

BÖLÜM 2

EKOLOJİDE VERİ
DEĞERLENDİRME VE
BİYOMETRİ

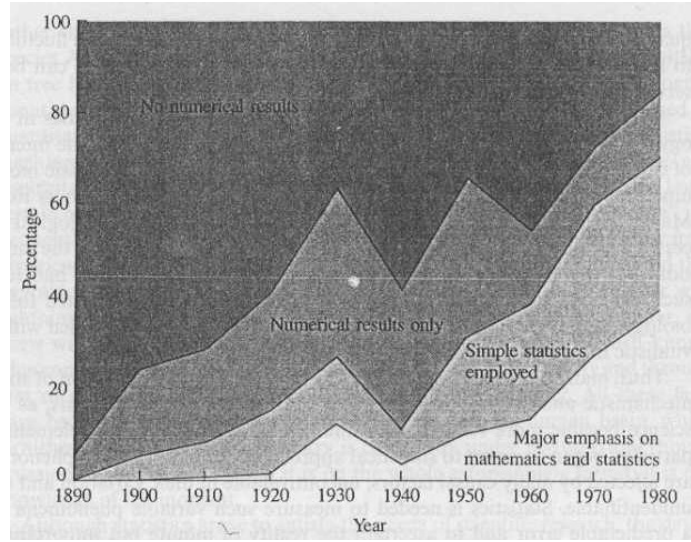
Ekolojik bir çalışmadan elde edilen verilerin değerlendirilmesi, çalışmanın en önemli kısımlarından birini oluşturur. Ekolojik çalışmalarda verilerin değerlendirilmesinde ve sunumundaki nesnelliği sağlayacak olan, çalışmada doğru bilimsel yöntemin ve doğru istatistiğin kullanılmasıdır. **Ham veri** halinde kalmış bir çalışmadan yapılabilecek yorumlar oldukça kısıtlıdır. Ayrıca, ham verileri yorumlamak zordur; yorumlansa bile hatalı yorumlara gitme olasılığı oldukça yüksektir. Bu noktada, verilerin değerlendirilmesi ve özet halinde belirtilebilmesi amacıyla en basitten en karmaşığa kadar birçok istatistiksel yöntemi uygulama olanağı vardır.

BİYOMETRİ VE EKOLOJİK VERİ DEĞERLENDİRMESİNDEKİ ÖNEMİ

Biyometri, geniş bir tanımla, istatistiksel yöntemlerin biyolojik sorunların çözümüne uygulanmasıdır. Biyolojide istatistiğin, yalnızca verilere uygulanıp çalışmamızın sonuçları hakkında nesnel bilgiye sahip olmamızı sağlamak gibi kısıtlı bir kullanımı olduğunu düşünebilirsiniz. Ancak, gerçekte biyometri, biyolojik değişkenliği açıklamamızda bize yol gösteren çok daha geniş faydaları olan bir araçtır. Doğadaki değişkenlik, tüm biyolojik bilimsel çalışmalarda bilim adamlarının araştırdığı asıl konudur. Bu nedenle, biyolojinin hangi alanında çalışılırsa çalışılsın, biyometri temeline dayanmayan çalışmaların günümüzde değeri giderek azalmaktadır (**Şekil 3.1**).

Biyolojide, çoğu olgu değişkenlikleri kontrol edilemeyen ve çoğunlukla da tanımlanamayan birçok faktör tarafından etkilenir. İstatistik, bu değişken olguları öngörülebilir bir hata payıyla ölçmek ve küçük ama önemli farklılıkların gerçeklik derecesini belirleyebilmek için gereklidir.

Biyometri bilgisinin, ekolojideki kullanımının getirdiği faydaların en önemlilerinden biri de, ekolojik çalışmaların planlanması aşamasında, bizi daha sonra gereksiz para ve zaman kaybına götüren yolları tıkamasıdır. Gereğinden fazla veri toplamak, çoğu deneyimsiz araştırmacının başına gelen bir durumdur. Biyometri bilgisi, daha çalışmanın başında deney tasarımında araştırmacıya büyük bir ufuk açarak, verilerin toplanma tarzından, gerektiği kadar veri toplanmasına kadar yol gösterici bir araç olabilir.



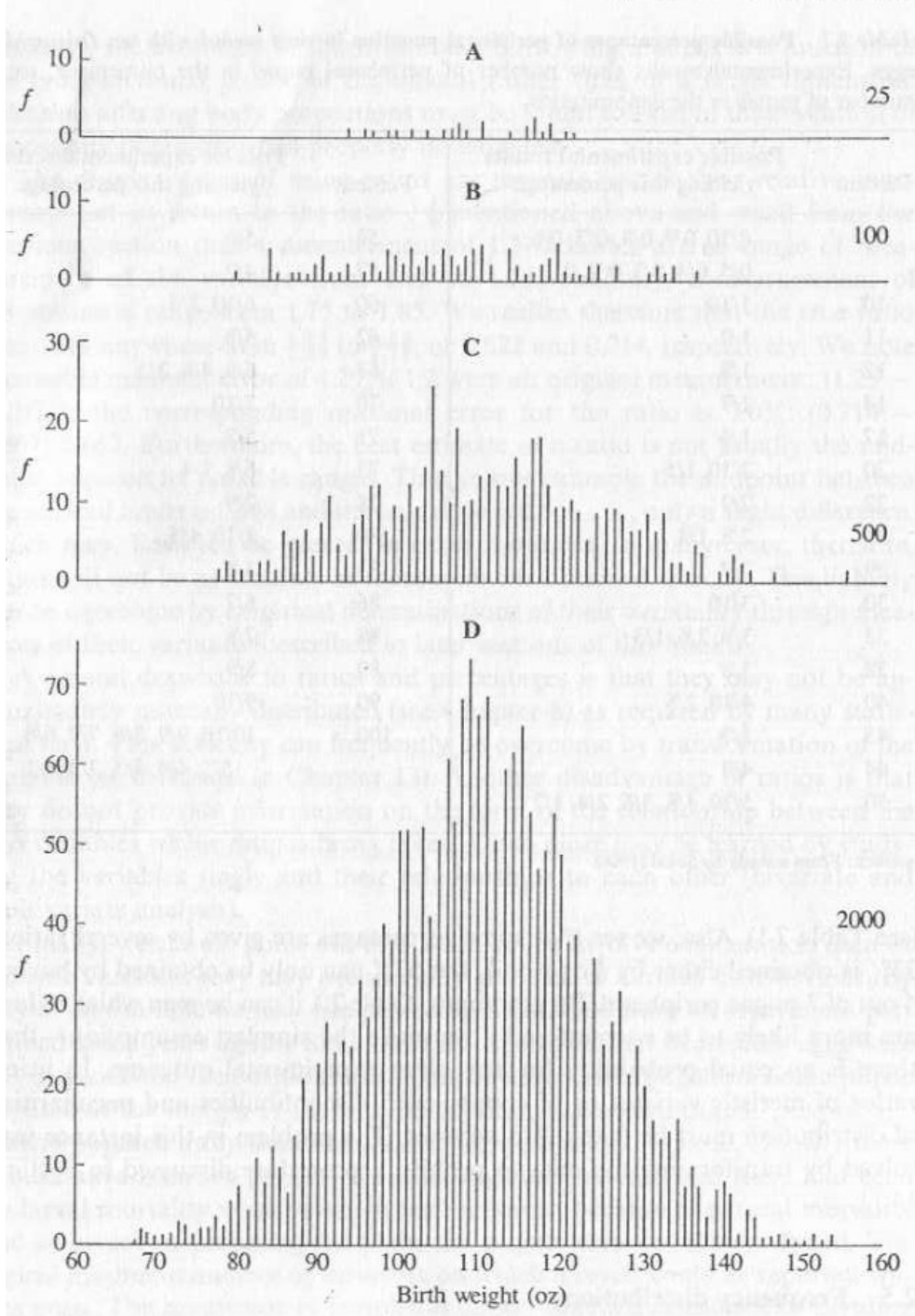
Şekil 3.1 Bilimsel bir dergi olan *The American Naturalist*'de istatistik kullanılan yayınların yıllar içindeki artışı (renkler koyudan açığa doğru sırasıyla; "sayısal sonuç olmayan", "yalnızca sayısal sonuçlar", "basit istatistik uygulanmış", "yüksek düzeyde matematik ve istatistikle desteklenmiş" yayınları belirtmektedir).

FREKANS DAĞILIMLARI

Bir populasyonda yeni doğan bireylerin vücut ağırlıklarını ölçmekle görevli olduğumuzu düşünelim. Her bir ölçümü, doğum ağırlık değerlerini gösteren bir eksen üzerinde, bir nokta ile gösterirebiliriz. Bu örnek Şekil 3.2A'de 25 tane doğum ağırlığı içeren bir örneklem için gösterilmiştir. Eğer bu populasyondan örnek almaya devam edip 100 tane doğum ağırlığı ölçümüne ulaşırsak elde ettiğimiz noktalardan bir kaçını muhtemelen bir birinin üstüne koymak zorunda kalacağız (Şekil 3.2B). örnek sayımızı 100'lere hatta 1000'lere çıkarttığımızda noktaların eksen üzerinde dizilimleri boyut olarak büyümeye devam edecektir (Şekil 3.2C ve D) ve hepimizin çok sıkça rastladığı bir şekle bürünecektir. Bu noktalar kümesinin ana hatlarını izleyen eğri ele aldığımız değişkenin dağılımını gösterecektir.

Bir değişkenin dağılımı biyolojik olarak oldukça önemlidir. asimetrik ve bir yöne doğru eğimli olan bir dağılımın, bizlere belki de dağılımın bir ucunda organizmaya karşı bir seçilim olduğunu belirtebilir. Bir başka örnek vermek gerekirse, ergin evreye ulaşmamış böcek örneklerinde, ölçümler sonucunda iki tepeli (bi-model) bir dağılım ortaya çıktığını varsayalım. Bu sonuç incelenen populasyonun dimorfik olduğu anlamına gelir (bir başka deyişle elimizde iki farklı ergin öncesi döneme ait bireyler

bulunmaktadır). Frekans dağılımları bize, çalıştığımız değişkenin yapısı hakkında bilgi verir fakat bu bilgiyi en doğru ve objektif şekilde ortaya koyabilmek için frekans dağılımı içerisinde saklı bulunan bilginin düzenlenmesi gerekmektedir.



Şekil 3.2 Farklı örneklem büyüklüklerinde çocuklarda doğum ağırlığı verisinin dağılımı.

TANIMLAYICI İSTATİSTİK

Tüm diğer bilimlerde olduğu gibi biyometride de ilk ve temel aşama tanımlayıcı aşamadır. **Tanımlayıcı istatistik**, bir örneklemden elde edilen ölçümlerin düzenlenmesi, özetlenmesi ve tanımlanmasında kullanılır.

Elimizdeki veriyi, bilimsel bir sunum ya da yayın haline dönüştürebilmemiz için gerekli olan şey, veriyi anlaşılabilir bir şekle dönüşecek şekilde özetlememizdir. Bunun da en basit yolu, verinin bir **tablo**da ya da bir **grafikte** daha net bir şekilde verilmesidir. Yüksek düzeyde istatistik uygulanmış bir veri setinin özetini bile bir grafikte göstermek, görselliği de artıracağından, ekolojik çalışmadan elde edilen sonuçların daha anlaşılır olabilmesini de sağlamaktadır. Verilerin özetlenmesi sırasında tablo ve grafiklere ek olarak, kullanılacak en basit istatistiksel ifadelerin hepsi **tanımlayıcı istatistik** kapsamına girmektedir.

Temel olarak iki tür Tanımlayıcı istatistik tanımlanabilir: **konum istatistiği** ve **dağılım istatistiği**. Konum istatistiği bir örneklemin, bir değişkeni temsil eden bir eksen üzerindeki konumunu belirler. Örneğin bir canlının boyu ile ilgili yaptığımız ölçümlerin 2cm civarında mı yoksa 20 cm civarında mı olduğunu merak edebiliriz. Buradaki en önemli nokta elde edilen konum istatistiğinin elimizdeki örnekleme doğru bir şekilde yansıtıyor olmasıdır. Fakat konum istatistiği frekans dağılımının şekli hakkında bize her hangi bir bilgi vermemektedir. bu veriye ait dağılım geniş veya dar olabilir, iki tepeli olabilir veya tamamen asimetrik olabilir. İşte dağılım istatistiği frekans dağılımının sahip olduğu bu gibi özellikleri ortaya koymak için kullanılır.

Konum İstatistikleri

Aritmetik Ortalama: Herkesin bildiği, en yaygın konum istatistiğidir. Ortalama, bir örneklemden tüm bireysel gözlemlerin toplanarak gözlem sayısına bölünmesiyle elde edilir.

$$\bar{Y} = \frac{\sum Y}{n}, \text{ burada } \bar{Y} : \text{aritmetik ortalama, } Y : \text{bireysel gözlemler, } n : \text{bireysel}$$

gözlem sayısıdır. Ortalama, aşırı değerlere duyarlı bir ölçüttür dolayısıyla çalışmalarda genel bir kural olarak verinin alt ve üst ucunda bulunan aykırı değerler çalışmadan çıkartılır.

Ortanca: Ortanca ekolojik araştırmalarda daha nadir olarak kullanılan bir konum istatistiğidir. Büyüklüklerine göre sıraya dizilmiş bireysel gözlemlerin orta noktasındaki gözlem değeri olan ortanca, aşırı değerlerden etkilenmez, bu nedenle

bazen ortalamanın yerine kullanılması daha uygundur. Örneğin, bir örneklemden bireysel gözlemler; 4, 8, 9, 12, 15, 25, 202 ise; bu durumda ortanca değeri 12 olacaktır, ancak bu verinin ortalaması hesaplanmış olsaydı ortaya çıkan sonuç gerçek değerden oldukça üstünde ve muhtemelen de yanlış bir değer olacaktır.

Dağılım İstatistikleri

Dağılım Aralığı: Bir örneklemden değişkenliğin en basit ölçütü dağılım aralığıdır dolayısıyla oldukça önemli bir dağılım istatistiğidir. Dağılım aralığı, bir dağılımdaki en yüksek ve en düşük gözlemleri gösterir; ayrıca, dağılımın yaygınlığı hakkında da bilgi verir.

Standart Sapma: Bir örneklemden değişkenliğin en yaygın olarak kullanılan ölçütüdür. Dağılım aralığının aksine, standart sapma, bir değişkenle ilgili bireysel gözlemlerin hepsinin doğrudan dikkate alınmasıyla hesaplanır.

$$s = \sqrt{\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1}}, \text{ burada } s: \text{ standart sapma, } (y_i - \bar{y}): \text{ ortalamadan sapma, } (n-$$

1): serbestlik derecesidir.

Serbestlik Derecesi: Standart sapma ve diğer birçok istatistiğin hesabında öne çıkan bir değerdir. Farz edelim ki $n=5$ bireysel gözlemden oluşan bir örneklemin ortalaması (\bar{y}) 50 olsun. Bu noktada, bizden bireysel gözlemleri uydurmamız istensin. Biliyoruz ki $\sum y = (\bar{y} \times n) = 250$. Dolayısıyla, eğer gözlemlerin toplamı 250 ise, sadece ilk dört gözlemi serbestçe belirleme şansımız vardır, çünkü beşinci gözlem mutlaka toplamı 250'ye tamamlayacak bir sayı (belki de negatif bir değer) olmalıdır. Serbestlik derecesi, bu örneğimizde gözlem sayısının bir eksigidir. Yani, $sd = (n - 1)$.

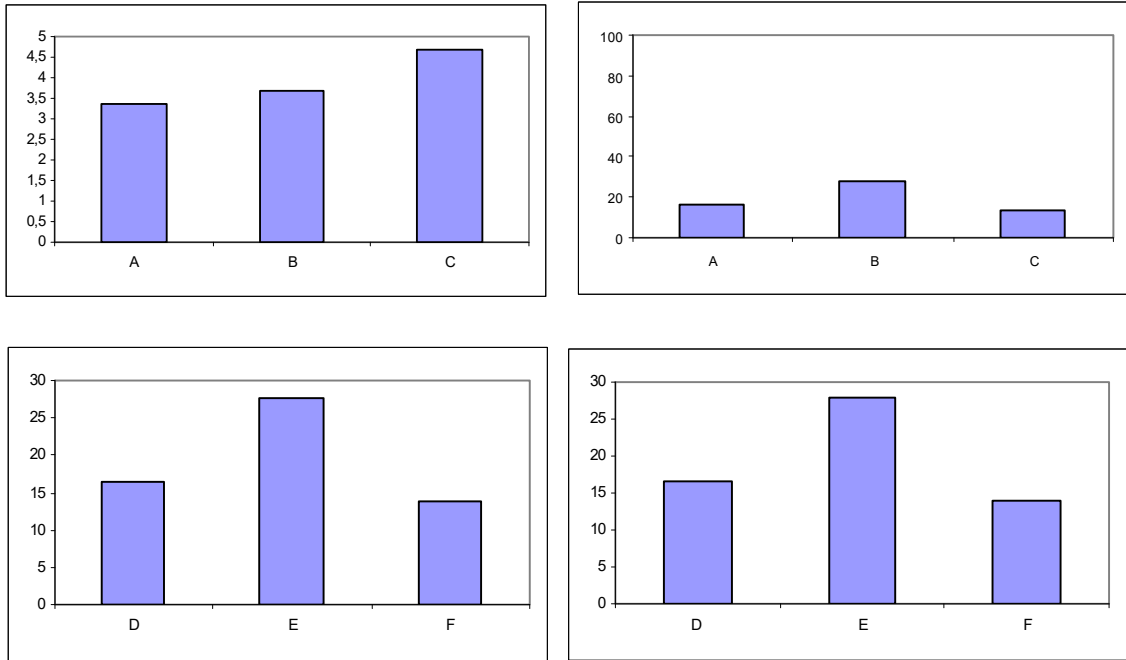
Varyans: Populasyondaki değişkenliğin standart sapmayla yakından ilişkili bir ölçütüdür. Varyans, standart sapmanın karesidir ve bazı parametrik istatistiksel tekniklerde kullanılır.

$$s^2 = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1}$$

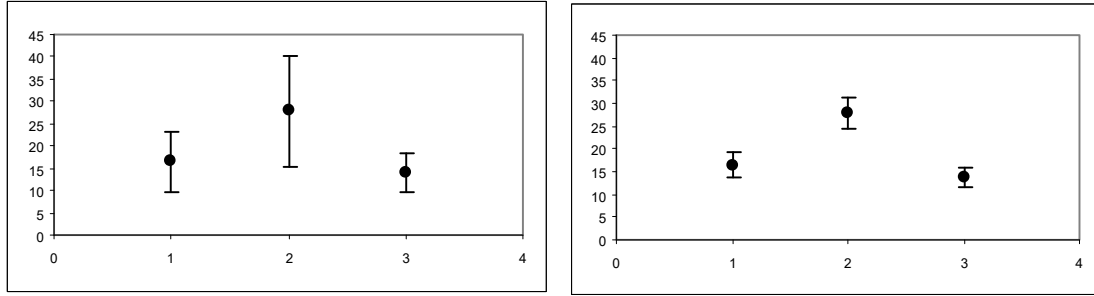
Bir dağılım frekansına (veri kümesine) ait konumsal istatistik ile dağılım istatistik bilgilerini sunmanın en kolay ve etkin yollarından biri bu değerleri grafikler ve tablolar şeklinde düzenlemektir. Grafikler ve tablolar görsellik açısından faydalı

olsalar ya da sonucun daha kolay anlaşılabilmesini sağlasalar bile, istatistiksel olarak bir değerlendirme içermiyorlarsa bir anlam ifade etmezler. Örneğin Şekil 3.2 ve Şekil 3.3'e bir bakalım. Şekil 3.2 ne olarak gözükmüşse gözüksün, aslında iki grafikteki değerler de aynıdır; yalnızca ölçek farklılığı vardır. Şekil 3.3'de ise ikinci grafikteki gruplar arasında daha büyük bir fark varmış gibi görünmekte. Açıkçası, görsel olarak buna karar vermek ne kadar öznelse, istatistiksel süreçlerle buna karar vermek de o kadar nesneldir. Ortalama değerler içeren bir grafik hakkında yorum yapabilmek için, *en azından* ortalama değerlerin değişkenliği hakkında da bilgi sahibi olunmalıdır. Örneklemin *standart sapmasını* grafikte göstermek bunun bir yolu olabilir (Şekil 3.4). Şekil 3.4'ü incelediğimizde iki grafiğin (değerler; ortalama \pm standart sapma) hangisinde gruplar arası fark daha büyük görünmektedir? Grupların sahip olduğu ortalama ve standart sapma gibi değerler arasındaki farkların anlamlı olup olmadığını hangi kritere göre karar verebilirsiniz?

Şekil 3.2 Yukarıdaki iki grafiğin (değerler; ortalama) hangisinde gruplar arası fark daha büyük görünmektedir?



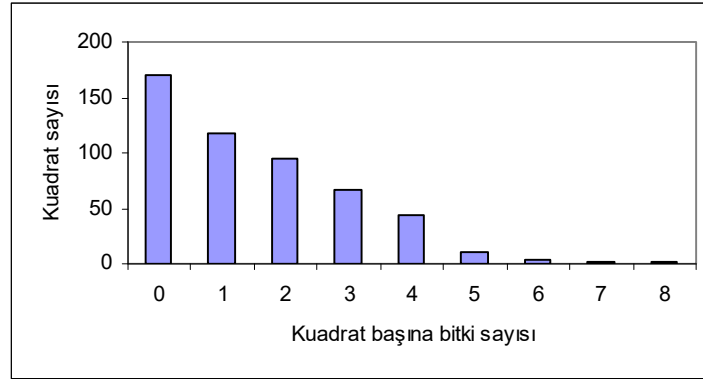
Şekil 3.3 Yukarıdaki iki grafiğin (değerler; ortalama) hangisinde gruplar arası fark daha büyük görünmektedir? Ne kadar eminsiniz?



Şekil 3.4 Yukarıdaki iki grafiğin (değerler; ortalama \pm standart sapma) hangisinde gruplar arası fark daha büyük görünmektedir? Ne kadar eminsiniz?

OLASILIK DAĞILIMLARI

Frekans dağılımları bize, çalıştığımız değişkenin yapısı hakkında bilgi verir ve bu değişken hakkında tanımlayıcı istatistikten yola çıkarak fonksiyonel hipotezler kurabilmemizi ya da bazı öngörülerde bulunabilmemizi sağlar. Bir frekans dağılımı örneği **Şekil 3.5**'te verilmiştir.



Şekil 3.5 Bir örneklemden elde edilen frekans dağılımına bir örnek

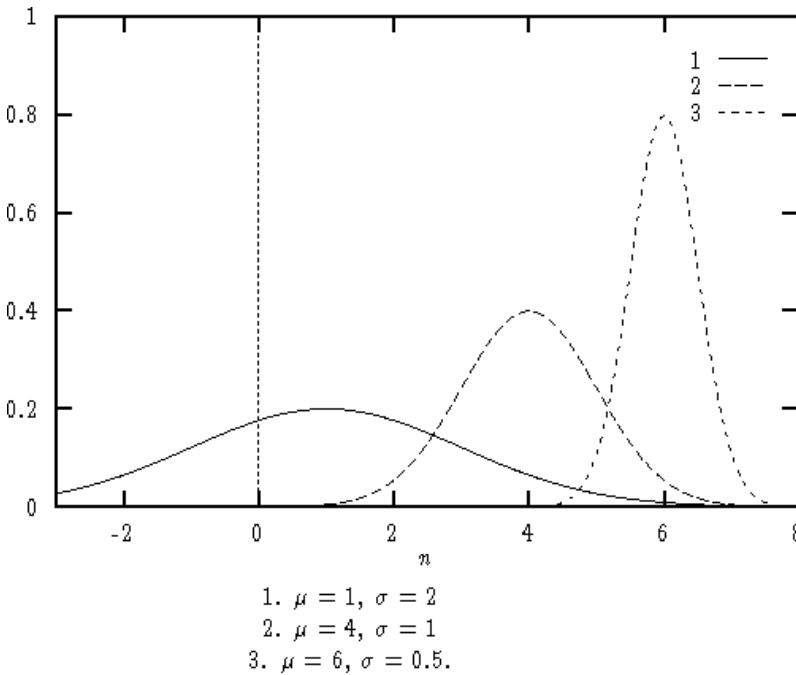
Ancak, çoğu durumda, bu şekilde öngörülerin deneysel dağılımlara değil de, bizim yargılarımıza daha çok uyan kuramsal düşünceye dayanmasını isteriz. Elimizdeki örnek üzerinde etkili olan güçlerin doğası hakkındaki temel tahminlerimizden yola çıkarak verinin belli bir şekilde dağılması gerektiğini hissederiz. Eğer gözlenen veriler bu tahminler ışığında beklenen değerlere yeterli bir şekilde uymuyorsa, tahminlerimiz üzerinde ciddi bir şüpheye düşeriz. Bu, frekans dağılımlarının biyolojideki yaygın bir kullanım tarzıdır. Sınanan tahminler genellikle bizi teorik bir frekans dağılımına götürür; bu aynı zamanda **olasılık dağılımı** olarak da bilinir. Bir olasılık dağılımı basit bir iki değerli dağılım olabileceği gibi (örn. bir Mendel çaprazında 3:1 oranı), bir

kuadrattaki bitkilerin sayısını öngörmeye çalışan daha karmaşık bir fonksiyon da olabilir.

Olasılık dağılımlarının arasında, binomial frekans dağılımı, Poisson dağılımı, normal dağılım, ki-kare (χ^2) dağılımı, t dağılımı sayılabilir.

Normal Dağılım

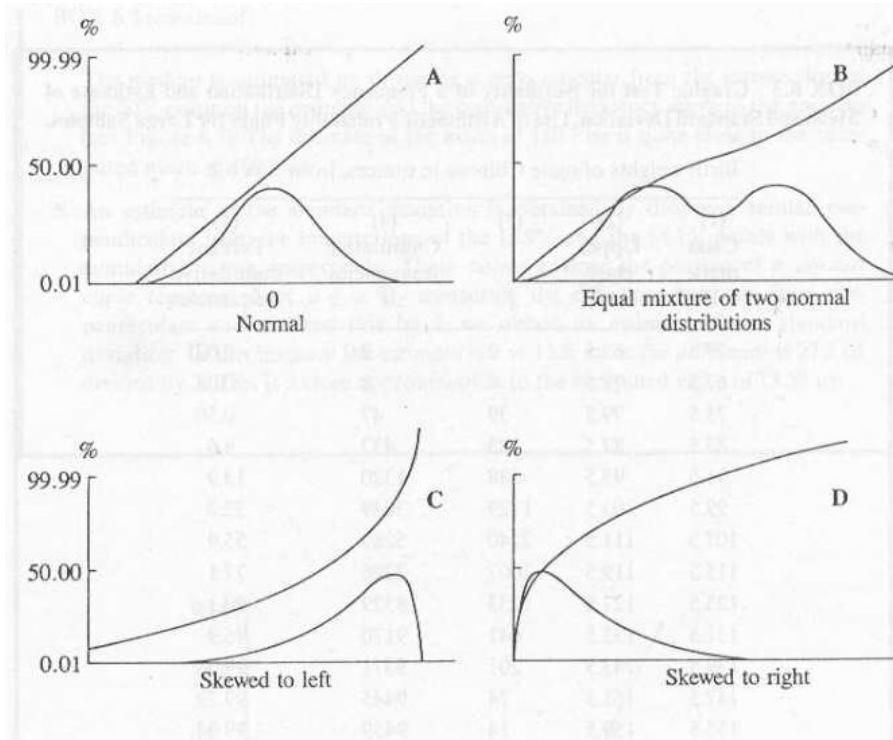
Normal dağılım, istatistikteki en önemli olasılık dağılımıdır. Biyolojide karşılaşılabilecek verilerin çoğu "sürekli sayısal değişken" veriler olacaktır. Sürekli değişken bir veri iki sabit nokta arasında sonsuz tane değere sahiptir. Örneğin iki ölçüm değeri olan 1.5 cm ile 1.6 cm arasında sonsuz sayıda daha ölçüm yapılabilir. Dolayısıyla sürekli sayısal değişken bir karakteri ölçtüğümüzde elde ettiğimiz sonuç (örneğin 1.5 cm) gerçek sonuca olan en yakın tahminimizdir. Çünkü ölçüm aletimizin hassaslığını arttırarak sonucu 1.57 cm veya 1.573 cm'de bulabilirdik. Sürekli sayısal değişkenler için teorik olasılık dağılımı, sürekli bir eğri ile ifade edilir ve bu eğri "Normal Dağılım Eğrisi" olarak adlandırılmıştır. Normal dağılım eğrisinin şekli popülasyonun sahip olduğu ortalama (μ) ve standart sapma (σ) tarafından belirlenir. Dolayısıyla tek bir normal olasılık dağılımı olmayıp her farklı (μ) ve (σ) değerlerine sahip popülasyonlar için farklı normal dağılım eğrileri çizilebilir. (Şekil 3.6). Biyolojideki çoğu değişken (uzunluk, ağırlık, alan, hacimi aç, sıcaklık, nem, yüzde vs) sürekli sayısal değişkendir. Bu nedenle biyolojide de en çok kullanılan dağılım tipi normal dağılımdır.



Şekil 3.6 Normal dağılım eğrisinin farklı ortalama ve standart sapmalar altında aldığı şekil.

Güçlü istatistiksel testlerin çoğunun en önemli önkoşullarından biri, veri setinin normal dağılımasıdır. Bir çalışmadan elde edilen veri normal dağılmıyorsa bile, veriye en güçlü istatistiksel testleri uygulayabilmek için, çeşitli *dönüşüm* yöntemleriyle verinin normal dağılıması sağlanmaya çalışılır. Dönüşüm, "veriyi farklı bir biçimde yeniden ifade etmek" şeklinde tanımlanabilir. Logaritmik dönüşüm ($\log y_i$ yada $\log[y_i + 1]$), karekök dönüşümü ($\sqrt{y_i}$) ve tersine dönüşüm ($1/y_i$), sıklıkla uygulanan dönüşüm tipleridir. Dönüşümler de veriyi normal dağılıma uygun hale getiremiyorsa, gücü nispeten daha düşük testler uygulamak zorunda kalınır.

Şekil 3.7'de normal dağılan ve normal dağılmayan veri setlerinin frekans dağılımı örnekleri ve normal olasılık dağılımına göre ölçeklendirilmiş kümülatif dağılımları gösterilmektedir. Burada, A ile gösterilen grafikte normal bir dağılım söz konusu iken, C ve D grafiklerinde, sırasıyla, soldan basık (negatif çarpık) ve sağdan basık (pozitif çarpık) dağılımlar söz konusudur. B grafiğinde ise iki normal dağılımın eşit bir şekilde karışmasıyla oluşmuş bir dağılım görülmektedir.



Şekil 3.7 Normal dağılan ve normal dağılmayan veri setlerinin frekans grafiklerine örnekler.

Verinin normal dağılımı, örneklem büyüklüğü ile doğrudan ilişkilidir. Örneklem büyüklüğü ne kadar artarsa, veri setinin dağılımı normale o kadar yaklaşır (Merkezi

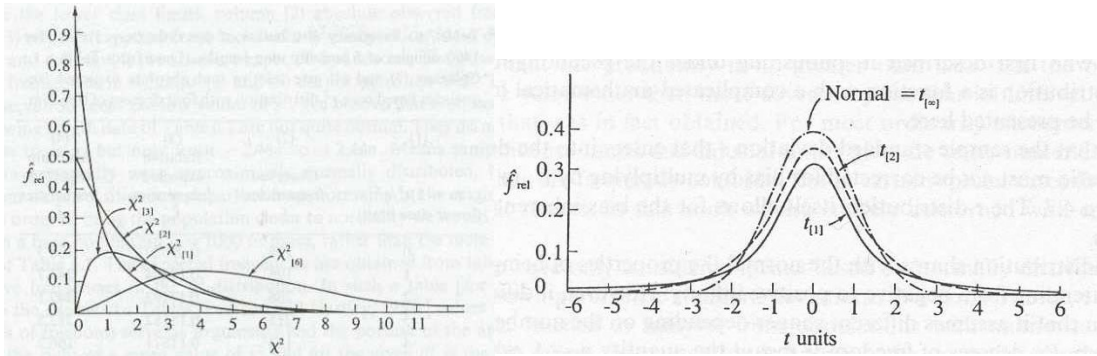
Limit Kuramı). Örneğin, Şekil 3.2’de örneklem büyüklüğünün 25, 100, 500 ve 2000 olduğu durumlarda, çocuklardaki doğum ağırlığı verisinin dağılımının nasıl değiştiği görülmektedir.

***t* -dağılımı**

Normal dağılımın şekli önceden de söylediğimiz üzere populasyonun sahip olduğu ortalama (μ) ve standart sapma tarafından belirlenir (σ). Fakat genelde (μ) ve (σ)’nın gerçek değerlerini bilmemiz imkansızdır ve çoğunlukla bunları elimizdeki örneklem üzerinden hesaplamamız gerekir. (örneğin değişkenimiz doğum ağırlıkları ise (μ) ve (σ) tam olarak hesaplayabilmek için dünya doğmuş olan bütün çocukların doğum ağırlıklarını ölçmemiz gerekir ki bunu pratikte gerçekleştirmek imkansızdır). Eğer örneklemimiz yeterince büyük ise örneklem üzerinden hesapladığımız ortalama ve standart sapmanın populasyonun gerçek (μ) ve (σ) (parametrik ortalama ve standart sapma) değerlerine çok yakın olduğunu varsayabiliriz. Fakat örneklemimiz küçük olduğu durumda ise bundan aynı oranda emin olamayız. Bu belirsizliğin üstesinde gelmek için küçük örneklem büyüklükleri için farklı bir olasılık dağılım eğrisi “*t*-dağılım eğrisi” kullanılır. *t*-dağılımının en büyük özelliği bu dağılımın şeklinin (μ) ve (σ) değerlerinin yanında örneklem büyüklüğü (n) tarafından da belirleniyor olmasıdır. Faka *t* dağılımı doğrudan “ n ” tarafından değil “ $n-1$ ” yani serbestlik derecesi (sd) tarafından belirlenir. Dolayısıyla her farklı “ $n-1$ ” (sd) değerleri için *t*-dağılım eğrisinin şekli farklı olacaktır. $sd=1$ olan bir *t*-dağılımı kural olarak normal dağılımdan en uzak olan dağılım şeklidir. Serbestlik derecesinin değeri arttıkça (örneklem sayımız arttıkça) *t*-dağılım eğrisi normal dağılım eğrisinin şekline gittikçe daha fazla benzemeye başlar. Genellikle $sd=30$ ve üstü değerleri için *t*-dağılım eğrisi ile normal dağılım eğrisini bir birinden ayırmak imkansızdır. $sd= \infty$ için *t*-dağılım eğrisi normal dağılım eğrisini verir. Dolayısıyla *t*-dağılım eğrisini genel durum olarak normal dağılım eğrisini ise *t*-dağılım eğrisinin $sd = \infty$ olduğu zamanki özel durumu olarak ele alabiliriz. Bu nedenle, örneklem büyüklüğünün düşük olduğu ($n < 100$) durumlarda daha kesin sonuçlara ulaşabilmek için *t* dağılımı kullanılır. Şekil 3.8 *t*-dağılım eğrisinin farklı sd değerlerindeki şeklinin normal dağılım eğrisinden olan farklılığını göstermektedir.

Ki-kare dağılımı

Sürekli sayısal değişkenler için biyolojide oldukça sıklıkla kullanılan bir başka dağılım tipi ki-kare (χ^2) dağılımıdır. Ki-kare dağılımı değerleri 0 ile $+\infty$ arasında değişen bir olasılık dağılım eğrisidir. T-dağılım eğrisinde olduğu gibi sadece tek bir χ^2 dağılım eğrisi yoktur fakat her serbestlik derecesi (sd) için farklı bir eğri çizilebilir ve aynen t-dağılımında olduğu gibi sd değeri sonsuza yaklaştıkça χ^2 dağılım eğrisi normal dağılım eğrisine yaklaşır (Şekil 3.8).



Şekil 3.8 Ki-kare ve t dağılımlarının örnek büyüklüğünün artışıyla normal dağılıma yaklaşması.

HİPOTEZ SINAMALARI

Hipotez sınavasının ilk aşaması, genellikle populasyon parametreleri arasında "fark yoktur" ya da "ilişki yoktur" şeklinde kurulan bir istatistiksel boş (null) hipotez (H_0) belirlenmesidir. İkinci aşama ise bu H_0 hipotezinin sınamak için çalışmanın başında seçilen bir test istatistiğinin uygulanmasıdır. Örneğin populasyon ortalamaları ile ilgili hipotezlerin sınavmasında yaygın olarak *t istatistiği* kullanılır.

Uygun istatistiğin seçiminden sonraki aşamada ise, H_0 doğru farz edilerek istatistik değerimizi o istatistiğin dağılımıyla karşılaştırırız. Eğer bu istatistik değerini (ya da daha büyüğünü) veren olasılık değeri, daha önceden belirlenmiş bir anlam (yanılma) düzeyindekinden (örn. 0.05) düşükse, H_0 hipotezinin yanlış olduğu ve onu reddettiğimiz sonucuna varırız ("anlamlı" sonuç).

Eğer, bu değeri veren olasılık belirlenmiş anlam düzeyine (örn. 0.05) eşit ya da büyükse, H_0 hipotezinin yanlış olması için bir kanıt bulunmadığı ve hipotezi reddetmediğimiz sonucuna varırız ("anlamsız" sonuç).

Örneğin ekte verilen *t* dağılımı tablosundan yararlanarak, belli bir anlam düzeyinde ve belli bir serbestlik derecesinde bir *t* değeri elde edilebilir. Eğer

karşılaştırılan iki grup varsa ve karşılaştırma sonucunda elde ettiğimiz t istatistiği değeri, bu tabloda elde ettiğimiz t değerinden (t_{tablo}) daha büyükse (ya da eşitse), biz sınıadığımız boş hipotezi (H_0) o anlam düzeyi için reddederiz. Örneğin, anlam düzeyimizi $\alpha=0.05$ olarak belirlemiş olalım ve örneklem büyüklüğümüz de birinci grup için 7, ikinci grup için 7 örneklem olsun. Bu durumda t testi için serbestlik derecesi $(7-1)+(7-1)= 12$ olmaktadır. Tabloda ise satırda $sd=12$ ve sütunda $\alpha=0.05$ bulunur ve kesiştikleri noktadaki değer bize t_{tablo} değerini verecektir (bu örnek için $t_{tablo} = 2.179$ 'dur). Eğer boş hipotezi t testi ile sınıamamız sonucu bulduğumuz t değeri 2.179'dan büyükse, boş hipotezi reddederiz.

İSTATİSTİKSEL TESTLER

Biyologlar tarafından kullanılan istatistiksel yöntemler parametrik ve parametrik olmayan (non-parametrik) yöntemler olarak iki sınıf içerisinde ele alınabilir.

Parametrik testler: "genel lineer model" olarak bilinen istatistik grubunun bir bölümüdür. Bu testleri uygulabilmeniz için yerine getirmeniz gereken en önemli varsayım veri grubunuz dağılımının normal olmasıdır. Dolayısıyla bu metotları kullanabilmek için örneklem büyüklüğünüzün yeterince yüksek olması ve örnekleri elde ederken rasgelelik kuralına uyulması şarttır.

T-testi: En basit istatistiksel metotlardan biridir. İki verinin ortalaması arasındaki farkın anlamlı olup olmadığını sınar. Örneğin bu testi kullanarak Beytepe Kampüsü merkezindeki günlük sıcaklık ortalaması ile Yeşil Vadinin günlük sıcaklık ortalaması arasında anlamlı bir fark olup olmadığını sınıyabilirsiniz.

Korelasyon analizleri: Bu iki değişkenin bir biri ile ne kadar ilişkili olup olmadığını sınar. Korelasyon analizler sizlere değişkenleriniz arasındaki ilişkiyi belirten bir (r) değeri verir.

Regresyon analizleri: Bir çok durumda ölçtüğünüz bir değişkenin başka bir bağımsız etken tarafından nasıl etkilendiğini bilmek istersiniz. Örneğin suyun derinliği arttıkça içindeki bitkilerin kütlelerinde nasıl bir varyasyon gözlemlenmektedir? Bu test sonucunda elde edeceğimiz r^2 değeri bitki kütlelerinde gördüğünüz varyasyonun yüzde kaçının su derinliği tarafından açıklandığını belirtecektir.

Varyans Analizleri (ANOVA): ekoloji veya biyolojide belkide en çok kullanılan istatistiksel araçtır. Bu analiz elinizde verideki varyasyonun ne kadar olasılıkla şans eseri ne kadar olasılıkla ise ele aldığınız bir bağımsız değişken

tarafından gerçekleştiğini ortaya koyar. Örneğin sulak alanlardaki bitki biyokütlesinde görülen varyasyonun su derinliği, gölün coğrafik konumu veya her iki değişken tarafından belirlenip belirlenmediğini öğrenmek istiyorsunuz. Burada bağımlı değişkeniniz bitki biyokütlesidir (niye bağımlı çünkü bu faktör bir takım etkenler ile örneğin derinlik ile değişmektedir) bağımsız değişkeniniz ise su derinliği ve göllerin coğrafik konumudur (bunlar bağımsız değişkenlerdir çünkü bunların değerleri bitki biyokütlesinden etkilenmez). ANOVA sizlere bağımlı değişkeninizde (biyokütle) görülen varyasyonun göllerin coğrafik konumundaki farklılıklardan mı yoksa su derinliğinden mi veya her ikisinden mi yoksa hiç birinden mi kaynaklandığını belirtir.

Parametrik Olmayan Testler: Yukarıda verilen parametrik istatistiksel testlerin aksine, parametrik olmayan testlerin uygulanabilmesi için verinin belirli bir şekilde dağılmış olması gerekmektedir. Parametrik olmayan testler biyologlar için çok önemlidir çünkü biyolojik veya ekolojik sistemlerden toplanmış bir çok önemli veri normal dağılım modeline uygunluk göstermemektedir. Parametrik olmayan istatistik testler ekologlara, kurdukları hipotezlerin anlamlılığını istatistiksel olarak kabul etme veya redetme imkanını yitirmeden, veri toplamada büyük oranda esneklik sağlamaktadır.

Ki-kare testi: Biyologlar Ki-kare (χ^2) testini genellikle elde ettikleri (**gözlenen**) değerlerle belli bir olasılık dağılımı sonucunda ortaya çıkan (**beklenen**) değerler arasındaki örtüşme miktarını belirlemek için kullanırlar. Ki-kare testi bir çok farklı ve özgül duruma uyarlanabildiği için oldukça önemlidir.

Mann-Whitnet testi: Bu testi parametrik olmayan t-testi olarak ele alabilirsiniz. Bu test iki örneklemin merkezi eğilimleri arasındaki farkın anlamlılığını verilerin dağılım şekliyle bağımsız olarak sınar. Nasıl t-testinin bir çok farklı biçimi varsa Mann-Whitney testinde de bir çok farklı biçimi vardır.

Kruskall-Wallis: Bu testi ANOVA'nın parametrik olmayan biçimi olarak ele alabilirsiniz.

İSTATİSTİKSEL PROGRAM PAKETLERİ

Son yıllarda kişisel bilgisayarların gelişimi ve istatistiksel program paketlerinin (örn. SPSS, Minitab, SYSTAT) ortaya çıkmasıyla, verilere istatistik uygulamak çok kolay hale gelmiştir. Bu durum özellikle bilimsel çalışmaların değerlendirilmesinde büyük kolaylık sağlamakla birlikte, bazı tuzaklar da içermektedir. İstatistik programlarında çoğu kez, programa girilen veriyi istediğimiz şekilde

değerlendirebiliriz. Ancak, eğer genel biyometrik bir altyapımız yoksa ve yapacağımız testler hakkında yeterli bilgiye sahip değilsek, verilerimize çok yanlış testler (ya da doğru testi yanlış bir şekilde) uygulayabiliriz ve bunun sonucunda tüm değerlendirmemizin baştan sona yanlış olması söz konusu olabilir.

Ayrıca, salt istatistiksel yazılımların yanı sıra, ekolojide sıklıkla kullanılan ve ekolojik çalışmalara özgü yazılımlar da vardır. Birçok ekolojik verinin değerlendirilmesinde yararlı olan indeksler ya da denklemleri içeren Krebs'in "Ecological Methodology" ve morfometrik analizlerde kullanışlı bir program olan Morphometrika bunlara örnek olarak verilebilir.

LABORATUAR UYGULAMASI 1

Amaç: Laboratuarda yetiştirdiğimiz yengeç populasyonlarımız ile doğada bulunan yengeç populasyonlarının sıcaklık tercihlerinin bir birinden farklı olup olmadığını öğrenmek istiyoruz. Aşağıda her iki populasyondan örneklenen 30 bireyin sıcaklık tercihleri, C° cinsinden verilmektedir. Sizden istenilen bu iki populasyonun sıcaklık tercihleri arasında istatistiksel olarak anlamlı bir farkın olup olmadığıdır

Laboratuar çalışması 4 aşamada gerçekleşecektir:

- 1- Tanımlayıcı istatistiklerden ortalama ve standart sapmanın bulunması.
- 2- İstatistiksel hipotezlerin kurulması.
- 3- t testinin uygulanması.
- 4- t tablosunun kullanılarak hesapla bulduğumuz t değeriyle tablodan bulunan t değerinin karşılaştırılması.
- 5- Hipotezlerin kabul edileceğine ya da reddedileceğine karar verilerek sonucun ortaya konulması.

Lab	Doğa
24,6	23,5
25,8	25,2
26,1	21,2
22,9	22,7
25,1	26,3
27,3	22,6
24,0	21,4
24,5	23,5
23,9	24,1
26,2	25,6
24,3	23,9
24,6	23,6
23,3	26,3
25,5	23,2
28,1	25,6
24,8	21,3
23,5	22,4
26,3	24,9
25,4	23,8
25,5	24,6
23,9	24,1
27,0	22,6
24,8	22,7
22,9	25,7
25,4	24,3
24,0	23,2
22,9	22,7
24,8	21,4
22,9	21,3
24,0	23,5

Yararlanılan ve Tavsiye Edilen Kaynaklar

Alpar, R., 2001. Spor Bilimlerinde Uygulamalı İstatistik. Nobel Yayın Dağıtım, Ankara.

Fowler, J., Cohen, L., 1990. Practical Statistics for Field Biology. Open University Pres, Buckingham.

Quinn, G.P., Keough, M.J., 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, UK.

Sokal, R.R., Rohlf, F.J., 1981. Biometry, 2nd edition. WH Freeman and Co., San Francisco.

T - testi (İki Ortalama Arasındaki Farkın Anlamlılık Testi)

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} \right] \left(\frac{n_1 + n_2}{n_1 n_2} \right)}}$$

 \bar{y}_1 : Birinci grubun ortalaması \bar{y}_2 : İkinci grubun ortalaması n_1 : Birinci grubun denek sayısı n_2 : İkinci grubun denek sayısı S_1^2 : Birinci grubun varyansı S_2^2 : İkinci grubun varyansı

$$\text{Ortak varyans: } S_0^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

Eğer örneklem büyüklükleri eşitse denklem aşağıdaki şekilde sadeleştirilebilir;

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}}$$

İstatistiksel karar için t_{hesap} istatistiği Ek'te verilen t tablosundan; $n_1 + n_2 - 2$ serbestlik derecesi ve belirlenen yanılma düzeyimizin (α) kesiştiği noktada yer alan t_{tablo} değeri ile karşılaştırılır.

$|t_{\text{hesap}}| \geq t_{\text{tablo}}$ ise "iki ortalama arasında fark yoktur" diye kurulan H_0 hipotezi reddedilir ve gruplar arasında fark olduğu söylenir ($P < \alpha$).

EK: *t* dağılımı kritik değerleri

$\nu \backslash \alpha$	0.9	0.5	0.4	0.2	0.1	0.05	0.02	0.01	0.001	α / ν
1	.158	1.000	1.376	3.078	6.314	12.706	31.821	63.657	636.619	1
2	.142	.816	1.061	1.886	2.920	4.303	6.965	9.925	31.598	2
3	.137	.765	.978	1.638	2.353	3.182	4.541	5.841	12.924	3
4	.134	.741	.941	1.533	2.132	2.776	3.747	4.604	8.610	4
5	.132	.727	.920	1.476	2.015	2.571	3.365	4.032	6.869	5
6	.131	.718	.906	1.440	1.943	2.447	3.143	3.707	5.959	6
7	.130	.711	.896	1.415	1.895	2.365	2.998	3.499	5.408	7
8	.130	.706	.889	1.397	1.860	2.306	2.896	3.355	5.041	8
9	.129	.703	.883	1.383	1.833	2.262	2.821	3.250	4.781	9
10	.129	.700	.879	1.372	1.812	2.228	2.764	3.169	4.587	10
11	.129	.697	.876	1.363	1.796	2.201	2.718	3.106	4.437	11
12	.128	.695	.873	1.356	1.782	2.179	2.681	3.055	4.318	12
13	.128	.694	.870	1.350	1.771	2.160	2.650	3.012	4.221	13
14	.128	.692	.868	1.345	1.761	2.145	2.624	2.977	4.140	14
15	.128	.691	.866	1.341	1.753	2.131	2.602	2.947	4.073	15
16	.128	.690	.865	1.337	1.746	2.120	2.583	2.921	4.015	16
17	.128	.689	.863	1.333	1.740	2.110	2.567	2.898	3.965	17
18	.127	.688	.862	1.330	1.734	2.101	2.552	2.878	3.922	18
19	.127	.688	.861	1.328	1.729	2.093	2.539	2.861	3.883	19
20	.127	.687	.860	1.325	1.725	2.086	2.528	2.845	3.850	20
21	.127	.686	.859	1.323	1.721	2.080	2.518	2.831	3.819	21
22	.127	.686	.858	1.321	1.717	2.074	2.508	2.819	3.792	22
23	.127	.685	.858	1.319	1.714	2.069	2.500	2.807	3.767	23
24	.127	.685	.857	1.318	1.711	2.064	2.492	2.797	3.745	24
25	.127	.684	.856	1.316	1.708	2.060	2.485	2.787	3.725	25
26	.127	.684	.856	1.315	1.706	2.056	2.479	2.779	3.707	26
27	.127	.684	.855	1.314	1.703	2.052	2.473	2.771	3.690	27
28	.127	.683	.855	1.313	1.701	2.048	2.467	2.763	3.674	28
29	.127	.683	.854	1.311	1.699	2.045	2.462	2.756	3.659	29
30	.127	.683	.854	1.310	1.697	2.042	2.457	2.750	3.646	30
40	.126	.681	.851	1.303	1.684	2.021	2.423	2.704	3.551	40
60	.126	.679	.848	1.296	1.671	2.000	2.390	2.660	3.460	60
120	.126	.677	.845	1.289	1.658	1.980	2.358	2.617	3.373	120
∞	.126	.674	.842	1.282	1.645	1.960	2.326	2.576	3.291	∞