K. H. Thompson

BU-98-M

ALGEBRA

Plant Breeding 210-211

1958 - 1959

INTRODUCTION.

A brief introduction to the postulational approach to algebra will be given since it plays the prominent role and has proven to be a powerful tool in the abstract concept of modern algebra.

The mathematician begins with a set of basic "postulates" which he does not prove, indeed cannot prove, but which he uses to build the structure of mathematics. Above all, it is required that this structure be consistent and free of contradiction, and it is this quality that gains for mathematics its reputation as an exact science. The best the mathematician may claim for his science, however, is that it has been tested by every known criterion and appears consistent -- he can never be absolutely certain.

We will now proceed to define the real numbers, and from this set of postulates and a few rules of logic we will be able to prove or deduce all of the known properties of real numbers. This approach seems worthwhile since throughout the course we will be proving or showing to be false a number of theorems, propositions, or statements concerning numbers.

Definition: A complete, ordered field is called a field of real numbers.

Definition: Let F be a set of elements $a,b,c,\cdots$ for which the operations + (addition) and $\cdot$ (multiplication) are defined for any two elements of F. F is then called a field if the following postulates hold:

(i) Closure. If a and b are in F, then $(a+b)$ and $(a\cdot b)$ are in F.

(ii) Commutative laws. For all a and b in F, $a+b = b+a$ and $a\cdot b = b\cdot a$.

(iii) Associative laws. For all a, b, and c in F, $a+(b+c) = (a+b)+c$ and $a\cdot(b\cdot c) = (a\cdot b)\cdot c$.

(iv) Distributive law. For all a, b, and c in F, $a(b+c) = ab+ac$.

(v) Existence of identity elements. F contains an element 0 such that $a+0=a$ for all a in F. F also contains an element $1\neq 0$ such that $a1=a$ for all a in F.

(vi) Existence of inverses. For each a in F, the equation $a+x=0$ has a solution x in F, and the equation $ax=1$ has a solution x in F for all $a\neq 0$.

Definition: A field F is said to be ordered if it contains a set P of "positive" elements such that:

(i) If a and b are in P, then a+b and a•b are also in P.

(ii) For each a in F, precisely one of the following statements is true: a is in P, -a is in P, or a=0.

Definition: An ordered field F will be called complete if and only if every non-empty set S of positive elements of F has a greatest lower bound in F.

Definition: By a lower bound of a set S of elements of an ordered field F is meant an element b (not necessarily in S) such that b is less than or equal to x for all x in S.

Definition: A lower bound b is said to be greatest if for any other element b' greater than b, there is an x in S such that b' is greater than x.

In deducing consequences from the above postulates, we must not contradict certain basic rules of logic. The three basic rules of logic for equality are:

(i) Reflexive law. a=a.

(ii) Symmetric law. If a=b, then b=a.

(iii) Transitive law. If a=b and b=c, then a=c.

We may now construct formal proofs for all of the well-known algebraic "rules" familiar to students of elementary algebra.

RANDOM VARIABLES.

The following definitions and concomitant notation have proven useful in mathematical statistics, and we will adopt them here.

Definition: A random or chance variable X is a measurable function defined from a sample space to the real numbers.

While we will not make this statement precise in a mathematical sense, we may consider an illustration. If X denotes the sum of the numerical values obtained by rolling two dice, then X is a random variable which assumes the integral values 2,3,•••,12. Here, the sample space consists of the 36 possible outcomes of the two dice, and X is a function which maps these outcomes into real numbers, namely the integers 2,3,•••,12.

As in this example, X is usually defined upon a sample space associated with physical experiments where the outcome of any one experiment is based upon chance -- hence the term, random variable. The outcome of an experiment is commonly called a random or a chance event.

A discrete random variable is a random variable that can assume only a finite number, or an infinite sequence, of distinct real numbered values. As an example of a discrete random variable which may assume an infinite sequence of distinct values, consider the sample space consisting of the sample points corresponding to the number of successive tosses of a fair coin before a head appears. These sample points may be designated H, TH, TTH, TTTH, $\cdots$. The number of sample points is infinite, yet the probability that the random variable X assumes the values $1,2,3,4,\cdots$ is given by $\frac{1}{2}$, $(\frac{1}{2})^2$, $(\frac{1}{2})^3$, $(\frac{1}{2})^4$, $\cdots$, and the sum of these sample point probabilities is 1 as it should be.

A continuous random variable is a random variable which may assume any real numbered value within a specified interval or intervals. Random variables associated with experiments involving physical measurements are typically considered as continuous random variables.

Notation: Let Σ stand for "the sum of". Thus, $\sum_{i=1}^{n} X_i = X_1 + \cdots + X_n$ may be read as "the sum of n random variables."

Let Π stand for "the product of." Thus $\prod_{i=1}^{n} X_i = X_1 \cdot \cdots \cdot X_n$ may be read as "the product of n random variables".

Consistent with our definition of a field, we will let $a,b,c,\cdots$ stand for arbitrary real numbers. Occasionally it will be more convenient to use $a_1,a_2,a_3,\cdots$ for the same arbitrary elements.

Examples:

$$\sum_{i=1}^{n} a_i X_i = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

$$\sum_{i=1}^{n} a^i X_i = aX_1 + a^2X_2 + \cdots + a^nX_n$$

$$\sum_{i=1}^{n} (-1)^i X_i = -X_1 + X_2 - \cdots + (-1)^n X_n$$

$$\sum_{i=1}^{n} X_{2i-1} = X_1 + X_3 + \cdots + X_{2n-1}$$

Notice that in the first example "i" is used to "index" the random variables, while in the second example it is used both as an index number and as an exponent of the constant a. This duplicity of use should not prove confusing.

Propositions:

1. $$\sum_{i=1}^{n} cX_i = c\sum_{i=1}^{n} X_i$$

2. $$\sum_{i=1}^{n} (X_i+Y_i) = \Sigma X_i + \Sigma Y_i$$

3. $$\sum_{i=1}^{n} (X_i+a) = \Sigma X_i + na$$

Problems:

1. Show $\sum_{i=1}^{n}(X_i-\bar{x})=0$ if $\bar{x}=(\sum_{i=1}^{n} X_i)/n$.

2. Show $(\Sigma X_i)^2/n=n\bar{x}^2$.

3. Show $\Sigma(X_i-\bar{x})^2=\Sigma X_i{}^2-\dfrac{(\Sigma X_i)^2}{n}$.

4. Show $\Sigma X_i(X_i-\bar{x})=\Sigma X_i{}^2-\dfrac{(\Sigma X_i)^2}{n}$.

5. Show $\Sigma(X_i-a)^2=\Sigma(X_i-\bar{x})^2+n(\bar{x}-a)^2$.

6. Use the result of problem 5 to show that $\Sigma(X_i-\bar{x})^2$ is a minimum sum of squares (i.e. $\bar{x}$ cannot be replaced by any other value c such that $\Sigma(X_i-c)^2$ is less than $\Sigma(X_i-\bar{x})^2$.

7. Let $Y_i=a+bX_i$ and $\bar{y}=(\Sigma Y_i)/n$.

   Show that $\bar{y} = a+b\bar{x}$, and
   $\bar{x} = (\bar{y}-a)/b$

8. If we define the sample variance as $s_y^2=\Sigma(Y_i-\bar{y})^2/n-1$,

   Show that $s_y^2 = b^2 s_x^2$, and
   $s_x^2 = s_y^2/b^2$

9. Show that when n=2
   $$\Sigma(X_i-\bar{x})^2/n-1 = (X_1-X_2)^2/2$$

EXPECTATION.

Definition: Let X be a discrete random variable assuming the values $x_1, x_2, \cdots$ with the corresponding probabilities $p(x_1), p(x_2), \cdots$. The mean or expected value of X is defined as

$$E(X) = \Sigma x_i p(x_i) \ .$$

The expected value of g(X) where g(X) is any function of X, and hence a random variable, is defined as

$$E(g(X)) = \Sigma g(x_i) p(x_i) \ .$$

Thus, expectation is defined to be a mean or a weighted average over all of the possible values that X or g(X) assumes.

Definition: A function p(x) which defines the probability that the discrete random variable X will assume any particular value x in its range is called the probability function of the random variable X.

Definition: The distribution function F(x) of a discrete random variable X is defined as

$$F(x) = P\left\{ X \leq x \right\} = \sum_{x_i \leq x} p(x_i)$$

Example: Let the random variable X be the sum of the two surfaces obtained by rolling two dice. The probability function of this random variable may be given by a set of equations: $P\left\{X=x_i\right\} = p(x_i)$ where $x_i = 2, 3, \cdots, 12$ and $p(x_i) = 1/36, 2/36, \cdots, 1/36$.

For this example

$$E(X) = \Sigma x_i p(x_i) = 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + \cdots + 12\left(\frac{1}{36}\right) = 7$$

$$E(X^2) = \Sigma x_i^2 p(x_i) = 2^2\left(\frac{1}{36}\right) + 3^2\left(\frac{2}{36}\right) + \cdots + 12^2\left(\frac{1}{36}\right) = \frac{329}{6}$$

Propositions:

1. $E(a) = a$
2. $E(aX) = aE(X)$
3. $E(\Sigma X_i) = \Sigma[E(X_i)]$
4. $E(\Sigma a_i X_i) = \Sigma a_i E(X_i)$

Problems:

1. Let the random variable X be the mean of the two surfaces obtained by rolling two dice.
   a) Tabulate the probability function.
   b) Compute $E(X)$
   c) Compute $E(X^2)$

2. Binomial distribution. Let X be the number of successes in n independent trials where at each trial we get a "success" with probability p and a failure with probability $q=(1-p)$. Such trials are called Bernoulli trials, and the "distribution" of this random variable X determined by the following probability function is commonly referred to as the "binomial distribution."

$$p(x:n,p)=\binom{n}{x}p^x q^{n-x} \ , \quad x=0,1,2,\cdots,n \ .$$

   Compute $E(X)=np$.

3. Poisson distribution. The random variable X is said to have the Poisson distribution, if for $x=0,1,2,\cdots$ the probability that X equals x, given a fixed parameter λ, may be expressed as

$$p(x:\lambda)=e^{-\lambda}\frac{\lambda^x}{x!} \ , \quad x=0,1,2,\cdots \ .$$

   Compute $E(X)=\lambda$.

4. Show $E(X)=\mu$ where $X=\Sigma Y_i/n$ and $E(Y_i)=\mu$, $i=1,2,\cdots,n$.

The Variance: If X is a random variable with a distribution defined by $p(x)$, then if it exists,

$$E(X^r) = \Sigma x_i^r p(x_i)$$

is called the r^{th} moment of X about the origin.

The first moment about the origin is called the mean; that is, $E(X)=\mu$.

Definition: The variance of the random variable X is defined to be

$$\mathrm{Var}(X)=E[(X-\mu)^2]=E(X^2)-\mu^2$$

The positive square root of the variance is called the standard deviation of X.

It is common to speak of the variance of the distribution of a random variable instead of the variance of the random variable itself as we have defined it. Thus, for simplicity, we say that the variance of the binomial distribution is npq, whereas a more formal statement would be

$$\mathrm{Var}(X) = npq \qquad \text{where}$$

$$\Pr\left\{X=x:n,p\right\} = \binom{n}{x}p^x q^{n-x} \;, \quad x=0,1,2,\cdots,n \;.$$

To illustrate the computation of the variance of a random variable, let X be the number scored with a fair die. Then,

$$\begin{aligned}\mathrm{Var}(X) &= E(X^2) - (EX)^2 \\ &= [1^2(\tfrac{1}{6})+2^2(\tfrac{1}{6})+\cdots+6^2(\tfrac{1}{6})]-(\tfrac{7}{2})^2 = \tfrac{35}{12} \;.\end{aligned}$$

Problems:

1. Show $\mathrm{Var}[a+bX] = b^2\,\mathrm{Var}(X)$.

2. For Problem 1, page 6, compute $\mathrm{Var}(X)$.

3. Binomial distribution. For Problem 2, page 6, compute $\mathrm{Var}(X) = npq$.

4. Poisson distribution. For Problem 3, page 6, compute $\mathrm{Var}(X) = \lambda$.

Proposition:

If X is a random variable with mean μ and variance $\sigma^2 (\sigma > 0)$, then the random variable Y, where $Y = (\frac{X-\mu}{\sigma})$, has mean 0 and variance 1.

Proof.

$$\begin{aligned}E(Y) &= E(\tfrac{X-\mu}{\sigma}) \\ &= \tfrac{1}{\sigma}[E(X)-\mu] \\ &= \tfrac{1}{\sigma}[\mu-\mu] \\ &= 0\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(Y) &= E(Y^2)-[E(Y)]^2\\ &= E(\frac{X-\mu}{\sigma})^2-0\\ &= \frac{1}{\sigma^2}E[X-\mu]^2\\ &= \frac{1}{\sigma^2}[E(X^2)-\mu^2]\\ &= \frac{1}{\sigma^2}[(\sigma^2+\mu^2)-\mu^2]\\ &= 1\end{aligned}$$

Independence: If X and Y are two random variables on the same sample space, then (XY) is also a random variable. If X and Y are independent random variables, then we say that $\Pr\{X=x:Y=y\} = P\{X=x\}$. In other words, the probability that the random variable X equals a specific value x does not depend in any way upon the random variable Y. This relationship constitutes a necessary and a sufficient condition of independence.

Definition: A function p(x,y) which defines the probability that X=x and at the same time Y=y, for all x in the range of X and all y in the range of Y, is called the joint probability function of X and Y.

We may now say that two random variables X and Y are independent if and only if

$$p(x,y) = p(x)p(y) .$$

Proposition: If X and Y are mutually independent random variables, then

$$E(XY) = E(X)E(Y) .$$

where

$$E(XY) = \sum_{i,j} x_i y_j p(x_i,y_j) ,$$

$$E(X) = \sum_{i,j} x_i p(x_i,y_j) , \quad \text{and}$$

$$E(Y) = \sum_{i,j} y_j p(x_i,y_j) .$$

Definition: The covariance of X and Y is defined as

$$\mathrm{Cov}(X,Y) = E(X-\mu_x)(Y-\mu_y) = E(XY) - \mu_x\mu_y$$

where $E(X) = \mu_x$ and $E(Y) = \mu_y$.

Proposition: If X and Y are mutually independent random variables, then

$$\mathrm{Cov}(X,Y) = 0$$

The converse of this proposition is not true; that is $\mathrm{Cov}(X,Y)=0$ is not sufficient to imply that X and Y are independent.

The computation of the covariance for two independent variables is now illustrated by an example. Let X be the score obtained on one die, and Y be the score obtained on a second die. The joint probability function of X and Y may be given by a set of equations: $P\{X=x_i, Y=y_j\} = p(x_i,y_j)$ where $x_i=1,2,\cdots,6$, $y_j=1,2,\cdots,6$ and $p(x_i,y_j)=1/36$ for each of the 36 unique values, (x_i,y_j). Using the definition $E(XY)=\sum_{i,j} x_i y_j p(x_i y_j)$, we compute

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y)\\ &= \frac{1}{36}[(1)(1)+\cdots+(1)(6)+(2)(1)+\cdots+(2)(6)+\cdots+(6)(1)+\cdots+(6)(6)]\\ &\quad - \left(\frac{7}{2}\right)\left(\frac{7}{2}\right) = \frac{441}{36} - \frac{49}{4}\\ &= 0 \ .\end{aligned}$$

The covariance of two independent random variables is always zero. The covariance of two dependent random variables may be negative, equal to zero, or positive.

We may now investigate the properties of the variance of a sum of n random variables.

Proposition: Let $X_1, X_2, \cdots, X_n$ be random variables such that $\mathrm{Var}(X_i)=\sigma_i^2$, $i=1,2,\cdots,n$. If $Z=\sum_{i=1}^{n} X_i$, then

$$\mathrm{Var}(Z) = \sum_{i=1}^{n}[\mathrm{Var}(X_i)] + 2\sum_{i<j}[\mathrm{Cov}(X_i,X_j)] \ .$$

Corollary: If the random variables X_i and X_j are mutually independent for all $i \neq j$, then we have the extremely useful result:

$$\mathrm{Var}(Z) = \sum_{i=1}^{n} \sigma_i^2$$

That is, the variance of the sum of independent random variables is equal to the sum of the individual variances.

Proposition: Let $X_1, X_2, \cdots, X_n$ be mutually independent random variables such that $E(X_i)=\mu$ and $\mathrm{Var}(X_i)=\sigma^2$ for all $i=1,2,\cdots,n$. Then, if $Y = \sum_{i=1}^{n} X_i/n$, $\mathrm{Var}(Y) = \sigma^2/n$.

Proof.

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}\left(\frac{\Sigma X_i}{n}\right) \\
&= \frac{1}{n^2}\,\mathrm{Var}(\Sigma X_i) \\
&= \frac{1}{n^2}[E(\Sigma X_i)^2 - [E(\Sigma X_i)]^2] \\
&= \frac{1}{n^2}[E(\Sigma X_i^2) + 2E(\sum_{i<j} X_i X_j) - (n\mu)^2] \\
&= \frac{1}{n^2}[\Sigma[E(X_i^2)] + 2\sum_{i<j}[E(X_i X_j)] - n^2\mu^2] \\
&= \frac{1}{n^2}[\Sigma(\sigma^2+\mu^2) + 2\sum_{i<j}[E(X_i)E(X_j)] - n^2\mu^2] \\
&= \frac{1}{n^2}[n\sigma^2 + n\mu^2 + 2\tbinom{n}{2}(\mu)(\mu) - n^2\mu^2] \\
&= \frac{1}{n^2}[n\sigma^2 + n\mu^2 + n(n-1)\mu^2 - n^2\mu^2] \\
&= \frac{1}{n^2}[n\sigma^2 + 0] \\
&= \frac{\sigma^2}{n} .
\end{aligned}$$

STATISTICAL INFERENCE.

In the discussion of the preceding sections concerning random variables and their probability distributions, we have obtained several propositions about random variables by purely logical deduction. No appeal to experience has been necessary, yet if these propositions have been logically deduced from a consistent axiom system, then they are true in this mathematical sense. It is important to realize, however, that mathematical theory can never prove anything about physical events that do, in fact, occur. At the same time, a great part of mathematical theory has been developed as a model for physical phenomena, and this is particularly true of the theory of probability and of random variables and their distributions.

We will now proceed to show how mathematical theory may be applied to problems of statistical inference; that is, how we may infer from physical events certain decisions, based upon an assumed mathematical model.

Consider a random variable X and connected with it a random experiment ξ. The physical outcome of a random experiment is called a random or chance event. To obtain a simple random sample of n observed values $x_1,\cdots,x_n$ of a random variable X, which may be discrete or continuous with a given distribution function F(x), we must perform n independent repetitions of the random experiment ξ. The set of all possible values x that X may assume is called the parent population. If X is continuous, the parent population will be infinite. If X is discrete, the parent population may be finite as X assumes only a finite number of values, or it may be infinite as X assumes an infinite number of discrete values. A simple random sample, derived from successive, independent repetitions of a random experiment, is one which gives each element of the parent population an equal chance of being included in the sample.

The values $(x_1,\cdots,x_n)$ observed in an actual sample may be thought of collectively as an observed value of the n-dimensional random variable $(X_1,\cdots,X_n)$. It follows that the sample mean or any other sample characteristic considered as a function $g(x_1,\cdots,x_n)$ of the n observed values, is itself an observed value of the random variable $g(X_1,\cdots,X_n)$. It is common usage to refer to a sample characteristic such as the mean, both as an actual value computed from the sample values and as a random variable. Considered in this latter sense, as a measurable function or transformation, the sample characteristic is called a statistic.

Theoretically, the distribution of the statistic $g(X_1,\cdots,X_n)$ is mathematically uniquely determined by the joint distribution of $X_1,\cdots,X_n$. Conceptually, if we repeatedly draw independent, random samples of size n from the same parent population and compute for each sample the sample characteristic $g(x_1,\cdots,x_n)$, the sequence of values obtained will constitute a sequence of observed values of the random variable $g(X_1,\cdots,X_n)$. The exact distribution of the statistic $g(X_1,\cdots,X_n)$, determined mathematically or by taking a totality of all possible samples of size n where possible, is called the <u>sampling distribution</u> of the corresponding sample characteristic.

We may now distinguish between sampling with replacement (simple random sampling) and sampling without replacement from a parent population. If the parent population is finite and we sample without replacement, it is clear that the composition of the parent population changes with each successive experiment. If the parent population is very large, the distinction between sampling with replacement and without replacement becomes less and disappears if the parent population is infinite. We will generally confine our attention to simple random samples for theoretical purposes.

The distribution of a sample of size n is defined as the distribution obtained by assigning a mass equal to $1/n$ to each of the sample points $x_1,\cdots,x_n$. This sample distribution may be thought of as a statistical image of the parent population defined by some random variable X and its probability function. Just as we defined and computed a mean and a variance as characteristics of the distribution of X, we may now define and compute the analogous characteristics of the distribution of the sample.

Many problems of statistical inference, determining the agreement between observed facts (the sample) and theory (the mathematical model), take the form of a <u>test of significance</u>. Thus, given a sample of n observed values from which we may compute some sample characteristic $g(x_1,\cdots,x_n)$, we may ask if this value can reasonably be considered as an observed value of some random variable $g(X_1,\cdots,X_n)$ with a known sampling distribution. We begin by assuming the value $g(x_1,\cdots,x_n)$ is in fact an observed value of $g(X_1,\cdots,X_n)$. Using this assumption and the sampling distribution of $g(X_1,\cdots,X_n)$, it is now possible to compute the numerical probability of a deviation between the observed sample characteristic $g(x_1,\cdots,x_n)$ and some hypothesized value of $g(X_1,\cdots,X_n)$, say $E[g(X_1,\cdots,X_n)]$. If the probability of observing a deviation as large or larger than the one actually observed is small, then we may wish to conclude that $g(x_1,\cdots,x_n)$ is <u>significantly different</u> from the hypothesized value of $g(X_1,\cdots,X_n)$. If the evidence is not sufficient to conclude a significant difference, then we may attribute the observed deviation to chance. It is important to realize that this procedure does not constitute a logical proof or disproof of any hypothesis, in marked contrast to the proofs of the various propositions encountered in the earlier sections on random variables.

Another important problem of statistical inference is to determine if the sample distribution itself can reasonably be expected to have been derived from some hypothesized parent distribution. A test of this hypothesis is called a goodness of fit test.

Example: The following test of significance procedure serves as an example of a statistical solution to one important problem in statistical inference. We hypothesize that the sample mean from a sample of size n is an observed value of the statistic $h(X_1,\cdots,X_n)=\sum_{i=1}^{n} X_i/n$, where we assume that $X_1,\cdots,X_n$ are independent and identically normally distributed with mean μ and variance σ^2 (symbolized by $N(\mu,\sigma)$). The sampling distribution of $\sum_{i=1}^{n} X_i/n$ is known and may be determined to be normal with mean μ and variance σ^2/n, $(N(\mu,\sigma/\sqrt{n}))$. This sampling distribution is continuous and is one of a general class of normal distributions. The population characteristics μ and σ^2/n are called parameters and together uniquely determine a specific normal distribution. If we let the random variable $Y=h(X_1,\cdots,X_n)=\sum_{i=1}^{n} X_i/n$, then the frequency function of Y is continuous and is given by the equation

$$F(y) = \frac{1}{\sqrt{\frac{2\pi}{n}}\,\sigma}\, e^{-\frac{n}{2}\left(\frac{y-\mu}{\sigma}\right)^2}, \quad -\infty < y < \infty \ .$$

However, for this problem we want a statistic whose sampling distribution permits us to test the simple hypothesis $H:\mu=c$ against the composite alternative hypothesis $A:\mu\neq c$. Since both μ and σ are unknown here, consider the random variable

$$t = \frac{Z}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} Z_i^2}}$$

where Z, $Z_1,\cdots Z_n$ are all independent and identically $N(0,\sigma)$. The sampling distribution of this random variable is known as <u>Student's t distribution</u>. The frequency function of t is continuous and is given by the equation

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} [1 + \frac{x^2}{n}]^{-(\frac{n+1}{2})} , -\infty < x < \infty \quad .$$

This distribution is symmetric about the point x=0 and is asymptotically normal; that is, as $n \to \infty$ the t distribution approaches the normal distribution $N(0,1)$. The t distribution is independent of σ and is a function of a single parameter n. The number of independent variables $Z_1,\cdots,Z_n$; each $N(0,\sigma)$, is called the number of <u>degrees of freedom</u>. For our definition of t, this number is n.

Assertion: The statistic $g(X_1,\cdots,X_n) = \frac{(\bar{x}-\mu)}{\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{x})^2}{n(n-1)}}}$

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

Proof: We have assumed $X_1,\cdots,X_n$ to be independent and $N(\mu,\sigma)$. It may be shown that $\bar{x}$ is independent of $\Sigma(X_i-\bar{x})^2/n(n-1)$. We have empirical evidence of this fact from the computing laboratory.

Problems:

1. Show that $Z = \sqrt{n}(\bar{x}-\mu)$ has mean 0 and variance σ^2. It is actually true that Z is $N(0,\sigma)$.
2. Show that $Z_1 = \sqrt{\frac{n}{n-1}}\,(X_1-\bar{x})$ has mean 0 and variance σ^2. It is actually true that Z_i is $N(0,\sigma)$ for $i=1,\cdots,n$.
3. Show that the variables $(X_i-\bar{x})$ and $(X_j-\bar{x})$ are not independent; i.e., show that $\mathrm{Cov}[(X_i-\bar{x}), (X_j-\bar{x})] \neq 0$. We may actually show by the use of the appropriate orthogonal transformations that $\sum_{i=1}^{n} \frac{(X_i-\bar{x})^2}{n-1}$ is distributed as the arithmetic mean of the squares of n-1 independent $N(0,\sigma)$ variables.

Hence the assertion is true.

To test the hypothesis $\mu=c$ against all possible alternatives $\mu\neq c$, we now compute a sample value of t; namely $g(x_1,\cdots,x_n)$, substituting the actual sample values x_i for X_i and c for μ in the statistic $g(X_1,\cdots,X_n) = \frac{(\bar{x}-\mu)}{\sqrt{\frac{\Sigma(X_i-\bar{x})^2}{n(n-1)}}}$.

If the x_i are, in fact, observed values of independent random variables X_i; each $N(c,\sigma)$, $i=1,\cdots,n$, then the sample value of t is a valid observed value of the statistic t whose exact sampling distribution has been seen to depend only upon the parameter n-1 and whose expected value is 0.

The probability of observing a deviation from 0 as great as $|g(x_1,\cdots,x_n)|$, assuming the hypothesis to be tested is true, has been conveniently tabulated for many values of n-1. We may now decide whether the deviation is large enough to conclude that the hypothesis to be tested is probably not true, or whether it could have occurred by chance from a sampling distribution for which the hypothesis is true.

ESTIMATION.

In the preceding chapter we discussed the concept of a sample distribution as a statistical image of some parent population, the distribution of which is defined by a random variable X and its frequency function. In the parametric estimation problem, the frequency function of X is unknown and may depend upon one or more unknown parameters. Frequently, it is assumed that the distribution of X belongs to a general class of distributions; e.g. $N(\mu,\sigma)$, and the problem is how we may "best" use the sample values $x_1,\cdots,x_n$ to form estimates of the population parameters.

There exist an infinite number of functions of the sample values, $h(x_1,\cdots,x_n)$, which may be used as estimates of the unknown parameters. To compare two estimates, however, we must compare the sampling distributions of the corresponding functions $h(X_1,\cdots,X_n)$, for it is the average and long run behavior of the estimators which is important and not individual observed values. Unfortunately, there does not exist a general solution to the problem of what properties the sampling distribution of a "good" estimator should possess. We will consider only a very limited aspect of the general theory of parametric estimation here.

Definition: A statistic $h(X_1,\cdots,X_n)$ is said to be an <u>unbiased</u> <u>estimator</u> of the parameter ζ if $E[k(X_1,\cdots,X_n)] = \zeta$.

Thus, an unbiased estimator $h(X_1,\cdots,X_n)$ has a sampling distribution whose mean is the parameter ζ being estimated.

We may also use the variance of $h(X_1,\cdots,X_n)$ as a basis for selecting good estimators within the general class of unbiased estimators.

Definition: An unbiased statistic $h = h(X_1,\cdots,X_n)$ is said to be a <u>minimum variance unbiased estimator</u> of the parameter ζ if

$$E[h-\zeta]^2 \leq E[h'-\zeta]^2 , \quad \text{where } h' \text{ is any unbiased estimator of } \zeta.$$

The heuristic argument associated with this basis of comparison is that an estimator with a small variance usually has a sampling distribution which concentrates a large part of its unit mass in a small neighborhood about the true parameter ζ. The larger this concentration, the smaller is the probability that an estimate will differ from ζ by a large amount. This is a worthwhile criterion, but one must appreciate that the variance of an estimator is not a foolproof measure of the concentration of the sampling distribution about ζ. The same criticism, however, may be applied to any other characteristic of the sampling distribution, and in fact, minimum variance estimators have proven very useful in application and theory.

While it seems reasonable to ask that the average value of an estimator be equal to the parameter being estimated, the restriction of the class of possible estimators exclusively to unbiased estimators is defective, since slightly biased estimators with even lower variances lose their claim for consideration.

There exist many other extremely useful criteria by which we may partition the totality of all estimators, but a unified solution to the general problem of estimation has not been set forth. In this course we will be concerned primarily with minimum variance unbiased estimators, since the chief method of estimation that we will use, the method of least squares, leads to estimators which have this property.

Proposition: If we define the statistic $s^2 = \frac{\Sigma(X_i-\bar{x})^2}{n-1}$, where $X_1,\cdots,X_n$ are independent and identically distributed with mean μ and variance σ^2, then $E(s^2) = \sigma^2$. Hence, s^2 is said to be an unbiased estimator of σ^2 since the mean of the sampling distribution of s^2 is equal to σ^2.

Proof.

$$E(s^2) = \frac{1}{n-1} E[\Sigma(X_i-\bar{x})^2]$$

$$= \frac{1}{n-1} E[\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}]$$

$$= \frac{1}{n-1} [E(\Sigma X_i^2) - \frac{1}{n} E(\Sigma X_i)^2]$$

$$= \frac{1}{n-1} [E(\Sigma X_i^2) - \frac{1}{n} E(\Sigma X_i^2 + 2 \sum_{i<j} X_i X_j)]$$

$$= \frac{1}{n-1} [(\frac{n-1}{n})E(\Sigma X_i^2) - \frac{2}{n} \binom{n}{2} EX_i X_j]$$

$$= \frac{1}{n-1} [(\frac{n-1}{n})\Sigma(EX_i^2) - (n-1)(EX_i)(EX_j)]$$

$$= \frac{1}{n-1}[(n-1)(\sigma^2+\mu^2) - (n-1)\mu^2]$$

$$= \frac{1}{n-1} [(n-1)\sigma^2]$$

$$= \sigma^2$$

Problems:

1. Let $X_1,\cdots,X_n$ be independent and identically distributed random variables with mean μ and variance σ^2.

   a) Show that $E\left[\frac{\Sigma a_i X_i}{\Sigma a_i}\right] = \mu$, where $a_1,\cdots,a_n$ are any constants such that $\Sigma a_i \neq 0$. Hence, $\Sigma a_i X_i/\Sigma a_i$ is a linear unbiased estimator of μ.

   b) Show that $\Sigma a_i X_i/\Sigma a_i$ is equal to $\Sigma X_i/n$ when $a_1=\cdots=a_n=a\neq 0$.

   c) Among all linear unbiased estimators of μ, of the general form $\Sigma a_i X_i/\Sigma a_i$, show that $\Sigma X_i/n$ has minimum variance equal to σ^2/n. First show that $\mathrm{Var}\left[\frac{\Sigma a_i X_i}{\Sigma a_i}\right] = \left[\frac{\Sigma a_i{}^2}{(\Sigma a_i)^2}\right] \sigma^2$. The problem now is to find $a_1,\cdots,a_n$ such that $\Sigma a_i^2/(\Sigma a_i)^2$ is a minimum. Differentiating $\Sigma a_i^2/(\Sigma a_i)^2$ with respect to a particular a_i, say $a_i = a_k$, setting the resulting equation equal to zero, and solving for a_k, we find that $a_k=\Sigma a_i^2/\Sigma a_i$ minimizes $\Sigma a_i^2/(\Sigma a_i)^2$ for each $k=1,\cdots,n$. Since the equation for a_k is independent of k, we may choose the a_i equal to minimize the variance. For the case of equal a_i we know from problem 1(b) and the proposition on page 10 that $\Sigma X_i/n$ is a minimum variance unbiased estimator with variance σ^2/n.

2. Let $X_{i1},\cdots,X_{in_i}$, $i=1,\cdots,n$, be independent and identically distributed random variables with mean μ and variance σ^2.

a) Show that $\frac{\Sigma n_i\bar{x}_i}{\Sigma n_i}$ is an unbiased estimator of μ, where

$$\bar{x}_i = \sum_{j=1}^{n_i} X_{ij}/n_i \quad .$$

b) Show that $\frac{\Sigma n_i\bar{x}_i}{\Sigma n_i} = \frac{\sum\limits_{i}\sum\limits_{j} X_{ij}}{\Sigma n_i}$.

c) Compute $\text{Var}\left[\frac{\Sigma n_i\bar{x}_i}{\Sigma n_i}\right] = \frac{\sigma^2}{\Sigma n_i}$.

d) Show that $E(s_i^2) = \sigma^2$, $i=1,\cdots,n$, where $s_i^2 = \sum\limits_{j=1}^{n_i}\frac{(X_{ij}-\bar{x}_i)^2}{n_i-1}$.

e) Show that $\frac{\Sigma(n_i-1)s_i^2}{\Sigma(n_i-1)}$ is an unbiased estimator of σ^2.

f) Show that $\frac{\Sigma(n_i-1)s_i^2}{\Sigma(n_i-1)} = \frac{\sum\limits_{i}\left[\sum\limits_{j}(X_{ij}-\bar{x}_i)^2\right]}{\Sigma(n_i-1)} = \frac{\sum\limits_{i}\sum\limits_{j} X_{ij}^2 - \sum\limits_{i}\left[\frac{(X_{i.})^2}{n_i}\right]}{\Sigma n_i - n}$

where $\sum\limits_{j=1}^{n_i} X_{ij} = X_{i.}$. You will recognize that the statistic in the last form is frequently used as an estimator of σ^2.

When sample values x_i are substituted for X_i in this expression, which is computationally quite easy, we get a pooled estimate of σ^2 from n samples.

3. Let $X_{11},\cdots,X_{1n_1},X_{21},\cdots,X_{2n_2}$ be independent and identically distributed random variables with mean μ and variance σ^2.

   a) Show that $\mathrm{Var}\,[\bar{x}_1-\bar{x}_2] = (\frac{n_1+n_2}{n_1 n_2})\sigma^2$.

   b) Show that $\left[\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}\right]\left[\frac{n_1+n_2}{n_1 n_2}\right]$ is an unbiased estimator of $(\frac{n_1+n_2}{n_1 n_2})\sigma^2$.

4. Let $X_{11},\cdots,X_{1n_1},X_{21},\cdots,X_{2n_2}$ be independent random variables such that $X_{11},\cdots,X_{1n_1}$ are identically distributed with mean μ and variance σ_1^2 and $X_{21},\cdots,X_{2n_2}$ are identically distributed with mean μ and variance σ_2^2 .

   a) Show that $\mathrm{Var}\,[\bar{x}_1-\bar{x}_2] = \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$.

   b) Show that $\left[\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right]$ is an unbiased estimator of $\left[\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right]$.

CONFIDENCE INTERVAL ESTIMATION.

Estimation of an unknown parameter ζ by an interval of values instead of a point estimate would be very useful if one could attach a probability measure to the statement that such an interval does, indeed, contain ζ. Clearly, to make any probability statement, we must consider the interval as a random variable, for otherwise the probability that ζ is contained in a specified interval (a,b) is either 1 or 0 depending upon whether ζ is or is not in the interval. While we will not consider the theory of estimation by confidence regions in any detail, the construction of a confidence interval for the mean of a normal distribution and the assumptions necessary for a probability statement are instructive.

Example: Given a set of sample values $(x_1,\cdots,x_n)$, assumed to be an observed value of the set of random variables $(X_1,\cdots,X_n)$ where $X_1,\cdots,X_n$ are independent and $N(\mu,\sigma)$, the problem is how to use the sample values to construct a confidence interval relative to the unknown parameter μ. Since σ^2 is also unknown, use is made of the statistic $(\frac{\bar{x}-\mu}{s_{\bar{x}}})$ whose sampling distribution, previously shown to be that of Student's t with n-1 degrees of freedom, is independent of σ^2. There exist an infinite number of solutions for c and d satisfying the following equation

$$\int_c^d f(x)dx = 1-\alpha,\ 0<\alpha<1$$

where $f(x)$ is the frequency function of the random variable t_{n-1}. Since $f(x)$ is a frequency distribution, we may write the equation above as a probability statement, namely

$$P\left[c<\frac{\bar{x}-\mu}{s_{\bar{x}}}<d\right] = 1-\alpha,\ 0<\alpha<1\quad .$$

This probability statement will be valid only if we consider $\bar{x}$ and $s_{\bar{x}}$ as random variables where $X_1,\cdots,X_n$ are independent and $N(\mu,\sigma)$.

Problem: Show that the following two probability statements are equivalent.

$$P\left[-t_{.05,n-1}<\frac{\bar{x}-\mu}{s_{\bar{x}}}<t_{.05,n-1}\right] = .95$$

$$P\left[\bar{x}-t_{.05,n-1}s_{\bar{x}}<\mu<\bar{x}+t_{.05,n-1}s_{\bar{x}}\right] = .95\quad .$$

In the last form we see the familiar confidence statement for the unknown mean of a normal distribution which gives, when sample values are substituted for $\bar{x}$ and $s_{\bar{x}}$, a symmetric interval about $\bar{x}$, the point estimate of μ. $t_{.05,n-1}$ is taken from the tabulated two-tailed t table where

$$\int_{-t_\alpha}^{t_\alpha} f(x)dx = 1-\alpha\quad .$$

The first statement may be deduced from distribution theory and gives the probability that a random variable $t = \frac{\bar{x}-\mu}{s_{\bar{x}}}$ will assume a value between two fixed end points. The second statement, although logically equivalent to the first, seems to be more difficult for students of statistics to interpret verbally; presumably because the concept of a fixed parameter μ in a variable interval appears to be quite different from that of a random variable t in a fixed interval. This difficulty is resolved if we consider the confidence interval, or fiducial interval in the terminology of R. A. Fisher, as a random variable.

It is worth noting that because σ^2 is unknown, we cannot obtain a confidence interval for μ of predetermined length in the fixed sample size case. The length of the interval, in fact, depends upon the random variable $s_{\bar{x}}$. There exist sequential sampling methods however; in particular, Stein's Two Stage Sampling Method, which permits estimation by means of a confidence interval of preassigned length.

LEAST SQUARES ESTIMATION.

It has been previously indicated in the discussion on estimation that the variance of a random variable is a convenient measure of the dispersion of the sampling distribution of that random variable. In general, the smaller the variance, the greater will be the concentration of the sampling distribution about the mean value of the random variable, and conversely. The classical principle of least squares, stating that the "most probable" estimate of an unknown value is the one which minimizes the sums of squares of the deviations of that estimate from the observed values, may be applied to the problem of finding random variables whose sampling distributions have minimum variance among all linear unbiased estimators. More precisely, the method of least squares consists of minimizing the quadratic form

$$\sum_{i=1}^{n} [X_i - \sum_{j=1}^{s} (a_{ij}\zeta_j)]^2$$

with respect to the unknown parameters $\zeta_1,\cdots,\zeta_s$ where $A = [a_{ij}]$ is a n x s matrix of real numbers, $n \geq s$, such that $EX_i = \sum_{j=1}^{s} a_{ij}\zeta_j$, $i=1,\cdots,n$. The random variables $X_1,\cdots,X_n$ are assumed to be uncorrelated and distributed with common variance σ^2; that is, $Cov[X_i,X_j] = \delta_{ij}\sigma^2$.