

PAPER DETAILS

TITLE: Örneklem dağılımlarının tarihsel gelişimi: Ki-kare, t ve F dağılımları

AUTHORS: Hakan SERİN, Seyit Mehmet TASDELEN, Davut SEYHAN, Tamer ÇAGLAYAN, Mehmet Emin TEKİN

PAGES: 96-109

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/2660008>



doi: 10.33188/vetheder.1177510

Derleme Makalesi / Review Article

Örnekleme dağılımlarının tarihsel gelişimi: Ki-kare, t ve F dağılımları

**Hakan SERİN^{1,a*}, Seyit Mehmet TAŞDELEN^{1,b}, Davut SEYHAN^{1,c}, Tamer ÇAĞLAYAN^{1,d},
Mehmet Emin TEKİN^{1,e}**

¹ Selçuk Üniversitesi, Veteriner Fakültesi, Biyoistatistik Anabilim Dalı, Konya, Türkiye

ORCID: 0000-0002-1290-4547^a; 0000-0002-4168-5304^b; 0000-0001-5577-1097^c; 0000-0002-5165-0877^d; 0000-0002-3449-9984^e

MAKALE BİLGİSİ /

ARTICLE

INFORMATION:

Geliş / Received:

20 Eylül 22

20 September 22

Revizyon/Revised:

27 Aralık 22

27 December 22

Kabul / Accepted:

28 Aralık 22

28 December 22

Anahtar Sözcükler:

İstatistik Dağılımlar

ANOVA

Örnekleme Çalışmaları

Yaşam öyküsü

Olasılık Fonksiyonları

Keywords:

Statistical Distributions

ANOVA

Sampling Studies

Biography

Likelihood Functions

ÖZET:

Örnekleme dağılımlarının temelini oluşturan Ki-kare, t ve F dağılımlarını sırasıyla Karl Pearson, William Sealy Gosset ve Ronald Aylmer Fisher istatistik bilimine kazandırmıştır. Karl Pearson, normal dağılım göstermeyen veriler için o dönemde yaptığı çalışmalar sonucu Ki-kare dağılımını elde etmiştir. Sonraki yıllarda Fisher, Ki-kare dağılımı üzerine yaptığı çalışmalar ile serbestlik derecesinde bir düzeltme yapmıştır. Bu durum Karl Pearson tarafından hoş karşılanmamış ancak çevresindeki bilim insanlarının etkisiyle bu modifikasyonu kabul etmek zorunda kalmıştır. Ayrıca, Fisher z dağılımının normalliği ile ilgili çalışmalar yapmış, bu konuda makale yazmış ve “Biometrika” dergisine göndermiştir. Ancak dönemin ünlü istatistik dergisi olan “Biometrika”ya başkanlık yapan Karl Pearson Fisher ile arasındaki rekabetten dolayı makalesini geri çevirmiştir. William Sealy Gosset o dönemde ünlü bir bira fabrikasında istatistikçi olarak çalışırken en iyi kalite ve verime sahip arpaları tespit etmek için yaptığı çalışmalar sonucu t dağılımını keşfetmiştir. Gosset’e mektup yazan Fisher, t dağılımında standart sapma hesaplamasında bir düzeltme tavsiye etmiştir. Bununla birlikte, Fisher makalelerinde t dağılımına değinerek t dağılımının bilim dünyasına tanıtımına katkıda bulunmuştur. Ronald Aylmer Fisher Rothamsted’te mahsul çalışmaları yaparken mevcut yöntemlerin yetersizliği ile karşılaşmış ve t dağılımından yola çıkarak yaptığı çalışmalar sonucu F dağılımını keşfetmiştir. F dağılımının keşfi ile birlikte ileri düzey veri analizinde yeni bir dönem başlamıştır.

Historical development of sampling distributions: Chi-square, t and F distributions

ABSTRACT:

Karl Pearson, William Sealy Gosset and Ronald Aylmer Fisher introduced the Chi-square, t and F distributions, which form the basis of the sampling distributions, to the science of statistics, respectively. Karl Pearson obtained the Chi-square distribution for the data that did not show normal distribution as a result of his studies at that time. In the following years, Fisher made a correction in the degrees of freedom with his studies on the chi-square distribution. This situation was not welcomed by Karl Pearson, but he had to accept this modification under the influence of the scientists around him. In addition, Fisher conducted studies on the normality of the z distribution, wrote an article on this subject and sent it to the “Biometrika” journal. However, he rejected his article due to the rivalry between him and Karl Pearson Fisher, who was the chairman of the famous statistical journal of the time, “Biometrika”. William Sealy Gosset, while working as a statistician at a famous brewery at that time, discovered the t distribution as a result of his studies to determine the best quality and yield barley. Writing to Gosset, Fisher recommended a correction in the calculation of the standard deviation of the t distribution. However, Fisher contributed to the introduction of the t distribution to the scientific world by referring to the t distribution in his articles. Ronald Aylmer Fisher, while doing crop studies in Rothamsted, encountered the inadequacy of the existing methods and discovered the F distribution as a result of his studies based on the t distribution. With the discovery of the F distribution, a new era has begun in advanced data analysis.

How to cite this article: Serin H, Taşdelen SM, Seyhan D, Çağlayan T, Tekin ME. Örnekleme dağılımlarının tarihsel gelişimi: Ki-kare, t ve F dağılımları. Vet Hekim Der Derg 2023; 94(1): 96-109 DOI: 10.33188/vetheder.1177510

* Sorumlu Yazar e-posta adresi / Corresponding Author e-mail address: 203159001003@lisansustu.selcuk.edu.tr

1. Giriş

İstatistik herhangi bir alanda verilerin toplanması, düzenlenip analiz edilmesi ve sonuçların sunulmasını sağlayan bir bilim dalıdır (59). Günümüzde istatistik, neredeyse tüm alanlarda, bilimsel bir çalışma yapabilmek için kullanılması zaruri bir araçtır. İstatistiğin önemli konularından biri olan örnekleme dağılımları, araştırmacıların temel ve ileri düzey veri analizinde sonuçları değerlendirmede kullandıkları önemli bir kriterdir. Örnekleme dağılımlarının temel yapı taşlarını F, t ve Ki-kare dağılımları oluşturmaktadır. Veri analizinde yaygın olarak kullanılan t testleri, Tek Yönlü Varyans Analizi, Çok Değişkenli Varyans Analizi ve Ki-kare gibi analizlerde test istatistiğini değerlendirmede kullanılan tablolar 19. ve 20. yüzyıllarda istatistikçilerin yaptığı araştırmalar neticesinde ortaya konulan formüller yardımıyla, bir seri hesaplama yapılarak oluşturulmuştur. Bu çalışmada araştırmacıları, örnekleme dağılımlarını keşfeden bilim insanları, dağılımların nasıl ortaya çıktığı, dağılımların birbirleriyle ilişkileri ve teorik yapıları hakkında bilgilendirmek amaçlanmıştır. Konunun daha iyi anlaşılabilmesi için örnekleme dağılımlarını bilime kazandıran bilim insanlarının hayatlarına da kısaca yer verilmiştir. Böylece bilim insanlarının birbirleriyle olan ilişkileri, dağılımların ortaya çıkışındaki sıralama ve diğer dağılımlara öncülük etmeleri gibi olaylar arasındaki bağlantı kurulmaya çalışılmıştır. T dağılımının en kaliteli birayı elde etme çalışmaları sırasında bulunması, Fisher'in Gosset'e düzeltme önerisinde bulunması ve Pearson'ın Fisher'i engelleme çabaları gibi bilimsel rekabetin olduğu dönemde ve ortamda örnekleme dağılımları istatistik bilimine kazandırılmıştır. Örnekleme dağılımlarının doğuşu hakkındaki bilgiler, konu bütünlüğünün bozulmaması için, bilim insanlarının yaş sırasına uygun olarak kapsamlı literatür bilgisi eşliğinde verilmiştir.

2. Karl Pearson'ın Hayatı ve Ki-kare Dağılımı

Karl Pearson, 27 Mart 1857 yılında Kuzey Londra'da doğmuştur. William Pearson ve Fanny Smith'in ikinci oğlu olan Karl Pearson'ın 1879'dan önce isminin yazılışı "Carl Pearson" şeklindeydi. Ancak Heidelberg Üniversitesi Carl'ı "Karl" olarak kaydettirerek isminin yazılışını değiştirmiştir. Pearson ailesinin her iki tarafının soyu da Yorkshire Quakers'dan gelmekteydi (1, 2).

Karl Pearson'ın eğitim hayatı erken başlamıştır. 4 yaşındayken Fransızca dersleri almış, 9 yaşındayken Gower Caddesindeki Londra Üniversitesi Kolejine başlamıştır. Burada 7 yıl eğitime devam ettikten sonra 1875'te Cambridge'deki King's Kolejine burslu olarak girmiştir. 1879'da matematik onur derecesi ile mezun olmuş, ertesi yıl Kuno Fischer ile anlaşarak Almanya'ya gitmiştir. Almanya'da metafizik, felsefe ve hukuk dersleri almış ve hukukçu olmaya karar vermiştir. 1880'de Londra'ya dönmüş ve 1881'de Baro'ya çağrılmış ancak teklifi kabul etmemiştir (1-3). 1880 yılında Pearson, Goethe'nin Genç Werther'in Acıları'ndan esinlenerek The New Werther adlı ilk kitabını yazmıştır. The New Werther, Loki takma adıyla yayınlanan idealizm ve materyalizm üzerine mektup tarzında romantik bir romandır (4). 1881'de Londra'daki King's Kolejine matematik profesörü olarak atanmıştır. 1884'te Londra Üniversitesi Kolejinde Uygulamalı Matematik ve Mekanik başkanlığını üstlenmiş, 1985 yılında "Common Sense of the Exact Sciences"ın editörü olmuştur (5, 12).

1889 yılında Gresham'da anlattığı derslerin de içerisinde yer aldığı "The Grammar of Science" adlı kitabı yayınlamıştır. 1890 yılında Gresham Kolejinde Geometri profesörlüğüne atanmıştır (8). Burada Walter Frank Raphael Weldon ile tanışmış, tanışmasının ardından 1901'de istatistiksel teorileri geliştirmek için birlikte "Biometrika" dergisini kurmuşlardır. Biometrika dergisi günümüzde halen istatistik ve matematik alanlarında yayın yapmakta olup, bu alanlarda metodolojik ve teorik yapılaraya vurgu yapan makalelere ilgi göstermektedir. Weldon, Pearson'ı kalıtım ve öjeni gibi konularla ilgilenen Charles Darwin'in kuzeni Francis Galton ile tanıştırmıştır (5, 7, 10).

Pearson, Weldon ile Napoli'de Ki-kare uyum iyiliği üzerine çalışırken Galton'un korelasyon ve regresyon ile ilgili çalışmalarında Pearson'dan yardım istemesi üzerine 1895 yılında korelasyon ve regresyon katsayıları üzerine bir makale yayımlamak istemiştir. Bu makaleyi yayımladıktan sonra makalenin özetini Weldon'a göstermiştir. Makale'de Galton'un da belirttiği gibi korelasyon katsayısının 0 ile +1 arasında değerler alabileceği yazmaktaydı. Ancak Weldon'un karidesler üzerine bir çalışma yapmasıyla korelasyon katsayısının -1 ve +1 arasında değerler alabileceği

sonucu ortaya çıkmıştır. Ardından Pearson korelasyon katsayısının -1 ve +1 arasında değerler alabileceğini kabul ederek, yayımladığı makalesine eklemeler yaparak tekrar yayımlamıştır (1, 9, 10). Pearson, 1906'da Galton'un yardımıyla kurulan "Galton Ulusal Öjeni Laboratuvarı"nın yöneticisi olmuştur (3, 6).

Galton'un 1911'de ölümü üzerine Pearson, Londra Üniversitesi Kolejinde Galton'un vasiyetinde bıraktığı fonlarla o yıl kurulan ilk Galton öjeni profesörü olmuştur (11). Weldon'ın hem Plymouth hem de Napoli kıyı yengeçleri hakkındaki verileri, Pearson'ın 1900 yılında Ki-kare uyum iyiliği testini tasarlamasına yardımcı olmuştur. Ki-kare uyum iyiliği testi, Pearson'ın modern matematiksel istatistik teorisine yaptığı en önemli katkıdır. Ardından Pearson, tüm değişkenlerin sürekli olmadığını ve aslında birçoğunun ayrık olduğunu fark etmiştir. Bunun sonucunda tetrakorik korelasyon, phi katsayısı, iki seri korelasyon ve Ki-kare istatistiği dahil olmak üzere ayrık değişkenler için bir dizi korelasyonel yöntem geliştirmiştir (1). 1911'den beri sürdürdüğü Londra Üniversitesi Kolejindeki "Francis Galton Ulusal Öjenik Kürsüsü" görevinden 1933'te emekli olmuştur. 27 Nisan 1936 yılında ise hayatını kaybetmiştir (13).

Pearson'ın çalışmaları, matematiksel istatistiklerin uygulanmasından geliştirilmesine kadar tüm aşamaları kapsamıştır. Pearson; biyoloji, antropometri, tıp, epidemiyoloji, psikoloji ve sosyal tarih alanlarında çalışmalar yapmıştır. Pearson'ın istatistik bilimine katkılarının sadece birkaçı aşağıda maddeler halinde verilmiştir.

- Korelasyon katsayısı
- Moment yöntemi
- Pearson'ın sürekli eğriler sistemi
- Chi mesafesi
- P değeri
- İstatistiksel hipotez testi teorisi ve istatistiksel karar teorisinin temelleri
- Pearson'ın Ki-kare testi
- Histogramın ilk tanıtımı (5).

Ki-kare (χ^2) dağılımı

Ki-kare dağılımının kökeni, 19. yüzyılın ortalarında astronomlar tarafından hatalar teorisinden türetilen en küçük kareler yöntemine dayanmaktadır. Pearson'ın, gama dağılımları ailesinden tam Ki-kare dağılımını buluşu modern istatistikte dönüm noktası olmuştur. Pearson'ın eğri uydurma ve asimetric dağılımlar için bir uyum iyiliği testi bulma konusundaki araştırmaları, 19. yüzyılın sonunda normal dağılımın kapsayıcılığına tepki olarak görülmektedir. Pearson, mekanik matematiğini ve moment yöntemini kullanarak, Weldon'ın verilerini yorumlamak için yeni bir istatistiksel yöntem olan χ^2 dağılımını geliştirmiştir. Daha sonra Weldon, Mendel'in Mendel dağılımları içinde analiz ettiği verilerini incelemek için Pearson'ın χ^2 uyum iyiliği testini kullanmış ve χ^2 uyum iyiliği testi genetikçiler tarafından yaygın bir şekilde kullanılmaya başlanmıştır (1, 14, 15).

Pearson, Weldon'ın Napoli'deki bir çalışmasından elde ettiği yengeç verilerinden türetilen asimetric eğrilerini inceledikten sonra, normal eğriye uymayan dağılımlar için uyum iyiliğini ölçmenin bir gereklilik olduğunu fark etmiştir. Pearson'ın uyum iyiliği testinin bir ölçüsünü belirleme konusundaki ilk düşüncesi Gresham konferansında aklına gelmiştir. 1894'ün başında χ^2 uyum iyiliği testini bulmuştur. Bu test ile Pearson, "düzeltme faktörü" olarak adlandırdığı serbestlik derecesi kavramının ortaya çıkmasına yardımcı olmuştur. Daha sonra χ^2 beklenmedik durum tablosu için düzeltmeleri (serbestlik dereceleri) belirlemiştir. Pearson ve öğrencisi (Alice Lee), 1900'de bir χ^2 olasılık tablosu oluşturmuş ve başka bir öğrencisi (William Palin Elderton) bu tabloyu düzenlemiştir (1, 9).

Fisher, 1922-1928 yılları arasında χ^2 istatistiği ile ilgili 5 tane makale yayımlayarak, Pearson'ın χ^2 istatistiğini hesaplamada kullandığı serbestlik derecesinin yanlış olduğunu vurgulayıp farklı bir serbestlik derecesiyle yeni bir χ^2 tablosu oluşturmuştur. Fisher, Pearson'ın χ^2 dağılımı için kullandığı (p x q - 1) serbestlik derecesi yerine bir (p x q) tablosunda (p - 1) x (q - 1) serbestlik dereceleri (satır sayısı - 1 x sütun sayısı - 1) için χ^2 dağılımı elde etmiştir. Yule, Greenwood, Bowley ve diğerleri daha önce testin önceki kullanımının geçerliliği konusunda şüphelerini dile getirirler de Pearson bu mantıklı modifikasyonu kabul etmekte zorlanmıştır. Bu durum Pearson'ın hoşuna

gitmemiş olsa da etrafındaki bilim adamlarının da Fisher'i doğru bulmasıyla bu düzeltmeyi kabul etmek zorunda kalmıştır. Günümüzde halen χ^2 istatistiği Fisher'in bulduğu serbestlik derecesi ile hesaplanmaktadır (16-18).

Aritmetik ortalamada olduğu gibi varyansın da bir dağılımı vardır; fakat bu dağılım normal bir dağılım değil, χ^2 dağılımıdır. Dolayısıyla varyans tahmininde χ^2 dağılımı kullanılır. Örneklem dağılımının popülasyon dağılımına uyup uymadığı, iki değişkenin birbirine bağımlı olup olmadığı veya bir değişkenin başka bir değişkenle ilişkili olup olmadığının belirlenmesinde χ^2 dağılımı kullanılır. χ^2 tesadüfi bir değişken olmakla birlikte bir test istatistiğidir (19).

χ^2 dağılımı, t dağılımı gibi serbestlik derecesi ile hesaplanmaktadır. χ^2 dağılımının şeklini de serbestlik derecesinin büyüklüğü belirlemektedir. Dağılım küçük serbestlik derecelerinde sağa eğikken, serbestlik derecesi arttıkça normal dağılıma benzemektedir (20).

χ^2 dağılımı z dağılımı ile yakından ilişkilidir. X, ortalaması μ ve standart sapması σ_x olan normal dağılıma sahip bir tesadüfi değişken olsun,

$$z = \frac{\bar{x} - \mu}{\sigma_x}$$

değerinin karesi olan

$$z^2 = \left[\frac{\bar{x} - \mu}{\sigma_x} \right]^2$$

veya

$$\chi^2 = \frac{(\bar{x} - \mu)^2}{\sigma_x^2} \text{ 1 serbestlik dereceli } \chi^2 \text{ dağılımına sahiptir.}$$

Bu dağılıma sahip olasılıklar z tablosu yardımıyla $(-z)^2 = z^2$ hesaplanabilmektedir (19).

En eski ve en iyi uygunluk testi olarak bilinen χ^2 uygunluk testi ilk olarak 1890 yılında Pearson tarafından ele alınmıştır. χ^2 uygunluk testi aşağıdaki formülle elde edilir; X'in Gamma fonksiyonu $\Gamma(x)$ olarak gösterilir ve şu şekilde ifade edilir:

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt \quad x > 0$$

Gamma $\Gamma(x)$ 'in olasılık fonksiyonu ise şu şekildedir:

$$f(x; a, n) = \begin{cases} \frac{a^n e^{-ax} x^{n-1}}{\Gamma(n)} & : X > 0 \\ 0 & : X < 0 \end{cases}$$

α ve n, gamma olasılık fonksiyonunun parametreleri olup, $\alpha > 0$ ve $n > 0$ koşulları da sağlanmalıdır.

Açıkça gösterilebilir ki,

$$f(x) \geq 0 \text{ ve } \int_{-\infty}^\infty f(x; a, n) dx = 1 \text{ dir.}$$

Gamma olasılık fonksiyonunda iki parametrenin özel değerler alması χ^2 olasılık fonksiyonunu ortaya çıkarır.

$$\alpha = \frac{1}{2} \text{ ve } n = \frac{m}{2}$$

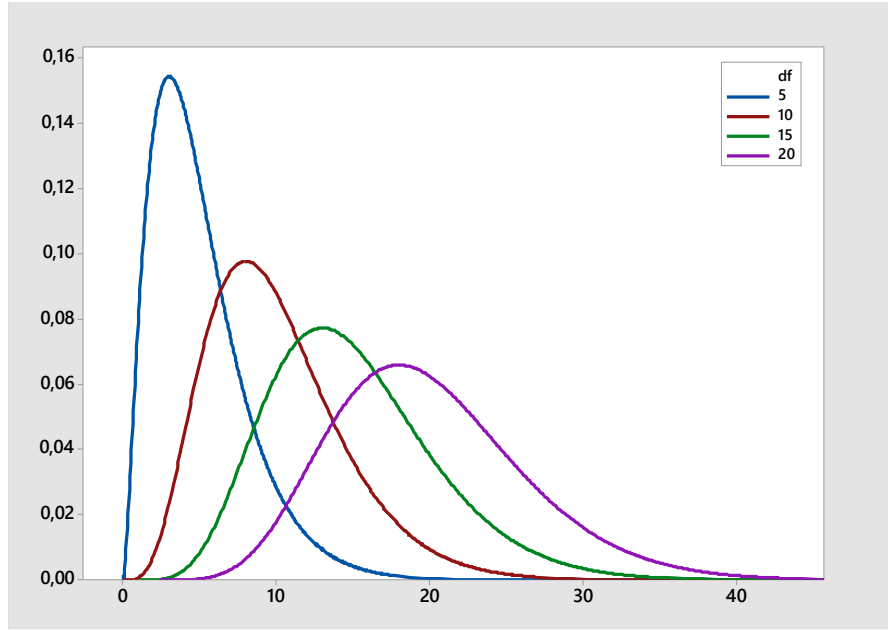
($m > 0$ ve tamsayı olduğu zaman $f(x; a, n)$ aşağıdaki gibidir.

$$f\left(x; \frac{1}{2}, \frac{m}{2}\right) = \begin{cases} \frac{e^{-1/2} x^{(m/2-1)}}{2^{(m/2)} \Gamma(m/2)} & X > 0 \\ 0 & X < 0 \end{cases}$$

$$f\left(x; \frac{1}{2}, \frac{m}{2}\right) \text{ olasılık fonksiyonu, serbestlik derecesi ve sürekli bir x değişkenine sahip } \chi^2 \text{ dağılımıdır. } \chi^2_{(m)}$$

ile gösterilen χ^2 olasılık fonksiyonunda (m) aynı zamanda fonksiyonun parametresidir. Dağılımın m parametresi serbestlik derecesidir.

χ^2 olasılık fonksiyonu yardımıyla serbestlik derecelerinden faydalanılarak χ^2 değerleri hesaplanabilmektedir. Dağılım serbestlik derecesine bağlı olduğundan her serbestlik derecesi için belli olasılıklara ait χ^2 dağılım tabloları oluşturulmuştur (19). Dağılım eğrilerinin çarpıklığı Şekil 1'de de görüldüğü gibi serbestlik dereceleri küçüldükçe artmaktadır. Serbestlik derecesi arttıkça dağılım normal dağılıma yaklaşmaktadır. Serbestlik derecesi 10'u geçtikten sonra dağılım simetrik bir görüntü oluşturmaya başlamaktadır (19).



Şekil 1: Farklı serbestlik dereceleri için χ^2 dağılımları
Figure 1: χ^2 distributions for different degrees of freedom

3. William Sealy Gosset'in Hayatı ve T Dağılımı

William Sealy Gosset'in soyu, 16. yüzyılda Fransa'da yaşayan Huguenot Protestanları mezhebine bağlı bir aileye dayanmaktadır (21-25). 1889 ile 1895 yılları arasında Winchester Koleji ve Oxford'a gitmiştir. Gosset ailesi, orta sınıf bir aile olduğu için William Sealy Gosset, kazandığı burslar ile eğitim hayatını sürdürmüştür. 1897 yılında Oxford'da matematik bölümünü birincilik ile bitirmiş, 1899 yılında ise kimya alanından birincilik ile mezun olmuştur (25-27).

Gosset'in okuldan mezun olduğu dönemde, Arthur Guinness, Son and Co. Ltd. adındaki bira üretim fabrikası en iyi ve ucuz birayı üretmek için matematikçi, istatistikçi ve kimyagerleri işe almıştır. Bu amaçla Guinness bira fabrikasının sahibi Arthur Guinness, William Sealy Gosset ile iletişime geçip Dublin'deki fabrikasında çalışması için işe almıştır. Gosset; hammaddelerin kalitesi ve üretim koşulları (örneğin, farklı aşamalarda sıcaklık değerleri) arasındaki ilişkileri analiz etmiştir. Daha sonra bu fabrikada yeni işe alınanlar arasında, en zeki kişi olarak deneysel bira fabrikasının başına getirilmiştir (26).

Gosset, matematik, kimya ve istatistik bilgilerini kullanarak en iyi verime sahip arpaları bulmak için deneyler yapmıştır. Gosset, bu deneylerden yola çıkarak birçok iç rapor yazmıştır. 1904 yılında Guinness bira fabrikası için "Hata Yasasının Bira Fabrikasındaki Çalışmalarına Uygulanması Üzerine" adlı bir iç rapor yazmıştır. Bu raporda, deneylerin sonuçlarına dayalı olarak kesin değerleri belirlemek için olasılık teorisinin kullanılmasının önemini vurgulamıştır. 1905 yılında, Guinness Kurulu tarafından da onaylanan "Pearson'ın Korelasyon Katsayısı" başlıklı bir iç rapor daha yazmıştır. Gosset'in bu raporları, geçerli sonuçlar elde etmek ve fabrikadaki üretim için kritik öneme sahip olmuştur. Bu raporlardaki yeni istatistiksel yöntemler sayesinde, üretilen ürünlerin kalitesi artmış ve Guinness bira fabrikası, bu dönemde en iyi bira üreten fabrika konumuna gelmiştir. Ayrıca bu raporlar, Gosset'in "Baş Bira Üreticisi" unvanını almasını sağlamıştır (26, 28-30).

Gosset, 35 yıl boyunca Guinness şirketinin fabrikalarında bira üreticisi olarak çalışmış ve bilimsel çalışmalarını bu dönemde yapmıştır. Gosset, aynı fabrikada çalışan Geoffrey Surrtees Phillpotts ile olan arkadaşlığı sayesinde Geoffret Surrtees Phillpotts'un kız kardeşi olan Marjory Surrtees Phillpotts ile tanışmıştır. Bu çift 16 Ocak 1906 yılında evlenmiştir. Gosset, 1937 yılında Guinness şirketinin yeni açtığı bira fabrikasında yönetici olmak için

Londra'ya taşınmıştır. Londra'ya taşındıktan kısa bir süre sonra, aynı yıl içinde, kalp krizi sonucu hayatını kaybetmiştir (25).

Gosset, yaptığı çalışmaları ve raporları yayımlamak istemiştir. Ancak Arthur Guinness, gizli belgelerin ifşa edilmesini engellemek için fabrikasında çalışan bilim adamlarının yaptıkları çalışmaları yayınlamalarını yasaklamıştır. Daha sonra gizli belgelerin ifşa edilmemesi ve yayınlarda takma ad kullanmak gibi şartlar ile yayınlamalarına izin vermiştir. Bu nedenlerden dolayı William Sealy Gosset, Guinness şirketinin bir çalışanı olduğu anlaşılmaması diye, bir çalışması hariç diğer tüm çalışmalarını “student” takma adıyla yayınlamıştır. Gosset, 1905 yılında, Biometrika dergisinin editörü olan Karl Pearson ile tanışmıştır. Çalışmalarını yayımlamak isteyen Gosset, bu durumu Pearson'a bildirince, Pearson'da bunu kabul etmiştir. Böylece 1905 yılında William Sealy Gosset, ilk makalesi olan “Hemositometre ile Sayma Hatası Üzerine” adlı makalesini “student” takma adıyla Biometrika dergisinde yayımlamıştır (29, 31). Gosset bu çalışmada, hücre sayılarının sonsuz hal aldığı durumlarda kendisinin geliştirdiği üstel serilerin kullanılmasını önermiştir. Ayrıca bu makalede, binom ve üstel serilerin karşılaştırılmasını yapmıştır. Karşılaştırma sonucunda, üstel serilerin normal dağılıma daha fazla yaklaştığını gözlemlemiştir. Gosset'in geliştirdiği bu üstel serilerin, 70 yıl öncesinde geliştirilen Poisson dağılımı olduğundan haberi olmamıştır (25, 32, 33).

Student'in kimliği, 16 Ekim 1937'de ki ölümüne kadar bir sır olarak saklanmıştır. Gosset'in ölümünden sonra, Karl Pearson'ın oğlu Eagon Pearson, Gosset'in makalelerini yayımlamak için Guinness şirketinden izin almıştır. Eagon Pearson, Gosset'in yazdığı makaleleri ve babası ile olan mektuplarını toplayarak Gosset'in biyografisini yazmıştır (28, 29).

T dağılımı

Gosset tarımda istatistiksel problemler üzerine çalışmalar yaparken elde ettiği t dağılımını z dağılımı adıyla ilk olarak “Bir Ortalamanın Muhtemel Hatası” adlı makalesinde 1908 yılında “student” takma adıyla yayımlamıştır (26). Ancak o dönemlerde küçük örneklemelerin istatistiğindeki genel bilgi eksikliği nedeniyle t dağılımı bira fabrikası dışında fazla kullanılmamıştır. Ronald Aylmer Fisher, Gosset'in “Bir Ortalamanın Muhtemel Hatası” adlı makalesini incelerken standart sapmadaki tutarsızlık nedeniyle standart sapmayı hesaplarken serbestlik derecesini “n” yerine “n-1” kullanmasını tavsiye etmiştir (34, 35). Ayrıca standart normal dağılımın da adı z dağılımı olduğundan dolayı Fisher, “t dağılımı” adını kullanmasını önermiştir. Gosset, bu önerilerden sonra makalesinde bazı düzeltmeler yaparak ve “z” yerine “t” kullanarak 1925 yılında bu makaleyi tekrar yayımlamıştır (36).

T dağılımı, William Sealy Gosset tarafından 1908'de “student” adıyla bir makalede tanıtılmıştır. Adını “student” kelimesinin son harfinden almaktadır. Bu makale normal dağılıma sahip bir popülasyondan örneklem oluşturduğumuzda t dağılımının normal dağılıma benzer olacağı ve ortalaması sıfır olan simetrik bir dağılımla karşılaşacağımızı belirtmektedir. T ve z dağılımları birbirine benzer olmakla aralarındaki temel farklılık t istatistiğinin örneklem dağılımının daha değişken olmasıdır. Bu durum t dağılımının $\bar{x}_i$ ve s gibi iki tane olasılık dağılımı hesaplamada kullanılan değişkene sahip olmasından kaynaklanmaktadır. Z dağılımında ise sadece $\bar{x}_i$ tek bir rassal değişken vardır. Örneklemdeki veri sayısı olan n küçüldükçe t dağılımının değişkenliği artmaktadır (20). Student t dağılımı, sürekli olasılık dağılımları grubunda yer almaktadır. T dağılımı, z puanlarının üzerinde yapılan düzenlemelerden sonra elde edilen t puanlarının dağılımıdır. T dağılımı, küçük örneklem boyutları için kullanılan bir normal dağılım türüdür. T dağılımında, serbestlik derecesi arttıkça t dağılımı puanlarının z dağılımı puanlarına yaklaştığı bildirilmiştir. Bundan dolayı, gözlem sayısının 30'dan büyük olduğu durumlarda, z dağılımının kullanılması önerilmiştir. Çünkü z dağılımında bilinen bir varyans kullanıldığından dolayı t dağılımından daha iyi tahminler yapılabilmektedir (37). T dağılımının kullanılabilmesi için her gruptaki veri sayısı 30'dan az ve veriler normal dağılım göstermelidir (20).

Z dağılımı, örneklem sayısının 30'dan büyük olduğu ve popülasyon varyansının bilindiği durumlarda kullanılan bir dağılımdır. Merkezi limit teoremine göre ortalaması μ , varyansı σ^2 olan bir popülasyondan seçilen $\bar{x}$ ortalamalı bir örneklem, yaklaşık olarak normal dağılmaktadır. Bu örneklem z puanı aşağıdaki formül ile hesaplanmaktadır (38).

$$Z = \frac{\bar{x} - \mu}{\sigma}$$

Popülasyon varyansının bilinmediği ve gözlem sayısının 30'dan küçük olduğu bir örneklemin dağılımı incelenmek istendiğinde, popülasyonun standart sapması yerine örneklemin standart sapması kullanılmaktadır. Örneklemin standart sapması kullanılarak incelenen bu dağılıma, t dağılımı denilmektedir (39). T dağılımı için gerekli olan t puanları, aşağıdaki formül ile hesaplanmaktadır (38).

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

T tablosundaki farklı serbestlik derecelerindeki t değerlerinin hesaplanması ve tablonun oluşturulması için olasılık yoğunluk fonksiyonu ile bir dizi hesaplama yapılmıştır. Student t dağılımında olasılık yoğunluk fonksiyonu, gama ve serbestlik derecesine bağlı olarak değişmektedir. Olasılık yoğunluk fonksiyonu formülü aşağıdaki gibidir (40).

$$f(t) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})} \left(1 + \frac{t^2}{v}\right)^{-\frac{(v+1)}{2}}$$

$f(t)$: T dağılımının olasılık yoğunluk fonksiyonu

Γ : Gama fonksiyonunu

π : Pi sayısını

v : Serbestlik derecesini temsil etmektedir (serbestlik derecesinin 1'den büyük olması gerekmektedir).

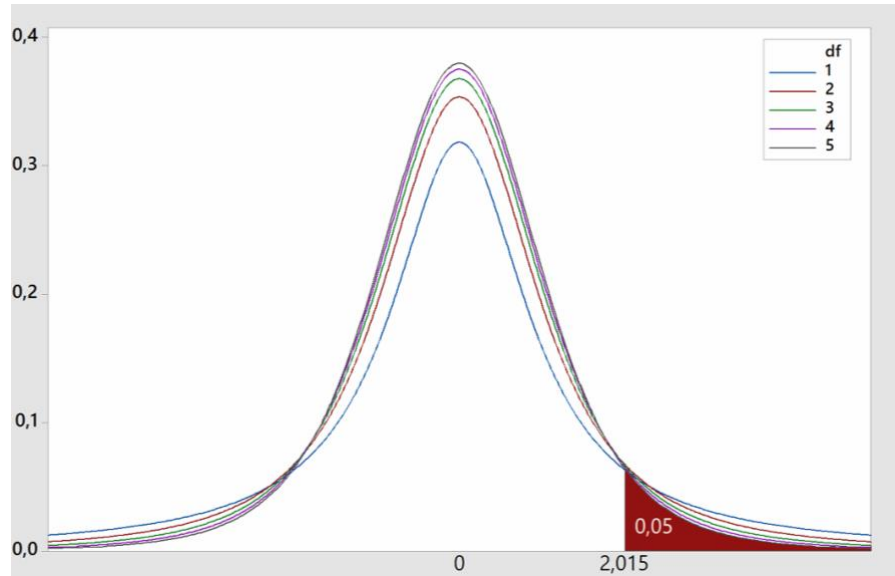
T puanı, t örneklem ortalaması $\bar{x}$, popülasyon ortalaması μ , örneğin standart sapması s , örneklem büyüklüğünün karekökü $\sqrt{n}$ olarak formülde verilmiştir. Formülde yer alan $s/\sqrt{n}$ standart hata olduğu için t değeri hesaplanırken, z değerinin aksine standart sapma değil standart hata kullanılmaktadır (40).

T dağılımı simetrik ve z dağılımına göre daha basık olmakla beraber dağılımın şekli serbestlik derecesine bağlıdır. İlk bakışta t ve z dağılımı birbirine çok benzerdir. Her iki dağılımda ortalama etrafında benzer dağılım gösterirler. T dağılımı yatay ekseninde daha yaygındır ve bu yaygınlık serbestlik derecesi ile ilişkilidir. $n \geq 30$ olduğu durumda t ve z dağılımları birbirine çok benzerdir (20). Özetle t dağılımının aşağıdaki iki yönü öne çıkmaktadır (41);

- Normal dağılıma göre değişkenliği daha fazladır.
- Değişkenlik derecesi serbestlik derecesi ile zıt orantılıdır.

Serbestlik derecesi: Popülasyon ortalaması incelenirken, genellikle popülasyonun standart sapması (σ) bilinmediğinden örneklemden elde edilen standart sapma (s) tahmin işleminde kullanılır. Dolayısıyla standart hata hesaplama işleminde σ yerine S kullanılmaktadır. S^2 σ^2 'nin tarafsız bir tahmin edicisi olduğundan tahmin edicinin yansızlık özelliğinden dolayı serbestlik derecesi hesaplanırken n yerine $(n-1)$ kullanılır. Ancak bu durum küçük örneklem için geçerlidir. Çünkü n 'in büyük olduğu durumda sonuca çok fazla etki etmez (41).

Bir veri dizisinde verilerin ortalamadan sapmalarının toplamının sıfır olduğu bilinmektedir. Yani ortalamadan $n-1$ tane sapma elimizde olduğundan son sapmada bilinmektedir. Bu sebeple t dağılımının şeklini örnekteki veri sayısı olan n yerine serbestlik derecesi olan $n-1$ belirler (20). Şekil 2'de de görüldüğü gibi veri sayısı azaldıkça dağılım eğrileri normal dağılıma göre kuyruklarda daha fazla yayılma eğilimi göstermektedir. Veri sayısı arttıkça normal dağılımdan farklılık azalmaktadır (19).



Şekil 2: Farklı serbestlik dereceleri için t dağılımları

Figure 2: t distributions for different degrees of freedom

4. Ronald Aylmer Fisher'in Hayatı ve F Dağılımı

Ronald Aylmer Fisher, 17 Şubat 1890'da Fisher ailesinin yedinci çocuğu olarak Birleşik Krallığa bağlı Doğu Finchley'de dünyaya gelmiştir(16). Henüz 16 yaşındayken bir matematik yarışmasında Neeld madalyasını kazanmıştır (44). Lise eğitiminden sonra Fisher 1909'da Cambridge'deki Gonville ve Caius Kolejinden burs kazanmıştır. Burada matematik bölümünü okumuş ve bir yandan da biyoloji ile ilgilenmiştir. Fisher Cambridge yıllarında genetik bilimine ilgi duymaya başlamış ve sonraki yıllarda yapacağı genetik çalışmalarını etkileyen Mendel genetiğinin teorisini öğrenmiştir (45). Fisher için genetik ve istatistik kadar önemli olduğunu düşündüğü sosyal bir sorun ise öjeniydi. Fisher, "Cambridge Üniversitesi Öjeni Derneği"ni kurup başkanlık yapmış ve öjeni ilkelerinin güçlü bir savunucusu olmuştur (17, 46).

Fisher 1912'de korelasyon katsayısı üzerine önemli çalışmalar yapmış ve maksimum olabilirlik yöntemini ortaya koyan ilk makalesini yayımlamıştır (47). 1915'te Biometrika dergisinde, p korelasyonlu iki değişkenli normal bir popülasyondan alınan örnekten türetilen korelasyon katsayısı olan r değerinin tam dağılımını yayımlamıştır (48). 1913'te mezun olduktan sonra Fisher, Büyük Britanya'nın I. Dünya Savaşı'na katılmasıyla orduya katılmak istemiş ancak görme bozukluğu nedeniyle tıbbi muayenelerden geçememiştir. Devletine hizmet etmek için Fisher bu dönemde devlet okullarında fizik ve matematik öğretmenini olarak görev yapmıştır. 1916'da yazılan ve 1918'de yayımlanan "Mendel Kalıtım Varsayımı Üzerine Akrabalar Arasındaki Korelasyon" dâhil olmak üzere birçok makale yayımlamıştır (43, 45).

Savaşın sonunda Fisher yeni bir iş ararken, iki iş teklifi birden almıştır. Bunlardan birisi Karl Pearson tarafından yönetilen dönemin ünlü Galton Laboratuvarı'ndan, diğeri de ülkede küçük bir tarım istasyonu olan Rothamsted Deney İstasyonudur. Ancak Fisher, Pearson ile gelişen rekabeti profesyonel bir engel olarak görmüş ve Galton Laboratuvarı teklifini kabul etmeyerek, özgür bir çalışma ortamı için 1919'da Rothamsted Tarım İstasyonu'nda istatistikçi olarak göreve başlamıştır (43). Fisher burada küçük örnekler için kesin anlamlılık testleri, tahmin teorileri özellikle maksimum olabilirlik yöntemi olmak üzere yeni istatistiksel yaklaşımlar geliştirmiştir. Ayrıca, modern deneysel analizlerin temellerini oluşturan, deney sonuçlarını ve hata düzeyini değerlendirmek için uygun bir yöntem sunan varyans analizinin iskeletini oluşturmuştur. Burada oluşturulan modern deneysel analiz yöntemleri tarım dışında diğer alanlarda da kullanılabilen yöntemler olarak nitelendirilmiştir (18, 45). Fisher Rothamsted'daki 14 yıllık

kariyerinde yaptığı çalışmaları “Mahsul Varyasyonu Çalışmaları” başlığı altında toplamıştır (49). Fisher, 1921 yılında Biometrika’ya gönderdiği makalesinde sınıf içi korelasyon katsayısına ilişkin öneriyi geliştirmiş ve z’nin örnekleme dağılımının normalliğe çok yakın olduğunu göstermiştir. Ancak Karl Pearson, Fisher’in yanıldığını ispatlamak ve yoluna zorluklar çıkarmak için Fisher’in makalesinin dergi tarafından reddedilmesine neden olmuştur. Daha sonra sınıf içi korelasyon ve z dönüşümüne ilişkin makalesini başka bir dergide yayımlamıştır (50).

Fisher’in varlığı Rothamsted’a çok şey katmıştır. Örneğin yağışın buğday verimi üzerine etkisi hakkında araştırmalar yapılmıştır. 1923’te büyük çapta tarımsal çalışmalar yapılarak deney tasarımı çalışmalarının kaynağı olan tarla denemeleri üzerine ilk makale yayımlanmıştır (18, 51). Fisher 1924’te kısmi korelasyon katsayısının dağılımını belirlemiş ve değişkenlerin ortadan kaldırılma etkisinin, ortadan kaldırılan her bir değişken için örneğin etki boyutunu bir birim kadar azaltmak olduğunu göstermiştir (42, 45).

Fisher 1925’te Varyans Analizi’ni tanıtmıştır. ANOVA (Analysis of Variance)’nın icadı bilim dünyasında yeni istatistiksel yöntemlerin geliştirilmesi için önemli bir adım olarak görülmüştür. Bu analizin kullanımları ilk olarak 1924’te Toronto’daki Uluslararası Matematik Kongresi’ne sunulan bir makalede sistematik olarak ortaya konulmuştur. Fisher’in bu dönemde yaptığı çalışmalar, 1925’te yayınlanan “Araştırma Görevlileri İçin İstatistiksel Yöntemler” adlı kitabında bir araya toplanmıştır (45, 52).

Fisher, genetik çalışmaları sayesinde 1933’te Karl Pearson’ın emekli olmasıyla onun yerine Londra Üniversitesinde Galton Öjenik Profesörü ve Galton Laboratuvarı başkanı olmuştur. Karl Pearson’ın emekli olmasından sonra İstatistik Bölümü’nün başına oğlu E. S. Pearson geçmiştir. Kayıtlarda Fisher’in deney analizine bilimsel açıdan yaklaşan ilk kişi olduğu belirlenmekte ve 1935 yılında yayınlanan “Deneylerin Tasarımı” adlı kitabının bu alana rehberlik ettiği yazılmaktadır. Fisher, yaklaşık olarak Gosset ile aynı zamanda ve Gosset’e uygun olarak, tedavi tekrarlarının bloklar halinde düzenlendiği deneylerde bir hata tahmini sağlamak için varyans analizinin kullanılmasını önermiş ve böylece Gosset’in grupta yöntemi arayışını çözmüştür (53, 54). Fisher 1943’te, mezun olduğu Cambridge Üniversitesinde, “Balfour Genetik Kürsüsü”nün başına geçmiş ve Cambridge’e taşınmıştır. Burada çalışmalarına devam etmiş ve 1949’da Cavalli Sforza ile birlikte “Akıbalı Yetiştirme Teorisi” adlı çalışmayı yayımlamıştır. 1947’de Fisher, Cyril Darlington ile beraber “Heredity: An International Journal of Genetics” dergisini kurmuştur. 1957’de Cambridge Üniversitesinden emekli olduktan sonra Fisher, Avustralya’nın Adelaide kentindeki CSIRO (Commonwealth Scientific and Industrial Research Organisation)’da kıdemli araştırma görevlisi olarak bir süre görev yaptıktan sonra 1962’de kolon kanserinden hayatını kaybetmiştir (45).

F Dağılımı

Fisher’in dağılım teorisi üzerine önemli istatistiksel çalışmaları olmuştur. O dönemde William Seally Gosset’in “student” t dağılımı üzerine çalışmalar yapmış ve kullanımını genişletmiştir. (55). Fisher’in t dağılımını kullanma ihtiyacı ise Rothamsted’daki deneysel çalışmaları sırasında olmuştur. Fisher, Rothamsted Deney İstasyonu’nda tarımsal çalışmalar sırasında değişkenler arasındaki ilişkiyi açıklamada ve sonuçların karşılaştırılması sırasında bazı zorluklarla karşılaşmıştır. Mevcut yöntemlerin, sonuçları değerlendirmede yetersiz kalması Fisher’in ANOVA’yı geliştirmesine ve bununla paralel olarak F dağılımının ortaya çıkmasına imkân sağlamıştır. Fisher “Araştırma Görevlileri İçin İstatistiksel Yöntemler” adlı kitabında F dağılımı hakkında bilgi verirken Gosset’in t dağılımının F dağılımının oluşmasındaki rolüne de vurgu yapmıştır. F dağılımının keşfi, ileri düzey veri analizinde yeni yöntemlerin ortaya çıkmasının önünü açmıştır (16).

F dağılımının kullanılabilmesi için öncelikle gruplarda varyansların eşit olması gerekmektedir (56). Aynı normal dağılıma sahip popülasyondan seçilen örneklerin varyanslarının oranının $(F = \frac{s_{max}^2}{s_{min}^2})$ 1’e eşit olması beklenir. Ancak bu oran her zaman 1’e eşit çıkmaz ve bir dağılım oluşturur. Varyansların oranlanması örneklem büyüklüğüne bağlıdır (57).

Popülasyon varyanslarının belirtilmediği ve birbirine eşit olup olmadığı bilinmediği durumlarda ortalamalar arası farka bakılmadan önce varyansların eşitliği varsayımını sağlamak için örneklem varyanslarından faydalanılarak

popülasyon varyanslarının eşit olup olmadığı kontrol edilir. Bir veri setini tanımlayan önemli iki parametreden birisi ortalama diğeri de varyanstır. Örneklem gruplarının karşılaştırılmasında ortalama yanında varyansında verilmesi gruplar arası farkın yanında verinin yaygınlığı hakkında da araştırmacının fikir sahibi olmasını sağlar. Bunun içinde F dağılımı kullanılır (58).

F dağılımında;

$F_{f_1;f_2;1-\alpha} = \frac{1}{F_{f_1;f_2;\alpha}}$ ilişkisi vardır. Eğer $f_1 = 1$ ise $F_{1;f_2;\alpha} = t_{f_2;\alpha/2}^2$ 'dir (58).

f_1 : Gruplar arası serbestlik derecesi

f_2 : Grup içi serbestlik derecesi

Teorem: Normal dağılıma sahip iki popülasyondan rastgele seçilen X_1 ve X_2 adında, n_1 ve n_2 sayıda veri içeren iki örneklem olduğunu varsayalım. Bunların aritmetik ortalamaları $\bar{x}_1$, $\bar{x}_2$ ve standart sapmaları $S_{\bar{x}_1}$, $S_{\bar{x}_2}$ olsun. Örnek varyansının popülasyon varyansına oranının serbestlik derecesi $(n-1)$ ile çarpımı $(n-1)$ serbestlik derecesinde bir χ^2 dağılımı oluşturmakta ve bu χ^2 değişkeni,

$$\chi_{(n-1)}^2 = \frac{(n-1) \cdot S_x^2}{\sigma_x^2}$$

şeklinde ifade edilmektedir. Daha sonra her iki taraf serbestlik derecesine bölünerek

$$\frac{\chi_{n-1}^2}{n-1} = \frac{S_x^2}{\sigma_x^2}$$

eşitliği elde edilir. Her iki örneklem için aynı işlem yapıldıktan sonra birbirine bölünerek

$$F = \frac{\frac{\chi_{n_1-1}^2}{n_1-1}}{\frac{\chi_{n_2-1}^2}{n_2-1}} = \frac{\frac{S_{x_1}^2}{\sigma_{x_1}^2}}{\frac{S_{x_2}^2}{\sigma_{x_2}^2}}$$

eşitliği bulunur. Eşitlikten de anlaşılacağı üzere F değeri negatif değer alamaz. Ayrıca F değeri χ^2 'den farklı bir değişkendir ve aynı bir olasılık dağılımına sahiptir. Yani iki χ^2 dağılımının serbestlik derecelerine oranlarının birbirine bölümü F olasılık dağılımını verir (19).

F dağılımının ortalaması ve varyansı;

$$E(X) = \frac{V_2}{V_2-2}, \quad V_2 > 2$$

$$\text{Var}(X) = \frac{2V_2^2(V_1+V_2-2)}{V_1(V_2-2)^2(V_2-4)}, \quad V_2 > 4$$

V_1 : Gruplar arası serbestlik derecesi

V_2 : Grup içi serbestlik derecesi

Yukarıdaki formülde de görüldüğü gibi F değeri negatif değerler alamaz ve sağa çarpık bir dağılımdır. V_1 ve V_2 serbestlik dereceli F dağılımına sahip X rastgele değişkeninin olasılık yoğunluk fonksiyonu;

$$f(x) = \int_0^{\alpha} \frac{\Gamma(\frac{V_1+V_2}{2})}{\Gamma(\frac{V_1}{2}) \Gamma(\frac{V_2}{2})} \left(\frac{V_1}{V_2}\right)^{\frac{V_1}{2}} x^{\frac{V_1}{2}-1} \left(1+\frac{V_1}{V_2}x\right)^{-\frac{V_1+V_2}{2}} dx = 1-\alpha$$

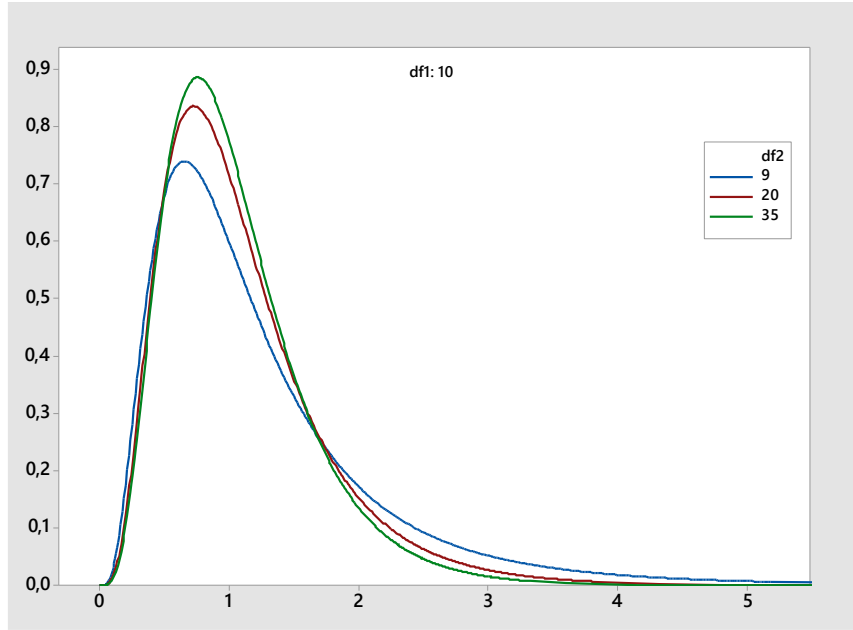
$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

V_1 = GASD

V_2 = GİSD

şeklindedir (59).

F dağılım tablosunda F değerinin belli aralıklarını oluşturan değerler yukarıdaki olasılık fonksiyonu yardımıyla hesaplanarak elde edilmiştir. Ancak manuel olarak bu hesaplamayı yapmak pratik olmadığından belli serbestlik dereceleri dikkate alınarak ve en çok kullanılan olasılıklar için hazır tablolar bulunmaktadır (19). Şekil 3'de de görüldüğü gibi F dağılımı t ve χ^2 dağılımı gibi tek bir dağılım değil her bir serbestlik derecesi çifti için bir dağılım olmak üzere dağılım eğrileri vermektedir(19).



Şekil 3: Farklı gruplar arası (df1) ve grup içi (df2) serbestlik dereceleri için F dağılımları
Figure 2: F distributions for different degrees of freedom (numerator (df1) and denominator (df2))

5. Ki-kare, t ve F dağılımları Arası İlişki

Dağılımlar arasındaki ilişkiye bakıldığında t dağılımı, sürekli değişkenlerin ortalamasına ilişkin hipotezi değerlendirmek için sıklıkla kullanılan bir olasılık dağılımıdır. Student t-dağılımı normal dağılıma oldukça benzer, ancak t-dağılımının tam şekli örneklem büyüklüğüne bağlı olarak değişmektedir. t dağılımı $-\infty, +\infty$ aralığında değer alır. Ayrıca F dağılımı ile t dağılımı arasında sıkı bir ilişki vardır. F dağılımı t dağılımının karesi şeklinde çözümlenir. $F = t^2$ şeklinde ifade edilir. Ancak F dağılımı, sadece pozitif değerlerden oluştuğundan olasılık fonksiyonu 0 ile $+\infty$ arasında bir dağılım oluşturur. $F = S_1^2 / S_2^2$ oranı F dağılımını oluşturur. Ayrıca F dağılımı t ve χ^2 dağılımı gibi bir tek dağılım değil her serbestlik derecesi çifti için bir dağılım vererek geniş bir dağılım kümesi oluşturur. F dağılımı t dağılımındaki gibi küçük örnek sayılarında basık bir görünüm almakta ancak örnek grubundaki veri sayısı arttıkça simetrik bir hal almakta ve dağılım normal dağılıma yaklaşmaktadır. χ^2 , t'de olduğu gibi tesadüfi değişken olmakla birlikte bir test istatistiği değeridir. χ^2 dağılımı da t dağılımında olduğu gibi tek bir eğri değil her serbestlik derecesi için bir eğri verir. χ^2 dağılımı F dağılımı gibi sadece pozitif değerler alabilmekte ve dağılım eğrisinin altında kalan alan 1'e eşit olmaktadır (19, 20).

$$\mu \rightarrow \chi^2 (f_1) \text{ ve } v \rightarrow \chi^2 (f_2)$$

$$f = \frac{\mu}{f_1} / \frac{v}{f_2} \quad \frac{1}{F_{S(f_1, f_2)}} = F_{1-S}(f_2, f_1)$$

f_2 ve f_1 : Serbestlik dereceleri

buradan, pay yada paydadan birinin serbestlik derecesini ∞ değerinde seçerek F dağılımı, (burada $f_2 = \infty$ alınarak f_1 serbestlik derecesine göre), $F(f_1, \infty) = \chi^2 (f_1)$ dağılımına eşit olduğu ve bu eşitlikle, F-dağılımı ile χ^2 dağılımı arasındaki ilişki görülmektedir. Özetle iki χ^2 dağılımının kendi serbestlik derecelerine oranlarının birbirine bölümü bir F olasılık dağılımı vermektedir (19, 59).

6. Sonuç

Ki-kare, t ve F dağılımları örnekleme dağılımları içerisinde önemli bir yer tutmaktadır. Günümüz bilim dünyasında kullanılan temel ve ileri düzey istatistiksel analizlerin bir çoğunun temel aldığı dağılım Ki-kare, t veya F dağılımlarından birisidir. Dolayısıyla herhangi bir bilimsel çalışmada istatistiksel analiz sonucunda elde edilen sonucun önemliliği testin temel aldığı dağılımla ilişkili olduğundan bu konunun bilinmesi tüm araştırmacılar için önemlidir. Dağılımların birbirlerine alternatif olarak ortaya çıkmaları ve sahip oldukları olasılık yoğunluk fonksiyonları dağılımlar arasındaki ilişki hakkında fikir vermektedir. Dağılımlar hakkında o dönemde yapılan çalışmalar bilim insanları arasında bazı anlaşmazlıkları meydana getirmiştir. T dağılımının serbestlik derecesinin belirlenmesi nispeten kolay olsa da Ki-kare dağılımının serbestlik derecesi bilimsel rekabet dolayısıyla bilim camiasında bir süre tartışmalara yol açmıştır. Pearson'ın normal dağılıma uymayan veriler için bir araştırma çabası içerisine girmesi ve Ki-kare dağılımını keşfetmesi nitel verilerin analizinde yeni bir başlangıç olmuştur. Gosset'in t dağılımını keşfetmesiyle iki grup karşılaştırmalarında normal dağılım gösteren veriler için bu dağılım kullanılmaya başlanmıştır. Ayrıca popülasyon varyansının bilinmediği durumda örnek varyansından hareketle standart hata kullanılarak analiz yapmak mümkün hale gelmiştir. Bu yönüyle t dağılımı z dağılımını tamamlayıcı rol oynamıştır. Fisher Rothamsted Tarım İstasyonu'nda yaptığı tarla deneyleri ile deney tasarımının temellerini oluşturmuştur. Ayrıca ANOVA'yı geliştirmesi ve korelasyon analiziyle ilgili yaptığı çalışmalar çok değişkenli istatistiksel analizlere geçişi sağlamıştır. Sonuç olarak yapılan çalışmalar dağılımların bugünkü halini almasında önemli rol oynamıştır. Örnekleme dağılımlarına temel teşkil eden Ki-kare, t ve F dağılımları ile ilgili literatür incelendiğinde bu dağılımlar hakkında ayrı ayrı teorik özelliklerine ait bilgiler bulunabilmekte ancak bu dağılımların kronolojik olarak tarihsel gelişimi ve birbirleriyle ilişkili yönleri hakkında kısıtlı bilgi bulunmaktadır. Dolayısıyla bu çalışma konuya ilgi duyan araştırmacılar için kaynak niteliğindedir. Gelişen teknoloji ve bilim dünyası ile birlikte dağılımların ortaya çıktığı ilk dönemdeki soruların cevapları daha kolay bulunabilmekte ve gelecek bilim dünyasında dağılımların tartışmalı yönlerinden hareketle yeni dağılım tiplerinin ortaya çıkması muhtemeldir.

Çıkar Çatışması Beyanı

Bu makalenin yazar/yazarları, çalışma kapsamında herhangi bir kişisel ve finansal çıkar çatışması bulunmamaktadır.

Finansal Kaynak Beyanı

Çalışmanın yürütülmesi sırasında alınan herhangi bir finansal kaynak bulunmamaktadır.

Yazar Katkısı Beyanı

Fikir/kavram: Mehmet Emin TEKİN

Denetleme/Danışmanlık: Mehmet Emin TEKİN, Tamer ÇAĞLAYAN

Kaynak taraması: Hakan SERİN, Davut SEYHAN, Seyit Mehmet TAŞDELEN

Makalenin yazımı: Hakan SERİN

Etik Onay

Bu makalededeki sunulan verilerin, bilgilerin ve dokümanların akademik ve etik kurallar çerçevesinde elde edildiği, tüm bilgi, belge, değerlendirme ve sonuçlarının bilimsel etik ve ahlak kurallarına uygun olarak sunulduğuna dair yazardan etik beyan alınmıştır.

Kaynaklar

1. Magnello ME. Karl Pearson and the establishment of mathematical statistics. *Int. Stat. Rev.* 2009 April;77:3-29.
2. Provine WB. The origins of theoretical population genetics 2. University of Chicago Press: Rosewood Drive; 2020.
3. Norton BJ. Karl Pearson and statistics: The social origins of scientific innovation. *Soc Stud Sci* 1978;8(1):3-34.
4. Pearson K. Karl Pearson: The Scientific Life in a Statistical Age. 2004.
5. Pearson ES. Karl Pearson. An Appreciation of Some Aspects of His Life Cambridge: Cambridge University. 1938.
6. Porter TM. Karl Pearson. British mathematician; 2022.
7. Yule GU, Filon LNG. Karl Pearson, 1857-1936. The Royal Society London; 1936.
8. Magnello ME. Karl Pearson's Gresham lectures: WFR Weldon, speciation and the origins of Pearsonian statistics. *The British journal for the history of science* 1996;29(1):43-63.
9. Pearson K. Notes on the history of correlation. *Biometrika* 1920;13(1):25-45.
10. Magnello ME. Karl Pearson's mathematization of inheritance: from ancestral heredity to Mendelian genetics (1895-1909). *Annals of Science* 1998;55(1):35-94.
11. Ögüş E. To be Together Medicine and Biostatistics in History. *Türkiye Klinikleri Biyoistatistik* 2017;9(1):74-83.
12. Pearson E. Pearson, Karl. An Appreciation of some aspects of his life and work, Part II. 1936:181-218.
13. Cain J. Karl Pearson praised Hitler and Nazi Race Hygiene. 2022.
14. Magnello ME. Karl Pearson, paper on the chi square goodness of fit test (1900). *Landmark Writings in Western Mathematics 1640-1940: Elsevier*; 2005. p. 724-31.
15. Pearson K. b. 27 March 1857-d 27 April 1936.
16. Rao CR. R. A. Fisher: The Founder of Modern Statistics. *Statistical Science* 1992;7(1):34-48, 15.
17. Fisher RA. The genetical theory of natural selection. Clarendon. Oxford; 1930.
18. Yates F. Sir Ronald Aylmer Fisher (1890-1962). *Jstor*; 1962.
19. Baykul Y. İstatistik: Metodlar ve uygulamalar: Anı Yayıncılık; 1999.
20. Gürsakal N. Çıkarımsal istatistik: SPSS-MINITAB uygulamalı: Dora Yayıncılık; 2009.
21. Adıbelli R. Katolik-Laik Tartışması Bağlamında Ortaya Çıkan Bir Disiplin Olarak Fransa'da Dinler Tarihinin Oluşum Süreci. *Çukurova Üniversitesi İlahiyat Fakültesi Dergisi*. 2011;11(2):149-92.
22. Ayfer G. Onyedinci Ve Onsekizinci Yüzyıl Düşünürlerinde Baskıya Karşı Direnme. *Journal of Istanbul University Law Faculty*. 2011;35(1-4):49-73.
23. Breathnach CS. Student. *Irish Journal of Psychological Medicine*. 1993;10(3):164-5.
24. Faraj QM. On altıncı yüzyıl Fransasında dini reformlar: Kahramanmaraş Sütçü İmam Üniversitesi, Sosyal Bilimler Enstitüsü; 2017.
25. Kemp CD, Pearson ES, Plackett RL, Barnard GA, editors. Student: A Statistical Biography of William Sealy Gosset; 1990.
26. Boland PJ. A Biographical Glimpse of William Sealy Gosset. *The American Statistician*. 1984;38(3):179-83.
27. Mane A. From a brewer to the faraday of statistics: William Sealy Gosset. *Muller Journal of Medical Sciences and Research* 2016;7:147.
28. Chan M, Lai J, Kolovos E, Bibra P. William Sealy Gosset: A History Of Beer And Statistics; 2017.
29. Raju TNK. William Sealy Gosset and William A. Silverman: Two "Students" of Science. *Pediatrics* 2005;116(3):732-5.
30. Boland PJ, editor William Sealy Gosset—An Inspiring 'Student'. *Proc 58th World Statistical Congress (Session STS028), Int Statistical Inst, Dublin*; 2011.
31. Pearson K. Student : a statistical biography of William Sealy Gosset. Oxford; New York: Clarendon Press; Oxford University Press; 1990.
32. Student. On the error of counting with a haemacytometer. *Biometrika* 1907;5(3):351-60.
33. Mather K. "Student's" Collected Papers. *Nature* 1943;152(3863):551-2.
34. Vyas SA, Desai SP. The Professor and the Student, Sir Ronald Aylmer Fisher (1890-1962) and William Sealy Gosset (1876-1937): Careers of two giants in mathematical statistics. *Journal of Medical Biography* 2015;23(2):98-107.

35. Wojnar R. Student's t-Distribution versus Zeldovich-Kompaneets Solution of Diffusion Problem. *Integration* 2012;510(1):15.
36. Box JF. Guinness, Gosset, Fisher, and small samples. *Statistical science* 1987;45-52.
37. Rumsey DJ. *Statistics essentials for dummies*: John Wiley & Sons; 2010.
38. Kim TK. T test as a parametric statistic. *Korean journal of anesthesiology*; 2015;68(6):540-6.
39. Pilling M. *Handbook of Statistical Distributions with Applications* 2nd edn K. Krishnamoorthy, 2016 Boca Raton, CRC Press 376 pp. 2017.
40. Zabell SL. On student's 1908 article "the probable error of a mean". *Journal of the American Statistical Association* 2008;103(481):1-7.
41. Gürtan K. *İstatistik ve araştırma metotları*: İstanbul Üniversitesi; 1971.
42. Mahalanobis P. Professor Ronald Aylmer Fisher. *Biometrics* 1964;20(2):238-52.
43. Rao C. Life and work of Ronald Aylmer Fisher. *Journal of Statistical Theory and Practice* 2007;1(3-4):489-99.
44. Fisher RA, Owen A. An Appreciation of the Life and Work of Sir Ronald Aylmer Fisher: FRS, FSS Sc. D. *Journal of the Royal Statistical Society Series D (The Statistician)* 1962;12(4):313-9.
45. Yates F, Mather K. Ronald Aylmer Fisher, 1890-1962. The Royal Society London; 1963.
46. Fisher RA. 007: *Biometrika*. 1916.
47. Hald A. On the history of maximum likelihood in relation to inverse probability and least squares. *Statistical Science* 1999;14(2):214-22.
48. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 1915;10(4):507-21.
49. Fisher RA. III. The influence of rainfall on the yield of wheat at Rothamsted. *Philosophical Transactions of the Royal Society of London Series B, Containing Papers of a Biological Character* 1925;213(402-410):89-142.
50. Fisher RA. 014: On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample; 1921.
51. Fisher RA, Mackenzie WA. Studies in crop variation. II. The manurial response of different potato varieties. *The Journal of Agricultural Science* 1923;13(3):311-20.
52. Williams R, Zumbo B, Zimmerman D. The scientific contributions of RA Fisher. *Edgeworth Series in Quantitative Educational and Social Sciences* 2001;7:1-23.
53. Zabell S. Ronald Aylmer Fisher. *Statisticians of the Centuries*: Springer; 2001. p. 389-97.
54. Skipper Jr RA. Sir Ronald Aylmer Fisher. *Philosophy of Biology*: Elsevier; 2007. p. 37-48.
55. Box JF. Gosset, Fisher, and the t distribution. *The American Statistician* 1981;35(2):61-6.
56. Gürsakal N. *Bilgisayar uygulamalı istatistik*: Alfa Yayınları; 2002.
57. Kendall MG. *Kendall's advanced theory of statistics*. Vol. 1. London: Hodder Arnold; 2007.
58. Gavcar E. *İstatistik Yöntemler 1*. Gazi Kitabevi Yayını, Muğla; 2006.
59. Erbaş SO. *Olasılık ve İstatistik*. Ankara: Gazi; 2013.