UNCLASSIFIED

AD NUMBER
AD865422
LIMITATION CHANGES
TO: **Approved for public release; distribution is unlimited.**
FROM: **Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; DEC 1969. Other requests shall be referred to Army Materiel Command, Alexandria, VA.**
AUTHORITY
USAMC ltr, 14 Jan 1972

THIS PAGE IS UNCLASSIFIED

AMCP-706-111-62

ORDP 20-111

ORDNANCE CORPS PAMPHLET

ORDNANCE ENGINEERING DESIGN HANDBOOK

EXPERIMENTAL STATISTICS

SECTION 2

ANALYSIS OF ENUMERATIVE AND CLASSIFICATORY DATA

RECEIVED
14 FEB 1963
REDSTONE SCIENTIFIC INFORMATION CENTER

FOR REFERENCE ONLY

LOAN COPY ONLY
DO NOT DESTROY
PROPERTY OF
REDSTONE SCIENTIFIC
INFORMATION CENTER

REDSTONE SCIENTIFIC INFORMATION CENTER
5 0510 00171641 1

ORDNANCE CORPS

JUNE 1962

[redacted]

ORDNANCE ENGINEERING DESIGN HANDBOOK SERIES

The Ordnance Engineering Design Handbook Series, when completed, is intended to provide a compilation of principles and fundamental data to supplement experience in assisting engineers in the evolution of new designs which will meet tactical and technical needs while also embodying satisfactory producibility and maintainability.

As of the date of this publication, the handbooks listed below have been published or publication is pending. Copies of these handbooks, and information on later publications which may be available, should be requested from the Ordnance Engineering Handbook Office, Ordnance Liaison Group, Durham, Box CM, Duke Station, Durham, N. C.

AMMUNITION SERIES

ORDP 20-210	Fuzes, General and Mechanical
ORDP 20-244	Section 1, Artillery Ammunition--General, with Table of Contents, Glossary and Index for Series
ORDP 20-245 (C)	Section 2, Design for Terminal Effects (U)
ORDP 20-246	Section 3, Design for Control of Flight Characteristics
ORDP 20-247 (C)	Section 4, Design for Projection (U)
ORDP 20-248	Section 5, Inspection Aspects of Artillery Ammunition Design
ORDP 20-249 (C)	Section 6, Manufacture of Metallic Components of Artillery Ammunition (U)

BALLISTIC MISSILE SERIES

ORDP 20-281 (S)	Weapon System Effectiveness (U)
ORDP 20-282	Propulsion and Propellants
ORDP 20-284 (C)	Trajectories (U)
ORDP 20-286	Structures

CARRIAGES AND MOUNTS SERIES

ORDP 20-341	Cradles
ORDP 20-342	Recoil Systems
ORDP 20-343	Top Carriages
ORDP 20-344	Bottom Carriages
ORDP 20-345	Equilibrators
ORDP 20-346	Elevating Mechanisms
ORDP 20-347	Traversing Mechanisms

EXPLOSIVES SERIES

ORDP 20-175	Solid Propellants, Part One
ORDP 20-176 (C)	Solid Propellants, Part Two (U)
ORDP 20-177	Properties of Explosives of Military Interest, Section 1
ORDP 20-178 (C)	Properties of Explosives of Military Interest, Section 2 (U)

EXTERIOR BALLISTICS SERIES

ORDP 20-140	Trajectories, Differential Effects, and Data for Projectiles

FIRE CONTROL SERIES

ORDP 20-331	Compensating Elements

GENERAL

ORDP 20-106	Elements of Armament Engineering, Part 1, Sources of Energy
ORDP 20-107	Elements of Armament Engineering, Part 2, Ballistics
ORDP 20-108	Elements of Armament Engineering, Part 3, Weapon Systems and Components

GENERAL (continued)

ORDP 20-110	Experimental Statistics, Section 1, Basic Concepts and Analysis of Measurement Data
ORDP 20-111	Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data
ORDP 20-112	Experimental Statistics, Section 3, Planning and Analysis of Comparative Experiments
ORDP 20-113	Experimental Statistics, Section 4, Special Topics
ORDP 20-114	Experimental Statistics, Section 5, Tables
ORDP 20-134	Maintenance Engineering Guide for Ordnance Design
ORDP 20-135	Inventions, Patents, and Related Matters
ORDP 20-136	Servomechanisms, Section 1, Theory
ORDP 20-137	Servomechanisms, Section 2, Measurement and Signal Converters
ORDP 20-138	Servomechanisms, Section 3, Amplification
ORDP 20-139	Servomechanisms, Section 4, Power Elements and System Design
ORDP 20-170 (C)	Armor and Its Application to Vehicles (U)
ORDP 20-270	Propellant Actuated Devices
ORDP 20-290 (C)	Warheads--General (U)
ORDP 20-355	The Automotive Assembly

ORDNANCE MATERIALS HANDBOOKS

ORDP 20-301	Aluminum and Aluminum Alloys
ORDP 20-302	Copper and Copper Alloys
ORDP 20-303	Magnesium and Magnesium Alloys
ORDP 20-305	Titanium and Titanium Alloys
ORDP 20-306	Adhesives
ORDP 20-307	Gasket Materials (Nonmetallic)
ORDP 20-308	Glass
ORDP 20-309	Plastics
ORDP 20-310	Rubber and Rubber-Like Materials
ORDP 20-311	Corrosion and Corrosion Protection of Metals

SURFACE-TO-AIR MISSILE SERIES

ORDP 20-291	Part One, System Integration
ORDP 20-292	Part Two, Weapon Control
ORDP 20-293	Part Three, Computers
ORDP 20-294 (S)	Part Four, Missile Armament (U)
ORDP 20-295 (S)	Part Five, Countermeasures (U)
ORDP 20-296	Part Six, Structures and Power Sources
ORDP 20-297 (S)	Part Seven, Sample Problem (U)

OFFICE OF THE CHIEF OF ORDNANCE
Washington 25, D. C., 29 June 1962

ORDP 20-111, Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data, forming part of the Ordnance Engineering Design Handbook Series, is published for the information and guidance of all concerned.

OFFICIAL:
ALFRED A. ROBINSON
Colonel, Ord Corps
Executive Officer

H. F. BIGELOW
Major General
Chief of Ordnance

OFFICE OF CHIEF OF ORDNANCE
OFFICIAL COPY
DEPARTMENT OF THE ARMY

DISTRIBUTION: Special

ORDNANCE ENGINEERING DESIGN HANDBOOK

EXPERIMENTAL STATISTICS

MARY G. NATRELLA
Statistical Engineering Laboratory
National Bureau of Standards

SECTION 2, STANDARD TECHNIQUES FOR ANALYSIS AND INTERPRETATION OF ENUMERATIVE AND CLASSIFICATORY DATA

CONSISTING OF
CHAPTERS 7 THROUGH 10

FOREWORD

INTRODUCTION

This is one of a group of handbooks covering the engineering information and quantitative data needed in the design, development, construction, and test of ordnance equipment which (as a group) constitute the Ordnance Engineering Design Handbook.

PURPOSE OF HANDBOOK

The Handbook on Experimental Statistics has been prepared as an aid to scientists and engineers engaged in Army Ordnance research and development programs, and especially as a guide and ready reference for military and civilian personnel who have responsibility for the planning and interpretation of experiments and tests relating to the performance of Army Ordnance equipment in the design and developmental stages of production.

SCOPE AND USE OF HANDBOOK

This Handbook is a collection of statistical procedures useful in ordnance applications. It is presented in five sections, viz:

ORDP 20-110, Section 1, Basic Concepts and Analysis of Measurement Data (Chapters 1-6)

ORDP 20-111, Section 2, Analysis of Enumerative and Classificatory Data (Chapters 7-10)

ORDP 20-112, Section 3, Planning and Analysis of Comparative Experiments (Chapters 11-14)

ORDP 20-113, Section 4, Special Topics (Chapters 15-23)

ORDP 20-114, Section 5, Tables

Section 1 provides an elementary introduction to basic statistical concepts and furnishes full details on standard statistical techniques for the analysis and interpretation of measurement data. Section 2 provides detailed procedures for the analysis and interpretation of enumerative and classificatory data. Section 3 has to do with the planning and analysis of comparative experiments. Section 4 is devoted to consideration and exemplification of a number of important but as yet non-standard statistical techniques, and to discussion of various other special topics. An index for the material in all four sections is placed at the end of Section 4. Section 5 contains all the mathematical tables needed for application of the procedures given in Sections 1 through 4.

An understanding of a few basic statistical concepts, as given in Chapter 1, is necessary; otherwise each of the first four sections is largely independent of the others. Each procedure, test, and technique described is illustrated by means of a worked example. A list of authoritative references is included, where appropriate, at the end of each chapter. Step-by-step instructions are given for attaining a stated goal, and the conditions under which a particular procedure is strictly valid are stated explicitly. An attempt is made to indicate the extent to which results obtained by a given procedure are valid to a good approximation when these conditions are not fully met. Alternative procedures are given for handling cases where the more standard procedures cannot be trusted to yield reliable results.

The Handbook is intended for the user with an engineering background who, although he has an occasional need for statistical techniques, does not have the time or inclination to become an expert on statistical theory and methodology.

The Handbook has been written with three types of users in mind. The first is the person who has had a course or two in statistics, and who may even have had some practical experience in applying statistical methods in the past, but who does not have statistical

ideas and techniques at his fingertips. For him, the Handbook will provide a ready reference source of once familiar ideas and techniques. The second is the person who feels, or has been advised, that some particular problem can be solved by means of fairly simple statistical techniques, and is in need of a book that will enable him to obtain the solution to his problem with a minimum of outside assistance. The Handbook should enable such a person to become familiar with the statistical ideas, and reasonably adept at the techniques, that are most fruitful in his particular line of research and development work. Finally, there is the individual who, as the head of, or as a member of a service group, has responsibility for analyzing and interpreting experimental and test data brought in by scientists and engineers engaged in ordnance research and development work. This individual needs a ready source of model work sheets and worked examples corresponding to the more common applications of statistics, to free him from the need of translating textbook discussions into step-by-step procedures that can be followed by individuals having little or no previous experience with statistical methods.

It is with this last need in mind that some of the procedures included in the Handbook have been explained and illustrated in detail twice: once for the case where the important question is whether the performance of a new material, product, or process exceeds an established standard; and again for the case where the important question is whether its performance is not up to the specified standards. Small but serious errors are often made in changing "greater than" procedures into "less than" procedures.

AUTHORSHIP AND ACKNOWLEDGMENTS

The Handbook on Experimental Statistics was prepared in the Statistical Engineering Laboratory, National Bureau of Standards, under a contract with the Office of Ordnance Research. The project was under the general guidance of Churchill Eisenhart, Chief, Statistical Engineering Laboratory.

Most of the present text is by Mary G. Natrella, who had overall responsibility for the completion of the final version of the Handbook. The original plans for coverage, a first draft of the text, and some original tables were prepared by Paul N. Somerville. Chapter 6 is by Joseph M. Cameron; most of Chapter 1 and all of Chapters 20 and 23 are by Churchill Eisenhart; and Chapter 10 is based on a nearly-final draft by Mary L. Epling.

Other members of the staff of the Statistical Engineering Laboratory have aided in various ways through the years, and the assistance of all who helped is gratefully acknowledged. Particular mention should be made of Norman C. Severo, for assistance with Section 2, and of Shirley Young Lehman for help in the collection and computation of examples.

Editorial assistance and art preparation was provided by John I. Thompson & Company, Washington, D. C.

Appreciation is expressed for the generous cooperation of publishers and authors in granting permission for the use of their source material. References for tables and other material, taken wholly or in part, from published works, are given on the respective first pages.

June 15, 1962

Comments on and requests for copies of this Handbook should be addressed to Ordnance Engineering Handbook Office, Ordnance Liaison Group, Durham, Box CM, Duke Station, Durham, N. C.

PREFACE

This listing is a guide to the Section and Chapter subject coverage in all Sections of the Handbook on Experimental Statistics.

Chapter No. *Title*

ORDP 20-110 (SECTION 1) — BASIC STATISTICAL CONCEPTS AND STANDARD TECHNIQUES FOR ANALYSIS AND INTERPRETATION OF MEASUREMENT DATA

1 — Some Basic Statistical Concepts and Preliminary Considerations
2 — Characterizing the Measured Performance of a Material, Product, or Process
3 — Comparing Materials or Products with Respect to Average Performance
4 — Comparing Materials or Products with Respect to Variability of Performance
5 — Characterizing Linear Relationships Between Two Variables
6 — Polynomial and Multivariable Relationships, Analysis by the Method of Least Squares

ORDP 20-111 (SECTION 2) — ANALYSIS OF ENUMERATIVE AND CLASSIFICATORY DATA

7 — Characterizing the Qualitative Performance of a Material, Product, or Process
8 — Comparing Materials or Products with Respect to a Two-Fold Classification of Performance (Comparing Two Percentages)
9 — Comparing Materials or Products with Respect to Several Categories of Performance (Chi-Square Tests)
10 — Sensitivity Testing

ORDP 20-112 (SECTION 3) — THE PLANNING AND ANALYSIS OF COMPARATIVE EXPERIMENTS

11 — General Considerations in Planning Experiments
12 — Factorial Experiments
13 — Randomized Blocks, Latin Squares, and Other Special-Purpose Designs
14 — Experiments to Determine Optimum Conditions or Levels

ORDP 20-113 (SECTION 4) — SPECIAL TOPICS

15 — Some "Short-Cut" Tests for Small Samples from Normal Populations
16 — Some Tests Which are Independent of the Form of the Distribution
17 — The Treatment of Outliers
18 — The Place of Control Charts in Experimental Work
19 — Statistical Techniques for Analyzing Extreme-Value Data
20 — The Use of Transformations
21 — The Relation Between Confidence Intervals and Tests of Significance
22 — Notes on Statistical Computations
23 — Expression of the Uncertainties of Final Results
Index

ORDP 20-114 (SECTION 5) — TABLES

Tables A-1 through A-37

TABLE OF CONTENTS

Paragraph		*Page*
	DISCUSSION OF TECHNIQUES IN CHAPTERS 7 THROUGH 10	x

CHAPTER 7

CHARACTERIZING THE QUALITATIVE PERFORMANCE OF A MATERIAL, PRODUCT, OR PROCESS

7-1	GENERAL	7-1
7-2	BEST SINGLE ESTIMATE OF THE TRUE PROPORTION P	7-1
7-3	CONFIDENCE INTERVAL ESTIMATES OF THE TRUE PROPORTION P	7-2
7-3.1	Two-Sided Confidence Intervals	7-2
7-3.1.1	Exact Limits for $n \leq 30$	7-2
7-3.1.2	Exact Limits for $n > 30$	7-2
7-3.1.3	Approximate Limits for $n > 30$	7-2
7-3.2	One-Sided Confidence Intervals	7-3
7-3.2.1	Exact Limits for $n \leq 30$	7-3
7-3.2.2	Exact Limits for $n > 30$	7-3
7-3.2.3	Approximate Limits for $n > 30$	7-3
7-4	SAMPLE SIZE REQUIRED TO ESTIMATE THE TRUE PROPORTION	7-4
7-4.1	Determining the Sample Size Required to Estimate the True Proportion *With a Specified Limit Of Error In Both Directions* (i.e., When It Is Required To Estimate P Within $\pm\delta$)	7-4
7-4.1.1	Graphical Method	7-4
7-4.1.2	Numerical Method	7-5
7-4.2	Determining the Sample Size Required To Estimate the True Proportion *With a Specified Limit Of Error In Only One Direction* (i.e., When It Is Required To Estimate P Within $+\delta$; or, To Estimate P Within $-\delta$)	7-5

Paragraph *Page*

CHAPTER 8

COMPARING MATERIALS OR PRODUCTS WITH RESPECT TO A TWO-FOLD CLASSIFICATION OF PERFORMANCE (COMPARING TWO PERCENTAGES)

8-1 COMPARING AN OBSERVED PROPORTION WITH A STANDARD PROPORTION 8-1

8-1.1 Does the New Product Differ From the Standard With Regard To the Proportion of Items Which Show the Characteristic of Interest? (Does P Differ From P_0?) 8-1

8-1.1.1 Procedure for $n \leq 30$ 8-1

8-1.1.2 Procedure for $n > 30$ 8-2

8-1.2 Does the Characteristic Proportion for the New Product *Exceed* That For the Standard? (Is $P > P_0$?) 8-3

8-1.2.1 Procedure for $n \leq 30$ 8-3

8-1.2.2 Procedure for $n > 30$ 8-4

8-1.3 Is the Characteristic Proportion for the New Product Less Than That for the Standard? (Is $P < P_0$?) 8-5

8-1.3.1 Procedure for $n \leq 30$ 8-5

8-1.3.2 Procedure for $n > 30$ 8-5

8-1.4 Sample Size Required To Detect a Difference Of Prescribed Magnitude From a Standard Proportion When the Sign of the Difference IS NOT Important 8-6

8-1.5 Sample Size Required To Detect a Difference Of Prescribed Magnitude From a Standard Proportion When the Sign of the Difference IS Important 8-7

8-2 COMPARING TWO OBSERVED PROPORTIONS 8-9

8-2.1 Comparing Two Proportions When the Sample Sizes Are Equal 8-9

8-2.1.1 Does the Characteristic Proportion for Product A Differ From That for Product B? (Does P_A Differ From P_B?) ... 8-10

8-2.1.2 Does the Characteristic Proportion for Product A Exceed That for Product B? (Is P_A Larger Than P_B?) 8-11

8-2.2 Comparing Two Proportions When the Sample Sizes Are Unequal and Small ($n_A \neq n_B$; Both No Greater Than 20) 8-12

8-2.2.1 Does the Characteristic Proportion for Product A Differ From That for Product B? 8-12

8-2.2.2 Does the Characteristic Proportion for Product A Exceed That for Product B? (Is P_A Larger than P_B?) 8-14

8-2.3 Comparing Two Proportions When the Sample Sizes Are Large 8-16

8-2.3.1 Does the Characteristic Proportion for Product A Differ From That for Product B? (Does P_A Differ From P_B?) ... 8-16

Paragraph *Page*

CHAPTER 8 (Cont)

8-2.3.2 Is the Characteristic Proportion for Product A Larger Than That for Product B? (Is P_A Larger Than P_B?) 8-18
8-2.4 Sample Size Required to Detect a Difference Between Two Proportions 8-18
8-2.4.1 Sample Size Required to Detect a Difference of Prescribed Magnitude Between Two Proportions When the Sign of the Difference IS NOT Important 8-18
8-2.4.2 Sample Size Required to Detect a Difference of Prescribed Magnitude Between Two Proportions When the Sign of the Difference IS Important 8-20

CHAPTER 9

COMPARING MATERIALS OR PRODUCTS WITH RESPECT TO SEVERAL CATEGORIES OF PERFORMANCE (CHI-SQUARE TESTS)

9-1 COMPARING A MATERIAL OR PRODUCT WITH A STANDARD 9-2
9-1.1 When the Comparison Is With a Standard Material or Product 9-2
9-1.2 When the Comparison Is With a Theoretical "Standard" 9-4
9-2 COMPARING TWO OR MORE MATERIALS OR PRODUCTS 9-6
9-3 A TEST OF ASSOCIATION BETWEEN TWO METHODS OF CLASSIFICATION 9-8

CHAPTER 10

SENSITIVITY TESTING

10-1 EXPERIMENTAL SITUATION 10-1
10-2 KÄRBER METHOD OF ANALYSIS 10-3
10-2.1 General Solution For the Kärber Method 10-4
10-2.1.1 Procedure 10-4
10-2.1.2 Example 10-5

Paragraph *Page*

CHAPTER 10 (Cont)

10-2.2 Simplified Solution (Kärber Method) For the Special Case When Test Levels Are Equally Spaced and Equal Numbers of Items Are Tested at Each Level....10-6
10-2.2.1 Procedure....10-6
10-2.2.2 Example....10-7
10-3 PROBIT METHOD OF ANALYSIS....10-8
10-3.1 Graphical Probit Solution....10-10
10-3.1.1 Procedure....10-10
10-3.1.2 Example....10-11
10-3.2 Exact Probit Solution....10-16
10-3.2.1 Procedure....10-16
10-3.2.2 Example....10-17
10-3.3 Testing Whether the Line Is An Adequate Representation of the Data....10-20
10-3.3.1 Procedure....10-20
10-3.3.2 Example....10-20
10-3.4 Using the Probit Regression Line For Prediction....10-21
10-3.4.1 Level of Stimulus x' At Which a Specified Proportion P' of the Individuals Would Be Expected To Respond....10-21
10-3.4.2 Level of Stimulus x' At Which 50% of the Individuals Would Be Expected To Respond....10-21
10-3.4.3 Proportion of Individuals Which Would Be Expected To Respond At a Specified Level of Stimulus....10-21
10-4 THE UP-AND-DOWN DESIGN....10-22
10-5 SENSITIVITY TESTS WHEN THE STIMULUS LEVELS CANNOT BE CONTROLLED....10-24

LIST OF ILLUSTRATIONS

Fig. No. *Title* *Page*

10-1. Probit regression line (fitted by eye) 10-13

LIST OF TABLES

Table No. *Title* *Page*

8-1. Observed frequencies from two samples in two mutually exclusive categories (a 2 × 2 table) 8-9
8-2. Rearrangement of Table 8-1 for convenient use in testing significance with Table A-29 8-12

9-1. Computational arrangement for Data Sample 9-1.1 9-3
9-2. Computational arrangement for Data Sample 9-1.2 9-5

9-3. Table of $\frac{f_{ij}^2}{n_i C_j}$ — computational arrangement for Data Sample 9-2 9-7

9-4. Table of $\frac{f_{ij}^2}{R_i C_j}$ — computational arrangement for Data Sample 9-3 9-10

10-1. Kärber method of analysis for fuze peak voltage test data 10-5
10-2. Simplified solution for the Kärber method of analysis when the test levels (x) are equally spaced and equal numbers of objects (n) are tested at each level 10-7
10-3. Graphical probit solution using Data Sample 10-1 10-11
10-4. Exact probit solution 10-17
10-5. Exact probit solution (second iteration) 10-19
10-6. Test of linearity — final probit equation 10-20

DISCUSSION OF TECHNIQUES IN CHAPTERS 7 THROUGH 10

For some kinds of tests, it may be impossible to obtain actual measurements. An item may be subjected to a test and the result of that particular test can be expressed only in terms of a pre-established classification of possible results. The simplest kind of classification, and the one most widely used in practice, consists of just two mutually exclusive categories. For example, the results of the test on each item may be recorded as pass or fail, hit or miss, fires or does not fire, larger than specification limit or less than specification limit, etc. Some other problems call for classification into more than two categories. In classifying types of metal fractures, we might establish such classes as smooth, rough, jagged, and splintery. Glass or plastic material after exposure to radiation might be classified as transparent, translucent, or opaque. In screening inspection, for example, we may have three established categories — accept, reject, or rework.

Once these qualitative observations have been recorded in the established categories of the classification scheme, we may count the number in any category, or we may compute the proportion of the total which falls in any category. In most of the analytical procedures given in these Chapters, we work with proportions, not with percentages, even though final presentation of results may be made in percentages.

The methods given in these Chapters also may be used for tests where exact measurements could have been obtained, but actually were not obtained because of the expense or inconvenience involved. For example, one always can measure a dimension; but, in large-scale production, go-no-go gauges may be used for routine checks. Whenever it is possible to obtain actual measurements, analysis of the measurements does provide more information than does analysis of counts. In planning experimental programs, various factors may contribute to the decision of whether to measure or to gauge — e.g., the availability of time, funds, and experienced personnel. When *measurements* are analyzed, the methods of these Chapters do not apply; the appropriate methods are given in ORDP 20-110, Chapters 2 through 6.

The problems considered in these Chapters parallel those of Chapters 2 through 6, as much as possible. Chapter 7 gives methods for making single estimates and interval estimates of a proportion. Instead of estimating the true average of a lot with respect to some property, we estimate the true proportion of items in the lot which have a particular property. Comparisons may be made between a new product and a standard product, or between any two products, with regard to the proportion of items which exhibit the characteristic in question. Chapter 8 gives methods for making such comparisons when the classification scheme consists of two categories. Chapter 9 gives methods for making such comparisons when the classification scheme consists of three or more categories. Chapter 10 gives methods of analysis for a particular experimental situation which has generally been called "sensitivity testing."

All A-Tables referenced in these Chapters are contained in ORDP 20-114, Section 5.

CHAPTER 7

CHARACTERIZING THE QUALITATIVE PERFORMANCE OF A MATERIAL, PRODUCT, OR PROCESS

7-1 GENERAL

The problem is that of estimating the true proportion (or percentage) of items that have a given quality characteristic. The tested items have been classified into two previously established categories of classification. Methods are given for obtaining:

(a) the best single estimate; and,

(b) confidence interval estimates* of the proportion which is of interest.

The following data will serve to illustrate the application of the procedures.

Data Sample 7-1 — Proportion of Defective Fuzes

Form: A sample of n items is selected at random from a much larger group. Upon examination or test, r of the n items show the characteristic of interest.

Example: Ten fuzes are taken at random from a production line, and are tested under a specified set of conditions. Four of the ten fail to function.

In general, what can we say on the basis of our sample about the larger group with regard to the proportion of defective items contained therein? We show how to answer two questions:

(a) What is the true proportion P of the fuzes produced that would be expected to fail under the specified conditions?

(b) What is an interval which we can expect, with prescribed confidence, to bracket the true proportion of defective fuzes?

7-2 BEST SINGLE ESTIMATE OF THE TRUE PROPORTION P

The best single estimate of the true proportion of items having a given characteristic in some well defined population is the observed proportion of items having this characteristic in a random sample from the population, i.e., the number of sample items which have the characteristic divided by the total number of items in the sample.

The best estimate from Data Sample 7-1 of the true proportion of fuzes that will fail is equal to the number of defective fuzes in the sample, divided by the total number of fuzes in the sample.

Procedure

(1) Compute the estimated proportion p, as follows:

$$p = \frac{r}{n}$$

Example

(1) From Data Sample 7-1,

$$p = 4/10$$
$$= .4$$

* The reader who is not familiar with the meaning and interpretation of confidence intervals should refer to Chapter 1 and to Paragraph 2-1.3 of ORDP 20-110.

7-3 CONFIDENCE INTERVAL ESTIMATES OF THE TRUE PROPORTION P

7-3.1 TWO-SIDED CONFIDENCE INTERVALS

Although the best single estimate of the true proportion of items having a given characteristic is the proportion of such items in a random sample, an interval estimate may be preferred. A two-sided confidence interval is an interval expected to give upper and lower limits for the true proportion with prescribed confidence.

7-3.1.1 Exact Limits for $n \leq 30$. For $n \leq 30$, two-sided confidence limits are given in Table A-22. For example, using Data Sample 7-1, where $n = 10$ and $r = 4$, a two-sided 95% confidence interval for the true proportion is the interval from .150 to .733.

7-3.1.2 Exact Limits for $n > 30$. For $n > 30$, use the charts of Table A-24 for 90%, 95%, and 99% confidence intervals, as desired. On the charts, there are two curves for each of a number of values of n. The upper and lower curve for a particular n constitute a confidence belt for the true proportion P. First, locate the observed proportion $p = r/n$, on the horizontal scale. From this point, travel up to the curves for the sample size n, and read off the upper and lower limits for the population proportion P. For example, in a sample of $n = 100$, where the observed proportion is .4, the interval from .31 to .51 gives 95% confidence limits for the true proportion P.

The three charts in Table A-24 give $(1 - \alpha)$ confidence interval estimates for $\alpha = .10$, $\alpha = .05$, and $\alpha = .01$. If we use these charts a large number of times to make interval estimates of the true proportion P, we can expect $100\,(1 - \alpha)$ % of these intervals to contain P. If the appropriate sample size requires interpolation on the charts, the procedure of Paragraph 7-3.1.3 should be used instead of the charts of Table A-24.

7-3.1.3 Approximate Limits for $n > 30$. This method should be used in lieu of interpolation on the charts (Table A-24).

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$.	(1) Let $1 - \alpha = .90$ $\alpha = .10$
(2) Look up $z_{1-\alpha/2}$ in Table A-2.	(2) $z_{.95} = 1.645$
(3) Compute: $p_1 = p - z_{1-\alpha/2}\sqrt{\dfrac{p(1-p)}{n}}$ $p_2 = p + z_{1-\alpha/2}\sqrt{\dfrac{p(1-p)}{n}}$	(3) Using $n = 150$, $p = .40$, for example, $p_1 = .40 - 1.645\sqrt{.0016}$ $= .40 - 1.645\,(.04)$ $= .40 - .07$ $= .33$ $p_2 = .40 + .07$ $= .47$
(4) The interval from p_1 to p_2 is a two-sided $100\,(1 - \alpha)$ % confidence interval estimate of the true proportion P.	(4) The interval from .33 to .47 is a 90% two-sided confidence interval estimate of P.

7-3.2 ONE-SIDED CONFIDENCE INTERVALS

A one-sided confidence interval estimate states that the true proportion P is less than a calculated proportion p_2' (or alternatively, that P is larger than p_1') and the statement is made at a prescribed confidence level.

7-3.2.1 Exact Limits for $n \leq 30$. For $n \leq 30$, one-sided confidence limits are given in Table A-23. For example, using Data Sample 7-1, where $n = 10$ and $r = 4$, the upper 90% one-sided confidence limit is .646. (The lower 90% one-sided confidence limit would be .188 in this case.)

7-3.2.2 Exact Limits for $n > 30$. Use the charts of Table A-24 to obtain .95, .975, or .995 one-sided confidence intervals, by using only the upper curve, or only the lower curve, of the belt for a given sample size. When used in this way, the chart labelled "Confidence Coefficient .90" yields one-sided 95% confidence intervals, the chart labelled "confidence coefficient .95" yields one-sided 97.5% confidence intervals, and the chart labelled "confidence coefficient .99" yields one-sided 99.5% confidence intervals.

If the appropriate sample size requires interpolation on the charts, the procedure of Paragraph 7-3.2.3 should be used instead of the charts of Table A-24.

7-3.2.3 Approximate Limits for $n > 30$. This method should be used in lieu of interpolation on the charts (Table A-24).

Procedure	Example
(1) Choose the desired confidence level, $1 - \alpha$.	(1) Let $1 - \alpha = .90$ $\alpha = .10$
(2) Look up $z_{1-\alpha}$ in Table A-2.	(2) $z_{.90} = 1.282$
(3) If a lower one-sided confidence limit is desired, compute $p_1' = p - z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}}$	(3) Using $n = 150$, $p = .40$, for example, $p_1' = .40 - 1.282\sqrt{.0016}$ $= .40 - 1.282\,(.04)$ $= .40 - .05$ $= .35$; this is the lower 90% confidence limit for P, the true proportion defective.
(4) Alternatively, if an upper one-sided confidence limit is desired, compute $p_2' = p + z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}}$	(4) Using $n = 150$, $p = .40$, for example, $p_2' = .40 + 1.282\,(.04)$ $= .40 + .05$ $= .45$; this is the upper 90% confidence limit for P, the true proportion defective.

7-4 SAMPLE SIZE REQUIRED TO ESTIMATE THE TRUE PROPORTION

We shall discuss two problems:

(a) Determining the sample size required to estimate the true proportion *with a specified limit of error in both directions;* i.e., when it is required to estimate P within $\pm\delta$. (See Paragraph 7-4.1.)

(b) Determining the sample size required to estimate the true proportion *with a specified limit of error in only one direction;* i.e., when it is required to estimate P within $+\delta$ (or to estimate P within $-\delta$). (See Paragraph 7-4.2.)

In (a), we are indifferent as to whether our estimate is too high or too low. In (b), we wish to protect ourselves against an overestimate, but do not worry about an underestimate (or vice versa).

7-4.1 DETERMINING THE SAMPLE SIZE REQUIRED TO ESTIMATE THE TRUE PROPORTION WITH A SPECIFIED LIMIT OF ERROR IN BOTH DIRECTIONS, i.e., WHEN IT IS REQUIRED TO ESTIMATE P WITHIN $\pm\delta$

7-4.1.1 Graphical Method. For the graphical method, the problem may be restated as follows: we wish to make a two-sided confidence interval estimate of P and the width of the interval should be not greater than 2δ. It, therefore, is possible to use the charts of Table A-24 in reverse; that is, to find the sample size belt whose maximum width (vertical distance on the charts) is equal to 2δ. The maximum width of confidence interval for a particular n will occur when the observed proportion is equal to 0.5. (If past records on the particular process indicate an upper or lower limit for the observed proportion, e.g., "the observed proportion always has been less than 0.1", one may use the widths of the intervals for this value of p rather than the maximum widths.)

Procedure

Problem: What is the sample size n required to estimate the true proportion P within $\pm\delta$?

(1) Choose $1 - \alpha$, the confidence coefficient to be associated with the resulting estimate. (The charts of Table A-24 can be used for confidence coefficients .90, .95, and .99).

(2) Specify δ, the error permitted in the estimate.

(3) Lacking knowledge of a safe upper or lower bound for P, look at the vertical line for $p = .50$. (If a safe upper or lower bound can be assumed, use this value of p.)

(4) Find the pair of n curves whose separation on this vertical line is not more than 2δ.

(5) n is the required sample size.

Example

Problem: What is the sample size n required to estimate the true proportion P within $\pm.10$?

(1) Let $1 - \alpha = .90$

Use Table A-24,
confidence coefficient = .90.

(2) Let $\delta = .10$

(3) Locate, on Table A-24, the vertical line for $p = .50$.

(4) At $p = .50$, n = 100 is the smallest n for which the interval is less than .20.

(5) $n = 100$ is the required sample size.

Note: $n = 50$ gives an interval approximately equal to .25, and $n = 100$ gives an interval approximately equal to .16, so that a sample somewhat less than 100 would be sufficient. The exact n, however, cannot be determined from the charts of Table A-24.

7-4.1.2 Numerical Method. The formula for sample size is

$$n = \frac{z^2_{1-\alpha/2}\, P(1 - P)}{\delta^2} .$$

A sample of size n guarantees a probability not greater than α that our estimate of P is in error by more than δ.

Since the true proportion P is unknown, we must substitute for it a value P' which is obtained as follows:

(a) If no prior information about P is available, or if P is believed to be in the neighborhood of 0.5, use $P' = 0.5$. The formula then simplifies to

$$n = \frac{z^2_{1-\alpha/2}}{4\delta^2} .$$

(b) If the true proportion P can safely be assumed to be less than 0.5, let P' be the *largest* reasonable guess for P.

(c) If the true proportion P can safely be assumed to be greater than 0.5, let P' be the *smallest* reasonable guess for P.

It is obvious that the largest sample size will be required when the true P is 0.5, and the purpose of these three rules is to be as conservative as possible.

Procedure	Example
Problem: What is the sample size n required to estimate the true proportion P within $\pm\delta$?	*Problem:* What is the sample size n required to estimate the true proportion P within $\pm.10$?
(1) Choose $1 - \alpha$, the confidence coefficient to be associated with the resulting estimate.	(1) Let $1 - \alpha = .90$ $\alpha = .10$
(2) Specify δ, the error permitted in the estimate.	(2) Let $\delta = .10$
(3) Look up $z_{1-\alpha/2}$ in Table A-2.	(3) $z_{.95} = 1.645$
(4a) If there is no prior information about the true proportion P, compute $n = \frac{z^2_{1-\alpha/2}}{4\delta^2}$	(4a) $n = \frac{(1.645)^2}{.04} = \frac{2.706}{.04} = 68,$ which is the required sample size.
(4b) If it is safe to assume that the true proportion P is less than some value P', compute $n = \frac{z^2_{1-\alpha/2}\, P'(1 - P')}{\delta^2}$	(4b) If it is safe to assume that the true proportion P is less than .40, for example, $n = \frac{(1.645)^2 (0.4)(0.6)}{.01} = 65,$ which is the required sample size.

7-4.2 DETERMINING THE SAMPLE SIZE REQUIRED TO ESTIMATE THE TRUE PROPORTION *WITH A SPECIFIED LIMIT OF ERROR IN ONLY ONE DIRECTION*, i.e., WHEN IT IS REQUIRED TO ESTIMATE P WITHIN $+\delta$ (OR TO ESTIMATE P WITHIN $-\delta$)

In some problems, we would be unconcerned if our estimate of P was too large, but would wish to be protected against an underestimate. Alternatively, in other problems, an underestimate is

tolerable, but not an overestimate. The error in the estimate is to be only in the direction that we choose.

The formula for sample size is

$$n = \frac{z_{1-\alpha}^2 P(1 - P)}{\delta^2} .$$

A sample of size n guarantees a probability not greater than α that our estimate of P is in error by more than $+\delta$ (or more than $-\delta$, as we choose).

Since the true proportion P is unknown, we must substitute for it a value P' which is obtained as follows:

(a) If no prior information about P is available, or if P is believed to be in the neighborhood of 0.5, use $P' = 0.5$. The formula then simplifies to

$$n = \frac{z_{1-\alpha}^2}{4\delta^2} .$$

(b) If the true proportion P can safely be assumed to be less than 0.5, let P' be the *largest* reasonable guess for P.

(c) If the true proportion P can safely be assumed to be greater than 0.5, let P' be the *smallest* reasonable guess for P.

The largest sample size will be required when $P = 0.5$, and the purpose of the rules is to be as conservative as possible.

Procedure	Example
Problem: What is the sample size n required to estimate the true proportion P within $+\delta$ (or, within $-\delta$)?	*Problem:* What is the required sample size? In estimating P, we wish to be protected against making an estimate that is too small by more than 0.05.
(1) Choose $1 - \alpha$, the confidence coefficient to be associated with the resulting estimate.	(1) Let $1 - \alpha = .90$ $\alpha = .10$
(2) Specify $+\delta$ (or $-\delta$) the error permitted in the estimate.	(2) Let $\delta = -.05$
(3) Look up $z_{1-\alpha}$ in Table A-2.	(3) $z_{.90} = 1.282$
(4a) If there is no prior information about P, compute $n = \frac{z_{1-\alpha}^2}{4\delta^2}$	(4a) $n = \frac{(1.282)^2}{.01}$ $= 164,$ which is the required sample size.
(4b) If it is safe to assume that the true proportion P is less than some value P', compute $n = \frac{z_{1-\alpha}^2 P'(1 - P')}{\delta^2}$	(4b) If it is safe to assume that the true proportion P is less than .40, for example, $n = \frac{(1.282)^2 (.4)(.6)}{.0025}$ $= 158,$ which is the required sample size.

CHAPTER 8

COMPARING MATERIALS OR PRODUCTS WITH RESPECT TO A TWO-FOLD CLASSIFICATION OF PERFORMANCE (COMPARING TWO PERCENTAGES)

In some situations, we are faced with the problem of comparing proportions or percentages. For example, the specification for a given kind of ammunition may prescribe the maximum allowable percentage of duds. Production lots of this ammunition will not be acceptable if they exceed this specified percent defective. The percentage of duds in a sample will provide an estimate of the percentage of duds in the lot, which then may be compared with the specified tolerance. When comparing an observed proportion with a specification or standard value, the procedures of Paragraph 8-1 are appropriate. The reader will note that the comparison is made by computing a confidence interval for the observed proportion and then looking to see whether the standard value is contained within this interval. This is a slightly different approach to answering the posed questions than was used in ORDP 20-110, Chapter 3, for example. Amplification of the relationship between confidence intervals and tests of significance of differences is given in ORDP 20-113, Chapter 21.

A different kind of comparison is involved when we compare two percentages with each other, without regard to any standard value — for example, in comparing two production processes with regard to the percentages of defective items produced. When two percentages are compared with each other, the methods of Paragraph 8-2 are appropriate.

8-1 COMPARING AN OBSERVED PROPORTION WITH A STANDARD PROPORTION

8-1.1 DOES THE NEW PRODUCT DIFFER FROM THE STANDARD WITH REGARD TO THE PROPORTION OF ITEMS WHICH SHOW THE CHARACTERISTIC OF INTEREST? (DOES P DIFFER FROM P_0?)

8-1.1.1 Procedure for $n \leq 30$.

Data Sample 8-1.1.1 — Defectives in Sample of New Product

Form: A sample of n items is selected at random from a much larger group. On examination, r of the n items show the presence of the pertinent characteristic. $p = r/n$ is the observed proportion, and is an estimate of P, the true proportion for the new product. P_0 is the known proportion of individual items in the standard product that show the pertinent characteristic.

Example: A sample of 20 components is taken from a production lot after a slight change in the process has been made. Three of the 20 items are classified as defectives. The observed

proportion, therefore, is $p = 3/20 = 0.15$. The proportion defective for this item with the standard process is known to be $P_0 = 0.10$.

The question to be answered is: Does the proportion defective in this lot, P, *differ* from the standard proportion defective (either an increase or a decrease being of interest)?

Procedure	Example
(1) Choose α, the significance level of the test. Table A-22 gives 90%, 95%, and 99% two-sided confidence limits, appropriate to $\alpha = .10$, $\alpha = .05$, and $\alpha = .01$, respectively.	(1) Let $\alpha = .10$ $1 - \alpha = .90$ Use the 90% confidence limits in Table A-22.
(2) Enter Table A-22 with observed n and r. Select appropriate column and read the limits.	(2) From Data Sample 8-1.1.1, $n = 20$ $r = 3$ From Table A-22, the 90% two-sided confidence limits for P are 0.056 to 0.328.
(3) If the tabled limits *do not* include P_0, conclude that P differs from P_0. If the tabled limits *do* include P_0, there is no reason to believe that P differs from P_0.	(3) Since the tabled limits *do* include $P_0 = 0.10$, there is no reason to believe that the proportion defective in the lot differs from the standard.

8-1.1.2 Procedure for $n > 30$.

Data Sample 8-1.1.2 — Performance of a New Type of Mine Fuze

Form: A sample of n items is selected at random from a much larger group. On examination, r of the n items show the presence of the pertinent characteristic. $p = r/n$ is the observed proportion, and is an estimate of P, the true proportion for the new product. P_0 is the known proportion of individual items in the standard product that show the pertinent characteristic.

Example: In a program of testing mine fuzes, 216 fuzes of a new type are buried, simulated "tanks" are run over them, and 160 "proper hits" are recorded. The observed proportion, p, of proper hits is $160/216 = 0.74$. The specified value for proportion of proper hits is $P_0 = 0.85$.

The question to be answered is: Does the proportion of proper hits for this fuze *differ* from the standard proportion (either an increase or a decrease being of interest)?

Procedure	Example
(1) Choose α, the significance level of the test. Table A-24 gives two-sided 90%, 95%, and 99% confidence limits appropriate to $\alpha = .10$, $\alpha = .05$, and $\alpha = .01$, respectively.	(1) Let $\alpha = .10$ $1 - \alpha = .90$ Use the Chart for confidence coefficient .90 in Table A-24.
(2) Compute $p = r/n$ and locate p on the horizontal scale. Locate curves for appropriate n.	(2) From Data Sample 8-1.1.2, $n = 216$ $r = 160$ $p = 160/216$ $= 0.74$
(3) Read off upper and lower limits for P. If these limits *do not* include P_0, conclude that P differs from P_0. If the limits *do* include P_0, there is no reason to believe that P differs from P_0.	(3) The chart does not show $n = 216$. Look at the belt for the next lower n, in this case $n = 100$. The belt for $n = 100$ will be wider than the belt for $n = 216$. Since the belt for $n = 100$ does not include $P_0 = 0.85$, the belt for $n = 216$ would not include P_0, and we conclude that the proportion of hits for this fuze does differ from the standard $P_0 = 0.85$.

8-1.2 DOES THE CHARACTERISTIC PROPORTION FOR THE NEW PRODUCT EXCEED THAT FOR THE STANDARD? (IS $P > P_0$?)

8-1.2.1 Procedure for $n \leq 30$. In terms of Data Sample 8-1.1.1, let us suppose that — in advance of looking at the data — the important question is: Does the characteristic proportion of defectives in this lot *exceed* that for the standard?

Procedure	Example
(1) Choose α, the significance level of the test. Table A-23 gives 90%, 95%, and 99% one-sided confidence limits appropriate to $\alpha = .10$, $\alpha = .05$, and $\alpha = .01$, respectively.	(1) Let $\alpha = .05$ $1 - \alpha = .95$ Use the 95% confidence limits in Table A-23.
(2) In Table A-23, follow directions at the beginning of the table to obtain p_1', a *lower* one-sided confidence limit for P.	(2) From Data Sample 8-1.1.1, $n = 20$ $r = 3$ $P_0 = 0.10$ is specified. The value in Table A-23, for $n = 20$, $n - r = 17$, is 0.958. The lower 95% limit for P is equal to $1 - 0.958 = 0.042$.

Procedure	Example
(3) If the *lower* limit p_1' obtained in Step (2) *exceeds* P_0, so that P_0 lies *outside* the confidence interval $p_1' \leq P \leq 1$, conclude that the characteristic proportion for the new product P exceeds that for the standard P_0.	(3)
If the *lower* limit obtained in Step (2) is *not larger* than P_0, so that P_0 lies *within* the confidence interval $p_1' \leq P \leq 1$, there is no reason to believe that the proportion for the new product P exceeds that for the standard P_0.	Since 0.042 is less than $P_0 = 0.10$, there is no reason to believe that the proportion of defectives in the lot exceeds the standard proportion.

8-1.2.2 Procedure for $n > 30$. In terms of Data Sample 8-1.1.2, let us suppose that — in advance of looking at the data — the important question is: Does the characteristic proportion of proper hits for this fuze *exceed* that for the standard?

Procedure	Example
(1) Choose α, the significance level of the test.	(1) Let $\alpha = .05$ $1 - \alpha = .95$ Use lower curves of the Chart labeled "confidence coefficient .90", in Table A-24.
By using only the lower curve of the confidence belt, Table A-24 gives 95%, 97.5%, and 99.5% one-sided confidence limits appropriate to $\alpha = .05$, $\alpha = .025$, and $\alpha = .005$, respectively.	
(2) Compute $p = r/n$, and locate p on the horizontal scale. Locate lower curve for appropriate n.	(2) From Data Sample 8-1.1.2, $n = 216$ $r = 160$ $p = 160/216$ $= 0.74$ $P_0 = 0.85$ is specified.
(3) Read off the *lower* confidence limit for P.	(3) From Table A-24, confidence coefficient .90, for $p = 0.74$ and $n = 216$, the lower 95% confidence limit for P is seen to be approximately 0.68.
If P_0 is *less* than this limit, p_1', so that P_0 lies *outside* the confidence interval $p_1' \leq P \leq 1$, conclude that the proportion for the new product exceeds that for the standard product.	
If P_0 is *larger* than p_1', and therefore is *included* in the confidence interval $p_1' \leq P \leq 1$, there is no reason to believe that P is larger than P_0.	Since $P_0 = 0.85$ is larger than 0.68, there is no reason to believe that P is larger than P_0.

8-1.3 IS THE CHARACTERISTIC PROPORTION FOR THE NEW PRODUCT LESS THAN THAT FOR THE STANDARD? (IS $P < P_0$?)

8-1.3.1 Procedure for $n \leq 30$. In terms of Data Sample 8-1.1.1, let us suppose that — in advance of looking at the data — the important question is: Is the characteristic proportion of defectives in this lot *less* than that for the standard?

Procedure	Example
(1) Choose α, the significance level of the test. Table A-23 gives 90%, 95%, and 99% one-sided confidence limits appropriate to $\alpha = .10$, $\alpha = .05$, and $\alpha = .01$, respectively.	(1) Let $\alpha = .10$ $1 - \alpha = .90$ Use the 90% confidence limits in Table A-23.
(2) Enter Table A-23 with n and r and chosen confidence. Read the upper one-sided limit p_2' for P.	(2) From Data Sample 8-1.1.1, $n = 20$ $r = 3$ $P_0 = 0.10$ is specified. The upper 90% limit for P is 0.304.
(3) If the tabled *upper* limit p_2' is *less* than P_0, so that P_0 lies *outside* the confidence interval $0 \leq P \leq p_2'$, conclude that the characteristic proportion for the new product is less than that for the standard. If the tabled limit is *larger* than P_0, so that P_0 lies *inside* the confidence interval $0 \leq P \leq p_2'$, there is no reason to believe that the proportion for the new product is less than the standard.	(3) Since the tabled limit (0.304) is larger than $P_0 = 0.10$, there is no reason to believe that the proportion of defectives in the lot is less than the standard.

8-1.3.2 Procedure for $n > 30$. In terms of Data Sample 8-1.1.2, let us suppose that — in advance of looking at the data — the important question is: Is the proportion of proper hits for this fuze *less than* that for the standard?

Procedure	Example
(1) Choose α, the significance level of the test. By using only the upper curve of the confidence belt, Table A-24 gives 95%, 97.5%, and 99.5% one-sided confidence limits appropriate to $\alpha = .05$, $\alpha = .025$, and $\alpha = .005$, respectively.	(1) Let $\alpha = .025$ $1 - \alpha = .975$ Use the upper curve of the Chart labeled "confidence coefficient .95", in Table A-24.
(2) Compute $p = r/n$, and locate p on the horizontal scale. Locate upper curve for appropriate n.	(2) From Data Sample 8-1.1.2, $n = 216$ $r = 160$ $p = 160/216$ $= 0.74$ $P_0 = 0.85$ is specified.

Procedure	Example
(3) Read off upper confidence limit p_2' for P. If P_0 is *larger* than this limit, so that P_0 lies *outside* the confidence interval $0 \leq P \leq p_2'$, conclude that the proportion for the new product is less than that for the standard product. If P_0 is *less* than this limit, and therefore is *included in* the one-sided confidence limit $0 \leq P \leq p_2'$, there is no reason to believe that P is less than P_0.	(3) From Table A-24, confidence coefficient .95, for $p = 0.74$ and $n = 216$, the upper 97.5% confidence limit for P is seen to be approximately 0.81. Since $P_0 = 0.85$ is larger than this limit, we conclude that the proportion of proper hits for this fuze is less than for the standard.

8-1.4 SAMPLE SIZE REQUIRED TO DETECT A DIFFERENCE OF PRESCRIBED MAGNITUDE FROM A STANDARD PROPORTION WHEN THE SIGN OF THE DIFFERENCE IS NOT IMPORTANT

Given:

P_0 = the known proportion of the population of standard items which exhibit the pertinent characteristic. P_0 may be known from the process history, or may be given by the requirements of a specification or a standard.

To be Specified for This Problem:

δ = the absolute magnitude of the difference which is considered important to detect.
α = the significance level, or the risk of announcing a difference when in fact there is none.
β = the risk of failing to detect a difference when in fact P, the true proportion for the new product, differs from the standard by an amount δ (i.e., $\delta = |P - P_0|$).

Tables to be Used:

Table A-25 gives the required sample size for a number of values of P_0 and P for $\alpha = .05$ and $1 - \beta = .50, .80, .90, .95$, and $.99$. This table is given largely for illustration, to demonstrate how the required sample size is affected by the magnitude of the P_0 and δ involved, and also by different choices of β. For desired values of α and β which are not included in Table A-25, use Table A-27, a table to convert the difference between the proportions into the form necessary for use with Table A-8.

Procedure	Example
(1) Specify δ, the absolute magnitude of the difference considered important to detect.	(1) Assume $P_0 = .30$ Specify $\delta = .10$
(2) Choose α and β.	(2) Let $\alpha = .05$ $\beta = .20$
(3) For $\alpha = .05$, and $1 - \beta = .50, .80, .90, .95$, and $.99$, go to Table A-25.	(3) Use Table A-25 with $1 - \beta = .80$
(4) Let $P = P_0 + \delta$, or $P = P_0 - \delta$, whichever makes P closer to 0.5.	(4) $P = .30 + .10$ $= .40$

Procedure	Example
(5) If *either* P or P_0 is less than 0.5, enter Table A-25 with P and P_0. If neither P nor P_0 is less than 0.5, enter Table A-25 with $1 - P$ and $1 - P_0$. In either case, the smaller of the two proportions determines the column and the larger determines the row in Table A-25.	(5) Enter Table A-25 in column .30 and row .40.
Read off n directly. n is the required sample size for the new product.	The required sample size is $n = 178$.
(6) For values of α, β, and P which are not included in Table A-25, go to Table A-27. Look up: $\theta_0 = \theta$ corresponding to P_0; $\theta = \theta$ corresponding to P	(6) Assume that we had wished to specify $\alpha = .01$, $\beta = .20$, $P_0 = .30$, and $P = .40$. From Table A-27, $\theta_0 = 1.16$; $\theta = 1.37$
(7) Compute $d = \lvert\theta - \theta_0\rvert$	(7) $d = 0.21$
(8) Enter Table A-8 with chosen α, $1 - \beta$, and d (from Step (7)). The tabled n is the required sample size for the new product. (The footnote to Table A-8 should be ignored.)	(8) From Table A-8, the sample size for $d = 0.2$ is $n = 292$. In this area of the table, interpolation is not recommended.
For values of d not given in Table A-8, the sample size may be computed using the formula $n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{d^2}$	To obtain the sample size for $d = 0.21$, compute $n = \frac{(2.576 + 0.84)^2}{(0.21)^2} = 265$

8-1.5 SAMPLE SIZE REQUIRED TO DETECT A DIFFERENCE OF PRESCRIBED MAGNITUDE FROM A STANDARD PROPORTION WHEN THE SIGN OF THE DIFFERENCE IS IMPORTANT

Given:

P_0 = the known proportion of the population of standard items which exhibit the pertinent characteristic. P_0 may be known from the process history, or may be given by the requirements of a specification or a standard.

To be Specified for This Problem:

δ = the absolute magnitude of the difference which is considered important to detect.
$P = P_0 + \delta$, if we wish to distinguish between P_0 and a proportion larger than P_0;
or,
$P = P_0 - \delta$, if we wish to distinguish between P_0 and a proportion smaller than P_0.
α = the significance level, or the risk of announcing a difference when in fact there is none.
β = the risk of failing to detect a difference when in fact the true proportion for the new product is P, where $P = P_0 + \delta$ or $P = P_0 - \delta$, depending on the choice made above.

Tables to be Used:

Table A-26 gives the required sample size for a number of values of P_0 and P for $\alpha = .05$ and $1 - \beta = .50, .80, .90, .95$, and $.99$. The Table is given largely for illustration, to demonstrate

how the required sample size is affected by the magnitude of the P_0 and δ involved, and also by different choices of β. For desired values of α and β which are not included in Table A-26, use Table A-27, a table to convert the difference between the proportions into the form necessary for use with Table A-9.

Procedure	Example
(1) Choose α and β.	(1) Let $\alpha = .05$ $\beta = .10$
(2) For $\alpha = .05$, and $1 - \beta = .50, .80, .90, .95$, and $.99$, go to Table A-26.	(2) Use Table A-26, with $1 - \beta = .90$
(3) Let $P = P_0 + \delta$, or $P = P_0 - \delta$, as specified.	(3) Assume $P_0 = 0.70$ Specify $P = 0.70 + 0.10$ $= 0.80$
(4) If either P or P_0 is less than 0.5, enter Table A-26 with P and P_0. If neither P nor P_0 is less than 0.5, enter Table A-26 with $1 - P$ and $1 - P_0$.	(4) Since neither P nor P_0 is less than 0.5, take $1 - P = 0.20$ $1 - P_0 = 0.30$.
In either case, the smaller of the two proportions determines the column and the larger determines the row in Table A-26.	Use column 0.20 and row 0.30 in Table A-26.
Read off n directly. n is the required sample size for the new product.	The required sample size is $n = 160$.
(5) For values of α, β, and P not included in Table A-26, go to Table A-27. Look up: $\theta_0 = \theta$ corresponding to P_0 $\theta = \theta$ corresponding to P	(5) Assume that we had specified $\alpha = .01$, $\beta = .10$, $P_0 = .10$, and $P = .40$. From Table A-27, $\theta_0 = .64$ $\theta = 1.37$
(6) Compute $d = \lvert\theta - \theta_0\rvert$	(6) $d = 0.73$
(7) Enter Table A-9 with chosen α, $1 - \beta$, and d (from Step (6)). The tabled n is the required sample size for the new product.	(7) From Table A-9, the sample size for $d = 0.6$ is $n = 37$, and for $d = 0.8$, is $n = 21$, so that the required sample size for $d = 0.73$ is greater than 21 and less than 37. In this area of the table, interpolation is not recommended.
For values of d not given in Table A-9, the sample size may be computed using the formula $n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{d^2}$	To obtain the sample size for $d = 0.73$, compute $n = \frac{(2.326 + 1.282)^2}{(0.73)^2}$ $= \frac{13.018}{.5329}$ $= 24.4$ As is conventional in sample size calculations, we round up to $n = 25$.

8-2 COMPARING TWO OBSERVED PROPORTIONS

We assume that n_A and n_B items are taken from products A and B, respectively. The items are to be examined or tested, and then classified into one of two mutually-exclusive categories. Some examples of two-category classifications are: hit or miss; pass or fail; white or black; damaged or not damaged; within tolerance or outside tolerance; etc. For purposes of illustration, we call the two categories Class I and Class II.

After examination of the n_A items, a number r_A are classified as Class I, and after examination of the n_B items, a number r_B are classified as Class I. The observed classification of the items is recorded in a two-row, two-column table (often called a 2 × 2 table) as shown in Table 8-1. Since there are just two mutually-exclusive classes, the entries for Class II can be filled in by subtracting the number recorded for Class I from the total number for each sample.

TABLE 8-1. OBSERVED FREQUENCIES FROM TWO SAMPLES IN TWO MUTUALLY EXCLUSIVE CATEGORIES (A 2 X 2 TABLE)

	Class I	Class II	Total
Sample from A	r_A	s_A	$n_A = r_A + s_A$
Sample from B	r_B	s_B	$n_B = r_B + s_B$
Total	r	s	n

The rows in the Table represent the two samples, and the columns are the two classes into which the observed items have been classified. Entries in the Table are counts — e.g., in the A sample (consisting of n_A items), r_A items are found to be Class I and s_A items ($s_A = n_A - r_A$) are Class II.

Although the problems will be posed in terms of proportions, and final results presented in terms of proportions, most of the techniques given use the observed *counts*. In terms of Table 8-1, if Class I is the property of interest, the observed proportions are $p_A = r_A/n_A$ and $p_B = r_B/n_B$.

Since the selection of available techniques depends on the sample sizes involved, this section is organized in three subparagraphs:

8-2.1 Comparing two proportions when the sample sizes are equal ($n_A = n_B$).

8-2.2 Comparing two proportions when the sample sizes are unequal and small ($n_A \neq n_B$; both less than 20).

8-2.3 Comparing two proportions when the sample sizes are unequal and large.

In each paragraph, procedures will be given for answering two questions:

(a) Does the characteristic proportion for product A differ from that for product B?

(b) Does the characteristic proportion for product A exceed that for product B?

As always, it is important to decide which question is appropriate before taking the observations. If this is not done, and if the choice of the problem is influenced by the observations, both the significance level of the test and the operating characteristics of the test may differ considerably from their nominal values.

8-2.1 COMPARING TWO PROPORTIONS WHEN THE SAMPLE SIZES ARE EQUAL

The solution involves three operations:

(a) Recording the observed counts in the form shown in Table 8-1;

(b) Selecting the proper pair of entries from among the four entries in the table; and,

(c) Comparing that pair with the "minimum contrast" pair given in Table A-28, to determine whether or not the observed contrast is significant at the chosen level. The procedure is detailed in Paragraphs 8-2.1.1 and 8-2.1.2; and, with a little practice, can be done quickly by eye.

Table A-28 gives "minimum contrasts" for $n_A = n_B = 1(1)20(10)100(50)200(100)500$ corresponding to significance levels $\alpha = .05$ and $\alpha = .01$ for two-sided tests (see Paragraph 8-2.1.1); or to $\alpha = .025$ and $\alpha = .005$ for one-sided tests (see Paragraph 8-2.1.2). By "minimum contrast" is meant the "least different"

pair which is significant at the chosen significance level. A "more different" pair is of course significant also. For example, look at the entries in Table A-28 for $n_A = n_B = 17$. The "minimum contrasts required" at significance level $\alpha = .05$ for the two-sided test (Does P_A differ from P_B?) are (0, 5), (1, 7), (2, 9), (3, 10), etc. Since (0, 5) is significant, so also is (0, 6), (0, 7), etc. Since (1, 7) is significant, so also is (1, 8), (1, 9), etc.

It is worth noting that Table A-28 can be used to give satisfactory answers for values of n intermediate to those tabulated (see *Note* to Table A-28).

Data Sample 8-2.1 — Small-scale Comparison Test of Two Types of Mine Fuzes

Seventeen impact fuzes of each of two different types are tested, and the number of successful firings are recorded, as follows:

Fuze Type	Success (Class I)	Failure (Class II)	Total
Type A	$r_A = 15$	$s_A = 2$	$n_A = 17$
Type B	$r_B = 7$	$s_B = 10$	$n_B = 17$
Total	$r = 22$	$s = 12$	$n = 34$

8-2.1.1 Does the Characteristic Proportion for Product A Differ From That for Product B? (Does P_A Differ From P_B?) In terms of Data Sample 8-2.1, we wish to compare the proportion of successful firings for the two types of fuzes. The question to be answered is: Does Type A *differ* from Type B with regard to the proportion of successful firings?

Procedure	Example
(1) Choose α, the significance level of the test. Table A-28 provides "minimum contrasts" corresponding to $\alpha = .05$ and $\alpha = .01$ for this two-sided test.*	(1) Let $\alpha = .01$
(2) Record the observed counts as in Table 8-1.	(2) See Data Sample 8-2.1.
(3) Let: a_1 = smallest of all four entries a_2 = entry in the same class as a_1 from the other sample. The "observed contrast" pair is the ordered pair (a_1, a_2). If $a_1 = a_2$, no further analysis is necessary. The data give no reason to believe that the two proportions differ.	(3) $a_1 = 2$ $a_2 = 10$ The "observed contrast" pair is (2, 10).
(4) Enter Table A-28 with sample size $n_A = n_B$.	(4) $n_A = n_B = 17$
(5) Call the tabled pairs (A_1, A_2). Find the tabled pair where $A_1 = a_1$; this is the "least different" pair which is significant at the chosen level.	(5) From Table A-28, with $\alpha = .01$ and $a_1 = 2$, the "least different" pair $(A_1, A_2) = (2, 11)$.
(6) If a_2 is equal to or larger than A_2, the observed contrast is significant at the chosen level, and we conclude that the two products differ with regard to the characteristic proportion considered. If a_2 is smaller than A_2, there is no reason to believe that the two proportions differ.	(6) Since $a_2 = 10$ is less than $A_2 = 11$, the observed contrast is not significant at the .01 level, and we conclude that there is no reason to believe that the two fuze types differ with regard to the proportion of successful firings.

* Table A-29 and the more complicated procedure of Paragraph 8-2.2.1 can be used to conduct equivalent two-sided tests corresponding to $\alpha = \underline{.10}, .05, \underline{.02}$, and .01, when $n_A = n_B \leq 20$, thus extending the present two-sided test procedure to the underscored values of α for equal sample sizes up to 20.

8-2.1.2 Does the Characteristic Proportion for Product A Exceed That for Product B? (Is P_A Larger Than P_B?) In terms of Data Sample 8-2.1, let us suppose that — in advance of looking at the data — the important question is: Does the proportion of successful firings for Type A exceed the proportion of successes for Type B?

Procedure	Example
(1) Choose α, the significance level of the test. Table A-28 provides "minimum contrasts" corresponding to $\alpha = .025$ and $\alpha = .005$ for this one-sided test.*	(1) Let $\alpha = .025$
(2) Record the observed counts as in Table 8-1.	(2) See Data Sample 8-2.1
(3) Compute p_A, the observed proportion for Product A, and p_B, the observed proportion for Product B. If Class I is the class of interest, $p_A = r_A/n_A$ and $p_B = r_B/n_B$. If Class II is the class of interest, $p_A = s_A/n_A$ and $p_B = s_B/n_B$.	(3) $p_A = 15/17 = 0.88$ $p_B = 7/17 = 0.41$
(4) If p_A is not larger than p_B, conclude at once that there is no reason to believe that the true proportion P_A is larger than P_B; otherwise, proceed to Step (5).	(4) Since p_A is larger than p_B, proceed to Step (5).
(5) If p_A is larger than p_B, let: a_1 = smallest of all four entries a_2 = entry in the same class as a_1 from the other sample. The "observed contrast" pair is the ordered pair (a_1, a_2).	(5) $a_1 = 2$ $a_2 = 10$ The "observed contrast" pair is (2, 10).
(6) Enter Table A-28 with sample size $n_A = n_B$.	(6) $n_A = n_B = 17$
(7) Call the tabled pairs (A_1, A_2). Find the tabled pair where $A_1 = a_1$; this is the "least different" pair which is significant at the chosen level.	(7) From Table A-28, with $\alpha = .025$ and $a_1 = 2$, the "least different" pair $(A_1, A_2) = (2, 9)$.
(8) If a_2 is equal to or is larger than A_2, the observed contrast is significant at the chosen level, and we conclude that the proportion for Product A exceeds that for Product B. If a_2 is smaller than A_2, there is no reason to believe that the two proportions differ.	(8) Since $a_2 = 10$ is larger than $A_2 = 9$, we conclude that the proportion of successes for type A exceeds that for type B.

* Table A-29 and the more complicated procedure of Paragraph 8-2.2.2 can be used to conduct equivalent one-sided tests corresponding to $\alpha = \underline{.05}, .025, \underline{.01}$, and .005, when $n_A = n_B \leq 20$, thus extending the present one-sided test procedure to the underscored values of α for equal sample sizes up to 20.

8-2.2 COMPARING TWO PROPORTIONS WHEN THE SAMPLE SIZES ARE UNEQUAL AND SMALL ($n_A \neq n_B$; BOTH NO GREATER THAN 20)

8-2.2.1 Does the Characteristic Proportion for Product A Differ From That for Product B?

Data Sample 8-2.2.1 — Small-scale Comparison Test of Two Types of Artillery Fuzes

The following data are recorded from an artillery fuze-testing program:

Fuze Type	Fires (Class I)	Does Not Fire (Class II)	Total
Type A	$r_A = 4$	$s_A = 2$	$n_A = 6$
Type B	$r_B = 8$	$s_B = 2$	$n_B = 10$
Total	$r = 12$	$s = 4$	$n = 16$

Procedure	Example
(1) Choose α, the significance level of the test. Table A-29 provides a listing of "significant contrasts" corresponding to $\alpha = .10, .05, .02$, and $.01$ for this two-sided test.*	(1) Let $\alpha = .02$
(2) Record the observed counts as in Table 8-1.	(2) See Data Sample 8-2.2.1
(3) In order to use Table A-29 for this problem, we must have the data arranged in a special way. Arrange the data as shown in Table 8-2 so that the results from the larger sample are in the first row, and re-label the entries r_1, r_2, etc., as shown in Table 8-2. Retain the original product identification of the samples.	(3) See Rearranged Data Sample 8-2.2.1A

TABLE 8-2. REARRANGEMENT OF TABLE 8-1 FOR CONVENIENT USE IN TESTING SIGNIFICANCE WITH TABLE A-29

	Class I	Class II	Total
Larger Sample	r_1	s_1	n_1
Smaller Sample	r_2	s_2	n_2
Total	r	s	n

* It should be noted that Table A-29 also could be used for equal sample sizes up to 20. For equal samples and $\alpha = .05$ or $.01$, however, the method of Paragraph 8-2.1.1 is recommended because of simplicity.

Rearranged Data Sample 8-2.2.1A

Fuze Type	Fires (Class I)	Does Not Fire (Class II)	Total
Larger sample (Type B)	$r_1 = 8$	$s_1 = 2$	$n_1 = 10$
Smaller sample (Type A)	$r_2 = 4$	$s_2 = 2$	$n_2 = 6$
Total	$r = 12$	$s = 4$	$n = 16$

Procedure	Example
(4) Compute the four proportions:	(4)
$p_1 = r_1/n_1$	$p_1 = 8/10 = .80$
$p_2 = r_2/n_2$	$p_2 = 4/6 = .67$
$q_1 = s_1/n_1$	$q_1 = 2/10 = .20$
$q_2 = s_2/n_2$	$q_2 = 2/6 = .33$
(5) If p_1 is larger than or is equal to p_2, focus attention on Class I. For use with Table A-29, take	(5) Since p_1 is larger than p_2, focus on Class I.
$a_1 = r_1$	$a_1 = r_1 = 8$
$a_2 = r_2$	$a_2 = r_2 = 4$
If q_1 is larger than or is equal to q_2, focus attention on Class II. For use with Table A-29, take	
$a_1 = s_1$	
$a_2 = s_2$	
(6) Enter Table A-29 with n_1, n_2, and a_1 (determined from Step (5)).	(6) From Table A-29, for $n_1 = 10$, $n_2 = 6$, $a_1 = 8$, and $\alpha = .02$, the tabled a_2 is 0.
The observed a_2 (from Step (5)) must be equal to or smaller than the tabled a_2 (bold-face in Table A-29) for significance at the chosen level. Therefore, if the observed a_2 is equal to or is smaller than bold-face a_2 in Table A-29, conclude that the two products differ with regard to the proportion of interest.	
If the observed a_2 is larger than the tabled a_2, there is no reason to believe that the two products differ.	Since the observed $a_2 = 4$ is larger than the tabled a_2, there is no reason to believe that the two fuzes differ in regard to the proportion which fire.

8-2.2.2 Does the Characteristic Proportion for Product A Exceed That for Product B? (Is P_A Larger Than P_B?)

Data Sample 8-2.2.2 — Small-scale Comparison Test of Two Types of Impact Fuzes

The following data are recorded from an impact fuze-testing program:

Fuze Type	Fires (Class I)	Does Not Fire (Class II)	Total
Type A	$r_A = 4$	$s_A = 2$	$n_A = 6$
Type B	$r_B = 0$	$s_B = 10$	$n_B = 10$
Total	$r = 4$	$s = 12$	$n = 16$

Procedure	Example
(1) Record the observed counts as shown in Table 8-1.	(1) See Data Sample 8-2.2.2.
(2) Focus on the class of interest. If this is Class I, compute: $p_A = r_A/n_A$ $p_B = r_B/n_B$ (If Class II were the class of interest p_A would equal s_A/n_A, and p_B would equal s_B/n_B). If p_A is larger than p_B, proceed to Step (3). If p_A is not larger than p_B, conclude at once that the data give no reason to believe that the true proportion P_A is larger than P_B.	(2) Since we are interested in comparing the proportions of fuzes which do fire, compute: $p_A = r_A/n_A = 4/6 = .67$ $p_B = r_B/n_B = 0/10 = 0$ Since p_A is larger than p_B, proceed to Step (3).
(3) Choose α, the significance level of the test. Table A-29 is used for this one-sided test* for $\alpha = .05, .025, .01$, and $.005$.	(3) Let $\alpha = .01$
(4) In order to use Table A-29 for this problem, we must have the data arranged in a special way. Arrange the data as shown in Table 8-2 so that the results from the larger sample are in the first row, and relabel the entries r_1, r_2, etc., as shown in Table 8-2. Retain the original product identification of the samples.	(4) See Rearranged Data Sample 8-2.2.2A.

* It should be noted that Table A-29 also could be used for equal sample sizes up to 20. For equal sample sizes and $\alpha = .025$ and $\alpha = .005$, however, the method of Paragraph 8-2.1.2 is recommended because of simplicity.

Rearranged Data Sample 8-2.2.2A

Fuze Type	Fires (Class I)	Does Not Fire (Class II)	Total
Larger sample (Type B)	$r_1 = 0$	$s_1 = 10$	$n_1 = 10$
Smaller sample (Type A)	$r_2 = 4$	$s_2 = 2$	$n_2 = 6$
Total	$r = 4$	$s = 12$	$n = 16$

Procedure	Example
(5) Compute the four proportions:	(5)
$p_1 = r_1/n_1$	$p_1 = 0/10 = 0$
$p_2 = r_2/n_2$	$p_2 = 4/6 = .67$
$q_1 = s_1/n_1$	$q_1 = 10/10 = 1.00$
$q_2 = s_2/n_2$	$q_2 = 2/6 = .33$
Note that: $p_1 = p_A$ if n_A is the larger sample; $p_1 = p_B$ if n_B is the larger sample.	
(6) If p_1 is larger than or is equal to p_2, focus attention on Class I. For use with Table A-29, take $a_1 = r_1$, $a_2 = r_2$.	(6)
If q_1 is larger than or is equal to q_2, focus attention on Class II. For use with Table A-29, take	Since q_1 is larger than q_2, focus attention on Class II.
$a_1 = s_1$	$a_1 = s_1 = 10$
$a_2 = s_2$	$a_2 = s_2 = 2$
(7) Enter Table A-29 with n_1, n_2, and the a_1 (determined from Step (6)). The observed a_2 (from Step (6)) must be equal to or smaller than the tabled a_2 (bold face in Table A-29) for significance at the chosen level.	(7) From Table A-29, for $n_1 = 10$, $n_2 = 6$, $a_1 = 10$, and $\alpha = .01$, the tabled a_2 is 2.
Therefore, if observed a_2 is equal to or is smaller than bold-face a_2 in Table A-29, conclude that the proportion of interest for product A exceeds the proportion for product B. If the observed a_2 is larger than the tabled a_2, there is no reason to believe that the two proportions differ.	Since the observed a_2 is equal to the tabled a_2, we conclude that the proportion of successful fuzes of type A exceeds that for type B.

8-2.3 COMPARING TWO PROPORTIONS WHEN THE SAMPLE SIZES ARE LARGE*

8-2.3.1 Does the Characteristic Proportion for Product A Differ From That for Product B? (Does P_A Differ From P_B?)

Data Sample 8-2.3.1 — Field Trials of Two Types of Mine Fuzes

In field trials of mine fuzes, 216 of each of two new types of fuze were buried, simulated tanks run over them, and the number of "proper hits" recorded. The results are as follows:

Fuze Type	Hit	Not Hit	Total
Type A	$r_A = 181$	$s_A = 35$	$n_A = 216$
Type B	$r_B = 160$	$s_B = 56$	$n_B = 216$
Total	$r = 341$	$s = 91$	$n = 432$

Let us assume with respect to Data Sample 8-2.3.1 that the important question is: Is the proportion of hits for Type A different from the proportion of hits for Type B?

Procedure	Example
(1) Choose α, the significance level of the test.	(1) Let $\alpha = .10$
(2) Look up $\chi^2_{1-\alpha}$ for one degree of freedom in Table A-3.	(2) $\chi^2_{.90}$ for 1 d.f. $= 2.71$
(3) Compute $$\chi^2 = \frac{n\left(\lvert r_A s_B - r_B s_A \rvert - \frac{n}{2}\right)^2}{n_A\, r\, n_B\, s}$$ (See *Note* below.)	(3) $$\chi^2 = \frac{432\,(4536 - 216)^2}{(73656)\,(19656)} = 5.57$$
(4) If $\chi^2 \geq \chi^2_{1-\alpha}$, decide that the two products differ with regard to the proportion having the given characteristic; otherwise, there is no reason to believe that the products differ in this respect.	(4) Since χ^2 is larger than $\chi^2_{.90}$, we conclude that the two types of fuzes do differ with regard to the proportion of "proper hits".

* The procedures of this paragraph *must be used* for large samples of unequal size, and *may be used* for samples of equal size. If the sample sizes are equal and are included in Table A-28, the procedures of Paragraph 8-2.1 are to be preferred because of simplicity.

Procedure

Note: The computation of χ^2 is most conveniently done in terms of the actual counts in the table, as given in Step (3) above. The formula can be expressed in terms of the observed proportions as follows:

$$\chi^2 = \frac{\left(n'|p_A - p_B| - \frac{1}{2}\right)^2}{n'p(1-p)}$$

where

$$p_A = r_A/n_A$$

$$p_B = r_B/n_B$$

$$p = \frac{r_A + r_B}{n_A + n_B}$$

and

$$n' = \frac{n_A n_B}{n_A + n_B}$$

This formula and the formula in Step (3) are algebraically equivalent, but use of the form given in the *Note* requires extra arithmetic and rounding. In spite of the fact that the question is put in terms of the difference between proportions, the answer is obtained more easily and more accurately using observed *counts*, i.e., the formula of Step (3) is preferred. Furthermore, using the formula in terms of counts (Step (3)) highlights the fact that one *cannot* judge the difference between two proportions without knowing the sample sizes involved.

Example

Note: Using Data Sample 8-2.3.1 with this formula:

$$\chi^2 = \frac{(108\,(.097) - 1/2)^2}{108\,(.789)\,(.211)} = \frac{(9.976)^2}{17.980} = 5.54$$

$$p_A = 181/216 = .838$$

$$p_B = 160/216 = .741$$

$$p = 341/432 = .789$$

$$1 - p = .211$$

$$n' = 46656/432 = 108$$

8-2.3.2 Is the Characteristic Proportion for Product A Larger Than That for Product B? (Is P_A Larger Than P_B?) In terms of Data Sample 8-2.3.1, let us suppose that — in advance of looking at the data — the important question is: Is the proportion of hits for type A larger than the proportion of hits for type B?

Procedure	Example
(1) Choose α, the significance level of the test.	(1) Let $\alpha = .025$
(2) Look up $\chi^2_{1-2\alpha}$ for one degree of freedom in Table A-3.	(2) $\chi^2_{.95}$ for 1 d.f. $= 3.84$
(3) Compute: $$\chi^2 = \frac{n\left(\lvert r_A s_B - r_B s_A \rvert - \frac{n}{2}\right)^2}{n_A\, r\, n_B\, s}$$	(3) $$\chi^2 = \frac{432\,(4536 - 216)^2}{(73656)\,(19656)} = 5.57$$
and $p_A = r_A/n_A$	$p_A = 181/216 = .84$
$p_B = r_B/n_B$	$p_B = 160/216 = .74$
(See *Note* at end of procedure of Paragraph 8-2.3.1)	
(4) If $\chi^2 \geq \chi^2_{1-2\alpha}$ *and* p_A is larger than p_B, decide that P_A exceeds P_B; otherwise, there is no reason to believe the proportions differ.	(4) Since χ^2 is larger than $\chi^2_{.95}$ and $p_A = .84$ is larger than $p_B = .74$, conclude that the proportion of hits for type A is larger than the proportion of hits for type B.

8-2.4 SAMPLE SIZE REQUIRED TO DETECT A DIFFERENCE BETWEEN TWO PROPORTIONS

8-2.4.1 Sample Size Required to Detect a Difference of Prescribed Magnitude Between Two Proportions When the Sign of the Difference Is Not Important. Unfortunately, the sample size required depends on the true but unknown values of the two proportions involved. Very often, the experimenter has some idea of the magnitude of (or an upper bound for) one of these values, and then must specify the size of the difference which the experiment should be designed to detect. For a fixed difference to be detected, the largest sample sizes will be required if the true proportions are in the neighborhood of 0.5. A look at Table A-25, however, will show that over-conservatism may not pay. Suppose, for example, that one of the proportions can *safely* be assumed to be less than 0.4. The most conservative assumption would be that it is equal to 0.4 (this being the closest reasonable guess to 0.5). Attempting to be over-cautious by using the value 0.45 will extract a heavy price in the number of tests to be run.

Given:

For this problem there is nothing given, but —

Assumed:

P' = an estimate of one of the two proportions.
To be conservative, make this estimate as close to 0.5 as is reasonable.

To be Specified for This Problem:

α = the significance level, or the risk of announcing a difference when in fact there is none.
β = the risk of failing to detect a difference when in fact the true proportions differ by an amount δ (i.e., $|P' - P''| = \delta$).
δ = the absolute magnitude of the difference which is considered important to detect.

Tables to be Used:

Table A-25 can be used for $\alpha = .05$ and $1 - \beta = .50, .80, .90, .95$, and $.99$, and for certain values of the proportions. The entry in Table A-25 must be doubled to give n'; and n' is the required sample size to be taken from each product.

For other desired values of α and β, use Table A-27, a table to convert the difference between the proportions into the form necessary for use with Table A-8.

The question to be answered by the experiment is: Does P_A differ from P_B?

Procedure	Example
(1) Specify δ, the absolute magnitude of the difference considered important to detect.	(1) Specify $\delta = .10$
(2) Choose α and β.	(2) Let $\alpha = .05$ $\beta = .20$
(3) For $\alpha = .05$ and $1 - \beta = .50, .80, .90, .95$, or $.99$, go to Table A-25.	(3) Use Table A-25 with $1 - \beta = .80$
(4) Let P' = an estimate of one of the proportions. Let $P'' = P' + \delta$ or $P' - \delta$, *whichever* makes P'' closer to 0.5.	(4) Let $P' = .20$ Let $P'' = .20 + .10$ $= .30$
(5) If *either* P' or P'' is less than 0.5, enter Table A-25 with P' and P''. If *neither* P' nor P'' is less than 0.5, enter Table A-25 with $1 - P'$ and $1 - P''$. In either case, the smaller of the two proportions determines the column and the larger of the two determines the row in Table A-25.	(5) Enter Table A-25 in column .20 and row .30.
Read off n, and double it to obtain n'. n' is the required sample size to be taken from each product.	$n = 146$ $n' = 292$, the required sample size to be taken from each product.
(6) For other values of α and β, and for values of P' and P'' not included in Table A-25, go to Table A-27. Look up: θ' = θ corresponding to P' θ'' = θ corresponding to P''	(6) Assume that we had specified $\alpha = .01$, $\beta = .20$, $P' = .34$, and $P'' = .44$. From Table A-27, $\theta' = 1.25$ $\theta'' = 1.45$
(7) Compute $d = \|\theta' - \theta''\|$	(7) $d = \|1.25 - 1.45\|$ $= .20$

Procedure	Example
(8) Enter Table A-8 with α, β, and d (from Step (7)). Read off n and double it to obtain n'. Then, n' is the required sample size to be taken from each product. (Rounding two-decimal values of d to the nearest value considered in Table A-8 may lead to excessively high (or low) values of n, and thence of n'. Interpolation for values of n corresponding to values of d not considered in Table A-8 is *not* recommended. For values of d not given in Table A-8, the sample size may be computed using the formula $n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{d^2}$	(8) From Table A-8 with $\alpha = .01$, $1 - \beta = .80$, and $d = .20$, $n = 292$. $n' = 584$, the required sample size to be taken from each product.

8-2.4.2 Sample Size Required to Detect a Difference of Prescribed Magnitude Between Two Proportions When the Sign of the Difference Is Important. Read the general discussion at the beginning of Paragraph 8-2.4.1.

Given:

For this problem, there is nothing given, but —

Assumed:

P' = an estimate of one of the two proportions. P' may be P_A', an estimate of P_A, or P_B', an estimate of P_B.
To be conservative, make this estimate as close to 0.5 as is reasonable.

To be Specified for This Problem:

α = the significance level, or the risk of announcing a difference when in fact there is none.
β = the risk of failing to detect a difference when in fact the true proportion for the other product is $P'' = P' + \delta$ or is $P'' = P' - \delta$.
δ = the absolute magnitude of the difference considered important to detect.

Tables to be Used:

Table A-26 can be used for $\alpha = .05$ and $1 - \beta = .50, .80, .90, .95$, and $.99$; and for certain values of the proportions.

For other desired values of α and β, use Table A-27, a table to convert the difference between the proportions into the form necessary for use with Table A-9.

The question to be answered by the experiment is: Is P_A larger than P_B?

Procedure	Example
(1) Specify δ, the absolute magnitude of the difference considered important to detect. If the estimate P_A' is available, then $P'' = P_A' - \delta$. If the estimate P_B' is available, then $P'' = P_B' + \delta$.	(1) Specify $\delta = .05$ Let P_A', the estimate of P_A, $= .10$ Then, $P'' = .10 - .05$ $= .05$

Procedure	Example
(2) Choose α and β.	(2) Let $\alpha = .05$ $\beta = .10$
(3) For $\alpha = .05$, $1 - \beta = .50$, .80, .90, .95, or .99, go to Table A-26.	(3) Use Table A-26 with $1 - \beta = .90$
(4) If either P' or P'' is less than 0.5, enter Table A-26 with P' and P''. If neither P' nor P'' is less than 0.5, enter Table A-26 with $1 - P'$ and $1 - P''$. In either case, the smaller of the two proportions determines the column and the larger determines the row in Table A-26.	(4) Since both P' and P'' are less than 0.5, enter Table A-26 in column .05 and row .10.
(5) Read off n, and double it to obtain n'. n' is the required sample size to be taken from each product.	(5) $n = 232$ $n' = 464$, the required sample size to be taken from each product.
(6) For other values of α and β, and for values of P' and P'' not included in Table A-26, go to Table A-27. Look up: $\theta' = \theta$ corresponding to P' $\theta'' = \theta$ corresponding to P''	(6) Assume that we had specified $\alpha = .01$, $1 - \beta = .90$, $P' = .70$, and $P'' = .50$. From Table A-27 $\theta' = 1.98$ $\theta'' = 1.57$
(7) Compute $d = \lvert\theta' - \theta''\rvert$	(7) $d = \lvert 1.98 - 1.57 \rvert$ $= .41$
(8) Enter Table A-9 with α, β, and d (from Step (7)). Read off n, and double it to obtain n'. n' is the required sample size to be taken from each product. Rounding two-decimal values of d to the nearest value considered in Table A-9 may lead to excessively high (or low) values of n, and thence of n'. Interpolation for values of n corresponding to values of d not considered in Table A-9 is *not* recommended. For values of d not given in Table A-9, the sample size may be computed using the formula $n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{d^2}$	(8) From Table A-9, for $d = .4$, $n = 82$. $n' = 164$ is an upper bound to the required sample size to be taken from each product. In the present instance, $n = \frac{(2.326 + 1.282)^2}{(.41)^2}$ $= \frac{13.018}{.1681}$ $= 78$ and $n' = 156$ is the required sample size to be taken from each product.

CHAPTER 9

COMPARING MATERIALS OR PRODUCTS WITH RESPECT TO SEVERAL CATEGORIES OF PERFORMANCE (CHI-SQUARE TESTS)

In some inspection and testing procedures, a two-category classification of performance (e.g., success or failure) is not sufficient. In inspection work, the classification scheme might consist of *three* categories; for example, (1) acceptable, (2) reworkable, and (3) unusable. In process control, we might wish to record occurrences of each of a number of types of defects, and to make comparisons between shifts or between time periods with regard to the distribution of the types of defects. Similarly, reports of types of failures of machinery, or of records of repairs, may call for a classification scheme with more than two categories. Classifications by size, color, and structure are other possible examples of classifications likely to require three or more categories.

Where the classification scheme provides for three or more categories, the procedures of this Chapter are appropriate. (The methods of Chapter 8 could be used only if we were to consider a single class as, for example, *success*, with all the other classes lumped together as *failure*.)

If the classification scheme has a large number of categories, and if we are interested in a special group of these classes, the individual classes in the group may be combined and considered as one grand category. For example, in records of the causes of aircraft accidents, we may consider the one large category *collision*, or we may have this information broken down into several classes, e.g., *between two in air*, *with ground*, *with water*, and *other types of collision*.

9-1 COMPARING A MATERIAL OR PRODUCT WITH A STANDARD

9-1.1 WHEN THE COMPARISON IS WITH A STANDARD MATERIAL OR PRODUCT

Data Sample 9-1.1 — Inspections and Tests of Clinical Thermometers

Form: A sample of n items is selected at random from a much larger group. After inspection or test, each sample item (or observation) is classified into one of k categories, according to some established classification scheme. The result is that n_1 items are observed to be in category 1, n_2 items in category 2, n_i items in the ith category, etc., and

$$n_1 + n_2 + \ldots + n_k = n.$$

Let: P_1 equal the known proportion of *standard* items that are classified in category 1; P_2 equal the known proportion in category 2; and P_i equal the known proportion of standard items in the ith category.

The relevant question to be asked is: "Does the new product differ from the standard with regard to the proportions in each category?"

Example: The inspection and testing of clinical thermometers provides an illustrative example. Clinical thermometers are classified into one of the following four categories, on the basis of inspection and test:

1. Non-defective;
2. Defective — class A (Defects in glass, defective markings, dimensional nonconformance, etc.);
3. Defective — class B (Defects in mercury column);
4. Defective — class C (Nonconformance to precision and accuracy requirements).

Over a period of time, it has been found that thermometers produced by a certain manufacturer are distributed among the four categories in the following average proportions:

1. Non-defective — 87 percent ($P_1 = 0.87$);
2. Class A — 9 percent ($P_2 = 0.09$);
3. Class B — 3 percent ($P_3 = 0.03$);
4. Class C — 1 percent ($P_4 = 0.01$).

A new lot of 1336 thermometers is submitted by the manufacturer for inspection and test, and the following distribution into the four categories results:

Category	No. of Thermometers Reported
1	1188 (n_1)
2	91 (n_2)
3	47 (n_3)
4	10 (n_4)
	$n = 1336$

The question asked is: "Does this new lot of thermometers differ from previous experience with regard to proportions of thermometers in each category?"

Procedure*

(1) Choose α, the significance level of the test.

(2) Look up $\chi^2_{1-\alpha}$ for $k - 1$ degrees of freedom in Table A-3.

(3) Compute nP_i, the theoretical value for each category.

(4) Compute

$$\chi^2 = \sum_{i=1}^{k} (n_i^2/nP_i) - n$$

(5) If $\chi^2 \geq \chi^2_{1-\alpha}$, conclude that the material, product, or process differs from the standard with regard to the proportions in the categories; otherwise, there is no reason to believe that they differ.

Example

(1) Let $\alpha = .05$
$1 - \alpha = .95$

(2) $k - 1 = 3$
$\chi^2_{.95}$ for 3 d.f. $= 7.81$

(3) For a convenient computational arrangement, see Table 9-1.

(4) See Table 9-1,

$$\chi^2 = 9.72$$

(5) Since χ^2 is larger than $\chi^2_{.95}$, we conclude that the new lot of thermometers is different from previous lots submitted by the same manufacturer with regard to the proportions in the respective inspection-test categories.

TABLE 9-1. COMPUTATIONAL ARRANGEMENT FOR DATA SAMPLE 9-1.1

Thermometer Class	No. of Thermometers in Each Category (n_i)	Known Proportion for Standard Product (P_i)	Expected No. (nP_i)	$\frac{n_i^2}{nP_i}$
1. Non-defective	1188	.87	1162.32	1214.25
2. Class A	91	.09	120.24	68.87
3. Class B	47	.03	40.08	55.11
4. Class C	10	.01	13.36	7.49
Total	$n = 1336$	1.00	1336.00	1345.72

$$\chi^2 = \Sigma(n_i^2/nP_i) - n = 1345.72 - 1336 = 9.72$$

* This χ^2 procedure is based on a *large-sample* approximation, but if $nP_i \geq 5.0$ for all categories, the approximation ordinarily is very good. If $nP_i < 5$ for several categories, these categories may be pooled to obtain a theoretical frequency of at least 5 for the combined cells, and a corresponding improvement in the accuracy of the approximate solution, but at the price of some loss of resolution.

9-1.2 WHEN THE COMPARISON IS WITH A THEORETICAL "STANDARD"

The following example illustrates an application of the foregoing procedure in which the "standard" is of a theoretical nature.

Data Sample 9-1.2 — Breakdowns of Electricity Meters

Form: There are k different types of meters in current use. The total number of each type in service at the beginning of the service period under consideration was $N_1, N_2, \ldots, N_k$, respectively. If the probability of a meter breaking down during the service period is the same for all k types, then we would expect the total number of breakdowns during this period to be distributed among the k types in proportion to their respective numbers in service. For example, suppose that a total of n meters break down. Among these n, we would expect to find the proportion

$$P_1 = \frac{N_1}{N_1 + N_2 + \ldots + N_k}$$

of them to be of type 1; the proportion

$$P_2 = \frac{N_2}{N_1 + N_2 + \ldots + N_k}$$

to be of type 2; etc. The actual number of each type that are found in the n breakdowns are $n_1, n_2, \ldots, n_k$ (and $n_1 + n_2 + \ldots + n_k = n$). The actual proportions $n_1/n, n_2/n, \ldots, n_k/n$, rarely will conform to the theoretical values, $P_1, P_2, \ldots, P_k$, even when the hypothesis that all types are equally likely to break down is true. The relevant question is: Are the differences between the observed and theoretical proportions sufficient to cast doubt on the supposition that the probability of a breakdown is the same for all k types of meters?

Example: There are (approximately) equal numbers of four different types of meters in service. If all types are equally likely to break down, the reported failures during a given period should be distributed (approximately) equally among the four types — i.e.,

$$P_1 = P_2 = P_3 = P_4 = \frac{1}{4}.$$

The actual number of breakdowns reported are given in the following list. Have we evidence to conclude that the chances of failure of the four types are not all equal?

Type of Meter	No. of Breakdowns Reported n_i
1	30
2	40
3	33
4	47
	$n = 150$

Procedure	Example
(1) Choose α, the significance level of the test.	(1) Let $\alpha = .10$ $1 - \alpha = .90$
(2) Look up $\chi^2_{1-\alpha}$ for $k - 1$ degrees of freedom in Table A-3.	(2) $k - 1 = 3$ $\chi^2_{.90}$ for 3 d.f. $= 6.25$
(3) Compute nP_i, the theoretical frequency for each category.	(3) For a convenient computational arrangement, see Table 9-2.
(4) Compute $\chi^2 = \sum_{i=1}^{k} (n_i^2/nP_i) - n$	(4) See Table 9-2, $\chi^2 = 4.62$
(5) If $\chi^2 \geq \chi^2_{1-\alpha}$, conclude that the probabilities of failure are not the same for all of the types*; otherwise, that there is no reason to believe that they differ.	(5) Since χ^2 is not larger than $\chi^2_{.90}$, we have no reason to discard the hypothesis that the probability of failure is the same for each type.

TABLE 9-2. COMPUTATIONAL ARRANGEMENT FOR DATA SAMPLE 9-1.2

Type of Meter	No. of Breakdowns Reported (n_i)	Expected No. of Breakdowns (nP_i)	$\frac{n_i^2}{nP_i}$
1	30	37.5	24.00
2	40	37.5	42.67
3	33	37.5	29.04
4	47	37.5	58.91
Total	$n = 150$		154.62

P_i = the theoretical proportion for each category. In Data Sample 9-1.2, $P_i = .25$ for all categories.
$\chi^2 = \Sigma(n_i^2/nP_i) - n = 154.62 - 150 = 4.62$.

* In reaching this conclusion on the basis of evidence that the P's are not all equal to their theoretical values, we are assuming, of course, that our information on the numbers of meters of each type in service is correct. In practice, this assumption should be checked before accepting the conclusion that the probabilities of failure are not the same for all of the types.

9-2 COMPARING TWO OR MORE MATERIALS OR PRODUCTS

Data Sample 9-2 — Causes of Rejection of Metal Castings

Form: There are m things to be compared with regard to the percentage of items distributed among several categories. The m things may be materials, products, processes, shifts, time periods (days, weeks, months, etc.) or any other such groups of interest. From each of the m groups, a sample is available, and each item in the sample is classified into one of k categories. The data is tabulated conveniently in the following form:

Material, Product, or Process	Category				Total
	1	2	. . .	k	
1	f_{11}	f_{12}	. . .	f_{1k}	n_1
2	f_{21}	f_{22}	. . .	f_{2k}	n_2
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
m	f_{m1}	f_{m2}	. . .	f_{mk}	n_m
Total	C_1	C_2	. . .	C_k	n

where:

m = number of materials, products, processes, etc., to be compared;
k = number of categories of classification;
n_i = size of sample for the ith material, product, or process;
f_{ij} = number of items of the ith kind which are classified in the jth category;
C_j = total number in the jth category;
n = total number of items.

The relevant question to be asked is: "Do the materials, products, etc., *differ* with regard to the proportion of items in the categories?"

Example: Rejects of metal castings were classified by cause of rejection for three different weeks, as given in the following tabulation. The question to be answered is: Does the distribution of rejects differ from week to week?

	Cause of Rejection							
	Sand	Misrun	Shift	Drop	Corebreak	Broken	Other	Total
Week 1	97	8	18	8	23	21	5	180
Week 2	120	15	12	13	21	17	15	213
Week 3	82	4	0	12	38	25	19	180
Total	299	27	30	33	82	63	39	573

Data adapted with permission from *Industrial Quality Control*, Vol. IV, No. 4, p. 26, 1948, from article entitled "A Training Program Becomes a Clinic," by George A. Hunt.

Procedure*

(1) Choose α, the significance level of the test.

(2) Look up $\chi^2_{1-\alpha}$ for $(k-1)(m-1)$ degrees of freedom in Table A-3.

(3) Compute

$$\chi^2 = n\left(\sum_{i=1}^{m}\sum_{j=1}^{k}\frac{f_{ij}^2}{n_iC_j} - 1\right)$$

That is, compute each column total C_j. Compute each row total n_i. For each cell in the table, square the number f_{ij} and divide by the product n_iC_j. Sum the resulting values for all cells in the table; subtract one, and multiply by n.

(4) If $\chi^2 \geq \chi^2_{1-\alpha}$, decide that the materials, products, or processes differ with regard to the proportions in the categories; otherwise, that there is no reason to believe that they differ in this regard.

Example

(1) Let $\alpha = .10$
$1 - \alpha = .90$

(2) $k = 7$
$m = 3$
$(k-1)(m-1) = 12$
$\chi^2_{.90}$ for 12 d.f. $= 18.55$

(3) See Table 9-3 for a convenient computational arrangement.
$\chi^2 = 45.84$

(4) Since χ^2 is larger than $\chi^2_{.90}$, we conclude that the weeks differ with regard to proportions of various types of rejections.

TABLE 9-3. TABLE OF $\frac{f_{ij}^2}{n_iC_j}$ — COMPUTATIONAL ARRANGEMENT FOR DATA SAMPLE 9-2

	Sand	Misrun	Shift	Drop	Corebreak	Broken	Other	Total
Week 1	$\frac{9409}{53820}$ = 0.175	$\frac{64}{4860}$ = 0.013	$\frac{324}{5400}$ = 0.060	$\frac{64}{5940}$ = 0.011	$\frac{529}{14760}$ = 0.036	$\frac{441}{11340}$ = 0.039	$\frac{25}{7020}$ = 0.004	0.338
Week 2	$\frac{14400}{63687}$ = 0.226	$\frac{225}{5751}$ = 0.039	$\frac{144}{6390}$ = 0.023	$\frac{169}{7029}$ = 0.024	$\frac{441}{17466}$ = 0.025	$\frac{289}{13419}$ = 0.022	$\frac{225}{8307}$ = 0.027	0.386
Week 3	$\frac{6724}{53820}$ = 0.125	$\frac{16}{4860}$ = 0.003	0	$\frac{144}{5940}$ = 0.024	$\frac{1444}{14760}$ = 0.098	$\frac{625}{11340}$ = 0.055	$\frac{361}{7020}$ = 0.051	0.356
Total	0.526	0.055	0.083	0.059	0.159	0.116	0.082	1.080

$$\chi^2 = n\left(\sum_{i=1}^{m}\sum_{j=1}^{k}\frac{f_{ij}^2}{n_iC_j} - 1\right)$$
$$= 573\,(1.080 - 1)$$
$$= 573\,(.080)$$
$$= 45.84$$

* The solution is approximate, but should be quite accurate if the smallest $n_iC_j/n \geq 5$.

Simplified Computation for the Special Case $m = 2$. In this case, the tabulation would consist of only the first two rows of the schematic table shown in Data Sample 9-2, and

$$\chi^2 = \sum_{j=1}^{k} \left[\frac{n_1 n_2}{f_{1j} + f_{2j}} \left(\frac{f_{1j}}{n_1} - \frac{f_{2j}}{n_2} \right)^2 \right].$$

The degrees of freedom for χ^2 is $k - 1$.

This form is convenient if the data are given in terms of proportions.

Further Simplification for $m = 2$ When $n_1 = n_2$. When there are only two rows, and the row totals are equal, then,

$$\chi^2 = \sum_{j=1}^{k} \frac{(f_{1j} - f_{2j})^2}{f_{1j} + f_{2j}}$$

with $k - 1$ degrees of freedom.

Note: This shortcut has an analog for $m = 3$ when $n_1 = n_2 = n_3$. For each category, take all three possible differences, sum the squares of the three differences, and divide by the sum of the three observations. Finally, sum this quantity over all of the categories, to obtain

$$\chi^2 = \sum_{j=1}^{k} \frac{(f_{1j} - f_{2j})^2 + (f_{1j} - f_{3j})^2 + (f_{2j} - f_{3j})^2}{f_{1j} + f_{2j} + f_{3j}}.$$

9-3 A TEST OF ASSOCIATION BETWEEN TWO METHODS OF CLASSIFICATION

There are situations in which individual items are classified into categories in terms of two different criteria. For example, in a study of tire wear, see Swan,[1] records of scrapping of tires were kept and tires were classified as front and rear, left and right. In another study of the cause of failure of vacuum tubes, see Day,[2] the two criteria of classification were position in shell and type of failure. In each study the question was: Is there any association or relation between the criteria of classification?

Basically, this is a different problem than the problem of Paragraph 9-2, but it is discussed here because of the similarity in analysis.

We assume that we have a total of n individual items, and that each item is classified by criteria A and B into k and m categories, respectively. Let f_{ij} be the number of individuals in the ith category of A *and* the jth category of B. Let R_i and C_j be the total numbers of individuals classified in the ith category of A and the jth category of B, respectively.

We tabulate these data as follows:

		Criterion B				Total
		1	2	...	k	
Criterion A	1	f_{11}	f_{12}	...	f_{1k}	R_1
	2	f_{21}	f_{22}	...	f_{2k}	R_2
		.	.		.	.
		.	.		.	.
		.	.		.	.
	m	f_{m1}	f_{m2}	...	f_{mk}	R_m
Total		C_1	C_2	...	C_k	n

The relevant question to be asked is: "Is there a relation or association between the two criteria of classification?"

Data Sample 9-3 — Vacuum Tube Failures

In the development of the VT fuze during World War II, a study was made of the causes of failure of vacuum tubes. The criteria of classification were: position in shell and type of failure. The following entries are the number of tubes that failed.

Position in Shell	Type of Failure			Total
	A	B	C	
Top block	75	10	15	$100 = R_1$
Bottom block	40	30	10	$80 = R_2$
Total	115	40	25	$180 = n$

Adapted with permission from *Review of the International Statistical Institute*, Vol. 17, Nos. 3 and 4, pp. 129-155, 1949, from article entitled "Application of Statistical Methods to Research and Development in Engineering" by Besse B. Day.

The question to be asked is: Is the type of tube failure associated with the position in the shell?

Procedure*	**Example**
(1) Choose α, the level of significance of the test.	(1) Let $\alpha = .10$ $1 - \alpha = .90$
(2) Look up $\chi^2_{1-\alpha}$ for $(k-1)(m-1)$ degrees of freedom in Table A-3.	(2) $k = 3$ $m = 2$ $(k-1)(m-1) = (2)(1)$ $= 2$ $\chi^2_{.90}$ for 2 d.f. $= 4.61$

* The solution is approximate, but should be quite accurate if the smallest of $R_iC_j/n \geq 5.0$.

Procedure

(3) Compute

$$\chi^2 = n\left(\sum_{i=1}^{m}\sum_{j=1}^{k}\frac{f_{ij}^2}{R_iC_j} - 1\right)$$

That is, compute each row total R_i. Compute each column total C_j. For each cell in the table, square the number f_{ij} and divide by the product R_iC_j. Sum the resulting values for all cells in the table; subtract one, and multiply by n.

(4) If $\chi^2 \geq \chi^2_{1-\alpha}$, conclude that there is an association between the two criteria of classification; otherwise, that there is no reason to believe that such an association exists.

Example

(3) See Table 9-4 for a computational arrangement.

$$\chi^2 = 180\,(1.1092 - 1) = 19.66$$

(4) Since χ^2 is greater than $\chi^2_{.90}$, we conclude that the type of failure is associated with the position in the shell.

TABLE 9-4. TABLE OF $\frac{f_{ij}^2}{R_iC_j}$ — COMPUTATIONAL ARRANGEMENT FOR DATA SAMPLE 9-3

Position in Shell	Type of Failure			Total
	A	B	C	
Top block	5625/11500 = .4891	100/4000 = .0250	225/2500 = .0900	.6041
Bottom block	1600/9200 = .1739	900/3200 = .2812	100/2000 = .0500	.5051
Total	.6630	.3062	.1400	1.1092

$$\begin{aligned}\chi^2 &= n\left(\sum_{i=1}^{m}\sum_{j=1}^{k}\frac{f_{ij}^2}{R_iC_j} - 1\right)\\ &= 180\,(1.1092 - 1)\\ &= 180\,(.1092)\\ &= 19.66\end{aligned}$$

REFERENCES

1. A. W. Swan, "The χ^2 Significance Test — Expected vs. Observed Results," *The Engineer*, Vol. 186, No. 4849, p. 679, December 31, 1948.
2. B. B. Day, "Application of Statistical Methods to Research and Development in Engineering," *Review of the International Statistical Institute*, Vol. 17, Nos. 3 and 4, pp. 129-155, 1949.

CHAPTER 10

SENSITIVITY TESTING

10-1 EXPERIMENTAL SITUATION

The term "sensitivity test" is commonly applied to the following situation:

1. A test item will *respond* or *not respond* to a certain level of test stimulus (e.g., a shell will explode or will not explode when subjected to a certain shock).
2. The test is *destructive* to the item being tested, no matter what the outcome of the test. Either the item is destroyed completely, or the characteristics of the item are so changed that further tests are meaningless.
3. The percentage of items expected to *respond* (fail, explode, die) increases as the severity of the test is increased.

In this general situation, there are variable (and usually controllable) levels of test which can be applied; e.g., height of drop in a shock test, dosage of a poison in tests of insecticides, etc. We assume that each object has an associated critical level or threshold value. If the test stimulus applied equals or exceeds this critical level, the object responds (fails, explodes, dies). If the test stimulus applied does not equal or exceed this critical level, the object does not respond. For any particular object, the *exact* critical level cannot be determined. More than one object may be tested at a given test level, however, and inferences may be made about the distribution of critical levels in a population of objects from which the tested samples came.

The experimenter obtains data of the following type: objects were tested at the k stimulus levels $x_1, x_2, \ldots, x_k$; of the n_i objects tested at level x_i, r_i responded and $n_i - r_i$ did not respond.

An ordnance example might involve the detonation of samples of an explosive in powder form by dropping a specified weight on them from various heights. If the weight is dropped from a height below the sample's critical level (in this case, the lowest height at which the weight will cause the sample to explode), the sample does not explode, but the powder may be packed more tightly than before and, therefore, the test cannot be repeated at increased height. If the weight is dropped from above the sample's critical level, the sample is destroyed.

A partial list of the many ordnance problem areas in which tests of increased severity can be used is as follows.

1. Sensitivity to mechanical shock:
   a. Impact tests of high explosives;
   b. Impact tests of artillery fuzes;
   c. Izod impact test of metals;
   d. Izod impact test of plastics;
   e. Impact or drop test of packing cases.
2. High explosives sensitivity to setback pressures.
3. Missile components sensitivity to acceleration.
4. Explosives sensitivity to friction.
5. Fuzes and explosives sensitivity to velocity.
6. Artillery fuzes and missile components sensitivity to voltage.
7. Pyrotechnic materials sensitivity to electric spark.
8. Explosives and missile components sensitivity to temperature.

The following ordnance example is used to illustrate the computational techniques.

Data Sample 10-1 — Peak-Voltage Test of Fuzes

Groups of fuzes are subjected to specified values of peak voltage.

For each group we observe the number which fire in less than a specified time. The observations are summarized as follows:

Peak Voltage	n	r	$p = r/n$
10.0	12	0	0
15.0	12	0	0
17.5	12	1	.08
20.0	13	2	.15
22.5	10	3	.30
25.0	13	6	.46
30.0	12	8	.67
35.0	13	9	.69
40.0	13	11	.85
50.0	11	10	.91
60.0	11	11	1.00

Several methods for collecting and analyzing such data are described and illustrated in this Chapter. Paragraphs 10-2 and 10-3 detail methods of analysis for the usual testing situation, where the levels of test are assigned before the test begins. Paragraph 10-4 details a special method applicable when the test levels can be different for each successive object tested, and can be changed easily during the course of the experiment. Paragraph 10-5 discusses the situation where the test levels cannot be completely controlled.

Most of the methods involve assumptions about the distribution of the critical levels such as, for example, that the distribution is *normal*. The distribution of critical levels as measured in the original units (or *natural* units) may not meet these assumptions, but there may exist a transformation such that the distribution of the transformed values does meet the assumption. The logarithm of the original value is perhaps the most frequently used transformation.

When a transformation is used, all comments on the selection of testing levels and all computational instructions refer to the transformed values, not the original ones. Usually, however, it is desirable to state the final results of the analysis in terms of the original units. For most transformations (including the logarithmic), the percentile estimates* and their associated confidence intervals are converted into the original units easily. Suppose that the stimulus levels are originally measured in "y" units, and transformed values, e.g., $x = \log y$, are used in the computations. If $\hat{x}_P$ is an estimate of x_P, as here defined,* and $[a, b]$ is a $1 - \alpha$ confidence interval estimate of x_P, then antilog $\hat{x}_P$ and [antilog a, antilog b] give, respectively, a point estimate and a $1 - \alpha$ confidence interval estimate of y_P, the (100 P)th percentile of the distribution of y values. This relationship does *not* hold for the means and standard deviations of the distributions. If m and s are estimates of the mean and standard deviation of the distribution of the x values, antilog m and antilog s should not be considered estimates of the (arithmetic) mean and standard deviation of the distributions of y's.†

Note: In this Chapter, *normal* means that the expected proportion of items responding at a stimulus level x is given by

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

i.e., the probability that an individual item has a critical level or threshold value $\leq x$.

* For any random variable X, the "(100 P)th percentile" of its distribution is the value x_P such that the probability that X is $\leq x_P$ is equal to P.

† If $x = \log y$, then

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} \log y_i$$

$$= \log (y_1 y_2 \ldots y_n)^{\frac{1}{n}}$$

$$= \log (\text{geometric mean of the } y\text{'s}).$$

Hence, antilog m will be an estimate of the *geometric mean* of the distribution of the y's. Since the geometric mean of a set of different numbers always is less than their arithmetic mean, it follows that antilog m will tend to underestimate the population mean of the y's.

10-2 KÄRBER METHOD OF ANALYSIS

This method gives simple nonparametric estimates of the mean and standard deviation of the distribution of critical levels. There are three situations for which this method might be chosen:

(1) the shape or mathematical form of the distribution is unknown;

(2) quick and easy procedures for routine laboratory calculations are desired; and,

(3) good *initial estimates* or *first approximations* are needed for iterative computational procedures; for example, the Exact Probit Solution given in Paragraph 10-3.2.

The Kärber method provides very good estimates of the mean and standard deviation of the distribution of critical levels in most laboratory situations. It must be remembered that the Kärber method gives an estimate of the mean of the distribution; the mean of the distribution is not equal to the 50th percentile (that level x_0 such that half the objects have critical levels less than x_0 and half have critical levels greater than x_0) unless it is known that the distribution is symmetrical about its mean.

For further discussion of the Kärber method, see Cornfield and Mantel.[1]

Selection of Stimulus Levels. Order the stimulus levels to be used in the test by their magnitude — in other words, let $x_1 < x_2 < \ldots < x_k$. For the Kärber method to be applicable, x_1 must be sufficiently low that there are no responses among the objects tested ($r_1 = 0$), and x_k must be sufficiently high that all objects tested respond ($r_k = n_k$). In other words, x_1 and x_k are to be chosen so that they are likely to cover the entire range of critical levels in the population. In addition, it is preferable to have more (and consequently, more closely spaced) test levels with fewer objects tested at each level than to have only a few test levels and a large number of objects tested at each level.

Note: The following segments of Chapter 10 are paged and spaced as necessary to allow a facing-page arrangement of Procedures and Examples. Thus, Procedure steps appear on left-hand pages, and their associated Example steps appear on facing right-hand pages. The only exception is in Paragraph 10-3.3, where both the Procedure and Example are complete on the same page.

10-2.1 GENERAL SOLUTION FOR THE KÄRBER METHOD

10-2.1.1 Procedure. Prepare a table with nine columns headed x, n, r, p, $(p_{i+1} - p_i)$, a, a^2, d, and d^2. (See Table 10-1.)

(1) In column 1, enter the stimulus levels used in the test from lowest to highest — that is, enter $x_1, x_2, \ldots, x_k$ where $x_1 < x_2 < \ldots < x_k$.

(2) In columns 2 and 3, for each x_i, enter the corresponding n_i (number of objects tested at that level) and r_i (number of responses).

(3) Corresponding to each x, compute $p_i = \frac{r_i}{n_i}$ and enter this in column 4. (Remember that p_1 must equal 0 and p_k must equal 1 if this solution is to be used.)

(4) Corresponding to each x_i (except x_k), compute $p_{i+1} - p_i$ and enter in column 5. There is no entry in this column corresponding to x_k.

(5) Corresponding to each x_i (except x_k), compute $a_i = \frac{x_{i+1} + x_i}{2}$, the midpoint of the interval from x_i to x_{i+1}. Tabulate the a_i values in column 6. There is no entry in this column corresponding to x_k.

(6) Corresponding to each a_i, enter a_i^2 in column 7.

(7) Corresponding to each x_i (except x_k), compute $d_i = x_{i+1} - x_i$, the length of the interval from x_i to x_{i+1}. Tabulate the d_i values in column 8. There is no entry in this column corresponding to x_k.

(8) Corresponding to each d_i, enter d_i^2 in column 9.

(9) Compute $m = \Sigma(p_{i+1} - p_i)a_i$, the sum of products of corresponding entries in the 5th and 6th columns. m is our estimate of the mean of the distribution of critical levels.

(10) Compute:

$S_1 = \Sigma(p_{i+1} - p_i)a_i^2$, the sum of products of corresponding entries in the 5th and 7th columns;

$S_2 = \Sigma(p_{i+1} - p_i)d_i^2$, the sum of products of corresponding entries in the 5th and 9th columns;

$$s^2 = S_1 - m^2 - \frac{S_2}{12}.$$

s is our estimate of the standard deviation of the distribution of critical levels.

10-2.1.2 Example. The observations of Data Sample 10-1 are used to illustrate the Procedure. The problem is to estimate the mean and standard deviation of the distribution of critical values of peak voltage. In this example, we assume that nothing is known about the distribution of critical levels, and the computations are performed in natural units. If the distribution of critical levels were known or presumed to be log-normal, and if the Kärber method were being used to provide a quick and easy answer, or to provide initial estimates for the exact probit solution (Paragraph 10-3.2), then the computations would have been performed on the transformed values x = log (peak voltage).

The entries and calculations of Steps (1) through (8) of the Procedure are shown in columns 1 through 9 of Table 10-1. The calculations of Steps (9) and (10) are shown at the bottom of Table 10-1.

TABLE 10-1. KÄRBER METHOD OF ANALYSIS FOR FUZE PEAK VOLTAGE TEST DATA (SEE DATA SAMPLE 10-1)

Level of Stimulus, Peak Voltage x	Number of Objects Tested n	Number of Objects Responding r	$r/n = p$	$p_{i+1} - p_i$	a	a^2	d	d^2
Col. (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
10.0	12	0	0	0	12.50	156.25	5.0	25.00
15.0	12	0	0	.08	16.25	264.06	2.5	6.25
17.5	12	1	.08	.07	18.75	351.56	2.5	6.25
20.0	13	2	.15	.15	21.25	451.56	2.5	6.25
22.5	10	3	.30	.16	23.75	564.06	2.5	6.25
25.0	13	6	.46	.21	27.50	756.25	5.0	25.00
30.0	12	8	.67	.02	32.50	1056.25	5.0	25.00
35.0	13	9	.69	.16	37.50	1406.25	5.0	25.00
40.0	13	11	.85	.06	45.00	2025.00	10.0	100.00
50.0	11	10	.91	.09	55.00	3025.00	10.0	100.00
60.0	11	11	1.00					

Step (9): $m = \Sigma (p_{i+1} - p_i)\, a_i$
$= 29.68$
= mean value of critical peak voltage.

Step (10): $S_1 = \Sigma (p_{i+1} - p_i)\, a_i^2$
$= 1002.4051$

$S_2 = \Sigma (p_{i+1} - p_i)\, d_i^2$
$= 27.6250$

$s^2 = S_1 - m^2 - (S_2/12)$
$= 1002.4051 - 880.9024 - 2.3021$
$= 119.2006$

$s = 10.92$
= estimated standard deviation of critical peak voltage.

10-2.2 SIMPLIFIED SOLUTION (KÄRBER METHOD) FOR THE SPECIAL CASE WHEN TEST LEVELS ARE EQUALLY SPACED AND EQUAL NUMBERS OF ITEMS ARE TESTED AT EACH LEVEL

10-2.2.1 Procedure. Prepare a table with columns headed x, r, p, and cumulative p. (See Table 10-2.)

(1) In the x column, enter the test levels from lowest to highest, i.e., enter $x_1, x_2, \ldots, x_k$, where $x_1 < x_2 < \ldots < x_k$.

(2) Enter r_i, the number of objects responding at each x_i.

(3) Corresponding to each x_i, compute $p_i = r_i/n$. (n is the number of objects tested at each level and is the same for all levels). Remember that p_1 must equal 0 and p_k must equal 1 for this solution to be used.

(4) In the last column, enter the cumulative p, i.e., at x_i, the sum of all p up to and including p_i.

(5) Let: x_k = highest test level.
d = interval between successive test levels.
S_1 = sum of column p.
S_2 = sum of column cumulative p.

(6) Compute $m = x_k - d(S_1 - \frac{1}{2})$.

(7) Compute $s^2 = d^2(2S_2 - S_1 - S_1^2 - \frac{1}{12})$.

10-2.2.2 Example. In order to demonstrate the computing procedure, the fuze data (Data Sample 10-1) have been changed arbitrarily, to have equal spacing in x and equal numbers of objects at each level. Assume that 12 objects were tested at each level, and that the responses were as shown in Table 10-2. Steps (1) through (4) of the Procedure consist of preparing and filling out the four columns of Table 10-2. Steps (5) through (7) are shown at the bottom of Table 10-2.

TABLE 10-2. SIMPLIFIED SOLUTION FOR THE KÄRBER METHOD OF ANALYSIS WHEN THE TEST LEVELS (x) ARE EQUALLY SPACED AND EQUAL NUMBERS OF OBJECTS (n) ARE TESTED AT EACH LEVEL

Peak Voltage x	Number of Objects Responding r	Proportion of Objects Responding $p = r/12$	Cumulative Proportion
10.0	0	0	0
15.0	0	0	0
20.0	2	.17	.17
25.0	6	.50	.67
30.0	8	.67	1.34
35.0	9	.75	2.09
40.0	10	.83	2.92
45.0	10	.83	3.75
50.0	11	.92	4.67
55.0	12	1.00	5.67
		$S_1 = 5.67$	$S_2 = 21.28$

Step (5) x_k = highest test level
= 55
d = interval in x
= 5
S_1 = sum of p
= 5.67
S_2 = sum of cumulative p
= 21.28

Step (6) $m = x_k - d\,(S_1 - \frac{1}{2})$
$= 55 - 5\,(5.67 - .50)$
$= 29.15$
= mean value of critical peak voltage.

Step (7) $s^2 = d^2\,(2S_2 - S_1 - S_1^2 - \frac{1}{12})$
$= 25\,(42.56 - 5.67 - 32.15 - .08)$
$= 116.50$
$s = 10.79$
= estimated standard deviation of critical peak voltage.

10-3 PROBIT METHOD OF ANALYSIS

When it is assumed that the distribution of critical levels is of a particular type, methods for estimating the properties of the distribution are tailored to it. There are three types of distributions for sensitivity data that have been studied extensively in the statistical literature — the *normal*, the *logistic*, and the *angular*. Only the *normal* will be discussed in detail here because it is most frequently used in ordnance sensitivity testing. For the *logistic*, see Berkson[2] and Hodges,[3] for the *angular*, see Knudsen and Curtis[4] and Fisher and Yates.[5] When the stimulus levels used in the test are between the levels which cut off the lower and upper 10% of the distribution (most testing is performed in this range), any one of these types will fit the data nearly as well (or as poorly) as another, no matter what the true distribution of critical levels is. However, estimates of the parameters of the *logistic* or the *angular* distributions involve simpler computations than are given here for the *normal;* for example, see the technique described in Hodges[3] for the logistic curve.

The procedures described here assume that the distribution of critical values is *normal* — that is, for all x the proportion of objects which have critical levels between x and $x + dx$ is equal to the area of some *normal* curve between x and $x + dx$. In general, the procedures are not very sensitive to moderate departures from normality, provided one does not extrapolate beyond the range of the data.

The problem may be summarized as follows: k different levels $x_1, x_2, \ldots, x_k$ of a stimulus are applied to $n_1, n_2, \ldots, n_k$ objects, with $r_1, r_2, \ldots, r_k$ responses, respectively. Let $p_i = r_i/n_i$. The questions to be answered are:

(1) At what level m of the stimulus would half of similar objects be expected to respond? Or equivalently, under the assumption of normality, what is the mean of the critical levels of all such objects?

(2) Estimate the relation between the level of the stimulus and the proportion of objects responding.

Selection of Stimulus Levels. There are no simple cut-and-dried rules. A general guide can be given in terms of the purpose of the experiment:

(1) If the experimenter is interested in estimating a specific percentage point, the stimulus levels to be used in the test should be fairly close to that point, and should bracket the point. It is pertinent here to emphasize that extrapolations may lead to serious error, particularly if the experimenter attempts to estimate an extreme percentage point (say, the 5% or 99% point) from observations at stimulus levels which all lie to one side of that point.

(2) The test levels should cover a range sufficiently wide so that the proportion responding p_i varies from near 0 to near 1, if:

(a) One is interested in the relation between stimulus level and percentage response over the entire range; or,

(b) One is interested in estimating the standard deviation of the distribution of critical levels (or equivalently, the slope of the regression line in the probit solution); or,

(c) One is interested in testing the assumption of *normality*.

Basis of the Method. If the critical levels are normally distributed (with unknown mean μ and standard deviation σ), then Y_i', which can be determined from tables of the normal integral, is a linear function of the corresponding stimulus level x_i.

Y_i' is determined from:

$$p_i = \int_{-\infty}^{Y_i'-5} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy .$$

The "5" in the upper limit of the integral is introduced for computational convenience.

Least squares procedures are used to estimate the best straight line passing through the k points (Y_i', x_i). The formulas take account of the fact that the points do not have equal weights. The line will be expressed as

$$Y_p = 5 + b(x - m)$$

where m and b are estimated from the data, x is the stimulus level, and Y_p is related to p, the probability that an object's critical level is $\leq x$, by the formula

$$\int_{-\infty}^{Y_p-5} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy = p .$$

m is an estimate of μ (the mean and 50th percentile of the underlying *normal* distribution), and b is an estimate of $1/\sigma$ (the reciprocal of the standard deviation of the critical values).

Solutions Described. We describe two methods of solution: the graphical probit method in Paragraph 10-3.1; and the computational (*exact*) probit method in Paragraph 10-3.2. The graphical method is much simpler, and is sufficiently precise for many purposes. When a more accurate solution is desired, the graphical method furnishes a first approximation for the exact probit method.

10-3.1 GRAPHICAL PROBIT SOLUTION

10-3.1.1 Procedure. To facilitate the calculations, prepare a table with nine columns, headed respectively, x, n, r, p, Y', Y, W, nW, nWx. (See Table 10-3.)

(1) In column x, enter the levels $x_1, x_2, \ldots, x_k$ used in the experiment.

(2) In columns n and r, record the values of n_i and r_i, corresponding to x_i.

(3) In column p, compute the respective proportions responding, $p_i = r_i/n_i$.

(4) Use Table A-2 to obtain z_p corresponding to p. In column Y', enter $Y' = z_p + 5$ corresponding to each p.

(5) Plot Y' as ordinate against x as abscissa, on ordinary rectangular coordinate graph paper. See Figure 10-1. If probit paper* is available, it can be used, and the column Y' is omitted from the table. The percentages responding at each x (% response = 100 p_i) are plotted on probit paper, using the left vertical scale; the right vertical scale gives the corresponding z_p, and hence the corresponding Y minus 5.

(6) Whichever graph paper is used, draw a straight line by eye to fit the k points. Only vertical deviations from the line are to be considered in fitting; and points for which the value of Y' is outside the interval 2.5 to 7.5 may almost be disregarded unless n_i for those points is much larger than for points inside the interval. (Points outside this interval are beyond the range of the probit paper).

(7) For each value of x plotted on the graph, read the ordinate Y of the line (on probit scale if on probit paper), and record the values in the Y column of the table.†

* "Probit paper" also is called "normal deviate paper", "normal ruling", etc.

† Instead of reading the Y values from the graph, we may complete step (8), and use the equation thus obtained to compute Y.

10-3.1.2 Example. The observations of Data Sample 10-1 were plotted on normal probability paper, using both x = peak voltage and $x = \log_{10}$ (peak voltage), as abscissa. The data plot more nearly as a straight line when transformed using the equation $x = \log_{10}$ (peak voltage); therefore, all of the probit method calculations are done on the transformed variable. (See Table 10-3.)

Steps (1) through (4) of the Procedure result in filling in the first 5 columns of Table 10-3 (through Y').

TABLE 10-3. GRAPHICAL PROBIT SOLUTION USING DATA SAMPLE 10-1

Level of Stimulus = $\log_{10}$peak Voltage x	Number of Objects Tested n	Number of Objects Responding r	Proportion of Objects Responding $p = r/n$	Y'	Y	W	nW	nWx
1.0000	12	0	0	—	2.0	.015	0.180	0.180
1.1761	12	0	0	—	3.2	.180	2.160	2.540
1.2430	12	1	.08	3.59	3.7	.336	4.032	5.012
1.3010	13	2	.15	3.96	4.1	.471	6.123	7.966
1.3522	10	3	.30	4.48	4.5	.581	5.810	7.856
1.3979	13	6	.46	4.90	4.8	.627	8.151	11.394
1.4771	12	8	.67	5.44	5.3	.616	7.392	10.919
1.5441	13	9	.69	5.50	5.8	.503	6.539	10.097
1.6021	13	11	.85	6.04	6.2	.370	4.810	7.706
1.6990	11	10	.91	6.34	6.9	.154	1.694	2.878
1.7782	11	11	1.00	—	7.4	.062	0.682	1.213

$\Sigma nW = 47.573$
$\Sigma nWx = 67.761$

Steps (5) through (7) — Plot Y' against x on ordinary graph paper, as shown in Figure 10-1. A straight line is fitted to the plotted points by eye. The ordinate Y of the line is read off at each observed x, and is entered in the Y column of Table 10-3.*

* Or, complete step (8), and use the equation $Y = 5 + b(x - m)$ to obtain the Y values.

10-3.1.1 Procedure (Cont)

Note: Before proceeding further, it is often desirable to determine whether the line is an adequate representation of the data. The procedure given in Paragraph 10-3.3 may be used for this purpose.

(8) Calculate b, the slope of the fitted line, as the increase in Y for a unit increase in x. Mark two convenient points, c and d, on the line. Read off the corresponding values for both x and Y.

$x_c =$ $\qquad$ $Y_c =$

$x_d =$ $\qquad$ $Y_d =$

Then $b = \dfrac{Y_d - Y_c}{x_d - x_c}$. Read off m, the value of x corresponding to $Y = 5$. (Probit = 5, on probit paper). We may then write the equation of the line as $Y = 5 + b(x - m)$.

(9) The relation between a given level of stimulus x' and the proportion of individuals responding p' is estimated by the relationship $z_{p'} = b(x' - m)$ where the p corresponding to $z_{p'}$ is given in Table A-1. (Or, this relationship can be read directly from the straight line drawn on probit paper.)

10-3.1.2 Example (Cont)

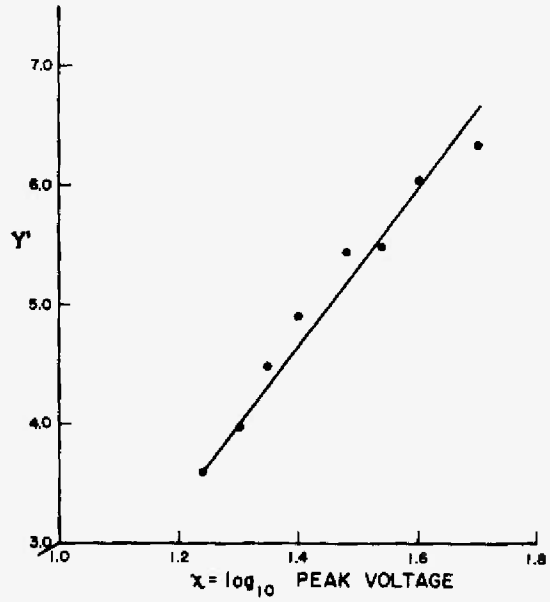


Figure 10-1. Probit regression line (fitted by eye).

Step (8) — The slope b of the line is calculated as follows:

$$x_c = 1.3 \qquad Y_c = 4.0$$
$$x_d = 1.6 \qquad Y_d = 6.0$$

$$b = \frac{Y_d - Y_c}{x_d - x_c} = \frac{6.0 - 4.0}{0.3} = 7$$

$$m \text{ (the value of } x \text{ at } Y = 5) = 1.43.$$

Equation of the line:

$$\begin{aligned} Y &= 5 + b(x - m) \\ &= 5 + 7x - 10.01 \\ &= -5.01 + 7x. \end{aligned}$$

Step (9) — If we wish to estimate the proportion of individuals responding at a peak voltage = 18, for example, then:

$$\begin{aligned} x' &= \log_{10} 18 \\ &= 1.255 \\ z_{p'} &= b(x' - m) \\ &= 7(1.255 - 1.43) \\ &= 8.785 - 10.01 \\ &= -1.22 \end{aligned}$$

$p' = .11$, the proportion which may be expected to respond at $PV = 18$.

10-3.1.1 Procedure (Cont)

m is our estimate of μ, the stimulus level at which we would expect half of similar individuals to respond. A $(1 - \alpha)$ confidence interval estimate of μ may be computed as follows:

(10) Look up $z_{1-\alpha/2}$ in Table A-2.

(11) In the column W, corresponding to each value of Y, enter the value of W obtained from Table A-18.

(12) Corresponding to each value of x, compute nW and nWx, and enter them in the last columns of the table.

(13) Compute $\Sigma n_i W_i$ and $\Sigma n_i W_i x_i$, the totals for the last two columns.

(14) Compute $s_m^2 = \frac{1}{b^2}\left(\frac{1}{\Sigma nW}\right)$, the estimate of the variance of m. This estimate is slightly too small on the average, but the bias is negligible provided that $\bar{x} = \Sigma nWx/\Sigma nW$ is approximately equal to m. If the two differ considerably, then the quantity

$$S_{xx} = \Sigma n_i W_i x_i^2 - (\Sigma nWx)^2/\Sigma nW$$

should be computed, and our estimate of the variance of m becomes

$$s_m^2 = \frac{1}{b^2}\left(\frac{1}{\Sigma nW}\right) + \frac{(m - \bar{x})^2}{S_{xx}}.$$

(15) A $(1 - \alpha)$ confidence interval estimate of μ is the interval from $m - z_{1-\alpha/2}\, s_m$ to $m + z_{1-\alpha/2}\, s_m$.

10-3.1.2 Example (Cont)

Steps (10) through (15) — A 95% confidence interval estimate of μ (the level at which we would expect half of the individuals to respond) is obtained as follows:

Columns W, nW, and nWx, are computed and entered in Table 10-3.

Compute:

$$\begin{aligned}
\Sigma nW &= 47.573 \\
\Sigma nWx &= 67.761 \\
\bar{x} &= \frac{\Sigma nWx}{\Sigma nW} \\
&= \frac{67.761}{47.573} \\
&= 1.424 \\
s_m^2 &= \frac{1}{b^2}\frac{1}{\Sigma nW} \\
&= \frac{1}{49}\left(\frac{1}{47.573}\right) \\
&= \frac{1}{2331.077} \\
&= .000429 \\
s_m &= .0207
\end{aligned}$$

(Since $\bar{x} = 1.424$, approximately equal to m ($= 1.43$), we do not bother to use the more complicated formula for s_m^2.)

Let: $\alpha = .05$
$z_{1-\alpha/2} = 1.96$

A 95% confidence interval estimate of μ is the interval $m \pm 1.96\, s_m = 1.43 \pm .04$, the interval from 1.39 to 1.47.

10-3.2 EXACT PROBIT SOLUTION

The graphical solution very often is adequate, but complications in the data may make this solution less satisfactory in some cases and an arithmetic technique may be necessary. For example, the points may be too irregular for us to place any confidence in a line drawn by eye; or, the weights (nW), that should be attached to each point, may be so different as to make it difficult to adjust for them visually.

10-3.2.1 Procedure. The probit solution given here involves a series of successive approximations, the first of which is given by the graphical probit solution or by the Kärber method described in Paragraph 10-2. For the exact solution, we need a table with ten columns headed, respectively, x, n, r, p, Y, W, nW, nWx, y, nWy. (See Table 10-4).

(1) In column x, enter the levels $x_1, x_2, \ldots, x_k$ used in the experiment.
(2) In columns n and r, record the values of n_i and r_i, corresponding to x_i.
(3) In column p, compute the respective proportions responding $p_i = r_i/n_i$.
(4) Obtain values for column Y by either method (a) or (b):
  (a) Follow instructions (1) through (7) of the graphical probit solution in Paragraph 10-3.1 to obtain the Y-values. These then are tabulated in the table for the exact solution.
  (b) Follow instructions (1) through (10) of the Kärber method described in Paragraph 10-2 to compute m and s. For initial estimates, take $m_0 = m$, $b_0 = \frac{1}{s}$; and corresponding to each x_i, use the equation $Y_i = 5 + b_0(x_i - m_0)$ to compute the values for the Y column of the table for the exact solution. Unless n is very large, 1 decimal in Y is sufficient. If 2 decimals in Y should be required, Tables A-18 and A-19 in ORDP 20-114 of this Handbook are not convenient; consult Finney[6] for more extensive tables.
(5) In column W, corresponding to each value of Y, enter the value of W obtained from Table A-18.
(6) Corresponding to each value of x, compute n_iW_i and $n_iW_ix_i$, and enter them in the nW and nWx columns, respectively, of the table.
(7) Corresponding to each "expected probit" Y in the Y column, use Table A-19 to compute the "working probit" y, as follows:

$$\text{For } Y \leq 5.0,\ y = y_0 + p\left(\frac{1}{Z}\right);$$

$$\text{For } Y \geq 5.0,\ y = y_{100} - (1 - p)\left(\frac{1}{Z}\right);$$

where, for each Y, p is the corresponding entry in the p column of the Table, and Table A-19 gives the values for y_0 (or y_{100}) and $\frac{1}{Z}$. Tabulate the values in column y.

(8) For each value of y, calculate the value $n_iW_iy_i$, and enter it in column nWy.
(9) Compute:

$\Sigma\, n_iW_i$, the sum of column nW
$\Sigma\, n_iW_ix_i$, the sum of column nWx
$\Sigma\, n_iW_iy_i$, the sum of column nWy
$S_1 = \Sigma\, n_iW_ix_i^2$, the sum of the products of elements in columns x and nWx.
$S_2 = \Sigma\, n_iW_ix_iy_i$, the sum of products of corresponding elements in columns nWx and y.
$S_3 = \Sigma\, nWy^2$
$S_{xx} = S_1 - (\Sigma\, nWx)^2/\Sigma\, nW$
$S_{xy} = S_2 - (\Sigma\, nWx)(\Sigma\, nWy)/\Sigma\, nW$
$S_{yy} = S_3 - (\Sigma\, nWy)^2/\Sigma\, nW$

10-3.2.2 Example. The observations of Data Sample 10-1 are used for illustration, and the line fitted graphically in Paragraph 10-3.1 is used as the first approximation.

Steps (1) through (3) consist of filling in the first four columns of Table 10-4.

Step (4) — The Y values entered are copied from the Y column of the graphical probit solution (Table 10-3). These are the ordinates of the line at each observed x.

Step (5) through (8) — fill in the remaining columns of Table 10-4.

TABLE 10-4. EXACT PROBIT SOLUTION

$\log_{10}PV$ x	No. of Objects Tested n	No. of Objects Responding r	Proportion of Objects Responding $p = r/n$	Expected Probit Y	W	nW	nWx	Working Probit y	nWy
1.0000	12	0	0	2.0	.015	0.180	0.180	1.695	0.305
1.1761	12	0	0	3.2	.180	2.160	2.540	2.745	5.929
1.2430	12	1	.08	3.7	.336	4.032	5.012	3.602	14.523
1.3010	13	2	.15	4.1	.471	6.123	7.966	3.972	24.321
1.3522	10	3	.30	4.5	.581	5.810	7.856	4.476	26.006
1.3979	13	6	.46	4.8	.627	8.151	11.394	4.900	39.940
1.4771	12	8	.67	5.3	.616	7.392	10.919	5.436	40.183
1.5441	13	9	.69	5.8	.503	6.539	10.097	5.461	35.709
1.6021	13	11	.85	6.2	.370	4.810	7.706	6.020	28.956
1.6990	11	10	.91	6.9	.154	1.694	2.878	5.966	10.106
1.7782	11	11	1.00	7.4	.062	0.682	1.213	7.766	5.296

Step (9) —

$$\Sigma\, nW = 47.573$$
$$\Sigma\, nWx = 67.761$$
$$\Sigma\, nWy = 231.274$$
$$S_1 = \Sigma\, nWx^2 = 97.4232$$
$$S_2 = \Sigma\, nWxy = 335.4415$$
$$S_{xx} = S_1 - (\Sigma\, nWx)^2/\Sigma\, nW = 0.9073$$
$$S_{xy} = S_2 - (\Sigma\, nWx)\,(\Sigma\, nWy)/\Sigma\, nW = 6.0244$$

10-3.2.1 Procedure (Cont)

(10) Compute:

$$b = S_{xy}/S_{xx}$$
$$\bar{x} = \Sigma\, nWx/\Sigma\, nW$$
$$\bar{y} = \Sigma\, nW_y/\Sigma\, nW$$

The equation of the probit regression line is

$$Y = \bar{y} + b(x - \bar{x}).$$

The procedure for obtaining the *best* line is an iterative one, and in theory we should repeat the above procedure until the same equation is obtained for two successive iterations. Practically, we often may be able to see that an additional iteration will not change the equation materially.

Next Iteration:

The procedure for obtaining an additional iteration is as follows: Make a table with nine columns headed x, n, p, Y, W, nW, nWx, y, nWy. (See Table 10-5.)

(11) Copy the three columns, x, n, p from Table 10-4.

(12) For each value of x, compute the corresponding value of Y, using the equation

$$Y = \bar{y} + b(x - \bar{x})$$

calculated in the previous iteration.

(13) Follow instructions of steps (5) through (10).

To test whether the line is a good fit to the data, use the test procedure outlined in Paragraph 10.3.3, using the $\bar{y}$, $\bar{x}$, and b obtained from the *last* iteration performed, to compute a new

$$Y = \bar{y} + b(x - \bar{x})$$

for each value of x used in the test.

10-3.2.2 Example (Cont)

Step (10) — $\bar{x} = \Sigma\, nWx/\Sigma\, nW = 1.4244$

$\bar{y} = \Sigma\, nWy/\Sigma\, nW = 4.8615$

$b = S_{xy}/S_{xx} = 6.640$

The equation of the probit regression line is

$$Y = \bar{y} + b(x - \bar{x}) = -4.5965 + 6.640\,x\,.$$

Next Iteration. An additional iteration is shown in Table 10-5. The first three columns (x, n, p) of Table 10-5 are copied from Table 10-4. The Y column is calculated by substituting observed values of x in the equation shown in Step (10) above. The remaining columns of Table 10-5 are filled in as described in Steps (5) through (8) of the Procedure.

TABLE 10-5. EXACT PROBIT SOLUTION (SECOND ITERATION)

$\log_{10}PV$ x	No. of Objects Tested n	Proportion of Objects Responding $p = r/n$	Y	W	nW	nWx	y	nWy
1.0000	12	0	2.0	.015	0.180	0.180	1.695	0.305
1.1761	12	0	3.2	.180	2.160	2.540	2.745	5.929
1.2430	12	.08	3.7	.336	4.032	5.012	3.602	14.523
1.3010	13	.15	4.0	.439	5.707	7.425	3.964	22.623
1.3522	10	.30	4.4	.558	5.580	7.545	4.477	24.982
1.3979	13	.46	4.7	.616	8.008	11.194	4.904	39.271
1.4771	12	.67	5.2	.627	7.524	11.114	5.432	40.870
1.5441	13	.69	5.7	.532	6.916	10.679	5.482	37.914
1.6021	13	.85	6.0	.439	5.707	9.143	6.036	34.447
1.6990	11	.91	6.7	.208	2.288	3.887	6.217	14.224
1.7782	11	1.00	7.2	.092	1.012	1.800	7.592	7.683

Step (9) — $\Sigma\, nW = 49.114$

$\Sigma\, nWx = 70.519$

$\Sigma\, nWy = 242.771$

$S_1 = \Sigma\, nWx^2 = 102.2663$

$S_2 = \Sigma\, nWxy = 355.3694$

$S_3 = \Sigma\, nWy^2 = 1247.7433$

$S_{xx} = 1.0135$

$S_{xy} = 6.7933$

$S_{yy} = 47.7238$

Step (10) — $\bar{x} = \Sigma\, nWx/\Sigma\, nW = 1.4358$

$\bar{y} = \Sigma\, nWy/\Sigma\, nW = 4.943$

$b = S_{xy}/S_{xx} = 6.703$

The equation of the probit regression line is:

$$Y = \bar{y} + b(x - \bar{x}) = -4.6812 + 6.703\,x\,.$$

Using this equation to calculate values of Y, the Y values obtained differ very little from those obtained on the first iteration, and no further iterations are considered necessary.

10-3.3 TESTING WHETHER THE LINE IS AN ADEQUATE REPRESENTATION OF THE DATA

10-3.3.1 Procedure. To facilitate the calculations, we require a table with ten columns: x, n, r, Y, P, nP, $1 - P$, $nP(1 - P)$, $r - nP$, and $\frac{(r - nP)^2}{nP(1 - P)}$. See Table 10-6.

(1) Choose α, the level of significance of the test. In columns x, n, r, copy the values from the probit solution table. If the graphical probit solution was used, copy the Y column of that table as the Y column here. If the exact probit solution was used, use the $\bar{y}$, $\bar{x}$, and b obtained from the last iteration performed to compute, corresponding to each x,

$$Y = \bar{y} + b(x - \bar{x})$$

and tabulate these values in the Y column here.

(2) Look up $\chi^2_{1-\alpha}$ for $k - 2$ degrees of freedom in Table A-3. k = number of rows in the table (see Table 10-6).

(3) Put $z_P = Y - 5$, and for each Y, using Table A-1, obtain the value of P corresponding to z_P.

(4) Compute the required quantities in the last columns of the table.

(5) Obtain $\chi^2 = \sum_{i=1}^{k} \frac{(r_i - n_iP_i)^2}{n_iP_i(1 - P_i)}$, the sum of the values in the last column.

(6) If $\chi^2 \leq \chi^2_{1-\alpha}$, decide that there is no reason to believe that the line does not adequately represent the data. If $\chi^2 > \chi^2_{1-\alpha}$, decide that the straight line does not adequately describe the relation between stimulus and response. If a significant value of χ^2 is obtained, check to see whether an unusually large contribution to χ^2 comes from one class or from a few classes with very small expected values, i.e., small nP or $n(1 - P)$. If this is the case, several such classes may be combined (for details, see Finney[6]).

10-3.3.2 Example. The test of the final probit equation is shown in Table 10-6. If $\alpha = .05$, we find in Step (2) that $\chi^2_{.95}$ for 9 d.f. = 16.92. In Step (5), the calculated $\chi^2 = 2.37$. Since this is not larger than $\chi^2_{.95}$, we accept the fitted line.

TABLE 10-6. TEST OF LINEARITY — FINAL PROBIT EQUATION

$\log_{10}PV$ x	n	r	Y	P	nP	$1 - P$	$nP(1 - P)$	$r - nP$	$\frac{(r - nP)^2}{nP(1 - P)}$
1.0000	12	0	2.02	.0014	0.02	.9986	0.02	−.02	.02
1.1761	12	0	3.20	.04	0.48	.96	0.46	−.48	.50
1.2430	12	1	3.65	.09	1.08	.91	0.98	−.08	.01
1.3010	13	2	4.04	.17	2.21	.83	1.83	−.21	.02
1.3522	10	3	4.38	.27	2.70	.73	1.97	.30	.05
1.3979	13	6	4.69	.38	4.94	.62	3.06	1.06	.37
1.4771	12	8	5.22	.59	7.08	.41	2.90	.92	.29
1.5441	13	9	5.67	.75	9.75	.25	2.44	−.75	.23
1.6021	13	11	6.06	.86	11.18	.14	1.57	−.18	.02
1.6990	11	10	6.71	.96	10.56	.04	0.42	−.56	.75
1.7782	11	11	7.24	.99	10.89	.01	0.11	+.11	.11

$\chi^2 = 2.37$

10-3.4 USING THE PROBIT REGRESSION LINE FOR PREDICTION

The procedures which follow describe how to use the probit regression line for estimation, including confidence interval estimates. If a *nonsignificant value* of χ^2 was obtained in the test of linearity, Paragraph 10-3.3, the formulas given here are directly applicable. If a *significant value* of χ^2 was obtained, these formulas do not apply without adjustment. For details, see Finney.[6]

10-3.4.1 Level of Stimulus x' At Which a Specified Proportion P' of the Individuals Would Be Expected To Respond.

Single Estimate

(1) Choose P'

(2) Use Table A-2 to find $z_{P'}$ corresponding to P'

(3) Let $Y' = 5 + z_{P'}$

(4) Compute $x' = \bar{x} + (Y' - \bar{y})/b$.

This x' is the value at which we would expect a proportion P' of the individuals to respond.

Confidence Interval Estimate

For a confidence interval estimate, see Finney.[6]

10-3.4.2 Level of Stimulus x' At Which 50% of the Individuals Would Be Expected To Respond. The estimate of μ, the mean and the 50th percentile of the distribution of critical levels, is

$$m = \bar{x} + \frac{1}{b}(5 - \bar{y}).$$

A $(1 - \alpha)$ confidence interval estimate of μ is computed as follows:*

(1) Look up $z_{1-\alpha/2}$ in Table A-2.

(2) Compute

$$s_m^2 = \frac{1}{b^2}\left[\frac{1}{\Sigma nW} + \frac{(m - \bar{x})^2}{S_{xx}}\right]$$

(3) A $(1 - \alpha)$ confidence interval estimate of μ is the interval from $m - z_{1-\alpha/2}\, s_m$ to $m + z_{1-\alpha/2}\, s_m$.

10-3.4.3 Proportion of Individuals Which Would Be Expected To Respond At a Specified Level of Stimulus. The probit regression equation $Y = \bar{y} + b(x - \bar{x})$ gives the expected value of $Y(Y')$ at a specified value of $x(x')$.

The variance of Y, s_Y^2, is given by the formula

$$s_Y^2 = \frac{1}{\Sigma nW} + \frac{(x' - \bar{x})^2}{S_{xx}}.$$

A $(1 - \alpha)$ confidence interval estimate for Y at a single specified value of x is given by the formula

$$Y' \pm t_{1-\alpha/2}\, s_Y.$$

If we want to make confidence interval statements for several values of x using the same fitted line, however, we must use the wider interval given by

$$Y' \pm \sqrt{2F}\, s_Y,$$

as discussed in Chapter 5 (see Paragraph 5-4.1.2), ORDP 20-110.

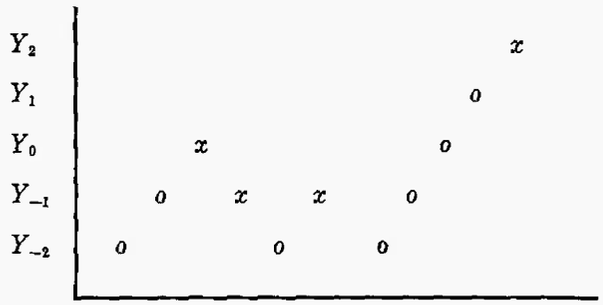
* This method is sufficiently good for most purposes. For an exact method, see Finney.[6]

10-4 THE UP-AND-DOWN DESIGN

The *up-and-down* design, sometimes called the "Bruceton" method, is one of a class of designs that are called *staircase methods* because the test level for the next trial or group of trials depends on the results of the preceding trial or group of trials. In the up-and-down design, only one object is tested at a time. Starting at a level where about 50% responses are expected, the test level is moved up one level after each non-response, and down one level after each response. The experiment is concluded after a specified number of trials.

The use of the up-and-down design, of course, presumes that it is convenient to test one object at a time and all staircase methods presume that the results of test can be known immediately, and that the test level can be adjusted quickly and easily.

If x's represent responses, and o's represent non-responses, then the pattern of the experiment looks like this:



The *up-and-down* design and its analysis are described in detail here. The Procedure assumes normal distribution of the critical levels and, in such a case, gives a more accurate estimate of the mean (which also is the 50th percentile) than any other method described in this Chapter. For further discussion of the up-and-down method, see Brownlee, Hodges, and Rosenblatt,[7] Dixon and Massey,[8] Dixon and Mood.[9] For a completely worked-out example of an ordnance application, see Culling.[10] Other staircase methods are described in Anderson, McCarthy, and Tukey.[11]

The up-and-down design requires initial guesses of the mean and standard deviation of the distribution of critical levels (x_0 = guess for mean, d = guess for standard deviation). The method of estimation makes some allowance for a poor initial guess of the mean x_0, and, in fact, is a particularly useful way of estimating the mean when the experimenter has little idea what the true mean is. So long as the initial guess for the standard deviation is between half and twice the true standard deviation of the distribution, the method of estimation described is appropriate.

Determine equally-spaced test levels, $\ldots, x_{-3}, x_{-2}, x_{-1}, x_0, x_1, x_2, x_3, \ldots$, so that the distance between two successive levels is d. ($d = x_1 - x_0 = x_0 - x_{-1}$, etc.). The first object is tested at level x_0; if it "responds," the second object is tested at level x_{-1}; if it "does not respond," the second object is tested at level x_1. Similarly, each succeeding object is tested at the level one step below the level used in the preceding test if it resulted in a "response," or, at the level one step above the level used in the preceding test if it resulted in "no response."

Procedure. Count the total number $R = \Sigma r_i$ of responses, and let $N = \Sigma n_i$ be the total number of objects tested. If $R \leq \frac{N}{2}$, perform steps (1) through (6); if $R > \frac{N}{2}$, perform steps (1′) through (6′).

When $R \leq \frac{N}{2}$:

(1) Prepare a four-column table with columns headed y, r, j, j^2.

(2) Let: y_0 = the lowest level at which a "response" occurred
y_1 = the level one step above y_0
y_2 = the level two steps above y_0
y_k = the highest level at which a "response" occurred.
Enter $y_0, y_1, \ldots, y_k$ in column y.

(3) In column r, corresponding to each y_j, enter r_j = the number of "responses" at level y_j.

(4) Enter the numbers 0, 1, . . . , k, in column j.

(5) Corresponding to each entry in column j, enter its square in column j^2.

(6) Compute:
$A = \Sigma j\, r_j$, the sum of products of corresponding entries in columns r and j.
$B = \Sigma j^2 r_j$, the sum of products of corresponding entries in columns r and j^2.

$$m = y_0 + d\left(\frac{A}{R} - \frac{1}{2}\right)$$

$$s = 1.620d\left(\frac{RB - A^2}{R^2} + .029\right).$$

When $R > \frac{N}{2}$:

(1′) Prepare a four-column table with columns headed y, $n - r$, j, j^2.

(2′) Let: y_0 = the lowest level at which "no response" occurred
y_1 = the level one step above y_0
y_2 = the level two steps above y_0
y_k = the highest level at which "no response" occurred.
Enter $y_0, y_1, \ldots, y_k$, in column y.

(3′) In column $n - r$, corresponding to each y_j, enter $n_j - r_j$, the number of "no response" at level y_j.

(4′) Enter the numbers 0, 1, . . . , k, in column j.

(5′) Corresponding to each entry in column j, enter its square in column j^2.

(6′) Compute:
$A = \Sigma j\,(n_j - r_j)$, the sum of products of corresponding entries in columns $n - r$ and j
$B = \Sigma j^2\,(n_j - r_j)$, the sum of products of corresponding entries in columns $n - r$ and j^2.

$$m = y_0 + d\left(\frac{A}{N - R} + \frac{1}{2}\right)$$

$$s = 1.62d\left\{\frac{(N - R)B - A^2}{(N - R)^2} + .029\right\}.$$

m is our estimate of the mean (and the 50th percentile) of the distribution of critical levels.

s is our estimate of the standard deviation of the distribution of critical levels.

10-5 SENSITIVITY TESTS WHEN THE STIMULUS LEVELS CANNOT BE CONTROLLED

The methods discussed in Paragraphs 10-2 through 10-4 assumed that the stimulus levels can be preassigned and accurately controlled. Although this is the usual case in experimental work, there are times when conditions cannot be sufficiently well controlled to insure that the level used is exactly the one that the experimenter intended to use. For example: he may intend to fire a group of 10 projectiles, each at a velocity of 2000 f/s; but, because of random variation in velocities for a fixed charge, the actual observed velocities range from 1975 to 2020 f/s.

In such a case, when the level used in the test can be measured directly, the experimenter has two choices for analyzing the data. If the actual levels used cluster so closely about the intended levels that:

(1) there is no overlapping between two of these clusters;

and,

(2) the range of any cluster is so small that the probability of "response" at any of the actual levels differs little from the probability of response at the intended level;

then the experimenter may simply assume that each test was conducted at the intended level, and use the methods already presented in Paragraphs 10-2 through 10-4. However, if one or both of these conditions are not met, none of the methods described in this Chapter are valid. Techniques for handling such data when the underlying distribution is normal are described in Golub and Grubbs.[12]

REFERENCES

1. J. Cornfield and N. Mantel, "Some New Aspects of the Application of Maximum Likelihood to the Calculation of the Dosage Response Curve," *Journal of the American Statistical Association*, Vol. 45, pp. 181-209, 1950.
2. J. Berkson, "A Statistically Precise and Relatively Simple Method of Estimating the Bio-assay with Quantal Response, Based on the Logistic Function," *Journal of the American Statistical Association*, Vol. 48, pp. 565-599, 1953.
3. J. L. Hodges, Jr., "Fitting the Logistic by Maximum Likelihood," *Biometrics*, Vol. 14, No. 4, pp. 453-461, December 1958.
4. L. F. Knudsen and J. M. Curtis, "The Use of the Angular Transformation in Biological Assays," *Journal of the American Statistical Association*, Vol. 42, pp. 282-296, 1947.
5. R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, Hafner Publishing Co., New York, N.Y., 5th ed., 1957.
6. D. J. Finney, *Probit Analysis*, Cambridge University Press, 1947.
7. K. A. Brownlee, J. L. Hodges, Jr., and M. Rosenblatt, "The Up-and-Down Method with Small Samples," *Journal of the American Statistical Association*, Vol. 48, pp. 262-277, 1953.
8. W. J. Dixon and F. J. Massey, Jr., *Introduction to Statistical Analysis* (2d ed.), Chapter 19, "Sensitivity Experiments," McGraw-Hill Book Co., Inc., New York, N.Y., 1957.
9. W. J. Dixon and A. M. Mood, "A Method for Obtaining and Analyzing Sensitivity Data," *Journal of the American Statistical Association*, Vol. 43, pp. 109-126, 1948.
10. H. P. Culling, "Statistical Methods Appropriate for Evaluation of Fuze Explosive Train Safety and Reliability," *NAVORD Report 2101*, U. S. Naval Ordnance Laboratory, White Oak, Maryland, October 1953.
11. T. W. Anderson, P. J. McCarthy, and J. W. Tukey, "Staircase Method of Sensitivity Testing," *NAVORD Report 65-46*, U. S. Navy Bureau of Ordnance, Washington, D. C., March 1946.
12. A. Golub and F. E. Grubbs, "Analysis of Sensitivity Experiments When the Levels of Stimulus Cannot Be Controlled," *Journal of the American Statistical Association*, Vol. 51, pp. 257-265, 1956.

☆ U. S. GOVERNMENT PRINTING OFFICE : 1963 O - 671314