A STUDY OF BAYESIAN INFERENCE IN MEDICAL DIAGNOSIS

A STUDY OF BAYESIAN INFERENCE IN MEDICAL DIAGNOSIS

By

MICHAEL HERZIG (B. ENG)

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Science

McMaster University

May 1976

MASTER OF SCIENCE (1976) McMASTER UNIVERSITY
(Statistics) Hamilton, Ontario

TITLE: A Study of Bayesian Inference in Medical Diagnosis

AUTHOR: Michael Herzig, B.Eng. (McGill University)

SUPERVISOR: Professor G. D. Anderson

NUMBER OF PAGES: xii, 125

ABSTRACT

Bayes' formula may be written as follows:

$$P(y_i|X) = \frac{P(X|y_i)\cdot P(y_i)}{\sum_{j=1}^{j=K} P(X|y_j)\cdot P(y_j)} \quad \text{where} \tag{1}$$

$$Y = \{y_1, y_2, \ldots, y_K\}$$

$$X = \{x_1, x_2, \ldots, x_k\}$$

Assuming independence of attributes $x_1, x_2, \ldots, x_k$, Bayes' formula may be rewritten as follows:

$$P(y_i|X) = \frac{P(x_1|y_i)\cdot P(x_2|y_i)\cdot \ldots \cdot P(x_k|y_i)\cdot P(y_i)}{\sum_{j=1}^{j=K} P(x_1|y_j)\cdot P(x_2|y_j)\cdot \ldots \cdot P(x_k|y_j)\cdot P(y_j)} \tag{2}$$

In medical diagnosis the y's denote disease states and the x's denote the presence or absence of symptoms. Bayesian inference is applied to medical diagnosis as follows: for an individual with data set X, the predicted diagnosis is the disease y_j such that

$$P(y_j|X) = \max_i P(y_i|X), \quad i=1,2,\ldots,K \tag{3}$$

as calculated from (2).

Inferences based on (2) and (3) correctly allocate a high proportion of patients (>70%) in studies to date, despite violations of the independence assumption. The aim of this thesis is modest, (i) to demonstrate the applicability of Bayesian inference to the problem of medical diagnosis (ii) to review pertinent literature (iii) to present a Monte Carlo method which simulates the application of Bayes' formula to distinguish among diseases (iv) to present and discuss the results of Monte Carlo experiments which allow statistical statements to be made concerning the accuracy of Bayesian inference when the assumption of independence is violated.

The Monte Carlo study considers paired dependence among attributes when Bayes' formula is used to predict diagnoses from among 6 disease categories. A parameter which measures deviations from attribute independence is defined by $DH=\left(\sum_{j=1}^{j=6} |P(x_B|x_A,y_j)-P(x_B|y_j)|\right)/6$, where x_A and x_B denote a dependent attribute pair. It was found that the correct number of Bayesian predictions, M, decreases markedly as attributes increasing diverge from independence, ie, as DH increases. However, a simple first order linear model of the form $M = B_0+B_1 \cdot DH$ does not consistently explain the variation in M.

ACKNOWLEDGEMENTS

I owe special thanks to Gary Anderson, who suggested this Bayesian inference study as a topic and has given me most generously of his time, advice, and encouragement. I also wish to thank the Department of Epidemiology and Biostatistics for allowing me the use of their resources. I appreciate the interest shown by Denis Lyons in leading me to recent literature, and am grateful to Bill Pawluk for his explanations of medical procedures. I also take this opportunity to thank Michael Stephens for his support and guidance during my stay at McMaster, and Rick MacLeod for sharing his insights in probability theory. I am grateful to Sue Anderson for typing the final copy rapidly and cheerfully from my less-than-perfect manuscript.

Most of all, thanks to Pat and Karen for their love and friendship, and for making life an adventure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Medicine is a science of uncertainty and an art of probability. *-Sir William Osler*

INTRODUCTION TO CHAPTER I

The diagnosis of disease from clinical data (signs, symptoms, laboratory tests, etc) is considered by the medical profession as a subtle art which may be mastered only after years of study and extensive personal experience. The reasoning process, and associated complexities, by which physicians arrive at a diagnosis have been outlined by several authors. See, eg., Feinstein (1963), Ledley and Lusted (1960). It has also been suggested that physicians could improve their utilization of clinical data if they used probability theory in the analysis of diagnostic problems. See, eg, Anderson and Boyle (1968), Crooks et al (1959), Gustafson and Throckmorton (1965), Hall (1967), Overall and Williams (1961), Wartak (1974).

The theme of Chapter I may be stated as follows: medical diagnosis is a logical reasoning process that can be simulated by Bayes' formula.

Section 1.1 outlines the traditional diagnostic process. Section 1.2 presents Boolean algebra as a simplistic mathematical model which parallels the physician's thinking process. Section 1.3 a develops Bayes' formula as a practical mathematical model to simulate the diagnostic

process. Section 1.3b demonstrates the application of Bayesian estimation to "diagnose" an individual described by a set of symptoms.

Let us weed through the old to discover the new.

- Mao Tse Tung

Section 1.1

An outline of the diagnostic method

Nash (1954) has described how a medical doctor arrives at a diagnosis:

1. The patient's history and physical examination may immediately bring about recognition of the disease. The capacity of recognition diagnosis is usually a valuable possession of a physician who has had exposure to the disease in his medical practice.
2. The patient may present signs and symptoms that enable the physician to formulate some hypothesis which he can test by laboratory investigations.
3. The physician may recall from his practice or from textbooks some diseases which may be responsible for the patient's signs and symptoms, and decide which disease best explains the patient's condition. Problems arising here are (i) the patient rarely presents a full set of symptoms described in a text (ii) the physician may temporarily forget some diseases which may be responsible for the patient's condition (iii) some diseases are not considered because they are not known to the physician (iv) the physician's diagnosis may be influenced by experiences with other patients in the recent past.

4. The physician may choose a small number of signs or symptoms which he feels are particularly significant and, perhaps with the aid of texts, attempt to discover what disease occurs most frequently as a cause. Such a procedure is an unwieldly task with or without the aid of texts.

The task of assessing the significance of all of the patient's signs and symptoms and relating them to a particular disease may prove to be far from trivial. Thus, motivation exists to find an aid to the physician which would improve his diagnostic capabilities. Nash (1954) and others (see Lipkin and Hardy (1958), Ledley and Lusted (1959)) proposed mechanical devices which return diagnostic possibilities for a given list of signs and symptoms. More recently, decision rules based on small numbers of key attributes[1] have been developed (see Teather (1974), Teather and Hilder (1975)).

If a man will begin with certainties he shall end in doubts.

- Francis Bacon

Section 1.2

Boolean algebra as a mathematical model of the diagnostic process.

Ledley and Lusted (1959), (1960) utilized symbolic logic to systemize the reasoning process that enables a physician to arrive at a diagnosis. For the purposes of medical diagnosis we consider attributes x and diseases y as shown in Table 1.2.1 below.

Table 1.2.1

Definition of Logical Terms

Symbol	Name	Interpretation
$\bar{x}_i$	negation	not x_i
$y_j \cdot x_i$	logical product	y_j and also x_i
$y_j + x_i$	logical sum	y_j or x_i or both
$y_j \rightarrow x_i$	implies	if y_j then x_i

The concepts inherent in medical diagnosis are (1) medical knowledge (2) attributes presented by the patient (3) the final diagnosis. Medical knowledge is information about relationships between diseases and attributes, while

attributes presented by the patient are information about that particular individual. A diagnosis is made with these two sources of information and by logical reasoning (formalized in Boolean algebra).

In general a set of k attributes $\{x_1, x_2, \ldots, x_k\}$ and a set of K diseases $\{y_1, y_2, \ldots, y_K\}$ will be under consideration. The relationships between attributes and diseases that compromise medical knowledge can be expressed as a Boolean function of the sets X and Y,

$$E\,[x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_K] .$$

The attributes presented by a patient can be expressed as a Boolean function of the set X,

$$G\,[x_1, x_2, \ldots, x_k].$$

The final diagnosis can be expressed as a Boolean function of the diseases,

$$F\,[y_1, y_2, \ldots y_K].$$

Once E and G are specified the logical aspect of medical diagnosis is to determine the function F such that the following Boolean equation is satisfied:

$$E \rightarrow (G \rightarrow F). \qquad (1.2.1)$$

In words, F represents the disease or diseases that a patient may have if we consider medical knowledge, E, together with what is known about the patient, G.

For an illustrative example, consider the case of two diseases and two attributes. Suppose medical knowledge is defined by E,

$$E=[y_2 \rightarrow x_1] \cdot [y_1 \cdot \bar{y}_2 \rightarrow x_2] \cdot [y_1 \cdot y_2 \rightarrow x_2] \cdot [x_1 + x_2 \rightarrow y_1 + y_2] \qquad (1.2.2)$$

In words,

1. if a patient has y_2 then he displays x_1, also
2. if a patient has y_1 and does not have y_2, then he displays x_2, also
3. if a patient has both y_1 and y_2, then he displays x_2, also
4. if a patient displays either x_1 or x_2 or both x_1 and x_2, then he has y_1 or y_2 or both y_1 and y_2.

Suppose the patient is described by G,

$$G = x_1 \cdot x_2 \qquad (1.2.3)$$

In words, the patient displays both x_1 and x_2.

A logical basis lists all conceivable attribute combinations and all conceivable disease combinations, as shown in Table 1.2.2 and Table 1.2.3. A zero indicates the attribute does not occur, a one indicates the attribute does occur.

Table 1.2.2
Logical Basis for x_1 and x_2

	C^0	C^1	C^2	C^3
x_1	0	1	0	1
x_2	0	0	1	1

$C^0 = \bar{x}_1 \cdot \bar{x}_2$, $C^1 = x_1 \cdot \bar{x}_2$, etc.

Table 1.2.3
Logical Basis for y_1 and y_2.

	C_0	C_1	C_2	C_3
y_1	0	1	0	1
y_2	0	0	1	1

$C_0 = \bar{y}_1 \cdot \bar{y}_2$, $C_1 = y_1 \cdot \bar{y}_2$, etc.

The columns in Table 1.2.2 and Table 1.2.3 represent an exhaustive and mutually exclusive list of all conceivable attribute combinations and disease complexes. For k attributes and K diseases there are 2^{k+K} conceivable attribute-disease combinations, for the case of 2 attributes and 2 diseases there are 16 attribute-disease combinations. They are listed in Table 1.2.4.

Table 1.2.4

Logical basis for x_1, x_2, y_1, y_2

	C^0	C^1	C^2	C^3	C^0	C^1	C^2	C^3	C^0	C^1	C^2	C^3	C^0	C^1	C^2	C^3
x_1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
x_2	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
y_1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
y_2	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
	C_0				C_1				C_2				C_3			

C_j^i $i = 1,2,3,4$, $j = 1,2,3,4$ will denote $C^i \cdot C_j$

$C_0^0 = \bar{x}_1 \cdot \bar{x}_2 \cdot \bar{y}_1 \cdot \bar{y}_2$, $C_1^0 = \bar{x}_1 \cdot \bar{x}_2 \cdot y_1 \cdot \bar{y}_2$, etc.

The role of E is to reduce the logical basis for attributes and diseases from the set of all conceivable to all possible attribute-disease combinations. For example, the first item of information in E(1.2.2) is $y_2 \rightarrow x_1$, thus columns C_2^0, C_2^2, C_3^0, and C_3^2 of Table 1.2.4 are inadmissible. E eliminates all but five attribute-disease combinations, as shown in Table 1.2.5.

Table 1.2.5

Reduced Basis Includes Medical Knowledge

	$c^0c^1c^2c^3$	$c^0c^1c^2c^3$	$c^0c^1c^2c^3$	$c^0c^1c^2c^3$
x_1	0	0 1	1	1
x_2	0	1 1	0	0
y_1	0	1 1	0	1
y_2	0	0 0	1	1
	C_0	C_1	C_2	C_3

By inspection of Table 1.2.5, the function F which statisfies $E \rightarrow (G \rightarrow F)$ when $G = x_1 \cdot x_2$ is $F = y_1 \cdot \bar{y}_2$, ie, the final diagnosis is the presence of disease y_1 only.

To summarize, the logical process by which a physician arrives at a diagnosis is susceptible to precise analysis, and the operations to obtain F, once E and G are specified, may be performed by a digital computer. However, the method presented here is not particularly suited for direct application as medical texts (correctly) use words such as "frequently", "very often", or "almost always" to describe attribute-disease associations[2]. The next step then is to consider a probabilistic analysis of medical diagnosis.

If a man will be content to begin with doubts he shall end in certainties. - *Francis Bacon*

Section 1.3a

An exposition of elementary concepts in probability theory. Bayes' formula is developed as a mathematical model which simulates the diagnostic problem more realistically than the previous Boolean algebra model.

Let Y define a set of K mutually exclusive and exhaustive disease entities which occur in a population. Thus an individual sampled from the population is assigned to one, and only one, disease category of the set Y. To allow for the exhaustive case a particular member of the set Y may identify (i) the normal health state, (ii) an undetermined disease state, or (iii) combinations of diseases.

The event y_i is the occurrence of a particular disease of the set Y in an individual, and the probability of y_i with respect to a specified population of size N is given by

$$P(y_i) = N_{y_i}/N \quad \text{where}$$

N_{y_i} is the number of individuals with disease y_i in the population.

To illustrate, suppose a clinic has been collecting data on referrals for several years. Then these patients can (in a very non-rigorous sense) be assumed to be a random sample of

referrals to the clinic for past and future times. The number of individuals with disease y_1 has been found to be m_1, the number of individuals with disease y_2 has been found to be m_2, etc. Then if $P(y_i)$ is the probability that a new referral to the clinic will have disease y_i, $P(y_i)$ can be estimated by

$$\hat{P}(y_i) = m_i / \sum_{j=1}^{j=K} m_j$$

$P(y_i)$ is termed the a priori or prior probability of disease y_i. Returning to the first paragraph of this section,

$$Y = \{y_1, y_2, \ldots, y_K\}$$

$$P(y_i) > 0 \qquad i = 1,2,\ldots,K$$

The exhaustive case implies

$$\sum_{i=1}^{i=K} P(y_i) = 1,$$

the mutually exclusive case implies

$$y_i \cap y_j = \emptyset \qquad i,j = 1,2,\ldots,K \qquad i \neq j$$

A patient may be described by a set X consisting of attributes which quantify or qualify symptoms, signs, laboratory findings and other data[3] which aid in assigning a patient to a particular disease category. It is convenient to assume that attributes occurring in a profile are independent; for the moment we may think of independence to mean that the occurrence of any attribute in a profile does not influence the occurrence of any other attribute in the profile. Suppose

there are k attributes in a profile, and the event x_i is the occurrence of a particular attribute: then

$$X = \{x_1, x_2, \ldots, x_k\}.$$

The probability of profile X occurring in an individual in a specified population is given by:

$$P(X) = N_X/N \quad \text{where}$$

N_X is the number of individuals with profile X in the population

N is the number of individuals in the population.

The probability of profile X occurring in an individual who is known to have disease y_i is given by

$$P(X|y_i) = N_{Xy_i}/N_{y_i} \quad \text{where}$$

N_{Xy_i} is the number of individuals in the population having both disease y_i and profile X

N_{y_i} as defined previously.

It follows that:

$$P(X|y_i) = P(X,y_i)/P(y_i) \quad \text{where} \qquad (1.3.1)$$

$$P(X,y_i) = N_{Xy_i}/N$$

Similarly,

$$P(y_i|X) = P(X, y_i) \; / \; P(X) \qquad (1.3.2)$$

If y consists of mutually exclusive and exhaustive events, then

$$P(X) = \sum_{j=1}^{j=K} P(X|y_j)\cdot P(y_j) \qquad (1.3.3)$$

Substituting (1.3.1) and (1.3.3) into (1.3.2) we arrive at an expression of Bayes' Theorem:

$$P(y_i|X) = P(X|y_i)\cdot P(y_i) / \sum_{j=1}^{j=K} P(X|y_j)\cdot P(y_j) \qquad (1.3.4)$$

If X consists of independent events then

$$P(X|y_i) = P(x_1|y_i)\cdot P(x_2|y_i)\cdot \ldots \cdot P(x_k|y_i) \qquad (1.3.5)$$

and (1.3.4) may be rewritten as

$$P(y_i|X) = \frac{P(x_1|y_i)\cdot P(x_2|y_i)\cdot \ldots \cdot P(x_k|y_i)\cdot P(y_i)}{\sum_{j=1}^{j=K} P(x_1|y_j)\cdot P(x_2|y_j)\cdot \ldots P(x_k|y_j)\cdot P(y_j)} \qquad (1.3.6)$$

$P(y_i|X)$ is termed the a posteriori or posterior probability of y_i. Further reference to Bayes' formula implies (1.3.6).

There is no more common error than to assume that, because prolonged and accurate calculations have been made, the application of the result to some fact of nature is absolutely certain. - *A.N. Whitehead*

Section 1.3b

An example illustrating the application of Bayes' formula, assuming attribute independence.

Suppose a population of 1000 individuals is examined, and it is found that the number of individuals with diseases y_1, y_2, and y_3 are 300, 200, and 500 respectively. Prior probabilities are:

$$P(y_1) = N_{y_1}/N = 300/1000 = .3$$

$$P(y_2) = N_{y_2}/N = 200/1000 = .2$$

$$P(y_3) = N_{y_3}/N = 500/1000 = .5$$

During the course of medical examinations various symptoms are noted and entered into the patient's medical record. Hypothetical data obtained for our population of 1000 is summarized in Table 1.3.1. The information available from Table 1.3.1 is used to construct Table 1.3.2.

Table 1.3.1

Number of Individuals with Indicated Sign or Symptom, By Disease Category

Sign or Symptom	In Disease Category y_1	In Disease Category y_2	In Disease Category y_3
1. Age			*
Under 21	240	20	48
21 to 50	30	40	384
Over 50	30	140	48
2. Sugar Level			
Normal	30	160	400
Abnormal	270	40	100
3. Weight, during past 3 months			
gain in excess of 7 lbs.	30	40	150
loss in excess of 7 lbs.	30	140	350
no weight change in excess of 7 lbs	240	20	0
4. Kinetic movements	*		
Present	175	20	50
Absent	75	180	450

*some individuals in the indicated disease category were not examined for the indicated sign or symptom.

Probability values in Table 1.3.2 are based upon the relative frequencies with which attributes are found among the patients in each disease category. For example,

$$P(x_1|y_1) = 240/(240+30+30) = .8$$

$$P(x_8|y_1) = 175/(175+75) = .7$$

$$P(X_5|y_3) = 100/(400+100) = .2$$

Table 1.3.2

Hypothetical Data of Table 1.3.1 Expressed As Conditional Probabilities

Characteristic*	Attribute Set Number	Attribute	$P(x_i\|y_j)$ j=1	j=2	j=3
Age	1	x_1: <21	.8	.1	.1
		x_2: 21-50	.1	.2	.8
		x_3: >50	.1	.7	.1
Sugar level	2	x_4: normal	.1	.8	.8
		x_5: abnormal	.9	.2	.2
Weight	3	x_6: gain in excess of 7 lbs during past 3 months	.1	.2	.3
		x_7: loss in excess of 7 lbs during past 3 months	.1	.7	.7
Kinetic Movements	4	x_8: present	.7	.1	.1
		x_9: absent	.3	.9	.9

*a sign, symptom, laboratory test, etc.

Some attributes (x_1-x_3, x_4-x_5, x_6-x_7, x_8-x_9) form groups which describe a particular characteristic. Such arrangements of attributes differ from profiles and will be

referred to as "attribute sets". Thus, Table 1.3.2 presents nine attributes distributed among four attribute sets. Typical patient profiles are:

$$S_1 = \{x_2, x_4, x_6, x_8\}$$
$$S_2 = \{x_1, x_5, x_8\}$$
$$S_3 = \{x_4, x_7, x_9\}$$

Before utilizing Bayes' formula to predict a diagnosis for an individual with profile S_2, let us consider the following:

1. While members of an attribute set are mutually exclusive they are not necessarily exhaustive. Attribute sets 1, 2, and 4 (see Table 1.3.2) consist of an exhaustive description of events for each disease category - for example, an individual whose age has been determined must be assigned one of the attributes of set 1, similarly, information about sugar level is described by x_4 or x_5, and information about kinetic movements is described by x_8 or x_9. However, since an individual may remain constant in weight, attribute set 3 does not describe an exhaustive set of events[4]. An attribute set is exhaustive for a particular disease y_j if $\sum_{i=1}^{i=\ell} P(x_i|y_j)=1$, where ℓ is the number of attributes in the set describing a particular characteristic.
2. We wish to enter as much information as possible into Bayes' formula when such information helps to discriminate among disease categories (see Section 3.3a). Letting x_{10} represent the attribute "no weight change in excess of

7 lbs during past 3 months",

$$P(x_{10}|y_1) = 1 - P(x_6|y_1) - P(x_7|y_1)$$
$$= 1 - .1 - .1$$
$$= .8$$

Similarly,

$$P(x_{10}|y_2) = .1$$
$$P(x_{10}|y_3) = 0$$

Profile S_2 should be replaced by $S_4 = \{x_1, x_5, x_8, x_{10}\}$. To summarize, attributes describing a characteristic should define an exhaustive set, and should be mutually exclusive.

3. An individual may not be examined for every characteristic, thus a profile may consist of fewer attributes than there are attribute sets. For example, age was not determined for the patient with profile S_3.

The application of Bayes' formula, assuming attribute independence, is as follows: replacing S_2 by S_4, $S_4 = \{x_1, x_5, x_8, x_{10}\}$,

$$P(y_1|S_4) = \frac{P(x_1|y_1) \cdot P(x_5|y_1) \cdot) (x_8|y_1) \cdot) (x_{10}|y_1) \cdot P(y_1)}{\sum_{j=1}^{j=3} P(x_1|y_j) \cdot P(x_5|y_j) \cdot P(x_8|y_j) \cdot P(x_{10}|y_j) \cdot P(y_j)}$$

$$= \frac{(.8)(.9)(.7)(.8)(.3)}{[(.8)(.9)(.7)(.8)(.3)+(.1)(.2)(.1)(.1)(.2) +(.1)(.2)(.1)(0.)(.5)]}$$

$$= .9997$$

Similarly,

$$P(y_2|S_4) = .0003$$

$$P(y_3|S_4) = 0$$

The presence of attribute x_{10} in profile S_4 logically (as well as mathematically) eliminates y_3, and the disease with the highest a posteriori probability, y_1, is the predicted diagnosis.

SUMMARY OF CHAPTER I

Chapter I has identified the problem of medical diagnosis: a physician is confronted with a set of symptoms, he considers the symptoms singly or in groups, and arrives at a diagnosis. Alternatively, he may wait for additional data such as laboratory tests, surgical procedures, or consultation before arriving at a diagnosis. Boolean algebra has been demonstrated to be particularly suited to analyze the logical process by which a physician arrives at a diagnosis, but unsatisfactory as a practical diagnostic aid. Bayes' formula has been developed as a mathematical model which parallels medical decision making, and a hypothetical example has been included to illustrate the type of data necessary to apply Bayes' formula.

CHAPTER II

To prepare biscuits: For one serving: 1/2 cup of ready mixed biscuit flour. Follow directions on the package for mixing the dough. - *Boy Scout Handbook, 1959.*

INTRODUCTION TO CHAPTER II

We have established, from theoretical considerations, that Bayes' formula is applicable to the problem of medical diagnosis. Chapter II marks a transition from theoretical considerations to issues associated with the practical application of Bayes' formula.

The theme of Chapter II may be stated as follows: if Bayes' formula is to serve as a practical diagnostic aid, then (i) probability values appearing on the right-hand side of (1.3.6) must be valid for the problem at hand, assuming attribute independence (ii) the digital computer should be used as a labour saving device.

Section 2.1 discusses properties of a priori and conditional probabilities used for diagnostic prediction. Section 2.2 considers the problem of obtaining data to estimate probability values, and introduces the computer as a valuable aid in storing and manipulating data. Section 2.3 discusses the methodology of research investigating the feasibility of Bayes' formula as a diagnostic aid.

It is never possible to step twice into the same river.

- Heraclitus

Section 2.1

Properties of a priori and conditional probabilities used for diagnostic prediction.

An attribute-disease (or S-D) matrix of the form shown in Table 2.1.1 is a convenient representation of statistical information concerning the occurrence of attributes and diseases in a specified population.

Table 2.1.1

The Usual Form of an S-D Matrix

Disease List	A Priori Probabilities	Attribute List x_1	x_2	x_k
y_1	$P(y_1)$	$P(x_1\|y_1)$*	$P(x_2\|y_1)$	$P(x_k\|y_1)$
y_2	$P(y_2)$	$P(x_1\|y_2)$	$P(x_2\|y_2)$	$P(x_k\|y_2)$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
y_K	$P(y_K)$	$P(x_1\|y_K)$	$P(x_2\|y_K)$	$P(x_k\|y_K)$

*$P(x_i|y_j)$ i=1,2,...,k, j=1,2,...,K are referred to as conditional probabilities.

The diseases listed in a S-D matrix usually identify sub-categories of a principle disease, eg., types of bone cancer, types of liver conditions, etc. A priori probabilities

take account of environmental aspects - geographical location, seasonal influence, occurrence of epidemics, community hygiene standards, and so forth. Operationally this means that for a specified disease y_j, $P(y_j)$ values estimated from large random samples of geographically separated populations may be quite different from one another. Documentation supports this view. See, for example, Kendrick (1974), Van Zwanenberg (1974). Some data illustrating the influence of environmental factors on prior probabilities are presented in the following tables and charts.

Table 2.1.2

Urban-rural differences in incidence of cancer of selected sites, Denmark, 1948 to 1952, and Iowa, 1950 †

	Denmark			Iowa	
Site and sex	Copenhagen	Provincial towns	Rural areas	Urban area	Rural areas
Lip (M)	2.9	4.7	8.1	17.3	15.7
Tongue (M)	1.4	1.0	0.3	1.6	1.1
Pharynx (M)	1.9	0.8	0.3	4.3	1.5
Esophagus (M)	8.0	5.1	3.8	3.9	2.1
Stomach (M)	37.0	43.4	41.8	24.7	24.7
Stomach (F)	23.6	26.7	29.8	13.1	14.1
Rectum (M)	25.2	22.5	17.4	16.3	12.4
Rectum (F)	13.9	12.2	11.2	10.0	11.0
Lung (M)	43.6	15.3	8.3	29.0	10.2
Kidney (M)	8.8	4.4	3.7	4.8	3.5
Kidney (F)	5.2	3.0	3.4	3.3	2.8
Bladder (M)	15.3	7.7	5.6	20.2	12.9
Bladder (F)	4.1	2.4	2.1	6.6	3.2
Leukemia‡ (M)	8.4	7.0	6.6	14.4	11.4
Leukemia‡ (F)	6.0	5.0	4.5	10.9	7.2
Cervix uteri (F)	38.4	34.9	20.2	43.4	23.6
Corpus uteri (F)	14.4	10.9	7.9	11.0	10.9
Breast (F)	57.4	49.3	40.9	78.0	62.4
All sites (M)	284.8	224.7	185.7	350.6	250.2
All sites (F)	272.5	241.5	208.6	351.6	261.5

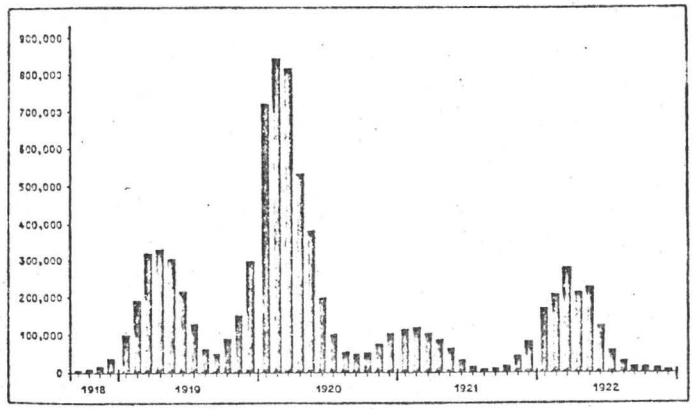
† Annual incidence rates per 100,000, standardized to the age distribution of the total United States population in 1950.

‡ For Iowa, these rates include other cancers of the hemopoietic system.

From MacMahon and Pugh, (1970).

Figure 2.1.1

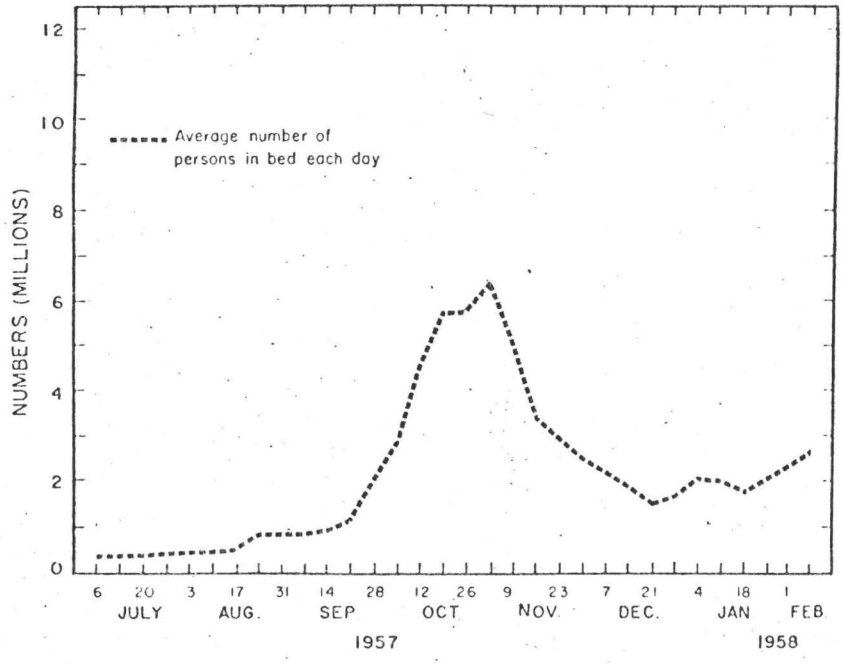
Seasonal and Yearly Variation in the Occurrence of Typhus During Epidemic Years in Russia, 1918-1922



From Rodenwaldt (1952)

Figure 2.1.2

Monthly Variation of Bed Disability Due to Acute Respiratory Disease in the United States, July 1957-Feb. 1958



From MacMahon and Pugh (1970)

Ledley and Lusted (1959) suggest that conditional probability values in a S-D matrix are relatively invarient to environmental factors and depend primarily on the physiological-pathological aspects of the disease. Operationally this means that for a specified disease y_j, $P(x_i|y_j)$ i=1,2,...,k values estimated from large random samples of different populations are not expected to vary significantly. Documentation which supports or refutes this view is rare; such conditional probability values have not been calculated for most diseases (although this view is implicitly affirmed in medical texts describing symptom-disease associations). Winkler et al (1967) defend the supposition put forth by Ledley and Lusted (1959) on the basis of Table 2.1.3.

Table 2.1.3

Conditional probabilities for attributes associated with hyperthyroidism. Comparison of Florida data and Florida data combined with Bonn, W. Germany data.

Attribute, x_i	$P(x_i \mid$ hyperthyroidism) Florida data, Fitzgerald and Williams (1964)	Bonn, W. Germany, and Florida data combined. Winkler et al (1967)
Nervousness	0.915 (108/118)	0.925 (195/210)
Heat sensitivity	0.742 (72/97)	0.746 (100/134)
Perspiration	0.678 (59/87)	0.747 (109/146)
Appetite gain	0.605 (69/114)	0.579 (84/145)
Weight loss	0.836 (112/134)	0.846 (187/221)
Hyperkinetic movements	0.755 (83/110)	0.688 (117/170)
Warm, moist skin	0.708 (80/110)	0.760 (130/171)
Light finger tremor	0.871 (108/124)	0.898 (193/125)
Lethargy	0.001 (0/119)	0.102 (16/157)
Cold sensitivity	0.051 (5/98)	0.094 (13/138)
Decreased perspiration	0.001 (0/98)	0.001 (0/157)
Appetite loss	0.133 (15/133)	0.160 (25/156)
Weight gain	0.023 (3/133)	0.032 (7/218)
Slower movements	0.018 (2/109)	0.012 (2/171)
Dry, rough skin	0.009 (1/112)	0.011 (2/185)
Face edema	0.008 (1/127)	0.026 (5/195)
Eye symptoms	0.300 (30/100)	0.446 (87/195)

From Winkler et al (1967)

Comparison of the results obtained when data gathered in Bonn, W. Germany, were combined with data obtained in Florida do not show any great changes in most clinical signs and symptoms. One notable exception is in lethargy, which went from .001 to .102. Winkler et al (1967) suggest that the Florida workers took lethargy to mean apathy, while the German workers understood lethargy to mean general tiredness. Zero values for conditional probabilities are usually interpreted to mean that the population sample was not large enough to include attributes which may possibly occur together with a given disease. Such zero values for conditional probabilities are often replaced by small finite values, .01 or .001. For example, refer to "lethargy" and "decreased perspiration" in Table 2.1.3 (also see Boyle et al (1966)). Table 2.1.4 is a S-D matrix developed by Lodwich (1965) as an aid to diagnose bone tumors. In this case, zero probability values exercise a Boolean function in eliminating certain categories of disease. For example, a tumor falling into Grade III cannot be a giant cell tumor, chondroblastoma, chondromyxoid fibroma, or parosteal sarcoma.

Table 2.1.4

An S-D Matrix for Predicting Bone Tumors

TUMOR TYPE		AGE			LOCATION						SIZE		MATRIX		GRADE				
							Long Bones												
	Incidence	Up to 21	21-30	31+	Small and flat	Long	Epiphysis	Plate	Metaphysis	Shaft	1-6 cm.	6+ cm.	Bone	Cartilage	I-A	I-B	I-C	II	III
Giant cell tumor	15	20	35	45	20	80	99	01	100	20	40	60	01	00	15	35	50	00	00
Chondroblastoma	05	75	20	05	20	80	100	50	75	00	90	10	00	30	50	35	15	00	00
Chondromyxoid fibroma	03	50	35	15	30	70	30	20	100	25	85	15	00	02	85	15	01	00	00
Chondrosarcoma	17	25	25	50	35	65	40	01	85	65	20	80	05	65	15	20	25	25	15
Fibrosarcoma	10	20	20	60	05	95	55	00	90	65	20	80	25	02	00	10	40	30	20
Osteosarcoma	25	65	25	10	10	90	30	05	95	75	15	85	98	05	00	00	10	30	60
Parosteal sarcoma	05	20	35	45	00	100	30	01	100	50	25	75	100	05	15	25	55	05	00
Ewing's tumor	15	70	25	05	35	65	20	05	85	90	15	85	00	00	00	05	10	20	65
Reticulum cell	05	10	25	65	20	80	50	01	85	80	15	85	00	00	00	00	20	30	50

From Lodwich (1965)

One of the disturbing, and at the same time fascinating, features of our human existence is the pervasiveness of uncertainty. - *J. Aitchison*

Section 2.2

Considerations in Data-Collecting

Digital computers are particularly suited to the task of Bayesian inference by reason of their:

1. capacity to store large masses of information,
2. ability to use all pertinent information in its memory,
3. ability to perform calculations rapidly,
4. flexibility in displaying information,
5. availability.

At present Bayes formula is advocated as an aid to the physician in the same sense as stethescopes or laboratory tests; the final diagnosis remains the responsibility of the physician. The performance of Bayes' formula in correctly predicting a disease category for an individual may be undermined by:

1. inaccurate and/or non-representative data regarding a priori probabilities,
2. inaccurate and/or non-representative data regarding conditional probabilities,
3. inaccurate data regarding an individual's profile,
4. the superficial use of attributes,

5. non-independence of attributes occurring in a profile.

These pitfalls may be circumvented by:

1. ensuring that a correct diagnosis is actually established. This is usually possible (i) through follow-up information (ii) by laboratory tests (iii) through agreement among several physicians (iv) at surgery (v) at autopsy. Assuming that a correct diagnosis can be established, prior probabilities may be based on a random sample of the population for whom posterior probabilities are required. The question of how large a random sample, n, is required may be discussed as follows: suppose a hospital or clinic has treated N individuals (say N = 10,000) and each of these individuals has been assigned to disease category y_1 or y_2 or...or y_K. Suppose y_K denotes the least common disease and N_{y_K}/N is thought to be about .10. It is reasonable to assume that we will be content if the final estimate of N_{y_K}/N, based on random sampling of n records, is correct within $\pm$ d (say d = .01) in the sense that if the sample shows n_{y_K}/n = a, then the interval a-d, a+d is 1-α% (say α = .05) sure to contain N_{y_K}/N. The following formula (Cochran (1963)) may be used to solve for n:

$$n = \frac{Z^2_{\alpha/2} \cdot P(y_K) \cdot (1-P(y_K))/d^2}{1 + 1/N \left[(Z^2_{\alpha/2} \cdot P(y_K) \cdot (1-P(y_K))/d^2) - 1 \right]} \quad \text{where}$$

$P(y_K)$ is the original estimate of N_{y_K}/N (.10)

α is the risk (.05) we are willing to incur that the actual error is larger than d. That is, $\alpha = [P(|n_{y_K}/n - N_{y_K}/N|) \geq d]$

d is the allowed margin of error in the estimated value of N_{y_K}/N, after random sampling (.01)

$Z_{\alpha/2}$ is the abscissa of the N(0,1) cumulative distribution curve that cuts off an area of $\alpha/2$ at each tail (1.96).

In this example,

$$n = \frac{(1.96)^2(.1)(.9)/(.01)^2}{1 + \frac{1}{10000}\left((1.96)^2(.1)(.9)/(.01)^2 - 1\right)}$$

$$= 2569$$

In practice n may be constrained by cost and time considerations.

2. Conditional probability values are usually based on the same random sample of the specified population used to calculate prior probabilities. The motivations for doing so are (1) such values are theoretically ideal in the sense that they are known to apply to the specified

population (2) since these individuals have already been examined for attributes in order to arrive at a clinical diagnosis they form a readily accessible data base on which conditional probabilities may be estimated (3) objective conditional probability values do not exist for most diseases, hence the alternative is to subjectively supply their values. To summarize, conditional probability values are usually based on (i) values known to apply to the specified population (see Figure 2.2.1, Option 1), (ii) literature values which are hypothesized to apply to any population (see Figure 2.2.1, Option 2), (iii) subjective estimates (see Figure 2.2.1, Option 3), (iv) combinations of i, ii, iii.

3. Only the most objective, readily recognizable and easily obtained attributes should be included in the attribute list. In practice a trade-off is made between detailed information associated with high costs and less precise but accurate information associated with low costs (cost as measured in time and dollars). Several authors have discussed the feasibility of using computer technology, with a minimum amount of physician involvement, to elicit and analyze the medical history of an individual. See for example, Payne (1963),(1964), Collen et al (1964), Slack et al (1966).

4. Attributes that are useful for discriminating among disease categories should be chosen, and in the case of measured (continuous) characteristics, intervals that best discriminate among disease categories should be defined. For example, 6-hr I^{131} uptake is a measure of high validity in distinguishing among thyroid diseases; low, moderate, and high 6-hr I^{131} uptake values are associated with hypothyroidism, euthyroidism, and hyperthyroidism respectively. Overall and Williams (1963) chose the intervals shown in Table 2.2.1 (Scheme 1) when applying Bayes' formula to predict thyroid diseases.

Table 2.2.1

Cutting Intervals for 6-hr I^{131} Uptake Chosen by Overall and Williams (1963)

Disease list, y_j	Characteristic: 6-hr I^{131} uptake %					
	Interval:	<2	2-7	8-27	27-35	>35
	Attribute:	x_1	x_2	x_3	x_4	x_5
y_1: hypothyroidism	Scheme 1*	.33	.57	.10	.00	.00
y_2: euthyroidism		.01	.10	.85	.03	.01
y_3: hyperthyroidism		.00	.01	.04	.16	.79

*Table entries are $P(x_i|y_j)$, i=1,2,3,4,5 j=1,2,3

Other schemes are possible, as shown in Table 2.2.2.

Table 2.2.2

An Alternative Scheme for Describing 6-hr I^{131} Uptake

Disease list, y_j	Characteristic: 6-hr I^{131} uptake, %			
	Interval:	< 7	8-27	> 27
	Attribute:	x_1	x_2	x_3
y_1: hypothyroidism	Scheme 2	.90	.10	.00
y_2: euthyroidism		.11	.85	.04
y_3: hyperthyroidism		.01	.04	.95

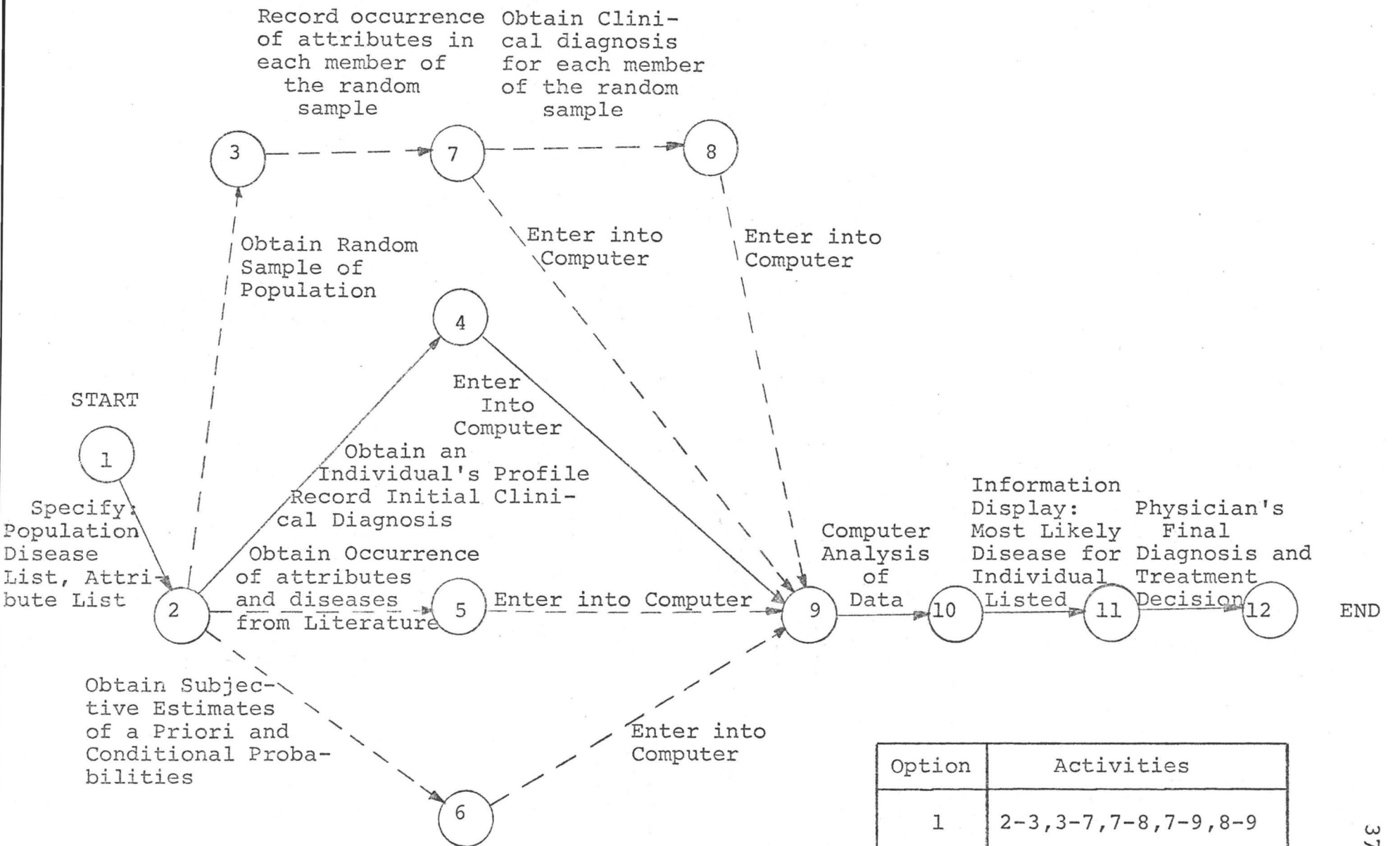
To the extent that such intervals are not optimally chosen, spurious a posteriori probability values may result (a topic beyond the scope of this project).

5. Equation 1.3.6 is valid if attributes appearing in a profile are independent. Of course, posterior probabilities may be calculated directly from equation 1.3.4. However, such a procedure increases data requirements. For example, consider a profile consisting of data on 20 characteristics with each characteristic reported as either present or absent. There are 2^{20} (ie, 1,048,576) possible combinations of attributes (profiles); several million cases of each disease category would be required to obtain reasonable estimates of the frequencies for each of the possible profiles. Data requirements for combinations of attributes may be lessened by (i)

considering groups of dependent attributes (see Section 3.3c) or (ii) excluding dependent attributes from profiles (see Section 3.3a). In short, although the assumption of attribute independence is often inappropriate (see Section 3.3b), the assumption of attribute independence reduces data requirements and complications in calculating posterior probabilities.

Arrow diagrams are useful for describing the essential steps in computer-aided Bayesian diagnostic procedure. Referring to Figure 2.2.1, full lines represent completed activities and dashed lines represent uncompleted activities. At least one option must be completed before posteriori and conditional probabilities can be determined. It is assumed that a computer has been programmed to handle the incoming data.

Figure 2.2.1 Activities in a computer aided Bayesian diagnostic scheme represented in an arrow diagram.



Option	Activities
1	2-3,3-7,7-8,7-9,8-9
2	2-5,5-9
3	2-6,6-9

Egad, I think the interpreter is the hardest to be understood...

- R.S. Sheridan, The Rivals

Section 2.3

The essential steps in a Bayesian diagnostic scheme are discussed in order of earliest completion, with reference to Figure 2.2.1.

Activity 1-2

In studies to date, the specified populations have consisted of individuals referred to a specialized department of a clinic or hospital. There are several reasons why specified populations have been restricted in this way:

1. a patient for whom a predicted diagnosis is required will be a member of this restricted population,
2. a high proportion of these individuals will present data on disease and attribute occurrence,
3. the organization to record and file medical data is operative at a medical center,
4. researchers investigating the applicability of Bayes' formula to medical diagnosis are usually associated with a medical center,
5. a specialized department is chosen as Bayes' formula is used to distinguish among sub-categories of a single disease entity, rather than among divergent disease categories.

Activities 2-3, 3-7, 7-8, 7-9, 8-9 (Option 1)

Most researchers have chosen Option 1 as the means to construct an S-D matrix. A "random sample" of the specified population usually consists of (i) individuals whose medical records have been on file for several years (ii) consecutive referrals to a clinic from some arbitrary date. The sample of the specified population on which a priori and conditional probabilities are estimated is referred to as the data base.

Activity 2-4

The information necessary to predict diagnosis for a new arrival from a specified population consists of (i) data organized into a S-D matrix (ii) the profile of the new arrival. Individuals for whom diagnoses are predicted, but who are not part of the data base, are referred to as a "trial group". The physician's clinical diagnosis for each member of a trial group is usually recorded and may be compared with Bayesian prediction.

Activities 9-10, 10-11, 11-12

Accuracy and sensitivity are two measures commonly used to measure the effectiveness of Bayesian prediction:

$$\text{overall accuracy} = \frac{\text{number of cases correctly identified by Bayes' formula}}{\text{total number of cases}}$$

$$\text{accuracy (in predicting disease } y_j) = \frac{\text{number of cases of } y_j \text{ correctly identified by Bayes' formula}}{\text{total number of cases of } y_j}$$

$$\text{sensitivity (in predicting disease } y_j) = \frac{\text{number of cases of } y_j \text{ correctly identified by Bayes' formula}}{\text{total number of Bayesian predictions for disease } y_j}$$

Accuracy measures the proportion of disease cases correctly identified, sensitivity is a measure of confidence that a particular prediction is correct.

SUMMARY OF CHAPTER II

Chapter II explores the reality represented by probability values appearing in Bayes' formula. It is emphasized that Bayes' formula is applied to restricted (specified) populations, and that information which is valid for one population may be invalid for other populations. Procedures for obtaining valid and useful data have been outlined. Finally, an overview of a computer aided Bayesian diagnostic scheme and the motivations for assuming attribute independence were presented.

CHAPTER III

busy, busy, busy - *Kurt Vonnegut, Cat's Cradle*

INTRODUCTION TO CHAPTER III

Chapter III reviews applications of Bayes' formula in medical literature.

Section 3.1 briefly describes Bayesian prediction of heart, thyroid, abdomen, and liver disorders. Applications of Bayes' formula in other diagnostic areas are listed in Appendix A. Section 3.2 describes an ongoing application of Bayes' formula as a diagnostic aid in a clinical setting. Section 3.3 discusses how some medical researchers have taken non-independence of attributes into account in Bayesian prediction procedures.

Attempt the end, and never stand to doubt;
Nothing's so hard but search will find it out.

- Robert Herrich, Seek and Find

Section 3.1

Literature review. Applications of Bayes' formula to 4 diagnostic areas - heart, thyroid, abdomen, and liver.

Heart

Warner and his colleagues (1961), (1964) were among the first researchers to give serious consideration to probability theory as a diagnostic aid. They use two characteristic lists to distinguish among 33 types of heart disorders. One list consists of 26 easily recognized characteristics, the other list includes 4 additional characteristics of a highly detailed nature - the type and location of heart murmurs. The researchers find that information on heart murmurs do not consistently help to identify disease categories. Their findings illustrate the general dilemma of (i) using as much information as possible and (ii) the limitations in accuracy with which highly detailed information can be obtained. Throughout their work attribute independence is assumed.

Hirschfeld and Judge (1965) compare Bayesian diagnoses of heart disorders to diagnoses arrived at by experienced

cardiologists. Each of 3 experienced cardiologists independently estimate S-D matrix entries, a fourth S-D matrix is constructed on the basis of a literature review. Diagnoses are then predicted for 28 patients by (i) each of the 3 physicians (ii) Bayes' formula using probability values provided by each physician (iii) Bayes' formula using data provided by a literature review. The three procedures achieve similar results, as shown in Table 3.1.1 below. Attribute independence is assumed.

Table 3.1.1

Summary of Findings by Hirschfeld and Judge (1965)

Number of Correctly Identified Cases, Using 20-30 Characteristics

		First ranked Diagnosis Correct		One of 3 Highest Ranked Diagnoses Correct	
Physician:	A	14	50%	21	75%
	B	18	64%	22	79%
	C	19	68%	19	68%
Bayes' formula using S-D matrix provided by:	A	12	43%	20	71%
	B	12	43%	19	68%
	C	11	39%	21	75%
Bayes' formula using S-D matrix provided by:	literature review	14	50%	20	71%

Templeton et al (1966) examine the occurrences of 20 common roentgenograph characteristics associated with 9 heart conditions for 231 cases of confirmed heart disease. The researchers are unsuccessful in isolating independent attributes. Their treatment of non-independence is discussed

in Section 3.3b.

Reale et al (1968) attempt to distinguish among 94 heart disorders using 25 characteristics. An overall accuracy of 82% is achieved in predicting diagnoses for a data base group of 1184 patients, 60% accuracy is achieved for a trial group of 125 patients. When diagnoses are calculated using a quasi-Bayesian approach in which all prior probabilities are arbitrarily made equal, it is found that the accuracy in predicting less common diseases increases. Columns 5 and 6 of Table 3.1.2 illustrate how prior probabilities may overwhelm incoming data (represented by patient profiles) in Bayesian prediction.

Table 3.1.2

Comparison of clinical and calculated diagnoses for a data base group of 1184 patients. Diseases are divided according to those represented by 12 or more cases, and less than 12 cases.

Category (1)	Number of Diseases (2)	Number of Patients (3)	Correct Diagnosis, %, by: Physician (4)	Bayes' (5)	quasi-Bayes' (6)
Total group	94	1184	73	82	70
Diseases with 12 or more cases	14	967	81	85	68
Diseases with less than 12 cases	80	217	37	69	82

When prior probabilities are made equal (column 6) the

predicted diagnosis is no longer weighed in favour of the more common diseases (as in column 5), and the less common diseases are correctly predicted more often, ie, prior probabilities are of no account, and predictions are based on profile data only. Attribute independence is assumed.

Thyroid

Boyle et al (1966) compare Bayesian, quasi-Bayesian, and clinical accuracy in distinguishing among 3 types of thyroid conditions on the basis of 30 attributes. Criterion for a correct diagnosis are more stringent than those of other workers. For example, a first ranked calculated diagnosis may arbitrarily be deemed "incorrect" if its posterior Bayes' probability does not exceed the posterior probability of the second ranked diagnosis by at least a factor of 3. The three diagnostic methods achieve similar results - 83% (Bayes'), 85% (quasi-Bayes'), and 77% (physician) accuracy for a trial group of 88 patients. Other workers who apply Bayes' formula to diagnose thyroid disorders include Billewicz et al (1969), Fitzgerald et al (1966), Gustafson et al (1971), Overall and Williams (1961) (1963), and Winkler et al (1967). In all cases these workers assume attribute independence.

Abdomen

Rinaldo et al (1963) apply Bayes' formula to distinguish among 6 types of abdomen disorders. The

characteristics (8 in all) are highly subjective and depend upon the patient as a witness. The authors ascribe poor Bayesian accuracy (52% for a trial group of 96 patients) to unreliable information obtained from patients. Scheinok and Rinaldo (1967) re-examine the data gathered by Rinaldo et al (1963) in an (unsuccessful) effort to identify subsets of characteristics of high diagnostic accuracy. Their study illustrates the subset concept as a generalized "attribute filter" to use medical data efficiently. Other workers who attempt to predict diagnoses with a minimum number of attributes include Knill-Jones et al (1973), Teather and Hilder (1975), and Templeton et al (1966). Here again, these workers apply Bayes' formula assuming attribute independence.

Horrocks et al (1972) and her colleagues (see De Dombal et al (1972), (1975)) have reported on the use of Bayes' formula as an ongoing diagnostic aid in a clinical setting. Their work is summarized in Section 3.2.

Liver

Lincoln and Parker (1967), Began and Dhumeaux (1971) and Knill-Jones (1975) apply Bayesian estimation to predict liver disorders. Lincoln and Parker (1967) take account of paired dependence; their work is discussed in detail in Section 3.3c. The study by Began and Dhumeaux (1971) is unique in that only data from laboratory tests appear in a profile; the authors state that the attributes appearing in a profile are independent, but give no quantitative evidence to substantiate

their claim. Knill-Jones (1975) (also see Knill-Jones et al (1973)) considers 102 characteristics and out of these identifies subsets of high diagnostic accuracy.

Others

Applications of Bayes' formula to other medically related areas are listed in Appendix A.

We must look at what we are seeking to nourish, and by the exercise of our thoughts seek for the proper aliment.

- I Ching

Section 3.2

A summary of Computer-aided diagnosis: description of an adaptable system, and operational experience with 2,034 cases, Horrocks et al (1972).

Horrocks et al (1972) and her colleagues (see De Dombal et al (1972), (1975)) initiated a system which provided computer technology and Bayes' formula as practical diagnostic aids to the clinician on the ward. Several measures were taken to accommodate the use of Bayes' formula in a clinical setting[5]:

1. Specialized forms were created in which information from the patient's physical examination could be formalized. An example of such a form is shown in Table 3.2.2.
2. Each possible attribute was allocated a three digit code. Thus, the physician need play no other role in the system besides examining the patient and reading the computer output.
3. The computer output was as close as possible to colloquial English. The patient's attribute list (profile) was also printed to allow the physician to check possible errors in coding. An example of computer output is shown in Table 3.2.3.

4. The physician received computer output within a few minutes after completing the patient's physical examination.

A summary of the findings by Horrocks et al (1972) is presented in Table 3.2.1.

Table 3.2.1

A Summary of the Findings by Horrocks et al (1972)

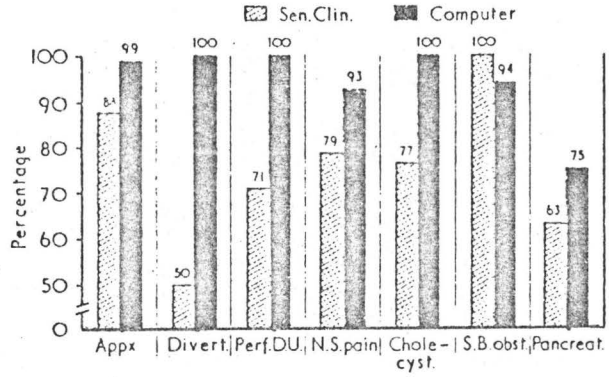
Diagnostic Area	Number of Cases Forming Data Base	Number of New Cases Forming Trial Group	Accuracy, %, of Bayesian Prediction for Trial Group
acute abdomen	600	376	91
lower G.I. Tract	642	82	88
dyspepsia	175	50	64
all areas	1417	508	88
estimates* (acute abdomen)		376	85

*Conditional and a priori probabilities are based on subjective estimates by 6 clinicians.

Figures 3.2.1 and 3.2.2 present detailed information concerning the performance of Bayes' formula in distinguishing among 7 diseases associated with acute abdomen patients.

Figure 3.2.1

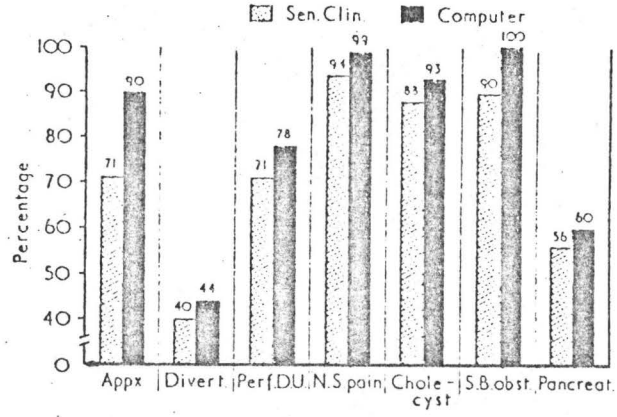
Accuracy of Diagnosis in 304 Patients: Comparison of Bayesian Prediction Versus Diagnosis of Most Senior Clinician to See the Case



From De Dombal et al (1972)

Figure 3.2.2

Sensitivity of Diagnosis in 304 Patients: Same Comparison as in Figure 3.2.1



From De Dombal et al (1972)

Overall, the error rate of Bayesian prediction (25 cases out of 304) compares favourably with that of the physician (62 cases out of 304), and this difference is

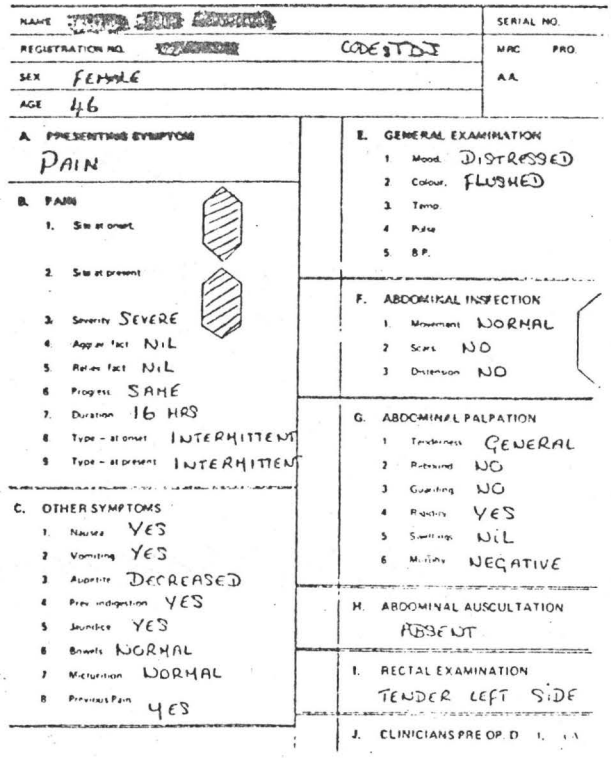
statistically significant ($\alpha < .01$). After 3 years of operational experience, the Leeds researchers have come to the following conclusions:

1. Computer-aided diagnosis of acute abdominal pain is feasible, using simple, inexpensive desk-top equipment.
2. Computer-aided systems in this sphere of clinical medicine appear to have an error rate about half that of the unaided clinician.
3. Addition of a computer-aided system to routine diagnostic practice appears significantly to improve the clinicians' own diagnostic performance.
4. The same considerations appear to apply also to decision-making.
5. The benefits of introduction of such a system are only realised after a training period of 4 to 8 weeks.
6. Equally, the benefits are seen only if the clinician receives regular feedback about his performance from the system.
7. There is no evidence that comparable benefit can be achieved using simple systems not aided by a computer.
8. It remains to be seen whether the system can be transposed from Leeds to other areas, though preliminary studies are modestly encouraging.

The Leeds workers assume attribute independence.

Table 3.2.2

Part of a Specialized Work-Sheet Used by Physicians



NAME [redacted] SERIAL NO.

REGISTRATION NO. [redacted] CODE STDJ MRC PRO.

SEX FEMALE A.A.

AGE 46

A. PRESENTING SYMPTOM
PAIN

B. PAIN
1. Site at onset
2. Site at present
3. Severity SEVERE
4. Aggravating factor NIL
5. Relieving factor NIL
6. Progress SAME
7. Duration 16 HRS
8. Type – at onset INTERMITTENT
9. Type – at present INTERMITTENT

C. OTHER SYMPTOMS
1. Nausea YES
2. Vomiting YES
3. Appetite DECREASED
4. Previous indigestion YES
5. Jaundice YES
6. Bowels NORMAL
7. Micturition NORMAL
8. Previous Pain YES

E. GENERAL EXAMINATION
1. Mood DISTRESSED
2. Colour FLUSHED
3. Temp
4. Pulse
5. B.P.

F. ABDOMINAL INSPECTION
1. Movement NORMAL
2. Scars NO
3. Distension NO

G. ABDOMINAL PALPATION
1. Tenderness GENERAL
2. Rebound NO
3. Guarding NO
4. Rigidity YES
5. Swelling NIL
6. Masses NEGATIVE

H. ABDOMINAL AUSCULTATION
ABSENT

I. RECTAL EXAMINATION
TENDER LEFT SIDE

J. CLINICIANS PRE OP. D

From Horrocks et al (1972)

Table 3.2.3

An example of the Computer Printout

Case ref-234

Female	
Age 70+	No previous abd. pain
Site onset-Central	Prev. abdominal surgery
Site present-Central	Not taking drugs
Aggrav. by nil	
Relieved by vomiting	Mood-distressed
Pain getting worse	Colour-pale
Duration over 48 hrs	Abd. mvt. normal
Onset pain colicky	Abdominal scar present
Pain now colicky	Distension present
Moderate pain	Tenderness-General abdomen
Nausea	No rebound tenderness
Vomiting	Guarding absent
Appetite decreased	Rigidity absent
No indigestion	Swelling absent
No jaundice	Murphy's sign negative
Bowels-Constipation	Bowel sounds hyperactive
Micturition-Dysuria	Normal rectal exam

Append	Divert	Perfdu	Nonsap	Chole	Smbobt	Pancre
.00	.00	.00	.00	.00	99.99	.00

Computer error analysis follows:-

Current position: Computer diagnosis-Small bowel obstruction

During 1971-3 computer made 21 such diagnoses
19 right, 2 wrong

Cases misdiagnosed as Sm. bowel obst.

Pancreatitis	1
Ischaemic colitis	1

From de Dombal et al (1975)

Things either are what they appear to be; or they neither are, nor appear to be; or they are, and do not appear to be; or they are not, and yet appear to be. Rightly to aim in all these cases is the wise man's task. - *Epictetus*

Section 3.3

Treatment of non-independence of attributes in applications of Bayesian estimation to medical diagnosis.

Section 3.3a

A summary of Probability theory in the diagnosis of Cushing's Syndrome. Nugent et al (1964).

Nugent et al (1964) apply Bayes' formula to predict the presence (y_1) or absence (y_2) of Cushing's Syndrome. Their data base consists of 211 patients examined for suspected disease presence. Conditional probability values on 19 characteristics considered in the study are shown in Table 3.3.1. Prior probabilities are: $P(y_1) = .25$, $P(y_2) = .75$.

Differences in the incidences of attributes between patients in the two diagnostic categories are examined by use of the Chi square test statistic (with Yates correction factor). The incidences of attributes differ significantly ($\alpha = .05$) from one another only in the case of characteristics 1-13, hence characteristics 14-19 are omitted from further consideration. The Chi square statistic is also used to

identify dependent pairs in each disease category. Pairs 4 & 12, 4 & 13, 10 & 12, 10 & 13, 12 & 14 are found to be significantly (α = .05) associated. Since characteristics 12 or 13 are included in each of these combinations, they too are omitted from further consideration. Thus, the maximum number of attributes which may possibly appear in a profile is 11.

In order to further examine the question of independence of the remaining characteristics (1-11), the contingency coefficient, which has a range of 0 - .707, is calculated for each pair of characteristics. Of 108 possible attribute pairs[6] in the two groups of patients, three have an absolute value for C greater than .3 and none is greater than .4. On this basis, independence of characteristics 1-11 is assumed.

Diagnoses are calculated for a trial group of 111 patients suspected of disease presence. The authors define a confident diagnosis as $P(y_1) \geq .99$ or $P(y_2) \geq .99$. They calculate 54 confident diagnoses, all correctly identify the patients' disease category.

Table 3.3.1

Characteristics considered and conditional probabilities

x_i	$P(x_i \mid y_1)$	$P(x_i \mid y_2)$
1. Osteoporosis	0.64	0.03
2. Central obesity	0.90	0.29
3. Generalized obesity	0.03	0.62
4. Weakness	0.65	0.07
5. Plethora	0.82	0.31
6. WBC 11,000 per mm³ or more	0.58	0.30
7. Acne	0.52	0.24
8. Striae (red or purple)	0.46	0.22
9. Diastolic BP 105 or above	0.39	0.17
10. Edema (pitting)	0.38	0.17
11. Hirsutism	0.50	0.29
12. Ecchymoses	0.53	0.06
13. Serum K 3.6 mEq/l or less	0.25	0.04
14. Oligomenorrhea	0.72	0.51
15. Headaches	0.41	0.37
16. VPRC 49 or above	0.37	0.32
17. Females	0.65	0.77
18. Abnormal GTT	0.88	0.77
19. Age 35 or less	0.55	0.52

From Nugent et al (1964)

Section 3.3b

A Summary of The Computer Evaluation and Diagnosis of Congenital Heart Diseases, Using Roentgenographic Findings. By Templeton et al, (1966).

Templeton et al (1966) examine the occurrences of twenty common roentgenograph characteristics (see Table 3.3.3) associated with 9 heart conditions. Conditional and a priori probabilities are based on 231 cases of confirmed heart disease (see Table 3.3.4). Bayes' formula is used to predict diagnosis for each member of the data base group. Some findings of the study are summarized in Table 3.3.2.

Table 3.3.2

Number of Correctly Identified Cases Using All Information Available From the Work-Sheet

Disease List	Number of Cases	Number Correctly Identified by Radiologist	Number Correctly Identified by Bayes' Formula
Primum IASD	15	10	12
Secondum IASD	44	41	38
IVSD	49	35	31
PDA	30	17	18
Tetrlogy of Fallot	38	35	35
Coarctation	11	10	11
Pul. Valve Stenosis	14	11	12
Aortic Valve Stenosis	14	11	11
Complete Trans.-Position	16	15	15
TOTAL	231	185 (80%)	183 (79%)

The authors discuss attribute independence as follows: suppose a large number of cases are predicted, and in each case the posterior probability assigned to disease y_i is about .7. We would expect about 70% of the cases to be actual cases of disease y_i. Similarly, we expect the number of cases correctly predicted to be about equal to the average of posterior probability values assigned to the most likely (first ranked) disease for each case. As the number of cases, N, becomes large the probability that the quantity t,

$$t = |\bar{P} - f| \cdot N / \left(\sum_{i=1}^{i=N} P_i(1-P_i) \right)^{1/2} \qquad \text{where}$$

f = fraction of correct diagnoses

N = number of cases

P_i = the highest posterior probability assigned to any disease for a particular case i

$$\bar{P} = \left(\sum_{i=1}^{i=N} P_i \right)/N$$

will exceed some value v is given by $P(t > v)$,

$$P(t > v) = 1 - (1/\sqrt{2\pi}) \int_{-v}^{v} e^{-\frac{u^2}{2}} \, du$$

In practice the value of t is calculated and the area under the N(0,1) cumulative distribution curve between -t and t is found from tables. This area represents the confidence level at which H_o (H_o: the attributes appearing in a profile

are independent) may be rejected. The authors test various subsets of the 21 characteristics. The best indication of attribute independence is associated with characteristics 1, 2, 3, 4, 12, 14, 16. They are given an 11% chance of being independent.

Table 3.3.3

CONGENITAL HEART WORK-SHEET

AGE_____years,_____months	SEX 1 Male 2 Female	CYANOSIS 1 Present 2 Absent

1. HEART SIZE
   1 Normal
   2 Large I
   3 Large II
   4 Large III

2. LEFT VENTRICULAR SIZE
   1 Normal
   2 Large I
   3 Large II
   4 Large III

3. RIGHT VENTRICULAR SIZE
   1 Normal
   2 Large I
   3 Large II
   4 Large III

4. LEFT ATRIUM
   1 Normal
   2 Enlarged

5. RIGHT ATRIUM
   1 Normal
   2 Enlarged

6. BOOT SHAPED HEART
   1 Present
   2 Absent

7. ASCENDING AORTA SIZE
   1 Small
   2 Normal
   3 Large
   0 Not Seen

8. AORTIC ARCH SIZE
   1 Small
   2 Normal
   3 Large
   0 Not Seen

9. CONVEX OR NOTCHED DESCENDING AORTA
   1 Present
   2 Absent
   0 Not Seen

10. INFUNDIBULUM SIGN
    1 Present
    2 Absent

11. NARROW BASE OF HEART
    1 Present
    2 Absent
    0 Not Seen

12. MAIN PULMONARY ARTERY
    1 Small
    2 Normal
    3 Large I
    4 Large II
    5 Large III

13. RIGHT PULMONARY ARTERY
    1 Small
    2 Normal
    3 Large

14. PERIPHERAL VESSELS
    1 Decreased
    2 Normal
    3 Increased

15. PULMONARY VEINS
    1 Small
    2 Normal
    3 Enlarged
    0 Not Seen

16. PERIPHERAL VESSEL DISPARITY
    1 Present
    2 Absent

17. RETICULAR PATTERN
    1 Present
    2 Absent

18. KYPHOSCOLIOSIS
    1 Present
    2 Absent

19. RIB NOTCHING
    1 Present
    2 Absent

20. HYPEREXPANSION
    1 Present
    2 Absent

RADIOLOGIST'S DIAGNOSIS
    1 Primum IASD
    2 Secundum IASD
    3 Interventricular Septal Defect
    4 Patent Ductus Arteriosus
    5 Tetralogy of Fallot
    6 Coarctation
    7 Pulmonary Valvular Stenosis
    8 Aortic Valvular Stenosis
    9 Complete Transposition

DEFINITIVE DIAGNOSIS _____
    (Same code as above)

From Templeton et al (1966)

Table 3.3.4

S-D Matrix Used by Templeton et al (1966)

PARAMETER BEING EVALUATED	PARAMETER GRADE	NINE "PURE" CONGENITAL HEART DEFECTS								
		PRIMUM IASD	SECUNDUM IASD	IVSD	PDA	TETRALOGY OF FALLOT	COARC-TATION	PUL. VALVE STENOSIS	AORTIC VALVE STENOSIS	COMPLETE TRANS-POSITION
AGE	Less than 3mo	07	02	14	17	16	00	07	08	94
	3 mo - 5 yr	33	11	31	37	24	09	36	07	06
	5 - 20 yr	47	37	37	33	34	46	43	50	00
	20 - 45 yr	13	25	14	13	26	45	14	35	00
	Over 45 yr	00	25	04	00	00	00	00	00	00
SEX	Male	40	30	35	30	50	55	43	79	56
	Female	60	70	65	70	50	45	57	21	44
CYANOSIS	Present	07	11	06	10	82	00	07	07	100
	Absent	93	89	94	90	18	100	93	93	00
HEART SIZE	Normal	00	09	18	28	37	73	43	50	06
	1+	13	34	39	34	29	18	50	43	38
	2+	27	37	27	24	29	09	00	00	25
	3+	60	20	16	14	05	00	07	07	31
LEFT VENTRICLE	Normal	07	73	20	30	92	09	93	29	37
	1+	47	18	70	54	08	73	07	57	63
	2+	46	09	10	13	00	18	00	14	00
	3+	00	00	00	03	00	00	00	00	00
RIGHT VENTRICLE	Normal	00	03	06	23	10	91	00	93	06
	1+	13	31	45	37	61	09	79	07	44
	2+	60	57	45	33	26	00	21	00	44
	3+	27	09	04	07	03	00	00	00	06
LEFT ATRIUM	Normal	33	95	40	46	92	91	100	100	87
	Large	67	05	60	54	08	09	00	00	13
RIGHT ATRIUM	Normal	27	95	96	100	100	100	100	100	87
	Large	73	05	04	00	00	00	00	00	13
BOOT SHAPE	Present	00	00	00	00	40	00	21	00	37
	Absent	100	100	100	100	60	100	79	100	63
ASCENDING AORTA	Small	91	67	64	10	19	20	50	00	43
	Normal	09	33	36	85	69	60	50	25	57
	Large	00	00	00	05	12	20	00	75	00
ARCH SIZE	Small	85	59	40	12	06	55	54	00	64
	Normal	15	37	58	38	44	45	46	67	36
	Large	00	04	02	50	50	00	00	33	00
NOTCHED AORTA	Present	00	00	00	00	00	100	00	00	00
	Absent	100	100	100	100	100	00	100	100	100
INFUNDIBU-LUM SIGN	Present	00	03	06	44	00	00	00	00	00
	Absent	100	97	94	56	100	100	100	100	100
NARROW HEART BASE	Present	00	00	00	00	05	00	00	00	81
	Absent	100	100	100	100	95	100	100	100	19
MAIN PULMONARY ARTERY	Small	00	00	00	00	55	09	00	00	38
	Normal	07	11	35	30	42	91	14	100	38
	1+	13	36	37	33	03	00	21	00	24
	2+	73	28	22	34	00	00	36	00	00
	3+	07	25	06	03	00	00	29	00	00
RIGHT PULMONARY ARTERY	Small	00	00	00	00	60	00	36	00	06
	Normal	00	16	28	43	37	100	64	100	38
	Large	100	84	72	57	03	00	00	00	56
PERIPHERAL VESSELS	Decreased	00	00	00	00	77	00	43	00	06
	Normal	00	03	06	23	18	100	50	93	25
	Increased	100	97	94	77	05	00	07	07	69
PULMONARY VEINS	Small	00	00	00	00	66	00	45	00	12
	Normal	00	00	31	39	31	100	46	100	38
	Large	100	100	69	61	03	00	09	00	50
VESSEL DISPARITY	Present	13	11	68	73	05	00	00	07	13
	Absent	87	89	32	27	95	100	100	93	87
RETICULAR PATTERN	Present	00	00	00	00	26	00	07	00	00
	Absent	100	100	100	100	74	100	93	100	100
KYPHO-SCOLIOSIS	Present	00	05	05	00	26	00	14	07	00
	Absent	100	95	95	100	74	100	86	93	100
RIB NOTCHING	Present	00	00	00	00	00	55	00	00	00
	Absent	100	100	100	100	100	45	100	100	100
HYPER-EXPANSION	Present	40	05	45	60	05	00	00	07	50
	Absent	60	95	55	40	95	100	100	93	50

Probabilities in per cent, eg, $P(\text{heart size } 3^{+} | \text{IVSD}) = .16$

Section 3.3c

A Summary of Medical Diagnosis Using Bayes' Theorom. Lincoln and Parker, (1967).

Lincoln and Parker (1967) use 16 characteristics to distinguish among 10 diseases associated with the liver (see Tables 3.3.5, 3.3.6). Paired dependent relationships are taken into account in the application of Bayesian prediction.

Attributes describing a particular characteristic[7] S_i are defined as follows:

$x_i = 0$ (normal) if $L_i \leqslant S_i < U_i$

$x_i = 1$ (abnormal) if $L_i > S_i$ or $S_i \geqslant U_i$

$x_i = -1$ (unknown) if no information available on S_i

For example, a patient having all characteristics outside their normal ranges is described by X = { 1, 1, ..., 1 }. Attribute dependence is taken into account as follows:

$$P(y_j|X) = \frac{P(x_1|y_j)\cdot P(x_2|x_1,y_j)\cdot \ldots \cdot P(x_{16}|x_{15},y_j)\cdot P(y_j)}{\sum_{\ell=1}^{\ell=10} P(x_1|y_\ell)\cdot P(x_2|x_1,y_\ell)\cdot \ldots \cdot P(x_{16}|x_{15},y_j)\cdot P(y_\ell)} \quad (3.3.1)$$

Probability values appearing on the R.H.S. of (3.3.1) are based on data gathered over a ten year period at The Department of

Medicine, Yale University. It was found that some probability values could not be estimated directly from frequency counts. For example, if no patients in a particular disease category y_j have an abnormal reading for S_i accompanying an abnormal reading for S_{i-1}, then $P(x_i = 1 | x_{i-1}=1, y_j)$ cannot be calculated. Probability values that could not be calculated because of lack of data were estimated subjectively.

Typical conditional probability values are shown in Tables 3.3.7 and 3.3.8. The authors report:

1. ...it became clear that a priori weights negated a portion of the usefulness of the program as an automated consultation, namely, to bring into consideration less common disease entities or less usual symptom-disease relationships. It was often more medically useful to consider an artificial population where all possibilities (prior probabilities) are made equally likely.

2. The choice of order of the symptoms was not as critical as initially presumed.

3. In 22 of 40 cases the program ranked the proper diagnosis correctly. The success of the program is considered sufficiently encouraging to warrant a more intensive study.

No other quantitative information is given to support their claims.

Table 3.3.5

Disease List and a Priori Probabilities

Disease Categories	Prior Probabilities
Viral hepatitis	.26
Nutritional cirrhosis	.10
Intrahepatic obstruction	.04
Extrahepatic obstruction	.06
Postnecrotic cirrhosis	.18
Cardiac cirrhosis	.01
Primary biliary cirrhosis	.06
Fatty liver	.07
Granuloma	.18
Reticuloendothelial respiratory reaction	.04

Table 3.3.6

Characteristics Considered and Intervals for Defining Attributes

			Units
S_1 = serum bilirubin (1 min.)	$I_{L_1} = 0.0$	$I_{U_1} = 0.3$	mg 100 ml
S_2 = serum bilirubin (total)	$I_{L_2} = 0.0$	$I_{U_2} = 1.3$	mg 100 ml
S_3 = serum B.S.P.	$I_{L_3} = 0.0$	$I_{U_3} = 7.0$	% retention
S_4 = serum ceph. flóc.	$I_{L_4} = 0.0$	$I_{U_4} = 2.0$	Hunger's
S_5 = serum thymol turbidity	$I_{L_5} = 0.0$	$I_{U_5} = 4.0$	Maclagan's
S_6 = serum alk. phos.	$I_{L_6} = 0.0$	$I_{U_6} = 8.6$	Shinowara's
S_7 = serum transaminase	$I_{L_7} = 0.0$	$I_{U_7} = 40.0$	Karmen's
S_8 = serum total protein	$I_{L_8} = 6.5$	$I_{U_8} = 8.4$	gm 100 ml
S_9 = serum albumin	$I_{L_9} = 3.6$	$I_{U_9} = 5.1$	gm 100 ml
S_{10} = serum globulin	$I_{L_{10}} = 2.2$	$I_{U_{10}} = 4.0$	gm 100 ml
S_{11} = serum total cholesterol	$I_{L_{11}} = 150.2$	$I_{U_{11}} = 250.0$	mg 100 ml
S_{12} = serum, % free cholesterol	$I_{L_{12}} = 28.0$	$I_{U_{12}} = 32.0$	mg 100 ml
S_{13} = liver size	$I_{L_{13}} = 2.0$	$I_{U_{13}} = 3.0$	
S_{14} = liver firmness	$I_{L_{14}} = 2.0$	$I_{U_{14}} = 3.0$	
S_{15} = spleen palpability	$I_{L_{15}} = 1.0$	$I_{U_{15}} = 2.0$	
S_{16} = stool color	$I_{L_{16}} = 2.0$	$I_{U_{16}} = 3.0$	

Table 3.3.7

Conditional Probability Values for y_1, Viral Hepatitis

Characteristic	$P(x_i=1\|y_1)$	$P(x_i=1\| x_{i-1}=1,y_1)$	$P(x_i=1\| x_{i-1}=0,y_1)$
1 % free cholesterol	1.		
2 Liver palpability	.80	.83	.09
3 Liver firmness	.64	.67	.50
4 B.S.P.	.89	.87	.92
5 Transaminase (SGOT)	.80	.83	0.
6 Serum bilirubin (Total)	.83	.87	.54
7 Serum bilirubin (1 min.)	.93	1.	.59
8 Spleen palpability	.33	.36	.43
9 Albumin	.79	.86	.86
10 Alk. phos.	.51	.45	.83
11 Stool color	.27	.32	.22
12 Total protein	.43	.21	.50
13 Ceph. floc.	.70	.63	.71
14 Globulin	.34	.36	.25
15 Thymol turbidity	.69	.89	.67
16 Cholesterol (total)	.44	.33	.60

Table 3.3.8

Conditional Probability Values for y_5, Postnecrotic Cirrhosis

Characteristic	$P(x_i=1\|y_5)$	$P(x_i=1\| x_{i-1}=1,y_5)$	$P(x_i=1\| x_{i-1}=0,y_5)$
1 % free cholesterol	.75		
2 Liver palpability	.69	.86	.33
3 Liver firmness	.69	.73	.60
4 B.S.P.	.88	.44	.70
5 Transaminase (SGOT)	.63	.66	0.
6 Serum bilirubin (Total)	.65	.79	.43
7 Serum bilirubin (1 min.)	.82	.93	.67
8 Spleen palpability	.56	.57	.17
9 Albumin	.86	.81	.84
10 Alk. phos.	.63	.61	.44
11 Stool color	.38	.38	.33
12 Total protein	.54	.71	.28
13 Ceph. floc.	.62	1.	.72
14 Globulin	.63	.66	.56
15 Thymol turbidity	.58	.63	.63
16 Cholesterol (total)	.37	.33	.50

SUMMARY OF CHAPTER III

Many researchers have investigated the premise that Bayesian procedures may be useful diagnostic aids; Chapter III has briefly reviewed a substantial portion of such research to date. Most investigators have conveniently assumed attribute independence; research papers which attempt to deal with non-independence of attributes have been presented in detail. The work of Horrocks et al (1972) is included as an example of an ongoing application of Bayes' formula in a clinical setting.

CHAPTER IV

All nature is but art, unknown to thee;
All chance, direction, which thou canst see.

- Alexander Pope

INTRODUCTION TO CHAPTER IV

Literature concerning the application of Bayes' formula to medical diagnosis has had little success in explaining why accurate predictions are often possible in spite of non-independence of attributes. Chapter IV examines the accuracy of predictions under various deviations from attribute independence. The study looks at paired dependence among 6 attributes when Bayes' formula is used to predict diagnoses from among 6 disease categories.

Section 4.1 describes a Monte Carlo method for simulating the usual Bayesian diagnostic process. Section 4.2 establishes 500 as a reasonable sample size (per disease category) on which to predict diagnoses. Section 4.3 illustrates decreasing Bayesian accuracy as deviations from independence increase. Section 4.4 presents a summary of the study.

Statistics? I can prove anything by statistics except the truth. - *G. Canning*

Section 4.1

Computer simulation of Bayes' formula as applied to medical diagnosis.

Introduction

A computer simulation for the Bayes' process of diagnosis was written for the CDC 6400. In the first portion of the program, $P(x_i|y_j)$ values and dependent relationships (ie, for the "true" state of nature) are specified. Profiles are then generated according to this actual probability distribution of attributes among diseases. The number of individual profiles (sample size) generated in each disease category is an input parameter denoted by LN. $P(x_i|y_j)$ values are then estimated by the frequency with which attributes occur in the generated random sample of the disease-attribute universe. A Bayesian diagnosis is calculated for each possible attribute profile according to (i) estimated $P(x_i|y_j)$ values, assuming attribute independence and (ii) actual $P(x_i|y_j)$ and joint probability values, ie, taking into account non-independence of attribute pairs. The resulting number of matches is denoted by M. For each value of LN, 5 "random samples" of the population are generated at a specified level of attribute dependence; thus, the computer program simulates

the process of predicting diagnoses on 5 distinct random samples of a disease-attribute population (see Tables 4.2.1, 4.2.2).

1. Describing the disease-attribute universe.

(i) prior probabilities, $P(y_j)$.

The frequencies with which diseases occur in the population are made equal in order to facilitate the interpretation of effects of deviations from independence on Bayesian accuracy. That is,

$P(y_j) = 1/K, \quad j = 1,2,\ldots,K$ where

K is the number of disease categories (6).

(ii) probabilities conditional on disease categories, $P(x_i|y_j)$.

Each characteristic i is defined as present (x_i) or absent ($\bar{x}_i$). Thus, $P(x_i|y_j)$ is the probability of attribute x_i being present in an individual having disease y_j and $P(\bar{x}_i|y_j)$ is the probability of characteristic i being absent in an individual having disease y_j. Of course, $P(\bar{x}_i|y_j) = 1-P(x_i|y_j)$. $P(x_i|y_j)$ is defined for each i,j combination. Values of $P(x_i|y_j)$ read into the program are chosen so that (i) they are between .01 and .5 in the majority of cases, conforming to literature values of $P(x_i|y_j)$, (ii) posterior probabilities do not exceed .5 in the majority of cases. This ensures that Bayes' formula is being applied to non-trivial problems of diagnosis, ie, it is unlikely that

correct diagnoses can be made by considering subsets of key attributes, or without recourse to mathematical computations.

(iii) dependent pairs, $P(x_B|x_A,y_j)$.

In each disease category the appearance of one characteristic, denoted by x_B, is made dependent on the appearance of another characteristic, denoted by x_A, ie, $P(x_B|x_A,y_j)$ is specified for each y_j, $j = 1,2,...6$. x_A and x_B may, and usually do, identify different pairs in different disease categories. This is done to conform with the nature of symptoms and diseases - different groups of symptoms occur jointly for different diseases. For a specified y_j, $P(\bar{x}_B|x_A,y_j)$ $= 1-P(x_B|x_A,y_j)$. $P(x_B|\bar{x}_A,y_j)$ may be determined from previously defined values as follows:

from

$$P(x_B|y_j)=P(x_B|x_A,y_j)\cdot P(x_A|y_j)+P(x_B|\bar{x}_A,y_j)\cdot P(\bar{x}_A|y_j) \qquad (4.1.1)$$

it follows

$$P(x_B|\bar{x}_A,y_j)=\left[\, P(x_B|y_j)-P(x_B|x_A,y_j)\cdot P(x_A|y_j)\right] /P(\bar{x}_A|y_j) \qquad (4.1.2)$$

Of course,

$$P(\bar{x}_B|\bar{x}_A,y_j) = 1-P(x_B|\bar{x}_A,y_j)$$

Two constraints are imposed upon assigned probability values, as suggested by (4.1.2):

(a) $P(x_B|y_j) \geq P(x_B|x_A,y_j) \cdot P(x_A|y_j)$

(b) $P(\bar{x}_A|y_j) \geq P(x_B|y_j) - P(x_B|x_A,y_j) \cdot P(x_A|y_j)$

The first constraint ensures $P(x_B|\bar{x}_A,y_j)$ is non-negative, the second constraint ensures $P(x_B|\bar{x}_A,y_j)$ is not greater than unity.

To summarize, the following values are read into the computer program and form a complete probabilistic description of a disease-attribute universe:

$P(y_j)$ $\quad j = 1,2,\ldots,6$

$P(x_i|y_j)$ $\quad i = 1,2,\ldots,6 \quad j = 1,2,\ldots,6$

$P(x_B|x_A,y_j)$ $\quad j = 1,2,\ldots,6$

A subroutine identifies the x_A, x_B pair in each disease category. The program returns an error message if constraints (a) and (b) are not satisfied.

2. Simulating simple random sampling of the disease-attribute universe.

In the second portion of the program attribute profiles for individuals are generated according to the actual distribution of attributes among diseases. For example, suppose we wish to generate a profile of an individual having disease y_j. A random number generator[8] returns a fraction $Z \sim U(0,1)$ for each x_i except x_B. x_i is assigned to the individual if $Z_i \leq P(x_i|y_j)$, otherwise $\bar{x}_i$ is assigned to the

individual. X_B is similarly assigned according to the presence or absence of characteristic A and the value of the random variable Z:

A present (ie, x_A) : x_B if $Z_B \leq P(x_B|x_A,y_j)$

$\bar{x}_B$ otherwise

A absent (ie, $\bar{x}_A$) : x_B if $Z_B \leq P(x_B|\bar{x}_A,y_j)$

$\bar{x}_B$ otherwise

The number of individual profiles to be generated in each disease category is an input parameter denoted by LN.

3. Calculating M.

The individual profiles generated in the second portion of the program are used to estimate $P(x_i|y_j)$ values. For example, suppose a_1 of the individuals in disease category y_j have x_1 present, a_2 have x_2 present, etc. Then $\hat{P}(x_1|y_j) = a_1/LN$, $\hat{P}(x_2|y_j) = a_2/LN$, etc. Prior probabilities are taken to be 1/K, according to the actual distribution. After a population of size LN has been generated for each disease category and $\hat{P}(x_i|y_j)$ values determined, a diagnosis is calculated for each possible attribute profile[9] using $\hat{P}(x_i|y_j)$ values and assuming attribute independence as follows:

$$\hat{P}(y_j|X) = \frac{\left[\prod_{i=1}^{k} (\hat{P}(x_i|y_j))^{n_i} \cdot (1-\hat{P}(x_i|y_j))^{1-n_i}\right] \cdot P(y_j)}{\sum_{\ell=1}^{K} \left[\prod_{i=1}^{k} (\hat{P}(x_i|y_\ell))^{n_i} \cdot (1-\hat{P}(x_i|y_\ell))^{1-n_i}\right] \cdot P(y_\ell)} \qquad (4.1.4)$$

where $n_i = 1$ if x_i

$n_i = 0$ if $\bar{x}_i$

k = number of attributes in profile (6)

K = number of disease categories (6)

Diagnoses are also calculated according to the actual distribution of attributes among diseases (4.1.5):

$$P(y_j|X) = \frac{\left[\prod_{\substack{i=1 \\ i\neq B}}^{k} (P(x_i|y_j))^{n_i} (1-P(x_i|y_j))^{1-n_i}\right] \cdot F_j \cdot P(y_j)}{\sum_{\ell=1}^{K} \left[\prod_{\substack{i=1 \\ i\neq B}}^{k} (P(x_i|y_\ell))^{n_i} (1-P(x_i|y_\ell))^{1-n_i}\right] \cdot F_\ell \cdot P(y_\ell)} \qquad (4.1.5)$$

where

$$F_\ell = \begin{cases} P(x_B|x_A,y_\ell) & \text{if } x_B, x_A \\ P(\bar{x}_B|x_A,y_\ell) & \text{if } \bar{x}_B, x_A \\ P(x_B|\bar{x}_A,y_\ell) & \text{if } x_B, \bar{x}_A \\ P(\bar{x}_B|\bar{x}_A,y_\ell) & \text{if } \bar{x}_B, \bar{x}_A \end{cases} \qquad F_j = F_\ell \text{ when } \ell = j$$

For each profile X the disease category with the highest posterior Bayes' probability is chosen as the predicted diagnosis. The number of matching predictions is denoted by M. Sections 2 and 3 of the program are repeated 5 times for each value of $P(x_B|x_A,y_j)$ $j = 1,2,\ldots 6$, and $\bar{M}$ is the average number of matching predictions for these 5 trials

(refer to Tables 4.2.2, 4.2.5).

4. Defining levels of dependence.

In the fourth section of the program $P(x_i|y_j)$ values remain unchanged while $P(x_B|x_A,y_j)$ values are altered. As $|P(x_B|x_A,y_j)-P(x_B|y_j)|$ increases for each y_j, the disease-attribute system may be said to be increasingly deviating from independence. The program resumes execution by ensuring constraints (a) and (b) are satisfied.

Further reference to Bayes' formula in the sections which follow imply (4.1.4).

It is my feeling that the chief duty of a statistician is to interpret data in such a way that they convey knowledge for the purposes of prediction. - *W. Edwards*

Section 4.2

The sample size, LN

The effects of sample size on the accuracy of Bayesian prediction were investigated for varying departures from attribute independence for two disease-attribute populations. Table 4.2.1 below shows the true distribution of attributes among 6 disease categories for the first population considered. The results of computer simulations are tabulated in Table 4.2.2 and presented graphically in Figure 4.2.1.

Table 4.2.1

Data Set 1. The True Distribution of Attributes Among Diseases. x_A and x_B Identify a Dependent Attribute Pair

	x_1	x_2	x_3	x_4	x_5	x_6	x_A	x_B
y_1	.05*	.10	.13	.16	.19	.22	x_1	x_6
y_2	.22	.10	.13	.19	.16	.05	x_2	x_5
y_3	.13	.22	.16	.19	.05	.10	x_3	x_4
y_4	.10	.16	.19	.22	.13	.05	x_1	x_5
y_5	.19	.05	.22	.13	.16	.10	x_2	x_6
y_6	.16	.13	.05	.10	.19	.22	x_3	x_6

*Table entries are $P(x_i|y_j)$

Table 4.2.2

Data Set 1. The Number of Correct Predictions for Sample Sizes of 50,250, and 1000 for Different Levels of Dependence and for Independence

$P(x_B \mid x_A, y_j)$	number of correct predictions, M max M = 64					
$j=1,2,\ldots,6$	LN = 50	$\bar{M}$	LN = 250	$\bar{M}$	LN = 1000	$\bar{M}$
Independence*	29,33,34,39,45	36.0	53,54,56,57,64	56.8	58,58,60,60,63	59.8
.2	25,30,33,36,41	33.0	47,48,49,50,55	49.8	51,52,53,53,58	53.4
.6	21,21,21,22,28	22.6	27,27,28,29,31	28.4	26,27,28,28,30	27.8
.99	17,17,18,18,19	17.8	19,21,21,22,22	21.0	17,19,19,20,22	19.4

*$P(x_B \mid x_A, y_j) = P(x_B \mid y_j)$ $j = 1,2,\ldots,6$

We would quite naturally expect Bayes' formula to perform best when the attributes are independent or when $P(x_B \mid x_A, y_j) \doteq P(x_B \mid y_j)$, $j = 1,2,\ldots,6$. This question will be addressed in Section 4.3. In this section we investigate the effects of sample size, LN, on the accuracy of predictions. For example, suppose we wish to test the null hypothesis H_o: Bayes' formula performs no better with LN=1000 than with LN = 250 when the attributes are independent. The Mann-Whitney U test is useful as a substitute for the unpaired t-test when assumptions underlying the t-test (normally distributed populations with equal variances) are not satisfied. The number of matches is a discrete random variable and hence does not follow a normal distribution;

since the data sets are small but do have ordinal measurement level, the Mann-Whitney U test is appropriate to analyze the data in Table 4.2.2:

$U = \min(U_1, U_2)$ where

$U_1 = n_1 \cdot n_2 + [n_1(n_1+1)/2] - R_1$

$U_2 = n_1 \cdot n_2 - U_1$

n_1 = size of one sample

n_2 = size of other sample

R_1 = sum of ranks assigned to sample 1

R_2 = sum of ranks assigned to sample 2

H_o: $\bar{M}_{LN=250} = \bar{M}_{LN=1000}$ at $P(x_B|x_A,y_j)=P(x_B|y_j)$

Sample 1, LN = 250		Sample 2, LN = 1000	
M	Rank	M	Rank
53	1	58	5.5
54	2	58	5.5
56	3	60	7.5
57	4	60	7.5
64	10	63	9
	$R_1=20$		$R_2=35$

U = 5, DO NOT REJECT H_o

$U_1=20$

$U_2=n_1 \cdot n_2 - U_1$

$=5 \cdot 5 - 20$

$=5$

$U=\min(20,5)$

$=5$

The critical value of U at $\alpha = .05$ is $U_{.05,\,5,\,5} = 4$ for a one tail test. Since the calculated value of U is greater than the critical value we do not reject H_o in favour of better performance with larger sample size. Table 4.2.3

summarizes the results of several such tests.

Table 4.2.3

Data Set 1. Comparisons of Bayesian Accuracy (assuming independence) for Different Sample Sizes

$P(x_B \mid x_A, y_j)$	LN = 50	U	LN = 250	U	LN = 1000
Independence	$\bar{M}$ = 36.0	0*	$\bar{M}$ = 56.8	5	$\bar{M}$ = 59.8
.2	$\bar{M}$ = 33.0	0*	$\bar{M}$ = 49.8	4	$\bar{M}$ = 53.4
.6	$\bar{M}$ = 22.6	2.5*	$\bar{M}$ = 28.4	10	$\bar{M}$ = 27.8
.99	$\bar{M}$ = 17.8	-.5*	$\bar{M}$ = 21.0	6	$\bar{M}$ = 19.4

*indicates significant difference between adjacent means, α = .05, one tail test.

We conclude that a sample size of 250 is preferable to a sample size of 50 and that variations in the number of correct predictions between sample sizes of 250 and 1000, at specified values of $P(x_B \mid x_A, y_j)$, are not significant. Similar computer simulations were performed with the population described in Table 4.2.4. Tables 4.2.5, 4.2.6 and Figure 4.2.2 summarize these results of computer simulations on Data Set 2.

Table 4.2.4

Data Set 2. The True Distribution of Attributes Among Diseases. x_A and x_B Identify a Dependent Attribute Pair

	x_1	x_2	x_3	x_4	x_5	x_6	x_A	x_B
y_1	.47	.37	.33	.34	.46	.50	x_1	x_6
y_2	.13	.29	.39	.34	.34	.18	x_2	x_5
y_3	.13	.31	.37	.45	.61	.09	x_3	x_4
y_4	.15	.37	.58	.38	.44	.45	x_1	x_5
y_5	.73	.20	.22	.34	.15	.28	x_2	x_6
y_6	.47	.18	.45	.54	.08	.53	x_3	x_6

Table 4.2.5

Data Set 2. The Number of Correct Predictions for Sample Sizes of 50, 250 and 1000 for Different Levels of Dependence and for Independence

$P(x_B \mid x_A, y_j)$	number of correct predictions, M												max M = 64					
$j=1,2,\ldots,6$	LN = 50					$\bar{M}$	LN = 250					$\bar{M}$	LN = 1000					$\bar{M}$
Independence	45	48	51	52	55	50.2	50	54	55	57	57	54.6	53	59	61	61	63	59.4
.05	34	34	35	36	39	35.6	34	35	35	36	38	35.6	35	38	40	40	43	39.2
.45	46	51	51	53	53	50.8	50	51	51	54	55	52.2	52	55	57	57	59	56.0
.99	24	29	30	31	34	29.6	29	30	31	32	33	31.0	29	29	30	31	34	30.6

Table 4.2.6

Data Set 2. Comparisons of Bayesian Accuracy (assuming independence) for Different Sample Sizes

$P(x_B \mid x_A, y_j)$	LN = 50	U	LN = 250	U	LN = 1000
Independence	$\bar{M}$ = 50.2	4.5	$\bar{M}$ = 54.6	4.0	$\bar{M}$ = 59.4
.05	$\bar{M}$ = 35.6	11.5	$\bar{M}$ = 35.6	3.5*	$\bar{M}$ = 39.2
.45	$\bar{M}$ = 50.8	8.5	$\bar{M}$ = 52.2	2.5*	$\bar{M}$ = 56.0
.99	$\bar{M}$ = 29.6	9.5	$\bar{M}$ = 31.0	10	$\bar{M}$ = 30.6

*indicates significant difference between adjacent means, α = .05, one tail test.

From Table 4.2.6 we note that increasing the sample size does not consistently result in increased Bayesian accuracy.

To date, applications of Bayes' formula to medical diagnosis have rarely taken LN above 300. The remaining computer runs are based on LN = 500 (i) since we wish to simulate practical applications in restricting the sample size (ii) since the performance of Bayes' formula when LN = 250 and LN = 1000 are not too different at extreme levels of dependence, it is expected that a sample size between 250 and 1000 is satisfactory.

FIGURE 4.2.1, DATA SET 1
PLOT OF $\bar{M}$ vs $P(x_B / x_A, y_i)$

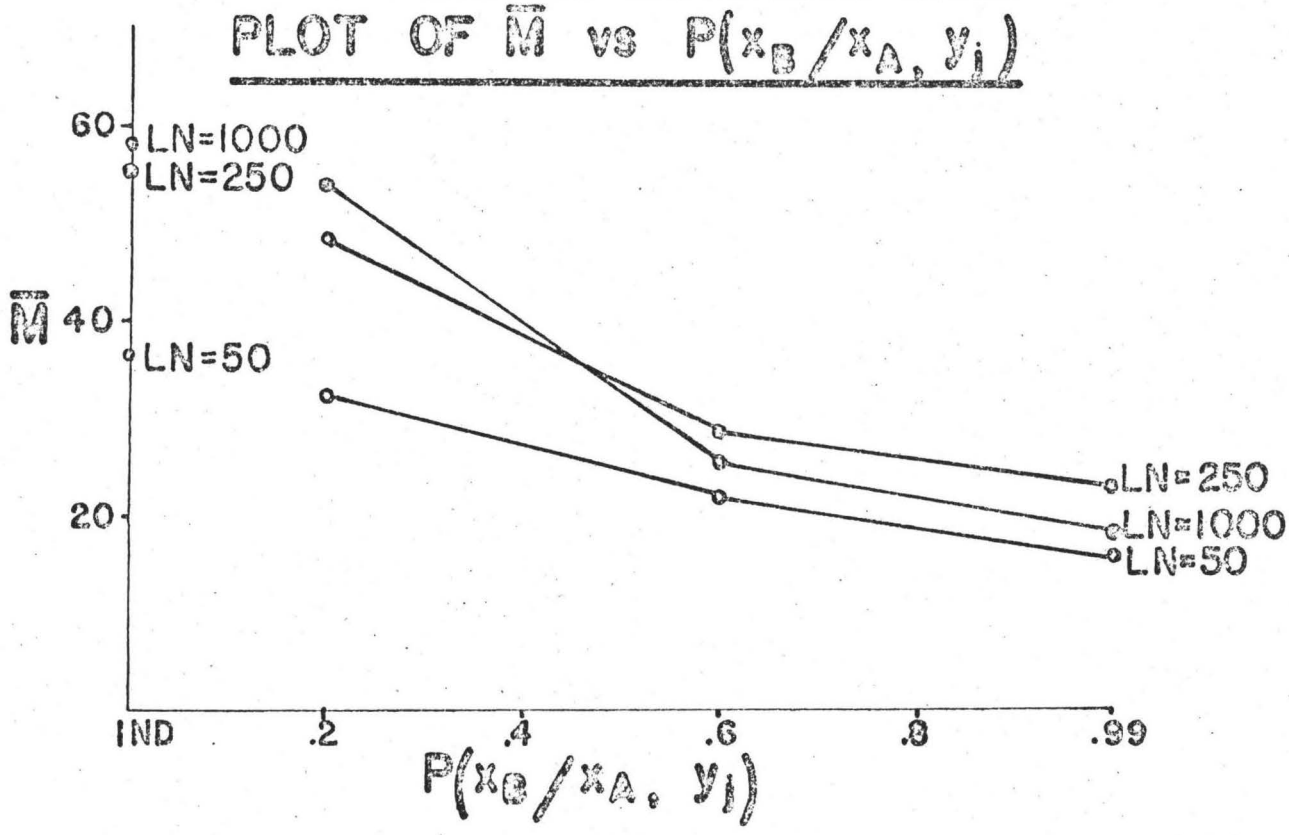
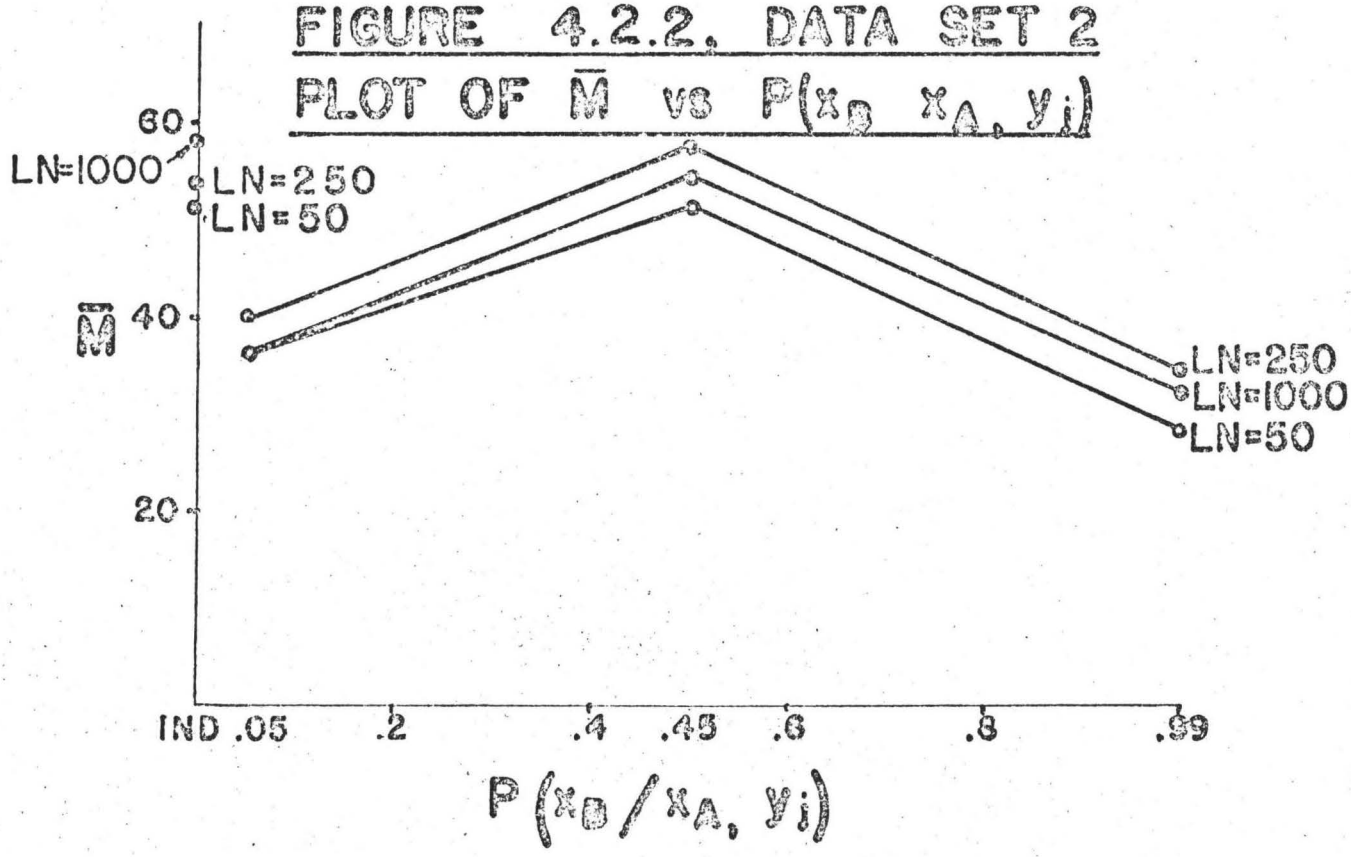


FIGURE 4.2.2, DATA SET 2
PLOT OF $\bar{M}$ vs $P(x_B\ x_A, y_i)$



Quoth she, I've heard old cunning stagers
Say fools for arguments use wagers. - *Samuel Butler*

Section 4.3

Effects of Non-Independence of Attribute Pairs on Bayesian Accuracy.

Introduction

In this section the effects of non-independence of attribute pairs on Bayesian accuracy are examined. A disease-attribute (D-A) population is defined by $P(x_i|y_j)$ and $P(x_B|x_A,y_j)$ values, i, j = 1,2,...6. The Monte Carlo program then generates 500 individual profiles in each disease category according to the actual probability distribution of attributes among diseases. $P(x_i|y_j)$ values are estimated from the frequency of attribute occurrence within each disease category. Bayesian predictions are calculated (i) from actual $P(x_i|y_j)$ and $P(x_B|x_A,y_j)$ values, accounting for attribute dependence (ii) from estimated $P(x_i|y_j)$ values, assuming attribute independence. In each case the predicted diagnosis is the disease category with the highest posterior probability. The average number of matching predictions for a specified value of $P(x_B|x_A,y_j)$ is denoted by $\bar{M}$.

A parameter, DH, which measures deviations from attribute independence is defined. A visual inspection of

plots of $\bar{M}$ versus DH suggests that M may be a linear function of DH, however, this hypothesis is not consistently supported by the data.

Notation

A word or two concerning notation may be helpful for the discussions which follow. Three distinct D-A populations are examined, they are referred to as Data Set 3, Data Set 4, and Data Set 5. $\bar{M}_\ell$ is the average number of correct Bayesian predictions (assuming attribute independence) when $P(x_B|x_A,y_j) = \ell$. Similarly, $\hat{\sigma}^2_\ell$ is the variance of M_ℓ values. For example, refer to Table 4.3.2 of Data Set 3: when $P(x_B|x_A,y_j) = .4$, we note $\bar{M}_{.4} = 30.80$ and $\hat{\sigma}^2_{.4} = 2.70$.

Defining DH

Figures 4.3.1, 4.3.2, and 4.3.3 are plots of $\bar{M}$ versus $P(x_B|x_A,y_j)$ for three distinct disease-attribute populations. While the shapes of the plots differ markedly, we may expect $\bar{M}$ to decrease with increasing deviations from attribute independence. Referring to Table 4.3.4 of Data Set 4, we note $P(x_B|y_j)$ values are .50, .34, .45, .44, .28, .53 for $j = 1,2,\ldots,6$ respectively. If deviations from attribute independence are to occur, then we may expect Bayesian prediction to perform best when $.28 \leq P(x_B|x_A,y_j) \leq .53$ since this range of $P(x_B|x_A,y_j)$ values minimize (in a very non-rigorous sense) deviations from attribute independence.

Figure 4.3.2 indicates that Bayes' formula predicts equally well at $P(x_B|x_A,y_j) = .4$ and .5, and that $\bar{M}$ decreases significantly as $P(x_B|x_A,y_j)$ assumes values outside the range .4 to .5.

Similarly, in Data Set 5, Table 4.3.7 indicates $P(x_B|y_j)$ values are .22, .16, .19, .13, .10, and .22 for $j = 1,2,\ldots,6$ respectively. We may expect Bayes' formula to perform best when $.10 \leq P(x_B|x_A,y_j) \leq .22$. Figure 4.3.3 indicates that this is so. While $\bar{M}_{.1}$ and $\bar{M}_{.2}$ do not differ significantly, both are significantly greater than other $\bar{M}$ values (except $\bar{M}_{IND}$, when the attributes are independent).

We wish to define a parameter, DH, which measures deviations from attribute independence such that plots of $\bar{M}$ versus DH will have similar configurations for any disease-attribute population. Defining DH by:

$$DH = \left(\sum_{j=1}^{j=K} |P(x_B|x_A,y_j) - P(x_B|y_j)| \right)/K \quad \text{where}$$

K is the number of disease categories in the population

DH measures an average deviation from attribute independence, $0 \leq DH < 1$. DH = 0 implies attribute independence. As DH increases, a disease-attribute population may be said to be increasingly deviating from attribute independence.

M as a linear function of DH.

Figures 4.3.1a, 4.3.2a, 4.3.3a show $\bar{M}$ plotted against DH for Data Sets 3, 4, and 5 respectively. $\bar{M}$ indicates the central tendency of repeated observations of M at various values of DH. These plots suggest that the response variable M may be a linear function of the independent variable DH. Since repeat observations of M are available at each value of DH, an estimate of pure error may be calculated, and this estimate may be used to judge the adequacy of the following model:

$$M_{ij} = B_0 + B_1 \cdot DH_{ij} + \varepsilon_{ij} \qquad \text{where} \qquad (1)$$

B_0, B_1 are independent parameters in the model

ε_{ij} are assumed to be independently and identically distributed $N(0,\sigma^2)$

$i = 1,2,\ldots,p$ p is the number of distinct points in the design space

$j = 1,2,\ldots n_i$ n_i is the number of replicates at each distinct point in the design space

The estimates of B_0 and B_1 shown in Figures 4.3.1a, 4.3.2a, 4.3.3a are obtained using standard least squares procedures (refer to Draper and Smith (1966)).

A detailed analysis of Data Sets 3, 4, and 5 follows.

Data Set 3

Table 4.3.1

Data Set 3. The True Distribution of Attributes Among Diseases. x_A and x_B Identify a Dependent Attribute Pair

y_j	x_1	x_2	x_3	x_4	x_5	x_6	x_A	x_B
y_1	.06*	.09	.11	.18	.29	.36	x_3	x_4
y_2	.04	.07	.13	.19	.50	.35	x_1	x_2
y_3	.03	.08	.14	.40	.26	.34	x_6	x_4
y_4	.04	.07	.15	.21	.27	.48	x_3	x_6
y_5	.05	.10	.11	.22	.28	.32	x_1	x_2
y_6	.06	.11	.12	.20	.24	.31	x_5	x_6

*Table entries are $P(x_i|y_j)$

Table 4.3.2

Data Set 3. The Number of Correct Bayesian Predictions, M, (assuming independence) at various levels of dependence and at Independence

	Dependence Levels, $P(x_B\|x_A,y_j)$									
Run	IND*	.1	.2	.3	.4	.5	.6	.7	.8	.9
1	32	15	20	27	32	27	25	24	23	22
2	39	31	26	30	32	24	23	22	19	18
3	50	40	32	34	29	25	23	22	19	19
4	52	38	29	32	29	22	21	20	20	18
5	32	21	30	30	32	23	26	24	21	23
Mean $\bar{M}$	41.00	29.00	27.40	30.60	30.80	24.20	23.60	22.40	20.40	20.00
$\hat{\sigma}^2$	92.0	116.50	21.80	6.80	2.70	3.70	3.80	2.80	2.80	5.50
DH	0	.167	.140	.140	.170	.243	.343	.443	.543	.643

*Independence, ie, $P(x_B|x_A,y_j) = P(x_B|y_j)$ $j = 1,2,\ldots,6$

Referring to Table 4.3.2, we wish to examine for significant differences among M values at different levels of dependence. 5 observations of M are available at each $P(x_B|x_A,y_j)$ value. Assuming M are normally distributed at each $P(x_B|x_A,y_j)$ value, we first establish homogeneity of group variances which is required before comparing $\bar{M}$ values. The $\grave{L}$ (see Croxton et al (1967)) and Q (see Anderson and McLean (1974)) statistics may be used to examine for homogeneity of group variances. If there is any difference between $\hat{\sigma}^2$ values, L will be less than 1, approaching 0 as its lower limit. L = 0 represents a condition of maximum non-uniformity which would not be approached in actual practice. The Burr-Foster Q test is also used to test the assumption of homogeneity of population variances which is required in the ANOVA technique. Large values of Q lead to rejection of the hypothesis of equal population variances. Both tests can be applied to groups of unequal size.

Hypothesis testing

Inspection of Table 4.3.2 suggests that $\hat{\sigma}^2_{.2}$, $\hat{\sigma}^2_{.3}$, ..., $\hat{\sigma}^2_{.9}$ may not be significantly different, while $\hat{\sigma}^2_{IND}$, $\hat{\sigma}^2_{.1}$ may differ significantly from other group variances. These suppositions are examined for the following tests of hypotheses:

$H_o(1)$: $\hat{\sigma}^2_{.2} = \hat{\sigma}^2_{.3} = \ldots = \hat{\sigma}^2_{.9}$

L TEST: $L = p(\prod_{i=1}^{p} \hat{\sigma}^2_i)^{1/p} \; / \; (\sum_{i=1}^{p} \hat{\sigma}^2_i)$ p = number of groups (8)

$= 8(21.80 \cdot 6.80 \cdot \ldots \cdot 5.50)^{1/8} / (21.80+6.80+\ldots+5.50)$

$= .76$ COMPARE TO CRITICAL VALUE, $L_{.05,5,8} = .620$

DO NOT REJECT $H_o(1)$

Q TEST: $Q = (\sum_{i=1}^{p} \hat{\sigma}^4_i) \; / \; (\sum_{i=1}^{p} \hat{\sigma}^2_i)^2$

$= (21.80^2+6.80^2+\ldots+5.5^2) / (21.80+6.80+\ldots+5.5)^2$

$= .24$ COMPARE TO CRITICAL VALUE, $Q_{.05,4,8} = .232$

DO NOT ACCEPT $H_o(1)$ AT $\alpha = .05$

COMPARE TO CRITICAL VALUE $Q_{.01,4,5} = .274$

DO NOT REJECT $H_o(1)$ AT $\alpha = .01$

$H_o(2)$: $\hat{\sigma}^2_{.1} = \hat{\sigma}^2_{.2} = \ldots = \hat{\sigma}^2_{.9}$

L TEST: CALCULATED L = .36 COMPARE TO CRITICAL VALUE
$L_{.05,5,9} = .626$
DO NOT ACCEPT $H_o(2)$

Q TEST: CALCULATED Q = .51 COMPARE TO CRITICAL VALUE
$Q_{.05,4,9} = .206$
DO NOT ACCEPT $H_o(2)$

$H_o(3)$: $\hat{\sigma}^2_{IND} = \hat{\sigma}^2_{.2} = \hat{\sigma}^2_{.3} = \ldots = \hat{\sigma}^2_{.9}$

L TEST: CALCULATED L = .42 COMPARE TO CRITICAL VALUE
$L_{.05,5,9} = .626$
DO NOT ACCEPT $H_o(3)$

Q TEST: CALCULATED Q = .45 COMPARE TO CRITICAL VALUE
$Q_{.05,4,9} = .206$
DO NOT ACCEPT $H_o(3)$

Referring to $H_o(1)$, the L tests accepts the null hypothesis at $\alpha = .05$, while the Q tests accepts $H_o(1)$ at an α level slightly greater than .05. Both tests agree in rejecting $H_o(2)$ and $H_o(3)$. These results are interpreted as follows: $\hat{\sigma}^2_{.2}$, $\hat{\sigma}^2_{.3}$, ..., $\hat{\sigma}^2_{.9}$ do not differ significantly, and $\hat{\sigma}^2_{IND}$, $\hat{\sigma}^2_{.1}$ differ significantly from other group variances. Having established homogeneity of group variances at $P(x_B|x_A,y_j) = .2, .3, \ldots, .9$ we may proceed to examine for differences between $\bar{M}$:

$$H_o: \quad \bar{M}_{.2} = \bar{M}_{.3} = \ldots = \bar{M}_{.9}$$

ANOVA TABLE FOR DATA SET 3

SOURCE	DF	SS	MS	F	$F_{.05}$(CRITICAL)
between groups	7	631.175	90.168	14.456	2.32: REJECT H_o
within groups	32	199.600	6.238		

Since the calculated F = 14.456 exceeds $F_{CRIT} = 2.32$, we conclude that at least two $\bar{M}$ values are significantly different. One test that allows investigation of all possible pairs of means in a sequential manner, and keeps the α level constant for each comparison, is the Newman-Keuls test (refer to Anderson and McLean (1974)).

The upper right triangular portion in Table 4.3.3 contains differences between $\bar{M}$ values, the lower left triangular section contains critical values for $\alpha = .05$. The calculated difference is compared to its corresponding

critical value, and a calculated difference greater than its critical value implies a significant difference between two means. For example, .40 is the first value in the first row of Table 4.3.3: $\bar{M}_{.8}-\bar{M}_{.9}$ = 20.40-20.00=.40. Since .40 < 3.22, H_o: $\bar{M}_{.8} = \bar{M}_{.9}$ is accepted. Similarly, 10.80 is the last value in the first row of Table 4.3.3: $\bar{M}_{.4}-\bar{M}_{.9}$ = 30.80-20.00=10.80. Since 10.80 > 5.12, H_o: $\bar{M}_{.4} = \bar{M}_{.9}$ is rejected in favour of $\bar{M}_{.4} > \bar{M}_{.9}$. The last column in Table 4.3.3 and Figure 4.3.1 summarize the findings of the multiple comparisons.

Table 4.3.3

Data Set 4. Newman-Keuls Multiple Comparison Test for Significant Differences Between Group Means.

Groups, Ascending Means Left to Right

	.9	.8	.7	.6	.5	.2	.3	.4	Groups having no sign. diff. for $\bar{M}$
.9		.40	2.40	3.60	4.20	7.40*	10.6*	10.8*	.5,.6,.7,.8&.9
.8	3.22		2.00	3.20	3.80	7.00*	10.2*	10.4*	.5,.6,.7&.8
.7	3.89	3.22		1.20	1.80	5.00*	8.2*	8.4*	.5,.6&.7
.6	4.29	3.89	3.22		.60	3.80	7.0*	7.2*	.2,.5&.6
.5	4.57	4.29	3.89	3.22		3.20	6.4*	6.6*	.2&.5
.2	4.79	4.51	4.29	3.89	3.22		3.2	3.4	.3,.4,&.2
.3	4.97	4.79	4.57	4.29	3.89	3.22		.2	.4&.3
.4	5.12	4.97	4.79	4.57	4.29	3.89	3.22		

* indicates significant difference between group means, α = .05.

In Figure 4.3.1a, $\bar{M}$ indicates the central tendency of repeated observations of the dependent variable M at 7

distinct values[10] of the independent variable DH. The fitted model is

$$\hat{M}_{ij} = 31.590 - 20.008\ DH_{ij}$$

The ANOVA Table is shown below.

ANOVA TABLE

Data Set 3, Model (1)

Source	DF	SS	MS	F	$F_{CRIT}(\alpha=.05)$
Model	2	25373.756	12686.878		
Mean	1	24850.225	24850.225		
Model(CFM)	1	523.531	523.531		
Residual	38	307.244	8.085		
Lack of fit	5	82.044	16.409	2.404	2.515
Pure Error	33	225.200	6.824		
Total	40	25681.000	642.025		
Mean	1	24850.225	24850.225		
Total(CFM)	39	830.775	21.302		

$R^2 = .988$ $R^2(CFM) = .630$

Since the calculated F = 2.404 is less than F_{CRIT}= 2.515, we conclude that the data supports the model at $\alpha = .05$. Values of R^2 and $R^2(CFM)$ close to 1 indicate that the model is doing a reasonable job of explaining the variation in the data.

FIGURE 4.3.1, DATA SET 3
PLOT OF $\bar{M}$ VS $P(x_B / x_A, y_j)$

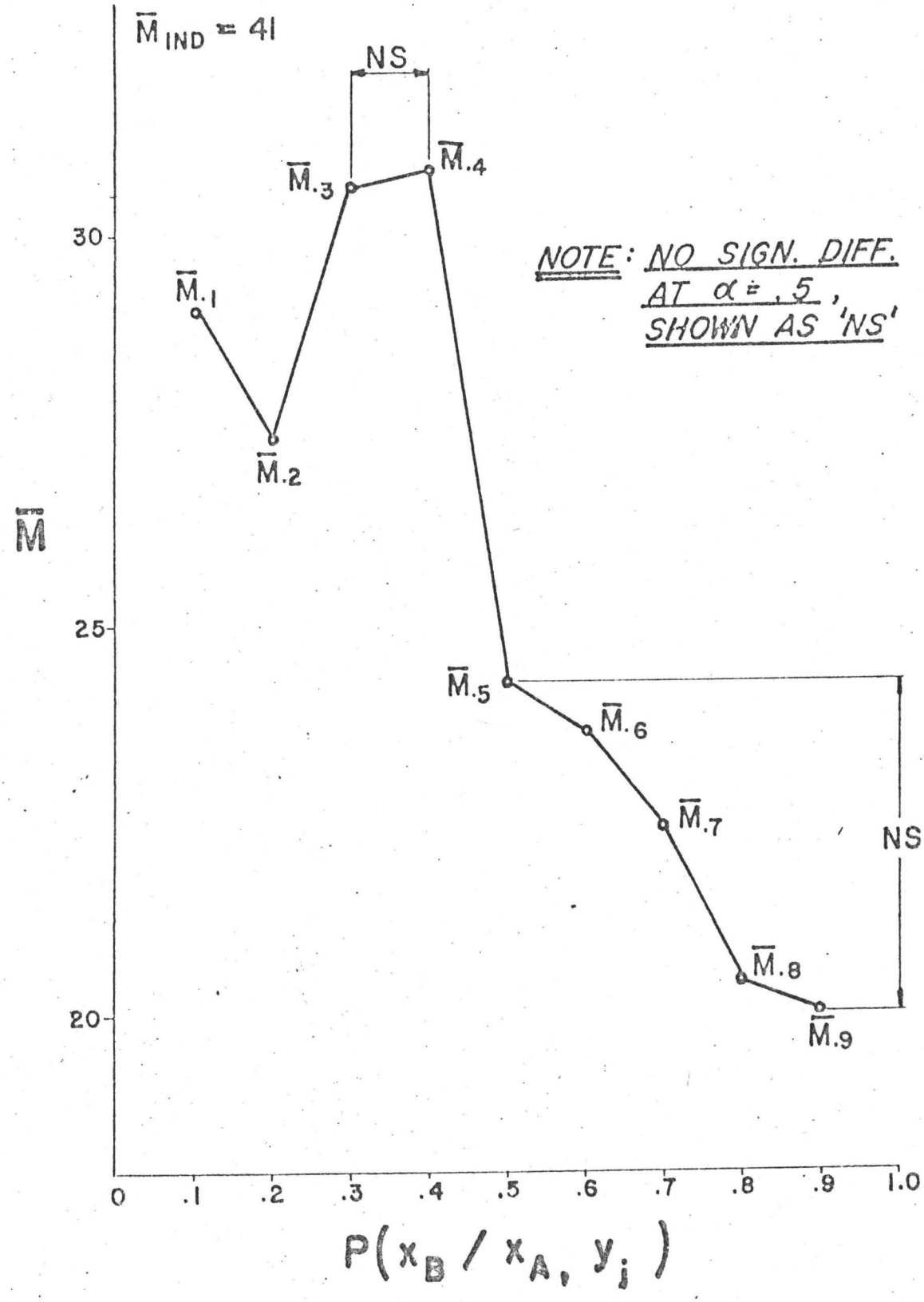
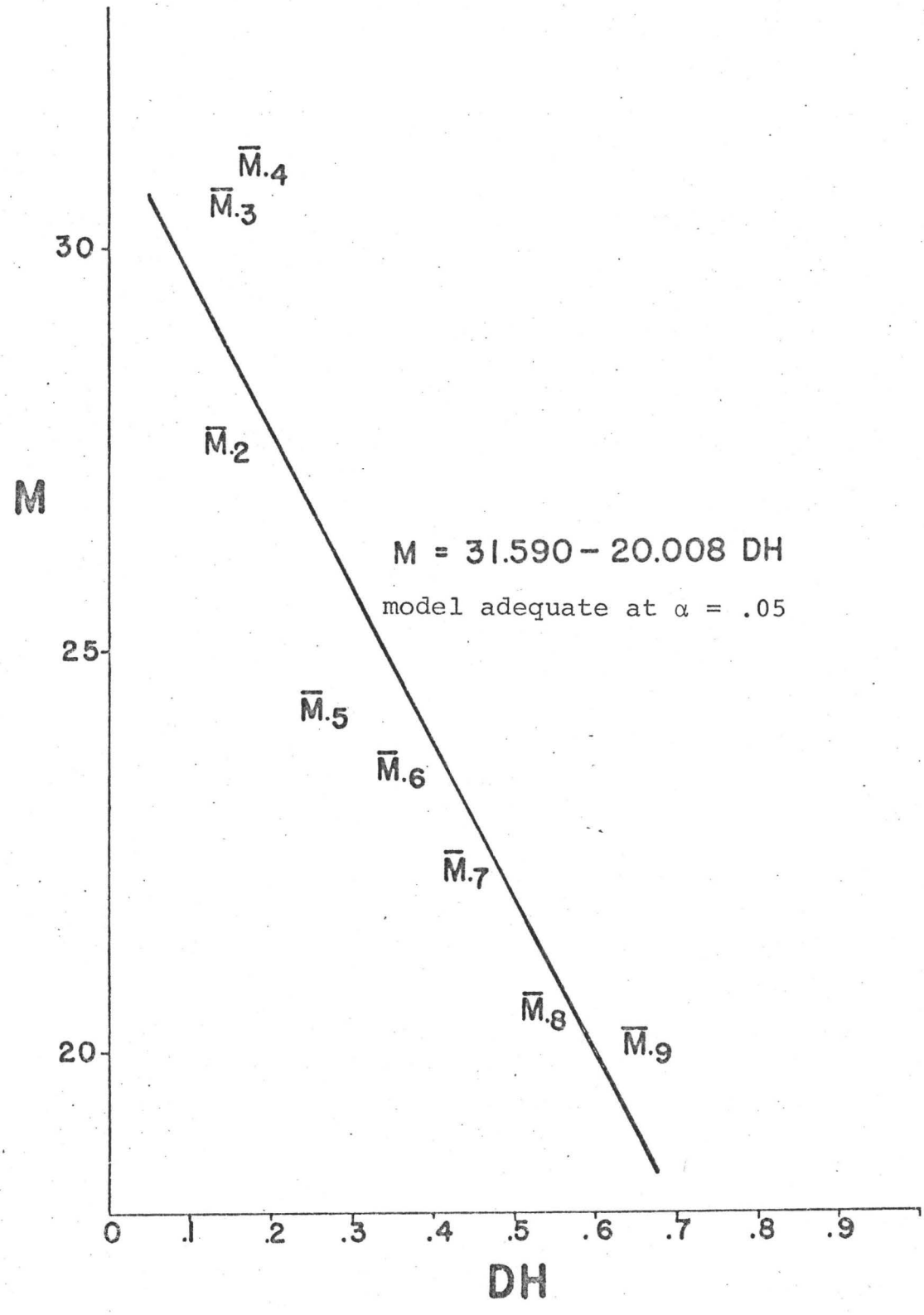


FIGURE 4.3.1a, DATA SET 3

M AS A LINEAR FUNCTION OF DH



Data Set 4

Table 4.3.4

Data Set 4. The True Distribution of Attributes Among Diseases. x_A and x_B Identify a Dependent Attribute Pair

y_j	x_1	x_2	x_3	x_4	x_5	x_6	x_A	x_B
y_1	.47*	.37	.33	.34	.46	.50	x_1	x_6
y_2	.13	.29	.39	.34	.34	.18	x_2	x_5
y_3	.13	.31	.37	.45	.61	.09	x_3	x_4
y_4	.15	.37	.58	.38	.44	.45	x_1	x_5
y_5	.73	.20	.22	.34	.15	.28	x_2	x_6
y_6	.47	.18	.45	.54	.08	.53	x_3	x_6

*Table entries are $P(x_i|y_j)$

Table 4.3.5

Data Set 4. The Number of Correct Bayesian Predictions, M, (assuming independence) at various levels of dependence and at independence

	Dependence Levels, $P(x_B \mid x_A, y_j)$									
Run	IND	.1	.2	.3	.4	.5	.6	.7	.8	.9
1	60	39	44	52	54	56	51	43	41	39
2	56	39	44	54	57	57	56	45	38	36
3	60	41	45	53	57	57	52	42	37	35
4	60	39	44	50	55	55	55	43	38	36
5	55	42	46	56	56	56	54	43	37	35
Mean $\bar{M}$	58.20	40.00	44.60	53.00	55.80	56.20	53.60	43.20	38.20	36.20
$\hat{\sigma}^2$	6.20	2.00	.80	5.00	1.70	.70	4.30	1.20	2.70	2.70
DH	0	.323	.223	.130	.083	.087	.177	.277	.377	.477

Hypothesis Testing

Inspection of Table 4.3.5 suggests homogeneity of all group variances. As before, the L and Q tests are applied to the data.

H_o: $\hat{\sigma}^2_{IND} = \hat{\sigma}^2_{.1} = \hat{\sigma}^2_{.2} = \ldots = \hat{\sigma}^2_{.9}$

TEST STATISTICS: L TEST

BURR-FOSTER Q TEST

L TEST: $L = p(\prod_{i=1}^{p} \hat{\sigma}_i^2)^{1/p} / (\sum_{i=1}^{p} \hat{\sigma}_i^2)$ p = number of groups (10)

$= 10(6.20 \cdot 2.00 \cdot \ldots \cdot 2.70)^{1/10}/(6.20+2.00+\ldots+2.70)$

$= .79$ COMPARE TO CRITICAL VALUE, $L_{.05,5,10} = .631$

DO NOT REJECT H_o

Q TEST: $Q = (\sum_{i=1}^{p} \hat{\sigma}_i^4) / (\sum_{i=1}^{p} \hat{\sigma}_i^2)^2$

$= (6.20^2+2.00^2+\ldots+2.70^2)/(6.20+2.00+\ldots+2.70)^2$

$= .142$ COMPARE TO CRITICAL VALUE $Q_{.05,4,10} = .284$

DO NOT REJECT H_o

Having established homogeneity of group variances, we proceed to test

H_o: $\bar{M}_{IND} = \bar{M}_{.1} = \bar{M}_{.2} = \ldots = \bar{M}_{.9}$

ANOVA TABLE FOR DATA SET 4

SOURCE	SS	DF	MS	F	$F_{.05}$ (CRITICAL)
between groups	3111.310	9	345.701	126.645	2.12: REJECT H_o
within groups	109.188	40	2.729		

The large value of the calculated F statistic indicates that at least two $\bar{M}$ values differ significantly ($\alpha = .05$). As before, the Newman-Keuls procedure examines for significant differences between all combinations of $\bar{M}$ pairs.

Table 4.3.6

Data Set 4. Newman-Keuls Multiple Comparison Test for Significant Differences Between Group Means.

Groups, Ascending Means Left to Right

	.9	.8	.1	.7	.2	.3	.6	.4	.5	IND	Groups having no sign. diff for $\bar{M}$
.9		2.0	3.8*	7.0*	8.4*	16.8*	17.4*	19.6*	20.0*	22.0*	.8&.9
.8	2.11		1.8	5.0*	6.4*	14.8*	15.4*	17.6*	18.0*	20.0*	.1&.8
.1	2.54	2.11		3.2*	4.6*	13.0*	13.6*	15.8*	16.2*	18.2*	
.7	2.80	2.54	2.11		1.4	9.8*	10.4*	12.6*	13.0*	15.0*	.2&.7
.2	2.99	2.80	2.54	2.11		8.4*	9.0*	11.2*	11.6*	13.6*	
.3	3.13	2.99	2.80	2.54	2.11		.6	2.8*	3.2*	5.2*	.6&.3
.6	3.24	3.13	2.99	2.80	2.54	2.11		2.2*	2.6*	4.6*	
.4	3.34	3.24	3.13	2.99	2.80	2.54	2.11		.4	2.4	IND&.4
.5	3.42	3.34	3.24	3.13	2.99	2.80	2.54	2.11		2.0	.5&.4
IND	3.49	3.42	3.34	3.24	3.13	2.99	2.80	2.54	2.11		IND&.5

*indicates significant difference between group means at α=.05

The findings in Table 4.3.6 are summarized in Figure 4.3.2. Referring to Figure 4.3.2a, $\bar{M}$ indicates the central tendency

of repeated observations of M at 10 distinct values of the independent variable DH. The fitted model (1) is:

$$\hat{M}_{ij} = 59.556 - 54.115\ DH_{ij}$$

The ANOVA Table is shown below.

ANOVA TABLE

Data Set 4, Model (1)

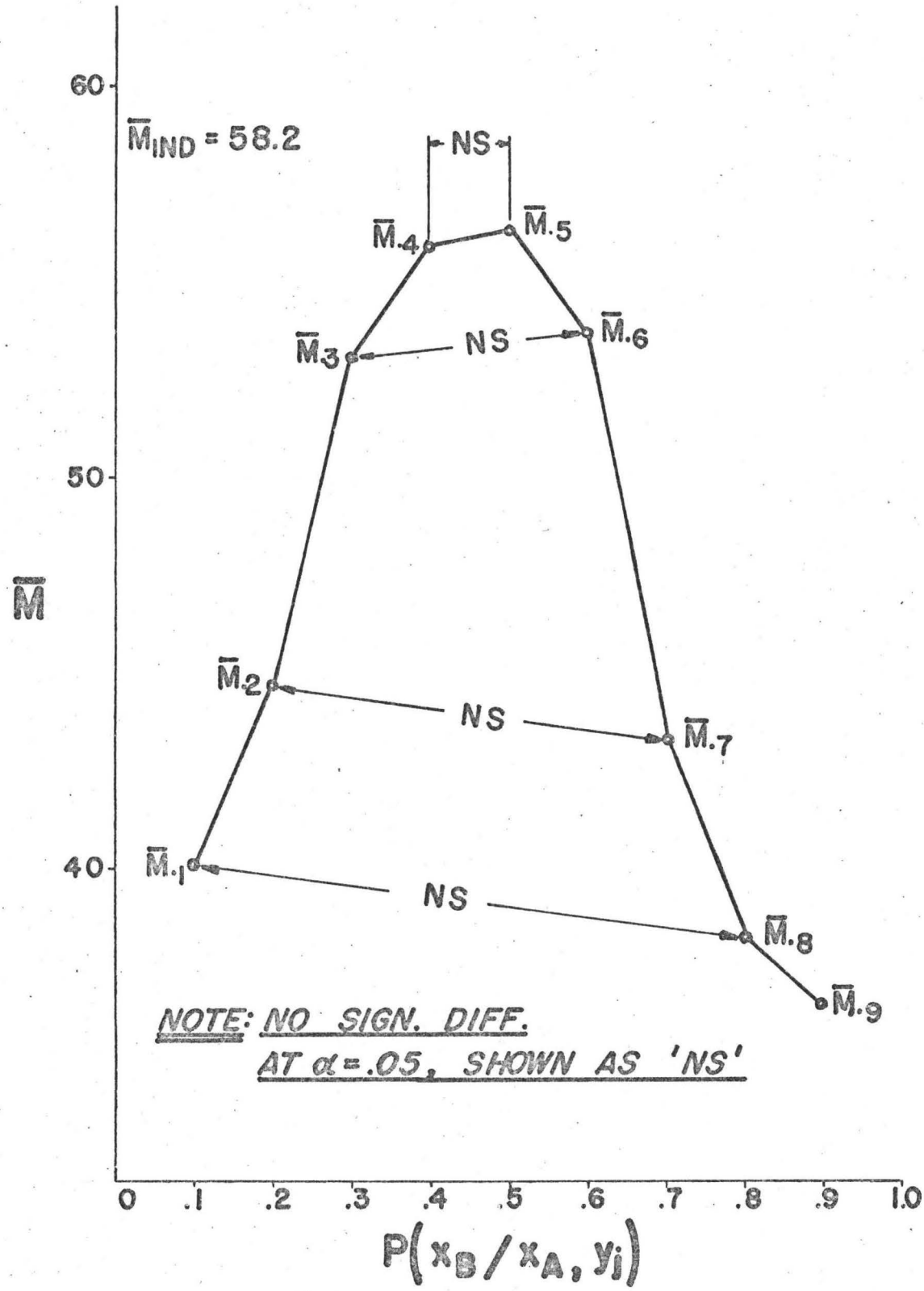
Source	DF	SS	MS	F	$F_{CRIT}(\alpha=.05)$
Model	2	117636.658	58818.329		
Mean	1	114720.500	114720.500		
Model(DFM)	1	2916.159	2916.159		
Residual	48	304.341	6.340		
Lack of fit	8	195.141	24.393	8.935	2.180
Pure Error	40	109.200	2.730		
Total	50	117941.000	2358.820		
Mean	1	114720.500	114720.500		
Total(CFM)	49	3220.500	65.72449		

$R^2 = .997$ $R^2(CFM) = .906$

The calculated F = 8.935 exceeds $F_{CRIT} = 2.180$. The proposed model (1) is clearly inadequate.

FIGURE 4.3.2, DATA SET 4

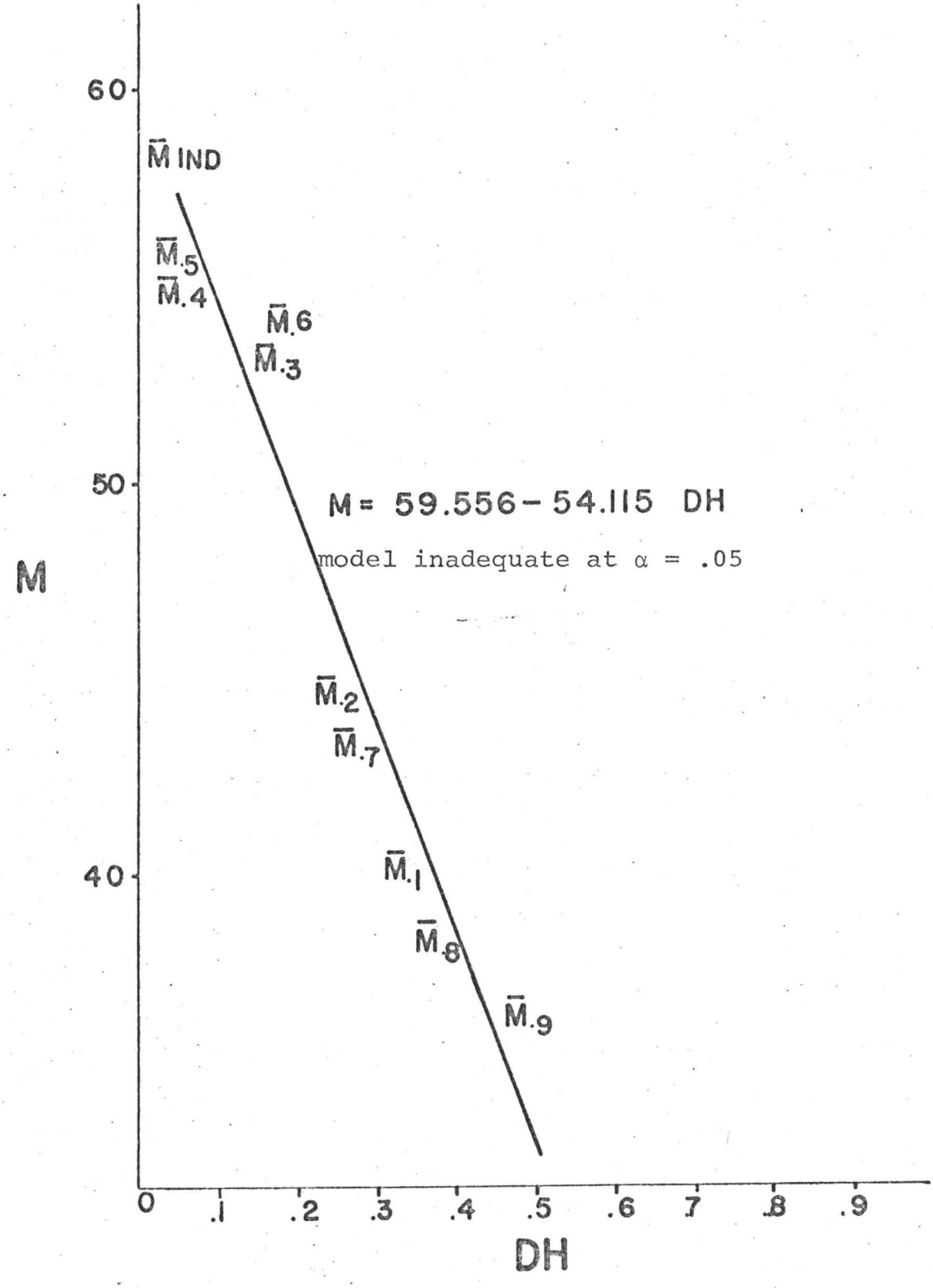
PLOT OF $\bar{M}$ vs $P(x_B/x_A, y_j)$



NOTE: NO SIGN. DIFF.
AT $\alpha = .05$, SHOWN AS 'NS'

FIGURE 4.3.2a, DATA SET 4

M AS A LINEAR FUNCTION OF DH



Data Set 5

Table 4.3.7

Data Set 5. The True Distribution of Attributes Among Diseases. x_A and x_B Identify a Dependent Attribute Pair

	x_1	x_2	x_3	x_4	x_5	x_6	x_A	x_B
y_1	.05*	.10	.13	.16	.19	.22	x_1	x_6
y_2	.22	.10	.13	.19	.16	.05	x_2	x_5
y_3	.13	.22	.16	.19	.05	.10	x_3	x_4
y_4	.10	.16	.19	.22	.13	.05	x_1	x_5
y_5	.19	.05	.22	.13	.16	.10	x_2	x_6
y_6	.16	.13	.05	.10	.19	.22	x_3	x_6

*Table entries are $P(x_i|y_j)$

Table 4.3.8

Data Set 5. The Number of Correct Bayesian Predictions, M, (assuming independence) at various levels of dependence and at independence

	Dependence Levels, $P(x_B|x_A,y_j)$									
Run	IND	.1	.2	.3	.4	.5	.6	.7	.8	.9
1	53	52	53	44	37	33	29	27	23	22
2	62	54	55	49	38	30	28	25	21	19
3	62	53	54	48	39	31	28	25	21	19
4	61	55	61	52	41	33	32	28	22	21
5	58	54	55	50	39	30	27	25	22	21
Mean $\bar{M}$	59.20	53.60	55.60	48.60	38.80	31.40	28.80	26.00	21.80	20.40
$\hat{\sigma}^2$	14.70	1.30	9.80	8.80	2.20	2.30	3.70	2.00	.70	1.80
DH	0	.07	.043	.128	.228	.328	.428	.528	.628	.728

Hypothesis Testing

Inspection of Table 4.3.8 suggests homogeneity of group variances. As before, the L and Q statistics are calculated from the data and compared to their critical values.

$$H_o: \quad \hat{\sigma}^2_{IND} = \hat{\sigma}^2_{.1} = \hat{\sigma}^2_{.2} = \ldots = \hat{\sigma}^2_{.9}$$

$$\text{L TEST:} \quad L = p\left(\prod_{i=1}^{p} \hat{\sigma}^2_i\right)^{1/p} \Big/ \left(\sum_{i=1}^{p} \hat{\sigma}^2_i\right)$$

$$= 10(14.70 \cdot 1.30 \cdot \ldots \cdot 1.80)^{1/10} / (14.70+1.30+\ldots+1.80)$$

$$= .65 \quad \text{COMPARE TO } L_{.05,5,10} = .631$$

$$\text{DO NOT REJECT } H_o$$

$$\text{Q TEST:} \quad Q = \left(\sum_{i=1}^{p} \hat{\sigma}^4_i\right) \Big/ \left(\sum_{i=1}^{p} \hat{\sigma}^2_i\right)^2$$

$$= (14.70^2+1.30^2+\ldots+1.80^2)/(14.70+1.30+\ldots+1.80)^2$$

$$= .189 \quad \text{COMPARE TO } Q_{.05,4,10} = .284$$

$$\text{DO NOT REJECT } H_o$$

Both tests support the null hypothesis. Having established homogeneity of group variances, we may proceed to test

$$H_o: \quad \bar{M}_{IND} = \bar{M}_{.1} = \bar{M}_{.2} = \ldots = \bar{M}_{.9}$$

ANOVA TABLE FOR DATA SET 5

SOURCE	SS	DF	MS	F	$F_{.05}$ (CRITICAL)
between groups	9791.000	9	1087.890	230.013	2.12
within groups	189.188	40	4.729		

Since the calculated value of F = 230.013 exceeds the critical value of F = 2.12 the null hypothesis H_o: $\bar{M}_{IND} = \bar{M}_{.1} = \ldots = \bar{M}_{.9}$ is rejected. As before, the Newman-Keuls procedure allows for comparisons of $\bar{M}$ pairs.

Table 4.3.9

Data Set 5. Newman-Keuls Multiple Comparison Test for Significant Differences Between Group Means.

Groups, Ascending Means Left to Right

	.9	.8	.7	.6	.5	.4	.3	.1	.2	IND	Groups having no sign. diff. for $\bar{M}$
.9		1.4	5.6*	8.4*	11.0*	18.4*	28.2*	33.2*	35.2*	38.8*	.8&.9
.8	2.78		4.2*	7.0*	9.6*	17.0*	26.8*	31.8*	33.8*	37.4*	
.7	3.35	2.78		2.8*	5.4*	12.8*	22.6*	27.6*	24.6*	33.2*	
.6	3.69	3.35	2.78		2.6	10.0*	19.8*	24.8*	26.8*	30.4*	.5&.6
.5	3.93	3.69	3.35	2.78		7.4*	17.2*	22.2*	24.2*	27.8*	
.4	4.11	3.93	3.69	3.35	2.78		9.8*	14.8*	16.8*	20.4*	
.3	4.27	4.11	3.93	3.69	3.35	2.78		5.0*	7.0*	10.6*	
.1	4.40	4.27	4.11	3.93	3.69	3.35	2.78		2.0	5.6*	.2&.1
.2	4.50	4.40	4.27	4.11	3.93	3.69	3.35	2.78		33.6*	
IND	4.60	4.50	4.40	4.27	4.11	3.93	3.69	3.35	2.78		

*indicates significant difference between group means at α=.05

The findings in Table 4.3.9 are summarized in Figure 4.3.3. In Figure 4.3.3a, $\bar{M}$ values indicate the central tendency of repeated observations of M at 10 distinct values

of the independent variable DH. The fitted model (1) is:

$$\hat{M}_{ij} = 55.665 - 55.220\ DH_{ij}$$

The ANOVA table is shown below.

ANOVA TABLE

Data Set 5, Model (1)

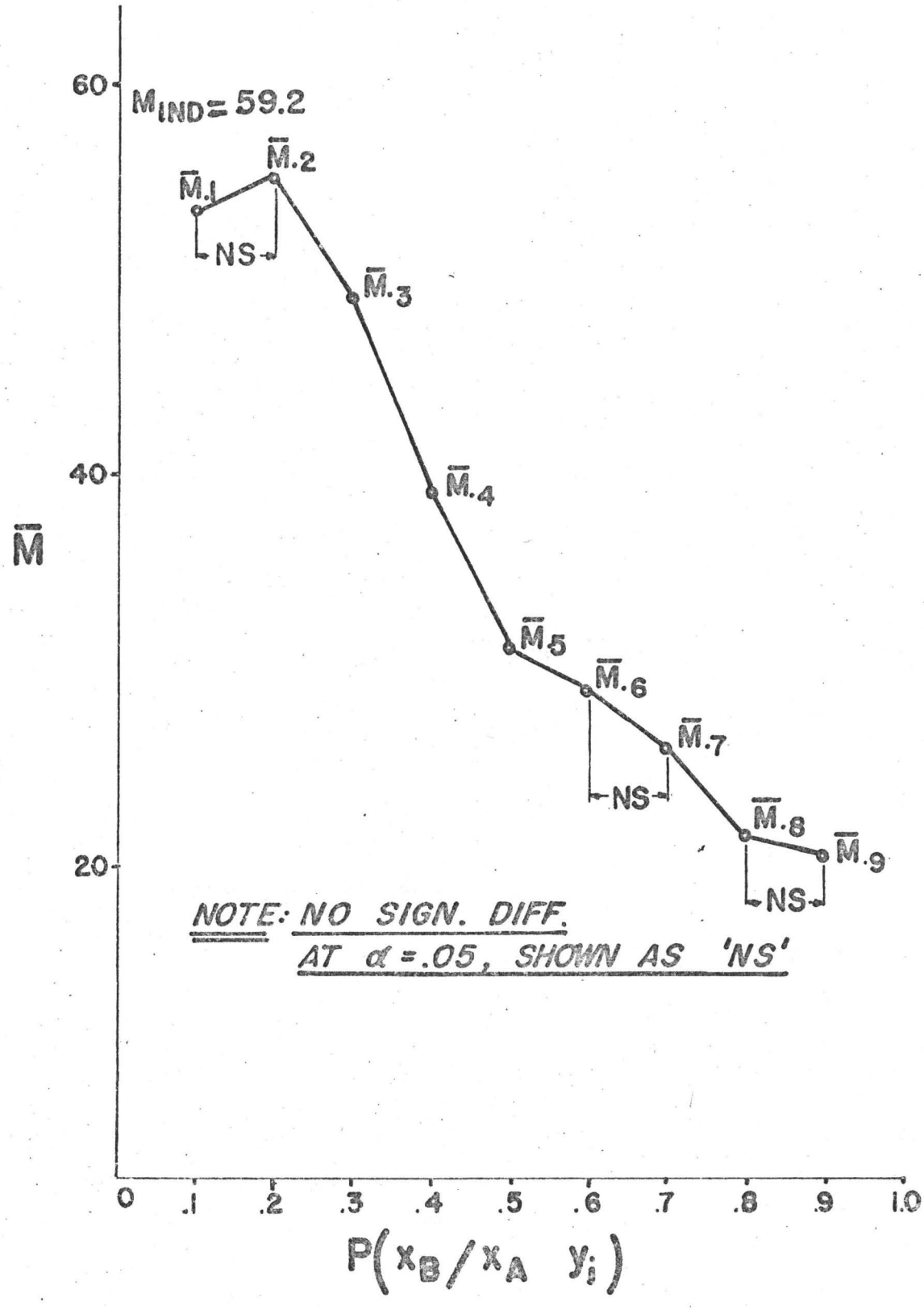
Source	DF	SS	MS	F	$F_{CRIT}(\alpha=.05)$
Model	2	83040.081	41520.041		
Mean	1	73804.820	73804.820		
Model(CFM)	1	9235.261	9235.261		
Residual	48	744.919	15.519		
Lack of fit	8	555.719	69.465	14.686	2.180
Pure Error	40	189.200	4.730		
Total	50	83785.000	1675.700		
Mean	1	73804.820	73804.820		
Total(CFM)	49	9980.180	203.677		

$R^2 = .991$ $R^2(CFM) = .925$

Since F = 14.686 exceeds $F_{CRIT} = 2.180$, the hypothesis of model adequacy is rejected.

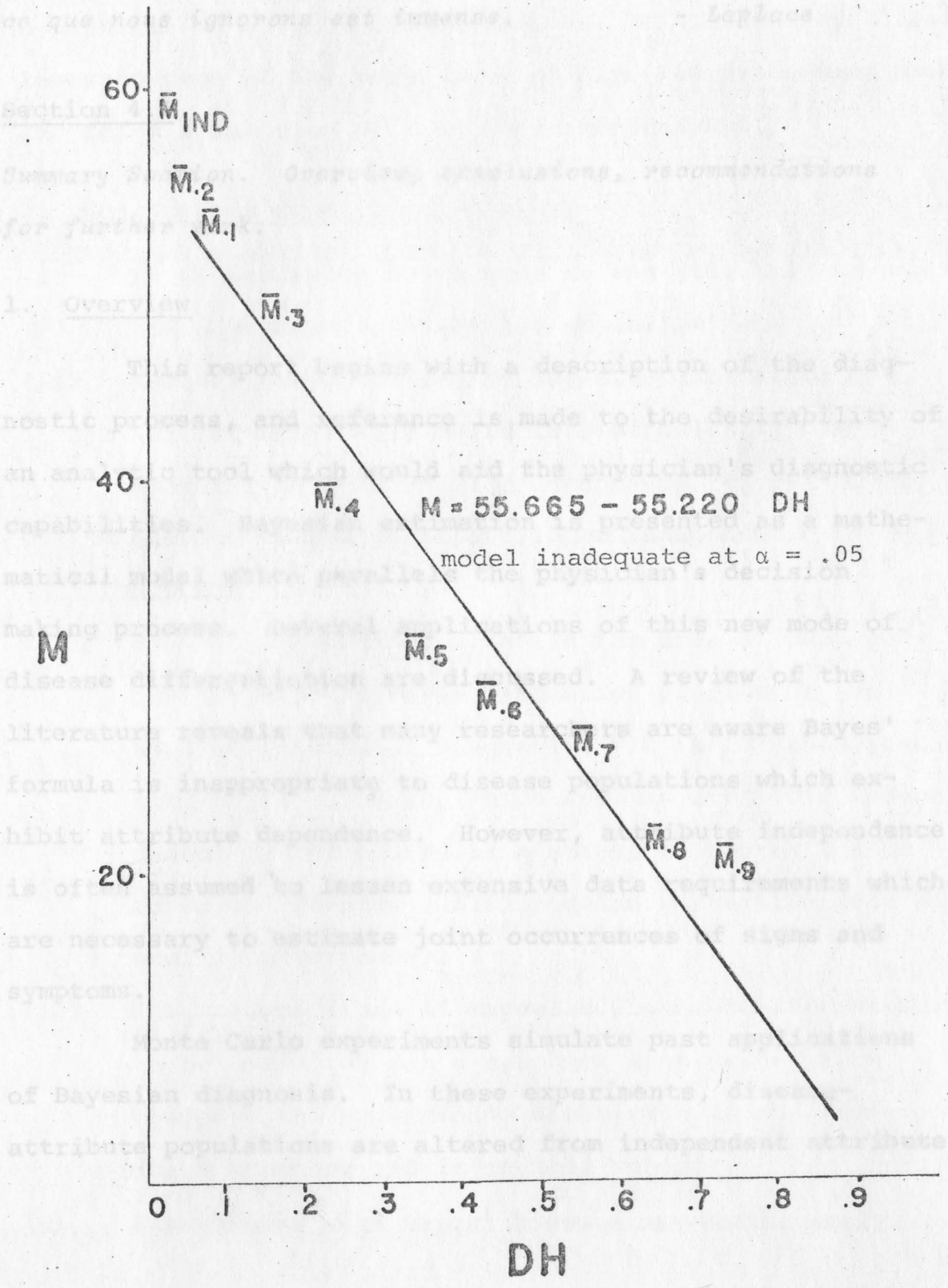
FIGURE 4.3.3, DATA SET 5

PLOT OF $\bar{M}$ vs $P(x_B/x_A, y_j)$



NOTE: NO SIGN. DIFF. AT $\alpha = .05$, SHOWN AS 'NS'

FIGURE 4.3.3a, DATA SET 5
M AS A LINEAR FUNCTION OF DH



Ce que nous connaissons et peu de chose,
ce que nous ignorons est immense. - *Laplace*

Section 4.4

Summary Section. Overview, conclusions, recommendations for further work.

1. Overview

This report begins with a description of the diagnostic process, and reference is made to the desirability of an analytic tool which would aid the physician's diagnostic capabilities. Bayesian estimation is presented as a mathematical model which parallels the physician's decision making process. Several applications of this new mode of disease differentiation are discussed. A review of the literature reveals that many researchers are aware Bayes' formula is inappropriate to disease populations which exhibit attribute dependence. However, attribute independence is often assumed to lessen extensive data requirements which are necessary to estimate joint occurrences of signs and symptoms.

Monte Carlo experiments simulate past applications of Bayesian diagnosis. In these experiments, disease-attribute populations are altered from independent attribute

structures to various levels of paired attribute dependence. Diagnoses are predicted according to the actual distribution of attributes among diseases, and also according to the false hypothesis of attribute independence. This strategy allows investigation of the robustness of Bayesian procedures under the false assumption of attribute independence.

2. Conclusions from the Literature

(i) It is desirable to develop an analytic tool to aid the physician's diagnostic capabilities.

(ii) Medical personnel regard Bayesian estimation as (a) conceptionally valid (b) sufficiently non-esoteric to receive wide acceptance (c) relatively easy to initiate, if attribute independence is assumed.

(iii) A major problem may arise in the area of data collection when Bayesian estimation procedures are applied to medical diagnosis. It is for this reason that Bayes' formula (which assumes attribute independence) is introduced as a surrogate approximation to Bayes' Theorem (which requires extensive data on profile occurrence).

(iv) The present state of Bayesian diagnosis is considered to be useful as (a) a feedback instrument which sharpens the physician's diagnostic capabilities (b) a teaching aid in medical schools.

(v) Experiences with Bayes' formula are sufficiently

encouraging to warrant further investigation. Refinements to the present state of Bayesian diagnosis may lead to a valuable diagnostic aid.

3. Conclusions From Monte Carlo Experiments

(i) Monte Carlo experiments indicate that Bayesian accuracy (under the false hypothesis of attribute independence) decreases markedly as disease-attribute populations diverge from attribute independence.

(ii) The data from Monte Carlo experiments suggest that M, the number of correct Bayesian predictions under the false assumption of attribute independence, decreases linearly as DH increases, where DH is a parameter which measures deviations from attribute independence. A simple first order linear model of the form $M_{ij} = B_0+B_1 \cdot DH+\varepsilon_{ij}$ was fitted to the data provided by Monte Carlo experiments. It was found that the model does not consistantly explain the variation in M.

(iii) Since R^2 and R^2(CFM) are close to 1 for all three Data Sets, it is unlikely that increasing the number of terms in the proposed model (by adding $B_2 \cdot DH_{ij}^2, B_3 \cdot DH_{ij}^3$, etc) will provide an adequate explanation of the variation in M. It may be that the variation in the data itself (M values) precludes curve-fitting.

4. Recommendations for Further Work

If the goal of further research is to improve the present state of Bayesian diagnosis, then the following courses of action are recommended:

(i) researchers may identify attributes which are independent in each disease category for which posterior probabilities are required. Of these, only those attributes which help to distinguish among disease categories should be considered. This strategy is used by Nugent et al (1964).

(ii) researchers may identify small subsets of dependent attributes, and modify Bayes' formula to account for attribute dependence within these subsets. Such a procedure requires estimates of joint occurrences of signs and symptoms. If the number of attributes in such subsets are small, then data requirements are managable. For example, suppose a profile is defined by $S = \{x_1, x_2, \ldots x_k\}$, and it is found that x_1, x_2, x_3 are not independent of each other, but are independent of other attributes in S (for a specific disease y_j). Then $P(y_j|S)$ may be estimated by:

$$\hat{P}(y_j|S) = \frac{\hat{P}(x_1,x_2,x_3|y_j)\cdot\left[\prod_{i=4}^{k}\hat{P}(x_i|y_j)\right]\cdot\hat{P}(y_j)}{\sum_{\ell=1}^{K}\hat{P}(S|y_\ell)\cdot\hat{P}(y_\ell)} \qquad (4.4.1)$$

Of course, this principle may be extended to include more than one subset of dependent attributes, and the denominator in (4.4.1) allows for different subset groupings in each of the disease categories $y_1, y_2, \ldots y_K$.

Such a procedure is seen as a refinement to the present state of Bayesian diagnosis, and is within the practical limitations of data collection.

FOOTNOTES

1. An attribute quantitatively or qualitatively describes a sign, symptom, result of laboratory test, etc.
2. $y_j \rightarrow x_i$ is interpreted as: if y_j then always x_i.
3. Such a set of attributes is referred to as a "profile", and is denoted by the letter S in Section 1.3b.
4. Attribute Set 3 describes an exhaustive set of events for disease category y_3, but not for y_1 or y_2.
5. At The University Department of Surgery, The General Infirmary, Leeds, England.
6. Attributes x_2 (central obesity) and x_3 (generalized obesity) are mutually exclusive. Thus, the number of possible attribute pairs in the two groups of patients is $2({}^{11}C_2-1) = 108$.
7. Here S_i denotes a characteristic, not a profile. The change in notation is made to conform with the notation used by Lincoln and Parker (1967) in Table 3.3.6.
8. Subroutine GGU1 of The IMSL Library, Volume 1. (1975). International mathematics and statistical libraries Inc., Houston, Texas.
9. There are $2^k = 2^6 = 64$ possible profiles, where k is the number of characteristics considered, and each characteristic is described as present (x_i) or absent ($\bar{x}_i$). Hence, k is also the number of attributes appearing in an individual's profile.

10. $M_{.2}$ and $M_{.3}$ values correspond to DH = .140. M_{IND} and $M_{.1}$ values are not considered in Model (1), according to the inferences from $H_o(2)$ and $H_o(3)$. Hence the distinct points in the design space are DH = .140, .170, .243, ..., .643, with 10 observations at DH = .140, and 5 observations for the remaining DH values.

REFERENCES

(1) Anderson, J.A. and Boyle, J.A. (1968). Computer diagnosis: statistical aspects. Br. Med. Bull. 24 230-235.

(2) Anderson, V.L. and McLean, R.A. (1974). Design of experiments: a realistic approach. Marcel Dekker, Inc. N.Y. 9-12, 22-23.

(3) Bailey, N.T.J. (1964). Mathematics and computer science in biology and medicine: Probability methods of diagnosis based on small samples. Proceedings of the Medical Research Council. H.M.S.O. London, England. 103-110.

(4) Began, F. and Dhumeaux, D. (1971). The application of computer techniques to the laboratory diagnosis. Minn. Med. 54 101-105.

(5) Billewicz, W.Z., Chapman, R.S., Crooks, J., Day, M.E., Grossage, J., Wayne, E. and Young J.A. (1969). Statistical methods applied to diagnosis of hypothyroidism. Q.J. Med. 38 255-266.

(6) Boyle, J.A., Greig, W.R., Franklin, D.A., Harden, R. McG., Buchanan, W.W. and McGirr, E.M. (1966). Construction of a model for computer assisted diagnosis: application to the problem of non-toxic goitre. Q.J. Med. 35 565-587.

(7) Collen, M.F., Rubin, L., Neyman, J., Dantzig, G.B., Baer, R.M. and Sieglaub, A.B. (1964). Automated

multiphasic screening and diagnosis. Am. J. Public Health 54 741-750.

(8) Crooks, J., Murray, I.P.C., and Wayne, E.J. (1959). Statistical methods applied to the clinical diagnosis of thyrotoxicosis. Q.J. Med. 28 211-234.

(9) Croxton, F.E., Cowden, D.J. and Klein, S. (1967). Applied General Statistics. Prentice-Hall, Inc. Englewood Cliffs, N.J. 604-605.

(10) DeDombal, F.T., Horrocks, J.C., Walmsley, G. and Wilson, P.D. (1975). Computer aided diagnosis and decision making in the acute abdomen. J. R. Coll. Physicians Lond. 9 211-218.

(11) DeDombal, F.T., Leaper, D.J., Staniland, J.R., McCann, A.P. and Horrocks, J.C. (1972). Computer aided diagnosis of acute abdomen pain. Br. Med. J. 2 9-13.

(12) Draper, N.R., and Smith, H. Jr. (1966). Applied Regression Analysis. J. Wiley and Sons, N.Y.

(13) Feinstein, A.R. (1963). Boolean algebra and clinical taxonomy: 1. Analytic synthesis of the general spectrum of human disease. N. Engl. J. Med. 269 929-938.

(14) Fitzgerald, L.T., Overall, J.E. and Williams, C.M. (1966). A computer program for diagnosis of thyroid disease. Am. J. Roentgenol. Radium Ther. Nucl. Med. 97 901-905.

(15) Gustafson, D.H., Kestly, J.J., Greist, J.H., and Jenson, N.M.

(1971). Initial evaluation of a subjective Bayesian diagnostic system. Health Serv. Res. 6 204-213.

(16) Gustafson, J.E. and Throckmorton, T.D. (1965). The computer as consultant. Medical Times 93 115-119.

(17) Hall, G.H. (1967). The clinical application of Bayes' theorem. Lancet 2 555-557.

(18) Hirschfeld, R.M.A. and Judge, R.D. (1965). Clinician vs computer: a study of the application of Bayes' formula to clinical diagnosis. Univ. Mich. Med. Center. J. 35 32-35.

(19) Horrocks, J.C., McCann, A.P., Staniland, J.R., Leaper, D.J. and DeDombal, F.T. (1972). Computer aided diagnosis: description of an adaptable system, and operational experience with 2034 cases. Br. Med. J. 2 5-9.

(20) Kenrick, M.A. (1974). Viral hepatitus in American missionaries abroad. J. Infect. Dis. 129 227-229.

(21) Knill-Jones, R.P. (1975). The diagnosis of jaundice by the computation of probabilities. J.R. Coll. Physicians Lond. 9 205-210.

(22) Knill-Jones, R.P., Stern, R.B., Girmes, D.H., Maxwell, J.D., Thompson, R.P.H. and Williams, R. (1973). The use of sequential Bayesian model in diagnoses of jaundice by computer. Br. Med J. 1 530-533.

(23) Ledley, R.S. and Lusted, L.B. (1951). Reasoning foundations of medical diagnosis. Science 130 9-21.

(24) Ledley, R.S. and Lusted, L.B. (1960). The use of electronic computers to aid in medical diagnosis. Proc. of the I.R.E. Transactions on Medical Electronics ME-7 31 1970-1977.

(25) Lincoln., T.L. and Parker, R.D. (1967). Medical diagnosis using Bayes' theorem. Health Serv. Res. 2 34-35.

(26) Lipkin, M. and Hardy, J.D. (1958). Mechanical correlation of data in differential diagnosis of hematological diseases. J.A.M.A. 166 113-125.

(27) Lodwich, G.S. (1965). A probabilistic to the diagnosis of bone tumors. Radiol. Clin. North Am. 3 487-497.

(28) MacMahon, B. and Pugh, T.F. (1970). Epidemiology - principles and methods. Little, Brown and Co. Boston.

(29) Nash, F.A. (1954). Differential diagnosis: an apparatus to assist the logical faculties. Lancet 1 874-875.

(30) Nugent, C.A., Warner, H.R., Dunn, J.T., and Tyler, F.H. (1964). Probability theory in the diagnosis of Cushing's Syndrome. J. Clin. Endocrinol. Metab. 24 621-627.

(31) Overall, J.E. and Williams, C.M. (1961). Models for medical diagnosis. Behav. Sci. 6 134-141.

(32) Overall, J.E. and Williams, C.M. (1963). Conditional probability program for diagnosis of thyroid function. J.A.M.A. 183 307-313.

(33) Payne, L.C. (1963). Medical automation. World Medical Electronics 2 7-11.

(34) Payne, L.C. (1964). The role of the computer in refining diagnosis. Lancet 2 32-35.

(35) Reale, A., Maccacaro, G.A., Rocca, E., D'Intino, S., Geoffre, P.M., Vestri, A. and Matolese, M. (1968). Computer diagnosis of congenital heart disease. Comput. Biomed. Res. 1 533-549.

(36) Rinaldo, J.A. Jr., Scheinok, P.A. and Rupe, C.E. (1963). Symptom diagnosis: mathematical analysis of epigastric pain. Ann. Intern. Med. 59 145-154.

(37) Rodenwalt, T. (1952). World atlas of epidemic diseases. Vollmer and Bentlin Gmph. Hamburg.

(38) Scheinok, P. and Rinaldo, J.A. Jr. (1967). Symptom diagnosis: optimal subsets for upper abdominal pain. Comp. Biomed. Res. 1 221-236.

(39) Slack, W.V., Hicks, G.P., Reed, C.E. and VanCura, L.J. (1966). A computer based medical history system. N. Engl. J. Med. 274 194-198.

(40) Teather, D. (1974). Diagnosis-methods and analysis. J. Inst. Math. Appl. Jan/Feb 37-41.

(41) Teather, D. and Hilder, W. (1975). The analysis of diagnostic data. J.R. Coll. Physicians Lond. 9 219-225.

(42) Templeton, A.W., Lehr, J.L. and Simmons, C. (1966). The computer evaluation and diagnosis of congenital heart disease, using roentgenographic findings. Radiology 87 658-670.

(43) Van Zwanenberg, D.F. (1974). The geography of disease in East Anglia. J.R. Coll. Physicians Lond. 8 145-153.

(44) Warner, H.R., Toronto, A.F. and Veasey, L.G. (1964). Experience with Bayes' theorem for computer analysis of congenital heart diseases. Ann. N.Y. Acad. Sci. 115 558-567.

(45) Warner, H.R., Toronto, A.F., Veasey, L.G. and Stephenson, R. (1961). A mathematical approach to medical diagnosis - application to congenital heart disease. J.A.M.A. 177 177-183.

(46) Wartak, J. (1974). Mathematical model for medical diagnosis. Comput. Biol. Med. 4 79-84.

(47) Winkler, C., Reichertz, P. and Kloss, G. (1967). Computer diagnosis of thyroid diseases: comparison of incidence data and considerations on the problem of data collection. Am. J. Med. Sci. 253 27-33.

APPENDIX A

(references by topic)

ABDOMEN

DeDombal, F.T., Horrocks, J.C., Walmsley, G. and Wilson, P.D. (1975). Computer aided diagnosis and decision making in the acute abdomen. J.R. Coll. Physicians Lond. 9 211-218.

DeDombal, F.T., Leaper, D.J., Staniland, J.R., McCann, A.P. and Horrocks, J.C. (1972). Computer aided diagnosis of acute abdomen pain. Br. Med. J. 2 9-13.

Horrocks, J.C., McCann. A.P., Staniland, J.R., Leaper, D.J. and DeDombal, F.T. (1972). Computer aided diagnosis: description of an adaptable system, and operational experience with 2034 cases. Br. Med. J. 2 5-9.

Rinaldo, J.A. Jr., Scheinok, P.A. and Rupe, C.E. (1963). Symptom diagnosis: mathematical analysis of epigastric pain. Ann. Intern. Med. 59 145-154.

Scheinok, P. and Rinaldo, J.A. Jr. (1967). Symptom diagnosis: optimal subsets for upper abdominal pain. Comp. Biomed. Res. 1 221-236.

BONE CANCER

Lodwich, G.S. (1965). A probabilistic to the diagnosis of bone tumors. Radiol. Clin. North Am. 3 487-497.

Lodwich, G.S., Cosmo, L.H., Walton, E.S., Keller, R.F. and

Robertson, E.D. (1963). Computer diagnosis of primary bone tumors: a preliminary report. Radiology 80 273-275.

HEART

Bruce, A.B. and Yarnell, S.R. (1966). Computer-aided diagnosis of cardiovascular disorders. J. Chron. Dis. 19 473-484.

Bruce, R.A., Portman, R.M., Blachman, J.R., Lamppert, R.V., Hofer, V., McGill, R., Mickey, M. and Harris, R. (1963). Computer diagnosis of heart disease. Fifth I.B.M. Med. Symp. 77-97.

Hirschfeld, R.M.A. and Judge, R.D. (1965). Clinician vs computer: a study of the application of Bayes' formula to clinical diagnosis. Univ. Mich. Med. Center. J. 35 32-35.

Reale, A., Maccacaro, G.A., Rocca, E., D'Intino, S., Geoffre, P.M., Vestri, A. and Matolese, M. (1968). Computer diagnosis of congenital heart disease. Comput. Biomed. Res. 1 533-549.

Templeton, A.W., Lehr, J.L. and Simmons, C. (1966). The computer evaluation and diagnosis of congenital heart disease, using roentgenographic findings. Radiology 87 658-670.

Warner, H.R., Toronto, A.F. and Veasey, L.G. (1964). Experience with Bayes' theorem for computer analysis of congenital heart diseases. Ann. N.Y. Acad. Sci.

115 558-567.

Warner, H.R., Toronto, A.F., Veasey, L.G. and Stephenson, R. (1961). A mathematical approach to medical diagnosis - application to congenital heart disease. J.A.M.A. 177 177-183.

LIVER

Began, F. and Dhumeaux, D. (1971). The application of computer techniques to the laboratory diagnosis. Minn. Med. 54 101-105.

Cattaneo, A.D., Lucchelli, P.E., Rocca, E., Mattioli, F. and Giuseppe, B. (1972). Computer versus clinical diagnosis of biliary tract diseases. Abdominal Surgery 14 71-75.

Knill-Jones, R.P. (1975). The diagnosis of jaundice by the computation of probabilities. J.R. Coll. Physicians Lond. 9 205-210.

Knill-Jones, R.P., Stern, R.B., Girmes, D.H., Maxwell, J.D., Thompson, R.P.H. and Williams, R. (1973). The use of sequential Bayesian model in diagnoses of jaundice by computer. Br. Med. J. 1 530-533.

Lincoln, T.L. and Parker, R.D. (1967). Medical diagnosis using Bayes' theorem. Health Serv. Res. 2 34-35.

Reale, A., Maccacaro, G.A., Rocca, E., D'Intino, S., Geoffre, P.M., Vestri, A. and Matolese, M. (1968). Computer diagnosis of congenital heart disease. Comput. Biomed. Res. 1 533-549.

Teather, D. (1974). Diagnosis - methods and analysis. J. Inst. Math. Appl. Jan/Feb 37-41.

Teather, D. and Hilder, W. (1975). The analysis of diagnostic data. J.R. Coll. Physicians Lond. 9 219-225.

THYROID

Billewicz, W.Z., Chapman, R.S., Crooks, J., Day, M.E., Grossage, J., Wayne, E. and Young, J.A. (1969). Statistical methods applied to diagnosis of hypothyroidism. Q.J. Med. 38 255-266.

Boyle, J.A., Greig, W.R., Franklin, D.A., Harden, R. McG., Buchanan, W.W. and McGirr, E.M. (1966). Construction of a model for computer assisted diagnosis: application to the problem of non-toxic goitre. Q.J. Med. 35 565-587.

Crooks, J., Murray, I.P.C., and Wayne, E.J. (1959). Statistical methods applied to the clinical diagnosis of thyrotoxicosis. Q.J. Med. 28 211-234.

Fitzgerald, L.T., Overall, J.E. and Williams, C.M. (1966). A computer program for diagnosis of thyroid disease. Am. J. Roentgenol. Radium Ther. Nucl. Med. 97 901-905.

Gustafson, D.H., Kestly, J.J., Greist, J.H., and Jenson, N.M. (1971). Initial evaluation of a subjective Bayesian diagnostic system. Health Serv. Res. 6 204-213.

Overall, J.E. and Williams, C.M. (1961). Models for medical diagnosis. Behav. Sci. 6 134-141.

Overall, J.E. and Williams, C.M. (1963). Conditional

probability program for diagnosis of thyroid function. J.A.M.A. 183 307-313.

Winkler, C., Reichertz, P. and Kloss, G. (1967). Computer diagnosis of thyroid diseases: comparison of incidence data and considerations on the problem of data collection. Am. J. Med. Sci. 253 27-33.

THEORETICAL TREATMENT OF NON-INDEPENDENCE OF ATTRIBUTES

Norusis, M.J. and Jacquez, J.A. (1975). Diagnosis. 1. Symptom non-independence in mathematical models for diagnosis. Comput. Biomed. Res. 8 156-172.

Norusis, M.J. and Jacquez, J.A. (1975). Diagnosis. 2. Diagnostic models based on attribute clusters: a proposal and comparisons. Comput. Biomed. Res. 8 173-188.

Vanderplas, J.M. (1967). A method for determining probabilities for correct use of Bayes' Theorem in medical diagnosis. Comput. Biomed. Res. 1 215-220.