A HISTORICAL AND MATHEMATICAL DEVELOPMENT OF STUDENT'S t-DISTRIBUTION

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
Edward George Marlin
February 1966

ABSTRACT

In 1908, "Student" published his investigation of a method of making exact probability statements by which the significance of means of small samples, drawn from normal populations, could be determined. Since that date, a large body of statistical theory, built on this foundation, has been developed, and has seen application in almost every area of modern scientific experimentation.

The purpose of this paper is to present an integrated view of the relationship between the historical development of uncertain inference with regard to small sample work, and "Student's" t-distribution. The distribution is derived, and its properties described. Some applications to tests of significance are presented, and some of the major areas of current research on problems related to this distribution are indicated.

TABLE OF CONTENTS

CHAPTER I

THE PROBLEM OF SMALL SAMPLES

Throughout the history of Mathematics, many attempts have been made to resolve the uncertainty associated with inductive inference. Some attempts, such as Bayes' Theorem and developments arising from the normal law of error, have provided partial solutions, but these solutions were of limited application due to the necessity of fulfilling rather restrictive conditions.

One notable lack in this regard was in the area of small sample work, a situtation commonly occurring in statistical investigations, where variation from sample to sample precluded reliable estimates of the population variance. Large sample methods where precise estimates of this variance could be obtained, or situations where this variance is known by previous experience, were available, but these were often impractical due to excessive cost, complexity, or to the impossibility of repeating an experiment.

In 1908, however, W. S. Gosset, writing under the pen-name of "Student", developed a test criterion for the mean independent of the population variance, and provided the first exact test of significance for small samples from normal populations. His contribution, while directly useful as a practical tool, was much more important in that it imparted a spectacular impetus to the field of statistics, leading to a successful generalization of his work which formed a basis for much of modern statistical theory.

The purpose of this paper is to present a comprehensive examination of Student's t-distribution from the point of view of its history, derivation, properties, applications and generalizations.

Historically, an account of the developments leading up to the t-distribution and of subsequent work, is presented in order to show the importance of this distribution as a method of inference and a major contribution to modern statistics.

Mathematically, the probability density function of this distribution is obtained by analytical and geometrical methods. A general form of the function is obtained, and the applications of the distribution are considered as particular cases of this general form.

Consideration is then given to current research on problems arising in investigations where certain assumptions regarding the distribution of the parent population and the assumption of a common variance in tests of significance are not met.

Lastly, a presentation of more general forms arising directly from the extension of the distribution is made, together with their applications.

CHAPTER II

HISTORICAL DEVELOPMENT OF THE t-DISTRIBUTION

Suppose there is a set of data concerning, for example, all events which happen under a given set of conditions. For these data to have any practical meaning, this population necessarily embodies a certain quantity of information. Now consider some proper subset of this population. If the information in this subset, or sample, is measured in the same way that the information contained in the population was measured, then it is apparent that not all of the information in the population is contained in the sample.

Thus, to reason deductively, that is, from a well defined population to a sample, logical statements may be made because all the information needed to define the population is known and consequently the component parts are known. There is, of course, some uncertainty inherent in the situation in that there is no way of knowing a priori which specific sample of all those samples possible will be obtained. This uncertainty cannot be removed, but the concept of probability provides a quantative measure of the uncertainty, and it can be dealt with rigourously. On the other hand, an absolute statement, where one reasons from sample to population, cannot be made. To do so would imply that the amount of information in the sample is identical to the amount of information in the population, which contradicts the premise that the sample is a <u>proper</u> subset of the population. Such a statement would further imply that all the information about a population is contained in any one observation, thus denying the existence of any inherent differences between one member of a population and another .

Since insufficient information about the population is available, classical probability theory can be of no help in resolving the uncertainty associated with inductive inference.

However, in the early 17th century, the theory of probability was in its infancy and was expected to be a very powerful weapon for attacking the problem. Fisher, outlining the history leading to recent developments in the logic of inductive reasoning noted:

> "For centuries, however, it was assumed that if uncertain inferences were to be made, they must be made in terms of mathematical probability. It was, I believe, this assumption, more than any other factor, which has led to efforts to define probability in more general, and usually in psychological terms, and has introduced infinite confusion into the use of this once well defined concept."[1]

1. Fisher[6]: p. 246

Bayes' Theorem

Thomas Bayes, recognizing the fundamental importance of this problem of uncertain inference, particularly in the era of blossoming scientific endeavour, attempted to bring inductive logic within the realm of deductive reasoning. He framed an axiom, defining a super-population, from which all possible types of populations had been drawn as samples. The latter populations supplied the information needed to apply classical probability theory in order to determine the probability that an observed sample was drawn from a particular population.

Bayes did not publish his work pending clarification of certain doubts regarding the validity of his axiom. These doubts were not cleared up apparently as the treatise was published posthumously by Price. Whatever the doubts, upon its appearance, the axiom was given a prominent place in the mathematics of the day perhaps because, as Fisher suggests, it met a very real need, and its unquestioned acceptance by some of the great names of mathematics gave the axiom an aura of authority.

No serious criticisms of the axiom arose for nearly 90 years, until the appearance of Boole's "Laws of Thought" in 1854. Boole noted that Bayes' axiom was a device for supplying, by means of an arbitrary hypothesis, information that the data lacked. His criticism was somewhat diffident, but more decisive rejection of the theory of inverse probability by Venn and Chrystal followed.

The following example of inference based on Inverse Probability, due to Edgeworth,[1] and cited by Welch,[1] will illustrate Boole's criticisms.

Let $x_1, x_2, \ldots x_n$ be a random sample of size n from a distribution which is $N(\mu, \sigma^2)$. If the sample is S, then the probability of the sample given μ and σ is

$$f(S\mid\mu,\sigma)\,dS = \frac{e^{-\frac{1}{2}\sigma^2 \sum_{i=1}^{n}(x_i-\mu)^2}\, dx_1, dx_2, \ldots dx_n}{(2\pi)^{n/2}\,\sigma^n}$$

Now by the definition of conditional probability

$$f(S\mid\mu,\sigma) = h(S,\mu,\sigma)/\,g(\mu,\sigma)$$

where the quantities on the right-hand side of this equation must be regarded as distributions of the probability that the parameters μ and σ take on specific values. If μ and σ are merely unknown constants, they can have no distribution other than a trivial one under the frequency definition of probability.

By Bayes' Theorem, the posterior distribution of μ and σ given that S has occurred is

$$P(\mu,\sigma|S)\,d\mu\,d\sigma = \frac{f(S|\mu,\sigma)\ g(\mu,\sigma)\ dS\,d\mu\,d\sigma}{\int_{-\infty}^{\infty} f\ (S|\mu,\sigma)\ g(\mu,\sigma)\quad d\mu\,d\sigma}$$

$$= f(S|\mu,\sigma)\ g(\mu,\sigma)\ dS\ d\mu\,d\sigma.$$

The probability distribution of μ given S is then

$$P(\mu|S)\ d\mu = \int_{\sigma=0}^{\infty} f(S|\mu,\sigma)\ \ g(\mu,\sigma)\ \ dS\ d\mu\,d\sigma.$$

But $p(\mu|S)$ is clearly dependent on $g(\mu,\sigma)$, the distribution of the probability that μ and σ each take on specified values. Choices of $g(\mu,\sigma)$ will obviously affect $p(\mu|S)$.

This is the key point of the criticism levelled at inverse probability. The prior distribution of μ and σ is seldom known. If it is not known, then this distribution is completely arbitrary as there is "insufficient reason" for the choice of one particular distribution over another. As Boole stated:

> "These results only illustrate the fact, that when the defect of data is supplied by hypothesis, the solution will, in general, vary with the nature of the hypothesis assumed; so that the question remains, only more definite in form, whether the principles of probabilities serve to guide us in the election of such hypotheses."[2]

In spite of the criticism, however, the axiom and the theory of statistical inference derived

2. Boole: Laws of Thought quoted by Fisher[6] p. 247

from it, retained a tenacious grip on its place in mathematics well into the twentieth century, becoming a point of bitter contention between two major figures in modern statistical theory. Fisher, an opponent of inverse probability in this controversy, attributes the retention of the theory to the fact that the mathematical world had nothing better with which to replace inverse probability (especially when it led to plausible conclusions in the case for which Bayes had developed it, specifically, for a finite set of exhaustive, mutually exclusive outcomes where the probability of any outcome could be determined by the frequency definition of probability), and to the mathematician's inexperience at conducting orderly retreats from false positions.

Inadequacy of Large Sample Method

An alternative approach to the problem of inference about the means of small samples from normal populations is the use of the quantity,

$$\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

as a test criterion.

If $\bar{x}$ is the observed mean of n observations, and μ the true mean of the population from which the sample was drawn, then it has long been known that $\bar{x}$ is distributed in different samples as the normal distribution, with its center at μ, and variance one nth of that of the population sampled. It follows that

$$\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

is normally distributed with mean zero and unit variance, and if σ^2, the true population variance is known, then exact probability statements about the population mean can be made. In practice, however, this variance is seldom known. It may be estimated by

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

but

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

does not necessarily follow the standard unit normal distribution. Lacking further knowledge of the distribution of this quantity, it was used as a normally distributed test criterion when n was large, for, as n increases, the population variance is more precisely estimated and consequently s^2 approaches the true value of σ^2. It will be shown in Chapter IV that the limiting form of this quantity as n becomes large is, in fact, the normal distribution.

Distribution of s^2

The next step was to investigate the sampling distribution of s, the estimate of the population standard deviation. This was first obtained by Helmert in 1876 and later reproduced by Czuber. These sources, however, were unknown to the English speaking statistical world.

In 1908 "Student" obtained the distribution of s^2 independently of Helmert's work. While Helmert's result was obtained analytically, "Student" obtained his result empirically by calculating the first four moments of the sampling distribution of s^2 and inferring a Pearson Type III curve. He then verified this by means of a series of sampling experiments.

Student's z

In this same paper, "Student" derived the probability density function f(z) of the quantity

$$z = \frac{\bar{x} - \mu}{s}.$$

He found that the distribution of z was independent of the population variance, and thus, tables of this function could be used to make probability statements about the difference between the observed mean of a sample and the true mean of the population from which the sample had been drawn, irrespective of the variance. If an hypothesized value of the true mean were μ_0, say, and

$$z_0 = \frac{\bar{x} - \mu_0}{s}$$

then

$$\int_{z_0}^{\infty} f(z)\ dz$$

was the probability of observing a sample mean larger than $\bar{x}$ if μ_0 were in fact the true mean of the population.

Fisher's Extension of Student's Work

"Student's" empirical result was obtained analytically by Fisher[1] in 1923 in the slightly modified form of $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{z}{\sqrt{n}}$

The density function was found to be

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

where n-1 was referred to as the degrees of freedom. Further, he showed that this modified distribution could be written as a statistic which is distributed as the ratio of a standard unit normal variate to s/σ, that is

$$t = \frac{N(0,1)}{s/\sigma} = \frac{N(0,1)}{\sqrt{V^2/(n-1)}}$$

where $s = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n-1}}$

and V^2 is a chi-square with n-1 degrees of freedom and t has Student's t-distribution with n-1 degrees of freedom.

With this more general expression for t, numerous tests of significance were developed encompassing a wide variety of problems. These will be discussed in Chapter V.

Confidence Intervals and Fiducial Intervals

One further development should be mentioned in connection with uncertain inference to conclude this discussion. It is, by no means, the last word in any discussion of the large topic of statistical inference, but it does represent a successful step in the search for a valid inductive inference. This development is confidence intervals, put forth by Neyman. While it is not a direct consequence of the t-distribution, its development can be attributed to the general advance in statistical theory which followed Student's work. Confidence intervals shall be considered in relation to the t-distribution, although the concept may be applied to any well-defined sampling distribution, and, indeed, has been applied to a wide variety of problems. Consider the test of the hypothesis $H_o: \mu = \mu_o$ against the alternative $H_1: \mu \neq \mu_o$.

Now,

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

in different samples will follow "Student's" t-distribution with n-1 degrees of freedom. The test will be judged to be significant if the difference between $\bar{x}$ and μ_o is sufficiently large to have a small probability of occurence. This probability (by convention less than 0.05 or 0.01) is known as the level of significance of the test and is usually designated α, and the fixed value of t, say t_α, associated with this level of significance is referred to as the critical value. These conditions may be written as

$$Pr\ (\ |t| > t_\alpha\ ) = \alpha$$

or

$$Pr\ (\ |t| < t_\alpha\ ) = 1 - \alpha.$$

Expanding this inequality, and substituting for t, then

$$Pr\ \left(\ -t_\alpha < \frac{\bar{x} - \mu_o}{s/\sqrt{n}} < t_\alpha\ \right) = 1 - \alpha$$

Rearrangement yields

$$Pr\left(\bar{x} - \frac{t_\alpha s}{\sqrt{n}} < \mu_o < \bar{x} + \frac{t_\alpha s}{\sqrt{n}}\right) = 1 - \alpha.$$

Thus, an interval has been calculated which covers the population mean, μ_o, with probability $1 - \alpha$.

A similar theory was being developed at approximately the same time by R. A. Fisher[5], which, in the case of estimation of a single parameter, led to the same result as Neyman's confidence limits.

Now, if $x_1, x_2, \ldots x_n$ is a random sample from $N(\mu, \sigma^2)$, then

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

is distributed in different samples as a distribution dependent on the sample size n. There exists a fixed value of t, say t_1, such that

$$\Pr \quad (t > t_1) = \alpha$$

where $0 < \alpha < 1$ is any specified probability. Further, from the above expression for t, t is a continuous function of the unknown parameter μ.

Upon substitution for t, the expression

$$\Pr \quad (t > t_1) = \alpha$$

becomes

$$\Pr \left[\frac{\bar{x} - \mu}{s/\sqrt{n}} > t_1 \right] = \alpha$$

or

$$\Pr \left[\mu < \bar{x} - \frac{st_1}{\sqrt{n}} \right] = \alpha .$$

Thus, by considering all possible values of t_1 (and the associated probability α), a probability distribution for μ can be constructed. To distinguish this distribution from Inverse Probability, Fisher termed it a fiducial distribution, and the limits of an interval which contains the parameter with a specified fiducial probability as fiducial limits.

Unlike Neyman's confidence limits, fiducial limits are not based on the idea that the assertion that a parameter is contained in a given interval is true in an assigned proportion, α, of cases in the long run. Instead, it is the range of conceivable values of the parameter which give rise to the observed statistic(s) with probability α.

A fiducial distribution is a probability distribution in the frequency sense of the word, only in that probabilities based on the frequency definition are attached to possible values of the parameter to be estimated. As pointed out in connection with Bayes' Theorem, a probability distribution of a parameter is incompatible with the frequency definition of probability. Hence the need for qualification.

A further distinction between Neyman's and Fisher's theories is that fiducial distributions admit only one set of fiducial limits. A fiducial distribution of a parameter is the one and only distribution of that parameter. A different distribution is possible using other estimates, but the existence of two different distributions of a parameter based on the same information is obviously self-contradictory. To avoid this, Fisher restricts its use to sufficient estimators.

The difference between these theories of inference and that of the Bayesian school of inference is that Neyman and Fisher rely on sampling distributions of observed statistics, rather than distributions independent of the sample work. Fisher and Bayes arrive at the same goal, a distribution of a parameter, from which inferences can be made, but their respective probability distributions and the question of validity are quite distinct.

Historical Importance of "Student's" Work

In a biographical paper on Gosset shortly after his death, E. S. Pearson noted:

> "It is probably true to say that this investigation published in 1908 has done more than any other single paper to bring these subjects within the range of statistical inquiry; as it stands it has produced an essential tool for the practical worker, while on the theoretical side it has proved to contain the seeds of new ideas which have since grown and multiplied a hundred fold..." [3]

Various authors have commented that Gosset's primary intention was practical--to develop a tool. The t-test is certainly that, and one of theoretical import as well, as has just been noted. In spite of this importance, however, it is interesting to note that this influence was not felt for some time. Although the test was in use immediately at Guiness Brewery in Dublin (Gosset was employed there

3. Pearson, E. S.: p. 224

as a brew-master from 1899 until his death in 1937), it did not come into common use until the early 1920's.

CHAPTER III

THEORETICAL DEVELOPMENT OF "STUDENT'S" t-DISTRIBUTION

In this chapter, "Student's" method of obtaining the t-distribution will be outlined and two rigourous developments of the distribution will be presented.

The first method is an analytical one in which the joint distribution of the sample mean, $\bar{x}$, and the sample variance, s^2, is established. The functional form of t is then obtained by a transformation of variable.

The second method is a parallel development of the first in that the joint distribution of $\bar{x}$ and s^2 is obtained, but by geometrical considerations. This development merits consideration for two reasons - first, it is the method presented by Fisher and thus represents the first rigourous derivation of the t-distribution and second, the representation of a sample as a point in n-dimensional Euclidean hyperspace constituted an important advance in the analysis of sampling problems.

"Student's" Approach

Although the t-distribution bears the name of "Student", the distribution in its present form is due to R. A. Fisher, who rigourously derived and extended the result presented by "Student" in 1908.

As has already been noted, "Student" obtained the distribution of the variable

$$z = \frac{\bar{x} - \mu}{s},$$

the integral of which was tabulated in terms of the sample size, n. His work was essentially empirical as he calculated the first four moments of the sampling distribution of z, from these quantities inferring that the frequency curve was that of a Pearson Type VII. After showing the correlation between $\bar{x}^2$ and s^2 to be zero, he further inferred that $\bar{x}$ and s were independent, a conclusion which does not necessarily follow such demonstration. His intuition was remarkable, as both inferences were later shown to be correct.

The transition from "Student's" z to "Student's" t was effected in collaboration with R. A. Fisher[3] in 1925. Fisher's research showed that the distribution had much wider application than "Student" had realized and that the form

$$t = \frac{z}{\sqrt{n}}$$

tabulated in terms of the degrees of freedom, n-1, rather than the sample size, n, was a more convenient one.

Independence of $\bar{x}$ and s^2

To show the independence of these quantities it is necessary to establish the following theorems.

Theorem III-1

Let x be a random variable which is $N(\mu, \sigma^2)$.

If $$V = \frac{x - \mu}{\sigma} ,$$

then V is N(0,1).

Proof:

The moment generating function of V is

$$E(e^{tV}) = \int_{-\infty}^{\infty} e^{t\frac{(x-\mu)}{\sigma}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx, \quad -\infty < x < \infty$$

Completing the square and setting

$$y = \frac{x-\mu}{\sigma} + t$$

then $$E(e^{tV}) = e^{t^2/2} \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = e^{t^2/2}$$

Thus V is N(0,1).

Theorem III-2:

Let x be a random variable which is $N(\mu, \sigma^2)$.

If $V = \frac{x-\mu}{\sigma}$, then V^2 has a chi-square distribution with 1 degree of freedom.

Proof

The moment generating function of V^2 is

$$E(e^{tV^2}) = \int_{-\infty}^{\infty} e^{t\frac{(x-\mu)^2}{\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \quad -\infty < x < \infty$$

Setting $y = \frac{(x-\mu)}{\sigma}\sqrt{1-2t}$, then

$$E(e^{tV^2}) = \frac{1}{\sqrt{1-2t}}\int_{-\infty}^{\infty}\frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy = (1-2t)^{-\frac{1}{2}}, \qquad -\infty < y < \infty$$
$$t < 1/2$$

which is the moment generating function of a chi-square variable with 1 degree of freedom. For notational simplicity, the use of V^2 shall be restricted throughout to indicate a variable which follows a chi-square distribution.

Theorem III-3 (Reproductive Property of Chi-Square)

If V_1^2 and V_2^2 are independent chi-square variables with r_1 and r_2 degrees of freedom respectively, then the distribution of

$$V^2 = V_1^2 + V_2^2$$

is chi-square with $r = r_1 + r_2$ degrees of freedom.

Proof

Since V_1^2 and V_2^2 are independent, the moment-generating function of V^2 is

$$\begin{aligned} M(V^2)(t) &= M(V^2)(t) \quad M(V^2)(t) \\ &= (1-2t)^{-r_1/2}(1-2t)^{-r_2/2} \\ &= (1-2t)^{-(r_1+r_2)/2} \end{aligned}$$

which is the moment-generating function of a chi-square variable with $(r_1 + r_2)$ degrees of freedom.

As a consequence of this theorem, it follows immediately that if $\{x_i\}$ is a random sample of size n from a distribution which is $N(\mu, \sigma^2)$

that

$$\sum_{i=1}^{n}\left[\frac{x_i-\mu}{\sigma}\right]^2$$

is distributed as V^2 with n degrees of freedom.

The following theorem, known as Cochran's Theorem, is stated without proof.

Theorem III-4

Let $x_1, x_2 \ldots x_n$ be independent standard normal variates. If $\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{k} q_i$, where q_i is a quadratic form of rank n_i, then the necessary and sufficient condition that $q_1, q_2 \ldots q_k$ are independently distributed as V^2 variables with respective degrees of freedom $n_1, n_2 \ldots n_k$ is that $\sum_{i=1}^{k} n_i = n$.

The independence of $\bar{x}$ and s^2 can now be established by applying the foregoing theorems to the expression for the sample variance

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

Upon rearranging, this expression becomes

$$\begin{aligned}(n-1)\ s^2 &= \sum_{i=1}^{n} (x_i-\bar{x})^2 \\ &= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2\end{aligned}$$

or $\sum_{i=1}^{n} x_i^2 = (n-1)s^2 - n\bar{x}^2$

If x' is the vector $(x_1, x_2 \ldots x_n)'$, then this last equation may be written in terms of quadratic forms as

$$\begin{aligned}x'\ I(n)x &= x'Ax - x'Bx \\ &= q_1' - q_2\end{aligned}$$

where

$$A = \begin{bmatrix} 1-1/n & -1/n & -1/n & \ldots & -1/n \\ -1/n & 1-1/n & -1/n & \ldots & -1/n \\ -1/n & -1/n & 1-1/n & \ldots & -1/n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -1/n & -1/n & -1/n & \ldots & 1-1/n \end{bmatrix}$$

$$B = \begin{bmatrix} 1/n & 1/n & 1/n & \ldots & 1/n \\ 1/n & 1/n & 1/n & \ldots & 1/n \\ 1/n & 1/n & 1/n & \ldots & 1/n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1/n & 1/n & 1/n & \ldots & 1/n \end{bmatrix}$$

and I(n) is the identity matrix of rank n.

Now, the determinant of A is

$$|A| = \begin{vmatrix} 1-1/n & -1/n & -1/n & \ldots & -1/n \\ -1/n & 1-1/n & -1/n & \ldots & -1/n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -1/n & -1/n & -1/n & \ldots & 1-1/n \end{vmatrix}$$

and by subtracting the 1st row from all other rows,

$$|A| = \begin{vmatrix} 1-1/n & -1/n & -1/n & \ldots & -1/n \\ -1 & 1 & 0 & \ldots & 0 \\ -1 & 0 & 1 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & 0 & 0 & \ldots & 1 \end{vmatrix}$$

and, by adding all columns to the 1st column

$$|A| = \begin{vmatrix} 0 & -1/n & -1/n & \cdots & -1/n \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

Clearly, the rank of A is (n-1), the order of the largest order non-vanishing sub-determinant in A. Hence q_1 is a quadratic form of rank (n-1).

Similarly, the determinant of B is

$$|B| = \begin{vmatrix} 1/n & 1/n & 1/n & \cdots & 1/n \\ 1/n & 1/n & 1/n & \cdots & 1/n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1/n & 1/n & 1/n & \cdots & 1/n \end{vmatrix}$$

$$= (1/n)^n \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & \cdots & 1 \end{vmatrix}$$

$$= (1/n)^n \; |1|$$

and the rank of quadratic form of q_2 is 1.

Now,

$$\text{rank } [q_1] + \text{rank } [q_2] = (n-1) + 1$$
$$= \text{rank } [I(n)]$$

which, by Cochran's Theorem, is the sufficient condition that q_1 and q_2 be independently distributed as V^2 variables with (n-1) and 1 degrees of freedom respectively. Hence $\bar{x}$ and s^2 are independent.

<u>The ($\bar{x}$, s^2) Frequency Surface</u>

The joint distribution of $\bar{x}$, s^2 is given by

$$f(\bar{x}, s^2)\, d\bar{x}\, d(s^2) = f_1(\bar{x})\, d\bar{x}\, f_2(s^2)\, d(s^2)$$

due to the independence of $\bar{x}$ and s^2.

It has been shown that

$$\frac{n(\bar{x} - \mu)^2}{\sigma^2} = \left[\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right]^2$$

is V^2 with 1 degree of freedom, and it is known that the square root of a V^2 variable with 1 degree of freedom is N(0,1). Thus, the probability density function of

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

is

$$\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\left[\frac{(\bar{x} - \mu)}{\sigma/\sqrt{n}}\right]^2} d\left[\frac{(\bar{x}-\mu)}{\sigma/\sqrt{n}}\right], \quad -\infty < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \infty$$

Hence the density function of $\bar{x}$ is

$$f_1(\bar{x})\, d\bar{x} = \sqrt{\frac{n}{2\pi}}\, e^{-\frac{1}{2}\left[\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right]^2} d\bar{x}, \quad -\infty < \bar{x} < \infty$$

Similarly, it has been shown that

$$(n-1)\, s^2 / \sigma^2$$

is V^2 with (n-1) degrees of freedom. Hence, its probability density function is given by

$$\frac{e^{-\frac{(n-1)s^2}{2\sigma^2}}}{\Gamma(\frac{n-1}{2})\ 2^{\frac{n-1}{2}}} \left[\frac{(n-1)s^2}{\sigma^2}\right]^{\frac{n-3}{2}} d\left[\frac{(n-1)s^2}{\sigma^2}\right],\ 0 < \frac{(n-1)s^2}{\sigma} < \infty$$

and

$$f_2(s^2)\ d(s^2) = \frac{(n-1)e^{-\frac{(n-1)}{\sigma^2}\frac{s^2}{2}}}{\Gamma(\frac{n-1}{2})\ 2^{\frac{n-1}{2}}\ \sigma^2} \left[\frac{(n-1)s^2}{\sigma^2}\right]^{\frac{n-3}{2}} d(s^2),\ 0<s^2<\infty$$

The joint density function of $\bar{x}$ and s^2, then, is given by

$$f_1(\bar{x})d\bar{x}f_2(s^2)\ d(s^2) = D\left[\frac{s^2}{\sigma^2}\right]^{\frac{n-3}{2}} e^{-[(n-1)\ s^2+n(\bar{x}-\mu)^2]/2\sigma^2} d\bar{x}d(s^2)$$

$$\text{where } D = \sqrt{\frac{n}{2\pi}}\left[\frac{n-1}{2}\right]^{\frac{n-1}{2}} \frac{1}{\sigma^3\Gamma(\frac{n-1}{2})}$$

<u>The t-distribution</u>

Let the variable t be defined

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

and consider the transformation

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}\ ,\quad s^2 = s^2$$

The Jacobian of this transformation is

$$J = \begin{vmatrix} \partial\bar{x}/\partial t & \partial s^2/\partial t \\ \partial\bar{x}/\partial s^2 & \partial s^2/\partial s^2 \end{vmatrix} = \begin{vmatrix} s/\sqrt{n} & 0 \\ t/2\sqrt{ns^2} & 1 \end{vmatrix} = s/\sqrt{n}$$

Hence $f(t,s^2)\ dt\ d(s^2) =$

$$\frac{\left(\frac{n-1}{2}\right)^{\frac{n-1}{2}} (s^2/\sigma^2)^{\frac{n-2}{2}} e^{-\frac{(n-1)s^2}{2\sigma^2}\left[1 + t^2/(n-1)\right]} \, dt \, d(s^2)}{\sqrt{2\pi}\,\sigma^2 \; \Gamma\left(\frac{n-1}{2}\right)}$$

$$-\infty < t < \infty$$

$$0 < s^2 < \infty$$

Integration of this function with respect to s^2 yields

$$f(t, s^2)\, dt\, d(s^2) = f(t)\, dt$$

$$f(t)\, dt = \frac{\Gamma\left(\frac{n}{2}\right)\left[1 + t^2/(n-1)\right]^{-\frac{n}{2}}}{\sqrt{\pi(n-1)}\;\Gamma\left(\frac{n-1}{2}\right)}\, dt \qquad -\infty < t < \infty$$

or $$f(t)\, dt = \frac{\left[1 + t^2/(n-1)\right]^{-\frac{n}{2}}\, dt}{\sqrt{(n-1)}\;\beta\left(1/2, \frac{n-1}{2}\right)} \qquad -\infty < t < \infty$$

where the function $\beta\left(1/2, \frac{n-1}{2}\right) = \dfrac{\Gamma(1/2)\,\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$

and $\Gamma(1/2) = \sqrt{\pi}$

The function $f(t)\, dt$ is the probability density function of "Student's" t-distribution with (n-1) degrees of freedom.

The use of the term "degrees of freedom" in this context requires some explanation. It has been shown that

$$\sum_i \left[\frac{x_i - \mu}{\sigma}\right]^2$$

is distributed as V^2 with (n) degrees of freedom, where the degrees of freedom have been interpreted as the number of independent squared standard normal variates. If, then, there were k linear restrictions among n variates, the sum of squares of these variates would be distributed with (n-k) degrees of freedom, each restriction reducing the "dimension" of the variation by unity.

Since the probability density function of t is dependent on the sample size, n, some specification of the sample size is necessary to completely determine t. The quantity (n-1) appears in the density function of t as a consequence of the degrees of freedom associated with s^2, which appears in the denominator of t. The degrees of freedom of s^2 are, in turn, a consequence of the sample size, where the "dimension" of the variation in s^2 is of size n subject to one linear restriction (that $\sum_i (x_i - \bar{x}) = 0$). Thus, by extension, the term "degrees of freedom" is associated with the t-distribution.

Geometrical Derivation of the $(\bar{x}, s^2)$ Frequency Surface

An alternative method of obtaining the joint distribution of $\bar{x}$ and s^2 is the following geometrical proof due to Fisher[3].

Let $x_1, x_2, \ldots x_n$ be n values of a random sample from a population that is $N(\mu, \sigma^2)$. The joint distribution of the sample is

$$\left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2} dx_1\, dx_2 \ldots dx_n$$

$$= \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n e^{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]} dx_1\, dx_2 \ldots dx_n$$

since it has been shown that

$$\sum(x_i-\mu)^2 = (n-1)s^2 + n(\bar{x}-\mu)^2$$

Let $(x_1, x_2 \ldots x_n)$ be co-ordinates of the sample point, P, in n-dimensional Euclidean hyper-space. In figure 1, let OA be the unit vector in the n-dimensional space with direction cosines proportional to 1, 1, ...1, and let PC be orthogonal to OA at C. If the co-ordinates of the point C are (a, a,...a), then the direction cosines of PC are proportional to x_1-a, $x_2-a, \ldots x_n-a$, and those of the unit vector are proportional to a, a,...a. By construction PC is orthogonal to OA, hence

$$a(x_1-a) + a(x_2-a) + \ldots a(x_n-a) = 0$$

Since $a \neq 0$, then $\sum(x_i-a) = 0$

and this gives $a = \sum \frac{x_i}{n} = \bar{x}$

so that the co-ordinates of C are $(\bar{x}, \bar{x}, \ldots \bar{x})$.

Then $(OC)^2 = \bar{x}^2 + \bar{x}^2 + \ldots \bar{x}^2 = n\bar{x}^2$ and $(PC)^2 = (x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \ldots + (x_n-\bar{x})^2 = \sum(x_i-\bar{x})^2 = (n-1)s^2$; from which $OC = \sqrt{n}\,\bar{x}$ and $PC = \sqrt{n-1}\,s$. The lines OC and PC are orthogonal, thus $\bar{x}$ and s are clearly independent.

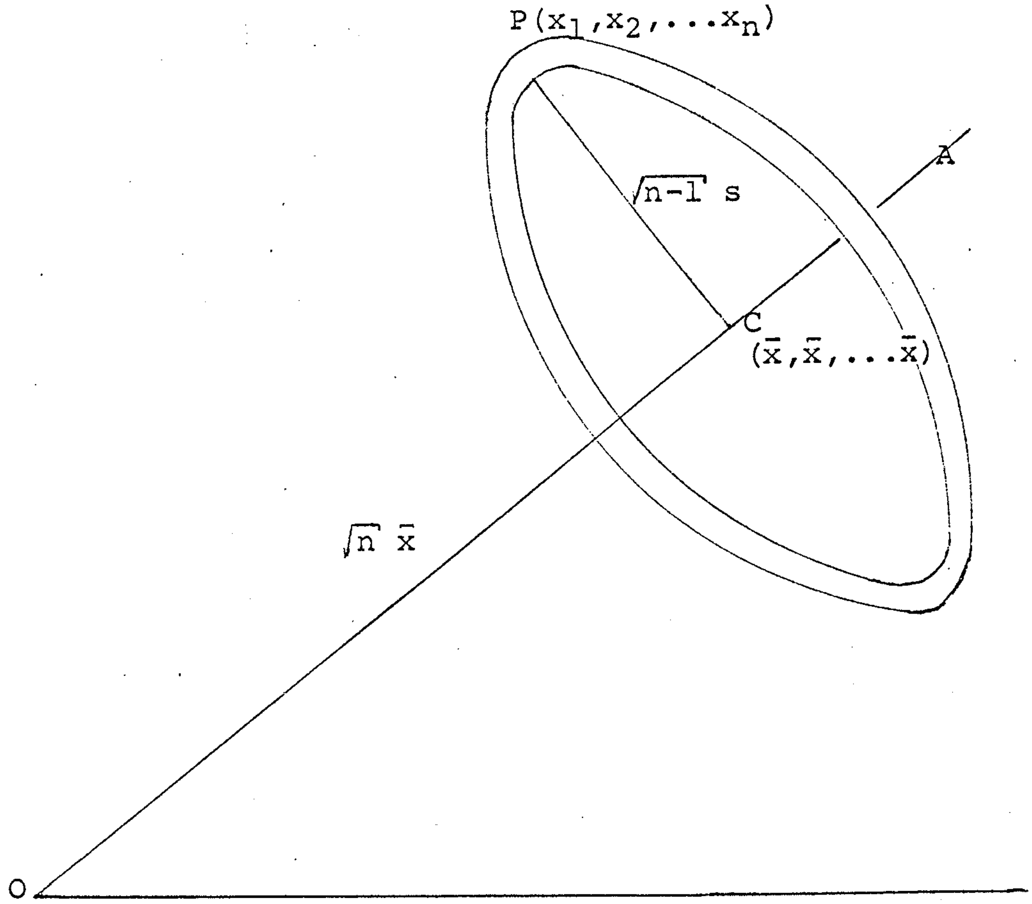

figure 1

To obtain the joint density function of $\bar{x}$ and s, it is necessary only to transform the volume element

$$dV = dx_1\ dx_2 \ldots dx_n$$

as the density factor is already expressed in terms of $\bar{x}$ and s from previous considerations.

For a given $\bar{x}$ and s, OC and PC are constant, and the sample point P moves on the surface of an (n-1)-dimensional hypersphere of radius PC = $\sqrt{n-1}\ s$, centred at C. The element of the "spherical" shell in which P moves has "dimensions" d(PC) = $\sqrt{n-1}\ ds$ and d(OC) $= \sqrt{n}\ d\bar{x}$ and therefore

$$dV = k_1\ (s\ \sqrt{n-1})^{n-2}\ \sqrt{n-1}\ ds\ \sqrt{n}\ d\bar{x}$$

where k_1 is a constant.

The joint distribution of $\bar{x}$ and s will be of the form

$$\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\frac{-1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]} k_1 (s\ \sqrt{n-1})^{n-2}\ \sqrt{n-1}\ ds \sqrt{n}\ d\bar{x}$$

$$= D\ e^{-\frac{1}{2\sigma^2}\left[(n-1)s^2 - n(\bar{x}-\mu)^2\right]} s^{n-2}\ ds d\bar{x} \qquad \begin{array}{l} 0 < s < \infty \\ -\infty < \bar{x} < \infty \end{array}$$

which is expressible in the form

$$f_1(\bar{x})\ d\bar{x}\ f_2(s) ds$$

due to the independence of $\bar{x}$ and s. Completing each factor with the necessary constants yields the joint density function of $\bar{x}$ and s^2, specifically,

$$D_1 \int_{-\infty}^{\infty} e^{\frac{-n}{2\sigma^2}(\bar{x}-\mu)^2} \, d\bar{x} = 1$$

and therefore $D_1 = \dfrac{\sqrt{n/2\pi}}{\sigma}$,

$$\text{hence } f_1(\bar{x})\, d\bar{x} = \frac{\sqrt{n/2\pi}}{\sigma} \, e^{-\frac{1}{2}\left[\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right]^2} d\bar{x} \qquad -\infty < x < \infty \quad ;$$

$$\text{and} \quad D_2 \int_0^{\infty} s^{n-2} \, e^{-\frac{1}{2}(n-1)s^2} \, ds = 1$$

$$\text{or} \quad D_2 \int_0^{\infty} \frac{s}{2}^{\,n-3} \, e^{-\frac{1}{2}(n-1)s^2} \, d(s^2) = 1.$$

$$\text{Therefore} \qquad D_2 = \frac{\left[\frac{n-1}{\sigma^2}\right]^{\frac{n-3}{2}}}{2^{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)} \quad ,$$

and hence $f_2(s^2)d(s^2)$

$$= \left(\frac{n-1}{2}\right)^{\frac{n-1}{2}} \frac{\frac{(s^2/\sigma^2)^{\frac{n-3}{2}}}{\sigma^2} \, e^{\frac{-1}{2\sigma^2}(n-1)s^2}}{\Gamma\left(\frac{n-1}{2}\right)} \, d(s)^2$$

$$0 < s^2 < \infty$$

The joint distribution is, therefore

$f_1(\bar{x})d\bar{x} \; f_2(s^2)d(s)^2$

$$= \frac{\sqrt{n/2\pi}}{\sigma^3} \left(\frac{n-1}{2}\right)^{\frac{n-1}{2}} \frac{(s^2/\sigma^2)^{\frac{n-3}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \, e^{\frac{-1}{2\sigma^2}\left[(n-1)s^2+n(\bar{x}-\mu)^2\right]} \, d\bar{x}d(s^2)$$

$$-\infty < x < \infty$$

$$0 < s^2 < \infty$$

from which the distribution of t can be obtained as before.

CHAPTER IV

PROPERTIES OF STUDENT'S t-DISTRIBUTION

The probability density function of "Student's" t, as shown in the preceding chapter is given by

$$f(t)dt = \frac{dt}{\sqrt{n}\;\beta(1/2,\, n/2)(1 + t^2/n)^{\frac{n+1}{2}}} \qquad -\infty < t < \infty$$

where n is the degrees of freedom. Since the distribution is the reciprocal of an even function, it is apparent that the distribution is symmetrical with median and mode at $t = 0$.

Following Craig's[1] notation, the distribution is of the form

$$y = C(U^2 + V^2)^{-m}$$

where

$$C = 1/\sqrt{n}\;\beta(1/2,\, n/2)$$

$$U = t/\sqrt{n}$$

$$V = 1$$

$$m = (n+1)/2$$

"Student's" t can be classified as a Transitional Type VII in the Pearson system, and is related to the Main Type IV and Transitional Type II, in that a Type VII curve may be derived from either of these distributions. The Transitional Type II can be shown to be a special case of Main Type I, of which class the β_1 variable is a member.

<u>Moments of the t-distribution</u>

The first non-central moment, the mean, is given by

$$\mu_1' = \int_{-\infty}^{\infty} \frac{t\,dt}{\sqrt{n}\ \beta(1/2,\ n/2)(1 + t^2/n)^{\frac{n+1}{2}}}$$

$$= K \int_{-\infty}^{\infty} t\ (\ 1 + t^2/n)^{-\frac{n+1}{2}}\ dt$$

where $K = \dfrac{1}{\sqrt{n}\ \beta(1/2,n/2)}$

thus $\mu_1' = \frac{n}{2}\ K \int \dfrac{\frac{2t}{n}}{(1 + t^2/n)^{\frac{n+1}{2}}}\ dt$

which is of the form

$$\int_{-\infty}^{\infty} \frac{dx}{x^m} = \frac{x^{-m+1}}{-m + 1}$$

Hence

$$\mu_1' = \frac{n}{n-1}\ K\ \left[1 + t^2/n\right]^{-\frac{(n-1)}{2}} \Bigg|_{-\infty}^{\infty}$$

$$= \frac{n}{n-1}\ K\ [0] = 0,$$

which result would seem obvious as it has already been noted that f(t) is symmetrical about t = 0.

From the property of symmetry, it follows that all odd order moments about the mean will vanish.

The even order moments about the mean are given by

$$E(t-\mu_1')^{2r} = E(t)^{2r} = \mu_{2r}'$$

and $\mu'_{2r} = \int_{-\infty}^{\infty} \frac{t^{2r} \quad (1 + t^2/n)^{-\frac{n+1}{2}}}{\sqrt{n}\, \beta(1/2,\ n/2)} \, dt$

$$= \frac{2n^r}{\sqrt{n}\, \beta(1/2,\ n/2)} \int_0^{\infty} (t^2/n)^r \quad (1+t^2/n)^{-\frac{n+1}{2}} \, dt.$$

Letting $t^2/n = y$ yields

$$\mu'_{2r} = \frac{n^r}{\beta(1/2,\ n/2)} \int_0^{\infty} \frac{y^{r-\frac{1}{2}}}{(1+y)^{\frac{n+1}{2}}}$$

$$= \frac{n^r}{\beta(1/2,\ n/2)} \quad \beta\,(n/2 - r,\ r + 1/2).$$

However, the integral

$$\beta(\ell, m) = \int_0^{\infty} \frac{y^{m-1}}{(1-y)^{\ell+m}} \, dy$$

converges if and only if ℓ, $m > 0$. Thus, since $m = r + 1/2$ where $r > 0$,

$$\ell = n/2 - r > 0 \quad \text{or } 2r < n$$

and the distribution of t possesses even order moments only up to a number less than the number of degrees of freedom.

The first moments, then are

$$\mu'_1 = \mu_1 = 0 = \text{mean}$$

$$\mu'_2 = \mu_2 = \frac{n}{n-2} = \text{variance}$$

$$\mu'_3 = \mu_3 = 0$$

$$\mu'_4 = \mu_4 = \frac{3n^2}{n^2 - 6n + 8}$$

and skewness = $\dfrac{\mu_3}{(\mu_2)^{3/2}} = 0$

$$\text{kurtosis} = \frac{\mu_4}{(\mu_2)^2} = \frac{3(n^2 - 2n + 4)}{n^2 - 6n + 8}$$

The kurtosis of the t-distribution is dependent on sample size, and since n is positive,

$$\frac{n^2 - 2n + 4}{n^2 - 6n + 8} = 1 + \frac{4n - 8}{n^2 - 6n + 8} > 1$$

and the curve has kurtosis > 3 and hence is leptokurtic.

It will be noted that

$$\lim_{n \to \infty} \left[\mu_2\right] = \lim_{n \to \infty} \left[\frac{n}{n-2}\right] = 1$$

and $\lim\limits_{n \to \infty} \left[\text{kurtosis}\right] = \lim\limits_{n \to \infty} 3 \left[\dfrac{n^2 - 2n + 4}{n^2 - 6n + 8}\right] = 3$

In the limit, as n increases, then, the value of these coefficients approaches that of the standard normal distribution, which suggests that the standard normal is the limiting form of the t distribution. This is, in fact, the case, as will be shown in the next section.

Limiting Form of the t-distribution

The limiting form of the t-distribution as $n \to \infty$ is the standard unit normal. To show this, consider

$$f(t) = \frac{1}{\sqrt{n}\ \beta(1/2,\ n/2)\ (1+t^2/n)^{\frac{n+1}{2}}}$$

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{2}{n}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} (1 + t^2/n)^{\frac{-n+1}{2}}$$

so that $\ln f(t) = -\frac{1}{2}\ln (2\pi) + \frac{1}{2}\ln 2 - \frac{1}{2}\ln(n) + \ln\Gamma(\frac{n+1}{2})$

$$- \ln \Gamma(\frac{n}{2}) - \frac{n+1}{2} \ln (1+t^2/n)$$

Applying Stirling's approximation

$$\ln \Gamma(1+x) = \ln(x!) = \frac{1}{2}\ln(2\pi) + (x+\tfrac{1}{2})\ln x - x + w_n$$

where $0 < w_n < 1/12x$, yields

$$\ln \Gamma(\frac{n+1}{2}) - \ln \Gamma(\frac{n}{2}) = -\tfrac{1}{2}\ln 2 + \tfrac{1}{2}\ln(n) - \tfrac{1}{2} + \tfrac{1}{2}\left[n \ln(1-1/n) - (n-1) \ln(1-2/n)\right]$$

Recalling that the Taylor expansion of $\ln(1+x)$ is

$$\ln (1+x) = x - \frac{x^2}{2!} + \frac{x^3}{3!} \; \ldots\ldots$$

then $\ln \Gamma(\frac{n+1}{2}) - \ln\Gamma(\frac{n}{2}) = -\frac{1}{2}\ln 2 + \frac{1}{2}\ln n - \frac{1}{4n} + \frac{1}{6n^2} + \frac{13}{24n^3} \; \ldots$

where succeeding terms are of order $1/n^4$.

Now $-\frac{n+1}{2} \ln (1+t^2/n) = -\frac{1}{2}(n-1) \ln(1+t^2/n)$, which, upon substitution of the expansion of $\ln(1+x)$, becomes

$$-\frac{t^2}{2}\left[1-1/n - \frac{t^2}{2}(\frac{1}{n}-\frac{1}{n^2}) + \frac{t^4}{3}(\frac{1}{n^2}-\frac{1}{n^3}) - \frac{t^6}{4}(\frac{1}{n^3}-\frac{1}{n^4}) - \ldots\right]$$

Hence $\lim\limits_{n\to\infty}\left[\ln f(t)\right] = -\frac{1}{2}\ln(2\pi) + \lim\limits_{n\to\infty}\left[\frac{-1}{4n} + \frac{1}{6n^2} + \frac{13}{24n^3} - \ldots\right]$

$$+ \lim_{n\to\infty}\left[-\frac{t^2}{2} (1-\frac{1}{n}) - \frac{t^2}{2} (\frac{1}{n}-\frac{1}{n^2}) \; \ldots\right]$$

where $\lim\limits_{n\to\infty} \left[- \frac{1}{4n} + \frac{1}{6n^2} + \frac{13}{24n^3} + \ldots \right] = 0$

and $\lim\limits_{n\to\infty} \left[-\frac{t^2}{2} (1-1/n) - \frac{t^2}{2} (\frac{1}{n}-\frac{1}{n^2}) - \ldots\right] = \frac{-t^2}{2} [1]$

Consequently,

$$\lim_{n\to\infty}\left[\ln f(t)\right] = -\tfrac{1}{2}\ln(2\pi) \quad -\frac{t^2}{2}$$

and therefore $\lim_{n\to\infty} f(t) = \frac{1}{\sqrt{2\pi}} \; e^{-t^2/2}$

which is the probability density function of the standard normal deviate.

Special case of the distribution for n = 1

It has been noted that the integral

$$\beta(\ell,m) = \frac{y^{m-1}}{(1+y)^{\ell+m}} \quad dy$$

converges if and only if ℓ, $m > 0$. Thus, the t-distribution is defined if and only if $m = n > 0$, where n is the degrees of freedom.

Consider the case where $n = 1$. The probability density function of t, is given by

$$f(t)\ dt = \frac{dt}{\sqrt{n}\ \beta(1/2,n/2)\,(1+t^2/n)^{\frac{n+1}{2}}}$$

reduces to

$$f(t)dt = \frac{dt}{\pi(1 + t^2)} \qquad -\infty < t < \infty$$

which is the probability density function of the Cauchy distribution. This distribution has no practical application, but, because of its unusual properties, was of great value historically in defining the necessary and sufficient conditions that a given frequency function be a probability density function.

The distribution is symmetrical about $t = 0$ but does not possess any moments. For example,

$$\text{mean} = \int_{-\infty}^{\infty} \frac{t}{\pi(1+t^2)} \, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} 2t(1+t^2)^{-1} \, dt$$

$$= \frac{1}{2\pi} \Big[\log |1+t^2| \Big] \Big|_{-\infty}^{\infty} \quad \text{does not exist}$$

although

$$\int_{-a}^{a} \frac{t \, dt}{\pi(1+t^2)}$$

does exist and is equal to zero in the limit as $a \to \infty$.

Regarding the higher order moments (odd and even) the integral

$$\mu_r' = \int_{-\infty}^{\infty} \frac{t^r}{\pi(1+t^2)} \, dt, \quad r = 1, 2, \ldots$$

does not converge for any integral value of r.

The characteristic function, however, does exist, and is equal to

$$\varphi(w) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itw}}{1+t^2} \, dt$$

$$= e^{-|w|}$$

<u>Generalized Definition of t</u>

Let the variable t be re-defined to be

$$t = \sqrt{\frac{V^2/1}{V_1^2/n_1}} = \frac{w}{\sqrt{V_1^2/n_1}}$$

where V^2 and V_1^2 are distributed as chi-square distributions with 1 and n_1 degrees of freedom, w is distributed as a standard unit normal variable, and numerator and denominator are independent.

This definition, proposed by Fisher in 1925, is much more general than that which has been considered to this point, and is to be preferred because of its generality, particularly in view of the applications of the t-distribution to be discussed in the next chapter. The definition

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

is of interest historically because it is the definition from which the work considered so far developed. Since the approach of this paper is essentially a historical one, the historical definition has been utilized thus far.

The historical definition can easily be shown to be a special case of the general definition. If

$$t = \sqrt{\frac{V^2/1}{V_1^2/n}}$$

then $$\frac{V^2/1}{V_1^2/n} = \frac{n\frac{(\bar{x}-\mu)^2}{\sigma^2} \Big/ 1}{\frac{(n-1)s^2}{\sigma^2} \Big/ (n-1)}$$

since the numerator and denominator of this expression have been shown to be distributed as V^2 variables with 1 and (n-1) degrees of freedom respectively if x is distributed normally.

$$\frac{\dfrac{n(\bar{x}-\mu)^2}{\sigma^2}}{\dfrac{(n-1)s^2}{(n-1)\sigma^2}} = \frac{n(\bar{x}-\mu)}{s^2} = t^2$$

After taking the square root of both sides

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

where t is distributed as "Student's" t with (n-1) degrees of freedom

<u>Inter-relationship of t, V^2 and F distributions</u>

Let V_1^2 and V_2^2 be independent chi-square variables with n_1 and n_2 degrees of freedom respectively. The joint distribution of V_1^2 and V_2^2 is

$$f(V_1^2,V_2^2)d(V_1^2)d(V_2^2) = \frac{e^{-\frac{1}{2}(V_1^2+V_2^2)}(V_1^2)^{\frac{n_1-1}{2}}(V_2^2)^{\frac{n_2-1}{2}}}{\Gamma(\frac{n_1}{2})\,\Gamma(\frac{n_2}{2})\,2^{\frac{n_1+n_2}{2}}}d(V_1^2)d(V_2^2)$$

$$0 < V_1^2 < \infty$$

$$0 < V_2^2 < \infty$$

Let $x = \dfrac{V_1^2}{V_2^2}$ and $y = V_2^2$. The Jacobian of this transformation is

$$J = \begin{vmatrix} \partial V_1^2/\partial x & \partial V_1^2/\partial y \\ \partial V_2^2/\partial x & \partial V_2^2/\partial y \end{vmatrix} = \begin{vmatrix} y & x \\ 0 & 1 \end{vmatrix} = y.$$

$$g(xy)\,dx\,dy = \frac{e^{-\frac{y}{2}(1+x)}\, x^{\frac{n_1}{2}-1}\, y^{\frac{n_1+n_2}{2}-1}}{\Gamma(\frac{n_1}{2})\,\Gamma(\frac{n_2}{2})\; 2^{\frac{n_1+n_2}{2}}}\,dx\,dy$$

$0 < x < \infty$

$0 < y < \infty$

If $w = \frac{y}{2}(1+x)$ then $dw = \frac{dy}{2}(1+x)$

and $$g(xw)\,dx\,dw = \frac{e^{-w}\, x^{\frac{n_1}{2}-1}\, w^{\frac{n_1+n_2}{2}-1}}{\Gamma(\frac{n_1}{2})\,\Gamma(\frac{n_2}{2})\,(1+x)^{\frac{n_1+n_2}{2}}}\,dx\,dw$$

Integration with respect to w yields

$$g(x)\,dx = \frac{x^{\frac{n_1}{2}-1}\,dx}{\beta(\frac{n_1}{2}, \frac{n_2}{2})\,(1+x)^{\frac{n_1+n_2}{2}}} \qquad 0<x<\infty$$

which is the probability density function of a β_2 variate. Thus the ratio of 2 independent V^2 variables with n_1 and n_2 degrees of freedom, is distributed as a β_2 variable with parameters $n_1/2$ and $n_2/2$, that is

$$\frac{x^{\frac{n_1}{2}-1}\,dx}{\beta(\frac{n_1}{2},\frac{n_2}{2})\,(1+x)^{\frac{n_1+n_2}{2}}} \text{ is } \beta_2\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$$

Now consider t with r degrees of freedom

$$t = \sqrt{\frac{V^2/1}{V_1^2/r}}$$

where V_1^2 has r degrees of freedom. Thus

$$t^2 = \beta_2\left(\frac{1}{2}, \frac{r}{2}\right)$$

since t^2 is the ratio of 2 independent chi-square variables divided by their corresponding degrees of freedom.

The F distribution is defined to be

$$F = \frac{s_1^2}{s_2^2}$$

where $s_1^2 = \sum \frac{(x_{1i}-\bar{x}_1)^2}{n_1-1}$

$$s_2^2 = \sum \frac{(x_{2i}-\bar{x}_2)^2}{n_2-1}$$

and these statistics are calculated from 2 independent random samples of size n_1 and n_2 from populations which are normally distributed.

Then

$$\frac{(n_1-1)}{(n_2-1)} F = \frac{\frac{(n_1-1)s_1^2}{\sigma^2}}{\frac{(n_2-1)\ s_1^2}{\sigma^2}}$$

$$= \frac{V_1^2}{V_2^2}$$

where V_1^2 and V_2^2 have (n_1-1) and (n_2-1) degrees of freedom respectively. Hence $\frac{n_1-1}{n_2-1}$ F is a $\beta_2\left(\frac{n_1-1}{2}, \frac{n_2-1}{2}\right)$

variate. If $(n_1-1) = 1$, then $\frac{(n_1-1)}{n_2-1} F = \frac{F}{n_2-1}$

$$= \beta_2(1/2, \frac{n_2-1}{2})$$

which is t^2 with (n_2-1) degrees of freedom. It is apparent, then, that t with (n_2-1) degrees of freedom is a special case of the F distribution with 1 and (n_2-1) degrees of freedom, specifically t^2 with n degrees of freedom $= F_{1,n}$ and that values of t^2 for n degrees of freedom at the α level of significance can be obtained from the values of F with 1 and n degrees of freedom, and level of significance α.

CHAPTER V

APPLICATIONS OF "STUDENT'S" t-DISTRIBUTION

The area of most frequent application of "Student's" t-distribution is that of testing statistical hypotheses. As has been pointed out in Chapter II, the extension of exact testing procedures to small sample work has been of particular importance to the development of statistical theory, and has facilitated the application of statistical techniques to the various disciplines of experimental science.

TESTS OF HYPOTHESES BASED ON THE t-DISTRIBUTION

1) Testing an assumed population mean

A common experimental situation is the one in which an experimenter wishes to test an assumed value, μ_o of the mean of a normal population.

Let $x_1, x_2, \ldots x_n$ be a random sample of size n from a normal population with unknown mean μ and variance σ^2. If

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

where
$$\bar{x} = \sum x_i/n$$

and
$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

then t is distributed as "Student's" t with n-1 degrees of freedom as has been shown in Chapters III and IV. Substituting for $\bar{x}$ and s, and given

$$H_o: \mu = \mu_o$$

then the statistic t can be evaluated and its significance determined with respect to a chosen level of significance.

As an example, consider the following artificial date--9.73, 5.42, 6.57, 6.01, 4.71, 6.97 5.73, 7.96, 8.05, 4.53, for which $\bar{x} = 6.57$, $s = 1.64$ and $n = 10$.

To test the hypothesis

$$H_o: \mu = 6.50$$
$$H_1: \mu \neq 6.50$$

compute

$$t = \frac{6.57 - 6.50}{1.64/\sqrt{10}} = \frac{0.07}{0.52} = 0.13$$

At the 5% level of significance, the values of t beyond which 5% of the area under the curve lie (2.5% in each tail) are ± 2.26 for 9 degrees of freedom. Since the calculated value of t does not exceed 2.26, the test is judged non-significant , and it may be concluded that the evidence in the sample does not refute the null hypothesis.

2a) <u>Testing the difference between unpaired sample means</u>

The t-distribution may be used to test the hypothesis that the means of two populations differ or the hypothesis that the difference between the population means, $\mu_1 - \mu_2$, is a specified value. If it can be shown that the variances of the samples are the same, then the hypothesis that $\mu_1 - \mu_2 = 0$ is equivalent to testing the hypothesis that the two samples have been drawn from a common normal population, $N(\mu, \sigma^2)$ where $\mu = \mu_1 = \mu_2$

Let $x_{11}, x_{12}, \ldots x_{1n_1}$ and $x_{21}, x_{22}, \ldots x_{2n_2}$ be independent random samples drawn from two populations, P_1 and P_2 which are, respectively, $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$.

Let

$$\bar{x}_1 = \frac{1}{n_1} \sum_i x_{1i} \qquad \bar{x}_2 = \frac{1}{n_2} \sum_i x_{2i}$$

and

$$s_1^2 = \frac{1}{n_1 - 1} \sum_i (x_{1i} - \bar{x}_1)^2 \quad s_2^2 = \frac{1}{n_2 - 1} \sum_i (x_{2i} - \bar{x}_2)^2$$

and consider the statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

Since x_{1i} and x_{2i} are normally distributed variates and are independent, $\bar{x}_1 - \bar{x}_2$ is normally distributed with mean $E\left[\bar{x}_1 - \bar{x}_2\right] = \mu_1 - \mu_2$

and variance $\operatorname{var}(\bar{x}_1 - \bar{x}_2) = \operatorname{var}\left[\sum \frac{x_{1i}}{n_1} - \sum \frac{x_{2i}}{n_2}\right]$

$$= \operatorname{var}\left[\sum \frac{x_{1i}}{n_1}\right] + \operatorname{var}\left[\sum \frac{x_{2i}}{n_2}\right]$$

$$= \frac{n_1\sigma_1^2}{n_1^2} + \frac{n_2\sigma_2^2}{n_2^2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Hence $\bar{x}_1-\bar{x}_2$ is $N(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$ and from Theorem III-1 it follows that

$$\frac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

is distributed as N(0,1).

Multiplying the numerator and denominator of t by $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, then t is given by

$$t = \frac{\dfrac{\bar{x}_1-\bar{x}_2 - (\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}}{\dfrac{\sqrt{s_1^2/n_1 + s_2^2/n_2}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}}$$

Now, the numerator of this latter expression is N(0,1) but the denominator is not distributed as $\sqrt{V^2/r}$ where V^2 has r degrees of freedom.

If, however, it is assumed that $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (i.e. that P_1 and P_2 have a common variance) then both s_1^2 and s_2^2 estimate σ^2, and the best estimate of this common variance will be obtained by pooling the individual sample variances. Let this estimate be

$$s^2 = \frac{\sum(x_{1i}-\bar{x}_1)^2 + \sum(x_{2i}-\bar{x}_2)^2}{n_1 + n_2 - 2}$$

where s^2 has n_1+n_2-2 degrees of freedom because two estimates, $\bar{x}_1$ and $\bar{x}_2$, have been calculated. By definition of s_1^2 and s_2^2,

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

In the expression for t, if s_1^2 and s_2^2 are replaced by s^2 then t becomes

$$t = \frac{\dfrac{\bar{x}_1-\bar{x}_2 - (\mu_1-\mu_2)}{\sqrt{\sigma^2(1/n_1+1/n_2)}}}{\dfrac{\sqrt{s^2\ (1/n_1+1/n_2)}}{\sqrt{\sigma^2\ (1/n_1+1/n_2)}}}$$

Now $s^2/\sigma^2 = \dfrac{\dfrac{(n_1-1)s_1^2}{\sigma^2} + \dfrac{(n_2-1)s_2^2}{\sigma^2}}{n_1 + n_2 - 2}$

But by Cochran's Theorem, it has been shown that $(n_1-1)s_1^2/\sigma^2$ and $(n_2-1)s_2^2/\sigma^2$ will each be distributed as V^2 with (n_1-1) and (n_2-1) degrees of freedom respectively. By the reproductive rule of chi-square, the sum of these variates will also be distributed as V^2 with $(n_1-1) + (n_2-1) = n_1+n_2-2$ degrees of freedom. Hence, the denominator of t is distributed as $\sqrt{V^2/r}$ where V^2 has $r = n_1+n_2-2$ degrees of freedom. Therefore, t will be distributed as Student's t with

n_1+n_2-2 degrees of freedom, and hypotheses about $\mu_1-\mu_2$ may be tested.

As an example, consider the following artificial data

Sample 1	Sample 2
9.73	11.19
5.42	6.56
6.57	9.06
6.01	8.03
4.71	7.10
6.97	8.87
5.73	7.91
7.96	7.46
8.05	8.36
4.53	6.90

where $\bar{x}_1 = 6.57$ $\quad\bar{x}_2 = 8.14$

and $s_1^2 = 2.6923$ $\quad s_2^2 = 1.8192$

Now $$s^2 = \frac{\sum (x_{1i}-\bar{x}_1)^2 + \sum (x_{2i}-\bar{x}_2)^2}{n_1 + n_2 - 2} = \frac{24.2310 + 16.3730}{18}$$

$$= 2.2558$$

Then for H_o: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$

$$t = \frac{6.57 - 8.14}{\sqrt{2.2558\ (1/10+1/10)}} = \frac{-1.57}{0.67} = -2.35$$

The critical value for 18 degrees of freedom at the 95% confidence level is ±2.10. Hence the test is significant and it is concluded that there is evidence to indicate that the means of the populations are different.

2b) <u>Testing the difference between paired sample means</u>

In the test outlined in 2(a), it was assumed that the samples $\{x_{1i}\}$ and $\{x_{2i}\}$ were independent. In some types of work, the two samples to be compared are not independent, but consist of values deliberately paired in order to reduce chance variation which might arise due to extraneous sources. Such an experiment might consist, for example, of performing a second set of trials on the same set of experimental units. This would ensure that differences between the trials were due to differences between the treatments applied to each set of units, and not influenced by variation caused by differences among the members of the sample if a new set of experimental units had been introduced for the second set of trials. Another situation where this technique is useful is the one in which sufficient homogeneous experimental material cannot be obtained, but homogeneous pairs can be formed.

Let $x_{11}, x_{12}, \ldots x_{1n}$ and $x_{21}, x_{22}, \ldots x_{2n}$ be two samples from two populations which are distributed as $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ where the pairs x_{1i} and x_{2i} have a correlation ρ. Define

$$x_i = x_{1i} - x_{2i}.$$

Now x_i is a linear combination of two normally distributed variables, and is itself, therefore, normally distributed with mean

$$E\left[x_i\right] = E\left[x_{1i}-x_{2i}\right] = \mu_1 - \mu_2 = \mu \text{ (say)}$$

and variance

$$E\left[x_{1i}-x_{2i}\right]\left[x_{1j}-x_{2j}\right] = \sigma_1^2 + \sigma_2^2 - 2\rho\,\sigma_1\,\sigma_2$$
$$= \sigma^2$$

Hence, hypotheses about μ may be tested by applying the test given in 1 as the variable x_i is $N(\mu, \sigma^2)$.

This procedure enables tests and estimates to be made which are more precise than those obtainable from 2(a). For

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\,\sigma_2$$

and if $\rho = 0$, the samples are independent and this test reduces to that of 2(a). If, however, the pairs are positively correlated, i.e. $\rho > 0$

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\,\sigma_2 < \sigma_1^2 + \sigma_2^2$$

and the variance from test 2(b) is less than that of 2(a), yielding a test which is more precise. It is easy to see that as ρ becomes closer to +1, the precision of the test increases until the maximum is reached for $\rho = +1$.

On the other hand, if the pairs are negatively correlated, then this test loses precision. A more precise test would be obtained by taking two independent random samples as in Test 2(a).

The following example based on the data given in example 2(a) will indicate this gain in precision. Considering Sample 1 and Sample 2 as paired variates, the mean of the differences between pairs will be found to be $\bar{x} = -1.57$ as would be expected in view of example 2(a). However, the estimated variance of a single difference is $s = 0.9902$ whereas the variance of the difference $x_{1i}-x_{2i}$ in example 2(a) is estimated to be 2.2558. This reduction is due to the fact that the values in samples 1 and 2 have a high positive correlation, having been drawn from a correlated bivariate parent population.

For the hypothesis $H_0: \mu = 0$

$H_1: \mu \neq 0$

$$t = \frac{-1.57}{\sqrt{\frac{0.9902}{10}}} = \frac{-1.57}{0.32} = -4.09$$

The critical value for 9 degrees of freedom at the 5% level of significance is ±2.26, and the test is significant.

3a) <u>Testing a correlation coefficient</u>

Let (x_1,y_1), $(x_2,y_2)\ldots$ (x_n,y_n) be an independent random sample of size n from a correlated bivariate population which is $N(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$.

The correlation coefficient, ρ, is estimated by

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2}}$$

The general expression for the sampling distribution of r can be shown to be

$$f(r)dr = \frac{(1-\rho)^{\frac{n-1}{2}}}{\pi} \frac{(1-r^2)^{\frac{n-4}{2}}}{\Gamma(n-2)} \frac{d^{n-2}}{d(r\rho)^{n-2}} \left[\frac{\cos^{-1}(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right] dr, \quad -1 \leq r \leq 1.$$

However, for the special case $\rho = 0$, the above distribution reduces to the simple form

$$f(r)dr = \frac{(1-r^2)^{\frac{n-4}{2}}\ dr}{\beta(1/2,(n-2)/2)}, \qquad -1 \leq r \leq 1$$

and a test of significance for r based on the t distribution can be obtained.

Let

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Solving for r gives the result

$$r = \left(\frac{t^2}{n-2}\right)^{\frac{1}{2}} \left[1 + \frac{t^2}{n-2}\right]^{-\frac{1}{2}}$$

Using this result to change the variable in the expression for the sampling distribution of r when $\rho = 0$, then

$$\frac{dr}{dt} = (n-2)^{-1/2} \left[1+t^2/(n-2)\right]^{-3/2}$$

and upon substitution f(r) dr becomes

$$g(t)dt = \frac{1}{\beta(1/2,\frac{n-2}{2})} \left[1-\frac{\frac{t^2}{n-2}}{1+\frac{t^2}{n-2}}\right]^{\frac{n-4}{2}} \qquad \frac{dt}{\sqrt{n-2}\ (1+\frac{t^2}{n-2})^{3/2}}$$

$$= \frac{1\qquad dt}{\beta(1/2,\frac{n-2}{2})\sqrt{n-2}\left(1+\frac{t^2}{n-2}\right)^{\frac{n-1}{2}}}$$

which is the probability density function of Student's t with n-2 degrees of freedom.

The hypothesis $H_o: \rho = \rho_o$ cannot, in general be tested by means of the t-distribution as the test statistic

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

follows the t-distribution only under the hypothesis that $\rho = \rho_o = 0$ when the sampling distribution of r reduces to the simple form given above.

Considering again the data of Example 2(a), the estimate of ρ is found to be

$$r = \frac{15.8463}{\sqrt{(24.2310)(16.3730)}} = 0.80$$

and the test statistic

$$t = \frac{(0.80)\sqrt{8}}{\sqrt{0.36}} = 3.77$$

which exceeds the critical value of ±2.31 for 8 degrees of freedom at the 95% confidence level. A significant test here might well have been expected because of the large reduction in sampling variance encountered in Example 2(b).

3b) <u>Testing a partial correlation coefficient</u>

A similar result can be obtained when dealing with multivariate populations where the t-distribution may be used to test the significance of a partial correlation coefficient.

Let $f(x_1, x_2, \ldots x_p)$ be a p-variate normal population. The partial correlation coefficient of order m is defined to be the correlation coefficient between two specified variables after the effects of m of the remaining variables have been eliminated. Partial correlations of order 1 to order p-2 exist in a p-variate population.

If the coefficients C_{ik} are elements of the pxp matrix

$$C = \begin{bmatrix} \sum(x_{1j}-\bar{x}_1)^2 & \sum(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2)\ldots & \sum(x_{1j}-\bar{x}_1)(x_{pj}-\bar{x}_p) \\ \sum(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2) & \sum(x_{2j}-\bar{x}_2)^2\ldots & \\ & \ldots\ldots\ldots\ldots\ldots\ldots & \\ \sum(x_{1j}-\bar{x}_1)(x_{pj}-\bar{x}_p) & & \sum(x_{pj}-\bar{x}_p)^2 \end{bmatrix}$$

then the partial correlation coefficient of order p-2 is

$$r_{ik}.1,\ldots k-1,k+1,\ldots i-1.i+1,\ldots p = \frac{-C_{ik}}{\sqrt{C_{ii}C_{kk}}}$$

Partial correlation coefficients of lesser order may be computed by deleting one of the variables $x_1, x_2, \ldots x_p$. If the variable h, say, is deleted, Cochran has shown that the coefficients of the new matrix may be obtained from the expression

$$C'_{ik} = C_{ik} - \frac{C_{ih}C_{kh}}{C_{hh}}$$

and then the partial correlation coefficient of order p-3 is

$$r_{ik \cdot 1,2,\ldots,h-1,h+1,\ldots i-1,i+1,\ldots k-1,k+1,\ldots p}$$

$$= \frac{-C'_{ik}}{\sqrt{C'_{ii}C'_{kk}}}$$

Fisher[2] has shown that the sampling distribution of the partial correlation coefficient of order m, when a sample of size n is taken from an uncorrelated multivariate normal population, is

$$f(r^2)d(r^2) = \frac{(r^2)^{-\frac{1}{2}}(1-r^2)^{\frac{n-m-2}{2}}}{\beta(1/2, \frac{n-m-2}{2})} \quad d(r^2) \qquad 0 < r^2 < 1$$

which is a $\beta_1(1/2, \frac{n-m-2}{2})$ distribution. Changing the variable to r, the sampling distribution of r is

$$f(r)d(r) = \frac{(1-r^2)^{\frac{n-m-2}{2}}}{\beta(1/2, \frac{n-m-2}{2})} \quad dr \qquad -1 \leqslant r \leqslant 1$$

where the factor 2 introduced by letting $r^2 = r$ has disappeared because the transformation is not 1:1, and the range of r is $-1 \leqslant r \leqslant 1$ rather than $0 \leqslant r^2 \leqslant 1$.

This distribution has the form of that obtained for the correlation coefficient in the test given in 3(a), with a factor n-m-2 instead of n-2. Hence, if

$$t = \frac{r\sqrt{n-m-2}}{\sqrt{1-r^2}}$$

then by the same argument as used in 3(a), the sampling distribution of t has "Student's" t-distribution with n-m-2 degrees of freedom.

This test is carried out in the same way as is the test for the correlation coefficient given in 3(a), subject to the same restriction that the only hypothesis that can be tested is that $\rho = 0$ and the variables are uncorrelated.

4) <u>Testing the significance of a regression coefficient</u>

Consider the experimental situation where the random variable y is believed to be a linear function of several other variables, say, $x_1, x_2, \ldots x_p$ in such a way that the population model is

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \ldots + \beta_p x_{pj} + \epsilon_j$$

where ϵ_j is a random error measuring the failure of the model to account for the variation in y, and is assumed to have a distribution which is $N(0,\sigma^2)$.

The sample model for an observed set $\{y_j, x_{1j}, x_{2j} \ldots x_{pj}\}$ is

$$y_j = b_o + b_1 x_{1j} + \ldots + b_p x_{pj} + e_j$$

where the constants $b_o, b_1, \ldots b_p$ estimate the regression coefficients $\beta_o, \beta_1, \ldots \beta_p$, and e_j estimates ϵ_j. For a sample of size n, the least squares solution for the estimates β_i, $i=0,1,\ldots p$, is

$$b_i = C_{ik} \sum_j (x_{kj} - \bar{x}_k)(y_j - \bar{y}) \quad i,k = 0,1,\ldots p$$

where the variable x_o is a dummy variable introduced for notational convenience and has a value 1 for all j, and where the coefficient C_{ik} are elements of the pxp symmetric positive definite matrix

$$C = \begin{bmatrix} \sum x_{oj} & \sum x_{1j} & \cdots & \sum x_{pj} \\ \sum x_{1j} & \sum x_{1j}^2 & \cdots & \sum x_{1j}x_{pj} \\ \sum x_{2j} & \sum x_{2j}x_{1j} & & \sum x_{2j}x_{pj} \\ \cdots & \cdots & \cdots & \cdots \\ \sum x_{pj} & & & \sum x_{pj}^2 \end{bmatrix}^{-1}$$

The variance of these estimates is

$$\operatorname{var}(b_i) = C_{ii}\,\sigma^2 \qquad i = 0, 1, \ldots p$$

Now y_j is a linear function of ϵ_j, which is $N(0,\sigma^2)$, and of x_i which are non-stochastic variates. Hence

$$E\,[y] = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

and $\operatorname{var}[y] = \sigma^2$

Similarly the estimates, b_i are linear functions of y_j and non-stochastic variables, and therefore, the distribution of b_i is of the same form as the distribution of y. Hence, since y is $N(\beta_o + \beta_1 x_1 + \cdots + \beta_p x_p, \sigma^2)$, the distribution of b_i is also normal with mean

$$E\,[b_i] = \beta_i$$

since the least squares solutions provide unbiassed estimates, and variance $= C_{ii}\sigma^2$. Thus, b_i is $N(\beta_i, C_{ii}\,\sigma^2)$, and it follows from Theorem III-1 that

$$\frac{b_i - \beta_i}{\sqrt{C_{ii}\sigma^2}}$$

is distributed as N(0,1).

Fisher, in developing the analysis of variance procedure, showed that the Total Sum of Squares, a quadratic form of rank n-1, may be partitioned into orthogonal quadratic components,

Total Sum of Squares = $Q_1 + Q_2$

which can be attributed to sources of variation present in the experiment. The Total Sum of Squares is, of course

$$\sum(y_j - \bar{y})^2$$

which is the basis of the estimate of s_w^2, the variance of the sample. This variance can be partitioned into a component, Q_1, due to the regression of y on $x_1, x_2 \ldots x_p$ which is of the rank p, and a component, Q_2, of rank n-p-1, which represents variation which cannot be accounted for by regression and is attributed to the residual or random error. Q_1 is referred to as the Sum of Squares due to Regression (or SSR) and Q_2 is the Sum of Squares for Error (SSE).

Now, $\quad TSS = \sum(y_j - \bar{y})^2$

and if y is distributed as $N(\mu, \sigma^2)$ where μ is estimated by $\bar{y}$, then

$$\frac{y - \bar{y}}{\sigma}$$

is distributed as N(0,1). Applying Cochran's Theorem to

$$\sum \frac{(y_j-\bar{y})^2}{\sigma^2} = Q_1/\sigma^2 + Q_2/\sigma^2,$$

It follows that each member of this equation is distributed as a V^2 variable. In particular, TSS/σ^2 is V^2 with (n-1) degrees of freedom (assuming the population regression coefficient = 0), $Q_1/\sigma^2 = SSR/\sigma^2$ is V^2 with p degrees of freedom and $Q_2/\sigma^2 = SSE/\sigma^2$ is V^2 with n-p-1 degrees of freedom. For SSE/σ^2,

$$E\left[SSE/\sigma^2\right] = n-p-1$$

and $\therefore$

$$E\left[SSE\right] = (n-p-1)\sigma^2$$

Hence an unbiassed estimate of σ^2 is

$$s^2 = \frac{SSE}{n-p-1}$$

$$\text{Let } t = \frac{b_i - \beta_i}{\sqrt{C_{ii}\, s^2}}$$

$$= \frac{\dfrac{b_i - \beta_i}{\sqrt{C_{ii}\,\sigma^2}}}{\dfrac{\sqrt{C_{ii}\, s^2}}{\sqrt{C_{ii}\,\sigma^2}}}$$

But it has been shown that

$$\frac{b_i - \beta_i}{\sqrt{C_{ii}\,\sigma^2}}$$

is N(0,1). Further, it has been shown that

$$\frac{SSE}{\sigma^2}$$

is distributed as V^2 with n-p-1 degrees of freedom.

Therefore

$$\sqrt{\frac{\frac{SSE}{\sigma^2}}{n-p-1}} = \sqrt{\frac{s^2}{\sigma^2}} = \sqrt{\frac{C_{ii}s^2}{C_{ii}\sigma^2}}$$

is distributed as $\sqrt{V^2/(n-p-1)}$ where V^2 has n-p-1 degrees of freedom, since multiplying by the ratio C_{ii}/C_{ii} will not affect the form of the distribution because the coefficients are functions of the non-stochastic variables $x_1, x_2, \ldots x_p$.

Hence

$$t = \frac{b_i - \beta_i}{\sqrt{C_{ii}\ s^2}}$$

is distributed as Student's t-distribution with n-p-1 degrees of freedom.

As an example, consider again the data of Example 2(a), where the data of sample 2 is the only independent variable, x_{1j}, and the data of sample 1, y_j, is believed to be dependent on x_j. For this case, known as simple linear regression, the coefficients of the matrix C are

$$C_{oo} = 1/n - \bar{x}^2/\sum(x_{1j}-\bar{x}_1)^2$$

and $$C_{11} = 1/\sum(x_{1j}-\bar{x})^2.$$

The expressions for the estimates of β_o and β_1 in this case are

$$b_o = \bar{y} - b_1\bar{x}_1$$

$$b_1 = \frac{\sum(x_{1j}-\bar{x}_1)(y_j-\bar{y})}{\sum(x_{1j}-\bar{x}_1)^2}$$

It was found that $b_o = -1.36$ and $b_1 = 0.97$. The tests of hypotheses $H_o : \beta_o = 0$ and $H_o : \beta_1 = 0$ are given below.

$$H_o : \beta_o = 0$$
$$H_1 : \beta_o \neq 0$$

and $t = \dfrac{-1.36}{\sqrt{(1.0436)(4.1468)}} = \dfrac{-1.36}{2.08} = -0.65$

where $C_{oo} = 4.1468$

and $s^2 = 1.0436$ and 8 degrees of freedom.

The critical value of t with 8 degrees of freedom at the 5% level of significance is ±2.306, therefore the test is not significant.

Similarly

$$H_o : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

and $t = \dfrac{0.97 - 0}{\sqrt{(1.0436)(0.0610)}} = \dfrac{0.97}{0.25} = 3.88$

where $C_{11} = 0.0610$.

Comparing the calculated value of t with the critical value, ±2.306, it is seen that this test is significant.

Thus, there is evidence to indicate that $\beta_1 \neq 0$ and that y and x_1 are related by a linear function of the form

$$y = \beta_1 x_1$$

Errors in Testing

Some care is required when interpreting

the result of a test of an hypothesis. Although the null hypothesis under test explicitly specifies only a value of a parameter θ, further implicit assumptions regarding the form of the population are generally involved. The effects of failures of the underlying assumptions will be examined briefly in a subsequent section.

As well as those considerations involved in the null hypothesis, there is the problem of the meaning of "statistical significance". A significant test means that the probability of observing the sample value $\hat{\theta}$ of the population parameter θ when the null hypothesis, $\theta=\theta_o$, is true, is less than some preassigned (generally, small) value, but not that the value $\hat{\theta}$ is impossible. There is evidence in the sample to indicate that the null hypothesis is not true, but it cannot be conclusive, for, due to some unlikely combination of circumstance, a significant value of $\hat{\theta}$ may occur even if $\theta = \theta_o$. Rejecting the null hypothesis when it is true is known as a Type I error, and the probability of this error is designated as α. This error is controlled in that the experimenter is at liberty to choose the magnitude of the risk he is willing to incur in a given experiment. The Type I error is defined by the equation

$$1 - \alpha = \int_{(A)} f(T,\theta_o)\,dT$$

where (A) is the interval in T in which $1-\alpha$ of the total probability of the sampling distribution of T is contained. The end points of this interval are the critical values by which the significance of the test is determined.

Another type of error in testing is the one in which the null hypothesis is accepted when the alternate hypothesis is true. This constitutes a Type II error, the probability of which is usually designated as β. This probability is defined by the equation

$$1-\beta = \int_{(A)} f(T,\theta_1)\, dT$$

where $f(T,\theta_1)$ is the sampling distribution of T under the alternate hypothesis $\theta = \theta_1$, and (A) is the same interval as defined in connection with a Type I error. The following diagram will illustrate these quantities and their physical meaning. Let the curves H_o and H_1 represent the sampling distribution of T for $\theta = \theta_o$ and $\theta = \theta_1$ respectively.

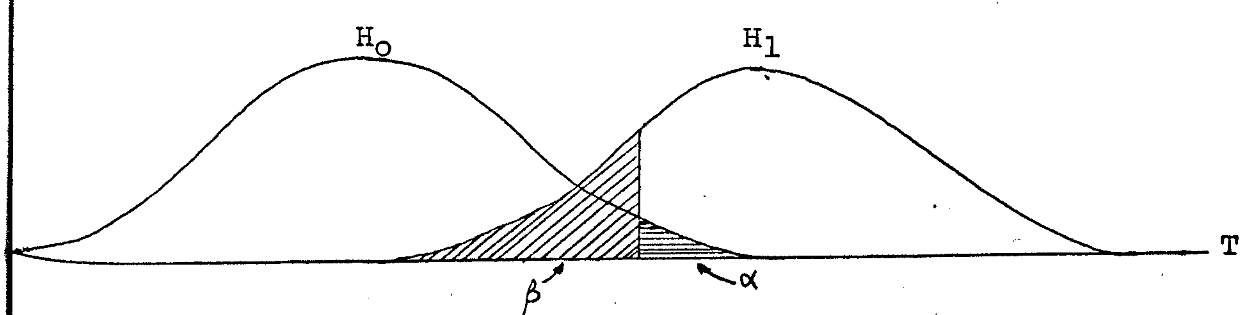

figure 2

It will be noted that as H_o "approaches" H_1 such that the curves are more nearly coincident that, if α is held constant, β increases. Conversely, α increases if β is held constant. Hence the probability of an incorrect decision is always present. The experimenter's only choice is to minimize the risk which is most serious.

The quantity $1-\beta$ as defined above is referred to as the power of a test. It reflects the ability of the test to discriminate between two hypotheses. Ideally, a test maximizes $(1-\beta)$ and minimizes α, or equivalently, minimizes both α and β, but in practice, desirable performance of a test with respect to one or the other of Type I or Type II errors has a corresponding undesirable performance with respect to the other quantity. As H_o and H_1 approach each other, $(1-\beta)$ decreases for a constant value of α until, when H_o and H_1 are coincident, $\alpha = 1-\beta$, and the test is unable to discriminate between H_o and H_1.

Power Function of the t-Test

Using a one-tailed test as an example, the power of the test H_o: $\theta = \theta_o$

H_1: $\theta = \theta_1$, $\theta_1 > \theta_o$

where $t = \dfrac{\bar{x} - \theta_o}{s/\sqrt{n}}$

is

$$1 - \beta = \int_{(A)} f(t,\theta_1)\, dt$$

where $f(t,\theta_1)$ is the sampling distribution of t under $\theta = \theta_1$.

Now

$$t = \frac{\bar{x} - \theta_1}{s/\sqrt{n}} = n \frac{\dfrac{(\bar{x} - \theta_1)}{\sigma}}{\dfrac{\sqrt{n-1}\; s}{\sqrt{n-1}\; \sigma}}$$

or if the alternate hypothesis is rewritten as

$$H_1\colon \theta = \theta_1 = \theta_o - (\theta_1 - \theta_o),\ \theta_1 - \theta_o > 0$$

then

$$t = \frac{\bar{x} - \theta_o - (\theta_1 - \theta_o)}{s/\sqrt{n}}$$

$$= \frac{\sqrt{n}\dfrac{(\bar{x} - \theta_o)}{\sigma} + \sqrt{n}\dfrac{(\theta_1 - \theta_o)}{\sigma}}{\dfrac{\sqrt{n-1}\; s}{\sqrt{n-1}\; \sigma}}$$

Writing

$$\delta = \sqrt{n}\frac{(\theta_1 - \theta_o)}{\sigma}$$

then the statistic t is of the form

$$\frac{w + \delta}{\sqrt{V^2/r}}$$

where w is N(0,1),

V^2 has a chi-square distribution with r degrees of freedom,

δ is a constant.

A statistic of this form is defined to have a sampling distribution of a non-central t, which has probability density function

$$h(f,\delta,t) = \frac{f!\ (1+t^2/f)^{\frac{f+1}{2}}}{2^{(f-1)/2}\Gamma(f/2)\sqrt{\pi f}}\ e^{-\frac{1}{2}\left(\frac{\delta^2}{1+t^2/f}\right)}\ Hh_f\left[\frac{-\delta t/\sqrt{f}}{\sqrt{1+t^2/f}}\right]$$

with $Hh_f(y) = \int_0^\infty \frac{V^f}{f!}\ e^{-\frac{1}{2}(v-y)^2}\ dV,$

known as Airey's Function, and f is the degrees of freedom, and δ the non-centrality parameter.

For $(\theta_1-\theta_0) = 0, \delta = 0$ and the non-central t reduces to the central t, which is Student's t-distribution. Hence the power of the t-test can be determined from the probability integral of the non-central t-distribution. Tables of this integral have been prepared by Lieberman and Resnikoff, and are of a three-entry type - α, δ and degrees of freedom. Other tables have been prepared by various authors and references to these are contained in the bibliography.

It has been shown by Dantzig that the power function of the t-test is not independent of the population variance. This can be seen from the expression for δ, the non-centrality parameter. However, Stein[1], in outlining his two-stage procedure, has shown that by using his procedure, power functions independent of the population variance can be constructured.

The foregoing considerations apply in general to all t-tests where the test statistic can be shown to follow a t-distribution under the alternate hypothesis. This eliminates such tests, however, where the test statistic is distributed as "Student's" t only due to an assumption implicit in the null hypothesis. Such a case is the test for the correlation coefficient, where

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

has Student's t-distribution if and only if $\rho = 0$.

<u>Remarks on Assumptions Underlying the t-test</u>

In discussing the applications of "Student's" distribution, it has been noted that various assumptions, both implicit and explicit, have been made about the parent distributions. Behind every test, for example, is the basic assumption that the parent population follows a normal distribution. Some tests require additional assumptions about population parameters, such as the variance or correlation coefficient. Failure of any of these underlying assumptions technically invalidates the tests and renders conclusions meaningless. However, some rectifying measures can be taken if the extent of the failure of the assumption is known.

1. Non-Normality In The Parent Population

One of the dangers of small sample work is that the assumption of normality is least likely to be met, and is most likely to escape notice because of the relatively small sample size.

If the parent distribution is not normal, the sampling distribution of the t-statistic is complicated by the appearance of parameters which express this deviation from normality. In addition, the sampling distributions of $\bar{x}$ and s^2, the sample mean and variance, are no longer independent.

Various writers have investigated the effect of non-normality on the validity of the t-test from both theoretical and empirical viewpoints. A detailed examination of this work will not be presented here, and the interested reader is referred to the bibliography or to Hey whose paper contains a bibliography of 36 papers on the subject.

In general, the results of these investigations show that significance levels of two tailed tests (Type I error) are not sensitive to skewness or kurtosis in the parent population. The one-tailed test, however, is affected by skewness, particularly if the test is for the differences of means where the groups are of unequal size, and if the skewness is different in each group. If the skewness is the same, and the groups are of the same size, the effect is small.

The effect of non-normality on Type II errors and the power function is not considered to be serious. For symmetrical populations, little effect is found on the power, while for asymmetrical populations, the effect is somewhat greater. For the t-test for the difference of means, skewness has little effect if group size is equal. The power here turns out to be greater or smaller than the normal-theory power depending on whether the sign of $\mu_1 - \mu_2$ is the same as, or opposite to, the sign of the skewness.

Other rectifying measures include the use of transformations of the raw statistical data. The merits of a particular transformation will, of course, depend on the parent population. Many papers dealing with the normalizing of data are available, and some of these have been included in the bibliography.

2. <u>Testing the significance of a difference of means when variances are unequal</u>

The test presented in 2(a) relies on the assumption that the variances σ_1^2 and σ_2^2 of the populations sampled are the same. Since serious errors in the t-test are introduced when these variances are unequal, the assumption should be tested by means of

$$F = s_1^2/s_2^2$$

for example, before applying the test. If the variances are found to be unequal, several alternate tests are available.

The oldest test is the Fisher-Behrens test based on the concept of fiducial probability. The procedure is to calculate

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

The fiducial limits of d for various significance levels have been tabulated by Sukhatme and are also contained in the fourth and subsequent editions of Fisher and Yates' Statistical Tables for Agricultural, Biological and Medical Research. The tables are of the three-entry type, depending on n_1, n_2 and the ratio $\sqrt{n_2}s_1/\sqrt{n_1}s_2$. If the calculated d falls within the tabulated limits, the test is judged non-significant. Considerable controversy exists in the literature regarding the validity of this test, the objection being that the probabilities given do not always reflect the correct value of Type I errors.

Welch has shown that the test criterion

$$t = \frac{u}{\sqrt{gr}}$$

where

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

$$g = \frac{a^2 r_1 + b^2 r_2}{a r_1 + b r_2}$$

$$a = \frac{\sigma_1^2}{n_1+n_2-2}\left[\frac{1/n_1+1/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}\right]$$

$$b = \frac{\sigma_2^2}{n_1+n_2-2}\left[\frac{1/n_1+1/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}\right]$$

and $$r = \frac{(a r_1 + b r_2)^2}{a^2 r_1 + b^2 r_2}.$$

is distributed approximately as Student's t with r degrees of freedom. Unlike other t-tests, this test clearly is dependent on σ_1^2 and σ_2^2, but by substituting for a and b in the expression for r, r becomes

$$r = \frac{(r_1\theta + r_2)^2}{r_1\theta^2 + r_2}$$

where $\theta = \sigma_1^2/\sigma_2^2$

then by considering particular values of θ, the effect of unequal variances on Student's t-test may be assessed and the result interpreted accordingly. Substituting for increasing values of θ has the effect of deviating further and further from the assumption of a common variance.

Two other approximate tests are quite commonly used. Cochran and Cox utilize a weighted mean of the tabular t values for the two samples. First calculate

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

and compare this value with

$$t' = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$$

where t_1 and t_2 are the usual critical values for a t-test at a given level of significance based on n_{i-1} degrees of freedom and $w_i = s_i^2/n_i$.

When $n_1 = n_2$, then $t_1 = t_2 = t$ since t_1 and t_2 are then based on the same degrees of freedom and

$$t' = \frac{t(w_1 + w_2)}{(w_1 + w_2)} = t$$

which is the usual Student t-distribution.

The other test is given by Smith and Satterthwaite. The test statistic t is calculated as in the Cochran and Cox approximation, but here the critical value is the usual tabular t with f degrees of freedom where

$$f = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1/n_1)^2}{n_1-1} + \frac{(s_1/n_2)^2}{n_2-1}}$$

For this same test, Dixon and Massey give a different value of f as

$$f = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} -2$$

This situation can also be attacked by considering the samples as paired values and applying the test given in 2(b). If sample sizes are unequal, however, information must be discarded by deleting n_2-n_1 observations if, for example, $n_2 > n_1$.

CHAPTER VI

RELATED ASPECTS OF THE t-DISTRIBUTION IN MULTI-VARIATE ANALYSIS

In the univariate case, it has been shown that the significance of a population parameter can be tested by comparing the estimate of the parameter with its observed standard deviation, the ratio being distributed as "Student's" t-distribution. For the multivariate case, the significance of a set of population parameters can be determined by an analogous test developed by Hotelling.

Hotelling's T

Where "Student's" t-distribution is the standardized measure of the departure of a sample mean from a population mean, Hotelling's generalization is the standardized measure of simultaneous departure of p sample means from their respective population means.

Let $x_1, x_2, \ldots x_p$ be a p-variate normal population with mean μ_i and covariance matrix σ_{ij}. When σ_{ij} is unknown, it may be estimated from a sample of size n by the matrix

$$(n-1)\ S_{ij} = \sum_k (x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j) \qquad i,j = 1, 2 \ldots p$$
$$k = 1, 2 \ldots n$$

Hotelling has shown that the statistic

$$T^2 = n(n-1) \sum_{ij} \left| S_{ij}^{-1} \right| \quad (\bar{x}_1 - \mu_i)(\bar{x}_j - \mu_j)$$

may be used to test the hypothesis Ho: $\mu_i = \mu_{io}$
Under the null hypothesis, the sampling distribution of T^2 is

$$g(T^2)\,d(T^2) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{n-p}{2})\,(n-1)^{p/2}} \; \frac{(T^2)^{\frac{p-2}{2}}}{[1+T^2/(n-1)]^{n/2}}\, d(T^2)$$

which is a $\beta_2(p/2, \frac{n-p}{2})$ variate. It will be recalled from Chapter IV that the relationship between the F-distribution and a $\beta_2(\frac{n_1-1}{2}, \frac{n_2-1}{2})$ distribution was shown to be

$$\frac{n_1-1}{n_2-1} F = \beta_2\left(\frac{n_1-1}{2}, \frac{n_2-1}{2}\right)$$

Hence, the significance of the above hypothesis can be determined by means of an F test.

For the case p = 2,

$$S = \begin{bmatrix} \sum_k (x_{1k}-\bar{x}_1)^2 & \sum_k (x_{1k}-\bar{x}_1)(x_{2k}-\bar{x}_2) \\ \sum_k (x_{1k}-\bar{x}_1)(x_{2k}-\bar{x}_2) & \sum_k (x_{2k}-\bar{x}_2)^2 \end{bmatrix}$$

$$\text{and } S^{-1} = \begin{bmatrix} 1/(n-1)s_1^2 & -r_{12}/(n-1)s_1 s_2 \\ -r_{21}/(n-1)s_1 s_2 & 1/(n-1)s_2^2 \end{bmatrix}$$

where, as usual

$$s_i^2 = \frac{1}{(n_i-1)} \sum_k (x_{ik}-\bar{x}_i)^2$$

$$r_{ij} = r_{ji} = \frac{\sum_k (x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j)}{\sqrt{\sum_k (x_{ik}-\bar{x}_i)^2 \sum_k (x_{jk}-\bar{x}_j)^2}}$$

thus $T^2 = n(n-1)\left\{\frac{(\bar{x}_1-\mu_1)^2}{(n-1)s_1^2} + \frac{(\bar{x}_2-\mu_2)^2}{(n-1)s_2^2}\right.$

$\left.\frac{-2r_{12}(\bar{x}_1-\mu_1)(\bar{x}_2-\mu_2)}{(n-1)\ s_1\ s_2}\right\}$, and is distributed

with 2 and (n-3) degrees of freedom.

For the case p = 1, Hotelling's T^2 clearly reduces to "Student's" distribution where

$$S = \sum(x_k-\bar{x})^2$$

and $S^{-1} = 1/(n-1)s^2$

so that $T^2 = \frac{n(n-1)(\bar{x}-\mu)^2}{(n-1)s^2} = \frac{n(\bar{x}-\mu)^2}{s^2} = t^2$

which is the square of Student's t.

When the variance-covariance matrix, σ_{ij}, is known, Hotelling has shown that the sampling distribution of the T^2 statistic is that of a V^2 with p degrees of freedom. It will be noticed that this situation exactly corresponds to the univariate case, where, if the population variance is known the statistic

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

will follow a normal distribution.

The properties of this test have been examined by various writers. Hsu and Wijsman have shown that the power function of this test can be determined from the distribution of the non-central F (which, for the case of 1 degree of freedom in the numerator is the square of a non-central t). Hsu has also shown that Hotelling's test is unbiassed, exact and most powerful. A more recent work on the properties of this test is that of Stein[2].

Another inter-relationship is of particular interest here. In a (p + 1)-variate normal population, Fisher[3] has shown that the sampling distribution of the multiple correlation coefficient, R^2, between 1 variable and the p remaining variables can be transformed to the sampling distribution of T^2 by setting

$$1-R^2 = \frac{1}{1 + T^2/(n-1)}$$

$$\text{where } g(R^2)\, d(R^2) = \frac{1}{\beta(\frac{n-p-1}{2}, \frac{p}{2})} (1-R^2)^{\frac{n-p-3}{2}} (R^2)^{\frac{p-2}{2}}\, d(R^2)$$

$$0 < R^2 < 1$$

The relevance of this transformation will be seen in connection with discriminant functions and Mahalanobis's Generalized Distance, Δ^2.

Discriminant Functions

Discriminant functions are an aspect of multivariate analysis developed by British statisticians, notably Fisher. The purpose of this type of analysis is to find a linear function of the sample measurements which will classify the measured object as belonging to a particular parent population. The discriminant function is then

$$x = b_1x_1 + b_2x_2 + \ldots + b_px_p$$

where p measurements on each object are taken. The function is obtained by assuming the existence of a dummy variable, y, which takes on specified values depending on the population with which the measured sample object is associated. A test of significance has been developed, based on the multiple correlation, R^2, between y and $x_1, x_2, \ldots x_p$ which, by means of the transformation given in the previous section, is identical to Hotelling's test.

Mahalanobis' D^2 Statistic

The D^2 statistic, also known as Mahalanobis generalized distance, measures the "distance" in p-dimensional space between two populations. If δ_i is the difference of population means for the ith variate, and σ_{ij} is the population dispersion matrix, assumed to be the same for both populations, then the distance, Δ^2, is

$$\Delta^2 = \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1}^{p} \left| \sigma_{ij}^{-1} \right| \delta_i \delta_j$$

where Δ^2 is considered to be a population parameter. The factor 1/p is due to the anthropological origins of this measure, being a "coefficient of racial likeness".

If σ_{ij} is known, then Δ^2 is estimated as

$$D^2 = \frac{1}{p} \sum_i \sum_j \left| \sigma_{ij}^{-1} \right| d_i d_j$$

where d_i estimates δ_i.

If σ_{ij} is not known, it is estimated from the sample as (n-1)S where S has been defined in connection with Hotelling's T^2 test. The distance is then

$$D_s^2 = \frac{1}{p} \sum_i \sum_j \left| S^{-1} \right| d_i d_j \, .$$

D_s^2 is referred to as the "studentized" distance.

The sampling distribution of the D^2 and D_s^2 statistics have been determined by Bose and Bose and Roy respectively, and have been found in the limiting case to have the distribution of the multiple correlation coefficient, R^2.

<u>Interrelationship of T^2, Discriminant Functions, and D^2</u>

These three lines of research are essentially the same, although differing in approach. Clearly, all are based on the same foundation. The main difference is that Mahalanobis's D^2 provides an estimate of a population parameter, the measure of divergence between two groups. Discriminant functions and Hotelling's T^2 test on the other hand, provide a test of group divergence rather than a measure of group divergence. For a more complete discussion of the various aspects of these tests, and for the proof of the relationships stated, the interested reader is referred to Fisher's [7] account of the uses of these analyses.

BIBLIOGRAPHY

A. BOOKS

Cochran, W. G. and Cox, G. M., Experimental Designs. John Wiley and Sons, New York, 1950.

Czuber, E., Beobachtungsfehler, Teubner, 1891

Dixon, W. J. and Massey, F. J. Jr., Introduction to Statistical Analysis, McGraw-Hill, New York, 1951.

Fisher, R. A., and Yates, F., Statistical Tables for Agricultural, Biological and Medical Research, 4th ed., Oliver and Boyd, London, 1957.

Lieberman, G. J. and Resnikoff, G. J., Tables of the Non-Central t-Distribution, Stanford University Press, 1957

B. PERIODICAL ARTICLES

Bose, R. C., On the Exact Distribution of the D^2 Statistic, Sankhya, Vol. 2:143-145, 1936

Bose, R. C., and Roy, S. N., The Exact Distribution of the Studentized D^2 Statistic, Sankhya, vol. 3, pt. 4, 1938.

Cochran, W. G., The Omission or Addition of an Independent Variate in Multiple Linear Regression, Jour. Roy. Stat. Soc. Supp. 5:171-76, 1938.

Craig, C. C., Chart for Pearson System of Frequency Curves, Ann. Math. Stat., Vol. 7:16, 1936

Dantzig, G. B., On the Non-Existence of Tests of Student's Hypothesis Having Power Functions Independent of σ, Ann. Math, Stat., Vol. 11:186, 1940

Edgeworth, F. Y., The Method of Least Squares, Phil. Mag., Vol. 16:360-75, 1883.

Fisher[1], R. A., Proc. Camb. Phil. Soc., Vol. 21:655-58, 1923.

Fisher[2], R. A., The Distribution of the Partial Correlation Coefficient, Metron, Vol. 3:329-32, 1923

Fisher[3], R. A., Applications of Student's Distribution, Metron, Vol. 5:90-194, 1925.

Fisher[4], R. A., The General Sampling Distribution of the Multiple Correlation Coefficient, Proc. Roy. Soc. A., Vol, 121:654-73, 1928

Fisher[5], R. A., The Fiducial Argument in Statistical Inferences, Ann. Eugenics, Vol. 6:391-98, 1935

Fisher[6], R. A., Uncertain Inference, Proc. Am. Acad. Arts and Sciences, Vol, 71, No. 4:245-258, 1936.

Fisher[7], R. A., The Statistical Utilization of Multiple Measurements, Ann. Eugenics, Vol. 8:376-86, 1938.

Helmert, F. R., Astronomiche Nachrichten, 88, No. 2096:122, 1876

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

Hotelling, Howard, The Generalization of Student's Ratio , Ann. Math. Stat., Vol. 2:360-78, 1931.

Hsu, P. L., Notes on Hotelling's Generalized T, Ann. Math. Stat., Vol. 9:231-243, 1938.

Pearson, E. S., Student as Statistician, Biometrika, Vol. 30:205-50, 1939

Satterthwaite, F. E., An Approximate Distribution of Estimates of Variance Components, Biometrics Bull, Vol 2:110-114, 1946.

Smith, H. F., The Problem of Comparing the Results of Two Experiments with Unequal Errors, Jour. Sci. Ind. Research (India), Vol. 9:211-12, 1936.

Stein[1], C. M., A Two Sample Test For a Linear Hypothesis Whose Power is Independent of the Variance, Ann. Math. Stat., Vol. 16:243-258, 1945.

Stein[2], C. M., The Admissibility of Hotelling's T^2 Test, Ann, Math. Stat., Vol. 27:616-23, 1956

"Student", The Probable Error of a Mean, Biometrika, Vol. 6:1, 1908

Sukhatme, P. V., On Fisher and Behrens Test of Significance for the Difference in Means of Two Normal Samples, Sankhya, Vol. 4:39-48, 1938.

Welch[1], B. L., The Significance of the Difference Between Two Means When the Population Variances are Unequal, Biometrika, Vol. 29:350=62, 1937.

Wijsman, R. A., Random Orthogonal Transformations and Their Use in Some Classical Distribution Problems in Multi-Variate Analysis, Ann. Math. Stat., Vol. 28:415-423, 1951.