

A Student Survey Market Research Project

James Bloomquist

Robert E. Lamberton High School

Overview

The purpose of this project is to expose the students to the type of [survey research](#) conducted by large [market research](#) and polling companies. This will include a presentation of the statistical considerations associated with understanding the [population](#) under consideration, [sampling](#) requirements, data collection techniques, data analyses, [error](#) estimations, and [sample](#) to population [projections](#). This thematic unit explores these concepts within survey research as utilized by marketing research companies, and how it can be adapted to the high school classroom in an experimental model.

The grade level for this project is expected to include students taking an Algebra 2 class, which includes minimal coverage of probability and statistics, or the high school statistics or economics classes.

Rationale

Most people are affected by the marketing research done by companies such as P&G, General Motors, and Coca Cola. Each company spends millions of dollars each year to advertise their products based upon survey research data gathered using sampling principles. Students who feel that they might eventually start their own small businesses, or enter the field of marketing or survey research, or just want to get a better understanding of why companies make some of their product and marketing decisions will gain a better understanding of some of the statistical and financial aspects and ramifications inherent in and resulting from sound [market](#) knowledge. The project explores the two questions, what are the underlying statistical principles, and how can they be explored in the classroom with simple survey experiments conducted by the students?

Standards - Pennsylvania

Mathematics

2.5. Mathematical Problem Solving and Communication

2.5.11. GRADE 11

- A. Select and use appropriate mathematical concepts and techniques from different areas of mathematics and apply them to solving non-routine and multi-step problems.

C. Present mathematical procedures and results clearly, systematically, succinctly and correctly.

2.6. Statistics and Data Analysis

2.6.11. GRADE 11

- A. Design and conduct an experiment using random sampling. Describe the data as an example of a distribution using statistical measures of center and spread. Organize and represent the results with graphs. (Use standard deviation, variance and t-tests.)
- B. Use appropriate technology to organize and analyze data taken from the local community.
- C. Determine the regression equation of best fit (e.g., linear, quadratic, exponential).
- D. Make predictions using interpolation, extrapolation, regression and estimation using technology to verify them.
- E. Determine the validity of the sampling method described in a given study.
- F. Determine the degree of dependence of two quantities specified by a two-way table.
- G. Describe questions of experimental design, control groups, treatment groups, cluster sampling and reliability.
- H. Use sampling techniques to draw inferences about large populations.

Thematic Unit Overview

The purpose of this unit is to serve as a resource for algebra 2, statistics, and economics classes at the high school level. This unit requires algebra as a pre-requisite. In the School District of Philadelphia, the curricula varies for the three courses from a simple exploration of probability, to a more in depth study of statistics and survey [methodology](#).

The overall focus here will cover the requisites necessary to consider acquiring and utilizing market data to make business decisions. The main focuses will include:

- [primary](#) data collection
- [errors in measurement](#)
- [population estimates](#)
- [biased](#) sampling methods
- [under-coverage](#)
- principles of experimental design
- [statistical significance](#)
- [confidence interval](#) for a population
- [mean](#).

Student involvement is crucial for the success of this unit. It allows all students to get involved, and it facilitates the transition from a traditional class format into one where discovery learning is taking place. As students construct their research project, they will be determining the type and scope of consumer products to consider, and the population of individuals to be surveyed for consumer preferences. This unit will provide an opportunity for students to become familiar with the use of the graphing calculator, such as the TI-83 plus, or a spreadsheet program to calculate sums, means, [standard deviations](#), and population projections. The lessons provide an opportunity for students to graph functions, create tables, and create [correlation](#) estimates.

For each of the various lessons, three to five days will be required to complete the assignments. The entire unit should take about 12 classroom days to complete. It is advised that the entire project be spread over several weeks and interspersed into other lessons.

Objectives

The units will enable students to explore the concepts of survey design, sample selection, data gathering, observation weighting, share of market estimation, level of market projection, and standard error of estimates calculations. Each concept will be first presented with a theoretical emphasis, and then follows with practical applications of these theoretical concepts.

Strategies

Most statistics classes cover the concepts of measures of central tendency, probability theory, and sample design at some depth. However, in most high school mathematics classes, these are not given more than a brief overview. Here these concepts are explored as students choose real-world domains that have some interest to them, such as what are the current clothing fashions, or what are the most popular fast food selections. Students will decide on a product class to explore, and up to five items within the product class. Students will determine the population from which they will gather their raw data. In some cases, census estimates may provide measures of the population such that sample to population projections can be made. Students will create a sample frame that best represents the desired population. They will design a data-gathering vehicle that can be administered by the student in a simple interviewing environment. Data will be gathered and evaluated by the students. Final [conclusions](#) will be generated and presentations of the findings prepared and presented.

Classroom Activities / Lesson Plans

Unit 1

The first unit will have the students form groups of four. They will decide on a consumer product class to research.

Since almost all students are addicted to fast food, I have chosen McDonalds as an illustration.

The web site <http://www.fatcalories.com/> contains most fast food chains and provides a listing of their menus including all of the caloric and fat information. This site is enjoyable for most students. Students can easily find a selection of no more than 20 items from this list by simply asking some of their best friends what their favorites are. Here is the list of food items from McDonald's:

Illustration 1

1% Lowfat Chocolate Milk
Jug
1% Lowfat Milk Jug
Apple Dippers
Apple Dippers with Low
Fat Caramel Dip

Bacon Ranch Salad
(without chicken)
Bacon Ranch Salad with
Crispy Chicken
Bacon Ranch Salad with
Grilled Chicken

Bacon, Egg & Cheese
Biscuit
Bacon, Egg, & Cheese
McGriddles®
Baked Apple Pie
Barbeque Sauce
Big Breakfast

Big Mac®
Big N' Tasty®
Big N' Tasty® with Cheese
Biscuit
Butter Garlic Croutons
Caesar Salad (without
chicken)
Caesar Salad with Crispy
Chicken
Caesar Salad with Grilled
Chicken
California Cobb Salad
(without chicken)
California Cobb Salad with
Crispy Chicken
California Cobb Salad with
Grilled Chicken
Cheeseburger
Chicken McGrill®
Sandwich
Chicken McNuggets® (10
piece)
Chicken McNuggets® (20
piece)
Chicken McNuggets® (4
piece)
Chicken McNuggets® (6
piece)
Chicken Selects® Premium
Breast Strips (10 pc)
Chicken Selects® Premium
Breast Strips (3pc)
Chicken Selects® Premium
Breast Strips (5 pc)
Chipotle Barbeque Sauce
Chocolate Chip Cookie
Chocolate Triple Thick
Shake (12 fl oz cup)
Chocolate Triple Thick
Shake (16 fl oz cup)
Chocolate Triple Thick
Shake (21 fl oz cup)
Chocolate Triple Thick
Shake (32 fl oz cup)
CocaCola Classic® (Child)
CocaCola Classic® (Large)
CocaCola Classic®
(Medium)

CocaCola Classic® (Small)
Creamy Ranch Sauce
Crispy Chicken Sandwich
Deluxe Breakfast
Deluxe Warm Cinnamon
Roll
Double Cheeseburger
Double Quarter Pounder®
with Cheese
Egg McMuffin®
English Muffin
Filet-O-Fish® Sandwich
Fruit 'n Yogurt Parfait
Fruit 'n Yogurt Parfait
(without granola)
Grape Jam
Hamburger
Hash Browns
HiC® Orange Lavaburst
(Child)
HiC® Orange Lavaburst
(Large)
HiC® Orange Lavaburst
(Medium)
HiC® Orange Lavaburst
(Small)
Honey
Hot Caramel Sundae
Hot Fudge Sundae
Hot Mustard Sauce
Hot 'n Spicy McChicken®
Sandwich
Hotcakes (Margarine 2 pats
& syrup)
Hotcakes and Sausage
Ketchup Packet
Kiddie Cone
Large French Fries
Low Fat Caramel Dip
M&M's® Candies
McFlurry (12 fl oz cup)
McChicken® Sandwich
McDonaldland® Chocolate
Chip Cookies
McDonaldland® Cookies
Medium French Fries
Minute Maid® Apple Juice
Box

Newman's Own® Cobb
Dressing
Newman's Own® Creamy
Caesar Dressing
Newman's Own® Low Fat
Balsamic Vinaigrette
Newman's Own® Ranch
Dressing
Oatmeal Raisin Cookie
Orange Juice (12 fl oz cup)
Orange Juice (16 fl oz cup)
Orange Juice (21 fl oz cup)
Oreo® McFlurry (12 fl oz
cup)
Peanuts (for Sundaes)
POWERade® Mountain
Blast (Child)
POWERade® Mountain
Blast (Large)
POWERade® Mountain
Blast (Medium)
POWERade® Mountain
Blast (Small)
Quarter Pounder®
Quarter Pounder® with
Cheese
Sausage Biscuit
Sausage Biscuit with Egg
Sausage Burrito
Sausage McGriddles®
Sausage McMuffin®
Sausage McMuffin® with
Egg
Sausage Patty
Sausage, Egg, & Cheese
McGriddles®
Scrambled Eggs (2)
Side Salad
Small French Fries
Spicy Buffalo Sauce
Sprite® (Child)
Sprite® (Large)
Sprite® (Medium)
Sprite® (Small)
Strawberry Preserves
Strawberry Sundae
Strawberry Triple Thick
Shake (12 fl oz cup)

Strawberry Triple Thick
Shake (16 fl oz cup)
Strawberry Triple Thick
Shake (21 fl oz cup)
Strawberry Triple Thick
Shake (32 fl oz cup)
Sugar Cookie
Sweet 'n Sour Sauce

Tangy Honey Mustard
Sauce
Vanilla Reduced Fat Ice
Cream Cone (3.2oz)
Vanilla Triple Thick Shake
(12 fl oz cup)
Vanilla Triple Thick Shake
(16 fl oz cup)

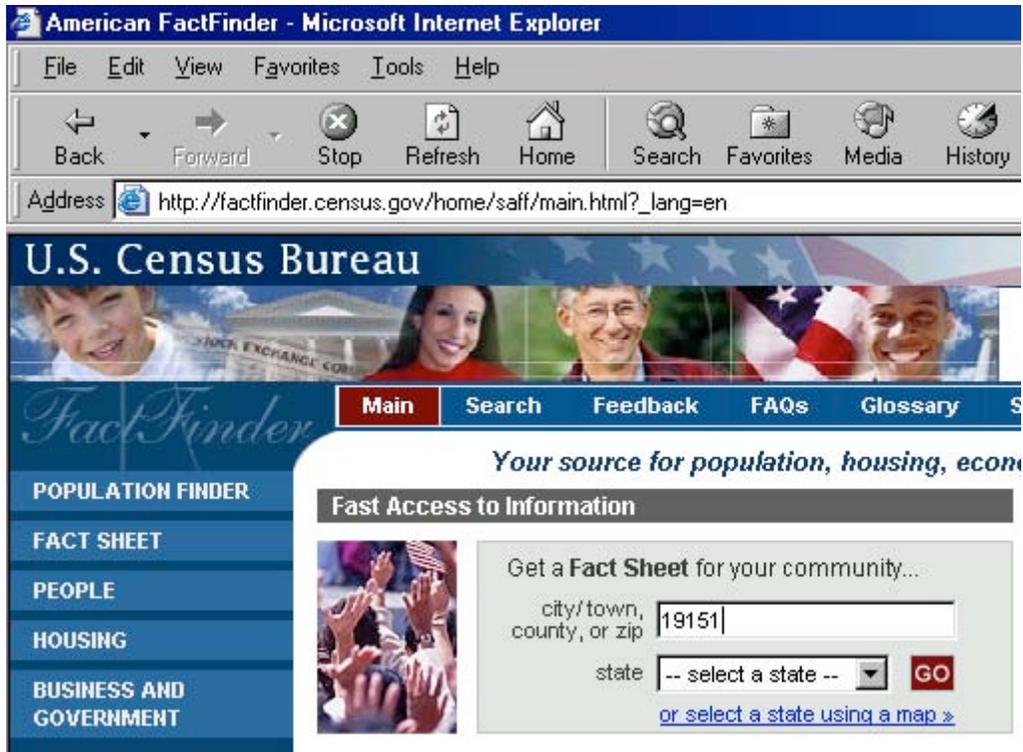
Vanilla Triple Thick Shake
(21 fl oz cup)
Vanilla Triple Thick Shake
(32 fl oz cup)
Warm Cinnamon Roll

Students should take the complete menu listing and narrow it to a dozen to 20 items. Then they need to create a [questionnaire](#). I have provided an example here of some general questions and item specific questions for the first ten items from McDonald's.

<C:\Questionnaire.xls>

They will then determine the appropriate population of consumers of these products. My suggestion is that they look at a [zip code](#) as their geographic boundary. [Demographics](#) are available by zip code at http://factfinder.census.gov/home/saff/main.html?_lang=en. I am demonstrating with zip code 19151.

Illustration 3



Within the zip code they can isolate a segment of the population, such as persons between the ages of 15-19. Under General Characteristics, use the **show more** option to get additional details. In general, students should find this or other similar web sites interesting.

Illustration 4

The screenshot shows the American FactFinder interface. The top navigation bar includes 'Main', 'Search', 'Feedback', 'FAQs', 'Glossary', 'Site Map', and 'Help'. The left sidebar contains various menu items such as 'POPULATION FINDER', 'FACT SHEET', 'PEOPLE', 'HOUSING', 'BUSINESS AND GOVERNMENT', 'ABOUT THE DATA', 'DATA SETS', 'DOWNLOAD CENTER', 'MAPS AND GEOGRAPHY', and 'TOOLS AND REFERENCES'. The main content area displays 'United States | 19151' and 'Zip Code Tabulation Area 19151'. A search box on the right allows for location input. Below the search box, a dropdown menu shows '2000' and a note that '2004 data not available for this geography'. The page title is 'View a Fact Sheet for a race, ethnic, or ancestry group'. The 'Census 2000 Demographic Profile Highlights' section includes a table with columns for 'General Characteristics', 'Number', 'Percent', and 'U.S.'. The table lists various demographic categories such as total population, gender, age groups, and race/ethnicity.

General Characteristics - show more >>	Number	Percent	U.S.		
Total population	31,255			map	brief
Male	13,890	44.4	49.1%	map	brief
Female	17,365	55.6	50.9%	map	brief
Median age (years)	34.1	(X)	35.3	map	brief
Under 5 years	2,212	7.1	6.8%	map	
18 years and over	22,712	72.7	74.3%		
65 years and over	4,255	13.6	12.4%	map	brief
One race	30,548	97.7	97.6%		
White	7,255	23.2	75.1%	map	brief
Black or African American	22,436	71.8	12.3%	map	brief
American Indian and Alaska Native	90	0.3	0.9%	map	brief
Asian	526	1.7	3.6%	map	brief
Native Hawaiian and Other Pacific Islander	3	0.0	0.1%	map	brief
Some other race	238	0.8	5.5%	map	brief
Two or more races	707	2.3	2.4%	map	brief
Hispanic or Latino (of any race)	586	1.9	12.5%	map	brief

This page shows that there were 2,315 persons in the 15-19 age category, or 7.1%, out of a total population of 31,255 for zip code 19151. This 15-19 age demographic will be used later to project the raw sample results to population estimates.

Illustration 5

Address http://factfinder.census.gov/servlet/QTTtable?_bm=y&-geo_id=86000US19151&-qr_name=DEC_2000_SF1_U_C



U.S. Census Bureau
American FactFinder

Main Search Feedback FAQs Glossary

Quick Tables

You are here: [Main](#) ▶ [Data Sets](#) ▶ [Data Sets with Quick Tables](#) ▶ [Geography](#) ▶ [Tables](#) ▶ **Results**

Use the links above to change your results | [Options](#) | [Print / Down](#)

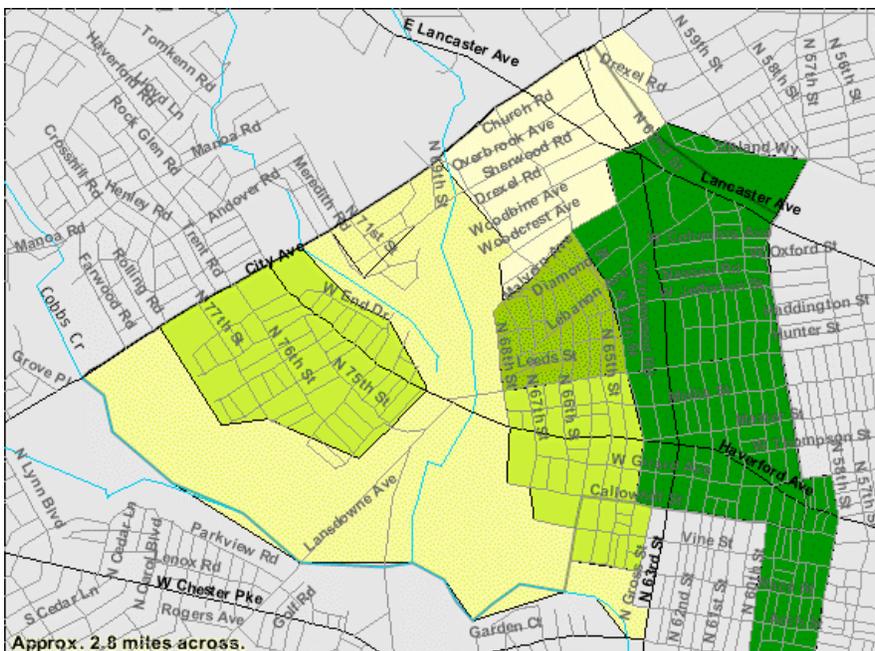
DP-1. Profile of General Demographic Characteristics: 2000
Data Set: [Census 2000 Summary File 1 \(SF 1\) 100-Percent Data](#)
Geographic Area: **19151 5-Digit ZCTA**

NOTE: For information on confidentiality protection, nonsampling error, definitions, and count corrections see <http://factfinder.census.gov/home/en/datanotes/expsf1u.htm>.

Subject	Number	Percent
Total population	31,255	100.0
SEX AND AGE		
Male	13,890	44.4
Female	17,365	55.6
Under 5 years	2,212	7.1
5 to 9 years	2,493	8.0
10 to 14 years	2,547	8.1
15 to 19 years	2,315	7.4
20 to 24 years	2,268	7.3

A map of the area can also be seen by using the **Reference Map** link.

Illustration 6



If students are curious, the web site <http://www.census.gov/dmd/www/pdf/d20ap0.pdf> shows what the United States 2000 Individual [Census report questionnaire](#) looked like.

Unit 2

In the second unit, students will approach the target [sample audience](#) and collect data manually. They will experience the ease and difficulties involved in attempts to collect accurate information from the target sample, including demographics with which to project sample data to population estimates. Priorities may have to be refocused as students realize that they are over-sampling some sectors and under-sampling other sectors of their population. As data are collected, they can be entered into forms that can be entered into hand held calculators or spreadsheets for the analyses to be conducted in Unit 3.

Conducting sixty surveys divided between four students in the group is a manageable task. Over the course of three to five days, each should be able to conduct three to five per day. The group needs to decide what their target population is and how to draw a representative sample within it. At this point assumptions must be made. An example: is the high school enrollment similar to the 15-19 age census category such that it is representative of the category? If it is felt that it is close enough, then a [random sample](#) of the student body would reflect the general population. If, however, the school is an all girls' school, it certainly would not be representative of the population. A random sample here would certainly be a [biased sample](#), completely under representing the male 15-19 population. In such a case, a random canvass of the zip code might be use to discover the location of sufficient 15-19 males. This however would be very time consuming, and therefore impractical. An alternative approach with its own bias would be to survey male 15-19 aged customers at the local McDonald's.

Step 1: Determining the [universe](#) or population to which the sample will ultimately be projected. My suggestion is that it should be the 15-19 age group for the zip codes that best represents the school. In my case it is zip code 19151, a racially mixed area, but the public school is 99% African-American. As a result, any sample drawn just from the students will [under-count](#) the non-African-American population in the zip code. There are two ways to address this sample bias. One is to do a random canvass of the zip code to supplement a random sample drawn in the school. The second is to suggest that the universe represented by the sample will only be the African-American 15-19 age population. It is issues of this nature that make sampling and survey research such a difficult thing. If it were easier to do a satisfactory job, more companies would be doing it. My suggestion here is to make a best estimate of the African-American 15-19 age population. The 2000 census estimate for the 15-19 age group was 2315. The African-American percentage of the total population was 71.8%. Estimating a similar percentage for the 15-19 age group, the estimate of the African-American 15-19 age population is $2315 \times .718$ or 1662 persons.

Step2: It would be a good idea to get the school administration's help when drawing a sample. Ask for a list of the names of all the students in some pre-defined order, for example by class, and within class, by gender, and within gender, alphabetically. In my example, the list would contain the approximately 600 students. As can be seen, this does not come close to the 1662 estimate we created for the 15-19 age group. First, the 15-19 age group actually covers 5 years, so the school population at best should be 80% of the 15-19 age group. An even bigger bias is the fact that many families in this economically middle class neighborhood, chose to send their children to private schools or to some other public school. Dealing with these biases is beyond the scope of this project, and will just be inherent within the data. Anyone attempting to do a similar project will have to make similar compromises.

If you chose to do a [systematic sample](#) of size 60, or 1 in 10 as in my example, arrange the student population in a pre-defined order. Since I want every 10th student, this becomes the skip ratio, and every 10th student is chosen, after choosing the first student. The first student is chosen by creating a random number between 1 and 10 inclusively, which means that any number from 1 to 10 has an equal chance of being chosen. If the random number is 6, then the sample will consist of the 6th, 16th, 26th, 36th, etc. students on the list. This will result in a sample of 60 students. These can then be distributed to the four students in the project group however they deem fit. The idea is to sample this exact group of students.

Step3: Our project group now becomes a team of four surveyors.

As the team begins to collect data they may find resistance among the students to be surveyed. If after a few attempts to get cooperation, a student still refuses to fill out a survey, then an alternate name should be drawn from the school population list. It should be done in this order. Pick a student either one before or one after the non-cooperating student. This gives two more chances to get cooperation for this single survey entry. If after exhausting the original student to be surveyed and the two students on either side, go to the next outer two students and seek cooperation from one of these. In total, this allows five opportunities to solicit cooperation.

An example of this is as follows. If the 56th student was originally chosen but refuses, student 55 or 57 could replace student 56. If all three refuse, students 54 or 58 could be replacements. If all five refuse, randomly chose another student of the same grade and gender. Continue until cooperation is reached.

Attempt to get the best cooperation possible.

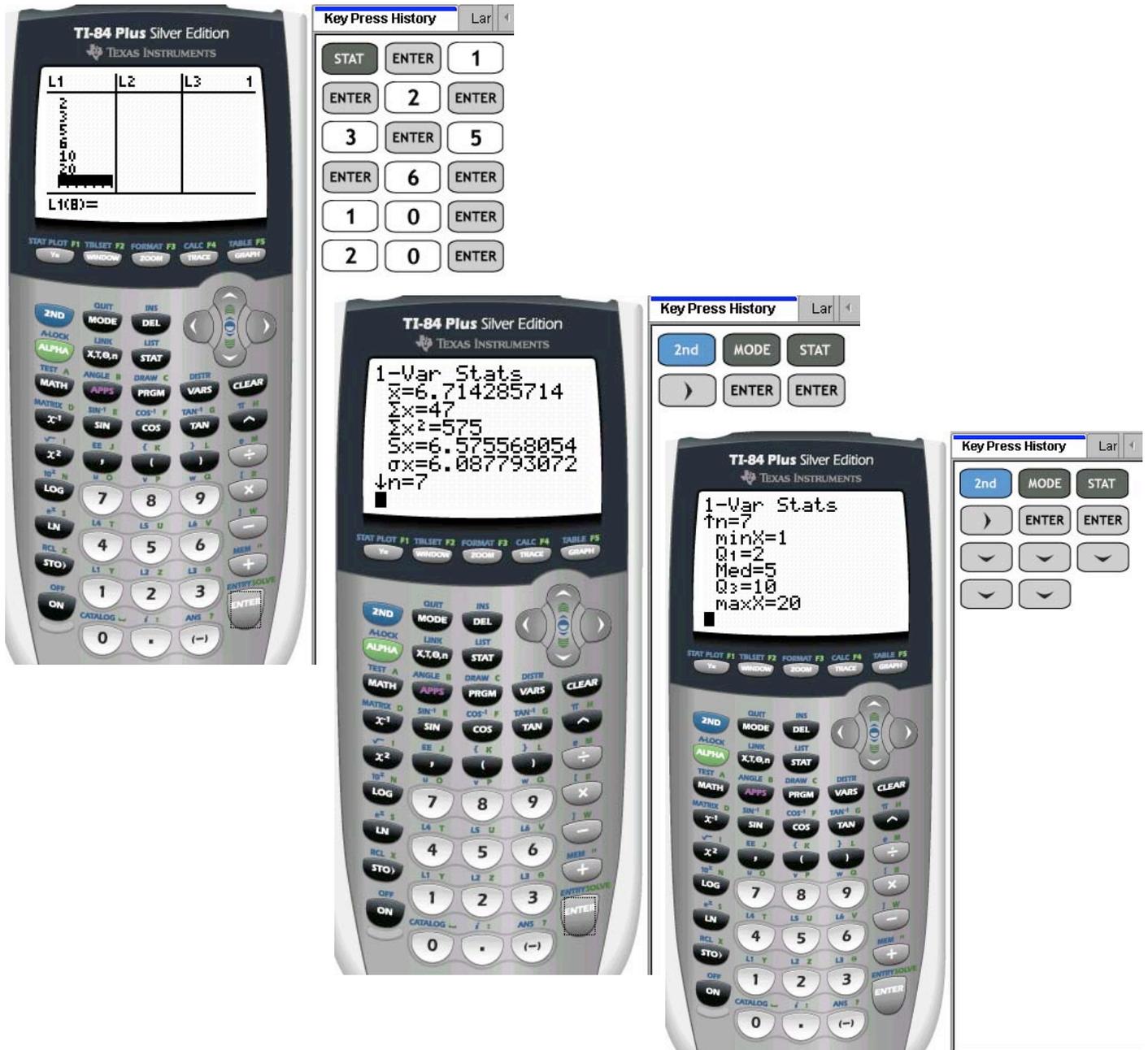
Unit 3

In the final unit, students will analyze the sample survey data that have been collected. They will create sample to population projection factors to be used to forecast population estimates of the product class and brand sales, and brand shares of the product class. Estimates of the standard errors of the population projections will be made. Confidence intervals will be determined. Final estimates of the brands sales, prices, and brand shares can be presented to the entire class as the students complete the unit.

The first step in Unit 3, students enter the survey data into a spreadsheet. Microsoft Excel is probably the best available option. Once the data are in the spreadsheet, statistical measures can be calculated. These will include totals, means, and standard deviations for all of the data. Data may also be projected to population estimates including item sales and dollar values of these sales. Correlations between various items can be calculated as well. Finally, a proposed “special” could be developed that could be presented to the McDonald’s manager as a possible item package to offer to customers. This might consist of two or three items that are purchased together in great frequency.

[C:\Questionnaire.xls](#) contains a suggested Excel spreadsheet format in the sheet named Analysis, but you should create one that fits your particular survey. A TI-83 or greater calculator can be used to create sums, means and standard deviations quite easily as well. The view shown here is from a TI-84, but the functions are the same. Data are entered using the STAT key and others as shown. The data enter here was 1, 2, 3, 5, 6, 10, 20. The mean is 6.71, the sum is 47, and the standard deviation of a sample is 6.57. By using the down arrow key, the minimum, first quartile, [median](#), third quartile, and maximum can also be see.

Illustration 7



Regardless of whether the students use Excel or a calculator, the data needs to be formatted into a useable worksheet so that sums of the answers for each question can be easily calculated. The spreadsheet I have provided is not sufficient for a full survey, since it only shows example data from two questionnaires. Students need to think about how best to layout their own data and then calculate the sums, means, and standard deviations of their sample data. Both Excel and the TI calculators use the following formula for the standard deviation.

$$\sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}}$$

$\sqrt{\quad}$ refers to taking the square root of the value underneath the symbol, n is the number of surveys, x refers to each data item for a given question, and \sum means to sum all of the x or x^2 values.

The following is the Excel description for correlation and covariance.

Correlation analysis tool and formulas

This analysis tool and its formulas measure the relationship between two data sets that are scaled to be independent of the unit of measurement. The population correlation calculation returns the covariance of two data sets divided by the product of their standard deviations:

$$\rho_{X, Y} = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

where

$$\sigma_X^2 = \frac{1}{n} \sum (X_i - \mu_X)^2$$

and

$$\sigma_Y^2 = \frac{1}{n} \sum (Y_i - \mu_Y)^2$$

Covariance analysis tool and formula

Covariance is a measure of the relationship between two ranges of data.

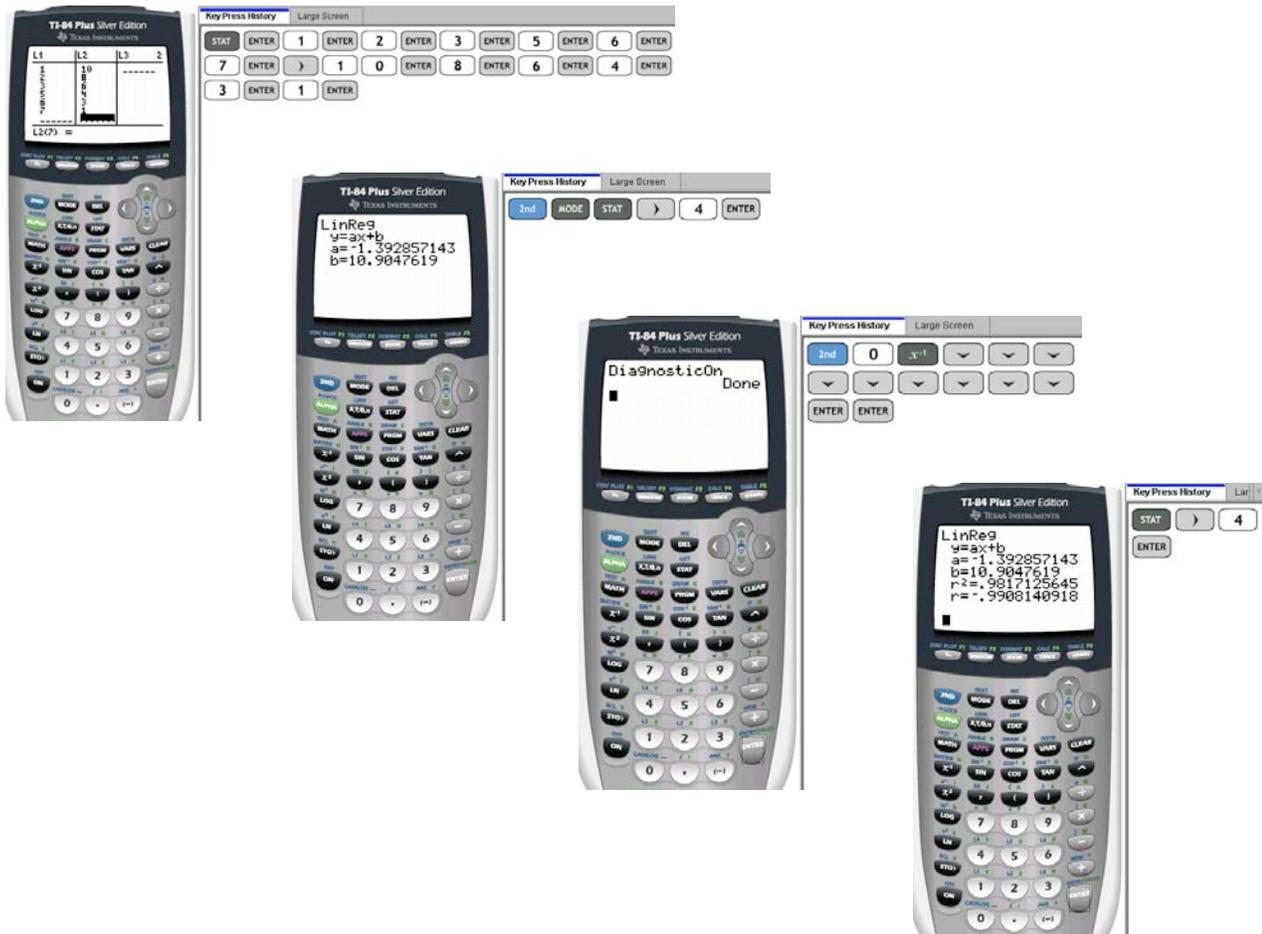
$$\text{cov}(X, Y) = \frac{1}{n} \sum (x_i - \mu_X)(y_i - \mu_Y)$$

You can use the Covariance and Correlation tools to determine whether two ranges of data move together — that is, whether large values of one set are associated with large values of the other (positive covariance), whether small values of one set are associated with large values of the other (negative covariance), or whether values in both sets are unrelated (covariance near zero).

The TI-83 produces a correlation each time a regression equation is calculated between two [variables](#). If the first variable is stored in L1 and the second in L2 and then the STAT CALC function 4:LinReg(ax+b) is executed, the $y=ax+b$ equation is generated, showing the a and b coefficients. If before the regression is executed, the DiagnosticOn is executed, both r and r^2 will be shown in addition to the equation.

The TI-84 images below illustrate the entry of a small amount of data into L1 and L2. Then a linear regression equation is generated, showing the a and b coefficients. Then the DiagnosticOn is executed, and the regression is executed a second time showing the correlation coefficient r and r^2 in addition to the regression equation. Note: The default condition in the TI calculators is to be in DiagnosticOff.

Illustration 8



In the event that two or more items have a correlation of $r^2 = .5$ or more, there is some relationship between the purchase of one with the other. The students should calculate correlations between all of the items. If the survey has 10 items, 45 correlations would result, or ${}_{10}C_2$, (10 choose 2). The students might want to create a product grouping of correlated items to present to the McDonalds store for consideration as a reward for their fine survey work.

The next step in Unit 3 is to create population totals. If the systematic sample is 1 in 10, then the individual items sales totals would be multiplied by a projection factor of 10 to reach the school population estimates. A further projection could be done to the [zip code population](#) by creating a projection factor of the ratio of the zip code 15-19 population to the sample size. This projection, as mentioned earlier may have one or more large biases, but could still be done as part of the project.

Finally, a **presentation of the survey finding** should be prepared. It should include descriptions of every step taken and rationale for each step. It should include a history of the decision making process to arrive at a set of products to survey, the selection of a geographic area, an overview of the demographics of the geographic area, the process used to select a sample, the creation of a survey document, anecdotal tales of the data collection process, examples of completed survey forms, a description of the data entry and analysis steps, final sales units and

dollar estimates calculated from the sample data, comparisons of price and sales volumes between various items surveyed, and any conclusions that the students feel would be a good business practice for the retailer to introduce.

The following is a glossary of terms that might be helpful when considering this project.

Glossary of Terms <http://www.quirks.com/resources/glossary.asp>

Bias

A systematic tendency of a sample to misrepresent the population. Biases may be caused by improper representation of the population in the sample, interviewing techniques, wording of questions, data entry, etc

Confidence interval for a population mean

Choose an SRS of size n from a large population of individuals having mean μ . The mean of the sample observations is \tilde{N} . When n is large, an approximate level C confidence interval for μ is

$$\tilde{N} \pm z^*s/\sqrt{n}$$

where z^* is the critical value for confidence level C from a confidence interval table.

Conclusions

The outcome or result; the section of the final report that contains the interpretation of the data in light of the research objectives.

Correlation analysis

Analysis of the degree to which changes in one variable are associated with changes in another.

Demographics

Description of the vital statistics or objective and quantifiable characteristics of an audience or population. Demographic designators include age, marital status, income, family size, occupation, and personal or household characteristics such as age, sex, income, or educational level.

Editing

The process of ascertaining that questionnaires were filled out properly, completely and accurately.

Errors in measurement

We can think about errors in measurements this way:

$$\text{measured value} = \text{true value} + \text{bias} + \text{random error}$$

A measurement process has bias if it systematically overstates or understates the true value of the property it measures. A measurement process has random error if repeated measurements on the same individual give different results. If the random error is small, we say the measurement is reliable.

Errors in sampling

[Sampling errors](#) are errors caused by the act of taking a sample. They cause sample results to be different from the results of a census. Random sampling error is the deviation between the sample statistic and the population parameter caused by chance in selecting a random sample. The margin of error in a confidence statement includes only random sampling error. Nonsampling errors are errors not related to the act of selecting a sample from the population. They can be present even in a census.

Estimate

A numerical value obtained from a statistical sample and assigned to the population parameter.

Expected value

The mean of a probability distribution. It is the value of the probability distribution we would expect in the long run.

Managing bias and variability

To reduce bias, use random sampling. When we start with a list of the entire population, simple random sampling produces unbiased estimates—the values of a statistic computed from an SRS neither consistently overestimate nor consistently underestimate the value of the population parameter.

To reduce the variability of an SRS, use a larger sample. You can make the variability as small as you want by taking a large enough sample.

Market

Total of all individuals or organizations that represent potential buyers.

Marketing

The process of planning and executing the conception, pricing, promotion and distribution of ideas, goods, and services to create exchanges that satisfy individual and organizational objectives.

Marketing research

The planning, collection, and analysis of data relevant to marketing decision making, and the communication of the results of this analysis to management.

Median and mean of a density curve

The median of a density curve is the equal-areas point, the point that divides the area under the curve in half.

The mean of a density curve is the balance point, at which the curve would balance if made of solid material.

The median and mean are the same for a symmetric density curve.

They both lie at the center of the curve. The mean of a skewed curve is pulled away from the median in the direction of the long tail.

Methodology

The research procedures used; the section of the final report in which the researcher outlines the approach used in the research, including the method of recruiting participants, the types of questions used, and so on. Methodology can also mean the approach a moderator uses to conduct focus groups.

Nonprobability sample

Subset of a population in which little or no attempt is made to ensure a representative cross section.

Non-random

Occurrences which do not have an equal probability of occurring; not mathematically predictable on the basis of the classical theory of probability.

Nonresponse bias

Error that results from a systematic difference between those who do and do not respond to the measurement instrument.

Nonsampling error

All the sources of bias or inaccuracy in a study besides sampling error. Examples: leading by the interviewer, recording/data entry errors.

Objectives

The information to be developed from a study to serve the project's purpose.

Observation

The value that the variable assumes for a single unit of the sample.

Participant

A person included in a focus group, survey or study. Also called respondent, unit, subject, experimental unit or unit of analysis.

Population

The collection of all objects that are of interest to the statistician. The elements of a population may be called units or subjects. Also known as the universe.

Primary research

Conducting research to collect new data to solve a marketing information need.

Probability sample

A sample in which every unit has an equal (nonzero) and known probability of being selected. Sometimes called a random sample.

Projection

An estimate, based on assumptions about future trends in births, deaths, and migration, of a demographic characteristic such as population or number of households. Forecasts and projections are terms that are often used interchangeably.

Questionnaire

A set of questions designed to generate data necessary for accomplishing the objectives of the research project.

Random sampling

A sample in which each unit has an equal and independent chance of selection. Also known as [probability sample](#).

Response bias

Error that results from the tendency of people to answer a question falsely, through deliberate misrepresentation or unconscious falsification.

Sample

A subset of the population of interest selected for a research study. It is a finite portion that is used to study the characteristics of concern in the population

Sample population

The population from which the sample is obtained.

Sampling

The method of selecting a specified portion, called a sample, from a population, from which information concerning the whole can be inferred.

Sampling error

The estimated inaccuracy of the results of a study when a population sample is used to explain behavior of the total population.

Short census form

U.S. Census Bureau questionnaire that all Americans answer every 10 years.

Standard deviation

The positive square root of average squared distance of the population or sample values from the mean. It is the most widely accepted measure of dispersion.

Statistical significance

An observed effect so large that it would rarely occur by chance is called statistically significant.

Survey objectives

The decision-making information sought through the questionnaire.

Survey research

Research in which an interviewer interacts with respondents to obtain facts, opinions and attitudes.

Systematic sample

A procedure that selects every Nth unit (skip interval) of a population until the desired sample size is reached. The starting point should be a random position.

Under-coverage

Under-coverage occurs when some groups in the population are left out of the process of choosing the sample.

Universe

The set of all the units from which a sample is drawn. Also called the population.

Variable

Any characteristic that can be measured on each unit of the population.

ZIP code

Registered trademark of the U.S. Postal Service; a five-digit or nine-digit code identifying regions in the United States.

ZIP code demographics

The demographic characteristics of a population living in a particular ZIP code.

Bibliography

Chamberlin, Michael. "The Fast Food Nutrition Fact Explorer". <www.fatcalories.com>. May 2006

US Department of Commerce, "U.S. Census Bureau Fact Finder".
<http://factfinder.census.gov/home/saff/main.html?_lang=en>. May 2006.

Bureau of the Census, "United States Census 2000 Individual Census Report".
<<http://www.census.gov/dmd/www/pdf/d20ap0.pdf>>. May 2006.

Quirk Enterprises, Inc. "Quirk's Market Research Review".
<<http://www.quirks.com/resources/glossary.asp>>. May 2006

