

# Chapter One

## Introduction to Biostatistics

### 1.2 What is the Field of Statistics?

Much of the joy and grief in life arises in situations that involve considerable uncertainty. Here are a few such situations:

1. Parents of a child with a genetic defect consider whether or not they should have another child. They will base their decision on the chance that the next child will have the same defect.
2. To choose the best therapy, a physician must compare the *prognosis*, or future course, of a patient under several therapies. A therapy may be a success, a failure, or somewhere in between; the evaluation of the chance of each occurrence necessarily enters into the decision.
3. In an experiment to investigate whether a food additive is *carcinogenic* (i.e., causes), the U.S. Food and Drug Administration has animals treated with and without the additive. Often, cancer will develop in both the treated and untreated groups of animals. In both groups there will be animals that do not develop cancer. There is a need for some method of determining whether the group treated with the additive has “too much” cancer.
4. In designing and planning medical care facilities, planners take into account differing needs for medical care. Needs change because there are new modes of therapy, as well as demographic shifts, that may increase or decrease the need for facilities. All of the uncertainty associated with the future health of a population and its future geographic and demographic patterns should be taken into account.

## 1.3 Why Biostatistics?

*Biostatistics* is the study of statistics as applied to biological areas. Biological laboratory experiments, medical research (including clinical research), and health services research all use statistical methods. Many other biological disciplines rely on statistical methodology.

There are primarily **three reasons** for focusing on biostatistics:

1. Some statistical methods are used more heavily in biostatistics than in other fields. For example, a general statistical textbook would not discuss the life-table method of analyzing survival data—of importance in many biostatistical applications.
2. Examples are drawn from the biological, medical, and health care areas; this helps you to maintain motivation. It also helps you understand how to apply statistical methods.
3. A biostatistical text is to teach the material to an audience of health professionals. In this case, the interaction between students and teacher, but especially among the students themselves, is of great value in learning and applying the subject matter.

## 1.4 Goals of this Course

1. Students will understand specified statistical concepts and procedures.
2. Students will be able to identify procedures appropriate (and inappropriate) to a given situation. They should also have the knowledge to recognize even they do not know an appropriate technique.
3. Students will be able to carry out appropriate specified statistical procedures.
4. Will be able to make inference about the unknown characteristic of the population.

## 1.5 Statistical Problems in Biomedical Research

There are several examples of situations in which biostatistical design and analysis have been of use. Here, we will discuss two of them.

### 1.5.3 Example 3: Use of Laboratory Tests and the Relation to Quality of Care.

An important feature of medical care are laboratory tests. These tests affect both the quality and the cost of care. The frequency with which such tests are ordered varies with the physician. It is not clear how the frequency of such tests influences the quality of medical care. Given the expense of laboratory tests and limited time and resources, there clearly is a need for evaluation of the relationship between the use of laboratory tests and the quality of care.

The study discussed here consisted of 21 physicians serving medical internships as reported by Schroeder et al. [1974]. The interns were ranked independently on overall clinical capability (i.e., quality of care) by five faculty internists who had interacted with them during their medical training. Only patients admitted with uncomplicated acute myocardial infarction or uncomplicated chest pain were considered for the study. "Medical records of all patients hospitalized on the coronary care unit between July 1, 1971 and June 20, 1972, were analyzed and all patients meeting the eligibility criteria were included in the study. . . ." The frequency of laboratory utilization ordered during the first three days of hospitalization was translated into cost. Mean costs of laboratory use were calculated for each intern's subset of patients, and the interns were ranked in order of increasing costs on a per-patient basis.

Ranking by the five faculty internists and by cost are given in Table 1.1. There is considerable variation in the evaluations of the five internists. This table still does not clearly answer the question of whether there is a relationship between clinical competence and the frequency of use of laboratory tests and their cost. Figure 1.1 shows the relationship between cost and one measure of clinical competence; on the basis of this graph and

some statistical calculations, the authors conclude that “at least in the setting measured, no overall correlation existed between cost of medical care and competence of medical care.”

**Table 1.1 Independent Assessment of Clinical Competence of 21 Medical Interns by Five Faculty Internists and Ranking of Cost of Laboratory Procedures Ordered, George Washington University Hospital, 1971–1972**

Intern	Clinical Competence <sup>a</sup>						Rank	Rank of Costs of Procedures Ordered <sup>b</sup>
	I	II	III	IV	V	Total		
A	1	2	1	2	1	7	1	10
B	2	6	2	1	2	13	2	5
C	5	4	11	5	3	28	3	7
D	4	5	3	12	7	31	4	8
E	3	9	8	9	8	37	5	16
F	13	11	7	3	5	39	7	9
G	7	12	5	4	11	39	7	13
H	11	3	9	10	6	39	7	18
I	9	15	6	8	4	42	9	12
J	16	8	4	7	14	49	10	1
K	17	1	17	11	9	55	11	20
L	6	7	21	16	10	60	12	19
M	8	20	14	6	17	65	13	21
N	18	10	13	13	13	67	14	14
O	12	14	12	18	15	71	15	17
P	19	13	10	17	16	75	16	11
Q	20	16	16	15	12	77	17	4
R	14	18	19	14	19	84	18	15
S	10	19	18	20	20	87	19	3
T	15	17	20	21	21	94	20.5	2
U	21	21	15	19	18	94	20.5	5

Source: Data from Schroeder et al. [1974]; by permission of Medical Care.

<sup>a</sup>1 = most competent.

<sup>b</sup>1 = least expensive.

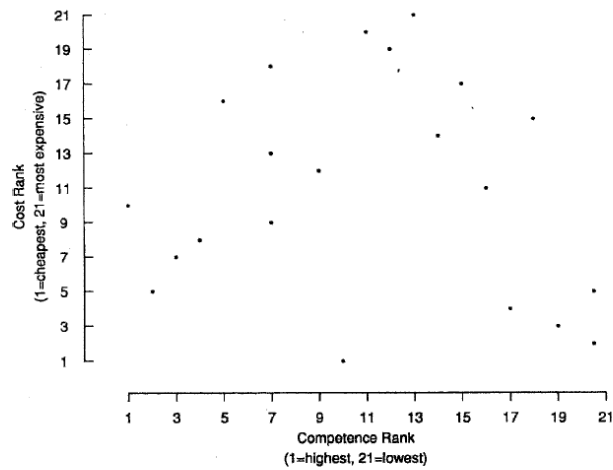


Figure 1.1 Rank order of clinical competence vs. rank order of cost of laboratory tests orders for 21 interns, George Washington University Hospital, 1971–1972. (Data from Schroeder et al. [1974].)

This study contains good examples of the types of (basically statistical) problems facing a researcher in the health administration area. For examples,

- What is the population of interest? In other words, what population do the 21 interns represent?
- There are difficult measurement problems: Is level of clinical competence, as evaluated by an internist, equivalent to the level of quality of care?
- How reliable are the internists? The variation in their assessments has already been noted.
- Is cost of laboratory use synonymous with cost of medical care as the authors seem to imply in their conclusion?

#### **1.5.4 Example 4: Internal Mammary Artery Ligation**

One of the greatest health problems in the world is coronary artery disease. The coronary arteries are the arteries around the outside of the heart. These arteries bring blood to the heart muscle (myocardium). Coronary artery disease brings a narrowing of the coronary arteries. As the coronary arteries narrow, the body often compensates by building *collateral circulation*, circulation that involves branches from existing coronary arteries that develop to bring blood to an area of restricted blood flow. The internal mammary arteries are arteries that bring blood to the chest. The tributaries of the internal mammary arteries develop collateral circulation to the coronary arteries. It was thus reasoned that by tying off, or *ligating*, the internal mammary arteries, a larger blood supply would be forced to the heart. An operation, internal mammary artery ligation, was developed to implement this procedure.

Early results of the operation were most promising. Battezzati et al. [1959] reported on 304 patients who underwent internal mammary artery ligation: 94.8% of the patients reported improvement; 4.9% reported no appreciable

change [Ratcliff, 1957; Time, 1959]. Still, the possibility remained that the improvement resulted from a placebo effect. A *placebo effect* is a change, or perceived change, resulting from the psychological benefits of having undergone treatment.

Two studies of internal mammary artery ligation were performed using a sham operation as a control. Both studies were *double blind* : Neither the patients nor physicians evaluating the effect of surgery knew whether the ligation had taken place. In each study, incisions were made in the patient's chest and the internal mammary arteries exposed. In the sham operation, nothing further was done. For the other patients, the arteries were ligated. Both studies selected patients having the ligation or sham operation by random assignment [Hitchcock et al., 1966; Ruffin et al., 1969].

Cobb et al. [1959] reported on the subjective patient estimates of "significant" improvement. Patients were asked to estimate the percent improvement after the surgery. Another indication of the amount of pain experienced is the number of nitroglycerin tablets taken for anginal pain. Table 1.2 reports these data.

Dimond et al. [1960] reported a study of 18 patients, of whom five received the sham operation and 13 received surgery. Table 1.3 presents the patients' opinion of the percentage benefit of surgery.

Both papers conclude that it is unlikely that the internal mammary artery ligation has benefit, beyond the placebo effect, in the treatment of coronary artery disease. Note that 12 of the 14, or 86%, of those receiving the sham operation reported improvement in the two studies. These studies point to the need for appropriate comparison groups when making scientific inferences.

**Table 1.2 Subjective Improvement as Measured by Patient Reporting and Number of Nitroglycerin Tablets**

	Ligated	Nonligated
Number of patients	8	9
Average percent improvement reported	32	43
Subjects reporting 40% or more improvement	5	5
Subjects reporting no improvement	3	2
Nitroglycerin tablets taken		
Average before operation (no./week)	43	30
Average after operation (no./week)	25	17
Average percent decrease (no./week)	34	43

Source: Cobb et al. [1959].

**Table 1.3 Patients' Opinions of Surgical Benefit**

Patients' Opinions of the Benefit of Surgery	Patient Number <sup>a</sup>
Cured (90–100%)	4, 10, 11, 12*, 14*
Definite benefit (50–90%)	2, 3*, 6, 8, 9*, 13*, 15, 17, 18
Improved but disappointed (25–50%)	7
Improved for two weeks, now same or worse	1, 5, 16

Source: Dimond et al. [1960].

<sup>a</sup>The numbers 1–18 refer to the individual patients as they occurred in the series, grouped according to their own evaluation of their benefit, expressed as a percentage. Those numbers followed by an asterisk indicate a patient on whom a sham operation was performed.

The use of clinical trials has greatly enhanced medical progress. Examples are given throughout the book, but this is not the primary emphasis of the text. Good references for learning much about clinical trials are Meinert [1986], Friedman et al. [1981], Tanur et al. [1989], and Fleiss [1986] among others.

## *Some Definitions of Statistics*

- The science of statistics is essentially a branch of Applied Mathematics, and may be regarded as mathematics applied to observational data. . . . Statistics may be regarded (i) as the study of populations, (ii) as the study of variation, (iii) as the study of methods of data reduction. [Fisher, 1950]
- Statistics is the branch of the scientific method which deals with the data obtained by counting or measuring the properties of populations of natural phenomena. [Kendall and Stuart, 1963]
- The science and art of dealing with variation in such a way as to obtain reliable results. [Mainland, 1963]
- Statistics is concerned with the inferential process, in particular with the planning and analysis of experiments or surveys, with the nature of observational errors and sources of variability that obscure underlying patterns, and with the efficient summarizing of sets of data. [Kruskal, 1968]
- Statistics = Uncertainty and Behavior. [Savage, 1968]
- . . . . . the principal object of statistics [is] to make inference on the probability of events from their observed frequencies. [von Mises, 1957]
- The technology of the scientific method. [Mood, 1950]
- The statement, still frequently made, that statistics is a branch of mathematics is no more true than would be a similar claim in respect of engineering . . . [G]ood statistical practice is equally demanding of appreciation of factors outside the formal mathematical structure, essential though that structure is. [Finney, 1975]

Collecting data (random sample) from a well-defined population, analyzing data with an appropriate statistical method, and making inference about the characteristics of the population is called *Statistics* . [KG, Jan 11, 2016]