktur
 - Bu varsayımın ihlali hatalı bir p değeri verecektir
3. Popülasyondan rastgele veri örneği
4. Test değişkeni üzerinde örneklem ve popülasyonun normal dağılımı (yaklaşık olarak)
 - Normal olmayan popülasyon dağılımları, özellikle de ağır çarpık olanlar, testin gücünü önemli ölçüde azaltır
 - Orta veya büyük örneklemde, normallik ihlali yine de doğru p değerleri verebilir
5. Varyansların homojenliği (yani, varyansların hem örneklemde hem de popülasyonda yaklaşık olarak eşit olması)
6. Aykırı değer olmayacak

Örnekler

- Bir montaj hattının beş kilo ağırlığında dizüstü bilgisayarlar üretip üretmediğini belirlemekle ilgilendiğinizi varsayalım. Bu hipotezi test etmek için, montaj hattından bir dizüstü bilgisayar örneği toplayabilir, ağırlıklarını ölçebilir ve tek örneklem t-testi kullanarak örneği beş değeriyle karşılaştırabilirsiniz.
- Bir hastanede erkekler için rastgele bir kolesterol ölçüm örneği bulunmaktadır. Bu hastalar kolesterol dışındaki sorunlar için görülmüştür. Yüksek kolesterol için herhangi bir ilaç kullanmıyorlardı. Hastane, hastalar için bilinmeyen ortalama kolesterolün hedef seviye olan 200 mg'dan farklı olup olmadığını bilmek istemektedir.
- Bir ilaç şirketi yeni ilacının kan basıncını bir haftada 10 mmHg düşürdüğünü vaat ediyor. Bunun doğru olup olmadığını öğrenmek istiyorsunuz. Bunu yapmak için 75 denegin kan basıncında gözlenen düşüşü beklenen 10 mmHg'lik düşüşle karşılaştırın.

Örnekler

- Bir enerji barı örneğinin protein gramajını ölçüyoruz. Etiket, barların 20 gram protein içerdiğini iddia ediyor. Etiketlerin doğru olup olmadığını bilmek istiyoruz.
- Kanada'daki yöneticilerin sağlık algısının nüfusun genelinden farklı olup olmadığını öğrenmek istiyorsunuz. Bu amaçla 50 yöneticiye sağlık algılarını soruyorsunuz.
- Şirketinizin ürettiği vidaların gerçekten ortalama 10 gram ağırlığında olup olmadığını öğrenmek istiyorsunuz. Bunu test etmek için 50 vidayı tartın ve gerçek ağırlıklarını olması gereken ağırlıkla (10 gram) karşılaştırın.

P-değerini anlamak...!

1. P-değeri: İstatistikte p-değeri, sıfır hipotezine karşı kanıtın gücünü değerlendirmek için kullanılan bir ölçüdür.
2. Sıfır Hipotezi (H_0): Bu, ölçülen iki olgu arasında bir ilişki olmadığına veya gruplar arasında bir ilişki bulunmadığına dair genel bir ifade veya varsayılan konumdur. Örneğin, regresör sonucu etkilemez.
3. Alternatif Hipotez (H_1): Test etmek istediğiniz şey budur. Genellikle boş hipotezin tersidir. Örneğin, regresörün sonucu etkilediği.
4. P-değerinin hesaplanması: Her bir katsayı için p-değeri tipik olarak t-testi kullanılarak hesaplanır. Bunun için birkaç adım vardır. Şimdi bunları inceleyelim.

5. Standart hata (SE): Bir örneğin bir popülasyonu temsil etme doğruluğunu ölçer
6. Test İstatistiği (T): Bir modelindeki her bir katsayı için test istatistiği, **Katsayı Tahmini / Katsayının Standart Hatası** bölünerek hesaplanır. Bu size bir t-değeri verir.
7. Serbestlik Derecesi: Bu test için serbestlik derecesi (df) genellikle gözlem sayısı eksi tahmin edilen parametre sayısı (kesişim dahil) olarak hesaplanır.
8. P-Değeri Hesaplaması: Daha sonra p-değeri, hesaplanan t-değerinin uygun serbestlik derecesine sahip t-dağılımıyla karşılaştırılmasıyla belirlenir. Hesaplanan t-değerinin ötesinde, t-dağılım eğrisinin altındaki alan p-değerini verir.
9. Yorumlama: Küçük bir p-değeri (genellikle $\leq 0,05$), sıfır hipotezi doğru olsaydı böyle bir veri örüntüsü gözlemlenmenin olası olmadığını gösterir ve tahmin edicinin modele önemli bir katkıda bulunduğunu düşündürür.

Örnek

- Bir araştırma hastanesinin yenidoğan bölümünden 24 adet bebeğin kiloları alınmıştır. Yapılan araştırmalarda, dünyaya gözlerini açan bir bebeğin normal kilosu 3200 gramdır. Bu 24 bebekler acaba normal kiloda mıdır?

Hipotezler

Boş hipotez (H_0) ve Alternatif hipotez (H_1) şu şekilde ifade edilebilir:

$H_0: \mu = \mu_0$ ("Popülasyon ortalaması belirtilen ortalama değere eşittir")

$H_1: \mu \neq \mu_0$ ("Popülasyon ortalamasının belirtilen ortalama değerden farklı olması")

- Burada μ "gerçek" popülasyon ortalaması ve μ_0 popülasyon ortalamasının önerilen değeridir.

Test İstatistiği

Tek Örneklem t Testi için test istatistiği t olarak gösterilir ve aşağıdaki formül kullanılarak hesaplanır:

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}},$$

burada $s_{\bar{x}} = s/\sqrt{n}$ ve

μ_0 : popülasyon ortalaması için önerilen sabit

\bar{x} : örneklem ortalaması (araştırma verilerinden elde edilir)

n : örneklem boyutu, gözlem sayısı, (N: Anakitle gözlem sayısı)

s : örneklem standart sapma

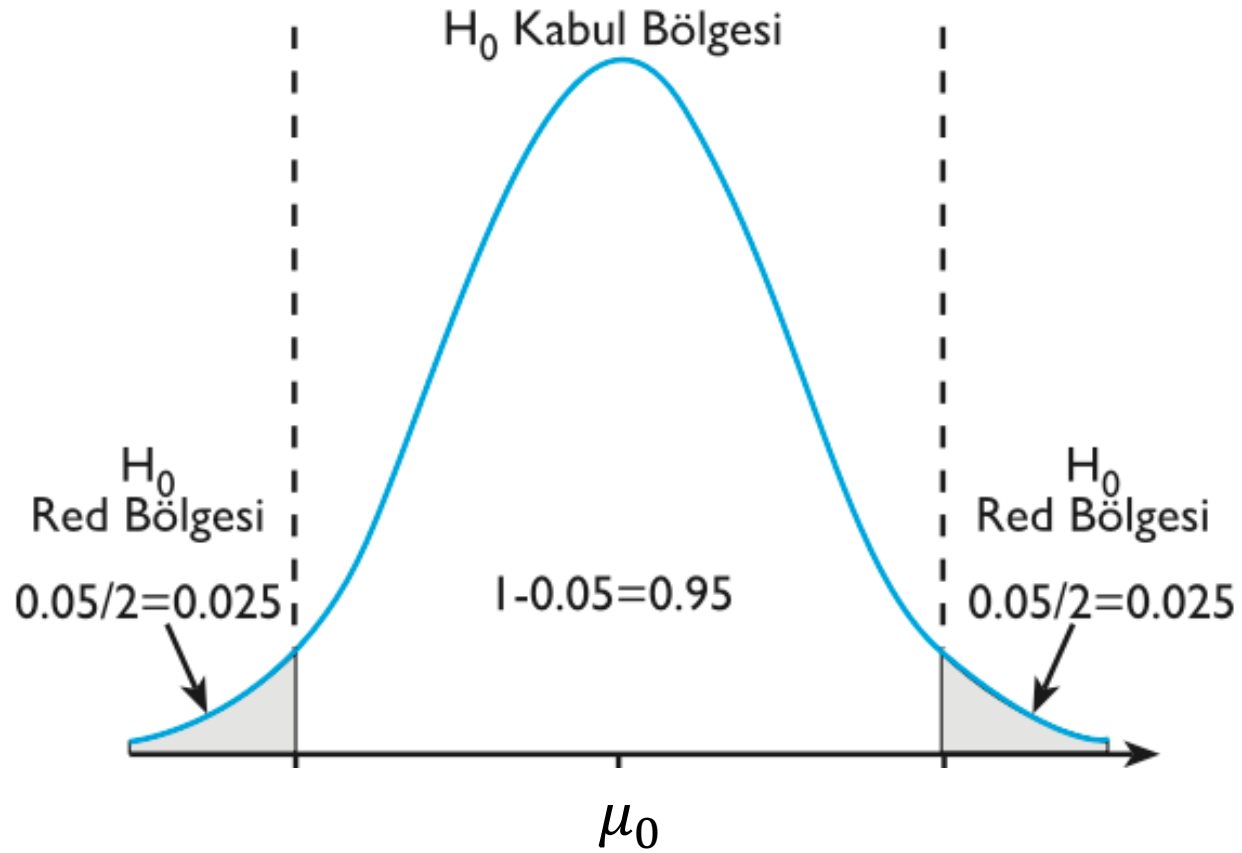
$s_{\bar{x}}$: Ortalamanın tahmini standart hatası

Hesaplanan t değeri daha sonra serbestlik derecesi $df = n - 1$ ve seçilen güven düzeyi ile t dağılım tablosundaki kritik t değeri ile karşılaştırılır. Hesaplanan t değeri $>$ kritik t değeri ise, sıfır hipotezi reddedilir.

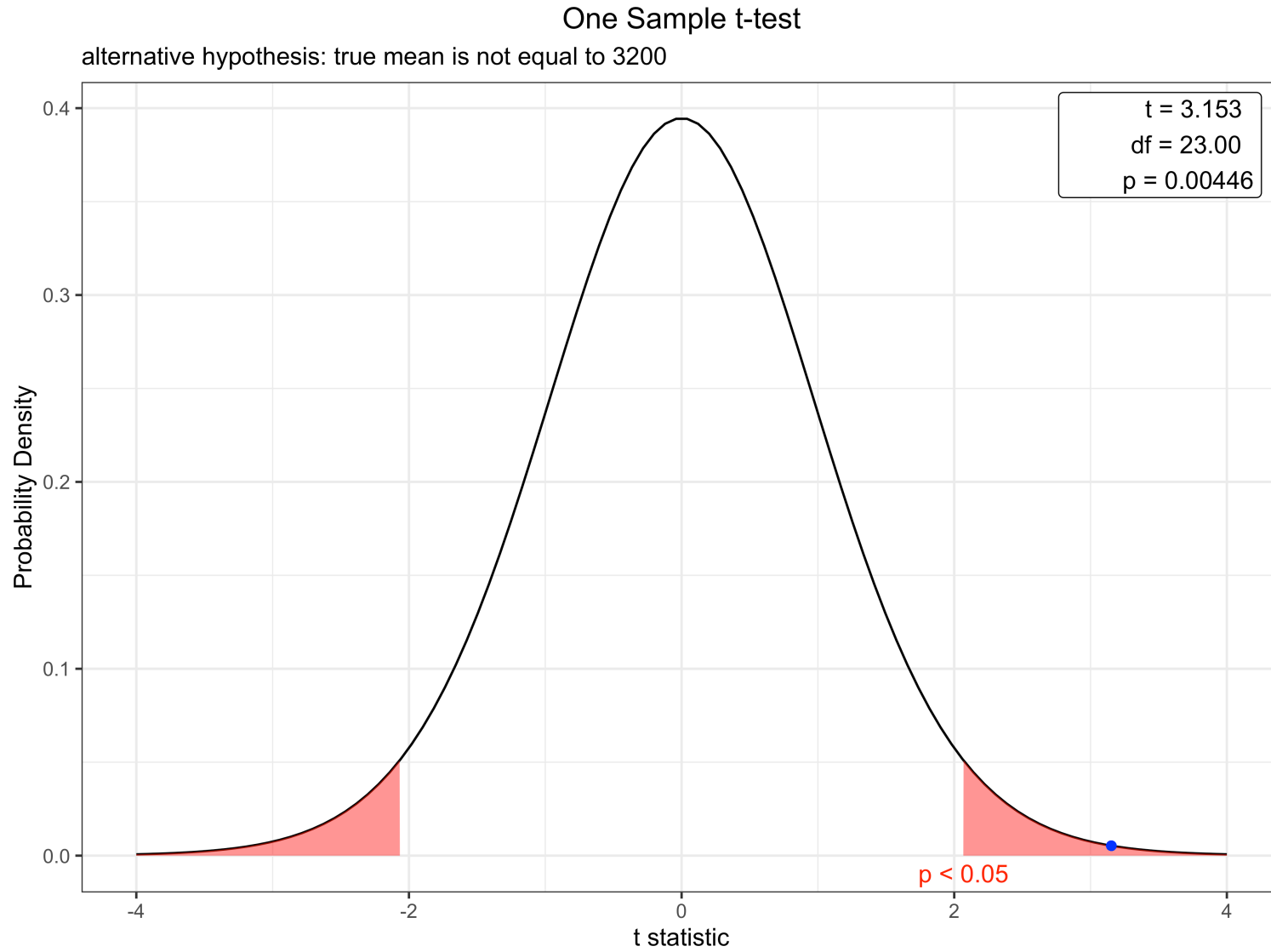
KRİTİK BÖLGENİN BELİRLENMESİ

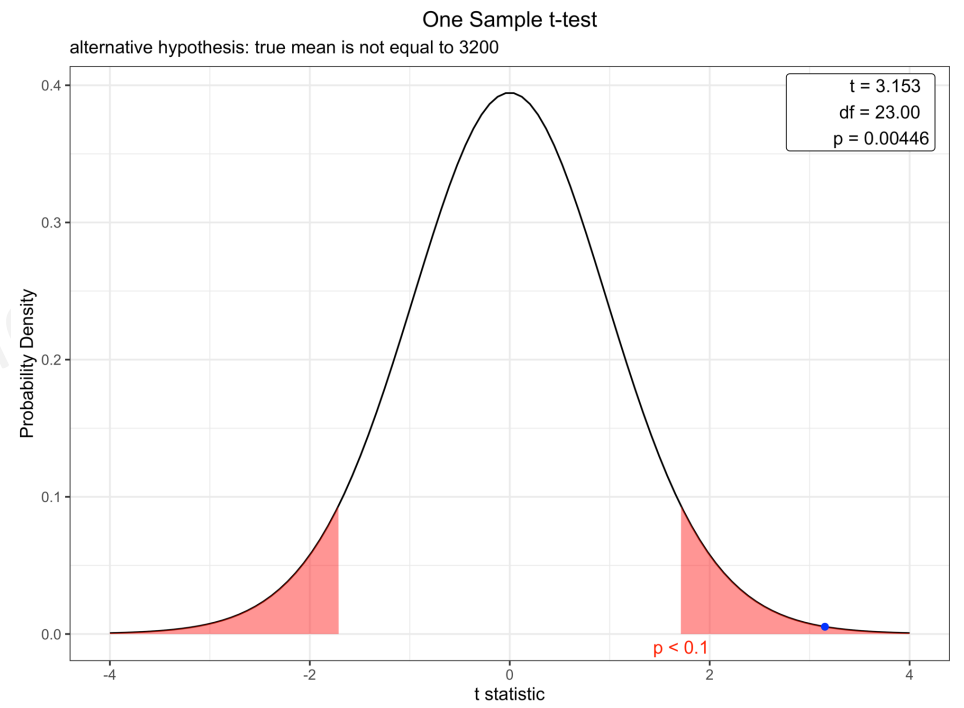
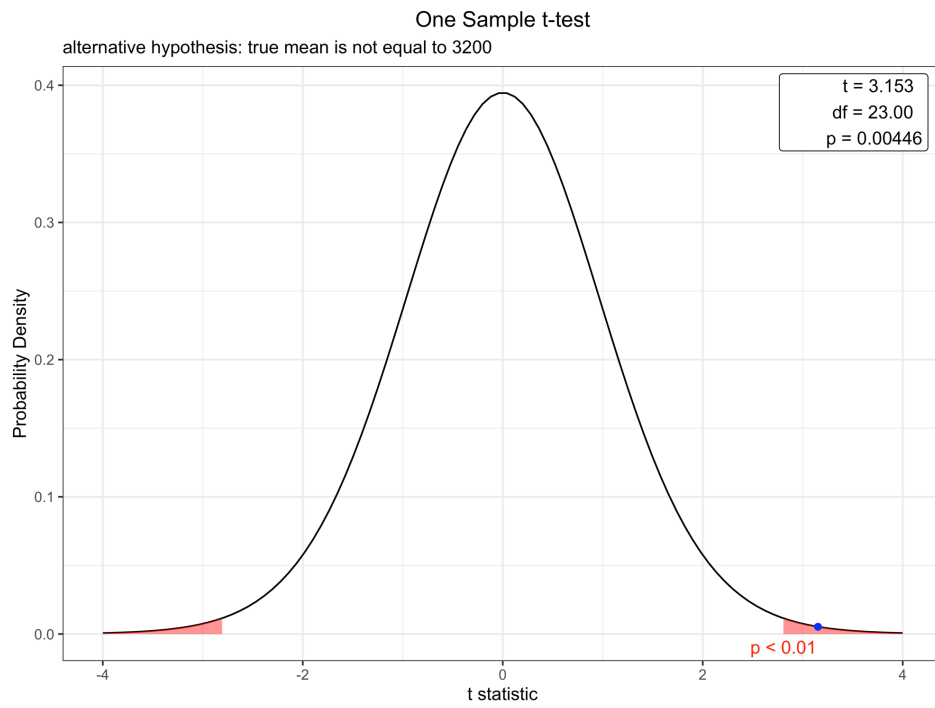
$\alpha=0.05$ anlamlılık düzeyi için

$1-\alpha=0.95$ testin güven düzeyini gösterir



bebek	kilo_dogum (gram)
1	3559,642
2	4427,398
3	3755,914
4	3674,335
5	4184,637
6	4004,514
7	4134,86
8	2501,323
9	3310,956
10	3618,676
11	4323,388
12	4462,213
13	2383,166
14	3571,462
15	3913,291
16	2847,233
17	2961,62
18	3913,627
19	4295,3
20	2761,597
21	3087,904
22	3286,258
23	4294,138
24	3156,766





T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
kilo_dogum	24	3601.25905	623.549771	127.281564

(1)

(2)

One-Sample Test

Test Value = 3200

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
kilo_dogum	3.153	23	.004	401.259054	137.95708	664.56103

(5)

(4)

(3a)(6)

(3b)

SPSS Çıktısı

(3b) Ortalama fark sütununda listelenen değer negatifse, ilgilendiğiniz değişkenin ölçüsü örnekleminiz için karşılaştırma popülasyonunuz için olduğundan daha *azdır*.

(5) *t* değerini iki ondalık basamağa kadar raporlayın. Eğer *t* değerinizi +1,00/-1,00'den küçükse, başına bir sıfır ekleyin (örneğin, .96 yerine 0,96 yazın)

(6) SPSS çıktısına göre tam *p* değerini iki veya üç ondalık basamağa kadar raporlayın. Ancak, *p* değeri .000 ise, < .001 olarak raporlayın. *p* değerimize baştaki sıfırı eklemeyin.

APA (American Psychological Association) Şablon Metni

Örneklemin tanımı ile popülasyonun tanımı arasında ilgililenen **değişken** açısından bir fark olup olmadığını değerlendirmek için tek örneklem *t* testi yapılmıştır. Örneklemin tanımı'ndaki ilgililenilen **değişken** ortalaması

($M = [ortalama]$ (1), $SD = [standart sapma]$ (2)) [**popülasyonun tanımı**]'na göre **önemli ölçüde daha**

fazlaydı/önemli ölçüde daha azdı/önemli ölçüde farklı değildi (3a ve 3b), $t([df](4)) = [t \text{ değeri}]$ (5), $p = [p \text{ değeri}]$ (6).

Bebek doğum kiloları ile genel popülasyona kıyasla bir fark olup olmadığını değerlendirmek için tek örneklem *t* testi yapılmıştır. Bebek doğum kiloları için ortalama (gram cinsinden) ($M=3601.35$, $SD=623.54$) genel popülasyona göre önemli ölçüde daha fazlaydı, $t(23)=1.15$, $p=.004$.

- Ortalama yeni doğan bebek ağırlığı (3601.25 ± 623.54) 3200 gram olan popülasyon 'normal' ağırlık değerinden daha yüksektir, $t(23)=3.15, p=.004$.
- Ortalama yeni doğan bebek ağırlığı, popülasyonun normal yeni doğan bebek ağırlığından istatistiksel olarak anlamlı derecede yüksektir, $t(23)=3.15, p=.004$.

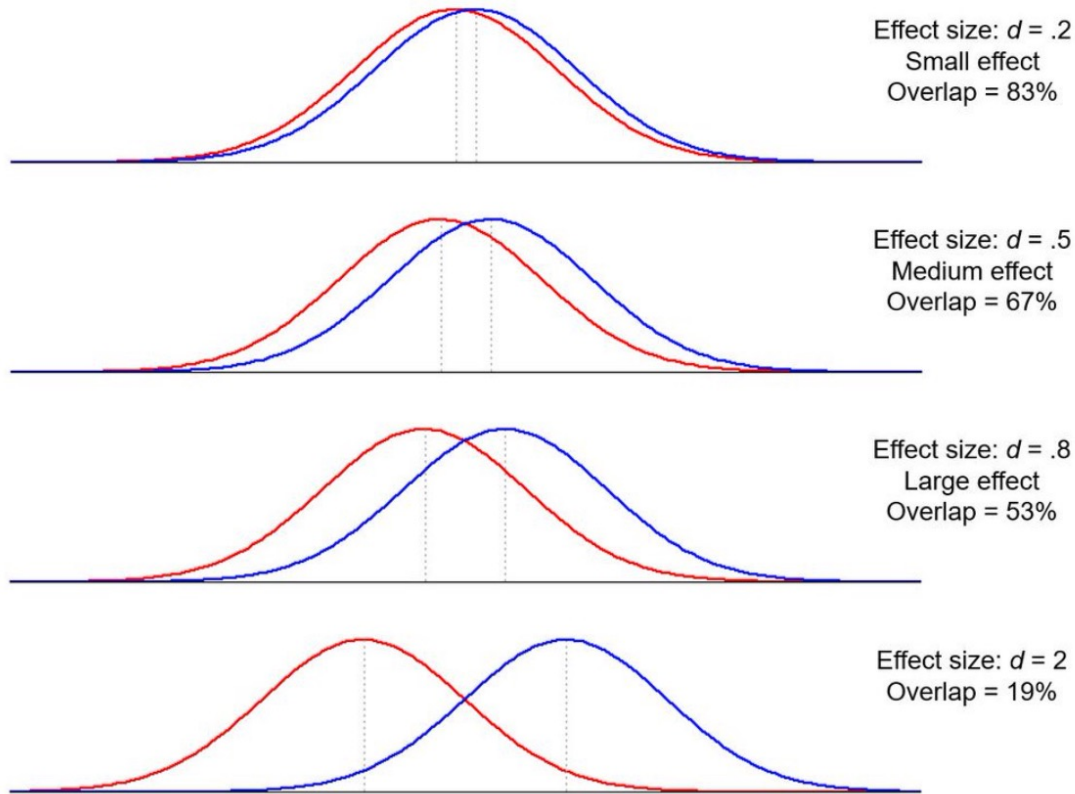
GÜVEN ARALIĞI RAPORA DAHİL EDİLİRSE

- Ortalama yeni doğan bebek ağırlığı, normal bebek ağırlığı olan 3200(gr)'dan istatistiksel olarak anlamlı şekilde 401.25 (%95 CI, 137.95 ila 664.56) daha yüksektir, $t(23)=3.15, p=.004$.
- Yeni doğan bebek ağırlığı, 3200 gram olan normal yeni doğan bebek skorundan ortalama 401.25 gram, %95 CI [137.95 ila 664.56] oranında istatistiksel olarak anlamlı derecede yüksekti, $t(23)=3.15, p=.004$.

Pratik Anlamlılık (Etki Büyüklüğü)

- Daha önce sıfır hipotezi anlamlılık testini kullanarak istatistiksel olarak anlamlı farklılıkları nasıl belirleyeceğinizi öğrenmiştiniz. Eğer p değeri α seviyesinden küçükse (tipik olarak $p < 0,05$), sonuçların **istatistiksel olarak anlamlı** olduğu kabul edilir. Varsayılan popülasyon parametresi ile gözlenen örneklem istatistiği arasındaki fark, şans eseri meydana gelmiş olma ihtimalinin düşük olduğu sonucuna varmak için yeterince büyük olduğunda sonuçların istatistiksel olarak anlamlı olduğu söylenir. Yeterince büyük bir örneklemimiz varsa, bu fark küçük olsa bile istatistiksel olarak anlamlı bir fark bulabiliriz. Dolayısıyla, pratik farklılıklarla da ilgilenmeliyiz.
- **Pratik anlamlılık, etki büyüklüğü** olarak bilinen farkın büyüklüğünü ifade eder. Sonuçlar, fark gerçek hayatta anlamlı olacak kadar büyük olduğunda pratik olarak anlamlıdır. Neyin anlamlı olduğu öznel olabilir ve bağlama göre değişebilir.

Etki Büyüklüğü



Etki büyüklüğü, ölçtüğümüz değişkenler arasındaki farkın veya ilişkinin büyüklüğünü veya boyutunu tanımlayan bir ölçüdür. Bu, pratik/anlamı önemi tanımladığı anlamına gelir. Bu bölümde göreceğimiz gibi, aslında her şey verilerin gruplarımız veya değişkenlerimiz arasında ne kadar örtüştüğüne bağlıdır. Eğer grupları farklılıklar açısından karşılaştırıyorsak, muhtemelen verilerde çok fazla örtüşme görmeyi beklemeyiz.

Verilerimizi benzerlik veya birliktelik açısından değerlendirirsek, daha fazla örtüşme ararız.

Bu çizgiler arasındaki bölge, her bir örnek için ortalamalardaki farkı temsil etmektedir.

Bu iki çizgi birbirinden ne kadar uzaksa, etki o kadar büyüktür. Başka bir şekilde ifade etmek gerekirse, grupların değerleri çok farklıysa, buradaki "tepeleri" birbirinden daha uzak olmalıdır, oysa benzer olsalardı (tepelere çok fazla örtüşür), bu gruplar arasındaki etkinin pratikte yararlı olmayacak kadar düşük olduğunu varsayabilirdik.

One-Sample Effect Sizes

		Standardizera ^a	Point Estimate	95% Confidence Interval	
				Lower	Upper
kilo_dogum	Cohen's d	623.549771	.644	.197	1.078
	Hedges' correction	644.847723	.622	.190	1.043

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Ortalama yeni doğan bebek ağırlığı, popülasyonun normal yeni doğan bebek ağırlığından istatistiksel olarak anlamlı derecede yüksektir, $t(23)=3.15$, $p=.004$, $d=0.644$.

Yeni doğan bebeklerin ortalama ağırlığı 3601.25, %95 CI [137.95, 664.56] olarak bulunmuştur.

Ortalama ağırlığın 3200 gr olduğu iddiası reddedilebilir, $t(23)=3.15$, $p=.004$, ihmal edilebilir bir etki büyüklüğü $d = .644$.