

# İki Değişkenli Normal Dağılımda Yeni Bir Regresyon Doğrusu ve Tatbikatı

*Dr. M. Emin ALTAN*

## Ö Z E T

Bilindiği gibi, iki değişkenli istatistik seriler bir frekans tablosu ile verilir. Bu tablonun her gözünde gösterilen sayılar, nazarı itibara alınan  $x, y$  değişken çiftlerinden ikisinin birden mevcut olma sayısını verir. Bu şekildeki sayıları, XY düzleminin gözönüne alınan gözünden düzleme çıkılan dikmeler üzerinde alırsak bir yüzey elde ederiz. Bu yüzeye ihtimal yüzeyi denir. Bu yüzeyin  $x = \text{sabit}$  ve  $y = \text{sabit}$  düzlemleri ile arakesiti birer Gauss eğrisi verir ise, dağılıma normal dağılım adı verilir ve buna ait yüzeye de Gauss tepesi denir.

Şu halde frekans tablosu ile verilen bir dağılımın maketi yapılsa Gauss tepesi elde edilince bu dağılımın normal dağılım olduğu sonucu çıkartılır. Aksi halde, yani Gauss tepesi elde edilmezse dağılım normal değildir.

Bu yazıda iki değişkenli dağılımın normal olduğunu kontrol edecek, düzlemsel bir metot verilmekte, ayrıca bu şekildeki dağılımlarda değişkenin bir diğerine mukabil değerinin en muhtemel değerinin bulunmasını temin eden ve pratikte raslanan birçok halde klâsik iki regresyon doğrusunun yerine geçen bir tek regresyon doğrusu verilmektedir.

Gerçekten, burada tarif edilen yeni regresyon doğrusunun eğimi, klâsik  $x_p = f(y)$  ve  $y_p = f(x)$  regresyon doğrularının eğimlerinin geometrik ortasıdır. Bu itibarla bu üç doğru daima aynı noktada kesişir ve XY düzleminde bu doğru klâsik  $x_p = f(y)$  ve  $y_p = f(x)$  doğrularının arasında bulunur. İki değişken arasındaki istatistik korelasyon ne kadar yüksekse,  $x_p = f(y)$  ve  $y_p = f(x)$  doğruları arasındaki açının o kadar küçük olduğu malumdur. O halde korelasyon katsayısının yüksek olması hallerinde  $x_p = f(y)$ ,

yahut  $y_p = f(x)$  regresyon denklemini kullanılarak hesaplanacak sonuçlar arasındaki farklar şüphesiz istatistik dalgalanmalardan daha küçüktür. İşte, yeni regresyon doğrusu bu halde karşılaşılan belirsizliği ortadan kaldırır ve iki ayrı doğru yerine bir tek doğruyu ikame eder.

İki değişkenli istatistik seriler için yukarıda açıklanan kurallar Bölüm II de 1. ve 2. paragraflarda normal dağılıma uygun iki örneğe tatbik edilerek sağlanmış, 3. paragrafta ise herhangi bir frekans cetveline tatbik edilerek, bunun normal dağılım olmadığı gösterilmiştir.

Bundan başka, Bölüm III de tek değişkenli bir dağılımın normal olup olmadığının çizimle anlaşılmasını temin edecek yeni bir test verilmiş ve iki örneğe de tatbik edilerek diğer normallik testleri ile mukayese yapılmıştır.

## Bölüm I

### TEORİK TEMELLER

#### § 1 — KULLANILAN TARİF VE FORMÜLLER

Esas konuya girmeden önce istatistiğin bazı temel tarifleri ile kullanılan sembollerin ifade ettiği mânalarının açıklanmasını, konunun takibinde kolaylık temin edeceği düşüncesi ile belirtilmesinde fayda görülmektedir. Zira istatistiğe ait literatürde oldukça farklı semboller kullanılmaktadır. Bu bakımdan her şeyden önce burada kullanılan sembollerin ifade ettiği anlamda bir birliğin temin edilmesi uygundur.

#### A. Tek değişkenli istatistik seriler

*Aritmetik ortalama.* — Tek değişkenli bir istatistik serinin terimleri  $x_1, x_2, \dots, x_n$  ile ve bu terimlerin frekansları  $n_1, n_2, \dots, n_n$  ile gösterilirse serinin toplam frekans sayısı

$$(1) \quad N = n_1 + \dots + n_n = \sum_{i=1}^n n_i$$

dir. Tek değişkenli bir istatistik serinin terimlerinin dizilişine *dağılım* da denir. Dağılımın  $\bar{x}$  ile gösterilen *ortalaması* N tane değişkenin *aritmetik ortalamasıdır*. Şu formülle bulunur:

$$(2) \quad \bar{x} = \frac{1}{N} (n_1x_1 + n_2x_2 + \dots + n_nx_n) = \frac{1}{N} \sum_{i=1}^n n_i x_i$$

*Dalgalanma (fluctuation), tipik sapma (écart-type)\*.*— Bir serinin terimlerinin  $\bar{x}$  ortalamasından sapmalarının karelerinin aritmetik ortalamasına *dalgalanma* denir. Bu da şu formülle hesaplanır:

$$(3) \quad \bar{x} m_2 = \sigma^2 = \frac{1}{N} \sum_{i=1}^n n_i (x_i - \bar{x})^2$$

Dalgalanma genel olarak  $\bar{x} m_2$  ile gösterilir.

Dalgalanmanın kare köküne *tipik sapma* denir.  $\sigma$  ile gösterilen tipik sapma

$$(4) \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^n n_i (x_i - \bar{x})^2}$$

formülü ile hesaplanır.

Dalgalanmanın,  $\bar{x}$  ortalama değeri yerine, başka bir  $x_0$  değerinden sapmalarının karelerinin aritmetik ortalamasını alarak da hesaplanması mümkündür. Bu halde iki dalgalanma arasındaki bağ şu şekildedir:

$$(5) \quad x_0 m_2 = \bar{x} m_2 + (\bar{x} - x_0)^2$$

*Momentler.*— Bir istatistik serinin  $x_0$  a nazaran  $q$  inci mertebeden momentini aşağıdaki formülle hesap edilir:

$$(6) \quad x_0 m_q = \frac{1}{N} \sum_{i=1}^n n_i (x_i - x_0)^q$$

\*) Ö. C. Sarc [1] «dalgalanma» yerine «varyans», «tipik sapma» yerine «standart sapma» terimlerini kullanmıştır. H. Furgaç [2] ise «tipik sapma» yerine «tipik inhiraf» terimini kullanmaktadır.

*Bazı momentler.*— I.  $x_0$  ne olursa olsun  $q=0$  ise, sıfırıncı mertebeden moment 1 e eşittir. Gerçekten :

$$x_0 m_0 = \frac{1}{N} \sum_{i=1}^n n_i (x_i - x_0)^0 = \frac{1}{N} \sum_{i=1}^n n_i = 1$$

II.  $x_0=0$  ve  $q=1$  olursa, başlangıca nazaran 1 inci mertebeden moment olup, değeri aritmetik ortalamadan ibarettir. Gerçekten :

$$(7) \quad {}_0 m_1 = \frac{1}{N} \sum_{i=1}^n n_i x_i = \bar{x}$$

III.  $x_0=\bar{x}$  ve  $q=2$  olursa, aritmetik ortalamaya göre 2 nci mertebeden moment, dalgalanmadan ibarettir. Gerçekten :

$$(8) \quad \bar{x} m_2 = \frac{1}{N} \sum_{i=1}^n n_i (x_i - \bar{x})^2 = \sigma^2$$

Momentlerde genel olarak  $x_0$  için başlangıca tekabül eden sıfır değeri, yahut da serinin aritmetik ortalaması olan  $\bar{x}$  değeri alınır.  $x_0=0$  alındığına göre hesaplanan  $q$  inci mertebeden moment  $m_q$  ile gösterilir; momente de sadece  $q$  inci mertebeden moment denir. Aynı şekilde  $x_0=\bar{x}$  alındığına göre hesaplanan  $q$  inci mertebeden moment  $\mu_q$  ile gösterilir; momente de  $q$  inci mertebeden *zahiri moment* denir.

Aşağıda bazı momentler yazılmıştır. Momentin mertebesini gösteren alt  $q$  indisinden sonra yazılan harfler değişkeni göstermektedir.

$$\begin{aligned}
 m_{qx} &= \frac{1}{N} \sum_{i=1}^n n_i x_i^q & m_{qu} &= \frac{1}{N} \sum_{i=1}^n n_i u_i^q \\
 m_{2x} &= \frac{1}{N} \sum_{i=1}^n n_i x_i^2 & m_{2u} &= \frac{1}{N} \sum_{i=1}^n n_i u_i^2 = \sigma_u^2 + \bar{u}^2 \\
 \mu_{qx} &= \frac{1}{N} \sum_{i=1}^n n_i (x_i - \bar{x})^q & \mu_{qu} &= \frac{1}{N} \sum_{i=1}^n n_i (u_i - \bar{u})^q \\
 \mu_{2x} &= \frac{1}{N} \sum_{i=1}^n n_i (x_i - \bar{x})^2 = \sigma_x^2 & \mu_{2u} &= \frac{1}{N} \sum_{i=1}^n n_i (u_i - \bar{u})^2 = \sigma_u^2
 \end{aligned}
 \tag{9}$$

B. İki değişkenli istatistik seriler. — Bir toplumda her aileyi baba ve oğul boylarına göre sınıflandırmak isteyelim. Her ailede oğulun boyu  $x$  ile babanın boyu da  $y$  ile gösterilsin. Bu halde her baba, oğul için değişkenleri  $(x_i; y_i)$  olan bir çift karşılık gelir. Her  $(x_i; y_i)$  çifti XY düzleminde  $P_i$  noktası ile gösterilir. Bu şekildeki grafik göstermeye bir *dağılım diyagramı* denir.

$P_i(x_i; y_i)$  noktaları hepsi birbirinden farklı olmayabilir. Eğer  $(x_i; y_i)$  çiftinden  $n_i$  tane varsa bu  $n_i$  ye çiftin *frekans*ı denir.

İki değişkenli dağılımda *moment*. — İki değişkenli dağılımda moment şu şekilde tarif edilir:

$$\mu_{rs}' = \frac{1}{N} \sum_{i=1}^n n_i x_i^r y_i^s
 \tag{10}$$

Burada  $x$  in  $r$  inci mertebeden,  $y$  nin de  $s$  inci mertebeden  $(0; 0)$  başlangıç noktasına göre momenti tarif edilmiştir.

*Bazı momentler* : I. Ortalamalar :

$$\mu_{10}' = \frac{1}{N} \sum_{i=1}^n n_i x_i = \bar{x}, \quad \mu_{01}' = \frac{1}{N} \sum_{i=1}^n n_i y_i = \bar{y}
 \tag{11}$$

Burada  $\bar{x}$  ile  $x$  dağılımının ortalaması ve  $\bar{y}$  ile de  $y$  dağılımının ortalaması gösterilmektedir.

II. Tipik sapmalar :

$$(12) \quad \mu_{20}' = \frac{1}{N} \sum_{i=1}^n n_i x_i^2 = \sigma_x^2 + \bar{x}^2, \quad \mu_{02}' = \frac{1}{N} \sum_{i=1}^n n_i y_i^2 = \sigma_y^2 + \bar{y}^2$$

Bunlar (9) formüllerinin ikinci satırına tekabül eder.

III. Değişken  $x$  ve  $y$  nin 1 inci dereceden kuvvetlerine göre moment :

$$(13) \quad \mu_{11}' = \frac{1}{N} \sum_{i=1}^n n_i x_i y_i$$

olur.

IV. Dağılımın  $(\bar{x}, \bar{y})$  ortalama değer çiftine nazaran birinci mertebeden moment  $\mu_{11}$  ile gösterilir ve değeri

$$\mu_{11} = \frac{1}{N} \sum_{i=1}^n n_i (x_i - \bar{x})(y - \bar{y})$$

yani

$$(14) \quad \mu_{11} = \mu_{11}' - \bar{x} \bar{y}$$

olur.

**C. Regresyon ve korelasyon : Regresyon doğruları.**— İstatistik iki  $x$  ve  $y$  değişkenleri ile belirtilen dağılım diyagramı genellikle, *regresyon eğrisi* adı verilen bir eğri civarında daha kesif dağılmış olur. Böyle bir eğrinin denklemi, değişkenlerin oldukça az hata ile birbirine yaklaştığı fonksiyonel bir bağıntıdır.

Özellikle, dikkatimizi verilen bir  $x$  değeri için  $y$  nin yakın bir ihtimalle değerini bulmayı temin eden doğru ile diğeri verilen bir  $y$  değeri için  $x$  in yakın bir ihtimalle değerini bulmayı temin eden doğru üzerine teksif edeceğiz.

Bu şekilde bulacağımız doğrunun fonksiyon tipinin,  $x$  serbest değişkeninin bir değerine mukabil  $y = mx + n$  ( $m$  ve  $n$  sabit) denklemi ile tekabül ettirilen  $y$  değerleri ve bu  $y$  değerlerine karşılık  $x = \frac{1}{m}y - \frac{n}{m}$  denklemi ile bulunan  $x$  değerlerini veren açık fonksiyon tipinden farklı karakterde olduğuna işaret etmeyi faydalı bulmaktayım [3].

Bu iki doğrudan birine  $y$  nin  $x$  e göre ve diğerine de  $x$  in  $y$  göre regresyon doğrusu denir.

Şimdi  $y$  nin  $x$  e göre regresyon doğrusunu gözönüne alalım. Bu doğrunun denklemi

$$(15) \quad y = bx + a$$

olsun. Burada  $a$  ve  $b$  sabit sayıları o şekilde tayin edilmelidir ki her  $x$  değeri için  $y$  nin *en iyi yaklaşık* değeri bu birinci derece doğru denklemi ile bulunabilsin. Buradaki en iyi yaklaşık değer ifadesinden en küçük kareler metoduna uygunluğu kastedilmektedir [4]. Yani istatistik seri ile verilen  $y$  değerleri ile (15) den hesap edilen  $y$  değerleri arasındaki farkların karelerinin toplamı en küçük olacak şekilde  $a$  ve  $b$  nin tayin edileceği anlaşılmaktadır.

Bu ise  $\sum_{i=1}^n n_i (y_i - a - bx_i)^2$  toplamının minimum olmasını temin eden

$a$  ve  $b$  nin araştırılması demektir.

Toplamın minimum olması için  $a$  ve  $b$  ye göre kısmî türevlerin sıfır olması gerektir. O halde şu denklemler elde edilir:

$$\sum_{i=1}^n n_i (y_i - a - bx_i) = 0 \quad \sum_{i=1}^n n_i x_i (y_i - a - bx_i) = 0$$

Bu denklemler de (11), (12), (14) yardımı ile aşağıdaki şekilleri alır:

$$(16) \quad \bar{y} - a - b\bar{x} = 0$$

$$(17) \quad \mu_{11} + \bar{x}\bar{y} - a\bar{x} + b(\sigma_x^2 + \bar{x}^2) = 0$$

Bu denklemlerden

$$b = \frac{\mu_{11}}{\sigma_x^2}$$

bulunur. (16) denklemi tayin edilecek doğrunun  $(\bar{x}, \bar{y})$  noktasından geçtiğini gösterir. O halde doğrunun denklemi

$$y - \bar{y} = \frac{\mu_{11}}{\sigma_x^2} (x - \bar{x})$$

olur. Buradaki  $\frac{\mu_{11}}{\sigma_x^2}$  eğimine  $y$  nin  $x$  e göre *regresyon katsayısı* denir.

Bu denklem ile iki değişkenli serinin her hangi bir  $x$  değerine karşılık en muhtemel  $y$  değeri elde edilir. Bulunan bu  $y$  değeri,  $x$  sayısının bir değerine tekabül eden muhtelif  $y$  değerlerinden bir tanesine eşit olmak zorunda değildir [5]. Bu bakımdan denklemi şu şekilde yazmak daha elverişlidir:

$$(18) \quad y_p - \bar{y} = \frac{\mu_{11}}{\sigma_x^2} (x - \bar{x})$$

Benzer muhakeme yolu ile  $x$  in  $y$  ye nazaran regresyon doğrusunu elde etmemiz imkân dahilindedir. Ancak simetri özeliğinden faydalanarak (18) denkleminde değişkenler arasında yer değiştirme yaparak  $x$  in  $y$  ye göre regresyon doğrusu hemen elde edilebilir:

$$(19) \quad x_p - \bar{x} = \frac{\mu_{11}}{\sigma_y^2} (y - \bar{y})$$

Burada da  $\frac{\mu_{11}}{\sigma_y^2}$  eğimine  $x$  in  $y$  ye göre *regresyon katsayısı* denir.

İki regresyon katsayısının çarpımının kare köküne *korelasyon katsayısı* denir ve  $\rho$  ile gösterilen bu katsayının ifadesi

$$\rho = \frac{\mu_{11}}{\sigma_x \sigma_y}$$

dir.



$\rho$  korelasyon katsayısı kullanılarak regresyon doğrularının (18) ve (19) ile gösterilen denklemleri aşağıdaki şekillerde yazılır:

$$(20) \quad y_p - \bar{y} = \rho \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

$$(21) \quad x_p - \bar{x} = \rho \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

Bu denklemler, koordinatların başlangıç noktası olarak  $(\bar{x}; \bar{y})$  noktasını alırsak, şu şekle girer:

$$y_p = \rho \frac{\sigma_y}{\sigma_x} x$$

$$x_p = \rho \frac{\sigma_x}{\sigma_y} y$$

*Tahmindeki tipik hatanın tayini.*

Şimdi de  $y$ 'nin  $x$ 'e nazaran regresyon doğrusu yardımı ile bir  $x_i$  değerine karşılık bulunan  $y_p$  değeri ile  $x_i$  lere mukabil gözlenen  $y_i$  değerleri arasındaki farkların kareleri toplamının en küçük olmasını istediğimiz değeri hesap edelim.

Bu toplam

$$\begin{aligned} \sum_{i=1}^n d_i^2 &= \sum_{i=1}^n (y_i - y_p)^2 = \sum_{i=1}^n \left( y_i - \rho \frac{\sigma_y}{\sigma_x} x_i \right)^2 \\ &= \sum_{i=1}^n y_i^2 - 2\rho \frac{\sigma_y}{\sigma_x} \sum_{i=1}^n y_i x_i + \rho^2 \frac{\sigma_y^2}{\sigma_x^2} \sum_{i=1}^n x_i^2 \end{aligned}$$

şeklinde yazılır. Diğer taraftan

$$\sum_{i=1}^n y_i^2 = N\sigma_y^2, \quad \rho = \frac{\sum_{i=1}^n y_i x_i}{N\sigma_x \sigma_y} \quad \text{ve} \quad \sum_{i=1}^n x_i^2 = N\sigma_x^2$$

olduğundan, toplam

$$\sum_{i=1}^n d_i^2 = N\sigma_y^2 - 2N\rho^2\sigma_y^2 + N\rho^2\sigma_y^2 = N\sigma_y^2(1 - \rho^2)$$

olur. Bu toplamı  $NS_y^2$  ile gösterelim. O halde

$$S_y^2 = \sigma_y^2(1 - \rho^2)$$

olur.

$S_y^2$  ifadesi noktaların,  $y$  nin  $x$  e göre regresyon doğrusuna nazaran sapmalarının karelerinin ortalaması olduğundan  $S_y$  ye  $y$  nin, denklemi (18) olan regresyon doğrusuna nazaran, *tahmindeki tipik hatası* denir [6].

#### D. İki değişkenli normal dağılımın özel gösterilişi.

İki  $X$ ,  $Y$  değişkenine göre normal olan bir dağılım gözönüne alalım. Bu değişkenlere tekabül eden tipik sapmalar  $\sigma_x$ ,  $\sigma_y$  ve korelasyon katsayısı da  $\rho$  olsun.

Dağılım bir Gauss yüzeyi ile temsil edilebilir; bunun denklemi bilindiği üzere [7] şu şekildedir:

$$(22) \quad P(X, Y) = \frac{1}{2\pi\sqrt{1 - \rho^2}\sigma_x\sigma_y} e^{-\frac{1}{2(1 - \rho^2)} \left[ \frac{(X - \bar{X})^2}{\sigma_x^2} - 2\rho \frac{(X - \bar{X})(Y - \bar{Y})}{\sigma_x\sigma_y} + \frac{(Y - \bar{Y})^2}{\sigma_y^2} \right]}$$

Burada  $\bar{X}$  ve  $\bar{Y}$  değişkenlerin ortalama değeri olup aynı zamanda  $X$  ve  $Y$  nin muhtemel değerleridir. (22) ifadesi  $X$  ve  $Y$  değerlerinin aynı zamanda zuhur etmesi ihtimalini göstermektedir.

Yukardaki (22) ifadesinde üstel terimin çarpanı  $X$  ve  $Y$  ye tabi değildir. Bunu bir  $C$  sabiti ile gösterelim ve (22) denklemini aşağıdaki yeni değişkenler cinsinden yazalım:

$$(23) \quad x = \frac{X - \bar{X}}{\sigma_x}, \quad y = \frac{Y - \bar{Y}}{\sigma_y}$$

Böylece (22) denklemi şu şekli alır:

$$P(X, Y) = C e^{-\frac{1}{2(1 - \rho^2)} (x^2 - 2\rho xy + y^2)}$$

Burada üs ifadeye  $\rho^2 y^2 - \rho^2 y^2$  ilâve edilir ve terimleri yeniden sıralanırsa

$$P(X, Y) = C e^{-\frac{1}{2(1-\rho^2)} [(1-\rho^2)y^2 + (x-\rho y)^2]}$$

olur. Bu ifade de iki üstel fonksiyonun çarpımı şeklinde yazılır :

$$(24) \quad P(X, Y) = C e^{-\frac{y^2}{2}} \cdot e^{-\frac{1}{2(1-\rho^2)} (x-\rho y)^2}$$

Bu çarpımdaki üsleri aşağıdaki özellikleri ifade edecek şekilde yazalım :

Verilen bir  $Y$  için  $X$  in en muhtemel değerini  $X_p$  ile göstere-  
lim: Bu halde  $x = \rho y$  olduğu zaman  $X$  de  $X_p$  olur. O halde bu  
 $X_p$  ve  $Y$  değerleri (21) denklemini sağlar. Yâni

$$X_p - \bar{X} = \rho \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})$$

veya

$$(25) \quad X_p = \bar{X} + \rho \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})$$

olur. Bundan başka (23) değişim formülleri kullanılarak

$$x - \rho y = \frac{X - X_p}{\sigma_x}$$

bulunur. O halde (24) denklemi şu şekli alır :

$$(26) \quad P(X, Y) = C e^{-\frac{(Y-\bar{Y})^2}{2\sigma_y^2}} \cdot e^{-\frac{1}{2(1-\rho^2)\sigma_x^2} (X-X_p)^2}$$

Buradaki  $C$  sabitini, her iki üstel çarpan normal bir dağılım  
olacak şekilde, iki çarpana ayırarak yazarsak

$$(27) \quad P(X, Y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(Y-\bar{Y})^2}{2\sigma_y^2}} \cdot \frac{1}{\sqrt{2\pi}(1-\rho^2)\sigma_x} e^{-\frac{(X-X_p)^2}{2(1-\rho^2)\sigma_x}}$$

elde edilir.

Buradaki ikinci çarpana  $Y$  nin belli bir değeri için  $X$  in  $X_p$  değerini aldığı zamanki *şartlı ihtimal yoğunluğu* denir [8].

Gözönüne aldığımız (22) ifadesi  $X$  ve  $Y$  ye nazaran simetrik olduğundan (27) ye tamamen benzeyen fakat  $X$  ile  $Y$  nin rolleri mübadele edilmiş bulunan bir denklemde yazılabilir.

Bu sebepten formel olarak şu iki fonksiyon tarif edilebilir:

$$F(X) = \frac{1}{\sqrt{2\pi} \sigma_x} e^{-\frac{(X-\bar{Y})^2}{2\sigma_x^2}}$$

$$G(Y) = \frac{1}{\sqrt{2\pi} (1-\rho^2) \sigma_y} e^{-\frac{(Y-Y_p)^2}{2(1-\rho^2)\sigma_y^2}}$$

Burada,  $X_p$  için yazılan (25) denklemi gibi,  $Y_p$  için de şu denklem yazılır:

$$(28) \quad Y_p = \bar{Y} + \rho \frac{\sigma_y}{\sigma_x} (X - \bar{X})$$

Demek ki (27) şeklinde yazılan (22) denklemini simetri hassasından dolayı şu şekilde yazabiliriz:

$$(29) \quad P(X, Y) = F(X) G(Y) = F(Y) G(X)$$

Böylece iki değişkenli normal dağılımın, değişkenlerin *şartlı ihtimal yoğunluk fonksiyonlarına göre*, özel bir yazılışı elde edilmiş olur.

## § 2 — İKİ DEĞİŞKENLİ NORMAL DAĞILIMDA YENİ BİR REGRESYON DOĞRUSUNUN TAYİNİ

İki değişkenli normal bir dağılımın frekans tablosunu gözönüne alalım. Bu değişkenlerin toplam dağılım fonksiyonlarını da sırasıyla  $f(X)$  ve  $h(Y)$  ile gösterelim. Herhangi birisinin belli bir değeri için diğerinin dağılımını veren *şartlı ihtimal fonksiyonları* da

$$\frac{1}{\sqrt{2\pi (1-\rho^2) \sigma_y}} e^{-\frac{(Y-Y_p)^2}{2(1-\rho^2) \sigma_y^2}}$$

ve

$$\frac{1}{\sqrt{2\pi (1-\rho^2) \sigma_x}} e^{-\frac{(X-X_p)^2}{2(1-\rho^2) \sigma_x^2}}$$

olsun. Bunlardan birincisi X in belli bir  $x$  değerine karşılık Y nin dağılım fonksiyonu, ikincisi de belli bir  $y$  değerine karşılık X in dağılım fonksiyonunu göstermektedir.

O halde bir  $(x, y)$  çiftinin birlikte seçilmesinin ihtimali, ihtimallerin çarpımı kuralına göre [9]

$$I(X, Y) = f(X) \frac{1}{\sqrt{2\pi (1-\rho^2) \sigma_y}} e^{-\frac{(Y-Y_p)^2}{2(1-\rho^2) \sigma_y^2}}$$

eşitliği ile, yahut simetriden dolayı

$$I(X, Y) = h(Y) \frac{1}{\sqrt{2\pi (1-\rho^2) \sigma_x}} e^{-\frac{(X-X_p)^2}{2(1-\rho^2) \sigma_x^2}}$$

eşitliği ile belli olur.

Sol tarafları eşit olan bu iki eşitlikten aşağıdaki eşitlik yazılabilir :

$$(30) \quad \frac{f(X)}{\sqrt{2\pi (1-\rho^2) \sigma_y}} e^{-\frac{(Y-Y_p)^2}{2(1-\rho^2) \sigma_y^2}} = \frac{h(Y)}{\sqrt{2\pi (1-\rho^2) \sigma_x}} e^{-\frac{(X-X_p)^2}{2(1-\rho^2) \sigma_x^2}}$$

Gözönüne aldığımız dağılımda  $f(X)$ ,  $h(Y)$  toplam dağılımları ile X ve Y değişkenlerinin arasında (30) eşitliği mevcut olduğuna göre toplam dağılımların oranı nasıl değişir? Bundan başka değişkenlerin birinin belli bir değerine karşılık diğerinin en muhtemel değeri nasıl bulunur?

Bu soruların cevaplarını sırasıyla araştıralım. Önce birinci sorunun cevabı olarak toplam dağılımların oranının dağılımını bulalım. Bunun için de aşağıdaki oranı teşkil edelim :

$$(31) \quad U(X, Y) = \frac{h(Y)}{f(X)} = \frac{\sigma_x}{\sigma_y} e^{-\frac{(Y-Y_p)^2}{2(1-\rho^2)\sigma_y^2} + \frac{(X-X_p)^2}{2(1-\rho^2)\sigma_x^2}}$$

Bu oran, serbest değişkenleri X ve Y olan bir ihtimal yüzeyini göstermektedir.

İkinci sorunun cevabını bulmak için U(X,Y) yüzeyini U=sabit düzlemi ile keselim.

Bu yüzeyin U=sabit düzlemi ile arakesiti olan eğrinin denklemi

$$U = \frac{\sigma_x}{\sigma_y} e^{-\frac{(Y-Y_p)^2}{2(1-\rho^2)\sigma_y^2} + \frac{(X-X_p)^2}{2(1-\rho^2)\sigma_x^2}}$$

veya

$$\ln U - \frac{\sigma_x}{\sigma_y} = -\frac{(Y-Y_p)^2}{2(1-\rho^2)\sigma_y^2} + \frac{(X-X_p)^2}{2(1-\rho^2)\sigma_x^2}$$

dir. Bu denklem de

$$2(1-\rho^2) \ln U - \frac{\sigma_y}{\sigma_x} = \frac{(X-X_p)^2}{\sigma_x^2} - \frac{(Y-Y_p)^2}{\sigma_y^2}$$

şeklinde yazılır. Bu denklemin sol tarafı  $\alpha$  ile gösterilir ve  $X_p, Y_p$  en muhtemel değerleri için (25) ve (28) deki değerleri yazılırsa

$$(32) \quad \alpha = \frac{\left[ X - \bar{X} - \rho \frac{\sigma_x}{\sigma_y} (Y - \bar{Y}) \right]^2}{\sigma_x^2} - \frac{\left[ Y - \bar{Y} - \rho \frac{\sigma_y}{\sigma_x} (X - \bar{X}) \right]^2}{\sigma_y^2}$$

denklemini bulunur. Bu denklemin sağ tarafı çarpanlarına ayrılırsa aşağıdaki şekli alır (köşeli parantezlere alt indisler birinci ve ikinci çarpanı göstermek için konulmuştur):

$$\alpha = \left[ \frac{X - \bar{X} - \rho \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})}{\sigma_x} + \frac{Y - \bar{Y} - \rho \frac{\sigma_y}{\sigma_x} (X - \bar{X})}{\sigma_y} \right]_1$$

$$\left[ \frac{X - \bar{X} - \rho \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})}{\sigma_x} + \frac{-Y + \bar{Y} + \rho \frac{\sigma_y}{\sigma_x} (X - \bar{X})}{\sigma_y} \right]_2$$

elde edilir. Birinci ve ikinci çarpanları basitleştirirsek

$$\left[ \begin{array}{l} 1 \\ 2 \end{array} \right] = \frac{(X-\bar{X}) \sigma_y - \rho \sigma_x (Y-\bar{Y}) + (Y-\bar{Y}) \sigma_x - \rho \sigma_y (X-\bar{X})}{\sigma_x \sigma_y}$$

$$\left[ \begin{array}{l} 1 \\ 2 \end{array} \right] = \frac{(X-\bar{X}) \sigma_y - \rho \sigma_x (Y-\bar{Y}) - (Y-\bar{Y}) \sigma_x - \rho \sigma_y (X-\bar{X})}{\sigma_x \sigma_y}$$

elde edilir. O halde (32) denkleminin sağ tarafı

$$\left[ \begin{array}{l} 1 \\ 2 \end{array} \right] = \frac{[(1-\rho)\sigma_y(X-\bar{X}) + (1-\rho)\sigma_x(Y-\bar{Y})][(1+\rho)\sigma_y(X-\bar{X}) - (1+\rho)\sigma_x(Y-\bar{Y})]}{\sigma_x^2 \sigma_y^2}$$

veya

$$\left[ \begin{array}{l} 1 \\ 2 \end{array} \right] = \frac{(1-\rho^2) [\sigma_y^2 (X-\bar{X})^2 - \sigma_x^2 (Y-\bar{Y})^2]}{\sigma_x^2 \sigma_y^2}$$

şeklinde yazılır. Demek ki (32) denklemi

$$2(1-\rho^2) \ln U \frac{\sigma_x}{\sigma_y} = (1-\rho^2) \left[ \frac{(X-\bar{X})^2}{\sigma_x^2} - \frac{(Y-\bar{Y})^2}{\sigma_y^2} \right]$$

veya

$$(33) \quad \ln U \frac{\sigma_y}{\sigma_x} = \frac{(X-\bar{X})^2}{\sigma_x^2} - \frac{(Y-\bar{Y})^2}{\sigma_y^2}$$

şeklini alır. Bu denklem ise merkezinin koordinatları  $\bar{X}$  ve  $\bar{Y}$  olan bir hiperbolü gösterir.

Bu hiperbol

$$(34) \quad \ln U \frac{\sigma_y}{\sigma_x} - 0$$

olunca kesişen iki doğru olur. Bu doğruların denklemleri de (33) denkleminde hemen aşağıdaki şekilde elde edilir:

$$(35) \quad Y-\bar{Y} = \frac{\sigma_y}{\sigma_x} (X-\bar{X}), \quad Y-\bar{Y} = -\frac{\sigma_y}{\sigma_x} (X-\bar{X})$$

dir. Bu denklemleri, koordinatların başlangıç noktası olarak  $(\bar{X}, \bar{Y})$  alırsa, şu şekilde yazabiliriz:

$$(36) \quad Y = \frac{\sigma_y}{\sigma_x} X, \quad Y = -\frac{\sigma_y}{\sigma_x} X$$

Değişkenlerden birinin belli bir değerine karşılık diğerinin değerini veren bu denklemler konumuzun başında ileri sürülen sorunun cevabını vermekte olup tayinini istediğimiz regresyon doğrularıdır.

*Bu regresyon doğrularının özellikleri:*

I. Bu şekilde tayin ettiğimiz yeni doğruların eğimlerinin mutlak değeri olan  $\frac{\sigma_y}{\sigma_x}$ , (25) ile gösterilen regresyon doğrusunun

$m_1 = \frac{1}{\rho} \frac{\sigma_y}{\sigma_x}$  eğimi ile (28) ile gösterilen regresyon doğrusunun

$m_2 = \rho \frac{\sigma_y}{\sigma_x}$  eğiminin arasında geometrik ortadır.

II. Çizilmesi.— (34) eşitliğinin mevcut olmasından

$$U \frac{\sigma_y}{\sigma_x} = 1$$

veya U nun yerine toplam dağılım fonksiyonları cinsinden değerini koyarsak

$$(37) \quad \frac{h(Y)}{f(X)} = \frac{\sigma_x}{\sigma_y}$$

eşitliği elde edilir. Bu bulduğumuz (37) denkleminin ancak X ve Y nin  $\bar{X}$  ve  $\bar{Y}$  değerleri için gerçekleştiği (32) den hemen görülür. Zira, ancak  $X = \bar{X}$  ve  $Y = \bar{Y}$  değerleri için (32) nin sol tarafı sıfır olur ve  $\alpha = 2(1 - \rho^2) \ln U \frac{\sigma_y}{\sigma_x} = 0$  olur. Bu da  $\rho = \pm 1$  özel halini bir tarafa bırakırsak  $\ln U \frac{\sigma_y}{\sigma_x} = 0$  olacağını gösterir.

O halde, (36) denklemlerindeki eğimler  $\frac{h(\bar{Y})}{f(\bar{X})}$  oranı ile bulunur. Bu duruma göre  $X - \bar{X}$  ve  $Y - \bar{Y}$  farklarının oranı sabittir. Bu bakımdan X ve Y değişkenlerinin dağılım fonksiyonlarının grafiği çizilmiş ise ve  $f(\bar{X})$  ile  $h(\bar{Y})$  değerleri normal dağılım eğrilerinde olduğu gibi kolayca bulunabileceğinden  $X - \bar{X}$  ile  $Y - \bar{Y}$  nin birbirine bağlılığını gösteren

$$(38) \quad Y - \bar{Y} = \frac{h(\bar{Y})}{f(\bar{X})} (X - \bar{X})$$

denklemin, doğrusu hemen çizilebilir.



Karşıt olarak X ve Y değişkenlerini dağılım grafiklerinden  $\bar{X}$  ve  $\bar{Y}$  değerlerinden itibaren uzaklıkları oranı,  $\frac{h(\bar{Y})}{f(\bar{X})}$  oranına eşit olan noktalar alırsak, bunlardan X i apsis, Y yi de ordinat olarak almakla çizilen noktaların (36) doğrusu üzerinde olduğu görülmelidir.

Bu karşıt çizimle elde ettiğimiz noktaların geometrik yeri bir doğru ise gözönüne aldığımız frekans tablosu bir normal dağılımı göstermektedir. Çizdiğimiz doğru da yeni bir regresyon doğrusu olarak mütalâa edilebilir.

*Tipik hatanın tayini.*— Normal dağılımı gösteren bir frekans tablosu için bulunan  $y$  ile  $y$  nin  $x$  e göre regresyon doğrusu  $y_p = \frac{\sigma_y}{\sigma_x} x$  dan elde edilen  $y$  nin tahminindeki tipik hatanın tayini. Burada da hesap, klâsik regresyon doğrusu için yapılan ve sahife 297 açıklandığı gibi yapılacaktır. Şöyle ki

$$\begin{aligned} \sum_{i=1}^n d_i^2 &= \sum_{i=1}^n (y_i - y_p)^2 = \sum_{i=1}^n \left( y_i - \frac{\sigma_y}{\sigma_x} x_i \right)^2 \\ &= \sum_{i=1}^n y_i^2 - 2 \frac{\sigma_y}{\sigma_x} \sum_{i=1}^n x_i y_i + \frac{\sigma_y^2}{\sigma_x^2} \sum_{i=1}^n x_i^2 \end{aligned}$$

olur. Diğer taraftan

$$\sum_{i=1}^n y_i^2 = N\sigma_y^2, \quad \sum_{i=1}^n x_i^2 = N\sigma_x^2, \quad \sum_{i=1}^n x_i y_i = \rho N\sigma_y\sigma_x$$

eşitlikleri gözönüne alınırsa, yukardaki toplam için

$$\begin{aligned} \sum_{i=1}^n d_i^2 &= N\sigma_y^2 - 2N\rho\sigma_y^2 + N\sigma_y^2 \\ &= 2N\sigma_y^2(1 - \rho) \end{aligned}$$

elde edilir. Buradan  $y$  nin tahminindeki tipik hatanın karesini  $T_y^2$  ile gösterirsek

$$T_y^2 = \frac{1}{N} \sum d_i^2 = 2\sigma_y^2(1 - \rho)$$

bulunur.

Bu ifade ile sahife 298 de bulunan  $S_y^2 = \sigma_y^2(1 - \rho^2)$  nin arasındaki bağıntıyı bulmak istersek, bu sonuncusu

$$S_y^2 = \sigma_y^2(1 - \rho^2) = \sigma_y^2(1 - \rho)(1 + \rho)$$

şeklinde yazılacağından

$$S_y^2 = \frac{1 + \rho}{2} T_y^2$$

eşitliği elde edilir. Buradan da

$$T_y^2 = \frac{2}{1 + \rho} S_y^2$$

$$\text{veya } T_y = \sqrt{\frac{2}{1 + \rho}} S_y$$

elde edilir.

Görülüyor ki yeni regresyon doğrusu kullanıldığı zaman  $y$  nin tahminindeki tipik hata genel regresyon doğrusu kullanıldığı zamanki tipik hatanın  $\sqrt{\frac{2}{1 + \rho}}$  katıdır.

Şu halde  $\rho$  korelasyon katsayısı 1'e yaklaştıkça tipik hatalar arasındaki fark da son derece azalmaktadır.

## T A T B İ K A T

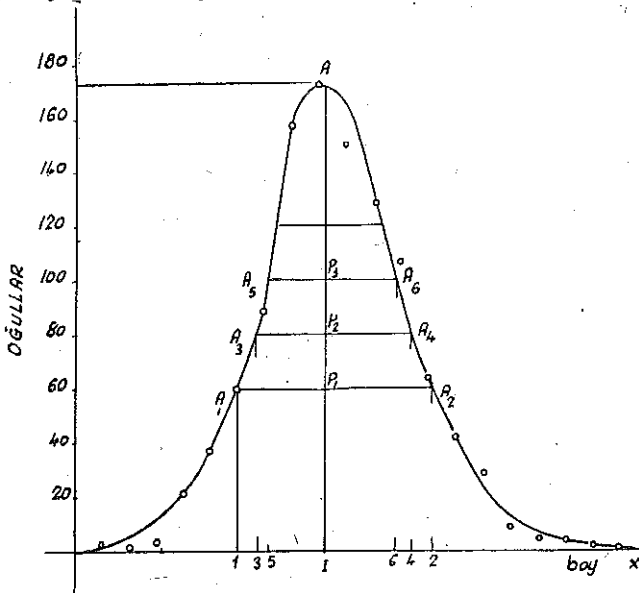
## § 1 — YENİ REGRESYON BAĞINTISININ BABA-OĞUL BOYLARI DAĞILIMINA TATBIKİ

Bir toplumda babaların ve oğulların boylarının dağılımını gösteren frekans tablosu sahife 308 de tablo I ile verilmiştir [10].

Bu frekans tablosu ile verilen dağılımın normal olup olmadığının tayini ile verilen bir çocuk boyuna mukabil en muhtemel baba boyunun, esasları birinci bölümde verilen metotla, nasıl tayin edilir ?

Frekans tablosu 1078 oğul, baba çiftinin boylarına göre dağılımını göstermektedir. Tablonun her gözündeki rakam, bu göze tekabül eden sıra ve sütunda bulunan boy uzunluklarındaki oğul, baba çiftinin sayısını vermektedir. Örnek olarak 16 sayısının bulunduğu bir gözü ele alalım : Bu göz oğul boylarının 68,5—69,5 inç aralığına, baba boyları da 66,5—67,5 aralığına tekabül etmektedir. Yâni 1078 oğul, baba çiftinin 16 sınıfın boyları bu aralıktadır.

Düşey toplam sütunundaki sayılar oğul boylarının frekansları olup ; bunlara göre boyların dağılım grafiği,  $f(X)$  in eğrisine tekabül eden, Şek. 1 de çizilmiştir. Bu grafikte inçle gösterilen boylar aralığı yatay ekseninde gösterilmiştir.



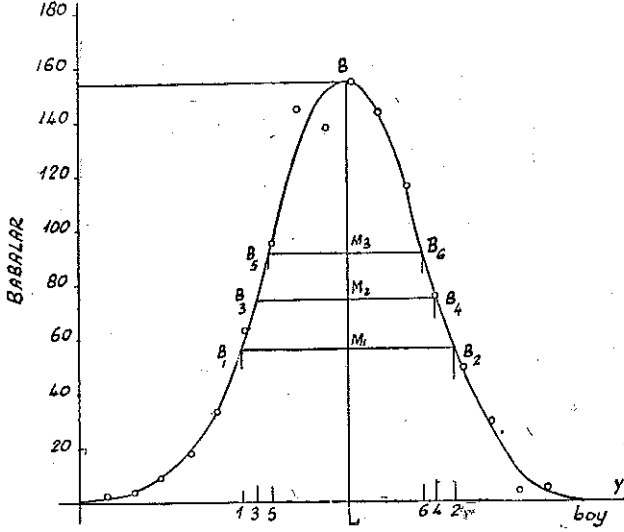
Şekil : 1

Tablo : I

		BABALARIN BOYLARI														TOPLAM f(X)				
		58,5	59,5	60,5	61,5	62,5	63,5	64,5	65,5	66,5	67,5	68,5	69,5	70,5	71,5		72,5	73,5	74,5	75,5
		59,5	60,5	61,5	62,5	63,5	64,5	65,5	66,5	67,5	68,5	69,5	70,5	71,5	72,5		73,5	74,5	75,5	
BOYLARI	59,5   60,5	-	-	-	0,5	0,5	1	-	-	-	-	-	-	-	-	-	-	-	2	
	60,5   61,5	-	-	-	0,5	-	-	-	1	-	-	-	-	-	-	-	-	-	1,5	
	61,5   62,5	-	0,25	0,25	-	0,5	1	0,25	0,25	0,5	0,5	-	-	-	-	-	-	-	3,5	
	62,5   63,5	-	0,75	0,25	2,25	2,25	2	4	5	2,75	1,25	-	0,25	0,25	-	-	-	-	20,5	
	63,5   64,5	1	-	1,5	3,75	3	4,25	9	9,25	3	1,25	1,5	0,75	1,25	-	-	-	-	38,5	
	64,5   65,5	2	1	0,5	2	3,25	9,5	13,5	10,75	7,5	5,5	3,5	2,5	-	-	-	-	-	61,5	
	65,5   66,5	-	0,5	1	2,25	5,25	9,5	10	16,75	17,5	16	5,25	2	2,5	1	-	-	-	89,5	
	66,5   67,5	-	1,5	2	4,75	3,5	13,75	19,75	26,5	25,75	19,5	12,5	13,75	3,25	0,5	1	-	-	148,5	
	67,5   68,5	-	-	1,5	2	7,5	10	10,25	24,25	31,5	23,5	29,5	13,25	8,5	9,5	2,25	-	-	173,5	
	68,5   69,5	-	-	1	-	5,25	5	12,75	18,25	16	24	29	21,5	10	3,5	2,25	-	1	149,5	
69,5   70,5	-	-	-	-	1	2,5	5,75	18,75	11,75	19,5	22,5	19,5	14,5	6,25	3,5	1,5	1	128		
BOYLARIN	70,5   71,5	-	-	-	-	-	3,25	5	9,75	10,25	19	14,75	20,75	10,75	8	5	1	1	108	
	71,5   72,5	-	-	-	-	-	0,25	3	1,25	7	7,75	10,75	11,25	10	8,5	2,75	0,5	-	63	
	72,5   73,5	-	-	-	-	-	-	0,75	0,75	2,5	7,5	6,5	6	7,5	6,25	3,25	0,5	0,5	42	
	73,5   74,5	-	-	-	-	1	-	1,5	1,5	-	5,25	2,25	2,5	6,5	3,25	3,25	-	2	29	
	74,5   75,5	-	-	-	-	-	-	-	-	-	1	2	-	2,5	0,75	1,75	0,5	-	8,5	
	75,5   76,5	-	-	-	-	-	-	-	-	-	-	1,25	0,25	-	0,5	1	1	-	4	
	76,5   77,5	-	-	-	-	-	-	-	-	-	-	1,25	0,25	1	-	-	1,5	-	4	
	77,5   78,5	-	-	-	-	-	-	-	-	-	-	-	1	1	-	0,25	0,75	-	3	
	78,5   79,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,25	0,25	-	0,5	
	TOPLAM h(Y)	3	3,5	5	17	33,5	61,5	95,5	142	137,5	154	141,5	116	78	49	28,5	4	5,5	1078	

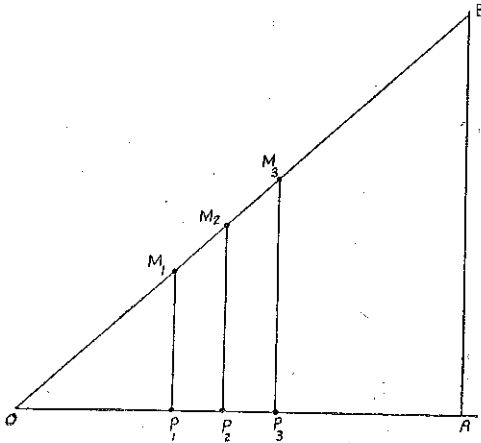
Yatay toplam satırındaki sayılar da baba boylarının frekansları olup bunlara göre boyların dağılım grafiği, bu da  $h(Y)$  nin eğrisine tekabül eder, Şek. 2 de çizilmiştir.

Bu iki grafikten bulunan  $IA=f(\bar{X})$  ve  $LB=f(\bar{Y})$  uzunlukları dik kenarlar olmak üzere bir dik OAB üçgeni çizelim (Şek. 3). Korelasyon üçgeni olarak isimlendirdiğim bu üçgen yardımı ile (36) ile gösterilen regresyon doğruları, aşağıda açıklandığı gibi çizilir:



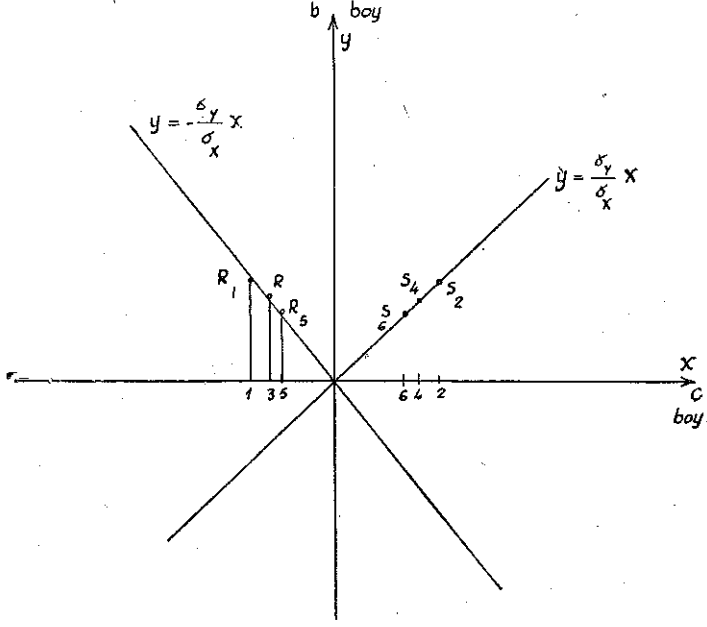
Şekil : 2

Dik OAB üçgeninin OA kenarı üzerinde alınan herhangi bir  $P_1$  noktasından OA ya çıkılan dikmenin OB yi kestiği nokta  $M_1$  olsun. Bu şekilde keyfi alınan her  $P_i$  noktası OA üzerinde hareket ettikçe  $M_i$  de OB üzerinde yer değiştirir. Bu metotla birbirine bağlı olarak bulunan  $OP_1$  ve  $P_1M_1$  uzunlukları sırası ile Şek. 1 ve



Şekil : 3

Şek. 2 de IA ve LB üzerinde alır ve bu noktalardan apsiler eksenine paraleller çizersek bu paraleller eğrileri sırasıyla  $A_1, A_2$  ve  $B_1, B_2$  noktalarında keser. Bu noktaların apsilerini, sırasıyla Şek. 1 ve Şek. 2 de apsiler için yeni başlangıç noktası olarak alacağımız, I ve L noktalarından itibaren alırsak ve  $A_i$  ( $i$  tek) lerin apsilerini



Şekil: 4

( $X_i$ ) apsis,  $B_i$  ( $i$  tek) lerin apsilerini ( $Y_i$ ) de ordinat olarak almakla gösterilen nokta  $R_i$  olsun (Şek. 4). Aynı şekilde  $A_i$  ( $i$  çift) ve  $B_i$  ( $i$  çift) noktalarına karşılık gelen noktalar da  $S_i$  ile gösterilsin.

Bu şekildeki çizimle elde edilen  $R_i$  ve  $S_i$  noktalarının geometrik yerlerinin birer doğru olacağını (36) denklemlerinden çıkartmıştık. Şekil 4 de bu metotla bulunan  $R_1, R_3, R_5$  noktalarının bir doğru üzerinde olduğu görülmektedir. Bu doğruyu bundan böyle R harfi ile göstereceğiz. Aynı şekilde  $S_2, S_4, S_6$  noktalarının da bir doğru üzerinde olduğu görülmekte olup, bu doğruyu da S ile göstereceğiz.

R ve S doğruları koordinat eksenlerine göre simetrik iki doğrudur.

Bu doğrulardan biri yardımıyla bulunan, belli bir  $x$  değerine karşılık gelen en muhtemel  $y$  değeri, diğer doğrudan bulunan değere eşit, ancak işaretçe farklıdır.

Şimdi gerek yukarıda açıkladığımız metotla çizilen S doğrusu, gerekse denklemleri (20) ve (21) de verilen regresyon doğruları yardımıyla, belli bir baba boyuna karşılık gelen en muhtemel oğul boyunu ve karşıt olarak belli bir oğul boyuna tekabül eden en muhtemel baba boyunu bulalım.

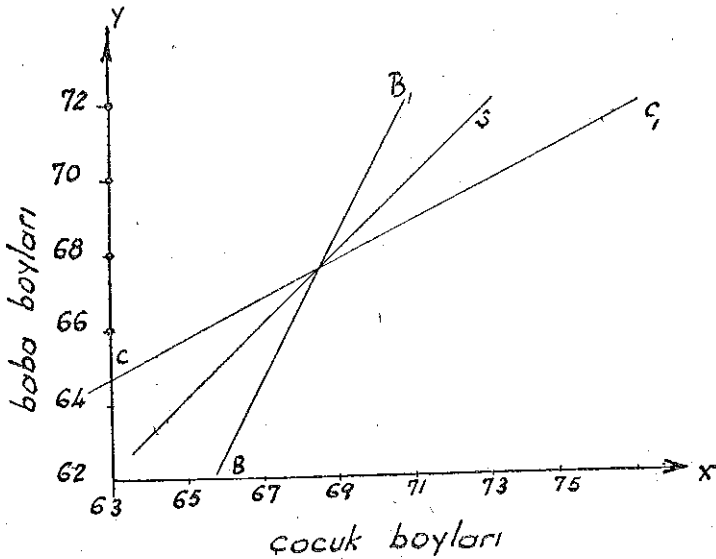
Bu üç doğru Şek. 5 de çizilmiştir.  $BB_1$  doğrusu bir oğul boyuna karşılık gelen en muhtemel baba boyunu veren ve denklemini

$$Y - \bar{Y} = \rho \frac{\sigma_y}{\sigma_x} (X - \bar{X})$$

olan doğrudur.  $CC_1$  doğrusu ise belli bir baba boyuna karşılık gelen en muhtemel oğul boyunu veren ve denklemini

$$X - \bar{X} = \rho \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})$$

olan doğrudur. S doğrusu ise,  $\sigma_x$ ,  $\sigma_y$  ve  $\rho$  sabitlerinin hesaplanmasını gerektirmeden, yukarıda açıklandığı gibi tayin edilen yeni bir regresyon doğrusudur.



Şekil : 5

Bu şekilden, baba boyu 69 inç ise en muhtemel oğul boyunun ne olacağını bulalım.  $BB_1$  doğrusu yardımı ile oğul boyu olarak 69,2 inç,  $S$  doğrusu yardımı ile de 69,8 inç bulunur.

İkinci olarak, oğul boyu 67 inç iken en muhtemel baba boyunu bulalım.  $CC_1$  doğrusu yardımı ile baba boyu olarak 65 inç,  $S$  doğrusu yardımı ile de 65,5 inç bulunur.

Bu metotla çizilen  $S$  veya  $R$  doğrusu, istatistik değişkenlerin birbirine olan tekabülünü bir türlü vermektedir. Bununla beraber tekabüldeki meyil, yâni baba boyuna tekabül eden oğul boyunun babanın boyundan büyük olacağı çıkarılmaktadır.

Korelasyon katsayısı 1 e yakın oldukça  $BB_1$  ve  $CC_1$  regresyon doğruları birbirine çok yakın olacağından bunların yerine bir tek  $S$  doğrusunu koymak en elverişli çözümdür.

Gerçekten korelasyon katsayısının yüksek olması hallerinde regresyon denklemleri kullanılarak hesaplanan sonuçlar arasındaki farklar şüphesiz istatistik dalgalanmalardan daha küçüktür.

## § 2 — YENİ REGRESYON BAĞINTISININ FİSYON OLAYINA TATBİKİ

Bilindiği gibi, bir atom çekirdeğinin içinde  $Z$  tane proton ve  $N$  tane de nötron mevcuttur. Bu ikisine birden *nükleon* denir.

Bir çekirdekte nükleon sayısını  $A$  ile gösterirsek

$$A = Z + N$$

eşitliği vardır. Bu  $A$  sayısı çekirdeğin kütleini verir.

Bir uranyum çekirdeğinin  ${}_{92}U^{235}$  şeklindeki gösterilişinde üst indis  $A$  nın, alt indis  $Z$  nin değerini göstermektedir.

Uranyum çekirdeği nötron bombardımanı ile çeşitli fisyon ürünleri meydana getirir.

Fisyon ürünlerinin verimleri, yani meydana gelme ihtimalleri aşağıdaki şekilde iki gruba ayrılır :

1 — İzobar verim: Bu ad altında  $A$  sı aynı olan ürünler serisinden bahsedilmektedir.

2 — İzotopik verim: Bu ad altında  $Z$  si aynı olan ürünlerin serisinden bahsedilmektedir.

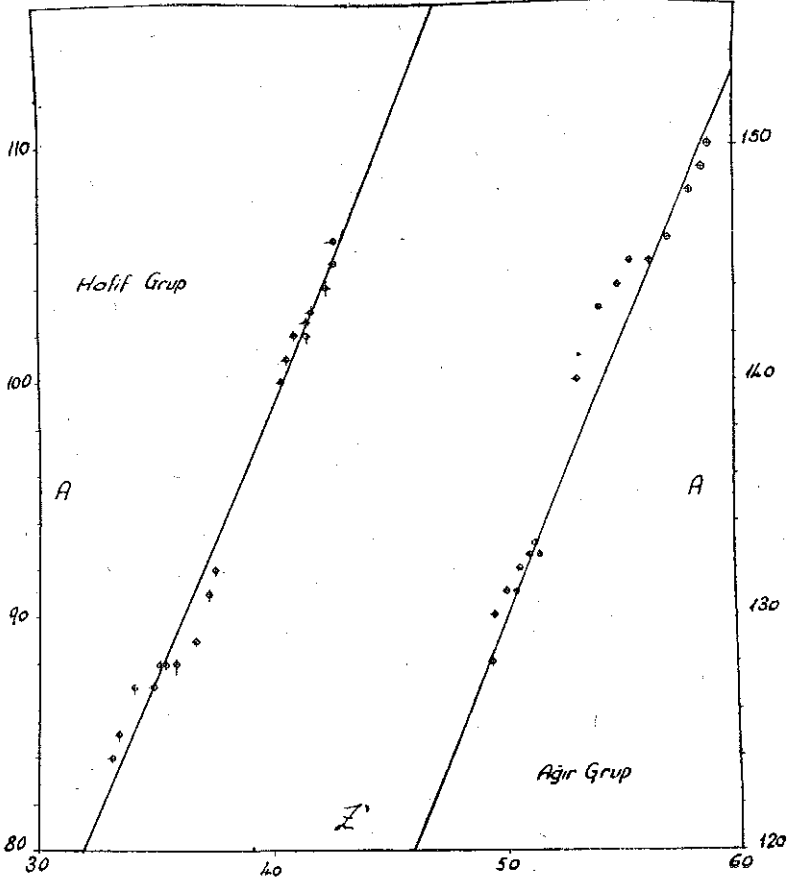


İzobar ve izotopik verimlerin dağılımları incelenirse, bunların da ağır ve hafif olmak üzere iki grupta toplandığı görülür.

İzobar verim eğrileri deneylerin sonuçlarına göre çizilmiştir<sup>[11]</sup>. Böylece gerek hafif grubun, gerekse ağır grubun dağılım eğrilerinin bir normal dağılım eğrisine yakın olduğu bulunmuştur.

Aynı şekilde izotopik verim eğrileri de çizilmiş ve bunların da hafif ve ağır gruplar için dağılımlarının gene bir normal dağılıma çok yakın olduğu görülmüştür<sup>[12]</sup>.

Bu eğrilerin mükemmel normal dağılım eğrisi olmayışı bazı fisyon ürünlerinin teşekkül meyillerinin fazla olmasından ve tesadüf (aléatoire) değişkenliği biraz bozmasından ileri gelmektedir.



Şekil : 6

Şimdi izobar ve izotopik verim eğrilerinden faydalanarak iki A ve Z değişkeni arasındaki regresyon doğrusunu arayalım. Bunun için de birinci paragraftaki metot gereğince izobar ve izotopik verim eğrileri yardımıyla korelasyon üçgeni çizilir ve sonra S ve R doğrularını belirten noktaların çizimine geçilir.

Şekil 6 daki ilk doğru hafif gruplar için ( $\bar{Z}=38$ ,  $\bar{A}=95$ ) noktasını başlangıç noktası alan eksen takımına göre çizilen S doğrusuna karşılık gelen doğru ile, bununla aynı doğrultuda olması gereken, R nin yatay eksene göre simetriğini çizerek elde edilen doğruyu göstermektedir. İkinci doğru ise ağır gruplar için ( $\bar{Z}=54$ ,  $\bar{A}=139$ ) noktasından çizilen, hafif grupta olduğu gibi, doğrusudur. Ancak bu sonuncusunu aynı grafikte göstermek için, A ordinatları, 120 den başlatarak sağ tarafta çizilmiştir.

Şu halde iki değişkenli A, Z serisinin ihtimal yüzeyi iki tepeli bir Gauss yüzeyidir.

### § 3 — HERHANGİ BİR İKİ DEĞİŞKENLİ FREKANS TABLOSUNUN NORMAL DAĞILIM OLUP OLMADIĞININ TAYİNİ

Bölüm I in 3. paragrafında bulduğumuz kural, iki değişkenli frekans tablosunun Gauss yüzeyi gösterdiği halde, değişkenlerin toplam dağılım fonksiyonları olan  $f(X)$  ve  $h(Y)$  nin grafiklerinden faydalanarak çizilen R ve S doğrularının eksene göre simetrik olduğu gösterilmiş ve bu şartı sağlayan iki örnek te önceki paragraflarda verilmiştir.

Şimdi herhangi bir, iki değişkenli frekans tablosunu ele alacağız, bunun bir Gauss yüzeyi göstermeyeceğini, toplam dağılım fonksiyonları yardımı ile çizilen R ve S doğrularının eksene göre simetrik olmamasını görerek, tesbit edeceğiz.

Örnek olarak aynı yaşta ve aynı cinsiyette olan 1000 şahsı gözönüne alalım. Bu şahısları kilo ve boylarına göre gruplara ayırarak bir frekans tablosu tesis edilmiş olsun [13]. Tablo II.

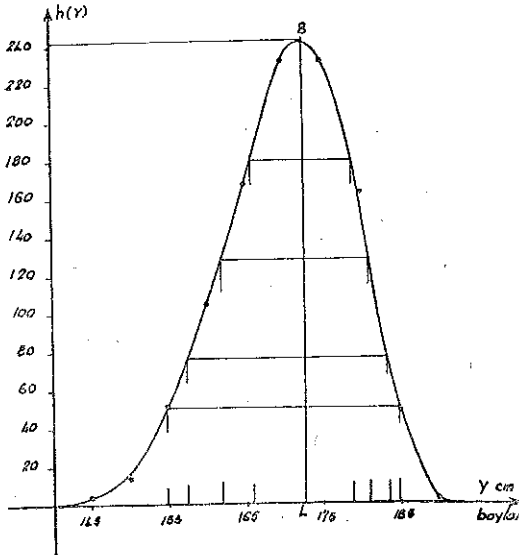
Bu tablo ağırlıkları 3 er kg aralıkla 13 sınıfa ve boyları da 5 er cm aralıkla 10 sınıfa ayırarak elde edilmiştir. Ağırlığı gösteren X sütununda yazılan sayılar aralığın merkez değerini, boyu gösteren Y satırındaki sayılar da aralığın merkez değerini göstermektedir. Merkez değerleri  $X=67$  ve  $Y=170$  olan bir sınıftan, ağır-

Tablo : II

## Boylar

	145	150	155	160	165	170	175	180	185	190	$\Sigma X$
46			1	2	1						4
49	1	2	2	2	1	1	1				10
52	1	4	4	12	13	10	6	1			5
55	2	2	13	15	18	19	8	1			78
58	2	1	7	24	18	21	16	7			96
61		2	8	22	26	43	37	8	1		147
64		2	6	11	30	33	31	15	3		131
67		1	7	6	34	36	60	29	6		179
70			3	7	16	27	38	27	15		133
73			1	3	6	23	19	24	9	1	86
76				1	3	13	11	22	9	1	60
79					1	3	4	8	6		22
82						2	1				3
$\Sigma Y$	6	14	52	105	167	231	232	142	49	2	1000

lıkları 55,5 ile 58,5 kg arasında ve boyları ise 167,5 ile 172,5 cm arasında olan şahıslar anlaşılır. Bunların sayısı 36 dır.

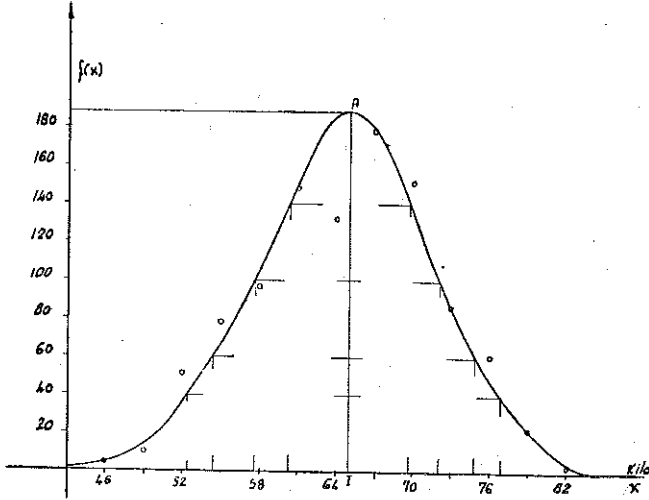


Şekil : 7

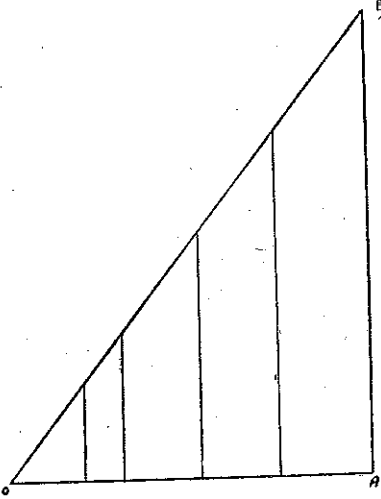
Gözönüne alınan 1000 şahsın yalnız kilolarına ve boylarına göre dağılımları  $x$  ve  $y$  sütun ve satırında gösterilmiştir.

Şimdi,  $f(X)$  e tekabül eden  $\Sigma x$  ve  $h(Y)$  ye tekabül eden  $\Sigma y$  lere ait eğrileri çizerek ve bunlardan faydalanarak R ve S doğrularını çizelim. Sonra da bu doğruların eksenlere göre simetrik olup olmadıklarına bakalım.

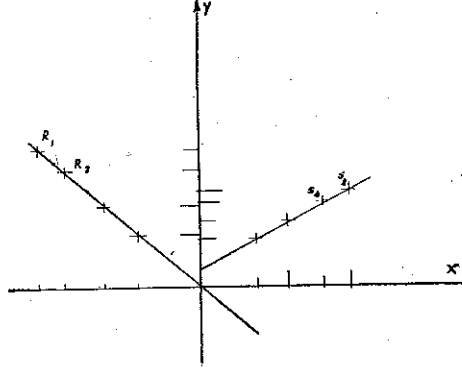
Bunun için de  $f(X)$  ve  $h(Y)$  nin eğrileri Şek. 7 ve Şek. 8 de çizilmiştir. Dik kenarları Şek. 7 ve Şek. 8 den bulunan LB ve LA maksimum değerlerine eşit olan OAB korelasyon üçgeni çizilmiştir (Şek.9). Bu üç şekilden faydalanarak R ve S doğruları çizilir (Şek. 10). Bu son şekilden R ve S doğrularının eksenlere göre simetrik olmadığı görülmektedir. O halde, gözönüne aldığımız frekans tablosu iki değişkenli normal dağılım göstermemektedir.



Şekil : 8



Şekil : 9



Şekil : 10

## Bölüm III

## NORMALLİK TESTİ

Bölüm I in 2. paragrafında bulduğumuz kural, her hangi bir tek değişkenli dağılımın normal dağılım olup olmadığını anlamak için de uygulanabilirse pratikliği ve çabuk netice vermesi bakımından diğer testlere tercih edilebilir.

Bu bölümde adı geçen uygulamanın mümkün olacağı gösterilmiş ve bir örnek ile de sağlama yapılmıştır.

Bundan başka diğer bir örnekte de bu tezdeki normallik testi ile ihtimal kâğıdı kullanarak yapılan test ve  $\chi^2$  (hi kare) testi arasında mukayese yapılmıştır.

§ 1 — BULUNAN METODUN TEK DEĞİŞKENLİ HERHANGİ BİR DAĞILIMININ NORMALLİĞİNİ BELİRTEN TEST OLARAK KULLANILMASI

Normal dağılım gösterdiğini bildiğimiz fonksiyon  $u=f(x)$  olsun. Kontrola tabi tuttuğumuz dağılımın fonksiyonu da  $v=h(y)$  olsun.

Her iki fonksiyondan faydalanarak, II. Bölümün birinci paragrafında açıklanan çizim yapılabilir. Bu halde elde edilen R ve S doğruları başlangıç noktasından geçen ve eksenlere göre simetrik olan iki doğru ise; ikinci  $h(y)$  fonksiyonu da bir normal dağılım gösteren fonksiyondur.

Gerçekten  $f(x)$  in uyguladığı şartlar  $x = \bar{x}$  için  $f(\bar{x})$  maksimum ve  $x < \bar{x}$  için monoton artan,  $\bar{x} < x$  için de monoton eksilendir. Aynı şekilde  $h(y)$  fonksiyonu da  $y = \bar{y}$  için  $h(\bar{y})$  maksimum ve  $y < \bar{y}$  için monoton artan,  $\bar{y} < y$  için de monoton eksilendir.

Bunlardan başka

$$(1) \quad \frac{h(y)}{f(x)} = \frac{h(\bar{y})}{f(\bar{x})} = \frac{a}{b} = k \text{ (sabit)}$$

olmak üzere alınan  $f(x)$  ve  $h(y)$  fonksiyonlarının  $x$  ve  $y$  ile gösterilen argümanları yardımıyla tayin edilen  $A(x; y)$  şeklindeki noktaların bir doğru üzerinde olduğu bilinmektedir. Diğer bir deyimle argümanlar arasında

$$y = mx + n$$

şeklinde lineer bir bağ vardır.

Şu halde  $f(x)$  ve  $h(y)$  fonksiyonları arasındaki (1) eşitliği şu şekilde yazılabilir:

$$(2) \quad h(mx + n) = kf(x)$$

Eğer  $f(x)$  ve  $h(y)$  fonksiyonlarının eğrilerini gösteren eksen rakamlarında başlangıç noktasını  $x = \bar{x}$  ve  $y = \bar{y}$  noktasına taşırsak; fonksiyonların sağladığı şartlar aşağıdaki şekilde ifade edilebilir:

$$(3) \quad \begin{array}{l} x = 0 \text{ için } f(0) = a \text{ ve } f'(0) = 0 \\ y = 0 \text{ için } h(0) = b \text{ ve } h'(0) = 0 \end{array}$$

Diğer taraftan (2) eşitliğinden türev alarak

$$mh'(mx + n) = kf'(x)$$

bulunur. Bu eşitlikten  $x = 0$  için

$$mh'(n) = kf'(0)$$

bulunur. (3) şartlarını göz önüne alırsak

$$(4) \quad mh'(n) = 0$$

elde edilir. Halbuki gene (3) eşitlikleri  $h(y)$  fonksiyonun türevinin  $y = 0$  için sıfır olacağını göstermektedir. O halde (4) eşitliğinden  $n = 0$  olacağı çıkartılır.

Bu duruma göre  $f(x)$  ve  $h(y)$  fonksiyonları arasındaki (1) eşitliği

$$h(mx) = k f(x)$$

şeklini alır.

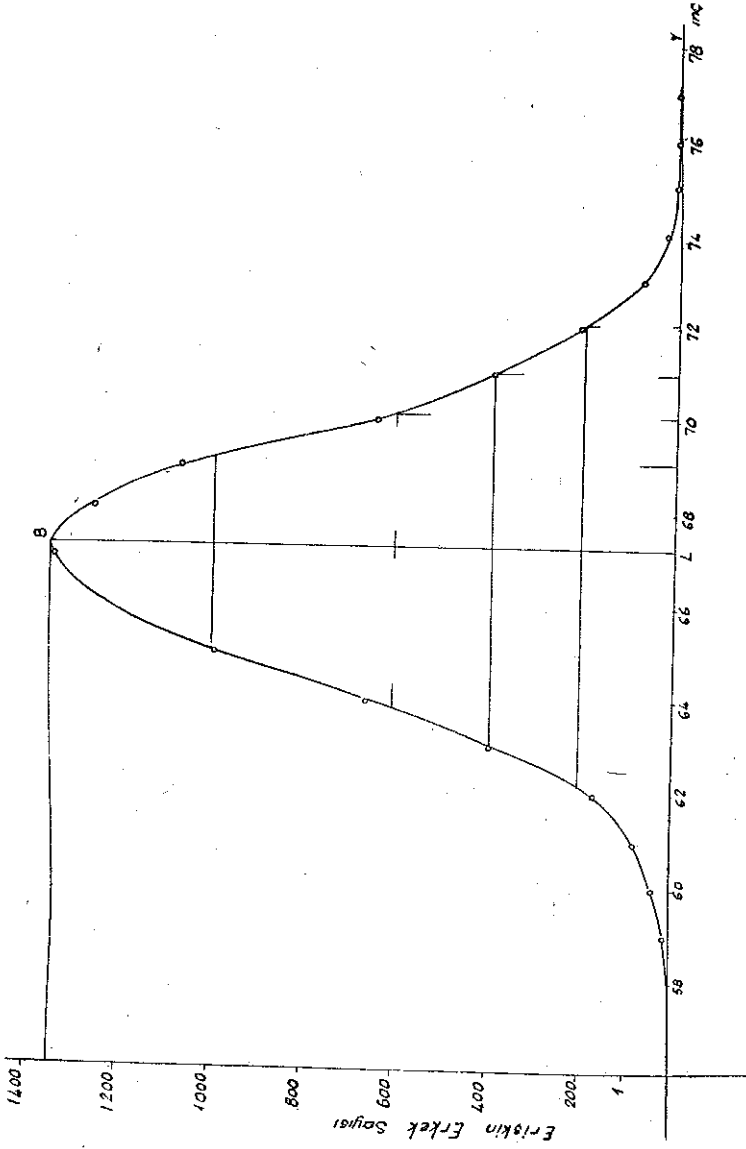
Şu halde  $f(x)$  fonksiyonu (3) şartlarına uyan bir normal dağılım ise,  $h(y)$  fonksiyonu da aynı şekilde bir normal dağılım gösteren fonksiyon olur.

## § 2 — MUHTELİF DAĞILIMLARA TATBİK VE BAZI TESTLERLE MUKAYESE

Örnek 1. İskoçya, Wales, İrlanda ve İngilterede doğan erişkin erkeklerin boylarının dağılımı tablo III olduğu gibi bulunmuştur<sup>[14]</sup>.

Tablo : III

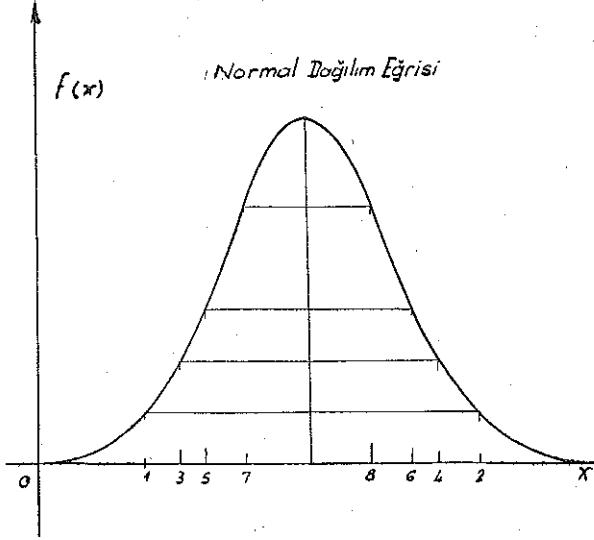
<i>Boylar (ayakkabısız) inç</i>	<i>frekans</i>
57	2
58	4
59	14
60	41
61	83
62	169
63	394
64	669
65	990
66	1223
67	1329
68	1230
69	1063
70	646
71	392
72	203
73	79
74	32
75	16
76	5
77	2



Şekil : 11

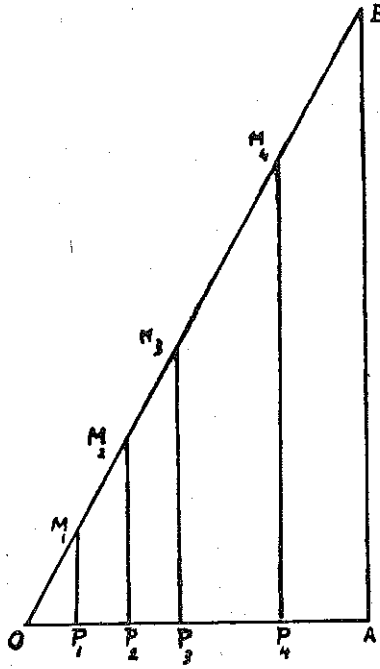


Bu tabloya göre 8585 erişkin erkeğin boylarının dağılım grafiği Şek. 11 de çizilmiştir. Şek. 12 ise her hangi bir normal dağılımı gösteren  $f(x)$  fonksiyonunun eğrisini göstermektedir.



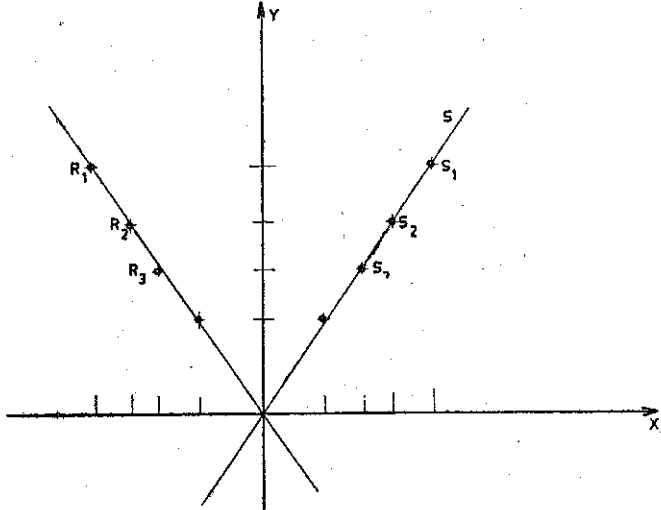
Şekil : 12

Bu ikisinden faydalanarak önce korelasyon üçgeni durumunda olan OAB dik üçgeni çizilir (Şek. 13). Sonra çizimleri I. inci paragrafda açıklandığı gibi  $R_i$  ve  $S_i$  noktaları çizilir (Şek. 14).



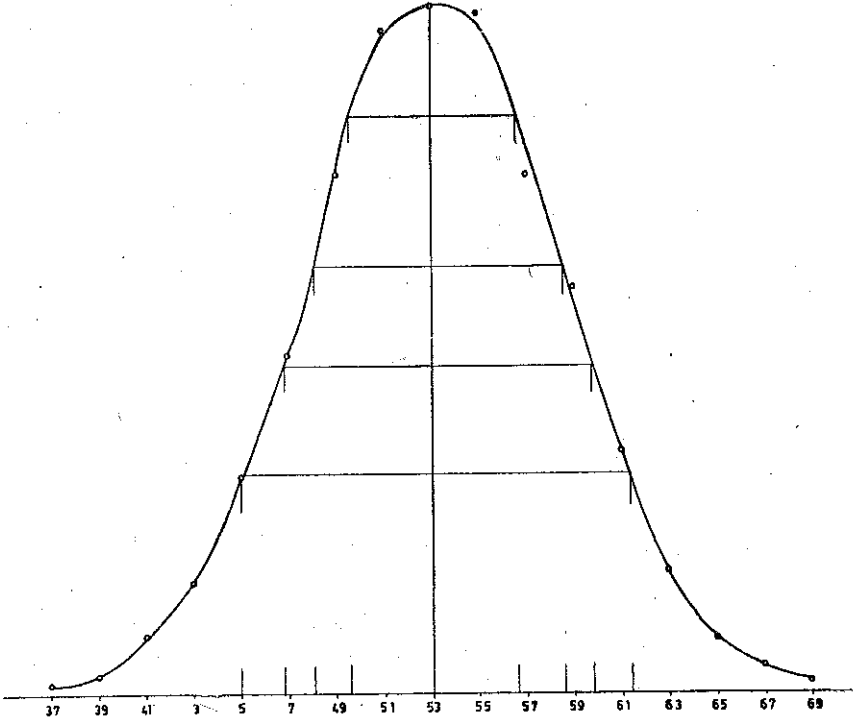
Şekil : 13

Son şekil gösteriyor ki,  $R_i$  ve  $S_i$  noktalarının geometrik yeri birer doğrudur ve bu doğrular eksenlere simetriktir. O halde boy-lar dağılımı normal dağılımdır.



Şekil : 14

Örnek 2. Türkiye’de 20273 kadının boya göre dağılımını gösteren Tablo IV deki seriyi ele alalım [15]:

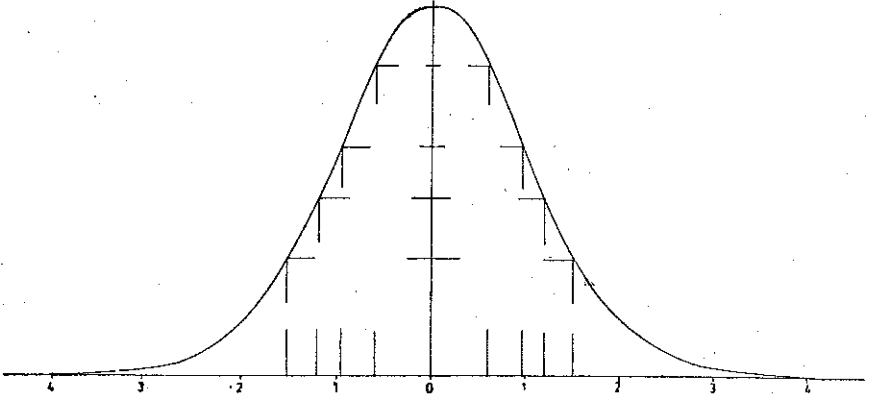


Şekil : 15

Bu tabloya göre 20263 kadının boylarının dağılım grafiği Şek. 15 de gösterilmiştir. Şek. 16 ise normal dağılımı gösteren  $f(x)$  fonksiyonunun eğrisini göstermektedir.

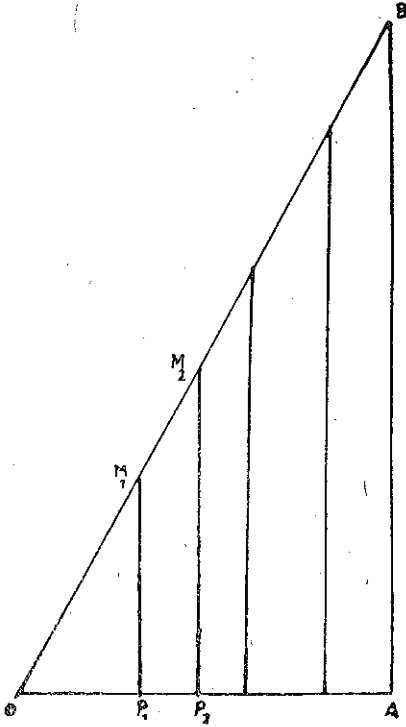
Tablo : IV

Boylar	N <sub>i</sub>
137 den az	57
137 — 139 dan az	85
139 — 141 »	262
141 — 143 »	477
143 — 145 »	928
145 — 147 »	1499
147 — 149 »	2218
149 — 151 »	2834
151 — 153 »	2929
153 — 155 »	2901
155 — 157 »	2214
157 — 159 »	1728
159 — 161 »	1047
161 — 163 »	579
163 — 165 »	263
165 — 167 »	153
167 — 169 »	59
169 ve +	31
	20263

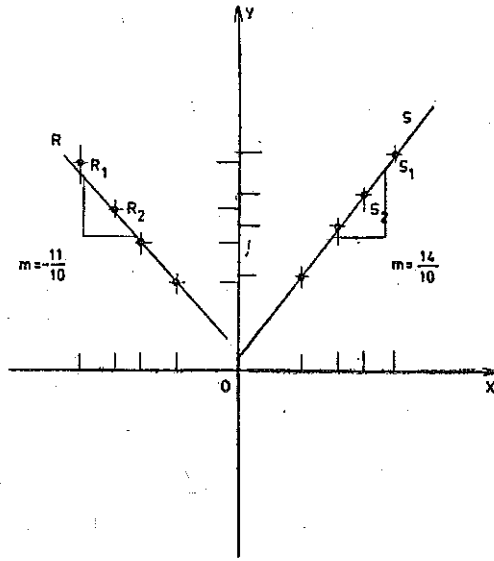


Şekil : 16

Bu ikisinden faydalanarak önce korelasyon üçgeni durumunda olan OAB dik üçgeni çizilir (Şek. 17). Sonra I. paragrafta açıklandığı gibi  $R_i$  ve  $S_i$  noktaları çizilir (Şek. 18). Bu son şeklin incelenmesinden şu neticeler çıkarılır :



Şekil : 17



Şekil : 18

a —  $S_i$  noktaları bir doğruya yaklaşmakta, ancak bu doğru başlangıçtan geçmemektedir.

b —  $S_i$  noktaları için yukarıda söylenen  $R_i$  noktaları için de mevcuttur.

c — Eğimleri işaretçe zıt, değerce eşit olması gereken S ve R doğrularının eğimleri farklı değerdedir. Bu eğimler şekilden ölçülerek; S doğrusu için  $m_S = \frac{14}{10}$ , R doğrusu için  $m_R = -\frac{11}{10}$  bulunur.

Bu neticelere göre kadın boylarının dağılımının normal dağılım olması şüphelidir. Bu bakımdan daha uzun hesaplar isteyen, fakat belli ihtimalle verilen serinin normal olup olmadığını tayin edebileceğimiz  $\chi^2$  testini tatbik edelim [16].

Bu test için serinin normale uydurulmuş halinin teorik mutlak frekanslar tablosunu hazırlamamız gerektir. Teorik mutlak frekanslar ve gerçek frekanslar ile test için gerekli sayılar tablo V te gösterilmiştir. Tablodaki teorik mutlak frekanslar Ömer Celâl Sarc'ın İstatistik II adlı eserinden alınmıştır [17].

Tablo : V

Teorik mutlak frekanslar	Gerçek frekanslar			$(N_i - m_i)^2$
$m_i$	$N_i$	$ N_i - m_i $	$(N_i - m_i)^2$	$m_i$
52	57	5	25	0,480
101	85	16	256	2,535
246	262	16	256	1,041
504	477	27	729	1,446
956	928	28	784	0,820
1556	1499	57	3249	2,088
2142	2218	76	5776	2,697
2731	2834	103	10609	3,885
2971	2929	42	1764	0,594
2752	2901	149	22201	8,067
2358	2214	144	20736	8,794
1715	1728	13	169	0,099
1068	1047	21	441	0,413
616	579	37	1369	2,222
300	263	37	1369	4,563
127	153	26	676	5,323
46	58	12	144	3,130
22	31	9	81	3,681
20263	20263			51,878

Burada frekans sayısı 18 olup serbestlik derecesi

$$r = 18 - 2 - 1 = 15$$

olur. Tabloya göre [18]

$$P[\chi^2 > 24,996] < \% 5 \text{ dir.}$$

O halde  $\chi^2$  için yukarda bulunan 51,878 değeri bu değerden büyük olduğundan dağılımın normal olduğunu kabul hatalıdır. Teklif edilen testde de aynı netice elde edilmişti.

Şimdi de ihtimal kâğıdı kullanarak yapılan normallik testini aynı örneğe uygulayalım [19].

İhtimal kâğıdındaki koordinat eksenlerinden biri aritmetik di-  
ziye, diğeri ise  $\int_{-\infty}^t e^{-\frac{t^2}{2}} dt$  değerlerine göre taksimatlanmıştır.  
Yâni OX eksenini aritmetik dizi ölçeğine göre, OY eksenini de ihtimal ölçeğine göre taksimatlanmıştır. Sınıfların  $X_i$  sınır değerleri apsislere ve bunlara tekabül eden kümüle edilmiş frekansların yüzde oranları olan  $\frac{cn_i \times 100}{N}$  değerleri de ordinatlara taşınmıştır.

Koordinatları bu şekilde işaret edilen noktalar bir doğruya yaklaştığı nisbette gözönüne alınan seri normal bölünmeye uymaktadır.

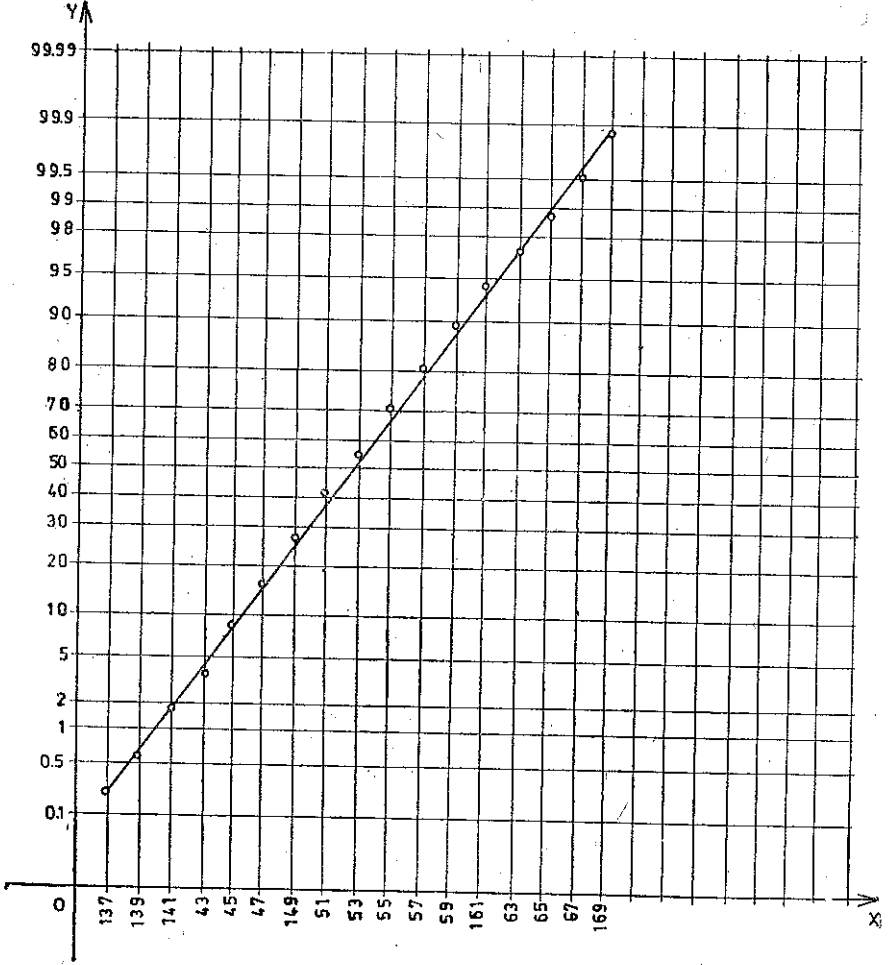
Tablo VI da birinci sütununa serinin sınıflarının  $X_i$  sınır değerleri, ikinci sütununa  $n_i$  frekanslarını, üçüncü sütununa  $cn_i$  kümüle edilmiş frekansların yüzde oranları olan  $\frac{cn_i \times 100}{N}$  değerleri yazılmıştır.

Tablo : VI

$x_i$	$n_i$	$cn_i$	$\frac{cn_i \times 100}{N}$
137	57	57	0,281
139	85	142	0,701
141	262	404	1,994
143	477	881	4,384
145	928	1809	8,927
147	1499	3308	16,325
149	2218	5526	27,271
151	2834	8360	41,257
153	2929	11289	55,712
155	2901	14190	70,029
157	2214	16404	80,955
159	1728	18132	89,483
161	1047	19179	94,650
163	579	19758	97,508
165	263	20021	98,815
167	153	20174	99,561
169	58	20232	99,847
	31	20263	

N = 20263

Tablonun birinci ve dördüncü sütunlarındaki değerler yardımı ile ihtimal kâğıdında noktalar çizilmiştir (şek. 19).



Şekil : 19

Grafiğin incelenmesinden kadın boylarının dağılımını gösteren serinin normal dağılıma uygun olduğu kanaati hakim olmaktadır. Diğer taraftan  $\chi^2$  testi aynı serinin normal dağılım olamayacağını işaret etmekte idi. Bu tezde teklif edilen normallik testi de aynı serinin normal dağılım olmasının şüpheli olduğuna işaret etmekte idi. Şu halde teklif edilen test yukarıda bahsedilen ihtimal kâğıdı testinden daha fazla hakikate uymaktadır. Bundan başka teklif edilen test yukarıdaki ihtimal kâğıdına çizilen grafik teste nazaran daha az zahmetlidir ve daha çabuk neticeye varmaktadır.



## BİBLİYOGRAFYA

- [1] *Ömer Celdil Sarc* : İstatistik, II, İstanbul Üniversitesi Yayınlarından 1962, sah. 168.
- [2] *Haydar Furgaç* : İstatistik Usulleri, I Rölöveler, İstanbul Üniversitesi Yayınlarından 1960, sah. 85.
- [3] *A. A. Tschuprow* : Principles of the Mathematical Theory of Correlation, William Hodge and Company, London 1939, sah. 45; *R. Risser* et *R. Traynard* : Les Principes de la Statistique Mathématique, II, Gauthier - Villars, Paris 1959, sah. 102.
- [4] *Victor Rouquet la Garrigue* : Les Problèmes de la Corrélation et L'Elasticité II, Hermann et Cie, Paris 1948, sah. 46; *E. Borel - R. Dettneil - R. Huron* : Probabilités Erreurs, Collection Armand Colin, Paris 1962, sah. 185; *M. G. Kendall* : The Advanced Theory of Statistics, Charles Griffin & Company limited, London 1952, sah. 328.
- [4] *A. A. Tschuprow*, [3] de a.g.k. sah. 46.
- [6] *E. Morice* et *F. Chartier* : Methode Statistique II, Imprimerie Nationale, Paris 1954, sah. 350; *Van Der Waerden* : Mathematische Statistik, Springer, Berlin 1957, sah. 297; *M. G. Kendall*, [4] de a.g.k. sah. 330; *A. Monjallon* : Eléments de Statistique Mathématique, Librairie Vuibert, Paris 1963, sah. 79; *Lancelot Hogben* : Zahl und Zufall, Verlag R. Oldenbourg, München 1956, sah. 425.
- [7] *A. Tortrat* : Calcul des Probabilités, Masson et Cie, Paris 1963, sah. 111; *E. Morice* et *F. Chartier*, [6] da a.g.k. say. 363; *Emanuel Parzen* : Modern Probability Theory and its Applications, John Wiley, New York 1960, sah. 325; *A. Renyi* : Wahrscheinlichkeitsrechnung, Veb Deutscher Verlag der Wissenschaften, Berlin 1962, sah. 156.
- [8] *J. Bass* : Eléments de Calcul des Probabilités, Masson et Cie, Paris 1963, sah. 83; *A. Blanc Lapierre* et *Robert Fortet* : Théorie des Fonctions Aléatoires, Masson et Cie, Paris 1953, sah. 60.
- [9] *A. Tortrat*, [7] de a.g.k. sah. 10; *A. Renyi*, [7] de a.g.k. sah. 45; *B. W. Gnedenko* : Lehrbuch der Wahrscheinlichkeitsrechnung, Akademi - Verlag, Berlin 1962, sah. 48.
- [10] *H. Yule* and *M. G. Kendall* : An Introduction to the Theory of Statistics, London 1940, sah. 199.
- [11] *E. P. Steinberg* ve *L. E. Glenderin* : Proceedings, International Conference on the Peacefull Uses of Atomic Energy, Geneva, 1955.
- [12] *M. Talât Erben* ve *Bınay Güven* : Physical Review, 129, 1762 (1963).
- [13] *A. Monjallon* : Introduction à la Methode Statistique, Librairie Vuibert, Paris 1961, sah. 164.
- [14] *G. U. Yule* and *M. G. Kendall*, [10] da a.g.k. sah. 94.
- [15] *Ömer Celdil Sarc*, [1] de a.g.k. sah. 213.

[16]  $\chi^2$  (hi kare) testi için: *R. Rissler et R. Traynard* [3] de a.g.k. I inci, sah. 49; *R. Deltheil et R. Huron*: Statistique Mathématique, Collection Armand Colin, Paris 1959, sah. 21; *R. A. Fisher*: Les Méthodes Statistiques, Presses Universitaires de France, Paris 1946, sah. 53. *A. Monjallon*, [9] da a.g.k. sah. 166; *F. C. Mills*: Statistical Methods, Henry Holt and Company, New York 1955, sah. 512.

[17] *Ömer Celâl Sarc*, [1] de a.g.k. sah. 213.

[18] *J. Bass*, [8] de a.g.k. sah. 191 deki tablodan *F. C. Mills*, [16] da a.g.k. sah. 773 deki tablodan.

[19] İntimal kâğıdı kullanılarak yapılan örnekler için: *A. Monjallon* [13] de a.g.k. sah. 138; *Gerd Bangen und Richard Stender*: Wahrscheinlichkeitsrechnung und Mathematische Statistik, Otto Salle Verlag, Hamburg 1961, sah. 47.