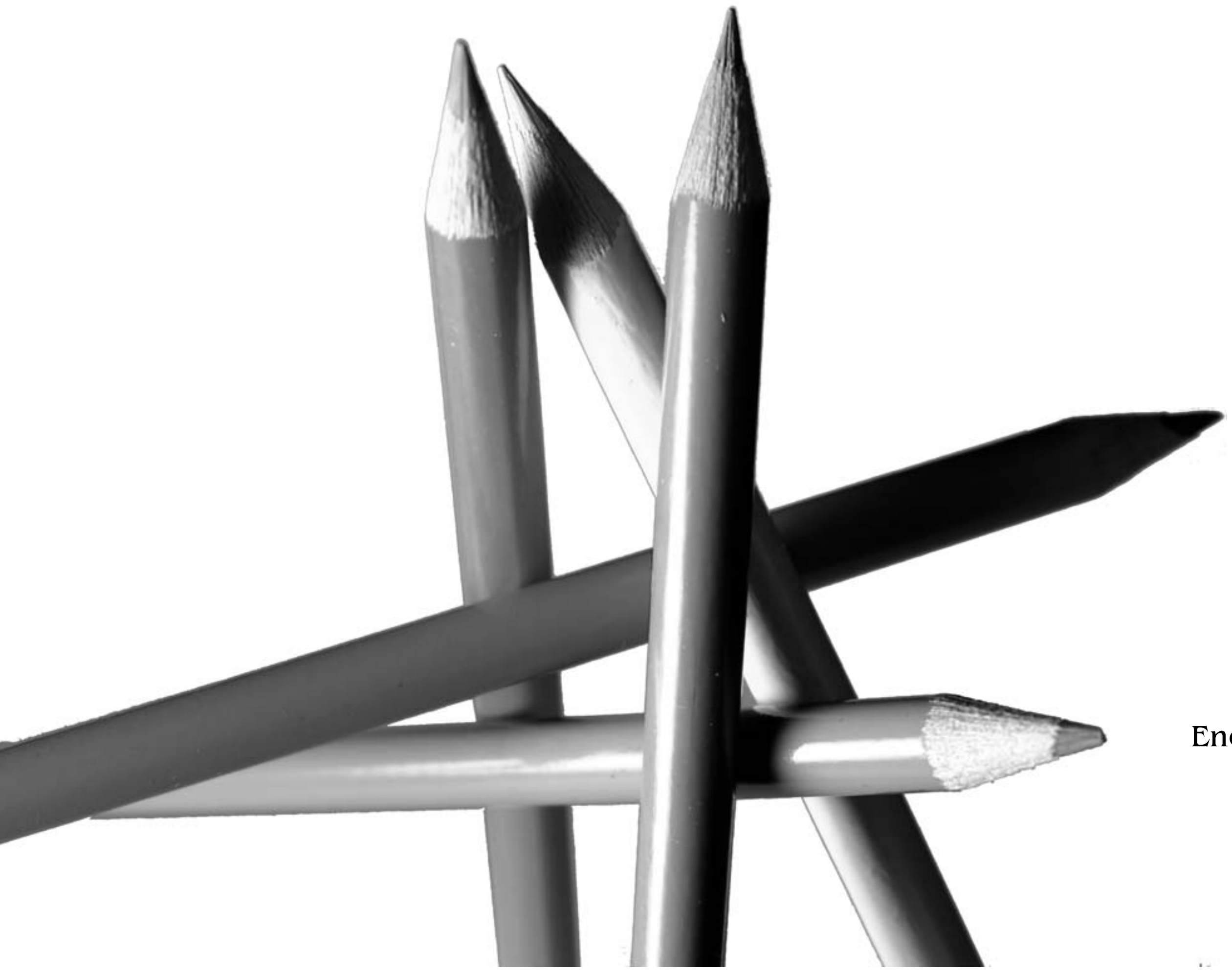
Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report

A PRE-K–12 CURRICULUM FRAMEWORK



CHRISTINE FRANKLIN
University of Georgia

GARY KADER
Appalachian State University

DENISE MEWBORN
University of Georgia

JERRY MORENO
John Carroll University

ROXY PECK
California Polytechnic State University, San Luis Obispo

MIKE PERRY
Appalachian State University

RICHARD SCHEAFFER
University of Florida

Endorsed by the American Statistical Association
August 2005

Library of Congress Cataloging-in-Publication Data
Guidelines for assessment and instruction in statistics education (GAISE) report: a pre-k–12 curriculum framework / Authors, Christine Franklin … [et al.].
p. cm.
Includes bibliographical references.
ISBN-13: 978-0-9791747-1-1 (pbk.)
ISBN-10: 0-9791747-1-6 (pbk.)
1. Statistics–Study and teaching (Early childhood)–Standards.
2. Statistics–Study and teaching (Elementary)–Standards.
3. Statistics–Study and teaching (Secondary)–Standards
I. Franklin, Christine A

QA276.18.G85 2007
519.5071–dc22

2006103096

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

978-0-9791747-1-1
0-9791747-1-6

Advisors

Susan Friel
The University of North Carolina

Landy Godbold
Westminster Schools

Brad Hartlaub
Kenyon College

Peter Holmes
Nottingham Trent University

Cliff Konold
University of Massachusetts/Amherst

Production Team

Christine Franklin
University of Georgia

Nicholas Horton
Smith College

Gary Kader
Appalachian State University

Jerry Moreno
John Carroll University

Megan Murphy
American Statistical Association

Valerie Snider
American Statistical Association

Daren Starnes
Fountain Valley School of Colorado

Special Thanks

The authors extend a special thank you to the American Statistical Association Board of Directors for funding the writing process of GAISE as a strategic initiative and to the ASA/NCTM Joint Committee for funding the production of the GAISE Framework.

Cover photo by Andres Rodriguez
Book design by Valerie Snider

Contents

Introduction **1**

Framework **11**

The Role of Variability in the Problem-Solving Process 11

Maturing over Levels 12

The *Framework* Model 13

Illustrations 16

Detailed Descriptions of Each Level 21

Level A **23**

Example 1: Choosing the Band for the End of the Year Party—Conducting a Survey 24

Comparing Groups 27

The Simple Experiment 28

Example 2: Growing Beans—A Simple Comparative Experiment 28

Making Use of Available Data 29

Describing Center and Spread 29

Looking for an Association 31

Example 3: Purchasing Sweat Suits—The Role of Height and Arm Span 31

Understanding Variability 33

The Role of Probability 33

Misuses of Statistics 35

Summary of Level A 35

Level B **37**

Example 1, Level A Revisited: Choosing a Band for the School Dance 38

Connecting Two Categorical Variables 40

Questionnaires and Their Difficulties 41

Measure of Location—The Mean as a Balance Point 41

A Measure of Spread—The Mean Absolute Deviation 44

Representing Data Distributions—The Frequency Table and Histogram 44

Comparing Distributions—The Boxplot 46

Measuring the Strength of Association between Two Quantitative Variables 48

Modeling Linear Association 51

The Importance of Random Selection 52

Comparative Experiments 54

Time Series 55

Misuses of Statistics 56

Summary of Level B 58

Level C **61**

An Introductory Example—Obesity in America 62

The Investigatory Process at Level C 64

Example 1: The Sampling Distribution of a Sample Proportion 67

Example 2: The Sampling Distribution of a Sample Mean 69

Example 3: A Survey of Music Preferences 71

Example 4: An Experiment on the Effects of Light on the Growth of Radish Seedlings 75

Example 5: Estimating the Density of the Earth—A Classical Study 79

Example 6: Linear Regression Analysis—Height vs. Forearm Length 80

Example 7: Comparing Mathematics Scores—An Observational Study 82

Example 8: Observational Study—Toward Establishing Causation 83

The Role of Probability in Statistics 84

Summary of Level C 87

Appendix for Level A **89**

Appendix for Level B **95**

Appendix for Level C **99**

References **108**

Index of Tables and Figures

Table 1: The Framework 14–15
Table 2: Frequency Count Table 24
Table 3: Frequencies and Relative Frequencies 39
Table 4: Two-Way Frequency Table 40, 95
Table 5: Grouped Frequency and Grouped Relative Frequency Distributions 46
Table 6: Hat Size Data 47
Table 7: Five-Number Summaries for Sodium Content 47
Table 8: Height and Arm Span Data 48
Table 9: Five-Number Summaries 55
Table 10: Live Birth Data 56
Table 11: Two-Way Frequency Table 72
Table 12: Lengths of Radish Seedlings 76
Table 13: Treatment Summary Statistics 77
Table 14: Heights vs. Forearm Lengths 81, 99
Table 15: NAEP 2000 Scores in Mathematics 82
Table 16: Cigarette Smoking and Lung Cancer 83
Table 17: Level of Cigarette Smoking and Lung Cancer 84
Table 18: Family Size Distribution 86
Table 19: 2x2 Two-Way Frequency Table 96
Table 20: Two-Way Frequency Table 97
Table 21: Two-Way Frequency Table 97
Table 22: Two-Way Frequency Table 98
Table 23: Result of Lifestyle Question 100
Table 24: Pulse Data 102
Table 25: Pulse Data in Matched Pairs 102
Table 26: U.S. Population (in 1,000s) 104
Table 27: U.S. Death Rates (Deaths per 100,000 of Population) 105
Table 28: Enrollment Data 106
Figure 1: Picture Graph of Music Preferences 25
Figure 2: Bar Graph of Music Preferences 26
Figure 3: Stem and Leaf Plot of Jumping Distances 27
Figure 4: Dotplot of Environment vs. Height 28
Figure 5: Parallel Dotplot of Sodium Content 29
Figure 6: Scatterplot of Arm Span vs. Height 32
Figure 7: Timeplot of Temperature vs. Time 32
Figure 8: Comparative Bar Graph for Music Preferences 39
Figure 9: Dotplot for Pet Count 42
Figure 10: Dotplot Showing Pets Evenly Distributed 42
Figure 11: Dotplot with One Data Point Moved 42
Figure 12: Dotplot with Two Data Points Moved 42
Figure 13: Dotplot with Different Data Points Moved 43
Figure 14: Dotplot Showing Distance from 5 43
Figure 15: Dotplot Showing Original Data and Distance from 5 43
Figure 16: Stemplot of Head Circumference 45
Figure 17: Relative Frequency Histogram 45
Figure 18: Boxplot for Sodium Content 47
Figure 19: Scatterplot of Arm Span vs. Height 49
Figure 20: Scatterplot Showing Means 49
Figure 21: Eyeball Line 51
Figure 22: Eighty Circles 53
Figure 23: Boxplot for Memory Data 55
Figure 24: Time Series Plot of Live Births 56
Figure 25: Histogram of Sample Proportions 68
Figure 26: Histogram of Sample Means 69
Figure 27: Dotplot of Sample Proportions from a Hypothetical Population in Which 50% Like Rap Music 72
Figure 28: Dotplot of Sample Proportions from a Hypothetical Population in Which 40% Like Rap Music 73
Figure 29: Dotplot Showing Simulated Sampling Distribution 74
Figure 30: Seed Experiment 75
Figure 31: Boxplot Showing Growth under Different Conditions 77
Figure 32: Dotplot Showing Differences of Means 78
Figure 33: Dotplot Showing Differences of Means 78
Figure 34: Histogram of Earth Density Measurements 80
Figure 35: Scatterplot and Residual Plot 81, 99
Figure 36: Random Placement of Names 89
Figure 37: Names Clustered by Length 90
Figure 38: Preliminary Dotplot 90
Figure 39: Computer-Generated Dotplot 91
Figure 40: Student-Drawn Graphs 92
Figure 41: Initial Sorting of Candies 93
Figure 42: Bar Graph of Candy Color 93
Figure 43: Scatterplot of Arm Span/ Height Data 95
Figure 44: Dotplot Showing Association 100
Figure 45: Dotplot Showing Differences in Sample Proportions 101
Figure 46: Dotplot of Randomized Differences in Means 103
Figure 47: Dotplot of Randomized Pair Difference Means 104
Figure 48: Scatterplot of Death Rates 105
Figure 49: Scatterplot of Actual Deaths 105
Figure 50: Distorted Graph 106
Figure 51: Plot of African-American vs. Total Enrollments 107
Figure 52: Plot of African-American Enrollments Only 107
Figure 53: Ratio of African-American to Total Enrollments 107

Introduction

The ultimate goal: statistical literacy. Every morning, the newspaper and other media confront us with statistical information on topics ranging from the economy to education, from movies to sports, from food to medicine, and from public opinion to social behavior. Such information guides decisions in our personal lives and enables us to meet our responsibilities as citizens. At work, we may be presented with quantitative information on budgets, supplies, manufacturing specifications, market demands, sales forecasts, or workloads. Teachers may be confronted with educational statistics concerning student performance or their own accountability. Medical scientists must understand the statistical results of experiments used for testing the effectiveness and safety of drugs. Law enforcement professionals depend on crime statistics. If we consider changing jobs and moving to another community, then our decision can be affected by statistics about cost of living, crime rate, and educational quality.

Our lives are governed by numbers. Every high-school graduate should be able to use sound statistical reasoning to intelligently cope with the requirements of citizenship, employment, and family and to be prepared for a healthy, happy, and productive life.

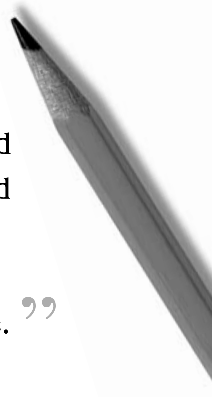
Citizenship

Public opinion polls are the most visible examples of a statistical application that has an impact on our lives. In addition to directly informing individual citizens, polls are used by others in ways that affect us. The political process employs opinion polls in several ways. Candidates for office use polling to guide campaign strategy. A poll can determine a candidate's strengths with voters, which can, in turn, be emphasized in the campaign. Citizens also might be suspicious that poll results might influence candidates to take positions just because they are popular.

A citizen informed by polls needs to understand that the results were determined from a sample of the population under study, that the reliability of the results depends on how the sample was selected, and that the results are subject to sampling error. The statistically literate citizen should understand the behavior of "random" samples and be able to interpret a "margin of sampling error."

The federal government has been in the statistics business from its very inception. The U.S. Census was established in 1790 to provide an official count of the population for the purpose of allocating representatives to Congress. Not only has the role of the U.S. Census Bureau greatly expanded to include the collection of a broad spectrum of socioeconomic data, but other federal departments also produce extensive "official" statistics concerned with agriculture, health, education, environment, and commerce. The information gathered by these agencies influences policy making and helps to determine priorities for

> "Every high-school graduate should be able to use sound statistical reasoning to intelligently cope with the requirements of citizenship, employment, and family and to be prepared for a healthy, happy, and productive life."



government spending. It is also available for general use by individuals or private groups. Thus, statistics compiled by government agencies have a tremendous impact on the life of the ordinary citizen.

Personal Choices

Statistical literacy is required for daily personal choices. Statistics provides information about the nutritional quality of foods and thus informs our choices at the grocery store. Statistics helps to establish the safety and effectiveness of drugs, which aids physicians in prescribing a treatment. Statistics also helps to establish the safety of toys to assure our children are not at risk. Our investment choices are guided by a plethora of statistical information about stocks and bonds. The Nielsen ratings help determine which shows will survive on television, thus affecting what is available. Many products have a statistical history, and our choices of products can be affected by awareness of this history. The design of an automobile is aided by anthropometrics—the statistics of the human body—to enhance passenger comfort. Statistical ratings of fuel efficiency, safety, and reliability are available to help us select a vehicle.

The Workplace and Professions

Individuals who are prepared to use statistical thinking in their careers will have the opportunity to advance to more rewarding and challenging positions. A statistically competent work force will allow the United States to compete more effectively in the global marketplace and to improve its position in the international economy. An investment in statistical literacy is an investment in our nation's economic future, as well as in the well-being of individuals.

The competitive marketplace demands quality. Efforts to improve quality and accountability are prominent among the many ways that statistical thinking and tools can be used to enhance productivity. Quality-control practices, such as the statistical monitoring of design and manufacturing processes, identify where improvement can be made and lead to better product quality. Systems of accountability can help produce more effective employees and organizations, but many accountability systems now in place are not based on sound statistical principles and may, in fact, have the opposite effect. Good accountability systems require proper use of statistical tools to determine and apply appropriate criteria.

Science

Life expectancy in the US almost doubled during the 20th century; this rapid increase in life span is the consequence of science. Science has enabled us to improve medical care and procedures, food production, and the detection and prevention of epidemics. Statistics plays a prominent role in this scientific progress.

The U.S. Food and Drug Administration requires extensive testing of drugs to determine effectiveness and side effects before they can be sold. A recent advertisement for a drug designed to reduce blood clots stated, "PLAVIX, added to aspirin and your current medications, helps raise your protection against heart attack or stroke." But the advertisement also warned, "The risk of bleeding may increase with PLAVIX..."

Statistical literacy involves a healthy dose of skepticism about "scientific" findings. Is the information about side effects of PLAVIX treatment reliable? A statistically literate person should ask such questions and be able to intelligently answer them. A statistically literate high-school graduate will be able to understand the conclusions from scientific investigations and offer an informed opinion about the legitimacy of the reported results. According to *Mathematics and Democracy: The Case for Quantitative Literacy* (Steen, 2001), such knowledge "empowers people by giving them tools to think for themselves, to ask intelligent questions of experts, and to confront authority confidently. These are skills required to survive in the modern world."

Statistical literacy is essential in our personal lives as consumers, citizens, and professionals. Statistics plays a role in our health and happiness. Sound statistical reasoning skills take a long time to develop. They cannot be honed to the level needed in the modern world through one high-school course. The surest way to help students attain the necessary skill level is to begin the statistics education process in the elementary grades and keep strengthening and expanding students' statistical thinking skills throughout the middle- and high-school years. A statistically literate high-school graduate will know how to interpret the data in the morning newspaper and will ask the right questions about statistical claims. He or she will be comfortable handling quantitative decisions that come up on the job, and will be able to make informed decisions about quality-of-life issues.

The remainder of this document lays out a curriculum framework for pre-K–12 educational programs that is designed to help students achieve statistical literacy.

The Case for Statistics Education

Over the past quarter century, statistics (often labeled data analysis and probability) has become a key component of the pre-K–12 mathematics curriculum. Advances in technology and modern methods of data analysis in the 1980s, coupled with the data richness of society in the information age, led to the development of curriculum materials geared toward introducing statistical concepts into the school curriculum as early as the elementary grades. This grassroots effort was given sanction by the National Council of Teachers of Mathematics (NCTM) when their influential document, *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989), included "Data Analysis

“Statistics education as proposed in this Framework can promote the ‘must-have’ competencies for graduates to ‘thrive in the modern world.’”

and Probability” as one of the five content strands. As this document and its 2000 replacement, *Principles and Standards for School Mathematics* (NCTM, 2000), became the basis for reform of mathematics curricula in many states, the acceptance of and interest in statistics as part of mathematics education gained strength. In recent years, many mathematics educators and statisticians have devoted large segments of their careers to improving statistics education materials and pedagogical techniques.

NCTM is not the only group calling for improved statistics education beginning at the school level. The National Assessment of Educational Progress (NAEP, 2005) was developed around the same content strands as the NCTM *Standards*, with data analysis and probability questions playing an increasingly prominent role on the NAEP exam. In 2006, the College Board released its *College Board Standards for College Success*™: *Mathematics and Statistics*, which includes “Data and Variation” and “Chance, Fairness, and Risk” among its list of eight topic areas that are “central to the knowledge and skills developed in the middle-school and high-school years.” An examination of the standards recommended by this document reveals a consistent emphasis on data analysis, probability, and statistics at each course level.

The emerging quantitative literacy movement calls for greater emphasis on practical quantitative skills that will help assure success for high-school graduates in life and work; many of these skills are statistical in nature. To quote from *Mathematics and Democracy: The Case for Quantitative Literacy* (Steen, 2001):

> Quantitative literacy, also called numeracy, is the natural tool for comprehending information in the computer age. The expectation that ordinary citizens be quantitatively literate is primarily a phenomenon of the late twentieth century. …Unfortunately, despite years of study and life experience in an environment immersed in data, many educated adults remain functionally illiterate. …Quantitative literacy empowers people by giving them tools to think for themselves [sic], to ask intelligent questions of experts, and to confront authority confidently. These are the skills required to thrive in the modern world.

A recent study from the American Diploma Project, titled *Ready or Not: Creating a High School Diploma That Counts* (*www.amstat.org/education/gaise/1*), recommends “must-have” competencies needed for high-school graduates “to succeed in postsecondary education or in high-performance, high-growth jobs.” These include, in addition to algebra and geometry, aspects of data analysis, statistics, and other applications that are vitally important for other subjects, as well as for employment in today’s data-rich economy.

Statistics education as proposed in this *Framework* can promote the “must-have” competencies for graduates to “thrive in the modern world.”

NCTM Standards and the Framework

The main objective of this document is to provide a conceptual *Framework* for K–12 statistics education. The foundation for this *Framework* rests on the NCTM *Principles and Standards for School Mathematics* (2000).

The *Framework* is intended to complement the recommendations of the NCTM *Principles and Standards,* not to supplant them.

The NCTM *Principles and Standards* describes the statistics content strand as follows:

Data Analysis and Probability

Instructional programs from pre-kindergarten through grade 12 should enable all students to:

→ formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them;

→ select and use appropriate statistical methods to analyze data;

→ develop and evaluate inferences and predictions that are based on data; and

→ understand and apply basic concepts of probability.

The "Data Analysis and Probability" standard recommends that students formulate questions that can be answered using data and address what is involved in wisely gathering and using that data. Students should learn how to collect data, organize their own or others' data, and display the data in graphs and charts that will be useful in answering their questions. This standard also includes learning methods for analyzing data and ways of making inferences and drawing conclusions from data. The basic concepts and applications of probability also are addressed, with an emphasis on the way probability and statistics are related.

The NCTM *Principles and Standards* elaborates on these themes somewhat and provides examples of the types of lessons and activities that might be used in a classroom. More complete examples can be found in the NCTM *Navigation Series on Data Analysis and Probability* (2002–2004). Statistics, however, is a relatively new subject for many teachers, who have not had an opportunity to develop sound knowledge of the principles and concepts underlying the practices of data analysis that they now are called upon to teach. These teachers do not clearly understand the difference between statistics and mathematics. They do not see the statistics curriculum for grades pre-K–12 as a cohesive and coherent curriculum strand. These teachers may not see how the overall statistics curriculum provides a developmental sequence of learning experiences.

This *Framework* provides a conceptual structure for statistics education that gives a coherent picture of the overall curriculum.

The Difference between Statistics and Mathematics

"Statistics is a methodological discipline. It exists not for itself, but rather to offer to other fields of study a coherent set of ideas and tools for dealing with data. The need for such a discipline arises from the *omnipresence of variability*." (Moore and Cobb, 1997)

A major objective of statistics education is to help students develop statistical thinking. Statistical thinking, in large part, must deal with this omnipresence of variability; statistical problem solving and decision making depend on understanding, explaining, and quantifying the variability in the data.

It is this focus on *variability in data* that sets apart statistics from mathematics.

The Nature of Variability

There are many sources of variability in data. Some of the important sources are described below.

Measurement Variability—Repeated measurements on the same individual vary. Sometimes two measurements vary because the measuring device produces unreliable results, such as when we try to measure a large distance with a small ruler. At other times, variability results from changes in the system being measured. For example, even with a precise measuring device, your recorded blood pressure could differ from one moment to the next.

Natural Variability—Variability is inherent in nature. Individuals are different. When we measure the same quantity across several individuals, we are bound to get differences in the measurements. Although some of this may be due to our measuring instrument, most of it is simply due to the fact that individuals differ. People naturally have different heights, different aptitudes and abilities, and different opinions and emotional responses. When we measure any one of these traits, we are bound to get variability in the measurements. Different seeds for the same variety of bean will grow to different sizes when subjected to the same environment because no two seeds are exactly alike; there is bound to be variability from seed to seed in the measurements of growth.

Induced Variability—If we plant one pack of bean seeds in one field, and another pack of seeds in another location with a different climate, then an observed difference in growth among the seeds in one location with those in the other might be due to inherent differences in the seeds (natural variability), or the observed difference might be due to the fact that the locations are not the same. If one type of fertilizer is used on one field and another type on the other, then observed differences might be due to the difference in fertilizers. For that matter, the observed difference might be due to a factor we haven't even thought about. A more carefully designed experiment can help us determine the effects of different factors.

This one basic idea, comparing natural variability to the variability induced by other factors, forms the heart of modern statistics. It has allowed medical science to conclude that some drugs are effective and safe, whereas others are ineffective or have harmful side effects. It has been employed by agricultural scientists to demonstrate that a variety of corn grows better in one climate than another, that one fertilizer is more effective than another, or that one type of feed is better for beef cattle than another.

Sampling Variability—In a political poll, it seems reasonable to use the proportion of voters surveyed (a sample statistic) as an estimate of the unknown proportion of all voters who support a particular candidate. But if a second sample of the same size is used, it is almost certain that there would not be exactly the same proportion of voters in the sample who support the candidate. The value of the sample proportion will vary from sample to sample. This is called sampling variability. So what is to keep one sample from estimating that the true proportion is .60 and another from saying it is .40? This is possible, but unlikely, if proper sampling techniques are used. Poll results are useful because these techniques and an adequate sample size can ensure that unacceptable differences among samples are quite unlikely.

An excellent discussion on the nature of variability is given in *Seeing Through Statistics* (Utts, 1999).

The Role of Context

"The focus on variability naturally gives statistics a particular content that sets it apart from mathematics, itself, and from other mathematical sciences, but there is more than just content that distinguishes statistical thinking from mathematics. Statistics requires a different kind of thinking, because *data are not just numbers, they are numbers with a context*. In mathematics, context obscures structure. In data analysis, context provides meaning." (Moore and Cobb, 1997)

> "In mathematics, context obscures structure. In data analysis, context provides meaning."

Many mathematics problems arise from applied contexts, but the context is removed to reveal mathematical patterns. Statisticians, like mathematicians, look for patterns, but the meaning of the patterns depends on the context.

A graph that occasionally appears in the business section of newspapers shows a plot of the Dow Jones Industrial Average (DJIA) over a 10-year period. The variability of stock prices draws the attention of an investor. This stock index may go up or down over intervals of time, and may fall or rise sharply over a short period. In context, the graph raises questions. A serious investor is not only interested in when or how rapidly the index goes up or down, but also why. What was going on in the world when the market went up; what was going on when it went down? Now strip away the context. Remove time (years) from the horizontal axis and call it "X," remove stock value (DJIA)

from the vertical axis and call it "Y," and there remains a graph of very little interest or mathematical content!

Probability

Probability is a tool for statistics.

Probability is an important part of any mathematical education. It is a part of mathematics that enriches the subject as a whole by its interactions with other uses of mathematics. Probability is an essential tool in applied mathematics and mathematical modeling. It is also an essential tool in statistics.

The use of probability as a mathematical model and the use of probability as a tool in statistics employ not only different approaches, but also different kinds of reasoning. Two problems and the nature of the solutions will illustrate the difference.

Problem 1:

Assume a coin is "fair."

Question: If we toss the coin five times, how many heads will we get?

Problem 2:

You pick up a coin.

Question: Is this a fair coin?

Problem 1 is a mathematical probability problem. Problem 2 is a statistics problem that can use the mathematical probability model determined in Problem 1 as a tool to seek a solution.

The answer to neither question is deterministic. Coin tossing produces random outcomes, which suggests that the answer is probabilistic. The solution to Problem 1 starts with the assumption that the coin is fair and proceeds to logically *deduce* the numerical probabilities for each possible number of heads: 0, 1,…, 5.

The solution to Problem 2 starts with an unfamiliar coin; we don't know if it is fair or biased. The search for an answer is experimental—toss the coin and see what happens. Examine the resulting data to see if it looks as if it came from a fair coin or a biased coin. There are several possible approaches, including toss the coin five times and record the number of heads. Then, do it again: Toss the coin five times and record the number of heads. Repeat 100 times. Compile the frequencies of outcomes for each possible number of heads. Compare these results to the frequencies predicted by the mathematical model for a fair coin in Problem 1. If the empirical frequencies from the experiment are quite dissimilar from those predicted by the mathematical model for a fair coin and are not likely to be caused by random variation in coin tosses, then we conclude that the coin is not fair. In this case, we *induce* an answer by making a general conclusion from observations of experimental results.

Probability and Chance Variability

Two important uses of "randomization" in statistical work occur in sampling and experimental design. When sampling, we "select at random," and in experiments, we randomly assign individuals to different treatments. Randomization does much more than remove bias in selections and assignments. Randomization leads to *chance variability* in outcomes that can be described with probability models.

The probability of something says about what percent of the time it is expected to happen when the basic process is repeated over and over again. Probability theory does not say very much about one toss of a coin; it makes predictions about the long-run behavior of many coin tosses.

Probability tells us little about the consequences of random selection for one sample, but describes the variation we expect to see in samples when the sampling process is repeated a large number of times. Probability tells us little about the consequences of random assignment for one experiment, but describes the variation we expect to see in the results when the experiment is replicated a large number of times.

When randomness is present, the statistician wants to know if the observed result is due to chance or something else. This is the idea of *statistical significance*.

The Role of Mathematics in Statistics Education

The evidence that statistics is different from mathematics is not presented to argue that mathematics is not important to statistics education or that statistics education should not be a part of mathematics education. To the contrary, statistics education becomes increasingly mathematical as the level of understanding goes up. But data collection design, exploration of data, and the interpretation of results should be emphasized in statistics education for statistical literacy. These are heavily dependent on context, and, at the introductory level, involve limited formal mathematics.

Probability plays an important role in statistical analysis, but formal mathematical probability should have its own place in the curriculum. Pre-college statistics education should emphasize the ways probability is used in statistical thinking; an intuitive grasp of probability will suffice at these levels.

In This Section

→ The Role of Variability in the Problem-Solving Process

→ Maturing over Levels

→ The *Framework* Model

→ Illustrations

I. ***Formulate Questions***

Word Length Example

Popular Music Example

Height and Arm Span Example

Plant Growth Example

II. Collect Data

Word Length Example

Plant Growth Example

III. Analyze Data

Popular Music Example

Height and Arm Span Example

IV. Interpret Results

Word Length Example

Plant Growth Example

Nature of Variability

Variability within a Group

Variability within a Group and Variability between Groups

Covariability

Variability in Model Fitting

Induced Variability

Sampling Variability

Chance Variability from Sampling

Chance Variability Resulting from Assignment to Groups in Experiments

→ Detailed Descriptions of Each Level

The Framework

Statistical problem solving is an investigative process that involves four components:

I. Formulate Questions

→ clarify the problem at hand
→ formulate one (or more) questions that can be answered with data

II. Collect Data

→ design a plan to collect appropriate data
→ employ the plan to collect the data

III. Analyze Data

→ select appropriate graphical and numerical methods
→ use these methods to analyze the data

IV. Interpret Results

→ interpret the analysis
→ relate the interpretation to the original question

The Role of Variability in the Problem-Solving Process

I. Formulate Questions

Anticipating Variability—Making the Statistics Question Distinction

The formulation of a statistics question requires an understanding of the difference between a question that anticipates a deterministic answer and a question that anticipates an answer based on data that vary.

The question, "How tall am I?" will be answered with a single height. It is not a statistics question. The question "How tall are adult men in the USA?" would not be a statistics question if all these men were exactly the same height! The fact that there are differing heights, however, implies that we anticipate an answer based on measurements of height that vary. This is a statistics question.

The poser of the question, "How does sunlight affect the growth of a plant?" should anticipate that the growth of two plants of the same type exposed to the same sunlight will likely differ. This is a statistics question.

The anticipation of variability is the basis for understanding the statistics question distinction.

II. Collect Data

Acknowledging Variability—Designing for Differences

Data collection designs must acknowledge variability in data, and frequently are intended to reduce variability. Random sampling is intended to reduce the differences between sample and population. The sample size influences the effect of sampling variability (error). Experimental designs are chosen to acknowledge the differences between groups subjected to different treatments. Random assignment to the groups is intended to reduce differences between the groups due to factors that are not manipulated in the experiment.

> "The formulation of a statistics question requires an understanding of the difference between a question that anticipates a deterministic answer and a question that anticipates an answer based on data that vary."

Some experimental designs pair subjects so they are similar. Twins frequently are paired in medical experiments so that observed differences might be more likely attributed to the difference in treatments, rather than differences in the subjects.

The understanding of data collection designs that acknowledge differences is required for effective collection of data.

III. Analyze Data

Accounting of Variability—Using Distributions

The main purpose of statistical analysis is to give an accounting of the variability in the data. When results of an election poll state "42% of those polled support a particular candidate with margin of error +/- 3% at the 95% confidence level," the focus is on sampling variability. The poll gives an estimate of the support among all voters. The margin of error indicates how far the sample result (42% +/- 3%) might differ from the actual percent of all voters who support the candidate. The confidence level tells us how often estimates produced by the method employed will produce correct results. This analysis is based on the distribution of estimates from repeated random sampling.

When test scores are described as "normally distributed with mean 450 and standard deviation 100," the focus is on how the scores differ from the mean. The normal distribution describes a bell-shaped pattern of scores, and the standard deviation indicates the level of variation of the scores from the mean.

Accounting for variability with the use of distributions is the key idea in the analysis of data.

IV. Interpret Results

Allowing for Variability—Looking beyond the Data

Statistical interpretations are made in the presence of variability and must allow for it.

The result of an election poll must be interpreted as an estimate that can vary from sample to sample. The generalization of the poll results to the entire population of voters looks beyond the sample of voters surveyed and must allow for the possibility of variability of results among different samples. The results of a randomized comparative medical experiment must be interpreted in the presence of variability due to the fact that different individuals respond differently to the same treatment and the variability due to randomization. The generalization of the results looks beyond the data collected from the subjects who participated in the experiment and must allow for these sources of variability.

Looking beyond the data to make generalizations must allow for variability in the data.

Maturing over Levels

The mature statistician understands the role of variability in the statistical problem-solving process. At the point of question formulation, the statistician anticipates the data collection, the nature of the analysis, and the possible interpretations—all of which involve possible sources of variability. In the end, the mature practitioner reflects upon all aspects of data collection and analysis as well as the question, itself, when interpreting results. Likewise, he or she links data collection and analysis to each other and the other two components.

Beginning students cannot be expected to make all of these linkages. They require years of experience and training. Statistical education should be viewed as a developmental process. To meet the proposed goals, this report provides a framework for statistical education over three levels. If the goal were to produce a mature practicing statistician, there certainly would be several levels beyond these. There is no attempt to tie these levels to specific grade levels.

The *Framework* uses three developmental Levels: A, B, and C. Although these three levels may parallel grade levels, they are based on development in statistical literacy, not age. Thus, a middle-school student who has had no prior experience with statistics will need to begin with Level A concepts and activities before moving to Level B. This holds true for a secondary student as well. If a student hasn't had Level A and B experiences prior to high school, then it is not appropriate for that student to jump into Level C expectations. The learning is more teacher-driven at Level A, but becomes student-driven at Levels B and C.

The *Framework* Model

The conceptual structure for statistics education is provided in the two-dimensional model shown in Table 1. One dimension is defined by the problem-solving process components plus the nature of the variability considered and how we focus on variability. The second dimension is comprised of the three developmental levels.

Each of the first four rows describes a process component as it develops across levels. The fifth row indicates the nature of the variability considered at a given level. It is understood that work at Level B assumes and develops further the concepts from Level A; likewise, Level C assumes and uses concepts from the lower levels.

Reading down a column will describe a complete problem investigation for a particular level along with the nature of the variability considered.

Table 1: The Framework

Process Component	Level A	Level B	Level C
I. Formulate Question	**Beginning awareness of the *statistics question distinction*** Teachers pose questions of interest Questions restricted to the classroom	**Increased awareness of the *statistics question distinction*** Students begin to pose their own questions of interest Questions not restricted to the classroom	**Students can make the *statistics question distinction*** Students pose their own questions of interest Questions seek generalization
II. Collect Data	Do not yet *design for differences* Census of classroom Simple experiment	Beginning awareness of *design for differences* Sample surveys; begin to use random selection Comparative experiment; begin to use random allocation	Students make *design for differences* Sampling designs with random selection Experimental designs with randomization
III. Analyze Data	*Use* particular properties of *distributions* in the context of a specific example Display variability within a group Compare individual to individual Compare individual to group Beginning awareness of group to group Observe association between two variables	Learn to *use* particular properties of *distributions* as tools of analysis Quantify variability within a group Compare group to group in displays Acknowledge sampling error Some quantification of association; simple models for association	Understand and *use distributions* in analysis as a global concept Measure variability within a group; measure variability between groups Compare group to group using displays and measures of variability Describe and quantify sampling error Quantification of association; fitting of models for association

Process Component	Level A	Level B	Level C
IV. Interpret Results	Students do not look *beyond the data* No generalization beyond the classroom Note difference between two individuals with different conditions Observe association in displays	Students acknowledge that *looking beyond the data* is feasible Acknowledge that a sample may or may not be representative of the larger population Note the difference between two groups with different conditions Aware of distinction between observational study and experiment Note differences in strength of association Basic interepretation of models for association Aware of the distinction between association and cause and effect	Students are able to *look beyond the data* in some contexts Generalize from sample to population Aware of the effect of randomization on the results of experiments Understand the difference between observational studies and experiments Interpret measures of strength of association Interpret models of association Distinguish between conclusions from association studies and experiments
Nature of Variability	Measurement variability Natural variability Induced variability	Sampling variability	Chance variability
Focus on Variability	Variability within a group	Variability within a group and variability between groups Covariability	Variability in model fitting

"The illustrations of learning activities given here are intended to clarify the differences across the developmental levels for each component of the problem-solving process."

Illustrations

All four steps of the problem-solving process are used at all three levels, but the depth of understanding and sophistication of methods used increases across Levels A, B, and C. This maturation in understanding the problem-solving process and its underlying concepts is paralleled by an increasing complexity in the role of variability. The illustrations of learning activities given here are intended to clarify the differences across the developmental levels for each component of the problem-solving process. Later sections will give illustrations of the complete problem-solving process for learning activities at each level.

I. Formulate Questions

Word Length Example

Level A: How long are the words on this page?

Level B: Are the words in a chapter of a fifth-grade book longer than the words in a chapter of a third-grade book?

Level C: Do fifth-grade books use longer words than third-grade books?

Popular Music Example

Level A: What type of music is most popular among students in our class?

Level B: How do the favorite types of music compare among different classes?

Level C: What type of music is most popular among students in our school?

Height and Arm Span Example

Level A: In our class, are the heights and arm spans of students approximately the same?

Level B: Is the relationship between arm span and height for the students in our class the same as the relationship between arm span and height for the students in another class?

Level C: Is height a useful predictor of arm span for the students in our school?

Plant Growth Example

Level A: Will a plant placed by the window grow taller than a plant placed away from the window?

Level B: Will five plants placed by the window grow taller than five plants placed away from the window?

Level C: How does the level of sunlight affect the growth of plants?

II. Collect Data

Word Length Example

Level A: How long are the words on this page?

The length of every word on the page is determined and recorded.

Level B: Are the words in a chapter of a fifth-grade book longer than the words in a chapter of a third-grade book?

A simple random sample of words from each chapter is used.

Level C: Do fifth-grade books use longer words than third-grade books?

Different sampling designs are considered and compared, and some are used. For example, rather than selecting a simple random sample of words, a simple random sample of pages from the book is selected and all the words on the chosen pages are used for the sample.

Note: At each level, issues of measurement should be addressed. The length of word depends on the definition of "word." For instance, is a number a word? Consistency of definition helps reduce measurement variability.

Plant Growth Example

Level A: Will a plant placed by the window grow taller than a plant placed away from the window?

A seedling is planted in a pot that is placed on the window sill. A second seedling of the same type and size is planted in a pot that is placed away from the window sill. After six weeks, the change in height for each is measured and recorded.

Level B: Will five plants of a particular type placed by the window grow taller than five plants of the same type placed away from the window?

Five seedlings of the same type and size are planted in a pan that is placed on the window sill. Five seedlings of the same type and size are planted in a pan that is placed away from the window sill. Random numbers are used to decide which plants go in the window. After six weeks, the change in height for each seedling is measured and recorded.

Level C: How does the level of sunlight affect the growth of plants?

Fifteen seedlings of the same type and size are selected. Three pans are used, with five of these seedlings planted in each. Fifteen seedlings of another variety are selected to determine if the effect of sunlight is the same on different types of plants. Five of these are planted in each of the three pans. The three pans are placed in locations with three different levels of light. Random numbers are used to decide which plants go in which pan. After six weeks, the change in height for each seedling is measured and recorded.

Note: At each level, issues of measurement should be addressed. The method of measuring change in height must be clearly understood and applied in order to reduce measurement variability.

III. Analyze Data

Popular Music Example

Level A: What type of music is most popular among students in our class?

A bar graph is used to display the number of students who choose each music category.

Level B: How do the favorite types of music compare among different classes?

For each class, a bar graph is used to display the percent of students who choose each music category. The same scales are used for both graphs so that they can easily be compared.

Level C: What type of music is most popular among students in our school?

A bar graph is used to display the percent of students who choose each music category. Because a random sample is used, an estimate of the margin of error is given.

Note: At each level, issues of measurement should be addressed. A questionnaire will be used to gather students' music preferences. The design and wording of the questionnaire must be carefully considered to avoid possible bias in the responses. The choice of music categories also could affect results.

Height and Arm Span Example

Level A: In our class, are the heights and arm spans of students approximately the same?

The difference between height and arm span is determined for each individual. An X-Y plot (scatterplot) is constructed with X = height, Y = arm span. The line Y = X is drawn on this graph.

Level B: Is the relationship between arm span and height for the students in our class the same as the relationship between arm span and height for the students in another class?

For each class, an X-Y plot is constructed with X = height, Y = arm span. An "eye ball" line is drawn on each graph to describe the relationship between height and arm span. The equation of this line is determined. An elementary measure of association is computed.

Level C: Is height a useful predictor of arm span for the students in our school?

The least squares regression line is determined and assessed for use as a prediction model.

Note: At each level, issues of measurement should be addressed. The methods used to measure height and arm span must be clearly understood and applied in order to reduce measurement variability. For instance, do we measure height with shoes on or off?

IV. Interpret Results

Word Length Example

Level A: How long are the words on this page?

The dotplot of all word lengths is examined and summarized. In particular, students will note the longest and shortest word lengths, the most common and least common lengths, and the length in the middle.

Level B: Are the words in a chapter of a fifth-grade book longer than the words in a chapter of a third-grade book?

Students interpret a comparison of the distribution of a sample of word lengths from the fifth-grade book with the distribution of word lengths from the third-grade book using a boxplot to represent each of these. The students also acknowledge that samples are being used that may or may not be representative of the complete chapters.

The boxplot for a sample of word lengths from the fifth-grade book is placed beside the boxplot of the sample from the third-grade book.

Level C: Do fifth-grade books use longer words than third-grade books?

The interpretation at Level C includes the interpretation at Level B, but also must consider generalizing from the books included in the study to a larger population of books.

Plant Growth Example

Level A: Will a plant placed by the window grow taller than a plant placed away from the window?

In this simple experiment, the interpretation is just a matter of comparing one measurement of change in size to another.

Level B: Will five plants placed by the window grow taller than five plants placed away from the window?

In this experiment, the student must interpret a comparison of one group of five measurements with another group. If a difference is noted, then the student acknowledges it is likely caused by the difference in light conditions.

Level C: How does the level of sunlight affect the growth of plants?

There are several comparisons of groups possible with this design. If a difference is noted, then the student acknowledges it is likely caused by the difference in light conditions or the difference in types of plants. It also is acknowledged that the randomization used in the experiment can result in some of the observed differences.

Nature of Variability

The focus on variability grows increasingly more sophisticated as students progress through the developmental levels.

Variability within a Group

This is the only type considered at Level A. In the word length example, differences among word lengths

“The focus on variability grows increasingly more sophisticated as students progress through the developmental levels.”

on a single page are considered; this is variability within a group of word lengths. In the popular music example, differences in how many students choose each category of music are considered; this is variability within a group of frequencies.

Variability within a Group and Variability between Groups

At Level B, students begin to make comparisons of groups of measurements. In the word length example, a group of words from a fifth-grade book is compared to a group from a third-grade book. Such a comparison not only notes how much word lengths differ within each group, but must also take into consideration the differences between the two groups, such as the difference between median or mean word lengths.

Covariability

At Level B, students also begin to investigate the "statistical" relationship between two variables. The nature of this statistical relationship is described in terms of how the two variables "co-vary." In the height and arm span example, for instance, if the heights of two students differ by two centimeters, then we would like our model of the relationship to tell us by how much we might expect their arm spans to differ.

Variability in Model Fitting

At Level C, students assess how well a regression line will predict values of one variable from values of another variable using residual plots. In the height and arm span example, for instance, this assessment is based on examining whether differences between actual arm spans and the arm spans predicted by the model randomly vary about the horizontal line of "no difference" in the residual plot. Inference about a predicted value of y for a given value of x is valid only if the values of y vary at random according to a normal distribution centered on the regression line. Students at Level C learn to estimate this variability about the regression line using the estimated standard deviation of the residuals.

Induced Variability

In the plant growth example at Level B, the experiment is designed to determine if there will be a difference between the growth of plants in sunlight and of plants away from sunlight. We want to determine if an imposed difference on the environments will induce a difference in growth.

Sampling Variability

In the word length example at Level B, samples of words from a chapter are used. Students observe that two samples will produce different groups of word lengths. This is sampling variability.

Chance Variability from Sampling

When random selection is used, differences between samples will be due to chance. Understanding this chance variation is what leads to the predictability of results. In the popular music example, at Level C, this chance variation is not only considered, but is also the basis for understanding the concept of margin of error.

Chance Variability Resulting from Assignment to Groups in Experiments

In the plant growth example at Level C, plants are randomly assigned to groups. Students consider how this chance variation in random assignments might produce differences in results, although a formal analysis is not done.

Detailed Descriptions of Each Level

As this document transitions into detailed descriptions of each level, it is important to note that the examples selected for illustrating key concepts and the problem-solving process of statistical reasoning are based on real data and real-world contexts. *Those of you reading this document are stakeholders, and will need to be flexible in adapting these examples to fit your instructional circumstances.*