

İstatistiksel Analize Yeni Başlayanlar İçin Bir Kılavuz

Prof.Dr. Ömer Cevdet Bilgin

İstatistiksel analiz *veri* kullanarak trendleri, biçimleri ve ilişkileri soruşturmak demektir. Sadece bilimcilerin değil aynı zamanda hükümetlerin, işletmelerin ve diğer organizasyonların da kullandığı önemli bir araştırma aracıdır.

Geçerli neticeler çıkarmak için, istatistiksel analizi, *araştırma prosesinin* en başında, dikkatlice planlaman gerekir. Hipotezini spesifikleştirmeli ve araştırma dizaynını, örnek hacmini ve örnekleme prosedürünü kararlaştırmalısın.

Hedef popülasyondan çektiğin örnekten verini topladıktan sonra, *betimleyici istatistikleri* kullanarak verini organize edebilirsin ve özetleyebilirsin. Daha sonra, formal olarak hipotezlerini test etmek için ve popülasyon hakkında kestirim yapmak için *çıkarımsal istatistikleri* kullanabilirsin. En sonunda, bulgularının interpretasyonunu yapabilirsin ve onları genelleştirebilirsin.

Bu yazıdan maksat, öğrenciler ve araştırmacılar için istatistiksel analize pratik bir giriş temin etmektir. İki araştırma misali üzerinden, 5-adım yürüyeceğiz. Birinci misalimiz değişkenler arasında bir potansiyel *neden-ve-etki ilişkisini* soruşturuyorken, ikincisi bir potansiyel *çağrışımsal ilişkiyi* soruşturuyor.

Misal 1. Nedenel araştırma sorusu: Hipnotik terapi gençlerin sınav performansını yükseltir mi? (Neden-ve etki ilişkisi, deneysel öğrenme-çalışması)

Misal 2. Çağrışımsal araştırma sorusu: Aile geliri ve üniversiteye giriş sınavı puanı arasında bir ilişki var mı? (Çağrışımsal ilişki, gözlemsel öğrenme-çalışması)

Adım 1. Hipotezleri yazmak ve araştırma dizaynını planlamak:

İstatistiksel analiz için geçerli veri toplamak üzere, ilk önce *hipotezlerini* spesifikleştirmeye ve araştırma dizaynını planlamaya ihtiyacın var.

İstatistiksel hipotezleri yazmak

Araştırmanın amacı çoğu zaman bir *popülasyondaki* değişkenler arasındaki ilişkileri soruşturmadır. Bir tahminle başlarsın ve bu tahmini test etmek için istatistiksel analiz kullanırsın.

Bir istatistiksel hipotez bir popülasyon hakkındaki bir tahmini yazmanın formal bir yoludur. Her araştırma tahmini örnek veri kullanılarak test edilebilecek bir *sıfır hipotezi* ve bir *alternatif hipotez* formunda yeniden-ifade edilir.

Sıfır hipotezi her zaman etkinin olmadığını veya değişkenler arasında ilişki olmadığını tahmin ederken, alternatif hipotez bir etkiyi veya ilişkiyi tahmin eder.

Misal 3. Bir etkiyi test etmek için kurulan istatistiksel hipotezler (neden-ve-etki-ilşkisi):

Sıfır hipotezi: Bir saatlik bir hipnoz seansının gençlerin matematik sınavı skorlarına etkisi olmayacaktır.

Alternatif hipotez: Bir saatlik bir hipnoz seansı gençlerin matematik sınavı skorlarını artıracaktır.

Misal 4. Bir korelasyonu test etmek için kurulan istatistiksel hipotezler (çağrışımsal ilişki):

Sıfır hipotezi: Öğrencilerin aile geliri ve üniversiteye giriş sınavı puanı arasında bir ilişki yoktur.

Alternatif hipotez: Öğrencilerin aile geliri ve üniversiteye giriş sınavı puanı arasında pozitif bir ilişki vardır.

Araştırma dizaynını planlamak

Araştırma dizaynı senin veri toplama ve analizi için uygulayacağın genel stratejindir. Bu strateji, aynı zamanda, daha sonra hipotezlerini test etmek için kullanabileceğin istatistiksel testleri belirler.

İlk önce, araştırmanın bir deneysel dizayn mı, bir gözlemsel dizayn mı yoksa bir betimleyici dizayn mı kullanacağına karar vermelisin. Deneyler *rastgele atama* vasıtasıyla açıklayıcı değişkende meydana getirilen değişikliğin yanıt değişkenine etkisini ölçerken (deneklere müdahale edilir), gözlemsel ve betimleyici öğrenme-çalışmaları değişkenleri kendi doğal ortamında gözleyerek çıkımlarını kaydetmekten ibarettir (deneklere müdahale edilmez).

- Bir *deneysel dizaynda*, bir neden-ve-etki ilişkisini (mesela sınav skorlarına hipnozun etkisi gibi) değerlendirebilmek için, istatistiksel karşılaştırma testleri veya regresyon analizi kullanırsın.
- Bir *gözlemsel dizaynda*, herhangi bir nedensellik varsaymaksızın, değişkenler arasındaki ilişkiyi (mesela aile geliri ve üniversiteye giriş sınavı puanı arasındaki çağrışım gibi) keşiflemek için, korelasyon katsayıları, çapraz tablolar ve anlamlılık testleri kullanırsın.
- Bir *betimleyici dizaynda*, bir fenomeni, bir durumu veya bir popülasyonu betimlemek için veri toplarsın. Değişkenleri yalnızca teşhis eder, gözler ve ölçersin. (Araştırma problemine dair niçin sorusundan daha ziyade, ne, ne zaman, nerede ve nasıl sorularını cevaplamaya matuftur.)

Araştırma dizaynının aynı zamanda denekleri grup düzeyinde mi, birey düzeyinde mi yoksa her iki düzeyde mi karşılaştıracağını alakalıdır.

- Bir *denekler-arası dizaynda*, farklı muamelelere tabi tutulan deneklerin (mesela hipnoz uygulananlar ve hipnoz uygulanmayanlar gibi) çıkımlarını grup-düzeyinde karşılaştırırsın.
- Bir *denekler-içi dizaynda*, öğrenme-çalışmasının tüm muamelelerine tabi tutulan deneklerin yinelenen ölçümlerini (mesela hipnoz uygulamasından önceki skorlar ve sonraki skorlar gibi) karşılaştırırsın.
- Bir *karışık dizaynda*, bir değişken denekler arasında ve bir diğer değişken denekler içinde değiştirilir (mesela hipnoz uygulanan ve uygulanmayan deneklerin uygulamadan önceki ve sonraki skorları gibi).

Misal 5. Deneysel araştırma dizaynı. Bir saatlik hipnoz seansının öğrencilerin matematik sınavı skorlarını yükseltip yükseltmediğini öğrenmeye-çalışmak için, bir denekler-içi deney dizaynı edebilirsin. Bu öğrenme-çalışmasında, bir grup denneğin yinelenen ölçümlerini alırsın.

Önce, deneklerinin baz-çizgisi sınav skorlarını alırsın. Sonra, deneklerine bir saatlik hipnoz uygulanır. En sonunda, deneklerinin ikinci bir matematik sınavındaki skorlarını kaydedersin.

Bu deneyde, *açıklayıcı değişken* hipnoz uygulamasıdır ve *yanıt değişkeni* müdahaleden önceki ve sonraki sınav skorlarıdır.

Misal 6. Gözlemsel araştırma dizaynı. Bir gözlemsel öğrenme-çalışmasıyla, liseden mezun olan öğrencilerin aile gelirleri ve üniversiteye giriş sınavı puanları arasında bir ilişki olup olmadığını test edebilirsin. Verini toplamak için, deneklerden bir anket formunu doldurmalarını ve ailelerinin gelirini ve üniversiteye giriş sınavı puanını kendilerinin rapor etmesini isteyeceksin.

Burada *açıklayıcı değişkenin* düzeyini kasten kendin belirlemediğin için ve denekleri muamele gruplarına rastgele atamadığın için, bir ilişkinin varlığına dair kanıt bulsan dahi, bunun yalnızca bir çağrışımsal ilişki olduğunu söyleyebileceksin.

Kontrol-altına-alınmış deneyler neden-ve-etki ilişkilerini tesis etmenin en iyi yoludur. Aralarındaki yegane manalı farkın uyguladığımız muameleden kaynaklandığını gösterecek şekilde, grupları oluşturmayı başatabildiğimiz ölçüde, deneyimizin içsel geçerliliğe sahip olduğunu söyleriz. Bu durumda, gruplar arasındaki içsel karşılaştırma geçerli bir karşılaştırmadır. İçsel geçerlilik ve rastgele atama arasındaki münasebete dikkatini çekmek isterim.

Değişkenleri ölçmek

Bir araştırma dizaynı planlarken, değişkenlerini *operasyonelleştirmelisin* ve onları tam olarak nasıl ölçeceğine karar vermelisin.

İstatistiksel analiz için, değişkenlerinin ölçüm düzeylerini mülahaza etmen önemlidir öyle ki bu sana değişkenlerinin hangi tipte veri üreteceğini söyleyecek:

- *Kategorik veri* sınıflandırmaları ihtiva eder. Bunlar *adlandırılmış* (mesela, cinsiyet, göz rengi vb.) veya *sıralanmış* (mesela, dil yeteneği düzeyi, memnuniyet düzeyi vb.) olabilirler.
- *Sayısal veri* miktarları ihtiva eder. Bunlar bir *aralık ölçeğinde* (mesela, sıcaklık vb.) veya *oran ölçeğinde* (mesela, ağırlık, uzunluk vb.) olabilirler.

Pek çok değişken farklı kesinlik düzeylerinde ölçülmeye müsaittir. Mesela, yaş verisi sayısal (13, 21, 48, 56 veya 86 yaşında) olabileceği gibi kategorik (genç, orta yaşlı veya yaşlı) te olabilir.

Sayısal değer alan değişkenlere *nicel değişken* ve kategorik değer alan değişkenlere *nitel değişken* adı verilir. Nitel bir değişkenin kategorik değerlerini sayısal değerler ile yeniden kodlamak, o değişkeni otomotikman nicel bir değişken yapmaz. Mesela bir futbol takımındaki oyuncuların formalarındaki numaralar, bu oyuncuları toplayıp, çıkarıp, çarpıp veya bölebileceğimiz ve hatta bu numaralara göre bir sıraya koyabileceğimiz manasına gelmez.

Ölçüm düzeylerinin teşhisi, mütenasip istatistiklerin ve hipotez testlerinin seçimi için önemlidir. Mesela, sayısal veri için bir aritmetik ortalama skor hesaplayabilirsin, fakat bunu kategorik veri için yapamazsın.

Bir araştırma öğrenme-çalışmasında, ilgilendiğin değişkenlere ait ölçümlerin yanısıra, çoğu zaman, deneklerin karakteristikleriyle münasebetli veri toplarsın.

Misal 7. Değişkenler (deney).

Değişken	Veri tipi
Yaş	Sayısal (oran)
Cinsiyet	Kategorik (adlandırılmış)
Baz-çizgisi sınav skoru	Sayısal (oran)
Son sınav skoru	Sayısal (oran)

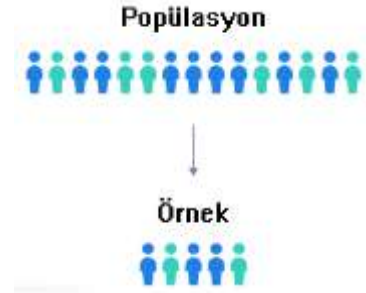
Misal 8. Değişkenler (Gözlemsel öğrenme-çalışması).

Değişken	Veri tipi
Aile geliri	Sayısal (oran)
Üniversiteye giriş sınavı puanı	Sayısal (oran)

Adım 2. Bir örnekten veri toplamak

Çoğu durumda, öğrenmeye-çalıştığın *popülasyonun* her bir üyesinden veri toplamak çok zor veya pahalıdır. Onun yerine, bir *örnekten* veri toplarsın.

Mütenasip *örnekleme prosedürlerini* kullandığın sürece, istatistiksel analiz, bulgularını popülasyona genelleştirebilmeni temin eder. İlgilendiğin popülasyonun temsilcisi olan bir örnek çekmeyi hedeflemelisin.

**İstatistiksel analiz için örnekleme**

Bir örnek seçerken, başlıca iki yaklaşım vardır.

- **Olasılıklı örnekleme:** rastgele seçim vasıtasıyla, popülasyonun her bir üyesinin öğrenme-çalışmasına seçilme şansı eşittir.
- **Olasıksız örnekleme:** Mesela elverişlilik veya gönüllü-seçilme gibi kriterlerden dolayı popülasyonun bazı üyelerinin öğrenme-çalışmasına seçilmesi şansı diğerlerinden daha fazladır.

Teorik olarak, yüksekçe genelleştirilebilir bulgular için, bir olasılıklı örnekleme metodu kullanmalısın. Rastgele seçim, mesela *örnekleme yanlılığı* gibi çeşitli tipteki *araştırma yanlılıklarını* azaltır ve örneğindeki verinin taşıdığı enformasyonu hesaplanabilir bir olasılıkla popülasyona genelleştirebilmeni sağlar. Yalnızca olasılıklı örnekleme kullanarak veriyi topladığın zaman, güçlü istatistiksel çıkarımlar yapmak için *parametrik testleri* kullanabilirsin.

Fakat pratikte, ideal örneği toplamak nadiren mümkündür. Olasıksız örnekler, mesela gönüllü-seçilme yanlılığı gibi yanlılık risklerine daha yatkın olmalarına rağmen, böyle örneklere ulaşmak ve onlardan veri toplamak çok daha kolaydır. Olasıksız örnekler için *parametrik-olmayan testler* daha mütenasiptir fakat popülasyon hakkında daha zayıf çıkarımlar yapabilirsin.

Eğer olasıksız örnekler için yine de parametrik testleri kullanmak istersen, şu iki şeyden emin olmak zorundasın:

- Örneğin bulgularını genelleştireceğin popülasyonun temsilcisidir.
- Örneğin sistematik yanlılığa sahip değildir.

Aklından çıkarmamalısın ki bir öğrenme-çalışmasının neticesini yalnızca senin örneğinle aynı karakteristikleri paylaşanlara genelleştirebilirsin. Buna *dışsal geçerlilik* adı verilir. Mesela lise öğrencilerine ait sonuçlar otomatik olarak emeklilerin popülasyonu için de geçerli değildir.

Eğer olasıksız örneklerden gelen veriye parametrik testleri icra edersen, sonuçlarının genelleştirilebilirliğinin sınırlarını araştırma raporunun *tartışma bölümünde* tafsilatıyla anlatmalısın.

Mütenasip bir örnekleme prosedürü yaratmak

Araştırmacı için kullanabileceği kaynaklara bağlı olarak, denekleri nasıl seçeceğine karar vereceksin. Bu esasen senin *örnekleme çerçeveni* tayin edecek.

- Hedeflediğin popülasyona ulaşabilecek, yeterli kaynağa sahip misin?
- Geniş bir popülasyonu temsil edecek çeşitlilikte bir örneği temin edecek vasıtalarla sahip misin?
- Ulaşılması zor grupların üyeleri ile temas edecek ve takip edecek zamana sahip misin?

Misal 9. Örnekleme (deney). İlgilendiğin popülasyon yaşadığın ildeki lise öğrencileri. Muhtelif ilçelerdeki üç özel okul ve yedi devlet okulu ile irtibat kurdun ve 11 inci sınıf öğrencileri ile bir deney yürüteceksin.

Deneklerin okullar tarafından kendiliğinden-seçilmiş olacaklar. Her ne kadar olasılıksız bir örnekleme kullanıyor olsan da, çeşitli ve temsilci bir örnek seçmeye çaba sarfedersin.

Misal 10. Örnekleme (gözlemsel öğrenme-çalışması). İlgilendiğin esas popülasyon Türkiye'deki erkek lise öğrencileri. Sosyal medya ilanlarını kullanarak, geri dönüş yapan öğrencilerin içerisinde örneğini oluşturdu.

Deneklerin yoklama için gönüllü olanlar ki bu da örneğini bir olasılıksız örnek yapar.

Bazı hallerde araştırmacı açıklayıcı değişkeni gönölünce manipüle edemez (yani, kasten değiştiremez). Böyle bir manipülasyon bu tür değişkenlerin doğasına aykırıdır (mesela cinsiyet) veya herhangi bir sebeple deney ünitelerinin muamele gruplarına rastgele ataması gerçekleştirilemez (mesela Misal 9). Böylesi öğrenme-çalışmalarına *yarı-deneysel öğrenme-çalışmaları* denir.

Yarı-deneysel öğrenme-çalışmaları, neden-ve-etki ilişkilerini açığa çıkarmada arızalıdır. Çünkü *içsel geçerlilik* yeterince sağlanamaz. Yani açıklayıcı değişkenin, yanıt değişkeninde yarattığı etkinin miktarı bu türden bir öğrenme-çalışması ile tam belirlenemez. Ancak, koşulların zorlaması ile, yine de bu türden araştırmalara çok sık başvurulur.

Yeterli örnek hacmini hesaplamak

Denekleri seçmeden önce, örnek hacmine karar vermelisin. Bunun için alanında önceden yapılmış öğrenme-çalışmalarına bakabilirsin veya istatistiksel hesaplama yaparsın. Çok dar bir örnek temsilci olmayabilirken, çok geniş bir örneğin lüzumundan fazla maliyeti olabilir.

Pek çok online *örnek hacmi hesaplayıcıları* bulabilirsin. Popülasyonun alt-gruplara sahip olup olmamasına veya öğrenme-çalışmasının ne kadar katı olacağına bağlı olarak, farklı formüller kullanılır. Bir başucu kuralı olarak, alt-grup başına minimum 30 gözlem ünitesi lazımdır.

Örnek hacmi hesaplamak için mevcut hesaplayıcıları kullanırken, şu kilit bileşenleri anlamak ve girmek zorundasın:

- *Anlamlılık düzeyi (α):* Hakikatte doğru bir sıfır hipotezini yanlışlıkla reddetme riski. Bu riski hangi düzeyde alacaksın? Çoğunlukla %5 tercih edilir.
- *İstatistiksel güç:* Öğrenme-çalışmanın, eğer varsa, belli miktardaki bir etkiyi tespit etmesi olasılığı. Çoğunlukla %80 veya üzeri tercih edilir.
- *Beklenen etki miktarı.* Öğrenme-çalışmanın beklenen sonucunun büyüklüğüne dair standartlaştırılmış bir belirtim, öyle ki başka benzer öğrenme-çalışmalarına dayanır.

- *Popülasyon standart sapması*: Popülasyonun değişkenliğinin bir kestirimi, öyle ki ya önceki bir öğrenme-çalışmasına veya kendi yaptığın bir pilot öğrenme-çalışmasına dayanır.

Adım 3. Veriyi betimleyici istatistiklerle özetlemek

Verinin tümünü topladıktan sonra, veriyi teftiş edebilir ve veriyi özetleyen betimleyici istatistikleri hesaplayabilirsin.

Veriyi teftiş etmek

Veriyi teftiş etmenin muhtelif yolları vardır, bazıları şunlardır:

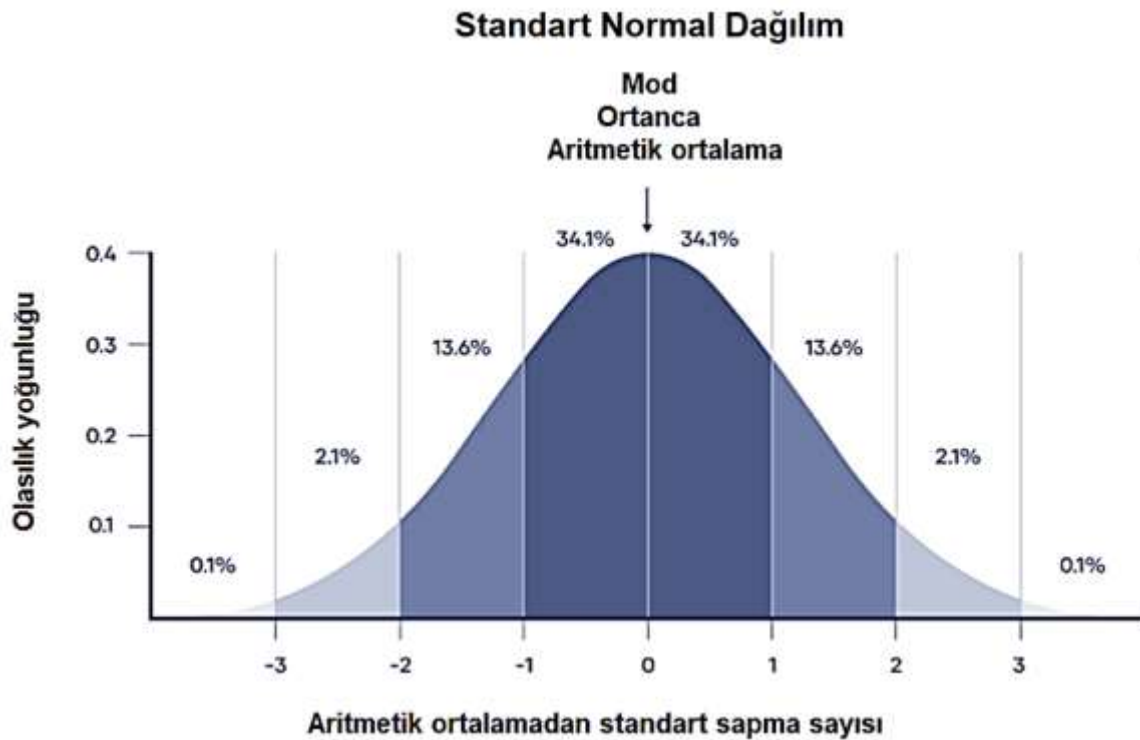
- Her bir değişkene ait veriyi *frekans dağılımı tablolarında* organize edersin.
- Yanıtların dağılımına bakmak için, kilit değişkenlere ait veriyi bir *histogramla* veya *çubuk grafik*le görselleştirsin.
- İki değişken arasındaki ilişkiyi bir *saçılım grafiği* veya bir *çapraz tablo* ile gösterirsin.

Verini tablo ve grafikler vasıtasıyla görselleştirmek suretiyle, verinin dağılımının çarpık mı yoksa normal mi olduğunu veya dışlaklar veya kayıp gözlemler olup olmadığını değerlendirebilirsin.

Bir *normal dağılım*, verinin bir merkez etrafında simetrikçe dağıldığı manasına gelir. Gözlem değerleri çoğunlukla bu merkeze yığılma temayülü gösterir ve uçlara doğru azalarak yayılır.

Aksine, bir çarpık dağılım asimetrik ve değerler bir uca diğerinden daha fazla yığılmıştır. Dağılımın şekline dikkat etmek önemlidir, çünkü yalnızca bazı betimleyici istatistikler çarpık dağılımlar için kullanılabilir.

Aşırı *dışlaklar* da yine yanıltıcı istatistikler ürebileceği için, böylesi gözlem değerlerini ele almak için sistematik bir yaklaşıma ihtiyacın olabilir.



Merkezileşme ölçülerini hesaplamak

Merkezileşme ölçüleri veri kümesindeki gözlem değerlerinin en çok nerede yığıldığını betimler. Merkezileşmenin başlıca üç ölçüsü çoğunlukla rapor edilir:

- Mod. Veri kümesinde en yüksek frekansla gözlenen yanıt veya değerdir.
- Ortanca. Veri kümesindeki gözlem değerleri küçükten büyüğe doğru sıralandığında, tam orta sıraya düşen değerdir.
- Aritmetik ortalama. Veri kümesindeki gözlem değerlerinin toplamının örnek hacmine bölümüdür.

Mamafih, dağılımın şekline ve ölçüm düzeyine bağlı olarak, bu ölçülerden bir veya ikisi senin verin için mütenasip olabilir. Mesela demografik özellikler gibi pek çok nitel değişkenin merkezileşmesini yalnızca mod veya ortanca ile betimleyebilirsin. Nicel değişkenler için, eğer dağılımın şekli normale yakınsa, aritmetik ortalama mütenasip olanıdır, fakat çarpık dağılımlar için ortancayı tercih etmelisin.

Değişkenlik ölçülerini hesaplamak

Değişkenlik ölçüleri sana veri kümesindeki gözlem değerlerinin merkez etrafında ne kadar uzağa yayıldığını söyler. Değişkenliğin başlıca dört ölçüsü çoğunlukla rapor edilir:

- Gidim. Veri kümesindeki en büyük değer ve en küçük değer arasındaki fark.
- Çeyrekliklerarası gidim. Veri kümesinin tam ortadaki yüzde ellisinin gidimi.
- Standart sapma. Veri kümesindeki her bir gözlem değeri ve aritmetik ortalama arasındaki ortalama uzaklık.
- Varyans. Standart sapmanın karesi.

Yine, dağılımın şekli ve ölçüm düzeyi, değişkenlik ölçülerinden hangisinin elindeki veri için daha mütenasip olacağını sana söyler. Standart sapma veya varyans normal dağılımlar için en iyi enformasyonu temin ederken, çeyrekliklerarası gidim çarpık dağılımlar için en iyi ölçüdür.

Misal 11. Betimleyici istatistikler (deney). İldeki 30 öğrenciden ön sınav ve son sınav verisini topladıktan sonra, betimleyici istatistikleri hesapla. Sahip olduğun veri oran ölçeğinde ve dağılımı normal olduğu için, aritmetik ortalamayı, standart sapmayı, varyansı ve gidimi tablolatacaksın.

Tabloyu kullanarak, betimleyici istatistiklerin değerlerini ön sınav ve son sınav için mukayese etmelisin. Mesela, grupların varyansları benzer mi? Aşırı değerler var mı? Eğer varsa, bunların dışlak olup olmadığına karar vermelisin. Eğer dışlak ise, ya veri kümesinden atmalı veya bir istatistiksel test için verinin varsayımları ne kadar karşıladığını denetlemelisin.

	Ön sınav skorları	Son sınav skorları
Aritmetik ortalama	68.44	75.25
Standart sapma	9.43	9.88
Varyans	88.95	97.96
Gidim	36.25	45.12
<i>n</i>	30	

Bu tabloda, görebiliyoruz ki aritmetik ortalama skor hipnoz seansından sonra artmıştır ve skorların iki grubunun varyansları barizce farklı değildir. Sonrasında, hipnoz seansının sınav skorlarında neden olduğu bu artışın istatistikselce anlamlı (yani, popülasyona genelleştirilebilir) olup olmadığını bulmak için bir hipotez testi icra edebiliriz.

Misal 12. Betitleyici istatistikler (gözlemsel öğrenme-çalışması). Altı yüz elli üç öğrenciden veri topladıktan sonra, yıllık aile geliri ve üniversiteye giriş sınavı puanı için betitleyici istatistikleri tablolaştırırın.

	Aile geliri (TL)	ÜGSP
Aritmetik ortalama	62100	312.7
Standart sapma	15000	45.2
Varyans	225000000	20.3
Gidim	370000	136
n	653	

Sonrasında, korelasyon katsayısını hesaplayabilirsin ve bu değişkenler arasındaki ilişkinin anlamlı (yani, popülasyona genelleştirilebilir) olup olmadığını bir istatistiksel test ile belirleyebilirsin.

Adım 4. Hipotezleri test etmek veya çıkarımsal istatistikler ile kestirimler yapmak

Bir örneği betitleyen bir sayıya *istatistik* adı verilir ve bir popülasyonu betitleyen bir sayıya *parametre* adı verilir. Çıkarımsal istatistikleri kullanarak, popülasyon parametreleri hakkında, örnek istatistiklerine dayanan neticelere varabilirsin.

Araştırmacılar çoğu zaman istatistiklerden çıkarım yapmak için başlıca şu iki metodu kullanırlar (eş zamanlı olarak).

- *Kestirim*: Örnek istatistiklerine dayanarak popülasyon parametrelerini hesaplamak.
- *Hipotez testi*: Popülasyon parametreleri hakkında, örnekleri kullanarak, araştırma tahminlerini test etmenin formal bir prosesi.

Kestirim

Popülasyon parametreleri için örnek istatistiklerinden iki tip kestirimde bulunabilirsin:

- *Nokta kestirimi*: Parametrenin değerini tam olarak temsil etmek üzere örnekten hesaplanan bir değer.
- *Aralık kestirimi*: Parametrenin değerini ihtiva etmek üzere örnekten hesaplanan bir aralık.

Eğer hedefin popülasyonun karakteristikleri hakkında örnek veriden *çıkarımda bulunmak* ve rapor etmek ise, makalende hem nokta kestirimlerini hem de aralık kestirimlerini kullanmak en iyisidir.

Eğer temsilci bir örneğin varsa, bir örnek istatistiğini popülasyon parametresinin bir kestirimi olarak telakki edebilirsin. Elbette bu halde örnek hacminin yeterince geniş ve rastgeleliğin yürürlükte olması elzemdir.

Popülasyonun tümünü örneklemeyi süreci, yapacağın kestirime daima hata dahil olacaktır. Öyleyse, nokta kestiriminin değişkenliğini de gösterecek bir aralık kestirimini, bir *güven düzeyi* ile beraber temin etmelisin. Buna *güven aralığı* adı verilir.

Bir güven aralığı örnek istatistiği, örnek istatistiğinin standart sapması (yani standart hata) ve tercih edilen güven düzeyine tekabül eden standart normal dağılıma ait bir z-puanı ile tesis edilir. Hesaplanan güven aralığının manası şudur ki mesela tercih edilen güven düzeyi %95 ise, aynı örneklem prosedürü eğer uzun vadede tekrarlanırsa, hesaplanan güven aralıklarının %95 i popülasyon parametresini ihtiva edecektir.

Hipotez testi

Bir örnek istatistiğinin anlamlılığına dair bir hipotez testi yapabileceğin gibi, örnekte gözlenen değişkenler arasındaki ilişkinin anlamlılığına dair bir hipotez testi de yapabilirsin. Hipotez testi sıfır hipotezinin popülasyonda hakiki olduğu varsayımı ile başlar ve sıfır hipotezinin ret edilebilip ret edilebilmeyeceğini değerlendirmek için istatistiksel testleri kullanacaksın.

İstatistiksel testler, senin örneğinin, sıfır hipotezinin hakiki olması halinde, *örnekleme dağılımının* neresine düştüğünü belirler. Bu testler başlıca iki çıktı verir:

- Bir *test istatistiği* sana senin verinin test edilen sıfır hipotezinden ne kadar farklı olduğunu söyler.
- Bir *P-değeri* sana sıfır hipotezinin hakiki olması halinde senin sonuçlarının elde edilmesi olasılığını söyler.

İstatistiksel testler senin araştırma soruna, araştırma dizaynına, örnekleme metoduna ve veri tiplerine göre çeşitlilik gösterecektir. Mesela, *karşılaştırma testleri* çıkımlardaki grup farklarını değerlendirirken, *regresyon testleri* ve *ki-kare testleri* değişkenler arasındaki ilişkileri değerlendirir.

İstatistiksel testler, öte taraftan, parametrik testler ve parametrik-olmayan testler olarak iki gruba ayrılır. Parametrik testler verinin çekildiği popülasyon hakkında güçlü varsayımlara sahiptir. Parametrik testlerin müşterek varsayımları şunlardır:

- Veri normal dağılır,
- Karşılaştırılacak grupların varyansları eşittir.
- Gözlemler bağımsızdır.

Parametrik testleri kullanabilmen için verinin bu varsayımları karşılaması ve değişkenlerinin uygun tipte olması gerekir. Eğer senin verin bu varsayımları karşılamıyorsa, mütenasip veri transformasyonlarını kullanabilirsin veya *parametrik-olmayan* testleri kullanabilirsin. Ancak, parametrik-olmayan testlerin varsayımsal gereksinimleri daha az olmakla beraber, çıkarımları daha zayıftır.

Parametrik testler

Bir *regresyon modeli* bir tahminci değişkende meydana gelen değişikliğe mukabil bir çıkım değişkeninde meydana gelen değişikliği modeller.

- *Basit doğrusal regresyon* bir tahminci değişken ve bir çıkım değişkeni ihtiva eder.
- *Çoklu doğrusal regresyon* iki veya daha çok sayıda tahminci değişken ve bir çıkım değişkeni ihtiva eder.

Karşılaştırma testleri çoğunlukla grup aritmetik ortalamalarını karşılaştırır. Bunlar iki bağımsız örneğin aritmetik ortalamaları (mesela, bir muamele grubu ve bir kontrol grubu), bir örneğin

farklı zamanlardaki aritmetik ortalamaları (mesela, ön-sınav ve son-sınav skorları) veya bir örnek aritmetik ortalaması ve bir popülasyon aritmetik ortalaması olabılır.

- *t-testi*: örnek dar (30 veya daha az) ise, 1 veya 2 grup aritmetik ortalaması için.
- *z-testi*: örnek geniş ise, 1 veya 2 grup aritmetik ortalaması ve her halükarda 1 veya 2 grup orantısı için.
- *ANOVA*: 3 veya daha çok sayıda aritmetik ortalaması için.

Eğer yalnızca bir örneğin var ve örnek aritmetik ortalamasını bir popülasyon aritmetik ortalaması ile karşılaştırmak istiyorsan, *bir-örnek testini* kullan.

Eğer aynı gruptan farklı zamanlarda alınan eşli ölçümlerin aritmetik ortalamalarını karşılaştırmak istiyorsan (denekler-içi dizayn), *bağımlı (eşli) örnekler testini* kullan.

Eğer birbirinden tamamen ayrı iki gruptan alınan ölçümlerin aritmetik ortalamalarını karşılaştırmak istiyorsan (denekler-arası dizayn), *bağımsız (eşsiz) örnekler testini* kullan.

Eğer örneğinin sıfır hipotezinden sapmasını yalnızca bir yönde bekliyorsan, *bir-kuyruklu test* kullan. Eğer sıfır hipotezinden sapmaya dair herhangi bir beklentin yoksa, *iki-kuyruklu test* kullan.

Tek parametrik korelasyon testi Pearson *r* dir. Korelasyon katsayısı *r* sana iki nicel değişken arasındaki çağrışımın yönünü ve gücünü söyler. Mamafih, örnekten hesapladığın korelasyonun popülasyonda da mevcut olup olmadığını teyit etmek için bir hipotez testi yapmaya ihtiyacın var. Bu test korelasyon katsayısının popülasyonda sıfırdan farkı için örnek hacmini kullanarak bir *P*-değeri hesaplar.

Misal 13. Deneyisel araştırma için eşli *t* testi. Senin araştırman bir denekler-içi deney olduğu için, hem ön-sınav hem de son-sınav skorları aynı gruptan geliyor. Bundan ötürü, bir bağımlı (eşli) *t*-testi yapman gerekiyor. Belirli bir yönde bir değişiklik (sınav puanlarında bir artış) tahmin ettiğin için, bir-kuyruklu bir teste ihtiyacın var.

Hipnoz seansının matematik sınavı puanlarını anlamlıca artırıp artırmadığını değerlendirmek için, bağımlı-örnekler bir-kuyruklu *t*-testi kullanıyorsun. Bu test sana şu sonucu verir:

- *t*-değeri = 3.
- *P*-değeri = 0.0028.

Misal 14. Korelasyon katsayısı ve anlamlılık testi. Örneğinde aile geliri ve üniversiteye giriş sınavı puanı arasındaki doğrusal ilişkinin gücünü hesaplamak için Pearson *r* kullanıyorsun. Pearson *r* nin değeri 0.12 ve örnekte küçük bir korelasyonun varlığını gösteriyor.

Pearson *r* bir test istatistiği olmakla beraber, popülasyonda korelasyonun ne kadar anlamlı olduğuna dair sana herhangi bir şey söylemez. Örnekten hesapladığın bu korelasyon katsayısının popülasyonda mevcut olan bir korelasyona delalet edecek kadar büyük olup olmadığını test etmeye ihtiyacın var.

Bir *t*-testi örnek hacmini baz alarak korelasyon katsayısının sıfırdan ne kadar anlamlıca farklı olduğunu belirleyebilir. Aile geliri ve üniversiteye giriş sınavı puanı arasında pozitif bir korelasyon beklediğin için, bir-örnek bir-kuyruklu *t*-testi kullanıyorsun. Bu test sana şu sonucu verir:

- *t*-değeri = 3.08.
- *P*-değeri = 0.001.

Adım 5. Sonuçların interpretasyonunu yapmak

İstatistiksel anlamlılık

Hipotez testinde, *istatistiksel anlamlılık* neticelerin teşkilinde ana kriterdir. Sonuçlarının istatistiksel olarak anlamlı olup olmadığına karar vermek için, veriden hesaplanan *P*-değerini tercih ettiğin bir anlamlılık düzeyi (çoğunlukla 0.05) ile karşılaştırırsın.

İstatistikselce anlamlı sonuçların salt şans eseri ortaya çıkmasının muhtemel olmadığı mülahaza edilir. Eğer popülasyonda sıfır hipotezi hakiki ise, böyle bir sonucun meydana gelmesinin yalnızca çok küçük bir şansı vardır.

Misal 15. Sonuçların interpretasyonu (deney). *P*-değerini 0.0028 hesaplamıştın. Bunu anlamlılık eşiğin 0.05 ile karşılaştırırsan, senin *P*-değerin daha küçük. Buna binaen, sıfır hipotezini reddetmeye karar veriyorsun ve sonuçlarının istatistiksel olarak anlamlı olduğunu addediyorsun. Bu şu manaya gelir ki rastgele faktörlerin değil, doğrudan hipnoz müdahalesinin sınav skorlarında artışa neden olduğuna inanıyorsun.

Misal 16. Sonuçların interpretasyonu (gözlemsel öğrenme-çalışması). *P*-değerini 0.001 hesaplamıştın. Bunu anlamlılık eşiğin 0.05 ile karşılaştırırsan, senin *P*-değerin daha küçük. Buna binaen, sıfır hipotezini reddedebilirsin. Bu şu manaya gelir ki erkek lise öğrencilerinin aile geliri ve üniversiteye giriş sınavı puanları arasında istatistikselce anlamlı bir korelasyon var.

Unutma ki korelasyon her zaman nedenlilik demek değildir, çünkü çoğu zaman üniversiteye giriş sınavı puanı gibi kompleks bir değişkene katkısı olan pek çok altta-yatan faktör vardır. Bir değişken bir başka değişken ile ilişkili olsa bile, belki de bu her ikisine de tesiri olan bir üçüncü değişken gibi doğrudan-olmayan bağlantılardan dolayıdır.

Yine, çok geniş bir örnek hacmi bir korelasyon katsayısının istatistiksel anlamlılığına kuvvetlice tesir edebilir ve çok küçük bir korelasyon katsayısının anlamlı görünmesini sağlayabilir.

Etki miktarı

İstatistikselce anlamlı bir sonuç bir bulgunun illaki önemli gerçek yaşam uygulamaları veya klinik çıkımları olduğu manasına gelmez.

Aksine, *etki miktarı* senin sonuçlarının pratik anlamlılığını belirtir. Sonuçlarının tam bir resmini çizmek için, çıkarımsal istatistiklerinin yanında etki miktarlarını da rapor etmen önemlidir. Eğer makalenin *APA stiline* yazıyorsan, aynı zamanda etki miktarlarının aralık kestirimlerini de rapor etmelisin.

Bekleneceği üzere, daha büyük bir etkinin örnek veride bulunabilirliği daha küçük bir etkinin bulunabilirliğine göre daha yüksektir ve bu da istatistiksel testin gücüne yansır.

Etki miktarı karşılaştırmaları kolaylaştırmak için standartlaştırılmış terimler ile tanımlanır. Aritmetik-ortalama farklar standart sapmalar cinsinden ifade edilir, öyle ki mesela 0.5 lik bir etki miktarı bize aritmetik-ortalama farkın standart sapmanın yarısı olduğunu belirtir. Korelasyonlar için, etki miktarı değişkenler arasındaki korelasyonun kendisidir çünkü korelasyon katsayısı zaten standartlaştırılmış bir terimdir.

Misal 17. Etki miktarı (deney). Ön sınav ve son sınav skorları arasındaki farkın miktarını bulmak için *Cohen-d* yi hesaplıyorsun.

Cohen-d 0.72 değerini alıyor öyle ki senin hipnoz seansının sınav skorlarını artırdığını söyleyen bulgunun orta-yüksek düzeyde pratik anlamlılığı var.

Misal 18. Etki miktarı (gözlemsel öğrenme-çalışması). Korelasyon katsayısının etki miktarını belirlemek için, hesapladığın Pearson r değerini Cohen etki miktarı kriteri ile karşılaştırmalısın.

Senin değerin 0.1 ve 0.3 arasında olduğu için, aile geliri ve üniversiteye giriş sınavı puanı arasında bir ilişki olduğunu söyleyen bulgun çok küçük bir etkiyi temsil eder ve sınırlı bir pratik anlamlılığı vardır.

Karar hataları

Tip I hata ve *Tip II hata* araştırma neticelerinde yapılan yanlışlardır. *Tip I hata* hakiki bir sıfır hipotezini yanlışlıkla ret etmek demektir. *Tip II hata* hakiki olmayan bir sıfır hipotezini yanlışlıkla ret edememek demektir.

Optimal bir anlamlılık düzeyi seçmek ve yüksek *güç* temin etmek suretiyle, bu hataların riskini minimize etmeyi hedefleyebilirsin. Mamafih, bu iki hata arasında bir takas vardır ve iyi dengelenmeleri lazımdır.