

# ENGR 2113 ECE Math

**Collection Editor:**

David Waldo



# ENGR 2113 ECE Math

**Collection Editor:**

David Waldo

**Authors:**

Susan Dean  
Kenny M. Felder  
Barbara Illowsky, Ph.D.

David Lane  
Paul E Pfeiffer  
Sunil Kumar Singh

**Online:**

< <http://cnx.org/content/col11224/1.1/> >

**C O N N E X I O N S**

Rice University, Houston, Texas

This selection and arrangement of content as a collection is copyrighted by David Waldo. It is licensed under the Creative Commons Attribution 2.0 license (<http://creativecommons.org/licenses/by/2.0/>).

Collection structure revised: August 27, 2010

PDF generated: June 20, 2012

For copyright and attribution information for the modules contained in this collection, see p. 346.

# Table of Contents

## 1 Sampling and Data

1.1 Key Terms	1
1.2 Data	3
1.3 Sampling	4
1.4 Variation	10
1.5 Answers and Rounding Off	12
1.6 Frequency	12
1.7 Summary	17
1.8 Practice: Sampling and Data	18
Solutions	21

## 2 Descriptive Statistics

2.1 Stem and Leaf Graphs (Stemplots), Line Graphs and Bar Graphs	23
2.2 Histograms	27
2.3 Measures of the Center of the Data	30
2.4 Skewness and the Mean, Median, and Mode	33
2.5 Measures of the Spread of the Data	35
2.6 Summary of Formulas	44
2.7 Practice 1: Center of the Data	45
2.8 Practice 2: Spread of the Data	48
Solutions	49

## 3 Sets and Counting

3.1 Sets	51
3.2 Subsets	55
3.3 Venn Diagrams (optional)	62
3.4 Union of sets	63
3.5 Intersection of sets	72
3.6 Difference of sets	82
3.7 Working with two sets	94
3.8 Working with three sets	101
3.9 Cartesian product	110
3.10 Cartesian Product (exercise)	119
3.11 Probability Concepts – Probability	126
3.12 Probability Concepts – Permutations	131
3.13 Probability Concepts – Combinations	132
3.14 Tree Diagrams (optional)	133
Solutions	137

## 4 Probability Topics

4.1 Probability Topics	139
4.2 Probability	140
4.3 Terminology	140
4.4 Two Basic Rules of Probability	143
4.5 Independent and Mutually Exclusive Events	146
4.6 Contingency Tables	149
4.7 Summary of Formulas	153
4.8 Practice 1: Contingency Tables	154
4.9 Practice 2: Calculating Probabilities	156

4.10 Homework .....	157
4.11 Review .....	168
Solutions .....	170
<b>5 Discrete Random Variables</b>	
5.1 Discrete Random Variables .....	175
5.2 Probability Distribution Function (PDF) for a Discrete Random Variable .....	176
5.3 Mean or Expected Value and Standard Deviation .....	177
5.4 Common Discrete Probability Distribution Functions .....	180
5.5 Binomial .....	180
5.6 Geometric (optional) .....	183
5.7 Hypergeometric (optional) .....	186
5.8 Poisson .....	188
5.9 Summary of Functions .....	191
5.10 Practice 1: Discrete Distribution .....	193
5.11 Practice 2: Binomial Distribution .....	194
5.12 Practice 3: Poisson Distribution .....	196
5.13 Practice 4: Geometric Distribution .....	197
5.14 Practice 5: Hypergeometric Distribution .....	199
5.15 Homework .....	200
5.16 Review .....	210
5.17 Lab 1: Discrete Distribution (Playing Card Experiment) .....	213
5.18 Lab 2: Discrete Distribution (Lucky Dice Experiment) .....	217
5.19 Binomial Distribution .....	220
Solutions .....	224
<b>6 Continuous Random Variables</b>	
6.1 Continuous Random Variables .....	231
6.2 Continuous Probability Functions .....	233
6.3 The Uniform Distribution .....	236
6.4 The Exponential Distribution .....	243
6.5 Summary of the Uniform and Exponential Probability Distributions .....	249
6.6 Practice 1: Uniform Distribution .....	250
6.7 Practice 2: Exponential Distribution .....	253
6.8 Homework .....	255
6.9 Review .....	261
6.10 Lab: Continuous Distribution .....	264
Solutions .....	267
<b>7 The Normal Distribution</b>	
7.1 The Normal Distribution .....	271
7.2 The Standard Normal Distribution .....	272
7.3 Z-scores .....	273
7.4 Areas to the Left and Right of $x$ .....	275
7.5 Calculations of Probabilities .....	275
7.6 Summary of Formulas .....	279
7.7 Practice: The Normal Distribution .....	280
7.8 Homework .....	282
7.9 Review .....	288
7.10 Lab 1: Normal Distribution (Lap Times) .....	290
7.11 Lab 2: Normal Distribution (Pinkie Length) .....	293
Solutions .....	295
<b>8 Appendix</b>	

8.1	Practice Final Exam 1 .....	299
8.2	Practice Final Exam 2 .....	308
8.3	English Phrases Written Mathematically .....	317
8.4	Symbols and their Meanings .....	318
8.5	Formulas .....	323
8.6	Some Mathematical Formulas .....	325
	Solutions .....	329
<b>9</b>	<b>Tables</b> .....	<b>333</b>
	<b>Glossary</b> .....	<b>334</b>
	<b>Index</b> .....	<b>342</b>
	<b>Attributions</b> .....	<b>346</b>





# Chapter 1

## Sampling and Data

### 1.1 Key Terms<sup>1</sup>

In statistics, we generally want to study a **population**. You can think of a population as an entire collection of persons, things, or objects under study. To study the larger population, we select a **sample**. The idea of **sampling** is to select a portion (or subset) of the larger population and study that portion (the sample) to gain information about the population. Data are the result of sampling from a population.

Because it takes a lot of time and money to examine an entire population, sampling is a very practical technique. If you wished to compute the overall grade point average at your school, it would make sense to select a sample of students who attend the school. The data collected from the sample would be the students' grade point averages. In presidential elections, opinion poll samples of 1,000 to 2,000 people are taken. The opinion poll is supposed to represent the views of the people in the entire country. Manufacturers of canned carbonated drinks take samples to determine if a 16 ounce can contains 16 ounces of carbonated drink.

From the sample data, we can calculate a statistic. A **statistic** is a number that is a property of the sample. For example, if we consider one math class to be a sample of the population of all math classes, then the average number of points earned by students in that one math class at the end of the term is an example of a statistic. The statistic is an estimate of a population parameter. A **parameter** is a number that is a property of the population. Since we considered all math classes to be the population, then the average number of points earned per student over all the math classes is an example of a parameter.

One of the main concerns in the field of statistics is how accurately a statistic estimates a parameter. The accuracy really depends on how well the sample represents the population. The sample must contain the characteristics of the population in order to be a **representative sample**. We are interested in both the sample statistic and the population parameter in inferential statistics. In a later chapter, we will use the sample statistic to test the validity of the established population parameter.

A **variable**, notated by capital letters like  $X$  and  $Y$ , is a characteristic of interest for each person or thing in a population. Variables may be **numerical** or **categorical**. **Numerical variables** take on values with equal units such as weight in pounds and time in hours. **Categorical variables** place the person or thing into a category. If we let  $X$  equal the number of points earned by one math student at the end of a term, then  $X$  is a numerical variable. If we let  $Y$  be a person's party affiliation, then examples of  $Y$  include Republican, Democrat, and Independent.  $Y$  is a categorical variable. We could do some math with values of  $X$  (calculate the average number of points earned, for example), but it makes no sense to do math with values of  $Y$  (calculating an average party affiliation makes no sense).

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m16007/1.16/>>.

**Data** are the actual values of the variable. They may be numbers or they may be words. Datum is a single value.

Two words that come up often in statistics are **mean** and **proportion**. If you were to take three exams in your math classes and obtained scores of 86, 75, and 92, you calculate your mean score by adding the three exam scores and dividing by three (your mean score would be 84.3 to one decimal place). If, in your math class, there are 40 students and 22 are men and 18 are women, then the proportion of men students is  $\frac{22}{40}$  and the proportion of women students is  $\frac{18}{40}$ . Mean and proportion are discussed in more detail in later chapters.

NOTE: The words "mean" and "average" are often used interchangeably. The substitution of one word for the other is common practice. The technical term is "arithmetic mean" and "average" is technically a center location. However, in practice among non-statisticians, "average" is commonly accepted for "arithmetic mean."

### Example 1.1

Define the key terms from the following study: We want to know the average amount of money first year college students spend at ABC College on school supplies that do not include books. We randomly survey 100 first year students at the college. Three of those students spent \$150, \$200, and \$225, respectively.

#### Solution

The **population** is all first year students attending ABC College this term.

The **sample** could be all students enrolled in one section of a beginning statistics course at ABC College (although this sample may not represent the entire population).

The **parameter** is the average amount of money spent (excluding books) by first year college students at ABC College this term.

The **statistic** is the average amount of money spent (excluding books) by first year college students in the sample.

The **variable** could be the amount of money spent (excluding books) by one first year student. Let  $X$  = the amount of money spent (excluding books) by one first year student attending ABC College.

The **data** are the dollar amounts spent by the first year students. Examples of the data are \$150, \$200, and \$225.

### 1.1.1 Optional Collaborative Classroom Exercise

Do the following exercise collaboratively with up to four people per group. Find a population, a sample, the parameter, the statistic, a variable, and data for the following study: You want to determine the average number of glasses of milk college students drink per day. Suppose yesterday, in your English class, you asked five students how many glasses of milk they drank the day before. The answers were 1, 0, 1, 3, and 4 glasses of milk.

## 1.2 Data<sup>2</sup>

Data may come from a population or from a sample. Small letters like  $x$  or  $y$  generally are used to represent data values. Most data can be put into the following categories:

- Qualitative
- Quantitative

**Qualitative data** are the result of categorizing or describing attributes of a population. Hair color, blood type, ethnic group, the car a person drives, and the street a person lives on are examples of qualitative data. Qualitative data are generally described by words or letters. For instance, hair color might be black, dark brown, light brown, blonde, gray, or red. Blood type might be AB+, O-, or B+. Researchers often prefer to use quantitative data over qualitative data because it lends itself more easily to mathematical analysis. For example, it does not make sense to find an average hair color or blood type.

**Quantitative data** are always numbers. Quantitative data are the result of **counting** or **measuring** attributes of a population. Amount of money, pulse rate, weight, number of people living in your town, and the number of students who take statistics are examples of quantitative data. Quantitative data may be either **discrete** or **continuous**.

All data that are the result of counting are called **quantitative discrete data**. These data take on only certain numerical values. If you count the number of phone calls you receive for each day of the week, you might get 0, 1, 2, 3, etc.

All data that are the result of measuring are **quantitative continuous data** assuming that we can measure accurately. Measuring angles in radians might result in the numbers  $\frac{\pi}{6}$ ,  $\frac{\pi}{3}$ ,  $\frac{\pi}{2}$ ,  $\pi$ ,  $\frac{3\pi}{4}$ , etc. If you and your friends carry backpacks with books in them to school, the numbers of books in the backpacks are discrete data and the weights of the backpacks are continuous data.

### Example 1.2: Data Sample of Quantitative Discrete Data

The data are the number of books students carry in their backpacks. You sample five students. Two students carry 3 books, one student carries 4 books, one student carries 2 books, and one student carries 1 book. The numbers of books (3, 4, 2, and 1) are the quantitative discrete data.

### Example 1.3: Data Sample of Quantitative Continuous Data

The data are the weights of the backpacks with the books in it. You sample the same five students. The weights (in pounds) of their backpacks are 6.2, 7, 6.8, 9.1, 4.3. Notice that backpacks carrying three books can have different weights. Weights are quantitative continuous data because weights are measured.

### Example 1.4: Data Sample of Qualitative Data

The data are the colors of backpacks. Again, you sample the same five students. One student has a red backpack, two students have black backpacks, one student has a green backpack, and one student has a gray backpack. The colors red, black, black, green, and gray are qualitative data.

NOTE: You may collect data as numbers and report it categorically. For example, the quiz scores for each student are recorded throughout the term. At the end of the term, the quiz scores are reported as A, B, C, D, or F.

### Example 1.5

Work collaboratively to determine the correct data type (quantitative or qualitative). Indicate whether quantitative data are continuous or discrete. Hint: Data that are discrete often start with the words "the number of."

<sup>2</sup>This content is available online at <<http://cnx.org/content/m16005/1.15/>>.

1. The number of pairs of shoes you own.
2. The type of car you drive.
3. Where you go on vacation.
4. The distance it is from your home to the nearest grocery store.
5. The number of classes you take per school year.
6. The tuition for your classes
7. The type of calculator you use.
8. Movie ratings.
9. Political party preferences.
10. Weight of sumo wrestlers.
11. Amount of money (in dollars) won playing poker.
12. Number of correct answers on a quiz.
13. Peoples' attitudes toward the government.
14. IQ scores. (This may cause some discussion.)

### 1.3 Sampling<sup>3</sup>

Gathering information about an entire population often costs too much or is virtually impossible. Instead, we use a sample of the population. **A sample should have the same characteristics as the population it is representing.** Most statisticians use various methods of random sampling in an attempt to achieve this goal. This section will describe a few of the most common methods.

There are several different methods of **random sampling**. In each form of random sampling, each member of a population initially has an equal chance of being selected for the sample. Each method has pros and cons. The easiest method to describe is called a **simple random sample**. Any group of  $n$  individuals is equally likely to be chosen by any other group of  $n$  individuals if the simple random sampling technique is used. In other words, each sample of the same size has an equal chance of being selected. For example, suppose Lisa wants to form a four-person study group (herself and three other people) from her pre-calculus class, which has 31 members not including Lisa. To choose a simple random sample of size 3 from the other members of her class, Lisa could put all 31 names in a hat, shake the hat, close her eyes, and pick out 3 names. A more technological way is for Lisa to first list the last names of the members of her class together with a two-digit number as shown below.

---

<sup>3</sup>This content is available online at <<http://cnx.org/content/m16014/1.17/>>.

**Class Roster**

<b>ID</b>	<b>Name</b>
00	Anselmo
01	Bautista
02	Bayani
03	Cheng
04	Cuarismo
05	Cunningham
06	Fontecha
07	Hong
08	Hoobler
09	Jiao
10	Khan
11	King
12	Legeny
13	Lundquist
14	Macierz
15	Motogawa
16	Okimoto
17	Patel
18	Price
19	Quizon
20	Reyes
21	Roquero
22	Roth
23	Rowell
24	Salangsang
25	Slade
26	Stracher
27	Tallai
28	Tran
29	Wai
30	Wood

**Table 1.1**

Lisa can either use a table of random numbers (found in many statistics books as well as mathematical handbooks) or a calculator or computer to generate random numbers. For this example, suppose Lisa chooses to generate random numbers from a calculator. The numbers generated are:

.94360; .99832; .14669; .51470; .40581; .73381; .04399

Lisa reads two-digit groups until she has chosen three class members (that is, she reads .94360 as the groups 94, 43, 36, 60). Each random number may only contribute one class member. If she needed to, Lisa could have generated more random numbers.

The random numbers .94360 and .99832 do not contain appropriate two digit numbers. However the third random number, .14669, contains 14 (the fourth random number also contains 14), the fifth random number contains 05, and the seventh random number contains 04. The two-digit number 14 corresponds to Macierz, 05 corresponds to Cunningham, and 04 corresponds to Cuarismo. Besides herself, Lisa's group will consist of Marcierz, and Cunningham, and Cuarismo.

Besides simple random sampling, there are other forms of sampling that involve a chance process for getting the sample. **Other well-known random sampling methods are the stratified sample, the cluster sample, and the systematic sample.**

To choose a **stratified sample**, divide the population into groups called strata and then take a **proportionate** number from each stratum. For example, you could stratify (group) your college population by department and then choose a proportionate simple random sample from each stratum (each department) to get a stratified random sample. To choose a simple random sample from each department, number each member of the first department, number each member of the second department and do the same for the remaining departments. Then use simple random sampling to choose proportionate numbers from the first department and do the same for each of the remaining departments. Those numbers picked from the first department, picked from the second department and so on represent the members who make up the stratified sample.

To choose a **cluster sample**, divide the population into clusters (groups) and then randomly select some of the clusters. All the members from these clusters are in the cluster sample. For example, if you randomly sample four departments from your college population, the four departments make up the cluster sample. For example, divide your college faculty by department. The departments are the clusters. Number each department and then choose four different numbers using simple random sampling. All members of the four departments with those numbers are the cluster sample.

To choose a **systematic sample**, randomly select a starting point and take every  $n$ th piece of data from a listing of the population. For example, suppose you have to do a phone survey. Your phone book contains 20,000 residence listings. You must choose 400 names for the sample. Number the population 1 - 20,000 and then use a simple random sample to pick a number that represents the first name of the sample. Then choose every 50th name thereafter until you have a total of 400 names (you might have to go back to the of your phone list). Systematic sampling is frequently chosen because it is a simple method.

A type of sampling that is nonrandom is convenience sampling. **Convenience sampling** involves using results that are readily available. For example, a computer software store conducts a marketing study by interviewing potential customers who happen to be in the store browsing through the available software. The results of convenience sampling may be very good in some cases and highly biased (favors certain outcomes) in others.

Sampling data should be done very carefully. Collecting data carelessly can have devastating results. Surveys mailed to households and then returned may be very biased (for example, they may favor a certain group). It is better for the person conducting the survey to select the sample respondents.

True random sampling is done **with replacement**. That is, once a member is picked that member goes back into the population and thus may be chosen more than once. However for practical reasons, in most populations, simple random sampling is done **without replacement**. Surveys are typically done without replacement. That is, a member of the population may be chosen only once. Most samples are taken from large populations and the sample tends to be small in comparison to the population. Since this is the case,

sampling without replacement is approximately the same as sampling with replacement because the chance of picking the same individual more than once using with replacement is very low.

For example, in a college population of 10,000 people, suppose you want to randomly pick a sample of 1000 for a survey. **For any particular sample of 1000**, if you are sampling **with replacement**,

- the chance of picking the first person is 1000 out of 10,000 (0.1000);
- the chance of picking a different second person for this sample is 999 out of 10,000 (0.0999);
- the chance of picking the same person again is 1 out of 10,000 (very low).

If you are sampling **without replacement**,

- the chance of picking the first person for any particular sample is 1000 out of 10,000 (0.1000);
- the chance of picking a different second person is 999 out of 9,999 (0.0999);
- you do not replace the first person before picking the next person.

Compare the fractions  $999/10,000$  and  $999/9,999$ . For accuracy, carry the decimal answers to 4 place decimals. To 4 decimal places, these numbers are equivalent (0.0999).

Sampling without replacement instead of sampling with replacement only becomes a mathematics issue when the population is small which is not that common. For example, if the population is 25 people, the sample is 10 and you are sampling **with replacement for any particular sample**,

- the chance of picking the first person is 10 out of 25 and a different second person is 9 out of 25 (you replace the first person).

If you sample **without replacement**,

- the chance of picking the first person is 10 out of 25 and then the second person (which is different) is 9 out of 24 (you do not replace the first person).

Compare the fractions  $9/25$  and  $9/24$ . To 4 decimal places,  $9/25 = 0.3600$  and  $9/24 = 0.3750$ . To 4 decimal places, these numbers are not equivalent.

When you analyze data, it is important to be aware of **sampling errors** and nonsampling errors. The actual process of sampling causes sampling errors. For example, the sample may not be large enough. Factors not related to the sampling process cause **nonsampling errors**. A defective counting device can cause a nonsampling error.

In reality, a sample will never be exactly representative of the population so there will always be some sampling error. As a rule, the larger the sample, the smaller the sampling error.

In statistics, a **sampling bias** is created when a sample is collected from a population and some members of the population are not as likely to be chosen as others (remember, each member of the population should have an equally likely chance of being chosen). When a sampling bias happens, there can be incorrect conclusions drawn about the population that is being studied.

#### Example 1.6

Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience).

1. A soccer coach selects 6 players from a group of boys aged 8 to 10, 7 players from a group of boys aged 11 to 12, and 3 players from a group of boys aged 13 to 14 to form a recreational soccer team.
2. A pollster interviews all human resource personnel in five different high tech companies.

3. A high school educational researcher interviews 50 high school female teachers and 50 high school male teachers.
4. A medical researcher interviews every third cancer patient from a list of cancer patients at a local hospital.
5. A high school counselor uses a computer to generate 50 random numbers and then picks students whose names correspond to the numbers.
6. A student interviews classmates in his algebra class to determine how many pairs of jeans a student owns, on the average.

**Solution**

1. stratified
2. cluster
3. stratified
4. systematic
5. simple random
6. convenience

If we were to examine two samples representing the same population, even if we used random sampling methods for the samples, they would not be exactly the same. Just as there is variation in data, there is variation in samples. As you become accustomed to sampling, the variability will seem natural.

**Example 1.7**

Suppose ABC College has 10,000 part-time students (the population). We are interested in the average amount of money a part-time student spends on books in the fall term. Asking all 10,000 students is an almost impossible task.

Suppose we take two different samples.

First, we use convenience sampling and survey 10 students from a first term organic chemistry class. Many of these students are taking first term calculus in addition to the organic chemistry class. The amount of money they spend is as follows:

\$128; \$87; \$173; \$116; \$130; \$204; \$147; \$189; \$93; \$153

The second sample is taken by using a list from the P.E. department of senior citizens who take P.E. classes and taking every 5th senior citizen on the list, for a total of 10 senior citizens. They spend:

\$50; \$40; \$36; \$15; \$50; \$100; \$40; \$53; \$22; \$22

**Problem 1**

Do you think that either of these samples is representative of (or is characteristic of) the entire 10,000 part-time student population?

**Solution**

**No.** The first sample probably consists of science-oriented students. Besides the chemistry course, some of them are taking first-term calculus. Books for these classes tend to be expensive. Most of these students are, more than likely, paying more than the average part-time student for their books. The second sample is a group of senior citizens who are, more than likely, taking courses for health and interest. The amount of money they spend on books is probably much less than the average part-time student. Both samples are biased. Also, in both cases, not all students have a chance to be in either sample.



**Problem 2**

Since these samples are not representative of the entire population, is it wise to use the results to describe the entire population?

**Solution**

**No.** For these samples, each member of the population did not have an equally likely chance of being chosen.

Now, suppose we take a third sample. We choose ten different part-time students from the disciplines of chemistry, math, English, psychology, sociology, history, nursing, physical education, art, and early childhood development. (We assume that these are the only disciplines in which part-time students at ABC College are enrolled and that an equal number of part-time students are enrolled in each of the disciplines.) Each student is chosen using simple random sampling. Using a calculator, random numbers are generated and a student from a particular discipline is selected if he/she has a corresponding number. The students spend:

\$180; \$50; \$150; \$85; \$260; \$75; \$180; \$200; \$200; \$150

**Problem 3**

Is the sample biased?

**Solution**

The sample is unbiased, but a larger sample would be recommended to increase the likelihood that the sample will be close to representative of the population. However, for a biased sampling technique, even a large sample runs the risk of not being representative of the population.

Students often ask if it is "good enough" to take a sample, instead of surveying the entire population. If the survey is done well, the answer is yes.

### 1.3.1 Optional Collaborative Classroom Exercise

**Exercise 1.3.1**

As a class, determine whether or not the following samples are representative. If they are not, discuss the reasons.

1. To find the average GPA of all students in a university, use all honor students at the university as the sample.
2. To find out the most popular cereal among young people under the age of 10, stand outside a large supermarket for three hours and speak to every 20th child under age 10 who enters the supermarket.
3. To find the average annual income of all adults in the United States, sample U.S. congressmen. Create a cluster sample by considering each state as a stratum (group). By using simple random sampling, select states to be part of the cluster. Then survey every U.S. congressman in the cluster.
4. To determine the proportion of people taking public transportation to work, survey 20 people in New York City. Conduct the survey by sitting in Central Park on a bench and interviewing every person who sits next to you.
5. To determine the average cost of a two day stay in a hospital in Massachusetts, survey 100 hospitals across the state using simple random sampling.

## Thank You for previewing this eBook

You can read the full version of this eBook in different formats:

- HTML (Free /Available to everyone)
- PDF / TXT (Available to V.I.P. members. Free Standard members can access up to 5 PDF/TXT eBooks per month each month)
- Epub & Mobipocket (Exclusive to V.I.P. members)

To download this full book, simply select the format you desire below

