A HISTORICAL AND MATHEMATICAL DESCRIPTION OF THE NORMAL PROBABILITY FUNCTION

A Thesis

Presented to

the Faculty of Graduate Studies

University of Manitoba

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Brian Gusztav Nemes

May 1965

ABSTRACT

This treatise explores the origins of the normal frequency function, integrates the history, proofs, properties, uses, and interrelationships of the findings into present day normal probability theory.

Although Laplace and Gauss have been associated with the normal probability function's origin, it was Abraham de Moivre who, in 1733, first deduced it as the limit of the Binomial distribution. Many authors have commented on the aspects of the methods of the "proof" of this function and conclude that it has been accepted as a foundation of statistical methods. The form of the normal frequency function is $ce^{-h^2x^2}$, where $h^2 = \frac{1}{2\sigma^2}$, and $c = \frac{1}{\sqrt{2\pi}\sigma}$. The properties of this function are:

1. It is symmetrical about the axis $x = \mu$.
2. The x-axis must be an asymptote.
3. The area under the curve is unity.
4. The mean is μ; and the variance is σ^2.
5. The skewness is 0, and the kurtosis is 3.
6. The first two cumulants completely determine the distribution.
7. The sum of independent normal variates is also a normal variate.

In statistical inference, a necessary assumption for tests of significance is that the errors of observation be a sample from a normally distributed population. When the assumption of normality is satisfied, then normal theory may be applied to the solution of the problem. The distributions which are the basis for the tests satisfy the assumption of normality in the limiting sense as the sample size becomes infinitely large.

Data are tested to determine whether or not they are a sample from a normally distributed population by two methods. The first is a method of areas procedure tested by the Chi-Square goodness of fit test. The second, a method of testing the skewness and kurtosis of the sample data by means of a normal deviate test. The problem of non-normal data may be solved through the choice of a transformation. If this fails, then non-parametric methods may be applied to make some sense of the data.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION AND HISTORICAL DEVELOPMENT

Introduction

In the whole of statistics, the most important continuous function is the normal probability function. Until the recent advent of distribution-free methods, the normal probability function was the unifying force behind the set of assumptions necessary for the interpretation of data. It is an integral part of the derivation of all of the sampling distributions.

If x is a random variable continuous in the interval (a, b) with probability density function $f(x)$ and cumulative distribution function $F(x)$, the probability that $x \leq A$ where $a \leq A \leq b$ is obtained by evaluating the integral

$$F(A) = \int_a^A f(x)dx$$

For the normal probability function

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \qquad -\infty < x < \infty$$

where the probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad -\infty < x < \infty$$

Functions that have this form were derived by mathematicians as early as the 18th century, and in some cases it was shown that the normal probability function is the limiting form of other important probability functions.

It is the purpose in this treatise to explore the origins of the normal function, and integrate the history, properties, uses, and interrelationships of these findings into present day normal probability theory.

In the literature of statistics today, authors use a variety of names for probability functions that describe the populations

frequency function
density function
distribution function
normal curve

Frequency function is a name given to a function that generates frequencies for an interval range of the random variable x. This name is more appropriate perhaps for functions that describe finite populations, or for functions of a discrete variable x. The total frequency may or may not sum to unity. The term distribution function is reserved to describe the value of the integral of the density function over a specified range of x. The density function is a non-negative function whose integral over the entire range of values x is equal to unity. The normal curve is a special density function. All of the functions listed must have certain properties

$$f(x) \geqslant 0$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\text{and } F(x_i) \geqslant F(x_j) \text{ for } x_i \geqslant x_j$$

In the case of discrete functions, the integral sign is replaced by the summation sign.

Historical Development

Abraham de Moivre (1667-1754)

Abraham de Moivre, though born in France, found it necessary in 1685 to take shelter in England, where he made his living solving problems for wealthy patrons and by private tutoring in mathematics. He is perhaps best known for his publications on trigonometry, probability, and annuities. In 1716, he published his *Doctrine of Chances* and in 1730, his *Miscellania Analytica*.

Karl Pearson found in his researches into de Moivre's works that de Moivre was in fact the first mathematician to derive the normal curve.

> . . . de Moivre published in 1730 his *Miscellania Analytica*. . . . Many copies have attached to them a *Supplementum* with separate pagination. . . . But only a very few copies have a second supplement, also with separate pagination (1-7) and dated Nov. 12, 1733. . . .[1]

The title of this second supplement is *Approximatio ad Summan Terminorium Binomii $(a + b)^n$ in Seriem Expansi.* A description of the method used by de Moivre is presented in Chapter II of this thesis. The supplement is written in Latin. However, de Moivre himself translated this into English and included it in the 1738 edition of his *Doctrine of Chances*.[2]

---

[1] "A Historical Note on the Origin of the Normal Curve of Errors." *Biometrika* 16, (1924) p. 402.

[2] pp. 235-243. The translation used here belongs to David Eugene Smith. "*A Source Book of Mathematics*" McGraw-Hill (1929) pp. 567-575.

Pearson states that for de Moivre, obtaining this most valuable result

> . . . was a theological problem, he was determining the frequency of irregularities from the Original Design of the Deity.[3]

Marquis Pierre Simon de Laplace (1749-1827)

Pierre Simon de Laplace is undeniably one of the greatest mathematicians of all time. However, as Glaisher comments:

> Laplace's investigations are rendered difficult, partly by want of fullness of explanation and partly by the peculiar character of the notation used and the novelty of the methods, which appear now so different from the way in which it is natural to consider the subject that it is not easy to feel any great confidence at first in the results as stated by Laplace.[4]

In his memoir of 1774, "Determiner le milieu que l'on doit prendre entre trois observations donnees d'un meme phenomeme." Mem. Acad Paris par divers savans (etrangers) vol 4, pp. 634-644.[5], Laplace sets forth the first attempt to deduce a rule for the combination of observations from the principles of probability. He says that the law of probability of observations may be represented by a curve whose equation is $y = \phi(x)$, x being any error and y its probability. He states that this curve must have three properties:

---

[3]Pearson op. cit. p.404.

[4]"On the Law of Facility of Errors of Observations and on the Method of Least Squares." Glaisher, J.W.L. Mem. Roy. Astr. Soc. 39, (1892) p.92.

[5]Merriman, Mansfield. "List of Writings Relating to the Method of Least Squares". Trans. Connect. Acad. 4, (1877) pp.151-232.

1. It must be symmetrical about the x-axis since positive and negative errors are equally probable.

2. The x axis must be an asymptote since the probability of an infinitely large error is zero.

3. The area under the curve is unity since it is certain that an error will be committed.

Laplace then takes as $\phi(x)$ the exponential function $\frac{1}{2} me^{-mx}$, where x is always regarded positive, without really justifying his choice.

Dr. Robert Adrain (1775-1843)

Dr. Adrain, as editor of the Analyst, offered a prize for a solution to the following problem:

> . . . to correct the distances and bearings of a survey, so as to deduce the most probable area of the enclosed field . . .[6]

A Dr. Bowditch, through special reasoning, was able to arrive at a solution of this problem based on the data supplied. His results coincided with what would have been deduced had the Gaussian method been applied. This solution was published in the Analyst of 1808, and immediately following it appeared Dr. Adrain's solution to the more general problem, the mathematics of which will be presented in Chapter II.

This work lay unknown for many years until Cleveland Abbe discovered it in 1871. It is of historical interest because

[6] "A Historical Note on the Method of Least Squares" Amer. Journ. Sci. ser 3. 1, (1871) p. 412.

it is another independent deduction of the normal curve. Its discovery prompted J. W. L. Glaisher to conduct research into the various methods of Least Squares. Glaisher used Adrain's proof as the introduction to this paper.

> Although some of the investigations of the law of facility $e^{-h^2x^2}$ are far from rigorous, still there is not one that is not of some importance, as throwing additional light on the properties of this law. . . . Dr. Adrain's proof, however, seems to me much inferior, both in point of rigor and conclusiveness . . . but . . . it appears worthwhile to notice the reasoning by which he obtained the law of facility.[7]

Dr. Adrain's proof depends upon the principle that the true errors of measured quantities are proportional to the quantities themselves. It is on this assumption that Glaisher bases his criticism of inferiority.

It is to the detriment of researches on the normal curve that mathematicians of the day concerned themselves primarily with researches on the Method of Least Squares as propounded by Legendre and Gauss. Thus, to these mathematicians, the normal curve was a means to an end rather than an end in itself. Many writers merely assumed the density function to be $e^{-h^2x^2}$, few of them actually included a deduction of this function in their demonstrations of the Method of Least Squares.

Karl Fredrich Gauss (1777-1855)

Karl Fredrich Gauss was one of the mathematicians who presented a derivation of the normal frequency function

---

[7] Glaisher *loc. cit.* p. 75.

in conjunction with his studies in the Method of Least Squares. He published his work under the title Theoria Motus Corporum Coelestium (Hamburg, 1809).

Gauss assumed that if several observations were made of a quantity, then the most probable value of that quantity would be the arithmetic mean of the observations. Taking this position, Gauss was then able to deduce the normal frequency function. Having established this, the Method of Least Squares follows almost immediately.

Those who criticize Gauss's approach do so from the point of view that there is no justification for assuming as axiomatic the postulate that the arithmetic mean is the most probable result. The basis for this argument is that the postulate yields a particular frequency function. Gauss's approach leaves little to be desired, even though it has failed to measure up to the standards of a rigorous proof. It is one of the earliest deductions of the normal curve.

Sir John Herschell (1792-1871)

Sir John Herschell, in his contributions to the method of Least Squares, also attempted a deduction of the frequency function $e^{-h^2x^2}$. In 1850, in a critique of Quetelet's methods,[8] Herschell presented his deduction of the frequency function in which his proof depends on the independence of the x and y deviations of a stone which has been dropped on a plane sur-

[8] "Quetelet on Probabilities." Edinburgh Review xcll (1850) pp. 1-57.

face. From this assumption, the form of the frequency function $ce^{-h^2x^2}$ is obtained. However, writers have commented upon the unwarranted nature of his assumption. Unless it can be proved that a deviation y occurs with the same comparative frequency when x has one value as when it has another, the assumption of independence should not be allowed.

The criticism, thus far leveled against the work of authors who wrote about the normal curve has been due in part to the difficulty or perhaps the impossibility in proving the normal curve. In all instances, the form of the frequency function was assumed, and it is assumed today, upon the strength of the Central Limit Theorem as a suitable model for a large number of statistical investigations, both theoretical and applied. There is not one instance in the literature where the normal curve is referred to as a theorem. This may be attributed to the fact that any basis for mathematical investigation when viewed as a starting point to be used to prove other concepts, may be taken as axiomatic and does not require proof.

Whittaker and Robinson comment on this situation:

> This state of things led in the past to much confusion of thought regarding the validity of the Normal Law, which was wittily referred to in Lippmann's remark to Poincare: "Everybody believes in the exponential law of errors: the experimenters, because they think that it can be proved by mathematics; and the mathematicians, because they believe that it has been established by observations."

---

[9]"<u>Calculus of Observations</u>" 4th edn. p.179 with the internal quote from Poincare, "Calcul des Probabilities" p.149.

The "Normal" Curve

Since most mathematical concepts are named after the mathematician who originated them, it is perhaps anomalous that the normal curve is not called the Gaussian curve or the Laplacean curve (ignoring for the moment the prior contribution of de Moivre). This situation is rendered less confusing after reading Karl Pearson's paper History of Correlation.

> . . . Many years ago, I called the Laplace-Gaussian curve the Normal Curve, which name, while it avoids an international question of priority, has the disadvantage of leading people to believe that all other distributions are in one sense or another 'abnormal' . . . It has led many writers to try and force all frequency by the aid of one or another process of distortion into a normal curve.[10]

Because Pearson can be considered to be the father of modern statistics in the English speaking world, it is no wonder that this theoretical distribution is known as the normal frequency curve.

Further Contributions to the Theory

This section deals with some authors whose memoirs did not deal directly with the normal curve, but did contribute to the general theory of errors which applies to the normal probability theory. Their contributions in the final analysis, combine to present a firmer theoretical basis.

---

[10] Biometrika 13, (1920) p.25.

Thomas Simpson (1710-1761)

Thomas Simpson in 1756 wrote, A Letter to the Right Honorable George Earl of Macclesfield, President of the Royal Society, on the Advantage of Taking the Mean of a Number of Observations in Practical Astronomy. This memoir is of great value in that it is among the first in which the theory of probability is applied to the discussion of errors of observations and in which there is an implied idea of a frequency function. Simpson wrote this letter because:

> . . . some persons of considerable note, have been of the opinion, and even publicly maintained, that one single observation taken with due care, was as much to be relied on as the Mean of a great number.[11]

There are two main propositions contained in this memoir. The first is a method of determining the probability that an error of the mean of 'n' observations shall be less than an assigned value provided that it is equally probable that the error of a single observation will be a specified quantity. The second proposition contains a method of determining that the probability, (provided that the probabilities of the single observations are proportional to a set of predetermined values) by taking the mean of 't' observations, does not exceed the given quantity m/t.

Simpson then illustrates his point with an example, after which he comments

> . . . it appears, that taking the Mean of a number of observations, greatly diminishes the chances for all of

[11]Phil. Trans. London xlix, pt. 1, (1755) p. 82.

> the smaller errors, and cuts off most all possibility for any great ones: which last consideration, alone, seems sufficient to recommend the method, not only to the astronomers, but to all others concerned in making experiments of any kind. . . .the more observations of experiments there are made, the less will the conclusions be liable to err, provided they admit of being repeated under the same circumstances.[12]

It was only one year later, when Simpson published a second memoir, An Attempt to Show the Advantage Arising by Taking the Mean of a Number of Observations in Practical Astronomy. This memoir is essentially the same as the one mentioned above, with the opening paragraph deleted and a second example incorporated. However, in this memoir, Simpson had found it necessary to lay down the following two axioms:

> 1. That there is nothing in the construction, or position of the instrument whereby the errors are constantly made to tend the same way, but that respective chances for their happening in excess, and in defect, are either accurately, or nearly, the same.
>
> 2. That there are certain assignable limits between which all these errors may be supposed to fall; which limits depend on the goodness of the instrument and the skill of the observer.[13]

The first axiom is considered to be the first instance where the equiprobability of positive and negative errors is set down, and there is mention of systematic errors in the measuring device. The second axiom sets out an outline for the idea of confidence intervals.

---

[12] "Miscellaneous Tracts on Some Curious Subjects." London, (1757) pp. 64-75.

[13] Ibid. p. 64.

Daniel Bernoulli (1700-1782)

Todhunter, in his A History of the Mathematical Theory of Probability,[14] while dealing with Daniel Bernoulli's Memoir Dijudicatio maxime probabilis plurium observationum discrepantium atque-versimillima inductio inde formanda (Acta. Acad. Petrop. 1777, pp.2-23), claims that Bernoulli was not acquainted with the researches of his predecessors when he wrote this memoir.

Bernoulli's treatment of the theory of errors is slightly different in that he claims that the method of taking the arithmetic mean of a set of discordant observations involves making the supposition that all of the observations have equal weight; a supposition to which Bernoulli takes exception. He prefers to consider the supposition that small errors are more probable than are large errors. Thus, allowing e to denote an error, he measures the probability by $\sqrt{r^2 - e^2}$, where r is a constant. Bernoulli then says that the best result will be determined from a number of observations when the product of their probabilities is a maximum. If he has the measurements $a, b, c, \ldots$ of the true value x then the following product must be a maximum for x.

$$\sqrt{r^2-(x-a)^2} \cdot \sqrt{r^2-(x-b)^2} \cdot \sqrt{r^2-(x-c)^2} \quad \ldots\ldots$$

---

[14] G. E. Stechert & Co. 1865, Arts. 424-427.

Todhunter concludes that:

> . . . Daniel Bernoulli agrees in some respects with modern theory. The chief difference is that modern theory takes for the curve of probability defined by the equation
>
> $$y = \sqrt{\frac{c}{\pi}}\; e^{-cx^2}$$
>
> while Daniel Bernoulli takes a circle.[15]

Some General Assumptions

Before proceeding to the mathematical development of the normal curve, it is convenient to mention the assumptions for the distribution function in general terms. This is because some of the methods do not make use of all of the assumptions. Thus, some degree of separation between methods of proof and general assumptions is necessary.

1. x is a continuous random variable,

2. positive and negative errors are equally probable,

3. the x_i are independent,

4. the probability of a certain deviation ε_i depends only on the magnitude of the deviation.

[15] *Ibid.* Art. 426, p. 237.

CHAPTER II

MATHEMATICAL DERIVATIONS OF THE NORMAL FREQUENCY FUNCTION

De Moivre's Derivation

De Moivre's derivation is based on his studies of the binomial expansion. He raises the binomial $(1 + 1)$ to a higher power n and finds that the ratio of the maximum term to the sum of all of the terms, may be expressed by the fraction

$$\frac{2A(n-1)^n}{n^n \sqrt{n-1}}$$

where $\log A = 1/12 - 1/360 + 1/1260 - 1/1680 + \cdots$

and for large n, $\frac{(n-1)^n}{n^n} = \left[1 - 1/n\right]^n = 1.$

If $\log B$ is defined as $-1 + \log A$,

$\log B = -1 + 1/12 - 1/360 + 1/1260 \cdots$ then $\frac{2A(n-1)^n}{n^n\sqrt{n-1}}$

may be written as $\frac{2B}{\sqrt{n-1}}$ or $\frac{2B}{\sqrt{n}}$.

Changing the signs in the series for $\log B$, yields the result $\frac{2}{B\sqrt{n}}$.

De Moivre attempted to calculate the value of B but was only able to determine that it converged rather slowly. Stirling showed that de Moivre's B was really the square root of the circumference of a circle whose radius is unity. Denoting the circumference by c, the ratio becomes

$$\frac{2}{\sqrt{nc}}.$$

The translation of this notation into modern notation indicates that de Moivre was aware that the maximum ordinate was given by

$$y_o = \frac{1}{\sigma\sqrt{2\pi}}$$

De Moivre proceeds to show that the ratio of the term at a distance ℓ from the maximum term to the maximum term has for its hyperbolic logarithm

$$-\frac{2\ell\ell}{n}$$

Again by translation of notation, the modern equivalent quantity is

$$y = y_o \quad e^{-\frac{x^2}{2\sigma^2}}$$

Thus de Moivre was the first to obtain the normal curve of errors, and in so doing also obtained an expression for n factorial similar to Stirling's approximation.

Adrain's Method

Dr. Adrain proposes the following problem:

> . . . Supposing AB to be the true value of any quantity of which the measure by observation or experiment is Ab , the error being Bb , what is the expression of the probability that the error Bb happens in measuring AB?[1]

Let AB, BC, etc. be successive distances the measures of which are Ab, bc, etc., the total error being Cc. Suppose the measures and the whole error are given. Assume

---

[1] "Research concerning the probabilities of the errors which happen in making observations." Analyst 1, (1808) p. 93

that AB, BC, are proportional to Ab, bc, and then the errors belonging to AB, BC, are proportional to their lengths. Let a, b, represent AB, BC, or Ab, bc,; E be the whole error Cc and x, y, the errors of the measures Ab, bc. For greatest probability, the equation

$$\frac{x}{a} = \frac{y}{b}$$

must hold.

Let $\phi(a, x)$ be the probability that an error x occurs in measuring a. Then the probability in measuring a, b is $\phi(a, x) \cdot \phi(b, y)$. From the relation $x + y = E$, and the condition that $\phi(a, x) \cdot \phi(b, y)$ is a maximum, the equation $\frac{x}{a} = \frac{y}{b}$ may be obtained by choosing the simplest forms of $\phi(a, x)$ and $\phi(b, y)$ possible.

For $\phi(a, x) \cdot \phi(b, y)$ to be maximum, subject to $x + y = E$,

$$\frac{\phi'(a, x)}{\phi(a, x)} dx + \frac{\phi'(b, y)}{\phi(b, y)} dy = 0; \; dx + dy = 0.$$

Thus
$$\frac{\phi'(a, x)}{\phi(a, x)} = \frac{\phi'(b, y)}{\phi(b, y)}.$$

This is equivalent to $\frac{x}{a} = \frac{y}{b}$

as shown by choosing $\frac{\phi'(a, x)}{\phi(a, x)} = \frac{mx}{a}$ and $\frac{\phi'(b, y)}{\phi(b, y)} = \frac{my}{b}$

where m is a constant.

By integration, $\phi(a, x) = e^{a' + \frac{mx^2}{2a}}$.

Dr. Adrain notes that m must be negative and the probability that the errors x, y, z, ... happen in the distances a, b, c, ... is

$$e^{a' + b' + c' + \cdots + \frac{mx^2}{2a} + \frac{my^2}{2b} + \frac{mz^2}{2c} + \cdots}$$

By choosing the constants carefully, Adrain obtains the formula $e^{f^2 - x^2}$, and by setting $f^2 = 0$, the curve takes its simplest form

$$y = e^{-x^2}.$$

Glaisher[2] says that it is not likely that the error should be directly proportional to the distance measured. He feels that there should be some neutralization of positive and negative errors. Thus he suggests $\frac{x}{\sqrt{a}} = \frac{y}{\sqrt{b}}$ would be a more reasonable assumption. As a general rule, he would prefer $\frac{x}{f(a)} = \frac{y}{f(b)}$ with $\frac{x}{a} = \frac{y}{b}$ and $\frac{x}{\sqrt{a}} = \frac{y}{\sqrt{b}}$ as special cases. Glaisher also notes that Adrain refers to $e^{a + \frac{mx^2}{2a}}$ as being the probability of an error x. However, he says that Adrain is not the only person who makes this mistake.

The significance of Adrain's development in the derivation of the normal function as it appeared to workers in his day is best summed up in a statement by Professor Abbe.

> . . . we must credit Dr. Adrain with the independent invention and application of the most valuable arithmetical process that has been invoked to aid the progress of the exact sciences.[3]

---

[2]"On the Law of Facility of Errors of Observations and on the Method of Least Squares." Mem. Roy. Astron. Soc. 39, (1892) p. 78.

[3]"A Historical Note on the Method of Least Squares." Amer. Journ. Sci. ser 3, 1, (1871) p. 415.

Gauss's Postulate of the Arithmetic Mean and Subsequent Derivation of the Normal Frequency Function

The Postulate of the Arithmetic Mean.

If apparently equally good observations are made of a certain quantity, then the most probable value of that quantity is the arithmetic mean of the observations, $\frac{x_1+x_2+x_3+\cdots+x_n}{n}$.[4]

Gauss's Derivation.

Let a be the true value. Thus $x_1 - a$, $x_2 - a$, , $x_n - a$ are the errors. Denote the frequency function by $\phi(x)$. Thus the *a priori* probability of these errors is proportional to

$$\phi(x_1-a)\cdot\phi(x_2-a)\cdot\phi(x_3-a)\cdot \cdots \cdot\phi(x_n-a). \qquad (1)$$

Note that after the observations have been made, the probability that a is the true value is also proportional to this same expression.

To find the value a , it is necessary to make (1) a maximum. Differentiating the logarithm of (1) with respect to a yields

$$\frac{\phi'(x_1-a)}{\phi(x_1-a)} + \frac{\phi'(x_2-a)}{\phi(x_2-a)} + \frac{\phi'(x_3-a)}{\phi(x_3-a)} + \cdots + \frac{\phi'(x_n-a)}{\phi(x_n-a)} \qquad (2)$$

---

[4] "Calculus of Observations." Whittaker and Robinson p. 216. The authors set up four axioms and derive the postulate of the arithmetic mean.

in which $a = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

Let $\frac{\phi'(x)}{\phi(x)} = \psi(x)$. The problem is to determine $\psi(x)$.

Let $x_2 = x_3 = x_4 = \cdots = x_n = x_1 - n\alpha$.

Thus $a = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = x_1 - (n-1)\alpha$.

Now, (2) reduces to $\psi\left\{(n-1)\alpha\right\} + (n-1)\psi\left\{(1-\alpha)\right\} = 0$

$$\text{or } \frac{\psi\left\{(n-1)a\right\}}{(n-1)a} = \frac{\psi(-a)}{-a} \quad \text{for any } a.$$

Thus $\frac{\psi(x)}{x} = \text{constant} = m$

$$\text{or } \frac{\phi'(x)}{x\phi(x)} = m$$

$$\frac{\phi'(x)}{\phi(x)} = mx$$

$\log_e \phi(x) = \frac{1}{2}mx^2 + \log_e A$ where A is a constant.

Finally, $\phi(x) = A\, e^{\frac{1}{2}mx^2}$.

In order that (2) be a maximum, it is necessary for m to be negative. Thus set $\frac{1}{2}m = -h^2$ $h > 0$.

Since all errors will lie between $\pm\infty$.

$$1 = \int_{-\infty}^{\infty} \phi(x)dx = \int_{-\infty}^{\infty} A\, e^{-h^2x^2}dx$$

$$= A\int_{-\infty}^{\infty} e^{-h^2x^2}dx$$

$$= \frac{A\sqrt{\pi}}{h}.$$

Thus $A = h/\sqrt{\pi}$.

Now, $\phi(x)\ \frac{h}{\sqrt{\pi}}\ e^{-h^2x^2}$ where Gauss called h the modulus of precision.

Herschell's Derivation

A simple line of reasoning was suggested by Sir John Herschell.[5] He supposed a board placed on the ground with two lines OX and OY ruled upon it at right angles to each other. A piece of metal is dropped on it with the intention of hitting the point O. The points of impact of several trials are marked and the problem is to find the law that will define the distribution of the points.

It is first assumed that the distribution is symmetrical about the point O and the probability that an impact point will fall between r and $r+\delta r$ may be represented by $\phi(r)\,\delta r$.

It is also assumed that the probability that an impact will fall between x and $x+\delta x$ is $f(x)\,\delta x$, independently of y; and the probability that it will fall between y and $y+\delta y$ is $f(y)\,\delta y$, independently of x.

The probability that the piece of metal will fall on the rectangle $\delta x\,\delta y$ is $f(x)\,f(y)\,\delta x\,\delta y$

and it follows that $f(x)\,f(y)\,\delta x\,\delta y = \phi(r) r\,\delta r\,\delta\theta$

where $x = r\cos\theta$ and $y = r\sin\theta$.

Taking the logarithm of the expression transformed into polar co-ordinates and differentiating with respect to θ yields

$$-\sin\theta\,\frac{f'(r\cos\theta)}{f(r\cos\theta)} + \cos\theta\,\frac{f'(r\sin\theta)}{f(r\sin\theta)} = 0$$

---

[5] "Quetelet on Probability." Edinburgh Review xcii (1850) pp. 1-57.

$$\text{or } \frac{f'(x)}{xf(x)} = \frac{f'(y)}{yf(y)} = \text{constant} = -2h^2.$$

The constant must be negative since $f(x)$ cannot increase indefinitely with x.

Thus $$\log f(x) = -h^2x^2 + \log C$$

$$f(x) = Ce^{-h^2x^2}$$

Similarly $$f(y) = Ce^{-h^2y^2}$$

and $$\phi(r) = f(x)\, f(y) = C^2 e^{-h^2(x^2+y^2)} = C^2 e^{-h^2r^2}.$$

The constant C is determined from the condition that the piece of metal must fall somewhere on the plane, so that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\, f(y)\, dxdy = \int_0^{2\pi}\int_0^{\infty} \phi(r) r dr d\theta = 1$$

$$\int_0^{2\pi}\int_0^{\infty} C^2 \; e^{-h^2r^2} \; r dr d\theta = 1$$

$$\frac{C^2}{2h^2}\int_0^{2\pi} d\theta = 1$$

$$\frac{C^2\pi}{h^2} = 1$$

Since $C^2\, \pi/h^2 = 1$

Thus $$\frac{C\sqrt{\pi}}{h} = 1$$

or $$C = \frac{h^2}{\sqrt{\pi}}.$$

Hence the law of distribution in either co-ordinate, which follows from the assumptions made, is of the form

$$f(x) = \frac{h}{\sqrt{\pi}} \quad e^{-h^2 x^2}.$$

The value of Herschell's method is that it is based upon simple assumptions. As noted in Chapter I, Ellis[6] takes exception to the assumption of the independence of the x and y deviations. However, the independence of the x and y deviations is a property of the frequency function $e^{-h^2 x^2}$.

[6]A paper in the Philosophical Magazine for Nov. 1850.

CHAPTER III

THE MATHEMATICAL DESCRIPTION OF THE NORMAL FREQUENCY FUNCTION

Some Special Mathematical Expectations

If the variable x is continuous, and if $f(x)dx$ is the probability of a value between x and $x + dx$, then the expected value of $p(x)$ is given by

$$E\left[p(x)\right] = \int_{-\infty}^{\infty} p(x)f(x)\,dx$$

provided the integral exists; and where $\int_{-\infty}^{\infty} f(x)\,dx = 1.$

Two elementary, but important properties of expected values should be noted.

1. The expected value of a sum is the sum of the expected values.

$$E\left[p(x) + q(x)\right] = E\left[p(x)\right] + E\left[q(x)\right].$$

Proof:

$$\begin{aligned} E\left[p(x) + q(x)\right] &= \int_{-\infty}^{\infty}\left[p(x) + q(x)\right] f(x)\,dx \\ &= \int_{-\infty}^{\infty} p(x)f(x)\,dx + \int_{-\infty}^{\infty} q(x)f(x)\,dx \\ &= E\left[p(x)\right] + E\left[q(x)\right] \end{aligned}$$

2. The expected value of a product is the product of the expected values, if the functions depend on sets of variates that are independent of each other.

$$E\left[p(x)\quad q(x)\right] = E\left[p(x)\right] \cdot E\left[q(x)\right]$$

Proof:

If $f_1(x_1)$ and $f_2(x_2)$ represent independent functions, such that $f_1(x_1)f_2(x_2) = f(x_1x_2)$, then

$$\begin{aligned} E\left[p(x_1)q(x_2)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1)q(x_2)f(x_1x_2)\,dx_1\,dx_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1)q(x_2)f_1(x_1)f_2(x_2)\,dx_1\,dx_2 \\ &= \left[\int_{-\infty}^{\infty} p(x_1)f_1(x_1)\,dx_1\right]\left[\int_{-\infty}^{\infty} q(x_2)f_2(x_2)\,dx_2\right] \\ &= E\left[p(x_1)\right] \cdot E\left[q(x_2)\right] \end{aligned}$$

It must be noted that property (2) holds only for independent variates; while property (1) holds for both dependent and independent variates.

Consider the form $E\left[p(x)\right] = \int_{-\infty}^{\infty} p(x)f(x)\,dx.$

1) If $p(x) = x^r$, then the integral over the range of x, defines the rth moment about the origin. This is sometimes referred to as the rth non-central moment about the origin.

$$\mu_r^1 = \int_{-\infty}^{\infty} x^r\, f(x)\,dx$$

2) If $p(x) = (x-\mu)^r$, the integral defines the rth moment about the mean. This is sometimes referred to as the rth central moment.

$$\mu_r = \int_{-\infty}^{\infty} (x-\mu)^r f(x)\,dx$$

At this stage, it is possible to define the mean and variance of a population in terms of the moments.

The arithmetic mean is the first moment about the origin

$$\mu_1' = \int_{-\infty}^{\infty} x\,f(x)\,dx$$

The usual practise is to drop the subscripts and the superscripts, and to designate the population mean as μ.

It may be easily verified that the moments of zero and first order, about the mean, are unit and zero, respectively. The second moment about the mean is known as the population variance.

$$\sigma^2 = \mu_2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx.$$

Since it is easier to calculate the non-central moments than it is to calculate the central moments, the former are used to obtain the latter. The relationship between the non-central and the central moments will be demonstrated on the basis of the second central moment.

$$\mu_2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

$$= \int_{-\infty}^{\infty} \left[x^2 - 2\mu x + \mu^2\right] f(x)\, dx$$

by property (1),

$$\mu_2 = \int_{-\infty}^{\infty} x^2 f(x) dx - 2\mu \int_{-\infty}^{\infty} x f(x) dx + \mu^2 \int_{-\infty}^{\infty} f(x) dx.$$

$$= \mu_2' - 2\mu^2 + \mu^2$$

$$= \mu_2' - \mu^2$$

In a similar fashion, the following central moments may be obtained.

$$\mu_3 = \mu_3' - 3\mu_2'\mu + 2\mu^3$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4$$

<u>The Moment Generating Function</u>

3) If in $E\left[p(x)\right] = \int_{-\infty}^{\infty} p(x) f(x) dx$, $p(x) = e^{tx}$, then the moment generating function of the variate x may be defined as:

$$M_x(t) = E\left[e^{tx}\right] = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

provided that the integral exists.

The relationship between the moment generating function and the moments follows

$$e^{tx} = 1+\frac{tx}{1!}+\frac{(tx)^2}{2!}+\frac{(tx)^3}{3!}+\cdots+\frac{(tx)^k}{k!}+\cdots$$

and

$$M_x(t) = \int_{-\infty}^{\infty}\left[1+tx+\frac{(tx)^2}{2!}+\frac{(tx)^3}{3!}+\cdots+\frac{(tx)^k}{k!}+\cdots\right] f(x)dx$$

$$= \int_{-\infty}^{\infty} f(x)dx+\int_{-\infty}^{\infty}\frac{tx}{1!}f(x)dx+\int_{-\infty}^{\infty}\frac{(tx)^2}{2!}f(x)dx+\cdots+\int_{-\infty}^{\infty}\frac{(tx)^k}{k!}f(x)dx+\cdots$$

$$= 1+\frac{t}{1!}\mu_1'+\frac{t^2}{2!}\mu_2'+\cdots+\frac{t^k}{k!}\mu_k'+\cdots$$

The moment generating function is expressed as an infinite power series in t, and the non-central moments are the coefficients of $\frac{t^k}{k!}$.

To obtain any particular non-central moment, $M_x(t)$ is differentiated with respect to t the number of times indicated by the subscript of the required non-central moment, and then t is equated to zero.

For example, in obtaining the second non-central moment

$$M_x^{(2)}(t) = \mu_2'+t\mu_3'+\frac{t^2}{2!}\mu_4'+\cdots+\frac{k(k-1)}{(k-2)!}t^{k-2}\mu_k'+\cdots$$

and by setting $t=0$, the right hand side reduces to μ_2'.

4) There are cases in which the moment generating function does not exist. In these cases, a slight modification in the expected value leads to the characteristic function, which does exist for all frequency functions.

If in $E\left[p(x)\right]=\int_{-\infty}^{\infty}p(x)f(x)dx$, $p(x) = e^{itx}$, then the characteristic function of the variate x may be defined as:

$$\phi_x(t) = E\left[e^{itx}\right] = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx. \qquad \begin{array}{l} -\infty < x < \infty \\ i = \sqrt{-1} \end{array}$$

Using the same method as that applied to obtain the moment generating function, the characteristic function may be expressed as an infinite power series in (it) where the non-central moments form the coefficients of $\frac{(it)^k}{k!}$.

$$\phi_x(t) = 1 + (it)\mu_1' + \frac{(it)^2}{2!}\mu_2 + \frac{(it)^3}{3!}\mu_3 + \cdots + \frac{(it)^k}{k!}\mu_k' + \cdots$$

Since the characteristic function is representative of the general theory of Fourier transforms, some of the procedures may be utilized here. The most fundamental theorem of the theory of characteristic functions is the Inversion Theorem, which shall be stated without proof here.[1]

$$\text{Since} \quad \phi_x(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx$$

$$\text{then} \quad f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi_x(t)\,dt.$$

Calculation of the Moment Generating Function

Before any of these calculations can be made, a slight change is necessary in the frequency function of the normal curve. Up to this point in the discussion of the frequency function $\frac{h}{\sqrt{\pi}} e^{-h^2 x^2}$, the symbol x has been used to represent

[1]For a proof of this theorem, see Kendal and Stuart, "Advanced Theory of Statistics." I, (1958) p. 94.

the random error ε. In order to simplify the derivation, x is redefined as a random observation or measurement of the true value a. Then $\varepsilon = x - a$ represents a random error. The form of the frequency function becomes

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2(x - a)^2}$$

The moment generating function of the normal frequency function is

$$M_x(t) = E\left[e^{tx}\right] = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

$$= \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{tx}\, e^{-h^2(x - a)^2} dx$$

$$= \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-h^2\left[(x-a)^2 - \frac{tx}{h^2}\right]} dx$$

and completing the square of

$$(x-a)^2 - \frac{tx}{h^2} = x^2 - 2ax + a^2 - \frac{tx}{h^2}$$

yields

$$x^2 - 2ax + a^2 - \frac{tx}{h^2} = x^2 - \left(2a + \frac{t}{h^2}\right)x + \left(a + \frac{t}{2h^2}\right)^2 - \left(a + \frac{t}{2h^2}\right)^2 + a^2$$

$$= \left[x - \left(a + \frac{t}{2h^2}\right)\right]^2 - a^2 - \frac{at}{h^2} - \frac{t^2}{4h^4} + a^2$$

$$= \left[x - \left(a + \frac{t}{2h^2}\right)\right]^2 - \frac{1}{h^2}\left[at + \frac{t^2}{4h^4}\right]$$

Now,

$$M_x(t) = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-h^2\left\{\left[x - \left(a + \frac{t}{2h^2}\right)\right]^2 - \frac{1}{h^2}\left[at + \frac{t^2}{4h^2}\right]\right\}} dx$$

$$= \frac{h}{\sqrt{\pi}} e^{\left[at + \frac{t^2}{4h^2}\right]} \int_{-\infty}^{\infty} e^{-h^2\left[x - \left(a + \frac{t}{4h^2}\right)\right]^2} dx$$

Let

$$z = x - \left(\frac{a + t}{4h^2}\right) \qquad dz = dx$$

then

$$M_x(t) = \frac{h}{\sqrt{\pi}} e^{\left[at + \frac{t^2}{4h^2}\right]} \frac{\sqrt{\pi}}{h}$$

$$= e^{\left[at + \frac{t^2}{4h^2}\right]}$$

with the result that the moment generating function of the normal frequency function is

$$e^{at+\frac{t^2}{4h^2}}$$

The mean or first non-central moment is calculated as follows:

$$M_x'(t)\Big|_{t=0} = \mu_1' = \mu$$

$$M_x'(t) = e^{at+\frac{t^2}{4h^2}}\left[a+\frac{t}{2h^2}\right]$$

at $t=0$, this reduces to a.

Thus $\mu = a$.

The variance or second central moment is defined as

$$\sigma^2 = \mu_2'-\mu^2$$

and is calculated as follows:

$$M_x^2(t) = e^{at+\frac{t^2}{4h^2}}\left[\frac{1}{2h^2}\right] + e^{at+\frac{t^2}{4h^2}}\left[a+\frac{t}{2h^2}\right]^2$$

which at $t=0$ reduces to $\frac{1}{2h^2} + a^2$

$$\text{and } \sigma^2 = \frac{1}{2h^2} + a^2 - a^2 = \frac{1}{2h^2}$$

In terms of the population parameters μ and σ^2 the moment generating function may be rewritten as

$$M_x(t) = e^{\mu t+\frac{\sigma^2 t^2}{2}}$$

and the frequency function for the normal curve may be written as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Continuing in this fashion, the results $\mu_3 = 0$ and $\mu_4 = 3\sigma^4$ may be obtained.

The characteristic function for the normal frequency function is

$$\phi_x(t) = E\left[e^{itx}\right] = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{itx}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\, dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left[(x-\mu)^2 - 2it\sigma^2 x\right]}\, dx$$

hence $\phi_x(t) = e^{\mu it - \frac{\sigma^2 t^2}{2}}$

The Inversion or Uniqueness Theorem may be illustrated in the following problem.

To find the distribution function of the random variable x given the characteristic function

$$\phi_x(t) = e^{\frac{-\sigma^2 t^2}{2}}$$

by the Inversion theorem,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}\, \phi_x(t)\, dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}\, e^{\frac{-\sigma^2 t^2}{2}}\, dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[t^2\sigma^2 + 2itx\right]}\, dt$$

and completing the square in

$$t^2\sigma^2 + \frac{2it\sigma x}{\sigma} = t^2\sigma^2 + \frac{2it\sigma x}{\sigma} + \left(\frac{ix}{\sigma}\right)^2 - \left(\frac{ix}{\sigma}\right)^2$$

$$= \left[t\sigma + \frac{ix}{\sigma}\right]^2 - \left(\frac{ix}{\sigma}\right)^2$$

with the result

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[\left(t\sigma+\frac{ix}{\sigma}\right)^2-\left(\frac{ix}{\sigma}\right)^2\right]}\,dt$$

$$= \frac{e^{\frac{1}{2}\left(\frac{ix}{\sigma}\right)^2}}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(t\sigma+\frac{ix}{\sigma}\right)^2}\,dt$$

Let $y = t\sigma + \frac{ix}{\sigma}$ and $dy = \sigma\, dt$

then $f(x) = \frac{e^{-\frac{x^2}{2\sigma^2}}}{2\pi}\ \frac{\sqrt{2\pi}}{\sigma} = \frac{e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}}{\sqrt{2\pi}\sigma}$

which is recognized as the frequency function of the normal curve when the mean of the errors is equal to zero.

A rather interesting property of the normal frequency function is brought forth if the moment generating function of the moments about the mean is considered.

$$M_{x-\mu}(t) = e^{\frac{\sigma^2 t^2}{2}}$$

This function yields the central moments directly.

For example, $\mu_2 = M^{(2)}_{(x-\mu)}(t)\Big|_{t=0} = \sigma^2$

$$\mu_3 = M^{(3)}_{(x-\mu)}(t)\Big|_{t=0} = 0$$

$$\mu_4 = M^{(4)}_{(x-\mu)}(t)\Big|_{t=0} = 3(\sigma^2)^2$$

$$\mu_5 = M^{(5)}_{(x-\mu)}(t)\Big|_{t=0} = 0$$

$$\mu_6 = M^{(6)}_{(x-\mu)}(t)\Big|_{t=0} = 15(\sigma^2)^3$$

Two properties of the moments of the normal probability function are of interest:

1) the moments obtained from odd powers of x are zero

$$\mu_{2k+1} = 0$$

2) the even moments about the mean can be expressed as functions of the second moments about the mean. This relationship takes the form

$$\mu_{2k} = \prod_{i=1}^{k} (2i-1) \ \mu_2^{\ k}$$

Skewness and Kurtosis

In a symmetrical distribution, as is the normal frequency function, measures of location such as the mean, median, and mode all fall on the same point. It is of interest therefore to determine the amount of deviation or the amount of skewness that exists when the measures of location do not all fall on the same point. The moment measure of skewness is given by the ratio

$$\frac{\mu_3}{\mu_2^{3/2}}$$

which Karl Pearson called $\sqrt{\beta}$.

As will be shown, for the normal frequency function, $\sqrt{\beta_1}$ assumes the value zero. Thus any departure from symmetry is given by either a positive or negative value of $\sqrt{\beta_1}$. In the case where $\sqrt{\beta_1}$ is negative, it is found that the distribution has a long tail to the left; and in the case where $\sqrt{\beta_1}$ is positive, the long tail is to the right; the former being called negative skewness, and the latter being called positive skewness.

For the normal frequency function, it has been shown that

$$\mu_3 = 0$$

$$\text{and} \quad \mu_2 = \sigma^2$$

$$\text{Thus} \quad \sqrt{\beta_1} = \frac{0}{\sigma^3} = 0$$

The symmetrical distribution also has a measure of 'peakedness' known as kurtosis. Originally, kurtosis was a measure of the peakedness of a curve, but now it is used to describe the behavior of the tails of the distribution. This means that both height of the tails from the axis of measurement and the manner in which the tails approach the axis of measurement contribute to the quantity known as kurtosis. The moment measure of kurtosis is given by the ratio

$$\frac{\mu_4}{\mu_2{}^2}$$

which Karl Pearson called β_2.

For the frequency function of the normal curve, β_2 has been found to assume the value three. If the value of β_2 is found to be less than three, the curve is said to be platykurtic; while if β_2 is greater than three, the curve is said to be peaked or leptokurtic. A curve with $\beta_2 = 3$ or nearly so is said to be mesokurtic.

It has already been shown for the normal frequency function that

$$\mu_4 = 3\sigma^4$$

and $\mu_2 = \sigma^2$

Thus $\beta_2 = \frac{3\sigma^4}{(\sigma^2)^2} = 3$

A Discussion of Cumulants or Semi-invariants

If the logarithm of the moment generating function $M_x(t)$ can be expanded as a convergent series in powers of t in the form

$$L(t) = \log M_x(t) = \lambda_1 t + \frac{\lambda_2 t^2}{2!} + \frac{\lambda_3 t^3}{3!} + \cdots$$

then $L(t)$ is defined to be the semi-invariant generating function, and the coefficients λ_i are called the semi-invariants of the function $f(x)$.

This is the definition for semi-invariants or cumulants as they are more often termed. It is not the same definition for semi-invariants that Theile used in 1903[2] when he first proposed their existence. Fisher[3] notes that the series of statistics termed semi-invariants are related to the population

[2] T.N. Theile "The Theory of Observations" 1903. Reprinted in the Annals of Math. Stat. 2, (1931) pp. 165-307.

[3] R.A. Fisher "Statistical Methods for Research Workers" 13th edn. (1958) p. 72.

moments. He claims that, as defined by Theile, the semi-invariants are no longer of any importance. In contrast to Fisher, most of the modern authors do not make this distinction between semi-invariants and cumulants. It is therefore not uncommon to see the terms used interchangeably. However, the Scandinavian school founded by Theile retains the term semi-invariant, while the English scholars under Karl Pearson retain the term cumulant.

It is of interest to note the reasoning behind the term "semi-invariant". This term came into existence when it was discovered that any linear transformation of the form

$$s = qx + m$$

leaves the cumulants unchanged, with the exception of the first cumulant which equals the mean and is therefore liable to change under a transformation.

Since with the exception of the first the cumulants are not influenced by a linear transformation, the cumulants stand out in contrast to the moments which exhibit some changes when under the influence of a linear transformation.

It is now proposed to show the relationship which exists between the first few moments and the first few cumulants. In terms of the cumulants (where $k_i = \lambda_i$ in the above definition), there is the following relationship

$$K_x(t) = \log_e M_x(t) = k_1 t + \frac{k_2 t^2}{2!} + \frac{k_3 t^3}{3!} + \cdots$$

where $M_x(t) = 1+\mu_1' t+\frac{\mu_2' t^2}{2!}+\frac{\mu_3' t^3}{3!}+ \cdots$

Now to transform the above moment generating function from one dealing with the non-central moments to one dealing with the central moments by the transformation of the origin to the mean

$$M_{x-\mu}(t) = 1+\frac{\mu_2 t^2}{2!}+\frac{\mu_3 t^3}{3!}+\frac{\mu_4 t^4}{4!}+ \cdots$$

Since $$\frac{dK_{x-\mu}(t)}{dt} = \frac{1}{M_{(x-\mu)}(t)} = \frac{dM_{x-\mu}(t)}{dt}$$

it follows that

$$\left[k_1+k_2 t+\frac{k_3 t^2}{2}+\cdots\right]\left[1+\frac{\mu_2 t^2}{2!}+\frac{\mu_3 t^3}{3!}+\cdots\right] = \left[\mu_2 t+\frac{\mu_3 t^2}{2!}+\frac{\mu_4 t^3}{3!}+\cdots\right]$$

Now, by equating the coefficients of like powers of t, there results

$$k_1 = 0$$
$$k_2 = \mu_2$$
$$k_3 = \mu_3$$
$$k_4+3\mu_2 k_2 = \mu_4$$
$$\mu_5+6\mu_2 k_3+4\mu_3 k_2 = \mu_5$$
$$\cdots\cdots\cdots\cdots$$

which in terms of the kappa's (k_i) becomes

$$k_2 = 0$$
$$k_3 = \mu_3$$
$$k_4 = \mu_4-3\mu_2^2$$
$$k_5 = \mu_5-10\mu_2\mu_3$$
$$\cdots\cdots\cdots$$

A similar method yields the following relationship between the cumulants and the non-central moments

$$k_1 = \mu_1'$$
$$k_2 = \mu_2' - \mu_1'^{2}$$
$$k_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^{3}$$
$$k_4 = \mu_4' - 4\mu_3'\mu_1' - 3\mu_2'^{2} + 12\mu_2'\mu_1'^{2} - 6\mu_1'^{4}$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

It is immediately apparent that for the moments about the mean (i.e. $\mu_1' = k_1 = 0$) the above relationships reduce to those previously noted.

For the normal frequency function, the first four cumulants are given as follows:

$$k_1 = \mu$$
$$k_2 = \sigma^2$$
$$k_3 = 0$$
$$k_4 = 0$$

It is a property of the normal frequency function that $k_r = 0$ for $r > 2$ and hence the first two cumulants completely determine the distribution.

An Evaluation of the Normal Probability Integral

Since the normal frequency function is, as shall be shown in Chapter IV, the theoretical basis for many other frequency curves, it is no wonder that Laplace[4] in 1783, merely from his personal studies, saw the need of determining a table

---

[4] "Theorie Analytique des Probabilities", p. 195 as noted by Todhunter "A History of the Mathematical Theory of Probability" Art. 910, p. 486.

of values for the probability integral. There is a problem involved in obtaining this set of values, as the normal frequency function contains the parameters μ and σ^2 which can vary. It is therefore advisable to introduce a standard normal variable.

The standard normal variate is the transformation denoted by

$$z = \frac{x - \mu}{\sigma}$$

which is equivalent to setting the mean equal to zero and the variance equal to one. Another way to express this same idea is the notation $x_i = N(0, 1)$. However, since the x_i are usually sample values, this statement is rewritten as $x_i \sim N(0, 1)$, which means that x_i is approximately normally distributed with zero mean and unit variance.

The frequency function

$$f(z) = \frac{1}{\sqrt{2\pi}} \quad e^{-\frac{1}{2} z^2}$$

has not changed through the process of the transformation, and it is still the normal frequency function. This is the frequency function which is evaluated using the process of integration to determine the areas under the normal curve.

The values of the integral

$$F(z_o) = \frac{1}{\sqrt{2\pi}} \int_0^{z_o} e^{-\frac{1}{2} z^2} \, dz$$

for different z_o values are the ones usually tabulated. The value of this integral is the area under the curve from $z = 0$ to $z = z_o$. Since the total area under the curve is unity, any area may be interpreted as the probability of a value, chosen at random, falling within the limits of the integral.

Although the integral cannot be evaluated in finite form, the integrand may be expressed as an infinite series. This series may then be integrated as far as is necessary to insure a prestated accuracy of result.

Since $\quad e^x = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots+\frac{x^p}{p!}+\cdots$

it follows that $\quad e^{-\frac{1}{2}z^2} = 1-\frac{z^2}{2}+\frac{z^4}{2^2\cdot 2!}-\frac{z^6}{2^3\cdot 3!}+\frac{z^8}{2^4\cdot 4!}-\frac{z^{10}}{2^5\cdot 5!}+\cdots$

and $\quad \int_0^{z_0} e^{-\frac{1}{2}z^2}\,dz = z_0-\frac{z_0^3}{6}+\frac{z_0^5}{40}-\frac{z_0^7}{336}+\frac{z_0^9}{3456}-\frac{z_0^{11}}{42240}+\cdots$

It is known that for small values of z_0, this series will converge very rapidly. However, for large values of z_0, even though the series is known to converge, it would take a digital computer to handle the number of terms that would be necessary to obtain a result with the prestated accuracy. As an illustration, consider the calculations that are involved in obtaining two values for the table of areas.

1. Evaluate the probability integral for $z_0 = 0.1$

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_0^{0.1} e^{-\frac{1}{2}z^2}\,dz &= .3989423\left[.1-\frac{.001}{6}+\frac{.00001}{40}\right]\\ &= .3989423\left[.1-.000166+.00000025\right]\\ &= .3989423\left[0.09983425\right]\\ &= .039828\end{aligned}$$

Note that with z_o this small, six place accuracy is obtained with the expansion of only the first two terms although the first three terms were summed here, and the final result rounded to six places.

2. Evaluate the probability integral for z_o - 1.0

$$\frac{1}{\sqrt{2\pi}}\int_0^{1.0} e^{-\frac{1}{2}z^2}\,dz = .3989425\left[1-\frac{1}{6}+\frac{1}{40}-\frac{1}{336}+\frac{1}{3456}-\frac{1}{42240}\right]$$

$$= .3989425\left[1-.1666666+.0250000-.0029761 +.0002893-.0000237\right]$$

$$= .3989425\left[0.8556228\right]$$

$$= .3413$$

In this example, since z_o is larger than it was in the first example, the expansion to six terms only gives four place accuracy. To obtain six place accuracy in the second example, the first eight terms would have to be summed.

The Additive Property of Normal Variates

If x_1, x_2, x_3, $\cdots$, x_n are n independent normal varietes with mean m_i and variance σ_i^2, the variate $\sum_{i=1}^{n} x_i$ is a normal variate with mean $\sum_{i=1}^{n} m_i$ and variance $\sum_{i=1}^{n}\sigma_i^2$,

Since the variates are independent, the moment generating function of their sum is the product of the separate moment generating functions. Thus the moment generating function of $\sum_{i=1}^{n} x_i = \prod_{i=1}^{n}$ (moment generating functions of x_i)

$$= \prod_{i=1}^{n} e^{m_i t+\frac{1}{2}t^2\sigma_i^2}$$

$$= e^{t \sum m_i + \frac{1}{2} t^2 \sum \sigma_i^2}$$

Thus $\sum_{i=1}^{n} x_i$ has a normal distribution with mean $\sum_{i=1}^{n} m_i$ and variance $\sum_{i=1}^{n} \sigma_i^2$.

CHAPTER IV

DISTRIBUTIONS WHICH HAVE THE NORMAL FREQUENCY FUNCTION AS THEIR LIMIT

There are many instances in the application of statistics to the procedures used in estimation and in tests of hypotheses where it is said that the conditions necessary for a normal approximation are sufficiently satisfied to apply normal theory to the solution of the problem. The conditions necessary for satisfactory approximation to the normal case vary with the nature of the basic population. It is worthwhile to investigate the particular circumstances under which a satisfactory solution is obtained, by using the normal approximation, for a number of commonly used frequency functions.

The Binomial Distribution

Consider the frequency function for the Binomial distribution:

$$f(x) = \binom{n}{x} p^x q^{n-x} \qquad \begin{array}{c} x = 0,1,2,\ldots,n \\ p+q = 1 \end{array}$$

The characteristic function of the Binomial distribution is found to be

$$\phi_x(t) = (q+pe^{it})^n.$$

Using the transformation to standard measure $y = \frac{x-\mu}{\sigma}$, for the Binomial distribution, it is easily verified that the mean is np and the variance is npq.

The properties of characteristic functions yield the relation: $\phi_y(t) = e^{-it\mu/\sigma}\, \phi_x(t/\sigma)$

The characteristic function for the Binomial distribution expressed in standard measure becomes

$$\phi_y(t) = e^{\frac{-itnp}{\sqrt{npq}}} \left[q+pe^{\frac{-itnp}{\sqrt{npq}}} \right]^n$$

Taking the natural logarithm

$$\log_e \phi_y(t) = \frac{-itnp}{\sqrt{npq}} + n \log_e \left[q+pe^{\frac{it}{\sqrt{npq}}} \right]$$

given that $e^y = 1+y+\frac{y^2}{2!}+\frac{y^3}{3!}+\cdots+\frac{y^k}{k!}+\cdots$

$$\log_e \phi_y(t) = \frac{-itnp}{\sqrt{npq}}+n \log_e \left[q+p+p\frac{it}{\sqrt{npq}}+\frac{p}{2!}\frac{(it)^2}{npq}+\frac{p(it)^3}{3!(npq)^{3/2}} + \cdots \right]$$

$$= \frac{-itnp}{\sqrt{npq}}+n \log_e \left[(q+p)+p\sum_{r=1}^{n} \frac{1}{r!}\left(\frac{it}{\sqrt{npq}}\right)^r \right]$$

Thus, $\log_e \phi_y(t) = \frac{-itnp}{\sqrt{npq}}+n \log_e \left[1+p\sum_{r=1}^{n} \frac{1}{r!}\left(\frac{it}{\sqrt{npq}}\right)^r \right]$

Given that $\log_e (1+v) = v-\frac{v^2}{2}+\frac{v^3}{3}-\frac{v^4}{4}+\cdots$

it follows that:

$$\log_e \phi_y(t) = \frac{-itnp}{\sqrt{npq}}+n \left[p \sum_1^n \frac{1}{r!}\left(\frac{it}{\sqrt{npq}}\right)^r - \frac{p^2}{2}\left\{\sum_1^n \frac{1}{r!}\left(\frac{it}{\sqrt{npq}}\right)^r\right\}^2 + \frac{p^3}{3}\left\{\sum_1^n \frac{1}{r!}\left(\frac{it}{\sqrt{npq}}\right)^r\right\}^3 - \cdots \right]$$

$$= \frac{-itnp}{\sqrt{npq}}+\frac{itnp}{\sqrt{npq}}-\frac{np(1-p)}{npq}\,\frac{t^2}{2}+\frac{np(1-p^2)}{2(npq)}\,(-i)\frac{(t)^3}{3}+\cdots$$

Finally, $\log_e \phi_y(t) = \frac{-t^2}{2}+\frac{np(1-p^2)}{2(npq)}\,(-i)\,\frac{t^3}{3}+\cdots$

In the limit, as n approaches infinity, it is found that:

$$\lim_{n\to\infty} \log_e \phi_y(t) = \frac{-t^2}{2}$$

Thus, in the limit, when expressed in standard measure, the Binomial distribution becomes the standard Normal distribution.

$$\lim_{n \to \infty} \phi_y(t) = e^{\frac{-t^2}{2}}$$

The Poisson Distribution

Consider the frequency function of the Poisson distribution

$$f(x) = \frac{e^{-m}m^x}{x!} \qquad x = 0,1,2,\ldots,n$$

The characteristic function of the Poisson frequency function is found to be

$$\phi_x(t) = e^{m(e^{it}-1)}$$

Using the transformation to standard measure, $y = \frac{x-\mu}{\sigma}$, for the Poisson distribution, it is easily verified that the mean and variance are equal to m.

$$\text{Thus } \phi_y(t) = e^{it\frac{\mu}{\sigma}}\,\phi_x(t/\sigma)$$

$$= e^{\frac{-itm}{\sqrt{m}}}\,e^{m(e^{\frac{it}{\sqrt{m}}}-1)}$$

$$\log_e \phi_y(t) = \frac{-itm}{\sqrt{m}}+m(e^{\frac{it}{\sqrt{m}}}-1)$$

$$= \frac{-itm}{\sqrt{m}}+m\left[\frac{it}{\sqrt{m}}+\frac{(it)^2}{2m}+\frac{(it)^3}{3\,m^{3/3}}+\frac{(it)^4}{4!m^2}+\cdots\right]$$

$$= \frac{-t^2}{2}+\frac{i^3t^3}{6m^{1/2}}+\frac{i^4t^4}{24m}+\cdots$$

$$\lim_{m \to \infty} \log_e \phi_y(t) = \frac{-t^2}{2}.$$

Therefore, as the sample size increases, and m approaches infinity, the limiting form of the Poisson distribution expressed in standard measure is the Normal distribution.

$$\lim_{m \to \infty} \phi_y(t) = e^{-\frac{t^2}{2}}$$

The Chi-Square Distribution

The frequency function of the Chi-Square distribution is given by

$$f(x) = \frac{x^{\frac{r}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{r}{2}} \Gamma\left(\frac{r}{2}\right)} \qquad 0 \leq x < \infty, \; r > 0$$

where r is the degrees of freedom.

The characteristic function of the Chi-Square distribution is given by

$$\phi_x(t) = \frac{1}{(1-2it)^{r/2}}$$

Transforming to standard measure by $y = \frac{x-\mu}{\sigma}$, where for the Chi-Square distribution the mean is r and the variance is $2r$, it follows that

$$\phi_y(t) = e^{\frac{-it\mu}{\sigma}} \phi_x(t/\sigma)$$

$$= e^{\frac{-itr}{\sqrt{2r}}} \frac{1}{\left(1-2it/\sqrt{2r}\right)^{r/2}}$$

$$\log_e \phi_y(t) = \frac{-itr}{\sqrt{2r}} - \frac{r}{2} \log_e \left(1-\frac{2it}{\sqrt{2r}}\right)$$

Since $\log_e(1+x) \quad x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots$ and letting $\frac{-2it}{\sqrt{2r}} = x$

$$\log_e \phi_y(t) = \frac{-itr}{\sqrt{2r}} + \frac{r}{2}\left[\frac{-2it}{\sqrt{2r}} - \frac{(-2it)^2}{2(2r)} + \frac{(-2it)^3}{3(2r)^{3/2}} - \frac{(-2it)^4}{4(2r)^2} + \cdots\right]$$

$$= \frac{-t^2}{2} + \frac{\sqrt{2}it^3}{3r} + \frac{t^4}{2r} + \cdots$$

$$\lim_{r \to \infty} \log_e \phi_y(t) = \frac{-t^2}{2}$$

$$\lim_{r \to \infty} \phi_y(t) = e^{\frac{-t^2}{2}}$$

Thus in the limit, as the degrees of freedom become infinitely large, the Chi-Square distribution approaches the Normal distribution.

There are two approximations to the Chi-Square distribution that may be found, both of which are based upon the assumption of Normality. R.A. Fisher, Statistical Methods for Research Workers (p.81), notes that $\sqrt{2X^2}$ is approximately normally distributed with mean $\sqrt{2r-1}$ and unit variance. Wilson and Hilferty (1931) have found that $(X^{2}/r)^{1/3}$ is approximately normally distributed with mean $1-\frac{2}{9r}$ and variance $\frac{2}{9r}$. Garwood (1936) shows that the second approximation is the more accurate of the two, but that to obtain this degree of accuracy, more calculations are involved.

Student's "t" Distribution

The frequency function for the sampling distribution of Student's "t" is given by

$$f(t) = \frac{\Gamma\left(\frac{r}{2}+\frac{1}{2}\right)}{(\pi r)^{1/2}\,\Gamma\left(\frac{r}{2}\right)\left(1+\frac{t^2}{r}\right)^{\frac{r+1}{2}}} \qquad -\infty < t < \infty$$

$$r > 0$$

where r is the degrees of freedom.

Consider

$$\frac{-r+1}{2} \log_e \left(1+\frac{t^2}{r}\right) = -\frac{r+1}{2}\left[\frac{t^2}{r} - \frac{t^4}{2r^2} + \frac{t^6}{3r^3} - \frac{t^8}{4r^4} + \cdots\right]$$

$$= \frac{-t^2}{2} + \left\{\frac{2t^2+t^4}{4r}\right\} + \left\{\frac{3t^4-2t^6}{12r^2}\right\} + \left\{\frac{-4t^6+3t^8}{24n^3}\right\} + \cdots$$

Now, consider $\log_e \sqrt{\frac{2}{r}}\, \frac{\Gamma\left(\frac{r}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)}$ which through the use of the approximation to factorials $n! = \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$ can be shown to equal $\frac{-1}{4r} + \frac{1}{6r^2} + \frac{13}{24r^3}$.[1]

Now,

$$\log_e f(t) = \frac{-1}{2}\log_e(2\pi) + \log_e\left[\sqrt{\frac{2}{r}}\, \frac{\Gamma\left(\frac{r}{2}+\frac{1}{2}\right)}{\Gamma(r/2)}\right] + \frac{-1}{2}(r+1)\log_e\left\{1+\frac{t^2}{r}\right\}$$

and $\lim\limits_{r\to\infty} \log_e f(t) = \frac{-1}{2}\log_e(2\pi) + 0 \quad \frac{-t^2}{2}$

Finally, $\lim\limits_{r\to\infty} f(t) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}$

Thus, in the limit, as the degrees of freedom become infinitely large, Student's "t" distribution becomes the standardized Normal distribution.

The F (or z) Distribution

The ratio of two independent variates, each distributed as a Chi-Square distribution, has an F distribution. One half of the natural logarithm of the F distribution has the z distribution.

---

[1]See Appendix.

Thus the F distribution is given by:

$$g(f) = \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} f^{\frac{r_1}{2}-1}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)\left(1+\frac{r_1 f}{r_2}\right)^{\frac{r_1+r_2}{2}}} \qquad \begin{array}{l} r_1, r_2 > 0 \\ 0 < f < \infty \end{array}$$

where r_1 and r_2 are the degrees of freedom in the numerator and denominator of the ratio respectively. With the use of the transformation $z = \frac{1}{2}\log_e F$, the distribution may be obtained.

If the logarithm of the characteristic function of the z distribution is considered, it can be shown that the z distribution is approximately normally distributed with mean $\frac{1}{2}\left(\frac{1}{r_2}-\frac{1}{r_1}\right)$ and variance $\frac{1}{2}\left(\frac{1}{r_1}+\frac{1}{r_2}\right)$.

CHAPTER V

SOME ASPECTS OF THE NORMAL DISTRIBUTION

The normal distribution is used by those who carry out research in disciplines such as: Education, Psychology, and Agriculture. From past experience, these researchers have found that their data resembles the classical bell shaped curve. Some of the researchers, when they obtain non-normal data, attempt to fit their data to the normal frequency curve by whatever means of distortion they have at hand. The basic rule that researchers should follow is that statistical tests involving the normal distribution should be used only if the data in question are assumed to be a sample from a normally distributed population. Two tests of the normality of data are outlined in this chapter.

The Role of the Normal Distribution in Probability

The probability of occurrence of a certain event is equal to the ratio between the number of cases which are favorable to this event and the total number of cases. This is called the classical definition of probability and it may be shown to be related to a frequency curve. The definition requires that the value assigned by this ratio must lie between zero and unity. Similarly, the restriction that the sum of the frequencies total unity may be placed on the frequency function. The normal frequency function deals with this restriction by either the presence or absence of a parameter representing the total frequency. The frequency function

$$f(x) = \frac{N}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

subject to the restriction that the frequency total N is unity becomes

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

It follows that the distribution function for the normal curve

$$F(A) = \int_A \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$ when integrated over the total range A, equals unity. Since the normal curve is symmetrical about the axis $x = 0$, only one side of the curve need be tabulated. To determine the probability that a value lies between $x = 0$ and $x = x$, the integral

$$F(x) = \int_{x=0}^{x=x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$ must be evaluated.

As mentioned in Chapter III, this evaluation is simplified if the variable is standardized by the transformation $Z = \frac{x-\mu}{\sigma}$.

The integral becomes

$$F(z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

It is this integral that has been tabulated, and it is found in all standard textbooks.

In order to determine a probability, the data must first be standardized, and the standardized value looked up in a set of tables to determine the probability.

Fitting Data to the Normal Frequency Curve

To determine whether or not data fits a theoretical frequency curve, it is necessary to test an hypothesis. There are two methods of testing the fit of data to the normal frequency curve. The first being the Chi-Square goodness of fit test of the hypothesis.

H_o: the data fits the normal frequency curve.

This test is made at the level of significance $\alpha = 0.05$, which means that under the hypothesis being true, there is a risk of one in twenty that the data do not fit the normal frequency curve.

The test is of the form:

$$\chi^2_{k-r} = \sum \frac{(\text{Observed frequency} - \text{Expected frequency})^2}{\text{Expected frequency}}$$

where k - r is the number of degrees of freedom.

The second or alternative test of the fit of the data to the normal frequency curve is a two stage test based on the moments of the data. Here, tests are made to determine whether or not there is marked skewness or kurtosis, either one of which may indicate a deviation from normality. The two hypotheses involved are:

H_o: $\sqrt{\beta_1} = 0$ $\qquad$ H_o: $\beta_2 = 3$.

In the instance that either or both tests yield significant results, the data may be said to deviate from normality.

The distribution of results of final examinations in Grade XII English as presented in Table I will be used to illustrate the two methods of testing the fit of data to the normal curve.

Method of Areas and the Chi-Square Goodness of Fit Test

Since the Chi-Square test involves the expected frequencies, these must be determined. The data are grouped in intervals, each interval being associated with a frequency. The problem is to determine the probability of each interval using tables of the normal curve areas. This probability is multiplied by the total observed frequency to yield the expected frequencies. The steps that are necessary to obtain this result are outlined in Table II on page 54.

Both the observed and the expected frequencies are contained in Table II. With this information, the Chi-Square goodness of fit test may be applied. Prior to its application, those cells whose expected frequency is less than five are combined, since without the combination, the Chi-Square tends to become inflated. Each combination of two cells carries a loss of one degree of freedom.

The steps necessary for the calculation of the Chi-Square value are contained in Table III on page 56. This calculated Chi-Square indicates that the results of the English final examinations do not follow the normal curve. The table contains the Chi-Square values for each integral so that the individual discrepancies from normality may be noted.

TABLE I

THE DISTRIBUTION OF THE GRADE XII ENGLISH FINAL EXAMINATION RESULTS IN MANITOBA FOR THE YEARS 1962, 1963, AND 1964

Class Boundries	Class Centers	Observed Frequency
99.5 - 94.5	97	3
94.5 - 89.5	92	10
89.5 - 84.5	87	29
84.5 - 79.5	82	115
79.5 - 74.5	77	229
74.5 - 69.5	72	568
69.5 - 64.5	67	1008
64.5 - 59.5	62	1715
59.5 - 54.5	57	2275
54.5 - 49.5	52	3051
49.5 - 44.5	47	1478
44.5 - 39.5	42	2075
39.5 - 34.5	37	1134
34.5 - 29.5	32	575
29.5 - 24.5	27	195
24.5 - 19.5	22	74
19.5 - 14.5	17	25
14.5 - 9.5	12	7
9.5 - 4.5	7	3
4.5 - -0.5	2	2
$\bar{x} = 52.042$	$s = 11.5911$	$n = 14,571$

TABLE II

FITTING THE DATA FROM ENGLISH EXAMINATIONS TO THE NORMAL CURVE BY THE METHOD OF AREAS

Class Boundry	Class Center	Observed Frequency	Standard Score	Area	Difference	Expected Frequency
				.50000		
					.00003	.3
99.5			4.09	.49997		
	97	3			.00010	1.5
94.5			3.66	.49987		
	92	10			.00040	7.1
89.5			3.23	.49938		
	87	29			.00193	28.3
84.5			2.80	.49744		
	82	115			.00633	92.2
79.5			2.37	.49111		
	77	229			.01730	252.1
74.5			1.94	.47381		
	72	568			.03933	573.2
69.5			1.51	.43447		
	67	1008			.07678	1118.8
64.5			1.07	.35769		
	62	1715			.11878	1730.7
59.5			0.64	.23891		
	57	2275			.15524	2269.3
54.5			0.21	.08317		
	52	3051			.17023	2480.4
49.5			-0.22	.08706		
	47	1478			.15509	2259.8
44.5			-0.65	.24215		
	42	2075			.11778	1716.2
39.5			-1.08	.35993		
	37	1134			.07456	1086.4
34.5			-1.51	.43447		
	32	575			.03934	573.2
29.5			-1.94	.47381		
	27	195			.01753	255.4
24.5			-2.38	.49134		
	22	74			.00618	90.0
19.5			-2.81	.49752		
	17	25			.00206	30.0

TABLE II (continued)

FITTING THE DATA FROM ENGLISH EXAMINATIONS TO THE NORMAL CURVE BY THE METHOD OF AREAS

Class Boundry	Class Center	Observed Frequency	Standard Score	Area	Difference	Expected Frequency
14.5			-3.34	.49958		
	12	7			.49981	4.2
9.5			-3.67	.49987		
	7	3			.00010	1.5
4.5			-4.10	.49997		
	2	2			.00002	.3
.5			-4.53	.49999		
					.00001	.1
		14,571		.50000		14,571.0

TABLE III

STEPS INVOLVED IN TESTING THE GOODNESS OF FIT BY THE CHI-SQUARE GOODNESS OF FIT TEST

Class Interval	Observed Frequency	Expected Frequency	O - E	$\frac{(O - E)^2}{E}$
99.5 - 89.5	13	8.9	4.1	1.888
89.5 - 84.5	29	28.3	.7	0.0173
84.5 - 79.5	115	92.2	22.8	5.638
79.5 - 74.5	229	252.1	-23.1	2.116
74.5 - 69.5	568	573.2	-5.2	0.047
69.5 - 64.5	1008	1118.8	-110.8	10.973
64.5 - 59.5	1715	1730.7	-15.7	0.1424
59.5 - 54.5	2275	2269.3	5.7	0.0143
54.5 - 49.5	3051	2480.4	570.6	131.2628
49.5 - 44.5	1478	2259.8	-781.8	270.4713
44.5 - 39.5	2075	1716.2	358.8	75.013
39.5 - 34.5	1134	1086.4	47.6	2.0856
34.5 - 29.5	575	573.2	1.8	0.0057
29.5 - 24.5	195	255.4	-60.4	14.2841
24.5 - 19.5	74	90.0	-16	2.844
19.5 - 14.5	25	30.0	-5	.833
14.5 - .5	12	6.1	5.9	5.7066
	14,571			523.343

at $\alpha = 0.05$, $\chi^2_{14} = 23.685$

The deviation from normality may be explained when the assumptions for normality are considered. In particular, the assumption of the randomness of the variable x will be considered.

In all academic marking, since no mark is allowed between and including the range from 46 to 49, the interval 44.5 to 49.5 is purposely deflated. The adjoining intervals receive a corresponding inflation. The data also shows that the markers were unwilling to allow a mark below 25 unless they found that there was no other recourse. This caused a corresponding increase in the interval from 24.5 to 29.5. Although not to the same extent the interval from 64.5 to 69.5 has an excess of expected frequency over observed frequency. All of these deviations contribute to the highly significant Chi-Square value. The greatest departure from expected value in the interval 44.5 to 49.5 is due to the fact that there are no marks assigned from 46 to 49 on any of the papers.

Tests of Skewness and Kurtosis

The Chi-Square goodness of fit test is extremely powerful in that it detects deviations from normality in the moments as high as the sixth moment. A less powerful though valid procedure for determining departures from normality are the tests on skewness and kurtosis. These two tests are less powerful because they only involve the second, third and fourth moments while the Chi-Square tests deals with all of the moments simultaneously.

Using procedure outlined in Chapter III, the skewness may be found to be 0.0339468 and the kurtosis to be 3.0863714. The standard error of $\sqrt{\beta_1}$ is $\sqrt{6/n}$ and that of β_2 is $\sqrt{24/n}$. The form of the standard errors indicates that these tests are basically large sample tests and that they are valid only when n is large.

$$H_o: \sqrt{\beta_1} = 0 \qquad \alpha = 0.05$$

$$z = \frac{.0339468 - 0.000000}{\sqrt{6/(14,571)}}$$

$$= 1.6728961$$

The skewness of the data is moderate and the data may be said to be normal with respect to skewness.

$$H_o: \beta_2 = 3 \qquad \alpha = 0.05$$

$$z = \frac{3.0863714 - 3.000000}{\sqrt{24/(14,571)}}$$

$$= 2.1281836^*$$

The significance of the test on kurtosis indicates that the data departs from normality with respect to kurtosis. It means that there is an excess of items about the mean of the distribution than is expected under the assumption of normality. As outlined in the discussion of the significance of the Chi-Square goodness of fit test, this significance is due largely to the deliberate distortion of the data.

Non-Normality

For random samples drawn from a non-normal population, it has been found that serious mathematical difficulties are encountered when attempting to obtain the exact distribution of F under the null-hypothesis. Generally, all work on the effects of non-normality fall into one category being the effect of non-normality upon tests of significance. The investigations have shown that no serious error is introduced by non-normality in the significance of the F test or the two tailed t test. As a general rule the tabular probability values are underestimates. For example, the true probability level corresponding to the tabular five percent level may lie at any point between four and seven percent. The effect of non-normality therefore, is to make results look more significant than they really are.

Studies have been carried out as to the effect on efficiency when a non-normal population has been sampled. Fisher[1] has shown that the efficiency is very slightly affected when a moderately non-normal population is sampled.

Transformations

Transformations are used to overcome defects which are inherent in certain types of data. If the data are non-normal, researchers will often attempt to convert the data by

---

[1] "On the Mathematical Foundations of Theoretical Statistics." Fisher, R.A. 1922 Contributions to Mathematical Statistics, paper #10.

some transformation to a form the errors of observation of which are normally distributed. When a transformation is applied, it is hoped that in attempting to use familiar methods, no serious loss in efficiency is incurred. A well applied transformation is usually accompanied by a gain in the efficiency of the estimators. In the instance that a transformation causes a loss in efficiency, some other means of analysis should be considered.

The Assumption of Normality Underlying Tests of Significance

Distributions on which tests of significance are based have been shown to have the normal frequency function as a limiting form. In order to make any valid tests of significance, the distribution of the errors of observation should therefore be assumed to be a sample from a normally distributed population of errors. The assumption is necessary only for tests of significance because up to the point of testing, only descriptive statistical techniques have been used.

An example of a technique which requires the assumption of normality for any tests of significance is the analysis of variance technique. If the procedures of analysis of variance are to be used only to the point of summarizing the data, then assumptions of any type are not necessary. On the other hand, if the analysis of variance techniques are used as a method of statistical inference, for making inferences about some population of values, then assumptions about the population from which the data were drawn must be made. In particular, if valid inferences based upon tests of significance are to be made, then the assumption of the normality of the errors of observation must be included.

Non-Parametric Methods

Methods have been developed that do not depend upon a particular form of the frequency function. The only assumption necessary is that the frequency function be continuous. The methods are not concerned with testing any particular parameter, and because of this, have been termed non-parametric methods. The first non-parametric method developed was the Chi-Square goodness of fit test as devised by Karl Pearson in 1900.

If the data are normally distributed, both non-parametric and parametric methods may be used as valid bases for statistical inferences. However, the efficiency of the estimators is greater for the parametric than for the non-parametric case. Non-parametric methods and tests deal with the problem by stating it generally, and then attempting to determine a solution using methods which do not involve known distributions. This class of probability distributions is so large that it cannot be indexed by parameters. The necessity of the generalization dictates that the estimators be less efficient than those used in the parametric methods.

When data are found to be non-normal, careful choice of a transformation may correct the sample data so that it appears to have been drawn from a normally distributed population. Should the transformation approach fail, then if any sense is to be gleaned from the data, researchers must revert to non-parametric methods of analysis.

BIBLIOGRAPHY

A. BOOKS

Aitken, A.C., Statistical Mathematics. Edinburgh: Oliver and Boyd, 1939.

Arley, N., and Buch, K.R., Introduction to the Theory of Probability and Statistics. New York: John Wiley and Sons, 1950.

Bowley, A.L., Elements of Statistics. Fourth edition; London: 1920.

Brunt, David, The Combination of Observations. Cambridge: 1917, pp. 11-24.

Cajori, Florian, A History of Mathematics. The Macmillan Company of New York, 1919.

Cornell, F.G., Essentials of Educational Statistics. New York: John Wiley and Sons, 1956, Chapter 6.

Cramer, Harald, Mathematical Methods of Statistics. Princeton University Press, 1950.

________, The Elements of Probability Theory. New York: John Wiley and Sons, 1955, pp. 96, 108-114.

Feller, William, An Introduction to Probability Theory and Its Applications. Second edition; New York: John Wiley and Sons, 1957, p. 164.

Fisher, Arne, The Mathematical Theory of Probabilities. The Macmillan Company, 1923.

________, Frequency Curves. The Macmillan Company, 1922.

Fisher, R.A., Contributions to Mathematical Statistics. New York: John Wiley and Sons, 1950.

________, Statistical Methods for Research Workers. Thirteenth edition; New York: Hafner Publishing Company, Inc., 1958.

Fraser, D.A.S., Nonparametric Methods in Statistics. New York: John Wiley and Sons, 1957.

Freeman, Harold, Introduction to Statistical Inference. Addison Wesley Publishing Company, 1963.

Hoel, P. G., Introduction to Mathematical Statistics. Third edition; New York: John Wiley and Sons, 1962.

Hogg, R. V., and Craig, A. T., Introduction to Mathematical Statistics. New York: Macmillan Company, 1959.

Kapur, J. N., and Saxena, H. C., Mathematical Statistics. Second edition; Delhi: S. Chand and Company, 1961.

Keeping, E. S., Introduction to Statistical Inference. Toronto: D. Van Nostrand Company, Inc., 1962.

Kendal, M. G., and Stuart, A., The Advanced Theory of Statistics. Vol I "Distribution Theory". New York: Hafner Publishing Company, 1958.

Kenny, J. K., and Keeping, E. S., Mathematics of Statistics Part II. New York: D. Van Nostrand Company, Inc., 1951.

Laplace, P. S., Theorie Analytique des Probabilities. 1820.

______, A Philosophical Essay on Probabilities. Dover, 1951.

Levy, H., and Roth, L., Elements of Probability. Oxford: Clarendon Press, 1936.

McNemar, Quinn, Psychological Statistics. New York: John Wiley and Sons, 1949.

Plummer, H. C., Probability and Frequency. London: The Macmillan Company.

Reitz, H. L., "Mathematical Statistics." Carus Mathematical Monographs #3. Chicago: Mathematical Association of America, Open Court Publishing Company, 1927.

Scarborough, James, B., Numerical Mathematical Analysis. Baltimore: The Johns Hopkins Press, 1930.

Simpson, T., "An Attempt to Show the Advantage Arising by Taking the Mean of a Number of Observations in Practical Astronomy." Miscellaneous Tracts on Some Curious Subjects. London: 1757, pp. 64-75.

Smith, David Eugene, A Source Book of Mathematics. McGraw-Hill, 1929, pp. 567-75.

Snedecor, G. W., Statistical Methods. Iowa State College Press, 1937, Chapter 5.

Tate, Merle, W., Statistics in Education. New York: The Macmillan Company, 1955.

Todhunter, I., A History of the Mathematical Theory of Probability. Second edition; New York: Chelsa Publishing Company, 1949.

Topping, J., Errors of Observations and Their Treatment. London: The Institute of Physics, 1955.

Wert, J.E., Neidt, C.O., and Ahmann, J.S., Statistical Methods in Educational and Psychological Research. New York: Appleton - Century - Crofts Inc., 1954.

Whittaker, E.T., and Robinson, G., The Calculus of Observations. Blackie and Son, 1924, pp. 165-180.

Woodward, Robert S., "Probability and Theory of Errors." Mathematical Monographs #7. New York: John Wiley and Sons.

Wright, Thomas Wallace, and Hayford, John Filmore, The Adjustment of Observations. New York: D. Van Nostrand Company, 1906.

B. PERIODICAL ARTICLES

Abbe, Cleveland, "A Historical Note on the Method of Least Squares." Amer. Journ. Sci., 1:411-15, 1871.

Adrain, R., "Research Concerning the Probabilities of the Errors Which Happen in Making Observations," Analyst or Mathematical Museum, 1:93-109, 1808.

Aroian, L.A., "The Probability Function of the Product of Two Normally Distributed Variables," Annals of Math. Stat., 18:265, 1947.

______, "A Study of R.A. Fisher's Z-distribution and the Related F distribution," Annals of Math. Stat., 12:429, 1941.

Baker, G.A., "Transformations of Non-normal Frequency Distributions Into Normal Distributions," Annals of Math. Stat. 5:113, 1934.

Bartlett, M.G., "The Use of Transformations," Biometrics, 3:39, 1947.

Cochran, W.G., "Note on An Approximate Formula for the Significance Levels of z," Annals of Math. Stat., 11:93, 1940.

_______, "Some Consequences When the Assumptions for the Analysis of Variance are Not Satisfied," Biometrics, 3:22, 1947.

Craig, C.C., "On a Property of the Semi-Invariants by Thiele," Annals of Math. Stat., 2:154, 1931.

Curtiss, J.H., "On Transformations Used in Analysis of Variance," Annals of Math. Stat., 14:107, 1943.

Deming, W.E., and Birge, R.T., "On the Statistical Theory of Errors," The Graduate School of the United States Department of Agriculture, Washington. Reprinted from the Reviews of Modern Physics, 6:119-61, July, 1934, with additional notes dated 1937.

De Moivre, A., "Approximatio ad Summam Terminorium Binomii $(a+b)^n$ in Seriem Expansi," Second Supplement to Miscellania Analytica, pp. 1-7, November, 1733.

Edgeworth, F.Y., "The Law of Error," Cambridge Phil. Trans., 20:36-65, 113-41, 1904.

Elderton, W. Palin, "Tables for Testing the Goodness of Fit to the Theory of Observations," Biometrika, 1:155, 1901-02.

Feller, W., "On the Normal Approximation to the Binomial Distribution," Annals of Math. Stat., 16:319, 1945.

Finney, D.J., "On the Distribution of a Variate Whose Logarithm is Normally Distributed," Jour. Roy. Stat. Soc. Suppl., 7:155-61, 1941.

Friedman, M., "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," Journ. Amer. Stat. Assoc., 32:675, 1937.

Glaisher, James Whitbread Lee, "Law of Facility of Errors of Observations and on the Method of Least Squares," Mem. Roy. Astro. Soc. 39, 1892.

Herschell, J. (Sir), "Quetelet on Probability," Edinburgh Review, xcii:1-57, 1850.

Hey, G.B., "A New Method of Experimental Sampling Illustrated on Certain Non-Normal Populations," Biometrika, 30:68-80, 1938.

Kaplansky, I., "A Common Error Concerning Kurtosis," Journ. Amer. Stat. Assoc., 40:259, 1945.

Lukacs, E., "A Characterization of the Normal Distribution," Annals of Math. Stat., 13:91, 1942.

Merriman, M., "List of Writings Relating to the Method of Least Squares with Historical and Critical Notes," Trans. Connecticut Acad., 4:151-232, 1877.

Nixon, J.W., "An Experimental Test of the Normal Law of Error," Journ. Roy. Stat. Soc., 176:702-6, 1913.

Pearson, E.S., "The Analysis of Variance in Cases of Non-normal Variation," Biometrika, 23:114, 1931.

———, "A Further Development of Tests for Normality," Biometrika, 22:239, 1930.

Pearson, Karl, "Notes on the History of Correlation," Biometrika, 13:25, 1921.

———, "A Historical Note on the Origin of the Normal Curve of Errors," Biometrika, 16:402, 1924.

Simpson, Thomas, "A Letter to the Right Honourable George, Earl of Macclesfield, President of the Royal Society as to the Advantage Arising by Taking the Mean of a Number of Observations in Practical Astronomy," Phil. Trans. xlix:82-93, 1755.

Wilkins, J.E. Jr., "A Note on Skewness and Kurtosis," Annals of Math. Stat., 15:333, 1944.

Wilson, E.B., and Hilferty, M.M., "The Distribution of the Chi-Square," Proc. Nat. Acad. Sci. 17:684, 1931.

APPENDIX

The expansion of the expression $\log R = \log\left\{\sqrt{\frac{2}{v}}\,\frac{\Gamma\left(\frac{v}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}\right\}$

into a convergent infinite series

Note:

$$1)\ \Gamma(1+x) = x!, \quad x \text{ an integer} \geq 0$$

$$2)\ \log x! = \frac{1}{2}\log(2\pi) + \left(x+\frac{1}{2}\right)\log x - x,\ x \neq 0$$

$$3)\ \log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots, \quad -1 < x < 1$$

$$\log R = \log\sqrt{\frac{2}{v}} + \log\left[\Gamma\left(\frac{v}{2}+\frac{1}{2}\right)\right] - \log\Gamma\left(\frac{v}{2}\right)$$

Using relations 1) and 2)

$$\log\Gamma\left(\frac{v}{2}+\frac{1}{2}\right) = \log\Gamma\left[\left(\frac{v}{2}-\frac{1}{2}\right)+1\right] = \frac{v}{2}\log\left(\frac{v}{2}-\frac{1}{2}\right) - \left(\frac{v-1}{2}\right) + \frac{1}{2}\log(2\pi)$$

$$\log\Gamma\left(\frac{v}{2}\right) = \log\Gamma\left[\left(\frac{v}{2}+1\right)+1\right] = \frac{v-1}{2}\log\frac{v-2}{2} - \frac{v-2}{2} + \frac{1}{2}\log(2\pi)$$

The difference between the last two terms of log R becomes

$$\log\left(\frac{v}{2}+\frac{1}{2}\right) - \log\left(\frac{v}{2}\right)$$

$$= \frac{v}{2}\log\left(\frac{v-1}{2}\right) - \left(\frac{v-1}{2}\right) + \frac{1}{2}\log(2\pi) - \left(\frac{v-1}{2}\right)\log\left(\frac{v-2}{2}\right) + \left(\frac{v-2}{2}\right) - \frac{1}{2}\log(2\pi)$$

$$= \frac{v}{2}\left[\log(v-1) - \log(2)\right] - \left(\frac{v-1}{2}\right)\left[\log(v-2) - \log(2)\right] - \frac{1}{2}$$

$$= \frac{v}{2}\left[\log(v-1)\right] - \left(\frac{v-1}{2}\right)\left[\log(v-2)\right] - \left[\frac{v}{2} - \left(\frac{v-1}{2}\right)\right]\log 2 - \frac{1}{2}$$

$$= \frac{v}{2}\left[\log v + \log\left(1-\frac{1}{v}\right)\right] - \left(\frac{v-1}{2}\right)\left[\log v - \log\left(1-\frac{2}{v}\right)\right] - \log\sqrt{2} - \frac{1}{2}$$

Now, the expression for log R is

$$\log R = \log\sqrt{2} - \frac{1}{2}\log v + \frac{v}{2}\left[\log v + \log\left(1-\frac{1}{v}\right)\right]$$

$$- \left(\frac{v-1}{2}\right)\left[\log v - \log\left(1-\frac{2}{v}\right)\right] - \log\sqrt{2} - \frac{1}{2}$$

$$= \frac{1}{2}\left[v\log\left(1-\frac{1}{v}\right) - (v-1)\log\left(1-\frac{2}{v}\right)\right] - \frac{1}{2}$$

The expansion of the logarithms using relation 3) yields

$$\log R$$

$$= \frac{1}{2}\left[v\left\{-\frac{1}{v}-\frac{\left(\frac{1}{v}\right)^2}{2}-\frac{\left(\frac{1}{v}\right)^3}{3}-\frac{\left(-\frac{1}{v}\right)^4}{4}+\cdots\right\}-(v-1)\left\{-\frac{2}{v}-\frac{\left(-\frac{2}{v}\right)^2}{2}+\frac{\left(-\frac{2}{v}\right)^3}{3}-\frac{\left(-\frac{2}{v}\right)^4}{4}+\cdots\right\}\right]-\frac{1}{2}$$

$$= \frac{1}{2}\left[v\left\{-\frac{1}{v}-\frac{1}{2v^2}-\frac{1}{3v^3}-\frac{1}{4v^4}-\cdots\right\}-(v-1)\left\{-\frac{2}{v}-\frac{4}{2v^2}-\frac{8}{3v^3}-\frac{16}{4v^4}-\cdots\right\}\right]-\frac{1}{2}$$

$$= \frac{1}{2}\left[\left\{-1-\frac{1}{2v}-\frac{1}{3v^2}-\frac{1}{4v^3}-\cdots\right\}-\left\{-2+\frac{2}{v}-\frac{4}{2v}+\frac{4}{2v^2}-\frac{8}{3v^2}+\frac{8}{3v^3}-\frac{16}{4v^3}+\frac{16}{4v^4}-\cdots\right\}\right]-\frac{1}{2}$$

$$= \frac{1}{2}\left[-1+2\right]-\frac{1}{2}+\frac{1}{2}\left[-\frac{1}{2v}-\frac{2}{v}-\frac{4}{2v}\right]+\frac{1}{2}\left[-\frac{1}{3v^2}-\frac{4}{2v^2}+\frac{8}{3v^2}\right]+\frac{1}{2}\left[-\frac{1}{4v^3}-\frac{8}{3v^3}+\frac{16}{4v^4}\right]+\cdots$$

$$= -\frac{1}{4v}-\frac{1}{6v^2}-\frac{13}{24v^3}-\cdots$$

Therefore, the expression $\log\left\{\sqrt{\frac{2}{v}}\;\frac{\Gamma\left(\frac{v}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}\right\}$ may be expressed as the infinite convergent series $-\frac{1}{4v}-\frac{1}{6v^2}-\frac{13}{24v^3}-\cdots$, $|v|>2$.

It should be noted that this series converges, but that this convergence is a very slow process.