2018

Introduction to "Probability"

Lawrence Leemis

Probability

Lawrence M. Leemis
Department of Mathematics
The College of William & Mary
Williamsburg, Virginia

Library of Congress Cataloging-in-Publication Data

Leemis, Lawrence M.
Probability / Lawrence M. Leemis.
Includes bibliographic references and index.
ISBN 978-0-9829174-0-4
1. Probability
QA 273.L44 2011

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 978-0-9829174-0-4

Chapter 1

Introduction

I was standing under a large tree on a pleasant and still day several years ago. The stillness was interrupted by a large branch, about six inches in diameter, crashing to the ground about ten feet from where I was standing. I walked away knowing that I was just feet away from a major, random catastrophe. We live in a world of events that occur at random, and probability is a way of measuring and predicting this randomness. Consider the following examples:

- a reliability engineer contemplates the number of spare parts needed to support a fleet of mining trucks;
- an actuary analyzes data compiled in life tables to determine appropriate premiums for term life insurance policies;
- a medical doctor decides what action to take based on two tests—one with a positive result and one with a negative result;
- a college student scans the sky to help determine whether to take an umbrella to class;
- a toddler slowly reaches up toward the counter to grab a forbidden cookie, contemplating likely outcomes;
- a young mother sends her child off to kindergarten for the first time, wondering if the school has adequate evacuation plans if it is hit by a meteor.

Each of these people, in one fashion or another, at one level of sophistication or another, and at one level of rationality or another, is assessing probabilities. The notion of probability is very intuitive, and all of us make dozens of decisions every day based on probability assessments.

The purpose of this book is to hone your already-intuitive probability notions into a mathematical framework. In this way, when you are confronted with a complex problem involving probability, you will be able to confidently use this framework to craft a solution.

Although intuition typically works well when it comes to probability, it occasionally breaks down. The following two examples are probability questions whose solutions defy intuition.

Example 1.1 The "birthday problem" is usually stated along the following lines:

> If 40 people are gathered in a room, what is the probability that two or more people have the same birthday?

The year that the 40 people were born is not considered in this problem. People typically guess too low when asked to estimate this probability. One of the more common guesses that I encounter is $40/365$. The probability is about 0.89, so it is *very* likely that one or more birthdays will match in a room of 40 people. This problem will be solved using complementary probabilities and the multiplication rule in Chapter 2.

Example 1.2 The second problem is alternatively called the "car and goats" problem, the "Monty Hall" problem, or the "Let's Make a Deal" problem after the popular television game show.

> Suppose you're on a game show and you're given the choice of three doors. A car is placed behind one door; goats are placed behind the other two doors. The car and the goats were placed randomly behind the doors before the show. The rules of the game show are as follows: After you have chosen a door, the door remains closed for the time being. The game show host, Monty Hall, who knows what is behind the doors, now has to open one of the two remaining doors, and the door he opens must have a goat behind it. If both remaining doors have goats behind them, he chooses one randomly. After Monty Hall opens a door with a goat, he will ask you to decide whether you want to stay with your first choice or to switch to the last remaining door. Imagine that you chose Door 1 and the host opens Door 3, which has a goat. He then asks you: Do you want to switch to Door Number 2? Is it to your advantage to change your choice?

The intuitive answer to this question is that there is no advantage to switching doors. The car is behind one door and the second goat is behind the other door, so the two results are equally likely. This problem was stated in a slightly different form in a letter to Marilyn vos Savant's *Ask Marilyn* column in *Parade* magazine in 1990. The solution, which states that changing doors doubles the probability of getting the car from $1/3$ to $2/3$, created a barrage of about 10,000 letters, nearly 1000 of which came from PhD's, stating that the solution was wrong. We will use Bayes' Rule in Chapter 2 to show that her solution was indeed correct.

The next section gives a sampling of a few more probability questions that will appear subsequently in the book, along with some pointers toward one of the most common applications of probability: the analysis of data using statistical techniques.

1.1 Applications

Probability is a branch of mathematics that describes experiments whose outcome can't be predicted with certainty prior to performing the experiment. It was first studied by Blaise Pascal and Pierre de Fermat in the 17th century and applied to gambling games. Here is an example of such a gambling game.

Example 1.3 Toss a pair of dice 24 times. You win if you roll double sixes at least once. Find the probability of winning.

This problem can be easily solved using the tools provided in this book. There are certain assumptions that can be made, for example, the dice are fair and the rolls are independent. Once these assumptions are made, the axioms and results provided in Chapter 2 will yield a solution to this problem of

$$\begin{aligned} P(\text{winning}) &= 1 - P(\text{losing}) \\ &= 1 - P(\text{tossing no double sixes}) \\ &= 1 - \left(\frac{35}{36}\right)^{24} \\ &\cong 0.4914. \end{aligned}$$

The fact that the probability is close to 0.5 will draw gamblers to the game; the fact that the probability is slightly less than 0.5 assures that the house will draw revenue from the game in the long run.

The questions that can be addressed via probability techniques apply to many important applied fields beyond gambling games. Application areas include genetics, actuarial science, casualty insurance, meteorology, stock market analysis, economics, quality control, reliability, medicine, biostatistics, marketing, strength of materials, human factors, and sociology, just to name a few.

To illustrate the variety of problems that can be addressed using the tools of probability, two more simple examples are presented below.

Example 1.4 Find the probability of dealing a five-card poker hand containing a full house from a well-shuffled deck of playing cards.

Questions of this nature also require assumptions. For example, assume that we are playing with a full deck and that all of the possible shufflings are equally likely. Again, using the techniques from Chapter 2, the probability of dealing a full house (three cards having one denomination and two cards having another denomination) is

$$P(\text{full house}) = \frac{\binom{13}{1}\binom{12}{1}\binom{4}{3}\binom{4}{2}}{\binom{52}{5}} = \frac{13 \cdot 12 \cdot 4 \cdot 6}{2{,}598{,}960} = \frac{3744}{2{,}598{,}960} = \frac{6}{4165} \cong 0.00144.$$

The binomial coefficient $\binom{n}{r}$ will be defined in the next section. So with a probability of dealing a full house being just a bit over 1 in 1000, one can conclude that this will not occur often.

Leaving the realm of gambling games, we now switch from working with problems involving discrete outcomes to a problem involving a continuous outcome.

Example 1.5 Probability problems involving sums of random quantities often arise. In this example, let $X_1, X_2, \ldots, X_{10}$ be independent random variables that are uniformly distributed between 0 and 1. That is, each of the random variables assumes a continuous value between 0 and 1 with equal likelihood. Most calculators and computer programs have a *random number generator* capable of producing such numbers. Find the probability that their sum lies between 4 and 6, that is,

$$P\left(4 < \sum_{i=1}^{10} X_i < 6\right).$$

A well-known approximation known as the central limit theorem (introduced in Chapter 8) yields only one digit of accuracy for this particular problem. Another approximation technique known as Monte Carlo simulation requires custom computer programming, and the result is typically stated as an interval around the true value. This problem can also be solved exactly using some of the techniques and software provided in this text yielding

$$P\left(4 < \sum_{i=1}^{10} X_i < 6\right) = \frac{655177}{907200} \cong 0.7222.$$

Notwithstanding the obvious benefit of probability calculations to a gambler, a more significant application of probability theory lies in the field of *inferential statistics*, which has the goal of drawing inferences (conclusions) about the population from which a data set was drawn. The field of statistics was first studied as numerical data was collected on political units (for example, a census). This has eventually evolved into what is now known as "political science." The following examples illustrate the graphical and numerical analysis of a data set using standard statistical techniques. The letter n is used nearly universally to denote the *sample size*, which is the number of data values collected.

Example 1.6 (Ball bearing failure times) Consider the data set of $n = 23$ ball bearing failure times (measured in 10^6 revolutions):

17.88	28.92	33.00	41.52	42.12	45.60	48.48	51.84
51.96	54.12	55.56	67.80	68.64	68.64	68.88	84.12
93.12	98.64	105.12	105.84	127.92	128.04	173.40.	

There are several things that one can do to analyze such a data set. Computing certain numerical measures that summarize a data set is common, particularly with large data sets. The two most commonly used sample statistics are the *sample mean* and the *sample variance*. Using the notation $x_1, x_2, \ldots, x_n$ to denote the data values, the formulas for the sample mean $\bar{x}$ and sample variance s^2 are

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 72.22 \qquad \text{and} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 1405.4.$$

The sample mean is a measure of the central tendency of a data set; the sample variance is a measure of the dispersion of a data set. The sample mean has the same units as the data values. The positive square root of the sample variance, known as the *sample standard deviation*, $s = 37.49$ for this data set, also has the same units as the data values. These two quantities are *random* in the sense that a data set of $n = 23$ other ball bearings would produce different values for $\bar{x}$ and s^2. These quantities are to be distinguished from the *population mean* μ and the *population variance* σ^2, which would be obtained if we sampled the entire population of ball bearings.

In addition to summarizing the data set with numerical values such as $\bar{x}$ and s, there are also some graphical procedures that can be applied to a data set. A *histogram* is useful for determining the shape of a probability distribution; it is the statistical analog of a function to be introduced in Chapter 3 known as a probability density function. A histogram for the ball bearing lifetimes is shown in Figure 1.1. The horizontal axis is the failure time and the vertical axis is the number of ball bearing failure times that fall in each of the cells of width 20. The histogram reveals a clumping of the data around 50 million revolutions and also reveals that the largest of the ball bearing failure times, 173.40 million revolutions, lies significantly to the right of the others. The data set appears to come from a population with a single mode (peak) near 50 million revolutions. Issues associated with a histogram include choosing the number of cells and cell boundaries, which are arbitrary decisions made by the data analyst. Unfortunately, histograms are not good graphical instruments for comparing two or more distributions.

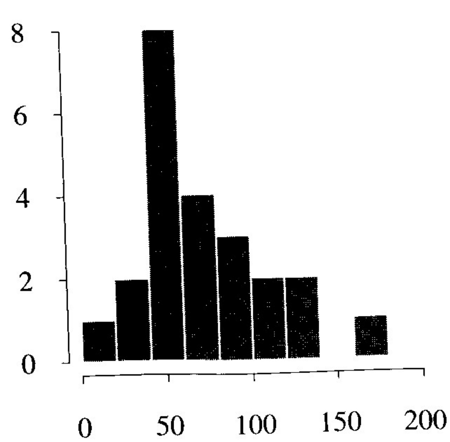

Figure 1.1: Histogram of ball bearing failure times.

Statistical packages are useful time-saving tools that can quickly perform numerical calculations and produce graphical displays associated with a data set. This text will use the statistical programming language R for such calculations and displays. R is available for free download on the web and is a powerful package that provides useful graphics, programming capability, and numerical calculations. It is becoming a standard that is used by statisticians. The code to compute the sample mean, compute the sample variance, and display the histogram for the ball bearing failure times is given below.

```
bearings = c(17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.48, 51.84,
             51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12,
             93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40)
mean(bearings)
var(bearings)
hist(bearings)
```

The ball bearing data illustrates what statisticians refer to as a *univariate* data set, since only a single variable has been collected on each ball bearing sampled. Many data sets involve collected *pairs* of data values, resulting in a *bivariate* data set, as illustrated by the following two examples.

Example 1.7 (Old Faithful Geyser eruptions) A data set of $n = 299$ data pairs, (x_i, y_i), $i = 1, 2, \ldots, 299$, has been collected on the waiting time x_i and the eruption duration y_i at the Old Faithful geyser in Yellowstone National Park in Wyoming. All observations are recorded in minutes. The data pairs are plotted in Figure 1.2. The data set exhibits some rather unique characteristics. First, there is an unusual clumping of the eruption duration around 4 minutes, which could be a natural phenomenon or could be due to rounding by those who collected the data. Second, there appears to be a tri-modal joint distribution of the (x_i, y_i) pairs. Look carefully at the scatterplot in Figure 1.2 to see if you can spot the three modes.

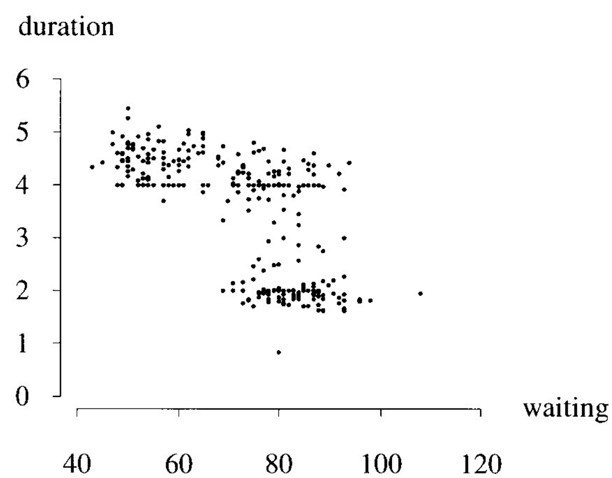

Figure 1.2: Geyser data.

Scatterplots like this are easily generated in R. The R command

```
plot(faithful$waiting, faithful$eruptions)
```

generates a scatterplot for a similar data set that is pre-loaded into R. The first argument to the `plot` function, the vector `faithful$waiting`, contains the waiting times for plotting on the horizontal axis; the second argument, the vector `faithful$eruptions`, contains the associated eruption times for plotting on the vertical axis.

Example 1.8 (Automobile warranty claims) As a second example of a bivariate data set, consider one particular make and model of an automobile that has a warranty that expires after 3 years or 36,000 miles, whichever occurs first. Warranty claim times (measured in both mileage and age in years) for a bivariate sample of size $n = 260$ are plotted in Figure 1.3.

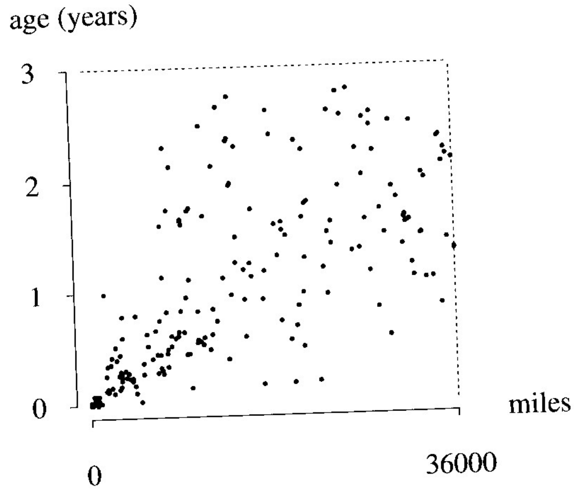

Figure 1.3: Scatterplot of warranty claim times.

The scatterplot of mileage and age reveals a significantly different pattern from the geyser data. First, we know the boundaries of the support, or the allowable values for the data pairs: they must fall in the rectangle with opposite corners at $(0, 0)$ and $(36000, 3)$. Second, there is a clustering of the data near the origin that corresponds to cars with problems that appear soon after being purchased. Third, there is what statisticians refer to as a *positive sample correlation* between the mileage and the age of an automobile that is taken to the dealer for a warranty claim. The two measures that reflect the aging of the automobile tend to increase together, resulting in (x_i, y_i) pairs that tend to be on the same sides of their means $\bar{x}$ and $\bar{y}$ more often than not. Fourth, if the histograms of the x_i and y_i values are plotted separately, as in Figure 1.4, one can

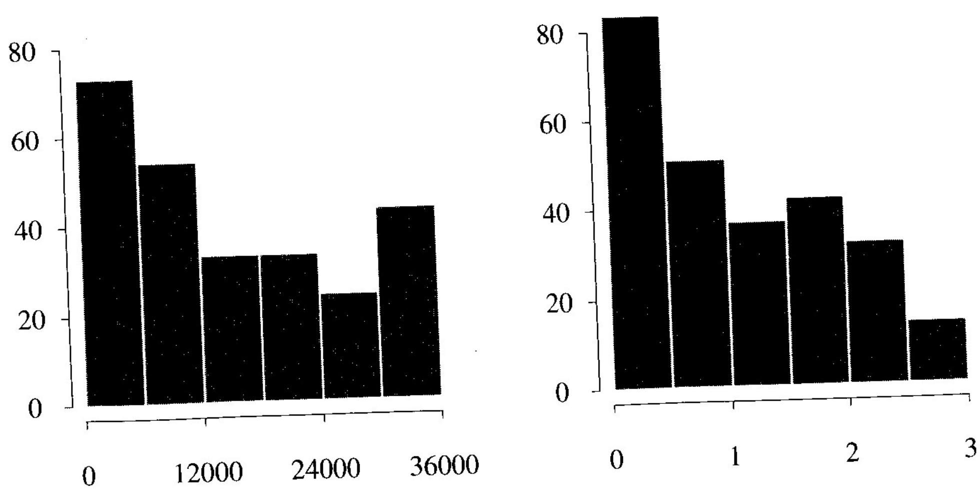

Figure 1.4: Histograms of warranty claim times (mileage on left; years on right).

clearly see the early failures on both histograms. The histogram on the left (mileage), however, also has a mode near 36,000 miles that corresponds to drivers being aware of a warranty expiring after 36,000 miles. The histogram on the right (age in years) does not have a second mode near 3 years because drivers are less aware of the 3-year anniversary of their purchase. In this case, the two histograms confirm our intuition about automobile owners and expiring warranties. The notion of the distribution of one of two variables (ignoring the other variable) will be developed in Chapter 6 as a *marginal* distribution.

The previous two examples have presented *bivariate* data sets for the Old Faithful geyser data and the car warranty claim data. When there are more than two data values collected, histograms and scatterplots can become problematic. A *boxplot* is a more compact way of looking at the shape of the distribution, as illustrated in the following example, where four data values are collected on each observational unit.

Example 1.9 (Cork deposits) The weights of cork deposits (in centigrams) of $n = 28$ trees is collected in the four directions: north, east, south, and west. The data is given in Table 1.1.

N	E	S	W
72	66	76	77
60	53	66	63
56	57	64	58
41	29	36	38
32	32	35	36
30	35	34	26
39	39	31	27
42	43	31	25
37	40	31	25
33	29	27	36
32	30	34	28
63	45	74	63
54	46	60	52
47	51	52	43
91	79	100	75
56	68	47	50
79	65	70	61
81	80	68	58
78	55	67	60
46	38	37	38
39	35	34	37
32	30	30	32
60	50	67	54
35	37	48	39
39	36	39	31
50	34	37	40
43	37	39	50
48	54	57	43

Table 1.1: Cork deposit weights.

A casual inspection of this data set reveals a positive correlation among the cork deposits in the four directions. Check out the tree with the spectacular 100 centigrams of cork deposits on its south side. All four of its cork deposit weights, that is (91, 79, 100, 75), dominate *all* of the deposits on the two trees above it on the list and the two trees below it on the list. In other words, when one of the four directions tends to have heavier cork deposits, the other directions are more likely to also have heavier cork deposits. The notion of positive correlation between the deposits captured in the four directions can be captured in a 4×4 *correlation matrix*, which for this data set is

$$\begin{bmatrix} 1.00 & 0.89 & 0.90 & 0.88 \\ 0.89 & 1.00 & 0.83 & 0.77 \\ 0.90 & 0.83 & 1.00 & 0.92 \\ 0.88 & 0.77 & 0.92 & 1.00 \end{bmatrix}.$$

The rows and columns of this matrix can be thought of as N, E, S, and W, rather than the usual 1, 2, 3, and 4. This is a *symmetric* matrix, so values on the opposite side of the diagonal are equal. The diagonal elements are all 1.0 and imply that there is perfect positive correlation between the data in each direction and itself. The off-diagonal elements are the correlations between the different directions. Consider the $(1, 2)$, or more exactly, the (N, E) element of the matrix, which has the value 0.89. This value indicates that there is a high positive correlation (all correlations must fall between -1 and 1) between the weights of the cork deposits in the north and east directions. In other words, the deposits in the north and east directions tend to be on the same sides of their means together. Another question that comes to mind is how the cork deposits in each direction vary individually. A histogram is not a good graphical device for comparing four distributions, but a *boxplot* can be used to compare several distributions simultaneously. The four boxplots displayed side-by-side in Figure 1.5 capture the essence of the four distributions for comparison. For each of the four directions, the middle half of the distribution is displayed vertically in a box. In other words, the top and the bottom of the box are the estimates of what is known as the *25th percentile* and the *75th percentile* of the probability distribution. Thus, the *quartiles* of a distribution are apparent from a boxplot. The horizontal line in the middle of the distribution shows the sample *median* or the middle value of the data set. This is an estimate of the *50th percentile* of the distribution. For a data set with an odd sample size n, this is just the middle value of the sorted data values. For an even sample size n, the two middle values are averaged. The fact that the median tends to fall consistently in the lower half of the box in all four directions indicates that the distribution of cork deposits is a *non-symmetric* distribution. The shortest box is associated with the weights of the deposits taken from the eastern side of the tree and the tallest box is associated with

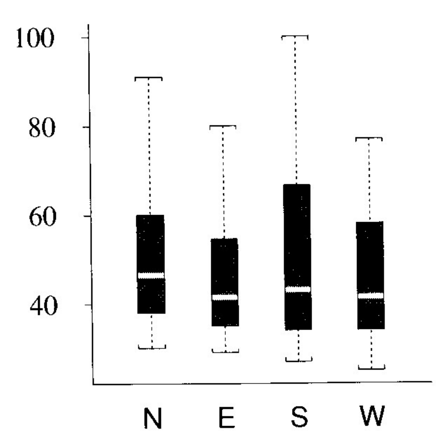

Figure 1.5: Boxplot of cork deposits in the four directions.

weights of the deposits taken from the southern side of the tree. This implies that the variability of the weights is greater on the southern side of the tree, even though the medians appear to be nearly identical. Finally, the *whiskers* in a boxplot extend to the smallest and largest data values, although this convention is not universal. The upward whisker in the southern direction extends to 100.

The R code to read the data from an external data set into an array named `cork`, calculate the correlation matrix, and generate the boxplot is given below.

```
cork = read.table("cork.dat")
cor(cork)
boxplot(cork, names = c("N", "E", "S", "W"))
```

Statisticians often encounter problems where a function needs to be fit to a data set. The simplest function to fit is a line. The next example illustrates a line being fit to data.

Example 1.10 (U.S. House of Representatives turnover) The average turnover percentages for the 12 decades following the end of the Civil War (these percentages are the averages of the five turnover percentages for elections held during the decade) are given in Table 1.2.

Decade	*Mean turnover percentage*
1870	50.3
1880	40.6
1890	40.8
1900	23.9
1910	28.5
1920	21.1
1930	26.3
1940	22.2
1950	14.6
1960	15.2
1970	16.6
1980	12.9

Table 1.2: U.S. House of Representatives turnover percentages.

It is clear that there is a downward trend in the data over time. It appears to be increasingly difficult to vote incumbent politicians out of office. It is impossible to fit a single line that will pass through all of these data values simultaneously, so we attempt to find the best line possible. One criterion for determining this best line is to find the *least squares* line that minimizes the sum of the squared vertical deviations between the line and the data points. Thus, the model for what is known as a *simple linear regression* is

$$Y = a + bX,$$

where X is the decade and Y is the turnover percentage in this particular setting. The slope and intercept of the regression line can be calculated by hand or by using any standard statistical package. Using the integers $1, 2, \ldots, 12$ to denote the $n = 12$ decades, the slope and intercept of the regression line are $b = -3.03$ and $a = 45.78$. The interpretation of slope is that the turnover percentage is decreasing by about 3% per decade on average. Two possible explanations for this decrease are the power of incumbency and the rise of the "career politician." A scatterplot of the data and the associated regression line is shown in Figure 1.6. The lengths of the vertical distances between the data points and the regression line are known as the *residuals*.

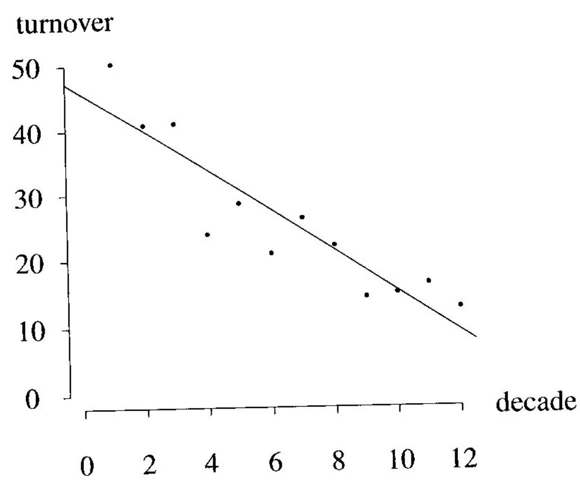

Figure 1.6: U.S. House of Representatives turnover percentages.

The R code to generate this figure is

```
decade   = 1:12
turnover = c(50.3, 40.6, 40.8, 23.9, 28.5, 21.1,
             26.3, 22.2, 14.6, 15.2, 16.6, 12.9)
plot(decade, turnover)
reg = glm(decade, turnover)
abline(reg$coef)
```

The first statement sets the R vector `decade` to the first 12 positive integers. The second statement assigns the 12 turnover percentages to the vector `turnover`. The 12 data pairs are then plotted using the `plot` statement. The `glm` function (for general linear model) performs the linear regression, and the least squares line is plotted using `abline` after extracting the regression coefficients (that is, a and b).

Before introducing probability in the next chapter, two important mathematical tools will be introduced that are often used in solving probability problems. These tools are *counting techniques* and *set theory*.

1.2 Counting

In many problems that arise in probability and statistics, it is useful to list (enumerate) or count the number of outcomes of an experiment. Counting is easy when there are only a handful of outcomes to count; when there are thousands or millions of outcomes, a more systematic approach is required.

Enumeration involves listing all of the possible outcomes to an experiment. Tree diagrams can be helpful, as will be seen in the following example.

Example 1.11 Imagine the unimaginable: The Chicago Cubs and the Chicago White Sox are playing in the World Series. The best-of-seven series is tied at two games apiece. What are the possible outcomes to the series?

The question is not asking for the *number* of possible outcomes, but rather a list of the possible outcomes. This will be accomplished using a *tree diagram* illustrated in Figure 1.7. The tree diagram reveals six possible endings to this particular World Series,

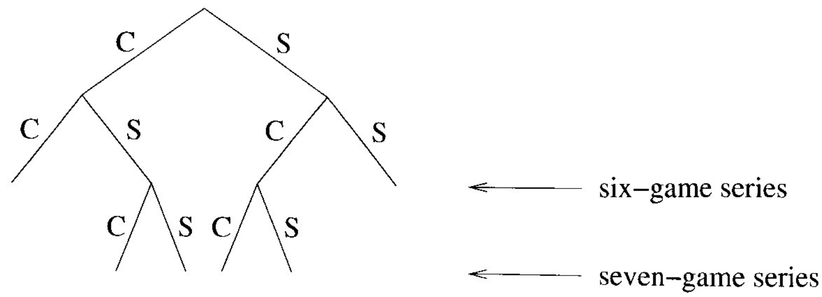

Figure 1.7: World Series outcomes.

two of which correspond to a six-game series and four of which correspond to a seven-game series, where C denotes a Cub victory and S denotes a Sox victory.

We now turn to *counting techniques* (or combinatorics or combinatorial methods), which are used when enumerating is cumbersome or infeasible. The field originated with a thirteenth-century Catalan missionary named Ramon Llull. We'll consider the following three techniques and some variations.

1. The multiplication rule.

2. Permutations (which is a special case of the multiplication rule).

3. Combinations.

Multiplication rule

We begin the discussion of counting techniques with the *multiplication rule*. The multiplication rule is also known as the fundamental theorem of counting, the basic principle of counting, the counting rule for compound events, and the rule for the multiplication of choices.

Theorem 1.1 (Multiplication rule) Assume that there are r decisions to be made. If there are n_1 ways to make decision 1, n_2 ways to make decision 2, ..., n_r ways to make decision r, then there are $n_1 n_2 \ldots n_r$ ways to make all decisions.

Proof To show why the multiplication rule holds, consider the case of $r = 2$ decisions. In this case, all of the potential outcomes can be displayed in the $n_1 \times n_2$ matrix given below. The rows represent the choices associated with decision 1; the columns represent the choices associated with decision 2.

	Choice 1	Choice 2	$\cdots$	Choice n_2
Choice 1				
Choice 2				
$\vdots$				
Choice n_1				

Thus, for $r = 2$, there are $n_1 \times n_2$ ways to make both decisions. To proceed from $r = 2$ to $r = 3$ decisions results in a rectangular solid consisting of $n_1 \times n_2 \times n_3$ cubes associated with the various ways of making the $r = 3$ decisions as illustrated in Figure 1.8.

The more general result for r decisions continues to follow this pattern, and is proved by induction. Assume that the theorem holds for r decisions. To show that this implies that

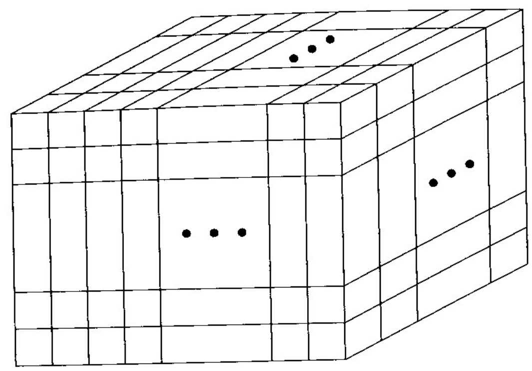

Figure 1.8: Multiplication rule justification for $r = 3$.

the theorem holds for $r+1$ decisions, each of the $n_1 n_2 \ldots n_r$ ways to make the r decisions is matched with one of the n_{r+1} ways to make decision $r+1$, resulting in $n_1 n_2 \ldots n_{r+1}$ total ways to make the $r+1$ decisions, thus proving the theorem by induction. □

Theorem 1.1 is applied to two simple counting problems involving coins and dogs. In each example, the number of choices for each decision is constant, that is, $n_1 = n_2 = \cdots = n_r$.

Example 1.12 How many different sequences of heads and tails are possible in 16 tosses of a fair coin?

Each of the 16 tosses can be considered a "decision" in terms of Theorem 1.1 with two possible outcomes: heads and tails. Since there are $r = 16$ tosses, there are

$$\underbrace{2 \cdot 2 \cdot 2 \cdot \ldots \cdot 2}_{16} = 2^{16} = 65,536$$

different sequences.

This problem is easily generalized to similar settings. Two such settings are

- the number of ways to answer a 16-question T/F test,
- the number of integers that can be stored on a 16-bit computer.

Example 1.13 How many ways can a mother give away 8 dogs to her 3 children?

In this example, Mom has $r = 8$ decisions on her hands, one for each dog. Furthermore, each decision can be made in three ways. Thus, there are

$$\underbrace{\overbrace{3}^{Fido} \cdot \overbrace{3}^{Inky} \cdot \overbrace{3}^{Suzy} \cdot \ldots \cdot \overbrace{3}^{Spot}}_{8} = 3^8 = 6,561$$

different ways for her to give away her 8 dogs to her 3 children.

As illustrated in the next two examples, occasions arise when the n_i values in the multiplication rule are not all identical.

Example 1.14 How many ways can a family of 5 line up for a photograph?

The photographer has $r = 5$ decisions to make. The first decision is whom to place on the left (which can be done five ways); the second decision is whom to place next to the person on the left (which can be done four ways), etc. Thus, there are

$$5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 120$$

different ways to line up the family for the photograph.

Example 1.15 How many ways can a family of 5 that consists of 3 men and 2 women line up for a photograph so that men and women alternate?

There are again $r = 5$ decisions for the photographer, but the decisions are restricted by gender, so we expect a lower count of outcomes than in the previous example. There is a choice of one of the 3 men to place on the left, a choice of 2 women to place next to him, etc. Continuing in this fashion yields a total of

$$\overset{M}{\overbrace{3}} \cdot \overset{W}{\overbrace{2}} \cdot \overset{M}{\overbrace{2}} \cdot \overset{W}{\overbrace{1}} \cdot \overset{M}{\overbrace{1}} = 12$$

different ways to line up the family members in the restricted fashion.

As a foreshadowing of the introduction of probability in Chapter 2, the previous two examples will be used to determine the probability that men and women alternate when the family of five sits for the photograph in a random order. Using the previous two examples, this probability is $P(\text{men and women alternate}) = 12/120 = 1/10$ assuming that all of the 120 possible orderings are equally likely. More details concerning the calculation of probabilities are given in Chapter 2.

Example 1.16 How many ways are there to arrange the letters in "dynamite"?

There are $r = 8$ decisions to be made, so there are

$$8 \cdot 7 \cdot 6 \cdot \ \ldots \ \cdot 2 \cdot 1 = 8! = 40,320$$

different ways to arrange the letters.

Example 1.17 How many ways can 3 men, 4 women, and 2 children arrange themselves in a row of nine chairs if

(a) the children insist on sitting together?
(b) the children insist on sitting on the leftmost and rightmost chairs?
(c) the men, women, and children must sit next to one another?

If there were no restrictions, there would be $9! = 362,880$ different ways to line up the $r = 9$ people. Since each part of this question places a restriction on the ordering, the number arrangements must be less than 362,880.

(a) When the children insist on sitting together, classify the people as 2 children and 7 adults. There are 2! ways to arrange the children. Since the leftmost child can occupy any one of 8 positions and there are 7! ways to arrange the adults, there are a total of

$$2! \cdot 8 \cdot 7! = 2 \cdot 8 \cdot 5040 = 80,640$$

different arrangements with the children sitting together.

(b) When the children insist on sitting on the leftmost and rightmost chairs, there are 2! ways to arrange the children at the extremes and 7! ways to arrange the adults in between them, so there are a total of

$$2! \cdot 7! = 2 \cdot 5040 = 10,080$$

different arrangements with the children sitting at the extremes.

(c) If the men sit in the three leftmost chairs, the women sit in the next four chairs, and the children sit in the two rightmost chairs, there are $3! \cdot 4! \cdot 2!$ arrangements. Since there are 3! arrangements of the men, women, and children as groups, there are

$$3! \cdot 3! \cdot 4! \cdot 2! = 6 \cdot 6 \cdot 24 \cdot 2 = 1728$$

different arrangements with the men, women, and children sitting next to one another.

Example 1.18 Cindy shuffles a deck of playing cards. Is it likely that she is the first person in history to achieve this particular ordering of the cards?

This is another of those problems that defies intuition. Of all of the people in history, almost surely *someone* must have attained the same shuffle as Cindy. First of all, by the multiplication rule, there are

$$52! = 80658175170943878571660636856403766975289505440883277824000000000000$$

different shufflings. Yikes! Perhaps Cindy's shuffle is likely unique after all. To address the likelihood of her shuffle being unique, some back-of-the-envelope calculations are required. The world population is about seven billion people. Approximately half of the people that have ever lived are currently alive, so assume that 14 billion people have lived through the ages. Now assume that everyone lives 100 years on average (dubious), and shuffles a deck of cards ten times a day on average (even more dubious), then there have been a total of a mere

$$14000000000 \cdot 100 \cdot 365 \cdot 10 = 5110000000000000$$

total shuffles. Hence Cindy's shuffle is almost certainly unique. Every shuffle of a deck of cards is almost always making playing-card history.

Although simple to state and use, the multiplication rule is a surprisingly versatile tool for addressing counting (combinatorics) problems. There is a special case of the multiplication rule that arises so often that it gets special treatment here. The object of interest is known as a *permutation.*

Permutations

The notion of whether a sample is taken with or without replacement is a critical notion in combinatorics and probability. When a sample of size r, for example, is selected at random and *with replacement* from a set of n distinct objects, there are n^r different ordered samples that can be taken. On the other hand, when the items are selected without replacement, the ordered items that are selected are a *permutation.*

Definition 1.1 A *permutation* is an arrangement of r objects selected from a set of n objects without replacement in a definite order.

One key question to be addressed in a counting problem is whether the ordering of the objects is relevant. If the ordering is relevant, then using permutations might be appropriate.

Example 1.19 List the permutations from the set $\{a, b, c\}$ selected 2 at a time.

Applying Definition 1.1 with $n = 3$ and $r = 2$ yields the 6 ordered pairs:

$$\begin{array}{cc} (a,b) & (b,a) \\ (a,c) & (c,a) \\ (b,c) & (c,b). \end{array}$$

The second column of permutations is the same as the first column in reverse order.

Theorem 1.2 The number of permutations of n distinct objects selected r at a time without replacement is

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+1) = \frac{n!}{(n-r)!}$$

for $r = 0, 1, 2, \ldots, n$ and n is a positive integer, and $0! = 1$.

Proof Consider the following two cases based on the value of r. Case I: When $r = 0$, there is only one way to choose the sample (don't select any items), and

$$\frac{n!}{(n-0)!} = 1.$$

Case II: When $r = 1, 2, \ldots, n$, the multiplication rule with r decisions yields

$$\underbrace{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+1)}_{r \text{ factors}} = \frac{n!}{(n-r)!}$$

which establishes the result. □

The proof shows that finding the number of permutations is just a special case of the multiplication rule.

Example 1.20 How many ways are there to pick a president, vice-president, and treasurer from 7 people?

The "objects" from the previous definition are the $n = 7$ candidates, and there are $r = 3$ of them being selected. The fact that sampling is performed without replacement is implicit in the problem statement in that one person could not occupy all three positions. Furthermore, the fact that order is relevant in their selection is implicit in the problem statement in that they are given distinct titles (president, vice-president, and treasurer). Thus, there are

$$\frac{7!}{(7-3)!} = \frac{7!}{4!} = 7 \cdot 6 \cdot 5 = 210$$

permutations associated with filling the three positions. The $7 \cdot 6 \cdot 5$ part of the equation serves as a reminder that this question could have been addressed directly by the multiplication rule.

Example 1.21 A ship has 3 stands and 12 flags to send signals. How many 3-flag signals can be sent?

Again, implicit in the statement of the problem is the fact that the 12 flags are distinct. Furthermore, to send a 3-flag signal requires sampling without replacement from the 12 flags. Proceeding with $n = 12$ and $r = 3$, there are

$$\frac{12!}{(12-3)!} = \frac{12!}{9!} = 12 \cdot 11 \cdot 10 = 1320$$

different signals.

Example 1.22 In the previous example, what if one or two flags also constitute a signal?

In this setting, the number of signals should simply be summed. Proceeding with $n = 12$ and $r = 1, 2, 3$, there are

$$\frac{12!}{11!} + \frac{12!}{10!} + \frac{12!}{9!} = 12 + 132 + 1320 = 1464$$

different signals that can be sent.

Example 1.23 How many ways are there to drive, in sequence, to four cities from a starting location?

Assuming that the starting location differs from the four cities, there are $n = 4$ objects (the cities) and all $r = 4$ of them must be selected. Thus, there are

$$\frac{4!}{(4-4)!} = \frac{24}{1} = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

different ways to drive to the four cities.

There are two minor tweaks that can be performed on permutations that are often useful in solving combinatorics problems: circular permutations and nondistinct permutations.

1. **Circular permutations:** Consider the placement of n objects in a circle.

Theorem 1.3 The number of permutations of n distinct objects arranged in a circle is $(n-1)!$

Proof Fix one object's position and use the multiplication rule with $n-1$ decisions to conclude that there are

$$(n-1) \cdot (n-2) \cdot \ \ldots \ \cdot 1 = (n-1)!$$

different ways to arrange the objects in a circle. □

Example 1.24 How many ways are there to seat 6 people around a round table for dinner?

There are $n = 6$ objects (the diners!) to place around the table. There are

$$5! = 120$$

ways to order them around the table.

Care should be taken when interpreting the solution to the dinner table question. The only thing that matters in the particular seating is who is on your left and who is on your right when considering a circular permutation. For example,

- What if all diners shift one chair clockwise? This would *not* be a new circular permutation.
- What if one seat is a blue throne and it matters who is sitting in the blue throne? In this case a clockwise shift does result in a new circular permutation, so there are $6! = 720$ circular permutations in the blue throne setting.
- What if the order of seating is reversed (clockwise vs. counterclockwise)? This is indeed a new ordering.

Example 1.25 How many circuits can a traveling salesman make of n cities? A reverse route is not considered a unique path.

Assuming that the traveling salesman is beginning at one of the cities, there are

$$\frac{(n-1)!}{2} = \frac{n!}{2n}$$

different routes. Dividing by two prevents double counting reverse circuits. The spectacular factorial growth in this quantity is shown in Table 1.3

n	3	4	8	10	15	50
$n!/(2n)$	1	3	2520	181,440	43,589,145,600	$\sim 10^{62}$

Table 1.3: The number of traveling salesman routes.

Even if you are going around town to run just 8 errands, you have plenty of options. If you need to run 15 errands, for example, there are over 43 billion routes. Finding the shortest of all of these routes is known as the *traveling salesman problem*, which is a classic optimization problem in a field known as *operations research*. The problem is faced daily by package delivery companies. It is particularly difficult to solve because of the factorial growth in the number of routes.

2. **Nondistinct permutations:** In a typical permutation counting problem, all of the n objects are distinct. We now consider the possibility of just r distinct types of objects.

Theorem 1.4 The number of nondistinct permutations of n objects of which n_1 are of the first type, n_2 are of the second type, ..., n_r are of the rth type, is

$$\frac{n!}{n_1!n_2!\ldots n_r!},$$

where $n_1 + n_2 + \cdots + n_r = n$.

Proof Let A be the number of nondistinct permutations. We want to show that

$$A = \frac{n!}{n_1!n_2!\ldots n_r!}.$$

If all n objects were distinct, the number of permutations is $n!$ or, by the multiplication rule, the number of permutations is $n_1!n_2!\ldots n_r!A$. Equating and solving for A yields the desired result. □

Example 1.26 Consider the case of

$$n = 9; r = 3, n_1 = 4, n_2 = 2, n_3 = 3.$$

The number of ways to order the objects

$$a_1a_2a_3a_4b_1b_2c_1c_2c_3$$

when the a, b, and c objects can't be distinguished from one another by their subscripts is

$$\frac{9!}{4!3!2!} = \frac{362,880}{24 \cdot 6 \cdot 2} = 1260.$$

Example 1.27 How many ways are there to arrange the letters in the word "door"?

There are $n = 4$ objects (the letters) and $r = 3$ of them are distinct. Thus, there are

$$\frac{4!}{1!2!1!} = \frac{24}{2} = 12$$

different arrangements. In the denominator, 2! accounts for swapping the two indistinguishable o letters.

Example 1.28 How many ways are there to arrange the letters in "puppet"?

Proceeding with the $n = 6$ objects with $r = 4$ distinct letters (p, u, e, t), there are

$$\frac{6!}{3!1!1!1!} = \frac{720}{6} = 120$$

different arrangements.

Example 1.29 How many ways are there to arrange the letters in "wholesome"?

In the word wholesome, there are $n = 9$ letters, $r = 7$ of which are distinct, leading to

$$\frac{9!}{2!2!} = \frac{362,880}{4} = 90,720$$

different arrangements.

Example 1.30 How many ways are there to line up identical twins and identical triplets for a photo if identical-looking people are nondistinct?
There are $n = 5$ people to line up with just $r = 2$ distinct looks. Since there are $n_1 = 2$ twins and $n_2 = 3$ triplets, there are

$$\frac{5!}{2!3!} = \frac{120}{12} = 10$$

different ways to line up the five.

This ends the discussion of permutations and two spinoffs (circular permutations and nondistinct permutations). We now switch to a discussion of *combinations*, which are closely related to permutations.

Combinations

In some situations, we are interested in the number of ways of selecting r objects without considering the *order* that they are selected (for example, a poker hand). These are called *combinations* and are a special case of nondistinct permutations when there are two types of objects.

Definition 1.2 A set of r objects taken from a set of n objects without replacement is a *combination*.

Example 1.31 List the combinations of 2 elements taken from $\{a, b, c, d\}$.

Since the order is not relevant, there will be fewer combinations than permutations:

$$\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}.$$

The number of combinations of r items selected from n items arises so often in combinatorics and probability that it gets its own symbol, as will be seen in the following theorem. The expression

$$\binom{n}{r}$$

is called "n choose r."

Theorem 1.5 The number of combinations of r objects taken without replacement from n distinct objects is

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}.$$

Proof This result can be proved in two different fashions. First,

$$\binom{n}{r} = \frac{n!}{(n-r)!r!},$$

which is just the formula for the number of permutations divided by $r!$ to account for the number of ways to order the r objects selected.

A second way to prove the theorem is to think of the two groups of objects (those selected and those not selected) as two different types of indistinguishable items. Using the result concerning the number of nondistinct permutations,

$$\binom{n}{r} = \frac{n!}{\underbrace{(n-r)!}_{n_1!}\underbrace{r!}_{n_2!}}. \qquad \square$$

We now illustrate the application of Theorem 1.5.

Example 1.32 How many ways are there to pick a *committee* of three people from seven "volunteers"?

This question differs fundamentally from the earlier example involving the selection of a president, vice-president, and treasurer from a group of seven people. The fact that titles were being assigned to the three people selected meant that order was important. In this case, the committee selection process makes no implication with respect to the order that the members are selected, so using combinations is appropriate. Since $r = 3$ people are being selected from the larger group of $n = 7$ people, there are

$$\binom{7}{3} = \frac{7!}{4!3!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35$$

different committees that can be formed.

Example 1.33 How many ways can a five-card hand be dealt from a standard deck of playing cards?

There are $r = 5$ cards to be selected from the $n = 52$ cards in the deck. Since the order in which the cards are dealt is not relevant, combinations should be used to solve the problem. Using Theorem 1.5, there are

$$\binom{52}{5} = \frac{52!}{47!5!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2{,}598{,}960$$

different five-card hands that can be dealt.

Example 1.34 How many ways are there to answer a 10-question true/false test with exactly two true answers?

One way of thinking about this problem is to consider all of the $2^{10} = 1024$ different sequences of 10 answers: T (for true) and F (for false). Consider dealing out the two *positions* for the T responses. Table 1.4 lists the responses associated with exactly two T responses, along with the positions of the T responses. The *order* that the two positions are selected is not relevant (for example, choosing positions 1 and 2 for the T responses is identical to choosing positions 2 and 1 for the T responses).

Proceeding using the selection of $r = 2$ positions for the T responses of the $n = 10$ positions means that there are

$$\binom{10}{2} = \frac{10!}{8!2!} = \frac{10 \cdot 9}{2 \cdot 1} = 45$$

different ways of answering the true/false exam with exactly two true responses.

Answers	Positions of the T responses
T T F F F F F F F F	1, 2
T F T F F F F F F F	1, 3
T F F T F F F F F F	1, 4
⋮	⋮
F F F F F F F F T T	9, 10

Table 1.4: Ten question true/false test responses.

Example 1.35 A ship has 3 stands and 12 flags to send signals. How many signals can be sent if one, two, or three flags constitute a signal and the stand(s) selected are relevant?

The problem implies that a red flag in stand 1 and a blue flag in stand 2 constitute a different signal than a red flag in stand 2 and a blue flag in stand 3. The solution to this problem requires the use of both combinations and permutations. Combinations are used to pick the stands, then permutations are used to place the flags in order in those stands. There are

$$\binom{3}{1}\frac{12!}{11!}+\binom{3}{2}\frac{12!}{10!}+\binom{3}{3}\frac{12!}{9!}=36+396+1320=1752$$

different signals that can be sent. The three quantities being added in the solution correspond to one-flag, two-flag, and three-flag signals.

Example 1.36 How many ways can 14 people split into two teams of seven for a game of ultimate frisbee?

Since the question concerns *teams*, the order of selection is not relevant, which implies that combinations should be used here. There are

$$\frac{\binom{14}{7}}{2}=1716$$

different ways to split the 14 people into two teams of seven. Division by two avoids double counting identical teams.

These examples illustrate the wide variety of problems that can be addressed using combinations. Combinations also have a number of interesting mathematical properties which will be given in an outline format below.

1. The well-known *binomial theorem* can be used to expand quantities such as

$$(x+y)^4 = 1x^4+4x^3y+6x^2y^2+4xy^3+1y^4.$$

The coefficients in the expansion (namely 1, 4, 6, 4, 1 in this case) happen to correspond to the number of combinations. For this reason $\binom{n}{r}$ is often referred to as a "binomial coefficient." The general statement of the binomial theorem is

$$(x+y)^n=\sum_{r=0}^{n}\binom{n}{r}x^{n-r}y^r.$$

2. There are several miscellaneous results that are associated with the binomial coefficients. Here are a few such results, stated without proof.

(a) Symmetry: $\binom{n}{r} = \binom{n}{n-r}$, for $r = 0, 1, \ldots, n$; and n is a positive integer

(b) $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$

(c) $\sum_{r=0}^{k} \binom{m}{r}\binom{n}{k-r} = \binom{m+n}{k}$

3. The binomial coefficient $\binom{n}{r}$ is defined to be 0 when $r < 0$ or $r > n$.

4. Pascal's triangle, which is given by

$$
\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1 \\
 & & & & \vdots & & & &
\end{array}
$$

consists entirely of binomial coefficients (notice the 1, 4, 6, 4, 1) in the fifth row corresponding to the coefficients in the expansion of $(x+y)^4$. Some other interesting tidbits about Pascal's triangle are listed below.

- The row number is determined by n and the position in the row is determined by r.
- Each row determines the subsequent row. Each entry that is not on the boundary of the triangle is the sum of the two closest entries in the previous row. This is equivalent to the result in 2(b) above.
- The row sums are powers of 2, that is, $\sum_{r=0}^{n} \binom{n}{r} = 2^n$.
- The sums of the first n diagonal elements are n, $n(n+1)/2$, $n(n+1)(2n+1)/6$, ... due to constant, linear, quadratic, ... growth of the diagonal elements.
- The Fibonacci sequence 1, 1, 2, 3, 5, 8, 13, ... can be found in the triangle. See if you can find a way to determine these values.
- Replacing odd numbers in the triangle by 1 and even numbers by 0 yields the "Sierpinski gasket."
- Try replacing each element modulo 3.
- If the digits of the first five rows are concatenated, they yield the powers of 11.

5. The binomial theorem can be extended to the "multinomial theorem" to handle the expansion of expressions like $(x+y+z)^8$. The coefficient for $x_1^{m_1}x_2^{m_2}\ldots x_k^{m_k}$ when expanding $(x_1+x_2+\cdots+x_k)^n$ is

$$\binom{n}{m_1, m_2, \ldots, m_k} = \frac{n!}{m_1!m_2!\ldots m_k!}.$$

6. Combinations are a special case of partitioning. Consider the following two examples.

Example 1.37 How many ways are there to deal a five-card poker hand?

This problem was encountered earlier and solved using combinations. The problem can also be considered as a partitioning problem. Dealing five cards from a 52-card deck is equivalent to partitioning the deck into five cards (those selected for the hand) and 47 other cards (those not selected for the hand), as shown below. The bar is used to denote the partitioning position.

$$\underbrace{1\ 2\ 3\ 4\ 5}_{r \text{ here}} \mid \underbrace{6\ 7\ \ldots\ 51\ 52}_{n-r \text{ here}}$$

The next example moves from partitioning a set of objects into two groups to partitioning a set of objects into three groups.

Example 1.38 You have a one, five, twenty, and hundred dollar bill to invest in three stocks: AT&T, Boeing, and Coke. How many ways are there to invest 2 bills in AT&T, 1 bill in Boeing, and 1 bill in Coke?

Let

- O denote the one dollar bill,
- F denote the five dollar bill,
- T denote the ten dollar bill,
- H denote the hundred dollar bill,

and let

- a bill to the left of the bars corresponds to an investment in AT&T,
- a bill between the bars corresponds to an investment in Boeing,
- a bill to the right of the bars corresponds to an investment in Coke.

The 12 possible investment strategies are enumerated below. The bars again represent the partition.

$OF\|T\|H$	$OF\|H\|T$	$OT\|F\|H$	$OT\|H\|F$
$OH\|F\|T$	$OH\|T\|F$	$FT\|O\|H$	$FT\|H\|O$
$FH\|O\|T$	$FH\|T\|O$	$TH\|O\|F$	$TH\|F\|O$

These two examples lead to a more general result which is stated without proof.

Theorem 1.6 The number of ways of partitioning a set of n distinct objects into k subsets with n_1 in the first subset, n_2 in the second subset, ..., n_k in the kth subset, is

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \ldots n_k!},$$

where $n_1 + n_2 + \cdots + n_k = n$.

The previous example concerning the number of investment strategies is solved using Theorem 1.6 with $n = 4$, $k = 3$, $n_1 = 2$, $n_2 = 1$, and $n_3 = 1$ yielding

$$\frac{4!}{2!1!1!} = 12$$

different investment strategies.

Example 1.39 The Glen family consists of 9 people. How many arrangements are there for them to watch the nightly news seated on four sofas: one that seats three and the others seat two?

Implicit in the problem statement is that the *position* (for example, left, right, middle on the big sofa) occupied by one of the Glens on a particular sofa is not relevant in terms of TV viewing arrangements. Applying Theorem 1.6 with $n = 9$ family members being partitioned onto $k = 4$ sofas with $n_1 = 3$, $n_2 = 2$, $n_3 = 2$, and $n_4 = 2$ family members occupying each sofa, there are

$$\binom{9}{3, 2, 2, 2} = \frac{9!}{3!2!2!2!} = 7560$$

different TV viewing arrangements. Assuming that the Glens have no better form of amusement, they could go over 20 years swapping different TV viewing arrangements each night.

The alert reader will have noticed that nondistinct permutations and partitioning problems both use

$$\frac{n!}{n_1!n_2!\ldots n_k!}.$$

The following two examples illustrate how these two approaches are actually solving fundamentally identical problems.

Example 1.40 (Nondistinct permutations) How many ways are there to arrange the letters in the word "bib"?

The b's in "bib" are considered nondistinct so that swapping the b's does not correspond to a new ordering. Enumerating the outcomes yields

bbi
bib
ibb

and the formula from Theorem 1.4 for nondistinct permutations yields

$$\frac{3!}{2!1!} = 3$$

different orderings. The "indistinguishable objects" here are the b's.

Example 1.41 (Partitioning) Preston, Jill and Gretchen are sisters. How many ways are there to sleep the three girls in a double and single bed?

Let P, J, and G denote the three girls. Also, place the two girls in the double bed on the left of the bar and the girl in the single bed on the right of the bar. Enumerating the outcomes yields

$PJ \mid G$
$PG \mid J$
$GJ \mid P$

Using Theorem 1.6 for partitioning problems, there are

$$\frac{3!}{2!1!} = 3$$

different orderings. The "indistinguishable objects" here are the two girls in the double bed. The problem of sleeping the sisters in the beds is fundamentally the same as the ordering of the letters in the word "bib."

We close this section with one final unifying example that stresses the importance of the following two questions associated with a counting problem. (a) Is the sampling performed *with replacement* or *without replacement*? (b) Is the sample considered *ordered* or *unordered*?

Example 1.42 How many ways are there to select 4 billiard balls from a bag containing the 15 balls numbered $1, 2, \ldots, 15$?

The question as stated is (deliberately) vague. It has not been specified whether

- the billiard balls are replaced (that is, returned to the bag) after being sampled, and
- the order that the balls are being drawn from the bag is important.

So there are really $2 \times 2 = 4$ different questions being asked here. The answers to these questions are given in the 2×2 matrix below.

	Without replacement	With replacement
Ordered sample	$15 \cdot 14 \cdot 13 \cdot 12$	$15 \cdot 15 \cdot 15 \cdot 15$
Unordered sample	$\binom{15}{4}$	$\binom{18}{4}$

These simplify to

	Without replacement	With replacement
Ordered sample	32,760	50,625
Unordered sample	1365	3060

There are several observations that can be made on the numbers in this 2×2 matrix. First of all, the entries in column 2 are always greater than the corresponding entries in column 1. This is because sampling with replacement allows for more possible draws due to the fact that the size of the population from which a draw is made remains constant rather than diminishing. Secondly, the entries in row 1 are always greater than the corresponding entries in row 2. This is because the count of ordered draws (permutations) will always exceed the corresponding number of unordered draws (combinations).

A further explanation of the lower-right entry of the matrix might be needed. Consider 15 bins and 4 balls, where ○ denotes a billiard ball. One draw of 4 balls is depicted below.

| ○ ○ | | ○ | | ··· | | ○

1 2 3 4 5 ··· 14 15

This arrangement of bins and markers corresponds to the unordered draw 2, 2, 4, 15 taken with replacement from the bag. We need to count the number of arrangements of 14 dividers plus 4 balls, or a total of 18 objects. Since the ○'s are indistinguishable, there are

$$\binom{18}{4}$$

different orderings (the outer walls are ignored).

The previous example has highlighted two important issues that arise in combinatorial problems: order and replacement. These concerns lead to a generic class of problems known as "urn models" in which objects are drawn sequentially from an urn. This has been an unusually long section, so it ends with an outline of the topics considered, and their associated formulas.

1. Multiplication rule: $n_1 n_2 \ldots n_r$
2. Permutations: $\dfrac{n!}{(n-r)!}$

(a) Circular permutations: $(n-1)!$

(b) Nondistinct permutations: $\dfrac{n!}{n_1!n_2!\dots n_r!}$

3. Combinations (binomial coefficients): $\displaystyle\binom{n}{r} = \frac{n!}{(n-r)!r!}$

1.3 Sets

Sets are often used in solving probability problems. The German mathematician Georg Cantor is generally credited with creating set theory. We provide a brief review of set theory here, and begin with basic definitions.

Definition 1.3 A *set* is a collection of objects (elements).

Upper-case letters are typically used to denote sets, for example, A, B. The notion of a set is a very general one, as will be seen in the example below.

Example 1.43 Consider the following four sets.

$$
\begin{aligned}
A &= \{1, 2, \dots, 100\} \\
B &= \{x \,|\, x \text{ is a positive integer less than } 101\} \\
C &= \{\text{Bulls, Trailblazers}\} \\
D &= \{(x, y) \,|\, 0 < x < 1, 0 < y < 2\}
\end{aligned}
$$

The elements of sets A and B are integers; the elements of set C are the names of basketball teams; the elements of set D are points in the interior of a rectangle in the Cartesian coordinate system. Sets B and D are defined by what is known as the *set-builder* notation, and the bar is read as "such that." Thus, the definition of B is read as "the set of all values x such that x is a positive integer less than 101." The sets A and B have identical elements, and their equality is written as $A = B$.

Definition 1.4 If an object belongs to a set, it is said to be an *element* of the set. The notation $\in$ is used to denote membership in a set.

Example 1.44 Using the sets defined in Example 1.43,

$$17 \in A \qquad 99 \in B \qquad \left(\frac{2}{3}, 1\right) \in D \qquad \text{Cubs} \notin C.$$

The notation $\in$ is read as "is a member of." The notation $\notin$ is read as "is not a member of."

Definition 1.5 If every element of the set A_1 is also an element of the set A_2, then A_1 is a *subset* of A_2. The notation $\subset$ is used to denote the subset relationship.

The subset symbol $\subset$ from Definition 1.5 allows for the two sets A_1 and A_2 to be equal. For any set A, for example, $A \subset A$. If A_1 is a subset of A_2, but A_1 is not allowed to equal A_2, then the relationship between A_1 and A_2 is known as a *proper subset*.

Example 1.45 The natural numbers **N**, also known as the positive integers, are a subset of the integers **Z**, which are a subset of the rational numbers **Q**, which are a subset of the real numbers **R**, which are a subset of the complex numbers **C**. These relationships are compactly stated as

$$\mathbf{N} \subset \mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R} \subset \mathbf{C}.$$

Venn diagrams are a useful tool in set theory and in probability for sorting out the relationships between various sets. An example of a Venn diagram containing the sets A and B is shown in Figure 1.9.

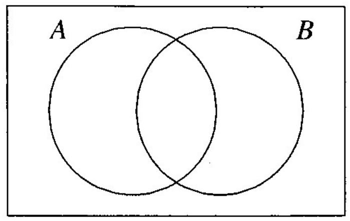

Figure 1.9: Venn diagram.

The external rectangle that is drawn outside of the two sets A and B is often called the *universal* set, and it contains all possible elements under consideration. If it is assumed that $A \subset B$, then the Venn diagram can be modified as in Figure 1.10.

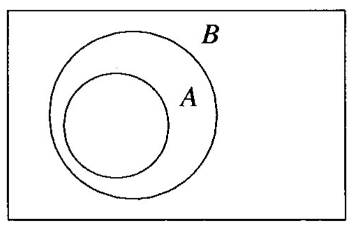

Figure 1.10: Venn diagram for $A \subset B$.

Example 1.46 When there are several subsets involved in a particular application, we often use subscripts, rather than individual letters to denote the sets. Thus, the relationship between

$$A_1 = \{x \mid 0 < x < 1\}$$

and

$$A_2 = \{x \mid 0 < x < 5\}$$

can be described by

$$A_1 \subset A_2.$$

Definition 1.6 A set containing no elements is called the *null set*. The notation $\emptyset$ is used to denote the null set.

Example 1.47 List the subsets of $\{a, b, c\}$.

By the multiplication rule, there are $2^3 = 8$ such subsets because there are three decisions to be made (whether to include or not include each element in the subset), and each decision can be made in two ways. The subsets are

$$\emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}.$$

To generalize, by the multiplication rule, there are always 2^n subsets of any set containing n elements.

This completes the statement of some basic definitions in set theory. We now define the operations that can be applied to a set. We consider just three: union, intersection, and complement.

Definition 1.7 (Union) $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$

The meaning of "or" in Definition 1.7 is not exclusive: the elements in $A \cup B$ are in A alone, B alone, or in both A and B simultaneously. A Venn diagram with the union of A and B shaded is given in Figure 1.11

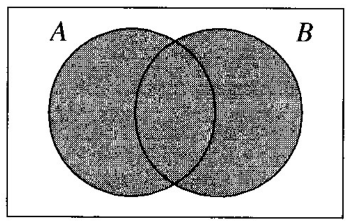

Figure 1.11: Venn diagram for $A \cup B$.

The notion of the union of two sets generalizes naturally to more than two sets.

Definition 1.8 $A_1 \cup A_2 \cup \ldots = \{x \mid x \in A_1 \text{ or } x \in A_2 \text{ or } \ldots\}$, which applies to a finite or infinite number of sets.

Example 1.48 For the sets

$$\begin{aligned} A_1 &= \{x \mid 0 < x < 1\} \\ A_2 &= \{x \mid 0 < x < 5\} \end{aligned}$$

find the union of A_1 and A_2.

Since $A_1 \subset A_2$, the union of A_1 and A_2 is just A_2, so

$$A_1 \cup A_2 = A_2.$$

Example 1.49 Let

$$A_k = \{k, k+1, k+2, \ldots, k^2\}$$

for $k = 1, 2, \ldots$. Find the union of A_3, A_4, and A_5.

Since the three-way union is all of the elements in A_3, A_4, or A_5,

$$A_3 \cup A_4 \cup A_5 = \{3, 4, \ldots, 25\}.$$

The next set operator to be introduced considers elements that belong to one set and a second set.

Definition 1.9 (Intersection) $A \cap B = (x \mid x \in A \text{ and } x \in B)$

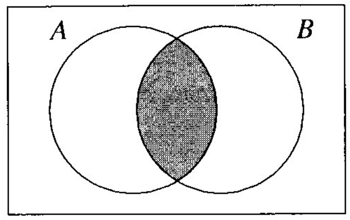

Figure 1.12: Venn diagram for $A \cap B$.

A Venn diagram with the intersection of A and B shaded is given in Figure 1.12. The notion of the intersection of two sets also generalizes naturally to more than two sets.

Definition 1.10 $A_1 \cap A_2 \cap \ldots = (x \,|\, x \in A_1, x \in A_2, \ldots)$, which applies to a finite or infinite number of sets.

Both union and intersection are symmetric operators, for example, $A \cap B = B \cap A$ and $A \cup B = B \cup A$.

Example 1.50 For the sets

$$\begin{aligned} A &= \{(x, y) \,|\, x^2 + y^2 \leq 16\} \\ B &= \{(0,0), (1,1), (2,4), (3,9), (4,16), \ldots\} \end{aligned}$$

find the intersection of A and B.

Since only the first two ordered pairs fall in the circle with radius 4 centered at the origin,

$$A \cap B = \{(0,0), (1,1)\}.$$

Definition 1.11 If $A \cap B = \mathbf{0}$, then A and B are *disjoint* or *mutually exclusive*.

A Venn diagram for the disjoint sets A and B is given in Figure 1.13

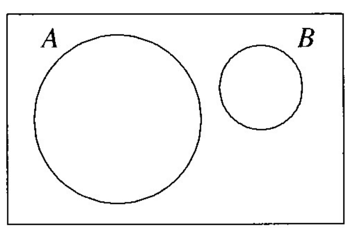

Figure 1.13: Venn diagram for disjoint sets A and B.

The final operation on a set that we present here is *complement*, which corresponds to the elements outside of a particular set.

Definition 1.12 (Complement) $A' = \{x \,|\, x \notin A\}$

A Venn diagram with the complement of A shaded is given in Figure 1.14

Example 1.51 Let the set A be the set of all real numbers on the open interval $(0, 1)$, that is

$$A = \{x \,|\, 0 < x < 1\}.$$

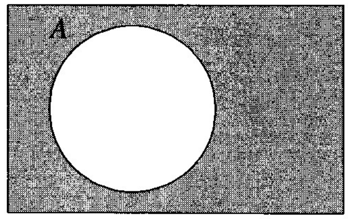

Figure 1.14: Venn diagram for A'.

Find A'.

Assuming that the universal set is the set of real numbers, the complement of A must be all of the real numbers not in the open interval $(0, 1)$, that is

$$A' = \{x \mid x \le 0 \text{ or } x \ge 1\}.$$

There are a number of comments on the small portion of set theory presented here that are given below in outline form.

1. Many authors use A^*, A^c, or $\bar{A}$ for complement, so the choice of A' used in this book is not universal. The symbols $\cup$ and $\cap$ are fairly universal.

2. Many authors use the shorthand $\bigcup_{i=1}^{n}$ and $\bigcap_{i=1}^{n}$ which parallels the use of $\sum_{i=1}^{n}$ for sums and $\prod_{i=1}^{n}$ for products. So, for example,

$$A_1 \cup A_2 \cup A_3 \cup A_4 = \bigcup_{i=1}^{4} A_i.$$

3. There are more operations on sets than the three presented here. One such operator which has applications in computer science is the *exclusive or* operator. The exclusive or of the sets A and B is denoted by $A \oplus B$ and includes all elements that are in set A or in set B, but not both. Figure 1.15 contains a Venn diagram for $A \oplus B$.

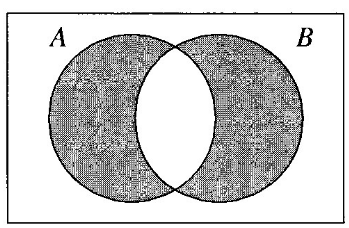

Figure 1.15: Venn diagram for $A \oplus B$.

4. DeMorgan's laws are given by

$$\left(\bigcup_{i=1}^{n} A_i\right)' = \bigcap_{i=1}^{n} A_i'$$

and

$$\left(\bigcap_{i=1}^{n} A_i\right)' = \bigcup_{i=1}^{n} A_i'.$$

5. The distributive laws are given by

$$A_1 \cap (A_2 \cup A_3) = (A_1 \cap A_2) \cup (A_1 \cap A_3)$$

and

$$A_1 \cup (A_2 \cap A_3) = (A_1 \cup A_2) \cap (A_1 \cup A_3).$$

Venn diagrams associated with sets can be useful for unscrambling befuddling counting problems, as illustrated in the following two examples.

Example 1.52 Of 100 boomers polled, 85 said they like Elvis, 62 said they like Zappa, and 5 said that they don't like either. How many of them like both Elvis and Zappa?

Let E be the set of boomers who like Elvis; let Z be the set of boomers who like Zappa. As shown in Figure 1.16, the 5 who said that they didn't like Elvis or Zappa are placed outside the sets E and Z. This leaves 95 for the remaining unfilled slots. Letting the cardinality function $N(\cdot)$ be a function that counts the number of elements in a set, the relationship between the three remaining blank spots on the Venn diagram is

$$N(E) + N(Z) - N(E \cap Z) = 95.$$

Since $N(E) = 85$ and $N(Z) = 62$, there are

$$N(E \cap Z) = 52$$

of the 100 that like both Elvis and Zappa. The remaining numbers have been placed into Figure 1.16.

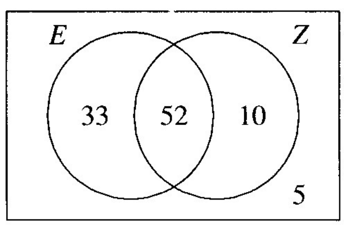

Figure 1.16: Venn diagram for counting fans.

Example 1.53 How many of the first 1000 positive integers are multiples of neither 6 nor 9?

There are 166 multiples of 6, which are

$$6, 12, 18, \ldots, 996$$

because $166 \cdot 6 = 996$. Likewise, there are 111 multiples of 9, which are

$$9, 18, 27, \ldots, 999$$

because $111 \cdot 9 = 999$. An integer is a multiple of both 6 and 9 if it is a multiple of the least common multiple of 6 and 9, which is $\text{lcm}(6, 9) = 18$. The 55 integers between 1 and 1000 that are multiples of both 6 and 9 are

$$18, 36, 54, \ldots, 990$$

because $55 \cdot 18 = 990$. Letting the set A denote the multiples of 6 and the set B denote the multiples of 9, the Venn diagram in Figure 1.17 shows the counts of the various numbers of integers in the four regions partitioned by the sets A and B.

To answer the original question, there are 778 integers between 1 and 1000 that are multiples of neither 6 nor 9.

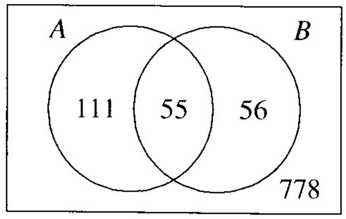

Figure 1.17: Venn diagram counting multiples.

1.4 Exercises

1.1 Each of the following questions has the same answer: 2^{32}. Write two more questions, (d) and (e), that also have 2^{32} as an answer.

(a) How many ways are there to answer a 32-question true/false test?

(b) How many integers can be represented on a 32-bit computer (ignoring the sign bit)?

(c) How many sequences of successes and failures can be recorded when Michael shoots 32 free throws?

1.2 Compute $\binom{4}{2}$.

1.3 A die is rolled four times consecutively. Find the number of possible outcomes if

(a) the order of the four outcomes is important,

(b) the order of the four outcomes is not important.

1.4 How many initial possible pairings are there for a single-elimination ping-pong tournament involving n players, $n = 2, 4, 8$?

1.5 How many triangles can be formed by connecting any three of ten distinct points that lie on an ellipse?

1.6 Chip has 4 pennies, 3 dimes, and 5 quarters in his pocket. How many different positive monetary values can he make with these coins? (*Hint*: Not all coins need to be used, but 0 cents is not a monetary value. He could, for example, make 30 cents with the three dimes.)

1.7 A license plate consists of two letters followed by three numbers.

(a) How many different license plates can be made?

(b) How many different license plates can be made if no two-character U.S. state abbreviations are allowed for the two letters?

1.8 Skip likes to use clichés. Here are some of his favorites

That's the greatest thing since sliced bread.
This fog is as thick as pea soup.
An apple a day keeps the doctor away.

Irritated by his "habit," several of Skip's friends and relatives decide to limit him to only three of his clichés per day. In an effort to keep his routine fresh, Skip vows to never use the same set of three clichés from his repertoire for the rest of his life. If Skip plans on living another 60 years (ignore leap years), what is the minimum size of his repertoire in order to achieve his goal and use exactly three clichés each day?

1.9 How many arrangements of six people in a row of six chairs are possible if

(a) there are no restrictions on the ordering?

(b) Bill and Sarah must sit together?

(c) Bill and Sarah must sit apart?

(d) there are three men and three women and no two people of the same gender can sit next to one another?

(e) there are four men and two women and the men must sit together?

(f) there are four men and two women and both the men and women must sit together?

(g) there are three married couples and the couples must sit next to one another?

1.10 Mr. Oliver North, Mr. Ray Southworth, Mrs. Mary Easterling, and Mr. Paul Westfield are playing cards.

(a) They decide to play bridge. The game of bridge begins by dealing 13 cards to each of the 4 players. How many different bridge *deals* are possible? Consider the players distinct (for example, if Mr. North and Mrs. Easterling swap their cards, it is a different deal).

(b) They decide to play poker. The game begins by dealing five cards to each person. How many *deals* are possible if the players are considered distinct?

1.11 How many ways are there to line up 8 people in a line for a photograph in a row of chairs if Rex and Laurie must sit next to one another and Bonnie and Clyde refuse to sit by one another?

1.12 Logan has 100 indistinguishable one dollar bills that he would like to invest in 4 banks. How many ways can he invest in these banks? (*Hint*: one way of investing is \$98 in the first bank, \$0 in the second bank, \$2 in the third bank, and \$0 in the fourth bank).

1.13 A committee must be chosen from 10 Republicans, 12 Democrats, and 5 Independents.

(a) How many committees of size three are possible if each member of the committee must have the same political affiliation?

(b) How many committees of size two are possible if the committee members must have a different political affiliation?

(c) How many committees of size five are possible that consist of two Republicans, two Democrats, and one Independent?

1.14 Of the $\binom{52}{5} = 2,598,960$ different five-card poker hands, how many contain

(a) the two of clubs?

(b) four of a kind?

(c) two pairs (for example, KK772 counts as a two pair hand, but KK777 does not, since it is a full house)?

1.15 A laboratory has seven female and six male rabbits. Three females and three males will be selected, then paired for mating. How many pairings are possible?

1.16 How many 6-digit numbers of the form $d_1d_2d_3d_4d_5d_6$, which range from 000000 to 999999, have the sum of the digits equal to 12? Use a combinatorics argument, then check your solution by enumeration.

1.17 How many ways can 8 people be seated around a round table?

1.18 Consider a sequence of n binary digits.

(a) How many sequences are possible?

(b) In how many of the possible sequences does the sum of the digits exceed j, where $j = 0, 1, 2, \ldots, n-1$?

1.19 Use the binomial theorem to show that for any positive integer n,

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} = 0.$$

1.20 Expand:

(a) $(x + 2y^3)^4$ using the binomial theorem,

(b) $(x + y + 3z)^3$ using the multinomial theorem.

Check the results with the Maple `expand` function.

1.21 Use the binomial theorem to show that for any positive integer n,

$$\sum_{i=0}^{n} \binom{n}{i} = 2^n.$$

1.22 Find the number of trailing 0's at the end of 100000! using a combinatorics argument.

1.23 A 12-digit number of the form $d_1d_2d_3d_4d_5d_6d_7d_8d_9d_{10}d_{11}d_{12}$ (these numbers range from 000000000000 to 999999999999) is said to be *wonderful* if

$$d_1 + d_2 + d_3 + d_4 + d_5 + d_6 = d_7 + d_8 + d_9 + d_{10} + d_{11} + d_{12}.$$

Prove that the number of wonderful numbers is even.

1.24 The game of *FreeCell* begins by dealing all of the cards from a standard deck into eight columns of cards. Four of these columns contain seven cards and four of these columns contain six cards. The order that the cards fall *within* a column is significant, but the order of the columns is not significant. How many different deals are possible?

1.25 A father has nine identical coins to give to his three children.

(a) How many allocations are possible?

(b) How many allocations are possible if each child must receive at least one coin?

(c) How many allocations are possible if each child must receive at least two coins?

1.26 There are n women who try out for a high school basketball team. Of the n, there are m that make the team. Of the m, there are five that start in the first game. Assuming that $n \ge m \ge 5$, use two different combinatorial approaches to find the number of possible ways to select a team and a starting lineup for the first game.

1.27 How many ways can a mother give n_1 identical coins to her n_2 children? Assume that n_1 and n_2 are positive integers satisfying $n_1 > n_2$. She must give all of the coins away. (*Hint*: if $n_1 = 5$ and $n_2 = 3$, then giving three coins to the firstborn, two coins to the middle child and no coins to the youngest is one possibility).

1.28 A family consists of n members. How many (pairwise) relationships are there between members?

1.29 A restaurant offers 3 appetizers, 4 entrees, and 5 desserts. How many ways are there to place an order for one appetizer, one entree, and one dessert?

1.30 How many ways can a seven-card hand be dealt from a standard 52-card deck?

1.31 Arthur, Ivah, Richard, Cindy, Larry, and Nancy need to cross a bridge at night. Exactly two may cross at a time, and they must carry a flashlight. There is only one flashlight. Assume that two people always cross the bridge and that one always returns with the flashlight. How many ways are there to get everyone across? *Illustration:* Here is *one sample* sequence: Arthur and Ivah cross, Arthur returns, Richard and Cindy cross, Richard returns, Larry and Nancy cross, Larry returns, Arthur and Richard cross, Cindy returns, Cindy and Larry cross.

1.32 How many ways can a committee of 11 people be subdivided into four subcommittees containing 4, 3, 2, and 2 people each?

1.33 A bag contains billiard balls numbered $1, 2, \ldots, 15$. How many ways can three balls be selected from the bag when the order of the selection is important?

1.34 The set A consists of all positive integers x from 1 to 15 inclusive such that $\gcd(x, 15) = 1$. List the elements of A.

1.35 A 6-letter "word" is formed by selecting 6 of the 26 letters without replacement. Two examples of such words are

FRISBE and XEALRY.

Let A_1 be the set of all words beginning with X and A_2 be the set of all words ending with Y. Find the numbers of distinct words in

(a) $A_1 \cap A_2$,

(b) $A_1 \cup A_2$.

1.36 A prime number is a positive integer that has exactly two distinct divisors: 1 and itself. Let A be the set of all prime numbers. Let B be the set of all even integers. Let C be the set of all negative integers. Draw a Venn diagram that describes the relationship among the sets A, B, and C.

1.37 Draw a Venn diagram with events A and B and shade $A' \cap B'$.

1.38 Draw a Venn diagram with events A_1, A_2, and A_3 and shade $A_1 \cap (A_2 \cup A_3)$.

1.39 Draw a Venn diagram with events A_1, A_2, and A_3 and shade $(A_1 \cap A_2') \cup A_3$.

1.40 If A and B are two events, use any of the set operations (for example, union, intersection, complement) to describe the event that neither A nor B occurs.

1.41 Let the set A be the perfect squares in the first 30 positive integers. Let the set B be the prime numbers in the first 30 positive integers. Find $N(A' \cap B')$, where N gives the cardinality (number of elements) of a set.