

**KATEGORİK DEĞİŞKEN**

İçerdiği ortak özelliklerine göre sınıflandırılan, belli kategoriler ya da gruplar altında toplanan verilere **kategorik değişken** denir. (cinsiyet, eğitim düzeyi, spor tercihi gibi)

İstatistiksel araştırmalarda **bağlam**, verilere dayalı bilgi üretme ihtiyacı duyulan gerçek yaşam durumlarıdır. Bağlam

- i) verilerin toplanma koşullarını
- ii) incelenen değişkenleri
- iii) araştırmanın amacını anlamamızı sağlamalıdır.

**İKİ KATEGORİK DEĞİŞKENİN İLİŞKİLİLİĞİNİ İÇEREN İSTATİSTİKSEL PROBLEM**

Kategorik veri dağılımları ile yürütülecek istatistiksel araştırmalarla

- i) bir veya daha fazla kategorik değişkenin dağılımını anlayabilir
- ii) değişkenler arasındaki ilişkiyi test edebilir
- iii) kategorik veriler arasındaki ilişkiye yönelik tahminler yapabiliriz

**Örnek...1 :**

Aşağıda verilen gerçek yaşam durumları için iki kategorik değişkenli veri toplamayı gerektiren bağlamlar oluşturunuz

Gerçek Yaşam Durumu	İki Kategorik Değişkenli Veri Toplamayı Gerektiren Bağlam
Eğitim	okul türü ve öğrencilerin ders dışı etkinliklere katılma durumu arası ilişki
Sağlık	Beslenme biçimi ve vücut kitle endeksi durumu arası ilişki

**İSTATİSTİKSEL ARAŞTIRMA SORUSU OLUŞTURMA**

İki kategorik değişken arasındaki ilişkinin ortaya koyulmasına yönelik istatistiksel araştırma soruları oluşturulurken uyulması gereken bazı temel ölçütler vardır. Bu ölçütlerle araştırmanın

- i) geçerliliği sağlanır
- ii) güvenilirliği sağlanır
- iii) elde edilen verilerin anlamlılığı arttırılır.

**TEMEL ÖLÇÜTLER**

1. **Amaç net olmalıdır.** Araştırma sorusunun amacı sorunun "ilişkilendiren" türde olmalıdır.
2. **Araştırmaya uygun olmalıdır.** Araştırma sorusu pratik olarak anlamlı bir problemi anlamaya veya çözmeye fayda sağlamalıdır.
3. **İlgilenilen grup (evren) açık olmalıdır.** Araştırmanın hangi grup ile ilgili olduğu açıkça belirtilmelidir.
4. **Değişken açık bir şekilde görülmelidir.** İki kategorik değişkenin ne olduğu net bir şekilde tanımlanmış olmalıdır.
5. **Veri toplanabilir olmalıdır.** İlgili veri ya mevcut olmalı ya da toplanabilir olmalıdır.
6. **Değişebilirliği yansıtmalıdır.** Değişkenlerde gözlemlenen farklılıklar verilerdeki değişebilirliği ortaya çıkarmaktadır.
7. **Odaklanılan grup, araştırma yapmaya imkân vermelidir.** Araştırmanın yapıldığı grup araştırmanın uygulanabilirliğini desteklemeli ve verilerin toplanması pratik olmalıdır.
8. **Kategorik veri toplamaya uygun olmalıdır.** Değişkenler kategorik olmalıdır yani belirli gruplara veya sınıflara ayrılabilir olmalıdır.

**Örnek...2 :**

Araştırma sorusu : A şehri B Lisesi öğrencilerinin cinsiyeti ile tercih ettikleri öğrenme yöntemi (örneğin: yüz yüze, çevrim içi, hibrit gibi) arasında anlamlı bir ilişki var mıdır?

Araştırma sorusunu iki kategorik değişkenli istatistiksel araştırma sorularında bulunması gereken ölçütlere göre inceleyiniz. Soruda ölçütlere uyulup uyulmadığını kontrol ediniz.

Araştırma Sorusu Ölçütleri	Durum
Açık ve nettir.	evet
Araştırmaya değerdir.	evet
İlgilenilen grup (evren) açıktır.	evet
Değişken açık bir şekilde görülmektedir.	evet
Veri toplanabilir.	evet
Değişebilirliği yansıtmaktadır.	evet
Odaklanılan grup araştırma yapılmasına imkân vermektedir	evet
Kategorik veri toplamaya uygundur.	evet

**Örnek...3 :**

A kentinde bulunan B alışveriş merkezi çalışanlarında sigara kullanma durumu (kullanıyor / kullanmıyor) ile kronik hastalık varlığı (var / yok) arasında anlamlı bir ilişki var mıdır?

Yukarıda verilen araştırma sorusunu iki kategorik değişkenli istatistiksel araştırma sorularında bulunması gereken ölçütlere göre inceleyiniz. Soruda ölçütlere uyulup uyulmadığını kontrol ediniz.

Araştırma Sorusu Ölçütleri	Durum
Açık ve nettir.	evet
Araştırmaya değerdir.	evet
İlgilenilen grup (evren) açıktır.	evet
Değişken açık bir şekilde görülmektedir.	evet
Veri toplanabilir.	evet
Değişebilirliği yansıtmaktadır.	evet
Odaklanılan grup araştırma yapılmasına imkân vermektedir	evet
Kategorik veri toplamaya uygundur.	evet

### İKİ KATEGORİK DEĞİŞKENLİ VERİ TOPLAMA PLANI YAPMA, VERİLERİ TOPLAMA VE ANALİZE HAZIR HÂLE GETİRME

Bir problemin çözümünde sistematik olmanın belirsizlikleri azaltma ve problemi çözmeye yönelik daha net bir yol haritası sunmasının yanında :

**Şeffaflık:** Sürecin açık ve izlenebilir olmasından dolayı ilerlemeyi ölçmenin kolaylaşması

**Hataları azaltma:** Adım adım ilerleme, hataların gecikmeden tespit edilmesi

**Etkinlik:** Zaman ve kaynakların daha verimli kullanılması.

**Karmaşıklığı azaltma:** Problemin yönetilebilir alt problemlere bölünmesi gibi yararları vardır.

### İKİ KATEGORİK DEĞİŞKEN İÇEREN İSTATİSTİKSEL ARAŞTIRMALARDA VERİ TOPLAMA PLANININ AŞAMALARI

Araştırmalarda verilerin toplanması güvenilir sonuçlar elde etmek için oldukça önemlidir. Veri toplamanın aşamalarını özetlersek:

#### 1. Veri Toplama Aracının Oluşturulması veya Belirlenmesi

Kullanılacak verilerin toplanması için gözlem formu , anket, test ve mülakat gibi uygun araçların belirlenmesi veya geliştirilmesidir.

#### 2. Evren ve Örneklem Belirlenmesi

Hedef kitle (evren) ve bu evrenden seçilecek örneklerin (örneklem) belirlenmesidir. Örneklem , sonuçların genellenebilmesine imkan verecek ölçüde yeterince büyük ve temsilî olmalıdır.

#### 3. Rastgeleliğin Sağlanması

Örneklem rastgele yöntemlerle seçilmesidir. Rastgelelik, her bir bireyin örnekleme seçilme şansının eşit olduğu bir yöntemle sağlanmalıdır.

#### 4. Değişkenlerin Belirlenmesi

İncelenmek istenen özelliklerin veya ölçütlerin tanımlanmasıdır. Değişkenler, araştırma sorusuna göre belirlenir.

#### 5. Verilerin Nerede, Ne Zaman, Nasıl ve Kimler Tarafından Toplanacağı Belirlenmesi

Verilerin sistematik bir şekilde toplanması için nasıl ve ne zaman toplanacağı, hangi süreçlerin izleneceği, kimlerin sorumlu olacağı gibi detaylar önceden planlanmalıdır.

#### 6. Verilerin Nasıl Kaydedileceğinin Belirlenmesi

Toplanan veriler elektronik tablolar veya istatistiksel yazılımlar kullanılarak dijital ortama aktarılır.

#### 7. Verilerin Gizliliğinin Sağlanması, Nesnel Ve Dürüst Olunmasına Yönelik Önlemler Alınması

Süreç boyunca etik kurallara uyulmalıdır. Araştırmaya katılım gönüllü olmalı, katılımcıların gizlilikleri korunmalı ve onayları alınmalıdır.

Veri toplama yöntemleri **birincil veri** ve **ikincil veri** olarak iki ana kategoride sınıflandırılabilir.

### 1. Birincil Veri Toplama Yöntemleri

Birincil veriler, araştırmacının özel gereksinimlerine göre araştırmacı tarafından anket, gözlem, deney, test, mülakat ve görüşmeler yoluyla doğrudan topladığı orijinal verileridir.

### 2. İkincil Veri Toplama Yöntemleri

İkincil veriler, başka bir araştırmacı veya kurum tarafından daha önce toplanmış olan resmî istatistikler, akademik yayınlar ve raporlar, kurumsal veri tabanları, medya ve haber kaynakları, dijital veri tabanları ve internet kaynakları gibi hazır verilerdir. Kaynakların güvenilirliğini, güncelliğini ve uygunluğunu değerlendirmek çok önemlidir.

Dijital veri kaynakları kullanılırken göz önünde bulundurulması gereken önemli noktalar:

1. Kaynağın Güvenilirliği
  2. Verilerin Güncelliği
  3. Verilerin Kapsamı ve Uygunluğu
  4. Veri Lisansı ve Erişim Hakları Verilerin telif hakkı, lisanslama durumu ve paylaşım politikaları
  5. Veri Kaynaklarının Referanslanması
  6. Veri Gizliliği ve Güvenlik
  7. Farklı Kaynaklarla Karşılaştırma
- gibi hususlara dikkat etmek gerekmektedir.

### Örnek...4 :

A lisesi B okulunun çalışanları için alışveriş tercihi (çevrimiçi /mağaza) ile medeni durum arasında ilişki var mıdır? Yukarıda verilen araştırma sorusu için bir veri toplama planı yapınız

### Veri Toplama Planı

- a) Araştırma sorularına cevap bulmayı sağlayacak veri toplama araçlarının belirlenmesi: **anket veya çevrim içi bir form**
- b) Evren ve örneklemin belirlenmesi  
Evren: B okulunun tüm çalışanları  
Örneklem : Evli ve bekarlardan seçilecek yeterli katılımcı
- c) Rastgeleliğin sağlanması  
Rastgele örnekleme yöntemi kullanılacaktır.

### d) Değişkenlerin belirlenmesi

Değişken 1 Medeni Durum  
Değişken 2 Alışveriş tercihi

### e) Verilerin nerede, ne zaman, nasıl ve kimler tarafından toplanacağını belirlenmesi

Anket veya çevrim içi bir form okullarda veya çevrim içi platformlarda eğitim-öğretim dönemi içerisinde araştırma ekibi tarafından toplanabilir.

### f) Verilerin nasıl kaydedileceğinin belirlenmesi

dijital olarak toplanan veriler elektronik tablo kullanılarak , anketler yapılırsa toplanan veriler dijital forma dönüştürülür.

### g) Verilerin gizliliğine dikkat edilmesi, dürüst ve nesnel olunması

Gizlilik: Çalışanların isimleri veya kimlik bilgileri yerine anonim kodlar kullanılacaktır.  
• Nesnellik ve Dürüstlük: Anket soruları tarafsız ve yönlendirici olmayan bir dille olacaktır  
Onay: Araştırmanın yapılabilmesi için gerekli etik onaylar alınacaktır

## İKİ KATEGORİK VERİYE DAYALI İSTATİSTİKSEL ARAŞTIRMALARDA VERİ ANALİZİ YAPMA VE SONUÇLARI YORUMLAMA

Kategorik veri analizi toplanan verilerden anlamlı bilgiler çıkarma sürecidir. Temel amaç, veriyi işlemek suretiyle belirli kategoriler arasındaki ilişkiyi ortaya çıkarmak, gruplar arası karşılaştırmalar yapmak ve sonuçları yorumlamaktır.

İki kategorik değişken arasındaki ilişkiyi hem sayısal hem de görsel olarak keşfetmeye ve analiz etmeye yardımcı olması için yaygın olarak

- i) iki yönlü tablo (çapraz tablo)
- ii) kümeli sütun grafikleri kullanılır.

### 1. İki Yönlü Tablo (Çapraz Tablo)

İki farklı kategorideki sıklıkları bir tablo halinde sunar. Tabloda kategorik değişkenlerden biri satırda diğeri ise sütunda temsil edilir. Tablodaki her hücre, iki değişkenin kesiştiği noktada gözlenen sıklıkları gösterir.

**Örnek...5 :**

Bir okulda öğrencilere hangi tür kitapları okumayı sevdikleri sorulmuş ve toplanan veri aşağıdaki iki yönlü bir tabloda gösterilmiştir.

Tabloda kategorilerden biri, öğrencilerin sınıf seviyesini, diğeri ise en çok okumayı sevdikleri kitap türünü göstermektedir. Bu tabloda iki değişkenin kesiştiği noktada gözlenen değerler öğrenci sayılarıdır.

**Tablo 1:** Sınıf seviyeleri ile sevilen kitap türlerini gösteren iki yönlü tablo

Okunan Kitap Türü \ Sınıf seviyesi	Macera	Korku	Romantik	Toplam
9	16	14	20	50
10	20	16	16	52
11	30	10	10	50
12	34	10	4	48
Toplam	100	50	50	200

Görelilik, bir kategorik değişkenin belirli bir değerinin, toplam gözlem sayısına oranıdır.

**Görelilik tablosu**, iki yönlü tablodaki kategorik değişkenlerin sıklıklarını oran olarak ifade ettiği tablodur. Tabloda her hücredeki değer, toplam öğrenci sayısına bölünerek hesaplanır.

**Tablo 2:** Sınıf seviyeleri ile sevilen kitap türlerini gösteren görelilik tablosu

Okunan Kitap Türü \ Sınıf seviyesi	Macera	Korku	Romantik	Toplam
9	$\frac{16}{200} = 0,08$	0,07	0,10	0,25
10	0,10	0,08	0,08	0,26
11	0,15	0,05	0,05	0,25
12	0,17	0,05	0,02	0,24
Toplam	0,5	0,25	0,25	1

**Koşullu Görelilik Sıklıkları**

Koşullu görelilik sıklık bir kategorik değişkene bağlı olarak diğ kategorik değişkenin dağılımını yüzde cinsinden temsil eder ve bu sayede iki değişken arasındaki ilişkililiği daha net bir şekilde ortaya koyar.

Koşullu görelilik sıklık, iki yönlü tabloda her bir hücredeki değer, ait olduğu kategorinin toplam değerlere oranlanmasıyla hesaplanır.

**Tablo 3:** Sınıf seviyesine göre sevilen kitap türleri koşullu görelilik sıklıkları tablodaki gibidir.

Okunan Kitap Türü \ Sınıflar seviyesi	Macera	Korku	Romantik	Toplam
9	$\frac{16}{50} \cdot 100 = \%32$	%28	%40	%100
10	%38,46	%30,76	%30,76	%100
11	%60	%20	%20	%100
12	%70,8	%20,8	%8,4	%100

**Tablo 4:** Sevilen kitap türlerine göre sınıf seviyeleri koşullu görelilik sıklıkları tablodaki gibidir.

Okunan Kitap Türü \ Sınıf seviyesi	Macera	Korku	Romantik
9	%16	%28	%40
10	%20	%32	%32
11	%30	%20	%20
12	%34	%20	%8
Toplam	100	100	100

**Özetlersek :** A kategorisine göre B kategorisi koşullu görelilik sıklık tablosu oluşturulurken A kategorisine ait sıklıkların B deki kategorilere göre dağılışı yüzde olarak ifade edilir.

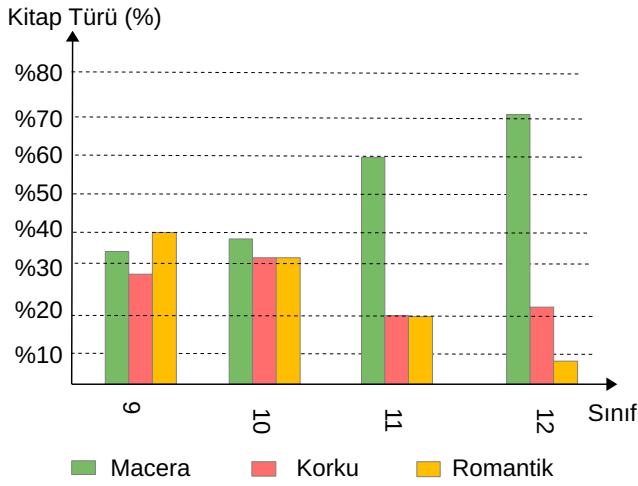
## 2. Kümeli Sütun Grafikleri

İki kategorik değişkenin karşılaştırmasını görsel olarak sunmak için kullanılan grafik türüdür. Grafikte her bir kategori için sütunlar kümelenmiş olarak gösterilir.

### Örnek...6 :

Tablo3 e bakarak sınıf seviyesine göre sevilen kitap türlerinin dağılımını aşağıdaki gibi kümeli sütun grafiğinde temsil edebiliriz.

**Grafik: 2.1**

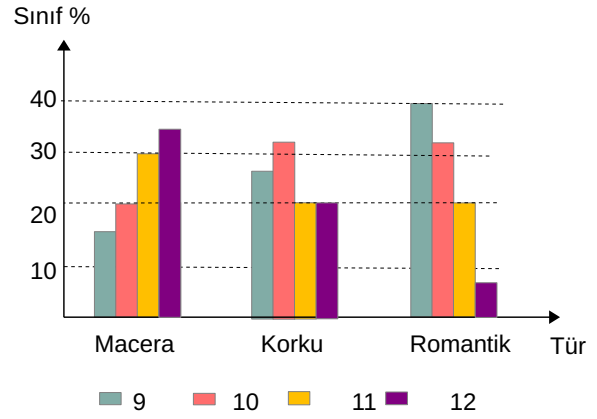


Kümeli sütun grafiklerinde her bir kategori için sütunlar kümelenmiş olarak gösterilmesi sayesinde her bir kategori içindeki dağılım farklılıklarını görmek mümkün olur.

### Örnek...7 :

Tablo4 e bakarak sevilen kitap türlerine göre sınıf seviyesi dağılımını aşağıdaki gibi kümeli sütun grafiğinde temsil edebiliriz.

**Grafik: 2.2** Sevilen kitap türlerine göre sınıf seviyelerinin dağılımı



İki kategorik değişkenli verilerde kullanılan görselleştirme araçları (çapraz tablolar, koşullu göreceli sıklık tabloları, kümeli sütun grafikler vb.) değişkenler arasında bir ilişki durumu olup olmadığını görsel olarak anlamaya olanak tanır.

**İlişkilsizlik** İki değişkenin birbiriyle bağımlı (bir değişkendeki farklı kategorilerin diğer değişken kategorileri ile belirli bir örüntüde ilişki göstermesi) olup olmamasını temsil eder. İlişkili olan iki kategorik değişken arasında neden sonuç ilişkisi olmayabilir.

Özetlersek, iki kategorik değişken arasındaki ilişki analiz edilirken yalnızca değişkenler arasında gözlemlenen bir ilişki bulunduğu söylenebilir, fakat bu ilişkinin neden sonuç anlamı taşıdığına karar verilmesi için ek araştırmalar ve analizler gereklidir.

**Örnek...8 :**

Bir tiyatro oyununa gösterimde olduğu süre içerisinde giden kişiler için kantinde satılan 3 çeşit içecek türüne ait tercihler merak edilmektedir.

Yaş gurubu (genç /yetişkin) ve içecek tüketim tercihi (çay/kahve/portakal suyu) arasında bir ilişki mevcut mudur?

Yukarıdaki araştırma sorusuna ait toplanan veri aşağıdaki iki yönlü bir tabloda gösterilmiştir.

İçecek Tercihi / Yaş Grubu	Çay	Kahve	Portakal Suyu	Toplam
Genç	20	30	50	100
Yetişkin	60	30	10	100
Toplam	80	60	60	200

a) Buna göre yaş grubu ile içecek tercihi türlerini gösteren görelî sıklık tablosunu yapınız

b) yaş grubuna göre içecek tercihinin koşullu görelî sıklık tablosunu yaparak kümeli sütun grafiğini çiziniz

c) içecek tercihinine göre yaş grubunun koşullu görelî sıklık tablosunu yaparak kümeli sütun grafiğini çiziniz

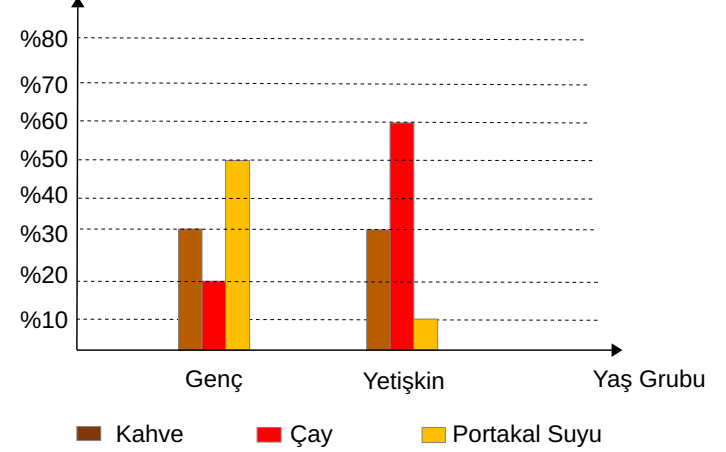
a)

İçecek Tercihi / Yaş Grubu	Çay	Kahve	Portakal Suyu	Toplam
Genç	0,1	0,15	0,25	0,5
Yetişkin	0,3	0,15	0,05	0,5
Toplam	0,4	0,3	0,3	1

b) yaş grubuna göre içecek tercihinin koşullu görelî sıklık tablosu

İçecek Tercihi / Yaş Grubu	Çay	Kahve	Portakal Suyu	Toplam
Genç	%20	%30	%50	%100
Yetişkin	%60	%30	%10	%100

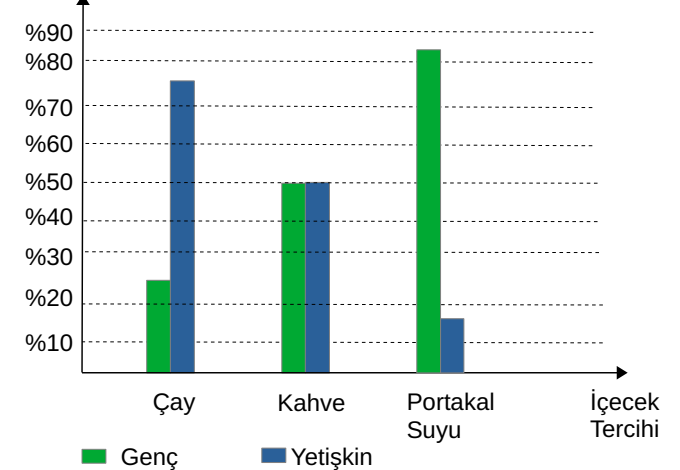
İçecek Türü (%)



c) içecek tercihinine göre yaş grubunun koşullu görelî sıklık tablosu

İçecek Tercihi / Yaş Grubu	Çay	Kahve	Portakal Suyu
Genç	%25	%50	%83,33
Yetişkin	%75	%50	%16,66
Toplam	%100	%100	%100

Yaş Grubu (%)



### BAŞKALARI TARAFINDAN OLUŞTURULAN İKİ KATEGORİK DEĞİŞKENLİ VERİ DAĞILIMLARINA DAYALI SONUÇ VEYA YORUMLARI TARTIŞMA

Başkaları tarafından oluşturulan veri dağılımlarına dayalı sonuç veya yorumlar, doğru bir şekilde değerlendirilmezse hatalı yönlendirmelere neden olabilir. Bu tür analizlere eleştirel bir gözle bakmak ve yapılan yorumların geçerliliğini sorgulamak büyük önem arz eder.

İki kategorik değişkenli verilere dayalı istatistiksel sonuç ve yorumların geçerliliğini değerlendirme süreci:

- istatistiksel yorumların doğruluğunu artırma
- yanlılığı en aza indirme
- sonuçların güvenilirliğini sağlama anlamında kritik öneme sahiptir.

Sonuç ve yorumların geçerliliğini değerlendirmede

- kullanılan veri toplama yöntemi
- örnekleme yöntemi
- analiz araçlarının uygunluğu
- yorumların verilere dayalı olması gibi unsurlar göz önünde bulundurulur.

Bu süreçte yapılabilecek hataların bazılarını şu şekilde sayabiliriz :

- Örnekleme yanlılığı
- Küçük örneklem büyüklüğü
- Yanlış ölçüm
- Analizde kullanılan görselleştirme araçlarının seçimi
- Yorumların varsayımlara dayalı olarak yapılması

Sonuç olarak elde edilen istatistiksel sonuç ve yorumlar, eldeki verilere ve yapılan analizlere dayalı olarak, verilerin bulunduğu bulgularla uyumluysa ve mantıksal temellere dayandırılıyorsa kabul edilir ya da aksi takdirde geçersizliği gösterilir ve doğru analiz yöntemleriyle düzeltilir.

### Örnek...9 :

Bir araştırmacı, her gün atılan adım sayısı ile kan kolesterol düzeyi arasında bir ilişki bulunmuş ve günde on bin adımdan fazla atanların kan kolesterol seviyelerinin daha düşük olduğu sonucuna varmıştır. Yapılan araştırmalar sonucu elde edilen veriler aşağıdaki iki yönlü tabloda verilmiştir.

	Kolesterol seviyesi <200 mg/dL	Kolesterol seviyesi >200 mg/dL
Adım Sayısı <10000	20	80
Adım Sayısı >10000	90	10

Buna göre

- Araştırmacının yaptığı çıkarımın dayanağı nedir?
- Araştırmacının yaptığı çıkarıma yönelik tespit ettiğiniz hatalar/yanlılıklar neler olabilir?
- Araştırmacının tespit ettiğiniz hatalar/yanlılıklar varsa gerekçesini sunarak doğru sonucu yazınız.

a) Uzmanın yaptığı çıkarım, atılan adım sayıları ile kan kolesterol seviyeleri durumu arasındaki verilere dayanmaktadır.

b) Nedensellik Yanılgısı: Atılan adım sayısı ile kilo alma arasında bir ilişki bulunsa bile bu durum atılan adım sayısının doğrudan kolesterol seviyesini düşürmeye neden olduğunu göstermez. Kolesterolü etkileyen faktörler sadece atılan adım sayısı ile ilgili değildir; beslenme ,genetik faktörler, yaşam tarzı fiziksel başka türlü aktiviteler gibi başka birçok değişken de kan kolesterol seviyesini etkileyebilir.

Eksik Değişken Yanlılığı: Kan kolesterol seviyesini etkileyen diğer değişkenler (diyet alışkanlıkları, fiziksel aktivite düzeyi, uyku düzeni vb.) dikkate alınmamıştır.

Seçim Yanlılığı: Verilerde fazla adım atanların durumu hakkında tam bir temsil durumu olmayabilir. Fazla adım atanlar daha aktif, diyetine özen gösteren veya kilo kontrolü yapan kişiler olması mümkün olduğundan bu durum verileri yanıltıcı bir biçimde etkileyebilir.

c) Araştırmacı, yalnızca iki kategori arasında gözlemlenen dağılımdan yola çıkarak nedensel bir ilişki kurmuş ve diğer etkileyici faktörleri göz ardı etmiştir. 10 bin adımdan daha fazla adım atmak tek başına düşük kolesterolün doğrudan bir sebebi olamaz, kan kolesterol seviyesini etkileyen daha fazla faktör dikkate alınmalıdır.

Doğru bir analiz yapabilmek için araştırmacının sadece atılan adım sayısı durumu ile kan kolesterol durumu arasındaki ilişkiye değil gün içerisinde yapılan fiziksel aktivite seviyesi, uyku düzeni gibi ve başka faktörleri de göz önünde bulundurması gerekir. Bu şekilde yapılacak çok değişkenli analizler, 10 bin adımdan fazla atmanın kan kolesterol seviyesi üzerindeki bağımsız etkisini daha doğru bir şekilde ölçmeye imkân tanır.

Doğru bir ifade şu şekilde olabilir:

"Veriler, 10 binden fazla adım atan bireylerin daha düşük kan kolesterol seviyesi değerine sahip olma eğiliminde olduğunu gösterse de bu durum doğrudan 10 binden fazla adım atılmaktan kaynaklanıyor demek için yeterli değildir. Kan kolesterol seviyesini etkileyen diğer faktörler de göz önünde bulundurulmalıdır."