Probability & Statistics

COURSE CODE: MT2005

Statistics

Statistics is the art of learning from data. It is concerned with the collection of data, its subsequent description, and its analysis, which often leads to the drawing of conclusions.

Data Collection & Descriptive Statistics

Sometimes a statistical analysis begins with a given set of data: For instance, the government regularly collects and publicizes data concerning earthquake occurrences, the unemployment rate and the rate of inflation. Statistics can be used to describe, summarize, and analyze these data.

In some situations, data are not yet available; in such cases statistical theory can be used to design an appropriate experiment to generate data. For instance, suppose that an instructor is interested in determining which of two different methods for teaching computer programming to beginners is most effective. To study this question, the instructor might divide the students into two groups, and use a different teaching method for each group. At the end of the class the students can be tested and the scores of the members of the different groups compared. If the data, consisting of the test scores of members of each group, are significantly higher in one of the groups, then it might seem reasonable to suppose that the teaching method used for that group is superior.

It is important to note, however, that in order to be able to draw a valid conclusion from the data, it is essential that the students were divided into groups in such a manner that neither group was more likely to have the students with greater natural aptitude for programming. For instance, the instructor should not have let the male class members be one group and the females the other. For if so, then even if the women scored significantly higher than the men, it would not be clear whether this was due to the them, or to the fact that women may be inherently better than men at learning programming skills. The accepted way of avoiding this pitfall is to divide the class members into the two groups “at random.” This term means that the division is done in such a manner that all possible choices of the members of a group are equally likely.

At the end of the experiment, the data should be described. For instance, the scores of the two groups should be presented. In addition, summary measures such as the average score of members of each of the groups should be presented. This part of statistics, concerned with the description and summarization of data, is called descriptive statistics.

Inferential Statistics & Probability Models

After the preceding experiment is completed and the data are described and summarized, we hope to be able to draw a conclusion about which teaching method is superior. This part of statistics, concerned with the drawing of conclusions, is called inferential statistics. To be able to draw a conclusion from the data, we must take into account the possibility of chance. For instance, suppose that the average score of members of the first group is quite a bit higher than that of the second. Can we conclude that this increase is due to the teaching method used? Or is it possible that the teaching method was not responsible for the increased scores but rather that the higher scores of the first group were just a chance occurrence? For instance, the fact that a coin comes up heads 7 times in 10 flips does not necessarily mean that the coin is more likely to come up heads than tails in future flips.

Indeed, it could be a perfectly ordinary coin that, by chance, just happened to land heads 7 times out of the total of 10 flips. (On the other hand, if the coin had landed heads 47 times out of 50 flips, then we would be quite certain that it was not an ordinary coin.)

To be able to draw logical conclusions from data, we usually make some assumptions about the chances (or probabilities) of obtaining the different data values. The totality of these assumptions is referred to as a probability model for the data.

Population & Samples

In statistics, we are interested in obtaining information about a total collection of elements, which we will refer to as the population. The population is often too large for us to examine each of its members. For instance, we might have all the residents of a given state, or all the television sets produced in the last year by a particular manufacturer, or all the households in a given community. In such cases, we try to learn about the population by choosing and then examining a subgroup of its elements. This subgroup of a population is called a sample.

The sample is to be informative because it is representative of the population.

Describing Data Set

The numerical findings of a study should be presented clearly, concisely, and in such a manner that an observer can quickly obtain a feel for the essential characteristics of the data. Over the years it has been found that tables and graphs are particularly useful ways of presenting data, often revealing important features such as the range, the degree of concentration, and the symmetry of the data.

Frequency Tables & Graphs

A data set having a relatively small number of distinct values can be conveniently presented in a frequency table. For instance, Table 1 is a frequency table for a data set consisting of the starting yearly salaries (to the nearest thousand dollars) of 42 recently graduated students with B.S. degrees in computer science. Table 1 tells us, among other things, that the lowest starting salary of $47,000 was received by four of the graduates, whereas the highest salary of $60,000 was received by a single student. The most common starting salary was $52,000, and was received by 10 of the students.

Table 1 *Starting Yearly Salaries*

Starting Salary	Frequency
47	4
48	1
49	3
50	5
51	8
52	10
53	0
54	5
56	2
57	3
60	1

Data from a frequency table can be graphically represented by a line graph that plots the distinct data values on the horizontal axis and indicates their frequencies by the heights of vertical lines. A line graph of the data presented in Table 1 is shown in Figure 1.

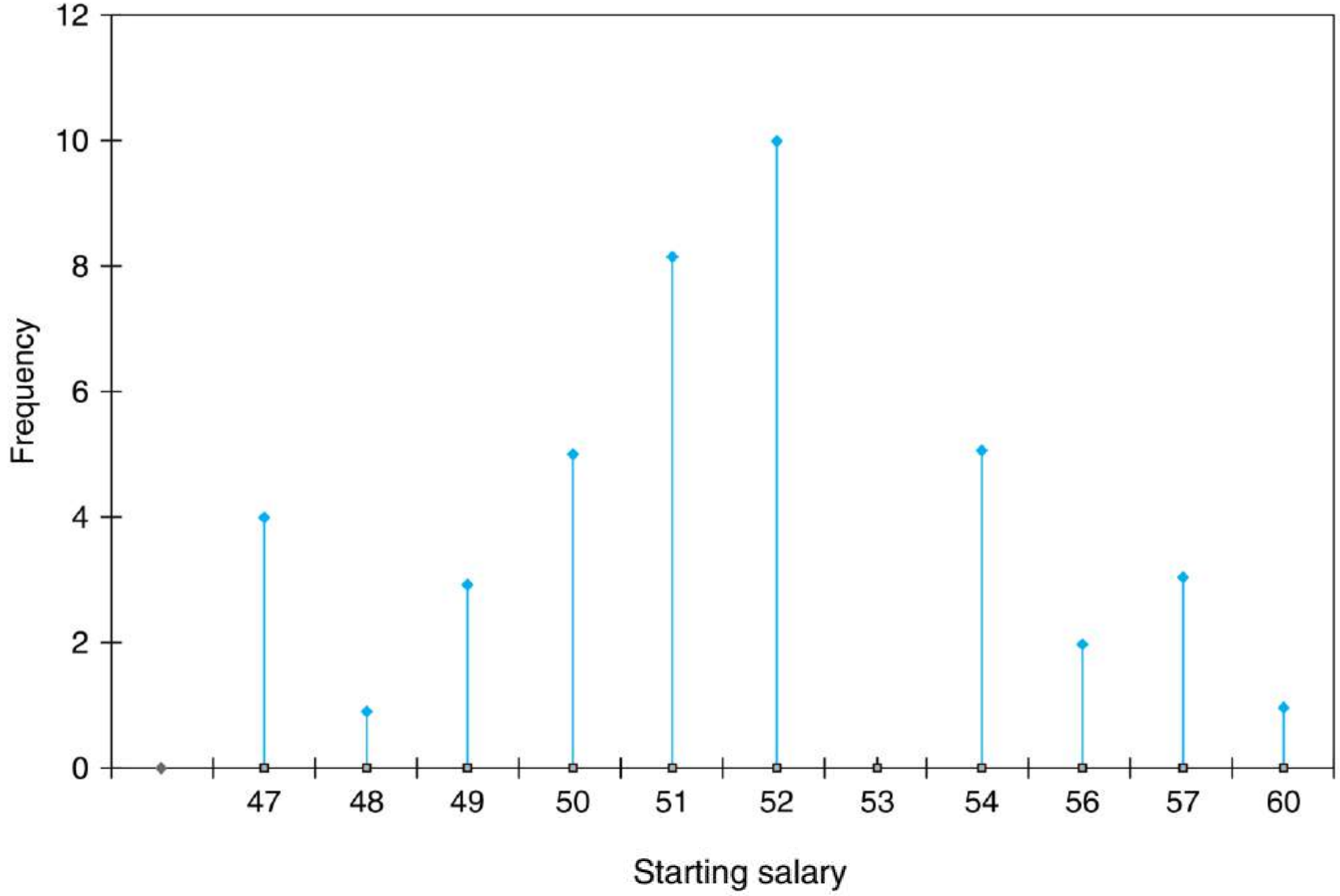


Figure 1

When the lines in a line graph are given added thickness, the graph is called a bar graph.

Figure 2 presents a bar graph.

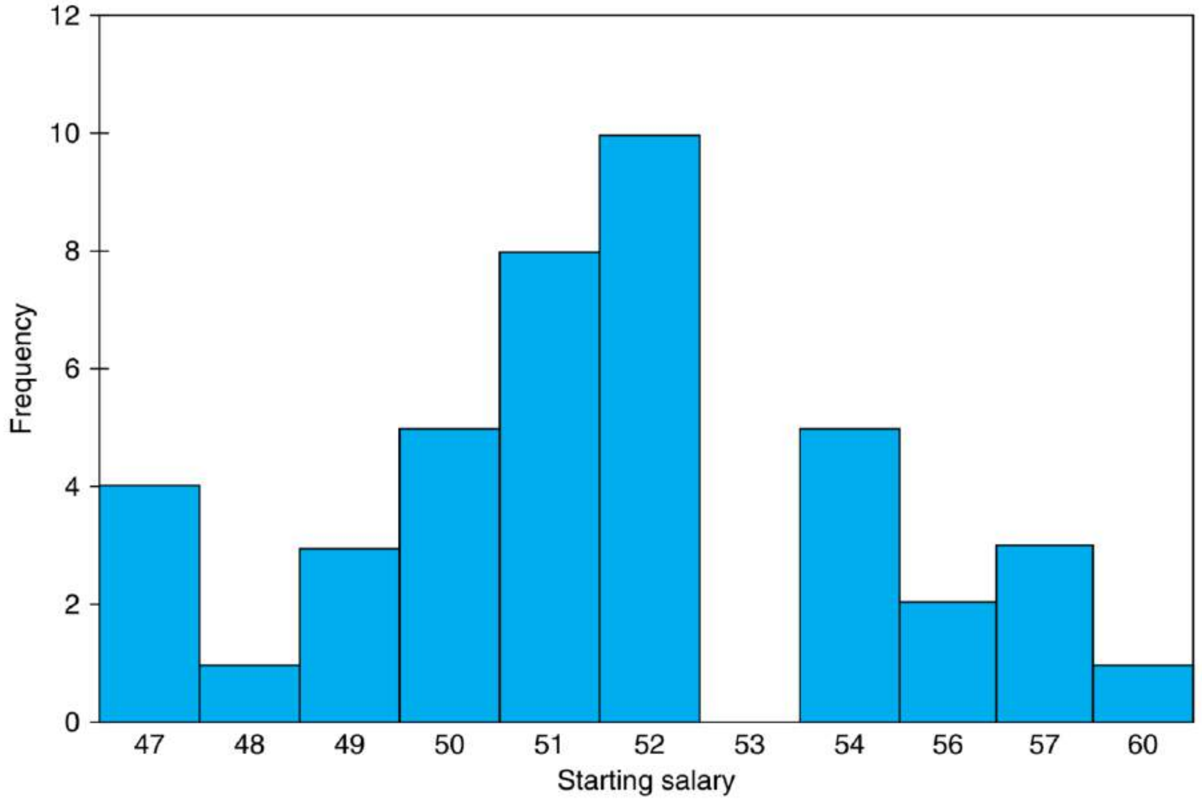


Figure 2 *Bar graph for starting salary data.*

Another type of graph used to represent a frequency table is the frequency polygon, which plots the frequencies of the different data values on the vertical axis, and then connects the plotted points with straight lines. Figure 3 presents a frequency polygon for the data of Table 1.

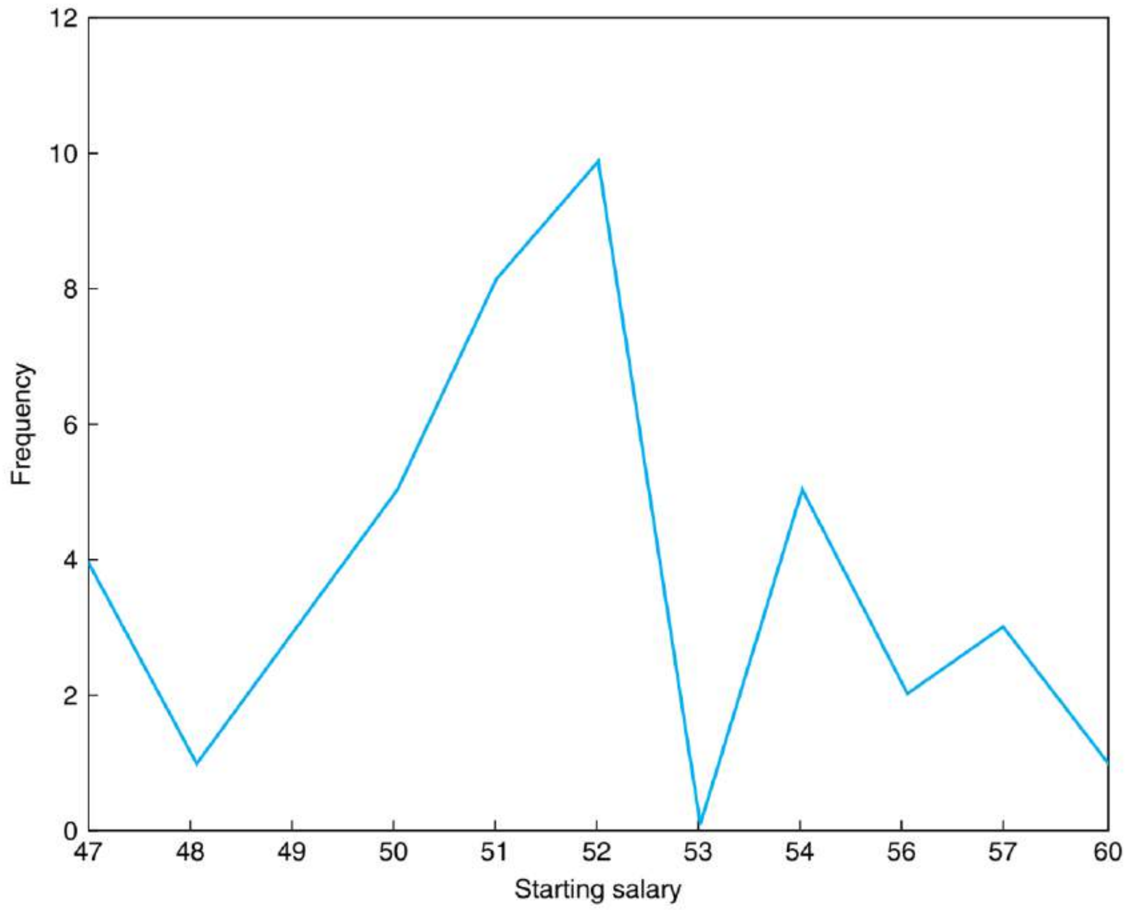


Figure 3

Frequency polygon for starting salary data.

Relative Frequency Tables & Graphs

Consider a data set consisting of n values. If f is the frequency of a particular value, then the ratio f/n is called its relative frequency. That is, the relative frequency of a data value is the proportion of the data that have that value. The relative frequencies can be represented graphically by a relative frequency line or bar graph or by a relative frequency polygon. Indeed, these relative frequency graphs will look like the corresponding graphs of the absolute frequencies except that the labels on the vertical axis are now the old labels (that gave the frequencies) divided by the total number of data points.

Table 4 is a relative frequency table for the data of Table 1. The relative frequencies are obtained by dividing the corresponding frequencies of Table 1 by 42, the size of the data set.

Starting Salary	Frequency
47	$4/42 = .0952$
48	$1/42 = .0238$
49	3/42
50	5/42
51	8/42
52	10/42
53	0
54	5/42
56	2/42
57	3/42
60	1/42

Figure 4

Pie Chart

We can construct pie chart by dividing a circle into various sections or slices. It should be used when we want to compare individual categories with the whole. If you want to compare the values of categories with each other, a bar chart may be more useful.

Problem

The following table shows the yearly budget of a family

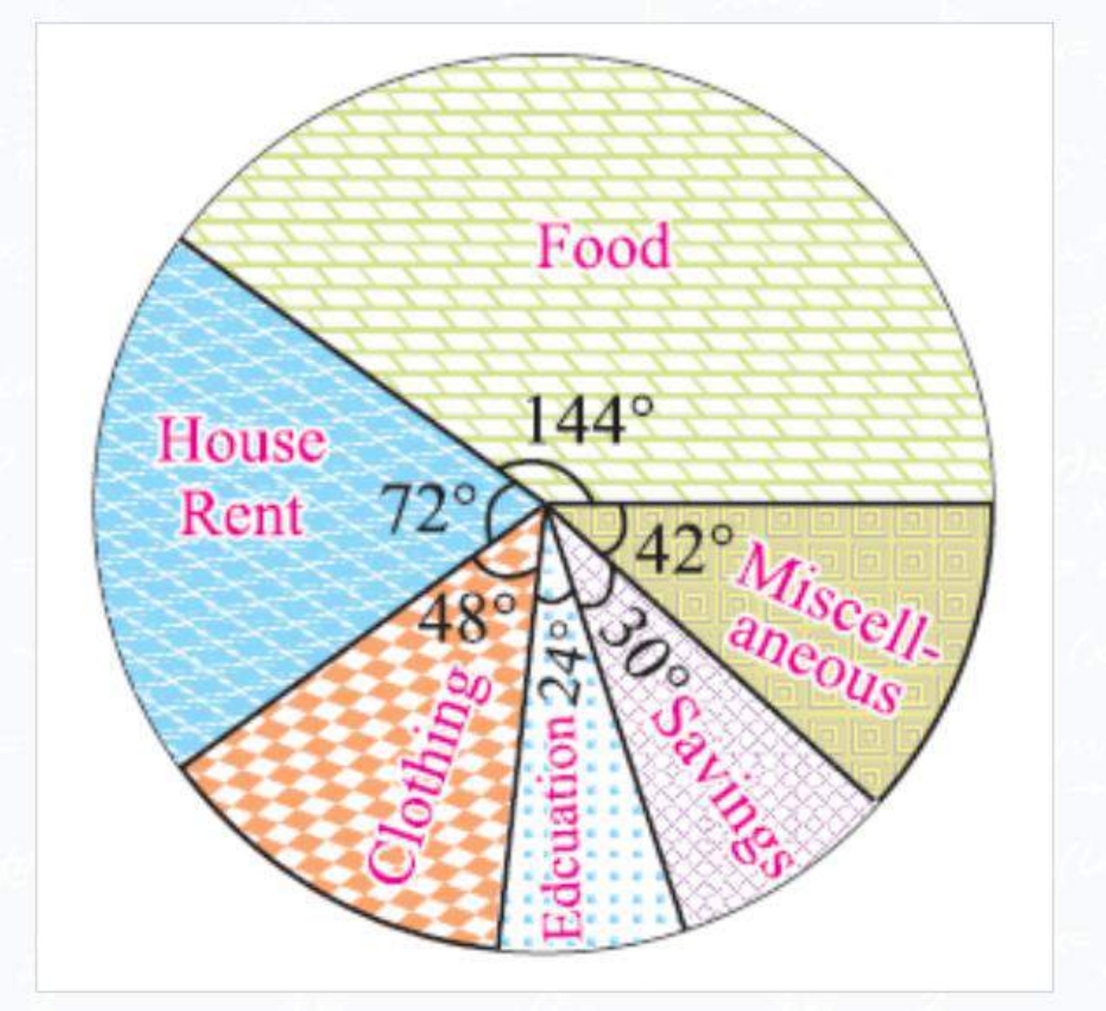
Particulars	Food	House Rent	Clothing	Education	Savings	Miscella-neous
Expenses (in $)	4800	2400	1600	800	1000	1400

Draw a pie chart to represent the above information.

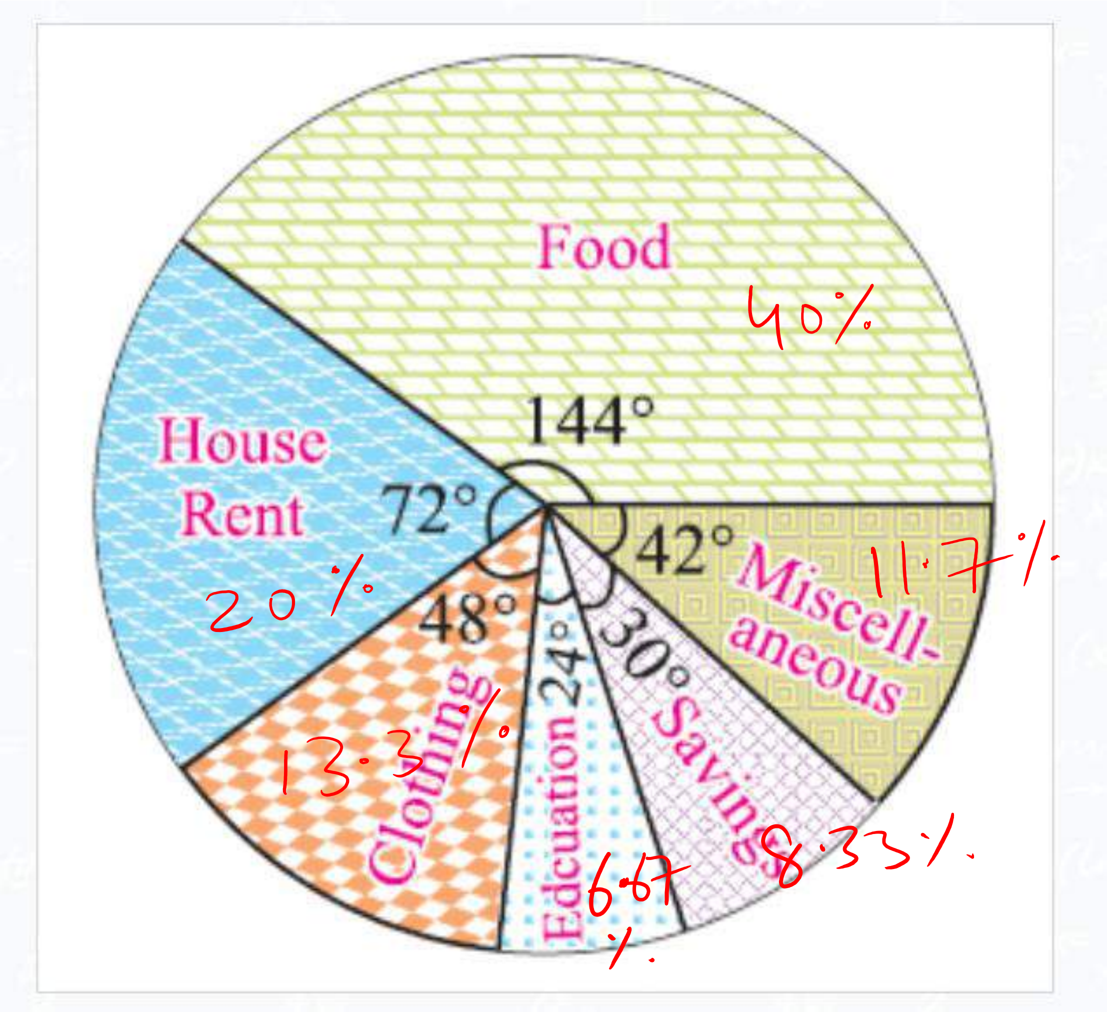
Solution

Particulars	Expenses ($)	Central angle
Food	4800	$\frac{4800}{12000} \times 360^\circ = 144^\circ$
House rent	2400	$\frac{2400}{12000} \times 360^\circ = 72^\circ$
Clothing	1600	$\frac{1600}{12000} \times 360^\circ = 48^\circ$
Education	800	$\frac{800}{12000} \times 360^\circ = 24^\circ$
Savings	1000	$\frac{1000}{12000} \times 360^\circ = 30^\circ$
Miscellaneous	1400	$\frac{1400}{12000} \times 360^\circ = 42^\circ$
Total	**12000**	**360°**

From the table, we obtain the required pie chart as shown below.



Particulars	Expenses ($)	%
Food	4800	$\frac{4800}{12000} \times 100 = 40$
House rent	2400	$\frac{2400}{12000} \times 100 = 20$
Clothing	1600	$\frac{1600}{12000} \times 100 = 13.3$
Education	800	6.67
Savings	1000	8.33
Miscellaneous	1400	11.7
Total	**12000**	100


Food
144°
40%
House
Rent
72°
20%
48°
Clothing
13.3%
24°
Edcuation
6.67
%
30°
Savings
8.33%
42°
Miscell-
aneous
11.7%

Grouped data, histograms, ogives, and stem and leaf plots

Using a line or a bar graph to plot the frequencies of data values is often an effective way of portraying a data set. However, for some data sets the number of distinct values is too large to utilize this approach. Instead, in such cases, it is useful to divide the values into groupings, or class intervals, and then plot the number of data values falling in each class interval. The number of class intervals chosen should be a trade-off between (1) choosing too few classes at a cost of losing too much information about the actual data values in a class and (2) choosing too many classes, which will result in the frequencies of each class being too small for a pattern to be discernible.

The endpoints of a class interval are called the class boundaries. We will adopt the left end inclusion convention, which stipulates that a class interval contains its left-end but not its right-end boundary point. Thus, for instance, the class interval 20–30 contains all values that are both greater than or equal to 20 and less than 30.

Table 2 (on next slide) presents the lifetimes of 200 incandescent lamps. A class frequency table for the data of Table 2 is presented in Table 3. The class intervals are of length 100, with the first one starting at 500.

Table 2 *Life in Hours of 200 Incandescent Lamps*

Item Lifetimes

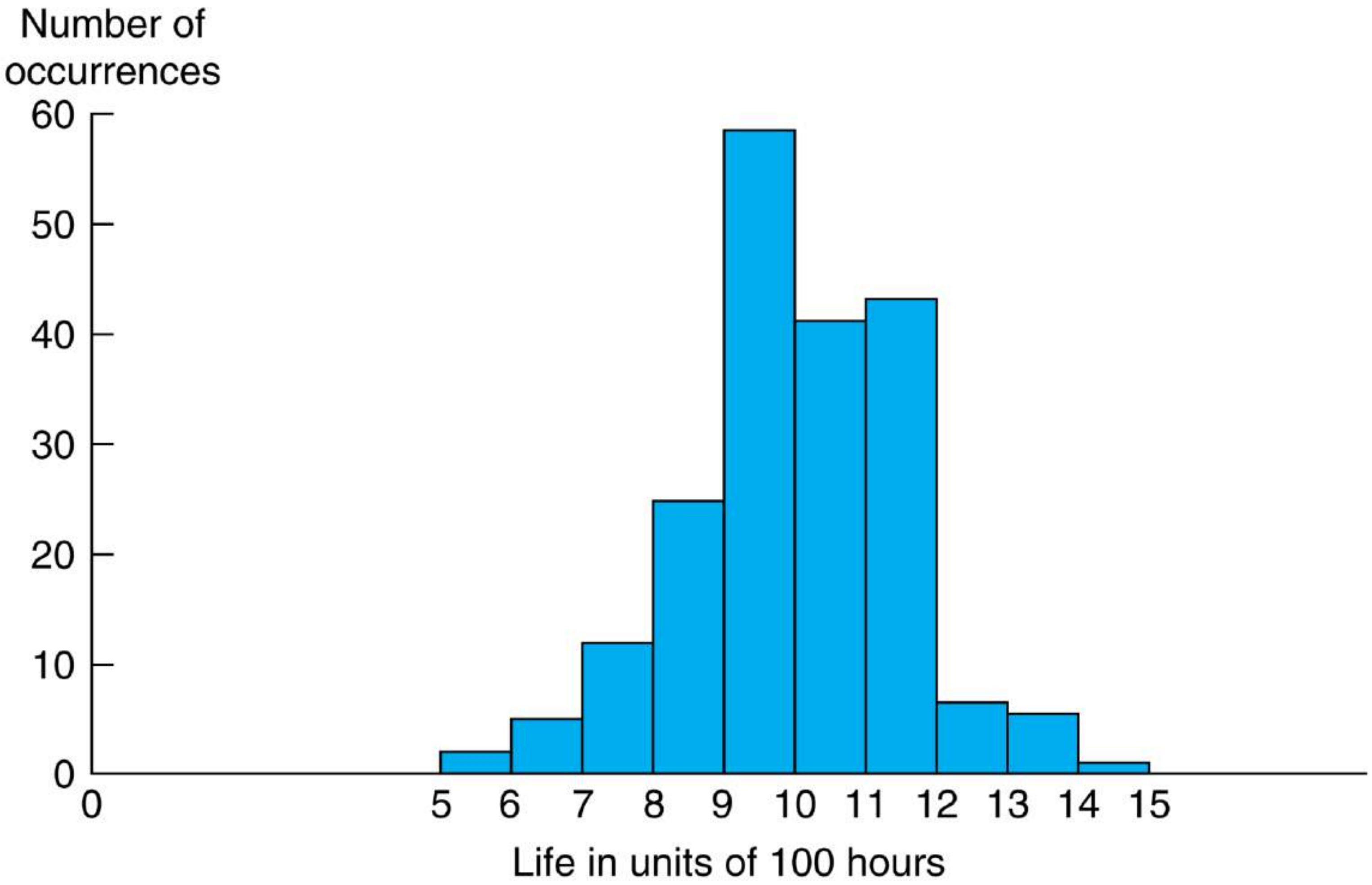
1,067	919	1,196	785	1,126	936	918	1,156	920	948
855	1,092	1,162	1,170	929	950	905	972	1,035	1,045
1,157	1,195	1,195	1,340	1,122	938	970	1,237	956	1,102
1,022	978	832	1,009	1,157	1,151	1,009	765	958	902
923	1,333	811	1,217	1,085	896	958	1,311	1,037	702
521	933	928	1,153	946	858	1,071	1,069	830	1,063
930	807	954	1,063	1,002	909	1,077	1,021	1,062	1,157
999	932	1,035	944	1,049	940	1,122	1,115	833	1,320
901	1,324	818	1,250	1,203	1,078	890	1,303	1,011	1,102
996	780	900	1,106	704	621	854	1,178	1,138	951
1,187	1,067	1,118	1,037	958	760	1,101	949	992	966
824	653	980	935	878	934	910	1,058	730	980
844	814	1,103	1,000	788	1,143	935	1,069	1,170	1,067
1,037	1,151	863	990	1,035	1,112	931	970	932	904
1,026	1,147	883	867	990	1,258	1,192	922	1,150	1,091
1,039	1,083	1,040	1,289	699	1,083	880	1,029	658	912
1,023	984	856	924	801	1,122	1,292	1,116	880	1,173
1,134	932	938	1,078	1,180	1,106	1,184	954	824	529
998	996	1,133	765	775	1,105	1,081	1,171	705	1,425
610	916	1,001	895	709	860	1,110	1,149	972	1,002

Table 3 *A Class Frequency Table*

Class Interval	Frequency (Number of Data Values in the Interval)
500–600	2
600–700	5
700–800	12
800–900	25
900–1000	58
1000–1100	41
1100–1200	43
1200–1300	7
1300–1400	6
1400–1500	1

A bar graph plot of class data, with the bars placed adjacent to each other, is called a histogram. The vertical axis of a histogram can represent either the class frequency or the relative class frequency; in the former case the graph is called a frequency histogram and in the latter a relative frequency histogram.

A frequency histogram



An efficient way of organizing a small- to moderate-sized data set is to utilize a stem and leaf plot. Such a plot is obtained by first dividing each data value into two parts —its stem and its leaf. For instance, if the data are all two-digit numbers, then we could let the stem part of a data value be its tens digit and let the leaf be its ones digit. Thus, for instance, the value 62 is expressed as

Stem	**Leaf**
6	2

and the two data values 62 and 67 can be represented as

Stem	**Leaf**
6	2, 7

Example

Table 4 (on next slide) presents the per capita personal income for each of the 50 states and the District of Columbia. We can represent the data by a stem and leaf plot.

Table 4 Capita Personal Income (Dollars per Person), 2002

State name		State name		State name	
United States	30,941	Kentucky	25,579	Ohio	29,405
Alabama	25,128	Louisiana	25,446	Oklahoma	25,575
Alaska	32,151	Maine	27,744	Oregon	28,731
Arizona	26,183	Maryland	36,298	Pennsylvania	31,727
Arkansas	23,512	Massachusetts	39,244	Rhode Island	31,319
California	32,996	Michigan	30,296	South Carolina	25,400
Colorado	33,276	Minnesota	34,071	South Dakota	26,894
Connecticut	42,706	Mississippi	22,372	Tennessee	27,671
Delaware	32,779	Missouri	28,936	Texas	28,551
District of Columbia	42,120	Montana	25,020	Utah	24,306
Florida	29,596	Nebraska	29,771	Vermont	29,567
Georgia	28,821	Nevada	30,180	Virginia	32,922
Hawaii	30,001	New Hampshire	34,334	Washington	32,677
Idaho	25,057	New Jersey	39,453	West Virginia	23,688
Illinois	33,404	New Mexico	23,941	Wisconsin	29,923
Indiana	28,240	New York	36,043	Wyoming	30,578
Iowa	28,280	North Carolina	27,711		
Kansas	29,141	North Dakota	26,982		

The data presented in Table 4 are represented in the following stem-and-leaf plot. Note that the values of the leaves are put in the plot in increasing order.

22	372
23	512, 688, 941
24	706
25	020, 057, 128, 400, 446, 575, 57
26	183, 894, 982
27	671, 711, 744
28	240, 280, 551, 731, 821, 936
29	141, 405, 567, 596, 771, 923
30	001, 180, 296, 578
31	319, 727
32	151, 677, 779, 922, 996
33	276, 404
34	071, 334
36	043, 298
39	244, 453
42	120, 706

Example

The following stem-and-leaf plot represents the weights of 80 attendees at a sporting convention. The stem represents the tens digit and the hundred digits, and the leaves are the ones digit.

10	2, 3, 3, 4, 7	(5)
11	0, 1, 2, 2, 3, 6, 9	(7)
12	1, 2, 4, 4, 6, 6, 6, 7, 9	(9)
13	1, 2, 2, 5, 5, 6, 6, 8, 9	(9)
14	0, 4, 6, 7, 7, 9, 9	(7)
15	1, 1, 5, 6, 6, 6, 7	(7)
16	0, 1, 1, 1, 2, 4, 5, 6, 8, 8	(10)
17	1, 1, 3, 5, 6, 6, 6	(7)
18	1, 2, 2, 5, 5, 6, 6, 9	(8)
19	0, 0, 1, 2, 4, 5	(6)
20	9, 9	(2)
21	7	(1)
22	1	(1)
23		(0)
24	9	(1)

The numbers in parentheses on the right represent the number of values in each stem class. These summary numbers are often useful. They tell us, for instance, that there are 10 values having stem 16; that is, 10 individuals have weights between 160 and 169. Note that a stem without any leaves (such as stem value 23) indicates that there are no occurrences in that class.

Summarizing data sets

Modern-day experiments often deal with huge sets of data. To obtain a feel for such a large amount of data, it is useful to be able to summarize it by some suitably chosen measures.

Sample mean, sample median, and sample mode

Let's introduce some statistics that are useful for describing the center of a set of data value.

Definition

The *sample mean*, designated by $\bar{x}$, is defined by

$$\bar{x} = \sum_{i=1}^{n} x_i/n$$

The computation of the sample mean can often be simplified by noting that if for constants a and b

$$y_i = ax_i + b, \qquad i = 1, \ldots, n$$

then the sample mean of the data set $y_1, \ldots, y_n$ is

$$\bar{y} = \sum_{i=1}^{n}(ax_i + b)/n = \sum_{i=1}^{n} ax_i/n + \sum_{i=1}^{n} b/n = a\bar{x} + b$$

Note: More often we use the term statistic for numerical quantity computed from a data set.

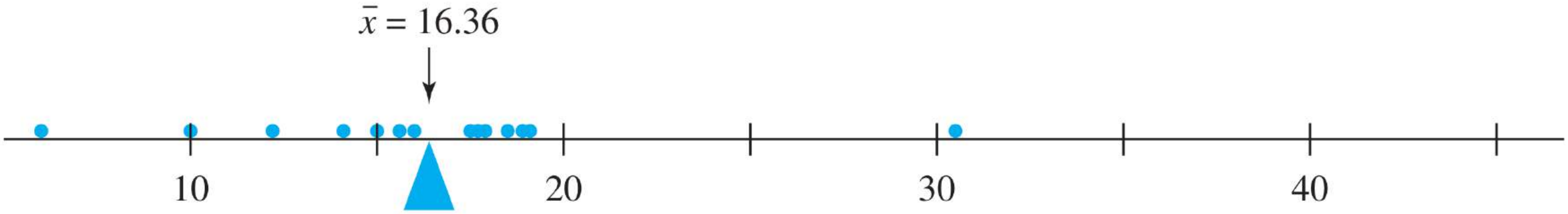
Consider the data

$$\begin{array}{lllll}
x_1 = 16.0 & x_2 = 30.5 & x_3 = 17.7 & x_4 = 17.5 & x_5 = 14.1 \\
x_6 = 10.0 & x_7 = 15.6 & x_8 = 15.0 & x_9 = 19.1 & x_{10} = 17.9 \\
x_{11} = 18.9 & x_{12} = 18.5 & x_{13} = 12.2 & x_{14} = 6.0 &
\end{array}$$

Figure 1.14 shows a dotplot of the data; a water-absorption percentage in the mid-teens appears to be "typical." With $\Sigma x_i = 229.0$, the sample mean is

$$\bar{x} = \frac{229.0}{14} = 16.36$$

A physical interpretation of the sample mean demonstrates how it assesses the center of a sample. Think of each dot in the dotplot as representing a 1 kg weight. Then a fulcrum placed with its tip on the horizontal axis will balance precisely when it is located at x. So the sample mean can be regarded as the balance point of the distribution of observations.



Problem

The winning scores in the U.S. Masters golf tournament in the years from 1999 to 2008 were as follows:

280, 278, 272, 276, 281, 279, 276, 281, 289, 280

Find the sample mean of these scores.

SOLUTION Rather than directly adding these values, it is easier to first subtract 280 from each one to obtain the new values $y_i = x_i - 280$:

$$0, -2, -8, -4, 1, -1, -4, 1, 9, 0$$

Because the arithmetic average of the transformed data set is

$$\bar{y} = -8/10$$

it follows that

$$\bar{x} = \bar{y} + 280 = 279.2$$

Sometimes we want to determine the sample mean of a data set that is presented in a frequency table listing the k distinct values $v_1, \ldots, v_k$ having corresponding frequencies $f_1, \ldots, f_k$. Since such a data set consists of $n = \sum_{i=1}^{k} f_i$ observations, with the value v_i appearing f_i times, for each $i = 1, \ldots, k$, it follows that the sample mean of these n data values is

$$\bar{x} = \sum_{i=1}^{k} v_i f_i / n$$

In expanded form, we have

$$\bar{x} = \frac{f_1}{n} v_1 + \frac{f_2}{n} v_2 + \cdots + \frac{f_k}{n} v_k$$

Problem

The number of iPhone sold daily by a small company for the past 6 days has been arranged in the following frequency table:

Value	Frequency
3	2
4	1
5	3

What is the sample mean?

Solution Since the original data set consists of the 6 values

$$3, 3, 4, 5, 5, 5$$

it follows that the sample mean is

$$\bar{x} = \frac{3 + 3 + 4 + 5 + 5 + 5}{6}$$

$$= \frac{3 \times 2 + 4 \times 1 + 5 \times 3}{6} = \boxed{\frac{25}{6}}$$

Another statistic used to indicate the center of a data set is the sample median; loosely speaking, it is the middle value when the data set is arranged in increasing order.

Definition

Order the values of a data set of size n from smallest to largest. If n is odd, the *sample median* is the value in position $(n+1)/2$; if n is even, it is the average of the values in positions $n/2$ and $n/2+1$.

Thus the sample median of a set of three values is the second smallest; of a set of four values, it is the average of the second and third smallest.

Order the data values from smallest to largest. If the number of data values is odd, then the sample median is the middle value in the ordered list; if it is even, then the sample median is the average of the two middle values.

Problem

The following data represent the number of weeks it took seven individuals to obtain their driver's licenses. Find the sample median.

$$2, 110, 5, 7, 6, 7, 3$$

Solution

First arrange the data in increasing order.

$$2, 3, 5, 6, 7, 7, 110$$

Since the sample size is 7, it follows that the sample median is the fourth-smallest value. That is, the sample median number of weeks it took to obtain a driver's license is $m = 6$ weeks.

Problem

The following data represent the number of days it took 6 individuals to quit smoking after completing a course designed for this purpose.

$$1, 2, 3, 5, 8, 100$$

What is the sample median?

Solution

Since the sample size is 6, the sample median is the average of the two middle values; thus,

$$m = \frac{3 + 5}{2} = 4$$

That is, the sample median is 4 days.

Comparison of Mean with Median

The sample mean and sample median are both useful statistics for describing the central tendency of a data set. The sample mean, being the arithmetic average, makes use of all the data values. The sample median, which makes use of only one or two middle values, is not affected by extreme values.

Mode

Another statistic that has been used to indicate the central tendency of a data set is the *sample mode*, defined to be the value that occurs with the greatest frequency. If no single value occurs most frequently, then all the values that occur at the highest frequency are called *modal values*.

Problem

The following frequency table gives the values obtained in 40 rolls of a die.

Value	Frequency
1	9
2	8
3	5
4	5
5	6
6	7

Find **(a)** the sample mean, **(b)** the sample median, and **(c)** the sample mode.

Solution

(a) The sample mean is

$$\bar{x} = (9 + 16 + 15 + 20 + 30 + 42)/40 = 3.05$$

(b) The sample median is the average of the 20th and 21st smallest values, and is thus equal to 3. **(c)** The sample mode is 1, the value that occurred most frequently.

Consider the following data sets

$$A: 1, 2, 5, 6, 6 \qquad B: -40, 0, 5, 20, 35$$

Although both data sets have the same sample mean and sample median there is a much greater variability or spread in the values of Set **B** than in the set **A**. For measuring variability in data we have statistics sample variance and standard deviation which we are discussing in coming slides.

Sample Variance

The *sample variance*, call it s^2, of the data set $x_1, \ldots, x_n$ is defined by

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n - 1)$$

Just as $\bar{x}$ will be used to make inferences about the population mean μ, we should define the sample variance so that it can be used to make inferences about σ^2. Now note that σ^2 involves squared deviations about the population mean μ. If we actually knew the value of μ, then we could define the sample variance as the average squared deviation of the sample x_i's about μ. However, the value of μ is almost never known, so the sum of squared deviations about $\bar{x}$ must be used. But *the x_i's tend to be closer to their average $\bar{x}$ than to the population average μ*. To compensate for this, the divisor $n - 1$ is used rather than the sample size n. In other words, if we used a divisor n in the sample variance, then the resulting quantity would tend to underestimate σ^2 (produce estimated values that are too small on the average), whereas dividing by the slightly smaller $n - 1$ corrects this underestimating.

Problem

Find the sample variance of data set **A**: 1, 2, 5, 6, 6

Solution

It is determined as follows:

x_i	1	2	5	6	6
$\overline{x}$	4	4	4	4	4
$x_i - \overline{x}$	–3	–2	1	2	2
$(x_i - \overline{x})^2$	9	4	1	4	4

Hence, for data set A,

$$s^2 = \frac{9+4+1+4+4}{4} = 5.5$$

Problem

Find the sample variance of data set $\boldsymbol{B}$: $-40, 0, 5, 20, 35$

Solution

The sample mean for data set B is also $\overline{x} = 4$. Therefore, for this set, we have

x_i	−40	0	5	20	35
$x_i - \overline{x}$	−44	−4	1	16	31
$(x_i - \overline{x})^2$	1936	16	1	256	961

Thus,

$$s^2 = \frac{3170}{4} = 792.5$$

The following algebraic identity is useful for computing the sample variance by hand:

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$$

The sample variance remains unchanged when a constant is added to each data value.

The above statement make sense because adding or subtracting *c* shifts the location of the data set but leaves distances between data values unchanged.

Problem

The following data give the yearly numbers of law enforcement officers killed in the United States over 10 years:

$$164, 165, 157, 164, 152, 147, 148, 131, 147, 155$$

Find the sample variance of the number killed in these years.

Solution

Rather than working directly with the given data, let us subtract the value 150 from each data item. (That is, we are adding $c = -150$ to each data value.) This results in the new data set

$$14, 15, 7, 14, 2, -3, -2, -19, -3, 5$$

Its sample mean is

$$\bar{y} = \frac{14 + 15 + 7 + 14 + 2 - 3 - 2 - 19 - 3 + 5}{10} = 3.0$$

The sum of the squares of the new data is

$$\sum_{i=1}^{10} y_i^2 = 14^2 + 15^2 + 7^2 + 14^2 + 2^2 + 3^2 + 2^2 + 19^2 + 3^2 + 5^2 = 1078$$

$$\sum_{i=1}^{10} (y_i - \bar{y})^2 = 1078 - 10(9) = 988$$

Hence, the sample variance of the revised data, which is equal to the sample variance of the original data, is

$$s^2 = \frac{988}{9} \approx \boxed{109.78}$$

Sample Standard Deviation

The quantity s, defined by

$$s = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 / (n - 1)}$$

is called the *sample standard deviation.*

The sample standard deviation is measured in the same units as the data.

It may appear to the reader that the use of both the sample variance and the sample standard deviation is redundant. Both measures reflect the same concept in measuring variability, but the sample standard deviation measures variability in linear units whereas the sample variance is measured in squared units. For instance, if the data are in feet, then the sample variance will be expressed in units of square feet and the sample standard deviation in units of feet.

Note

If we shift the measurements by adding or subtracting a constant, then the measure of center gets shifted by the same amount, but the measure of variation is unaffected by any shift in measurements.

Sample Percentiles and Box Plots

The *sample 100*p *percentile* is that data value such that $100p$ percent of the data are less than or equal to it and $100(1-p)$ percent are greater than or equal to it. If two data values satisfy this condition, then the sample $100p$ percentile is the arithmetic average of these two values.

Method for finding Sample Percentile

- First arrange the data in increasing order.
- If np is not an integer, then the data value whose position is the smallest integer exceeding np is the sample $100p$ percentile.
- On the other hand, if np is an integer, then the sample $100p$ percentile is the average of the values in positions np and $np + 1$.

Problem

Table 2.6 lists the populations of the 25 most populous U.S. cities for the year 1994. For this data set, find (a) the sample 10 percentile and (b) the sample 80 percentile.

TABLE 2.6 *Population of 25 Largest U.S. Cities, July 2006*

Rank	City	Population
1	New York, NY	8,250,567
2	Los Angeles, CA	3,849,378
3	Chicago, IL	2,833,321
4	Houston, TX	2,144,491
5	Phoenix, AR	1,512,986
6	Philadelphia, PA	1,448,394
7	San Antonio, TX	1,296,682
8	San Diego, CA	1,256,951

9	Dallas, TX	1,232,940
10	San Jose, CA	929,936
11	Detroit, MI	918,849
12	Jacksonville, FL	794,555
13	Indianapolis, IN	785,597
14	San Francisco, CA	744,041
15	Columbus, OH	733,203
16	Austin, TX	709,893
17	Memphis, TN	670,902
18	Fort Worth, TX	653,320
19	Baltimore, MD	640,961
20	Charlotte, NC	630,478
21	El Paso, TX	609,415
22	Milwaukee, WI	602,782
23	Boston, MA	590,763
24	Seattle, WA	582,454
25	Washington, DC	581,530

Solution

(a) Because the sample size is 25 and 25(.10) = 2.5, the sample 10 percentile is the third smallest value, equal to 590, 763.

(b) Because 25(.80) = 20, the sample 80 percentile is the average of the twentieth and the twenty-first smallest values. Hence, the sample 80 percentile is

$$\frac{1{,}512{,}986 + 1{,}448{,}394}{2} = 1{,}480{,}690$$

The sample 50 percentile is, of course, just the sample median. Along with the sample 25 and 75 percentiles, it makes up the sample quartiles.

Quartiles

The sample 25 percentile is called the *first quartile*; the sample 50 percentile is called the sample median or the *second quartile*; the sample 75 percentile is called the *third quartile*.

The quartiles break up a data set into four parts, with roughly 25 percent of the data being less than the first quartile, 25 percent being between the first and second quartile, 25 percent being between the second and third quartile, and 25 percent being greater than the third quartile.

Problem

Noise is measured in decibels, denoted as dB. One decibel is about the level of the weakest sound that can be heard in a quiet surrounding by someone with good hearing; a whisper measures about 30 dB; a human voice in normal conversation is about 70 dB; a loud radio is about 100 dB. Ear discomfort usually occurs at a noise level of about 120 dB.

The following data give noise levels measured at 36 different times directly outside of Grand Central Station in New York.

82, 89, 94, 110, 74, 122, 112, 95, 100, 78, 65, 60, 90, 83, 87, 75, 114, 85
69, 94, 124, 115, 107, 88, 97, 74, 72, 68, 83, 91, 90, 102, 77, 125, 108, 65

Determine the quartiles.

Solution A stem and leaf plot of the data is as follows:

6	0, 5, 5, 8, 9
7	2, 4, 4, 5, 7, 8
8	2, 3, 3, 5, 7, 8, 9
9	0, 0, 1, 4, 4, 5, 7
10	0, 2, 7, 8
11	0, 2, 4, 5
12	2, 4, 5

The first quartile is 76, the average of the 9th and 10th smallest data values; the second quartile is 89.5, the average of the 18th and 19th smallest values; the third quartile is 104.5, the average of the 27th and 28th smallest values.

Box Plot

A box plot is often used to plot some of the summarizing statistics of a data set. A straight line segment stretching from the smallest to the largest data value is drawn on a horizontal axis; imposed on the line is a “box,” which starts at the first and continues to the third quartile, with the value of the second quartile indicated by a vertical line.

Table 2.1 Starting Yearly Salaries.

Starting Salary	Frequency
57	4
58	1
59	3
60	5
61	8
62	10
63	0
64	5
66	2
67	3
70	1

For instance, the 42 data values presented in Table 2.1 go from a low value of 57 to a high value of 70. The value of the first quartile (equal to the value of the 11th smallest on the list) is 60; the value of the second quartile (equal to the average of the 21st and 22nd smallest values) is 61.5; and the value of the third quartile (equal to the value of the 32nd smallest on the list) is 64. The box plot for this data set is shown in Figure 2.7.

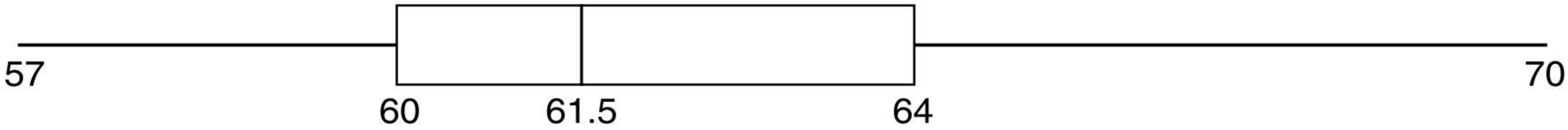


FIGURE 2.7

The length of the line segment on the box plot, equal to the largest minus the smallest data value, is called the **range of the data**. Also, the length of the box itself, equal to the third quartile minus the first quartile, is called the **interquartile range**.

Normal Data Set

Many of the large data sets observed in practice have histograms that are similar in shape. These histograms often reach their peaks at the sample median and then decrease on both sides of this point in a bell-shaped symmetric fashion. Such data sets are said to be *normal* and their histograms are called *normal histograms*. Figure 2.8 is the histogram of a normal data set.

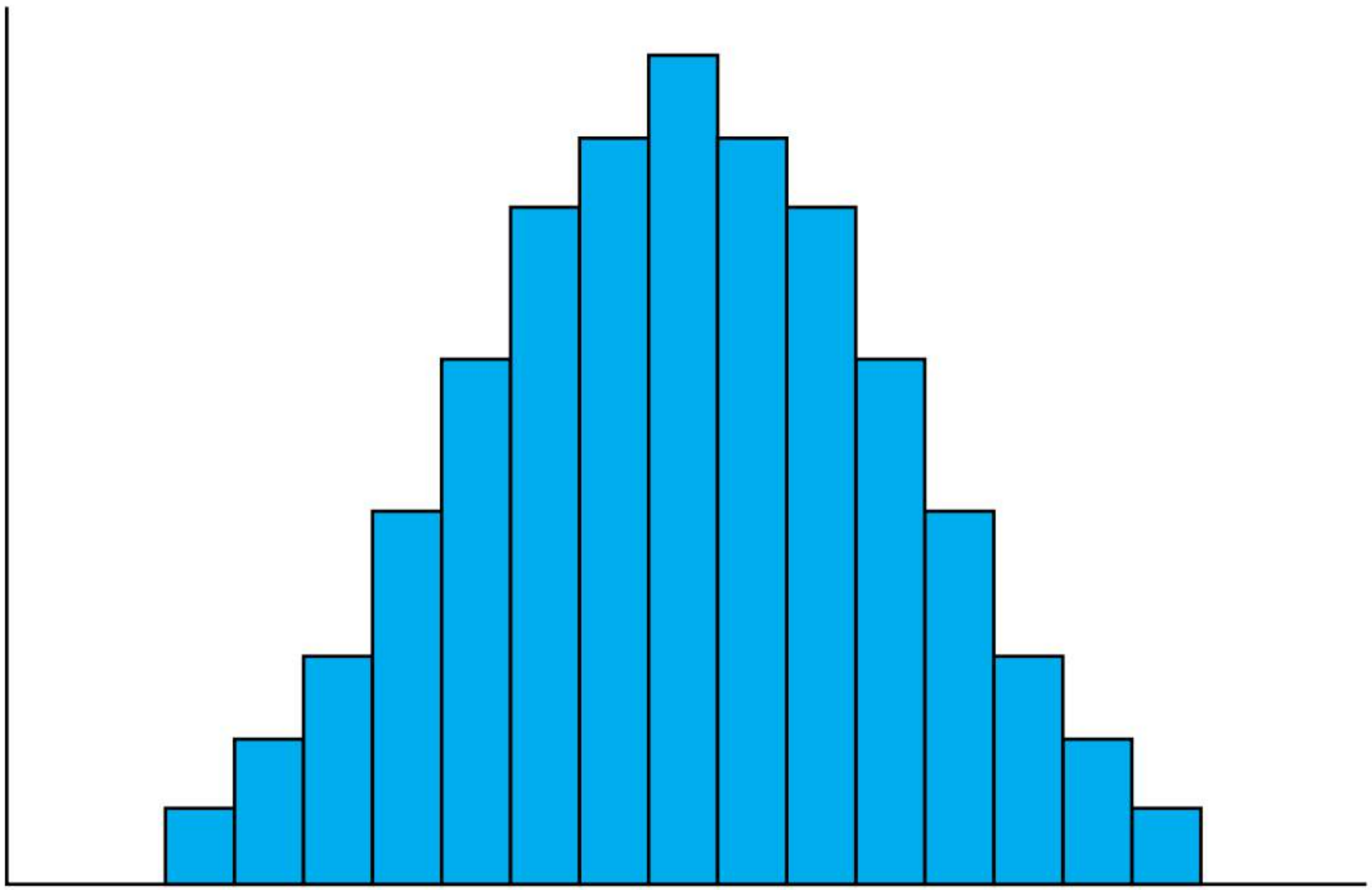


FIGURE 2.8 *Histogram of a normal data set.*

Definition *A data set is said to be* normal *if a histogram describing it has the following properties:*

1. *It is highest at the middle interval.*
2. *Moving from the middle interval in either direction, the height decreases in such a way that the entire histogram is bell-shaped.*
3. *The histogram is symmetric about its middle interval.*

If the histogram of a data set is close to being a normal histogram, then we say that the data set is approximately normal. For instance, we would say that the histogram given in Figure 2.9 is from an approximately normal data set, whereas the ones presented in Figures 2.10 and 2.11 are not (because each is too nonsymmetric).

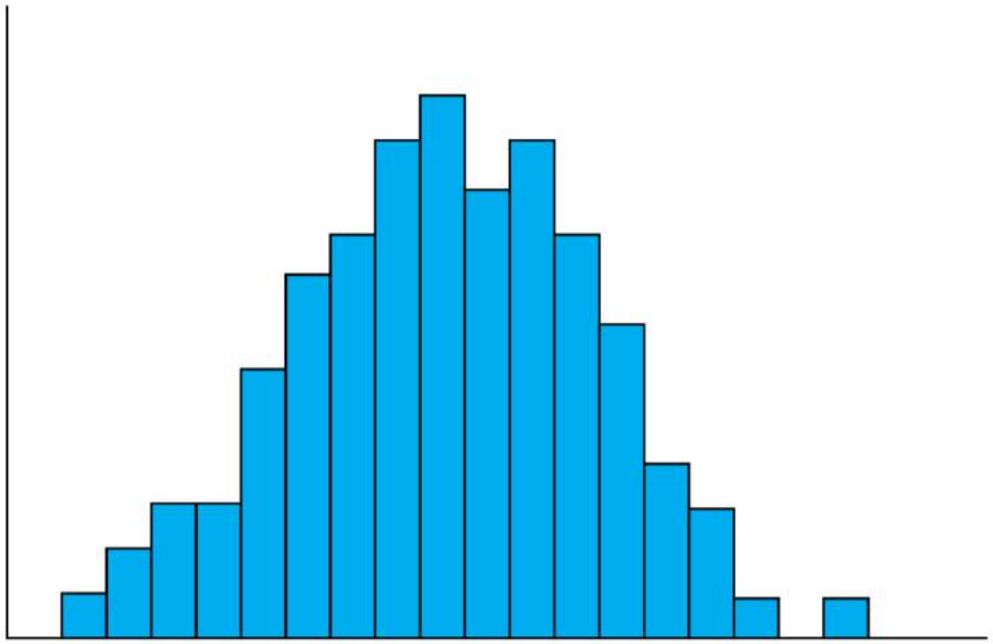


FIGURE 2.9 *Histogram of an approximately normal data set.*

Any data set that is not approximately symmetric about its sample median is said to be skewed. It is "skewed to the right" if it has a long tail to the right and "skewed to the left" if it has a long tail to the left. Thus the data set presented in Figure 2.10 is skewed to the left and the one of Figure 2.11 is skewed to the right.

It follows from the symmetry of the normal histogram that a data set that is approximately normal will have its sample mean and sample median approximately equal.

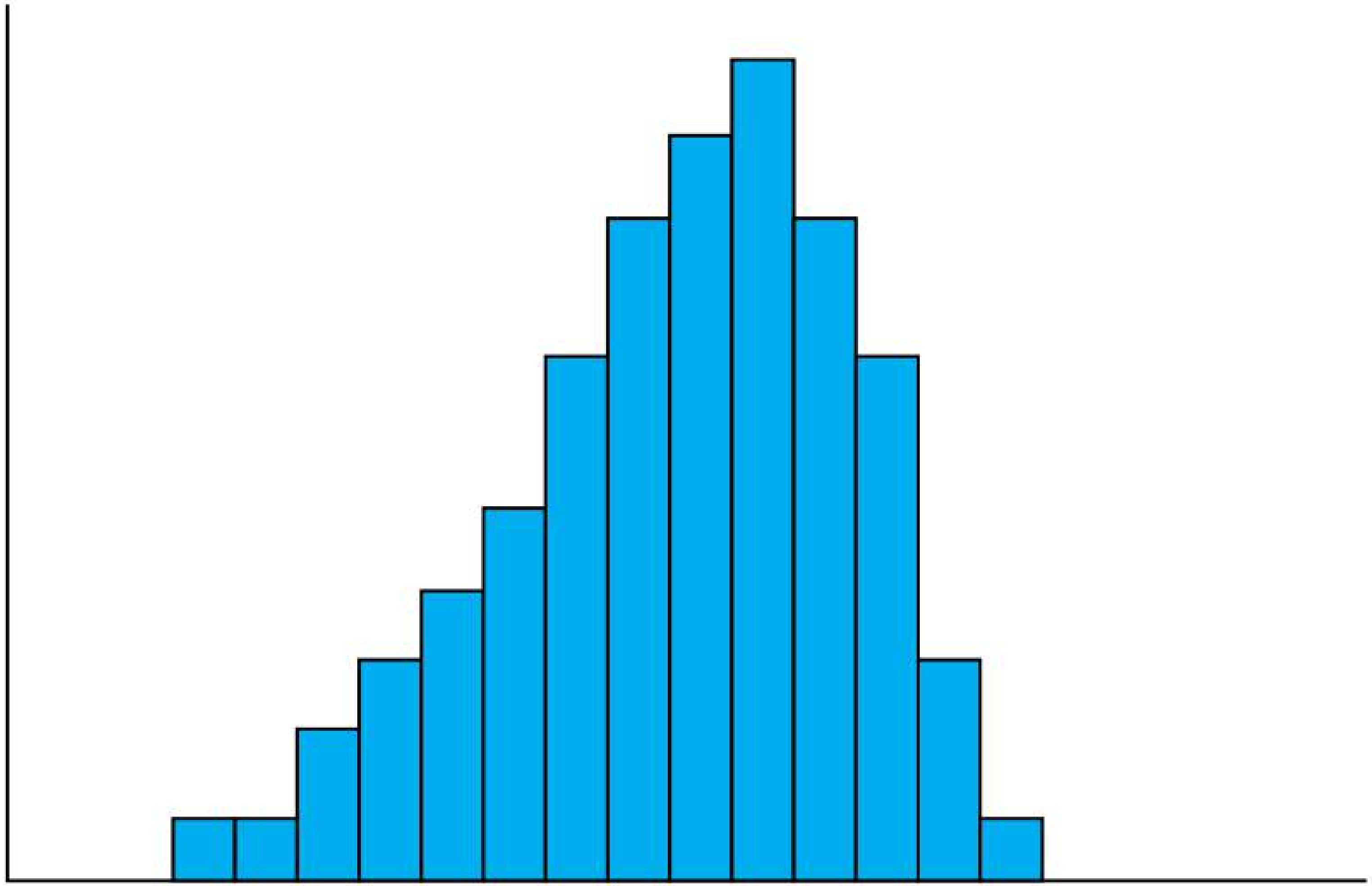


FIGURE 2.10 *Histogram of a data set skewed to the left.*

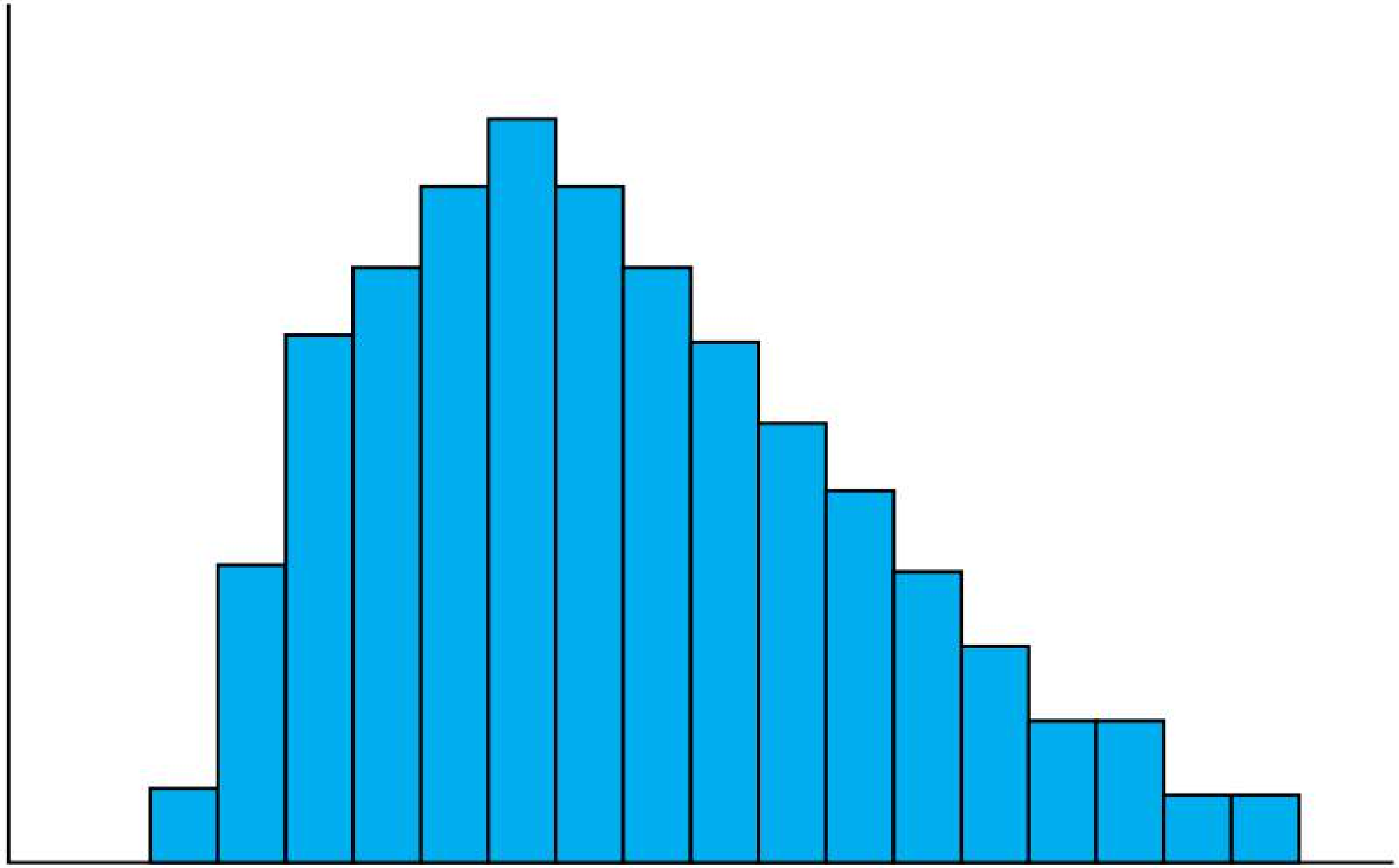


FIGURE 2.11 *Histogram of a data set skewed to the right.*

The Empirical Rule

If a data set is approximately normal with sample mean $\bar{x}$ and sample standard deviation s, then the following statements are true.

1. Approximately 68 percent of the observations lie within

$$\bar{x} \pm s$$

2. Approximately 95 percent of the observations lie within

$$\bar{x} \pm 2s$$

3. Approximately 99.7 percent of the observations lie within

$$\bar{x} \pm 3s$$

Problem

The scores of 25 students on a history examination are listed on the following stem-and-leaf plot.

9	0, 0, 4
8	3, 4, 4, 6, 6, 9
7	0, 0, 3, 5, 5, 8, 9
6	2, 2, 4, 5, 7
5	0, 3, 5, 8

By standing the stem and leaf plot on its side we can see that the corresponding histogram is approximately normal. Use it to assess the empirical rule.

Solution

A calculation yields that the sample mean and sample standard deviation of the data are

$$\bar{x} = 73.68 \quad \text{and} \quad s = 12.80$$

The empirical rule states that approximately 68 percent of the data values are between $\bar{x} - s = 60.88$ and $\bar{x} + s = 86.48$. Since 17 of the observations actually fall within 60.88 and 86.48, the actual percentage is $100(17/25) = 68$ percent.

Similarly, the empirical rule states that approximately 95 percent of the data are between $\bar{x} - 2s = 48.08$ and $\bar{x} + 2s = 96.28$, whereas, in actuality, 100 percent of the data fall in this range.

Paired data sets and the sample correlation coefficient

We are often concerned with data sets that consist of pairs of values that have some relationship to each other. If each element in such a data set has an x value and a y value, then we represent the ith data point by the pair (x_i, y_i).

For instance, in the data set presented in Table 2.12, x_i represents the score on an intelligence quotient (IQ) test, and y_i represents the annual salary (to the nearest $1000) of the ith chosen worker in a sample of 30 workers from a particular company. Let's effectively display data sets of paired values.

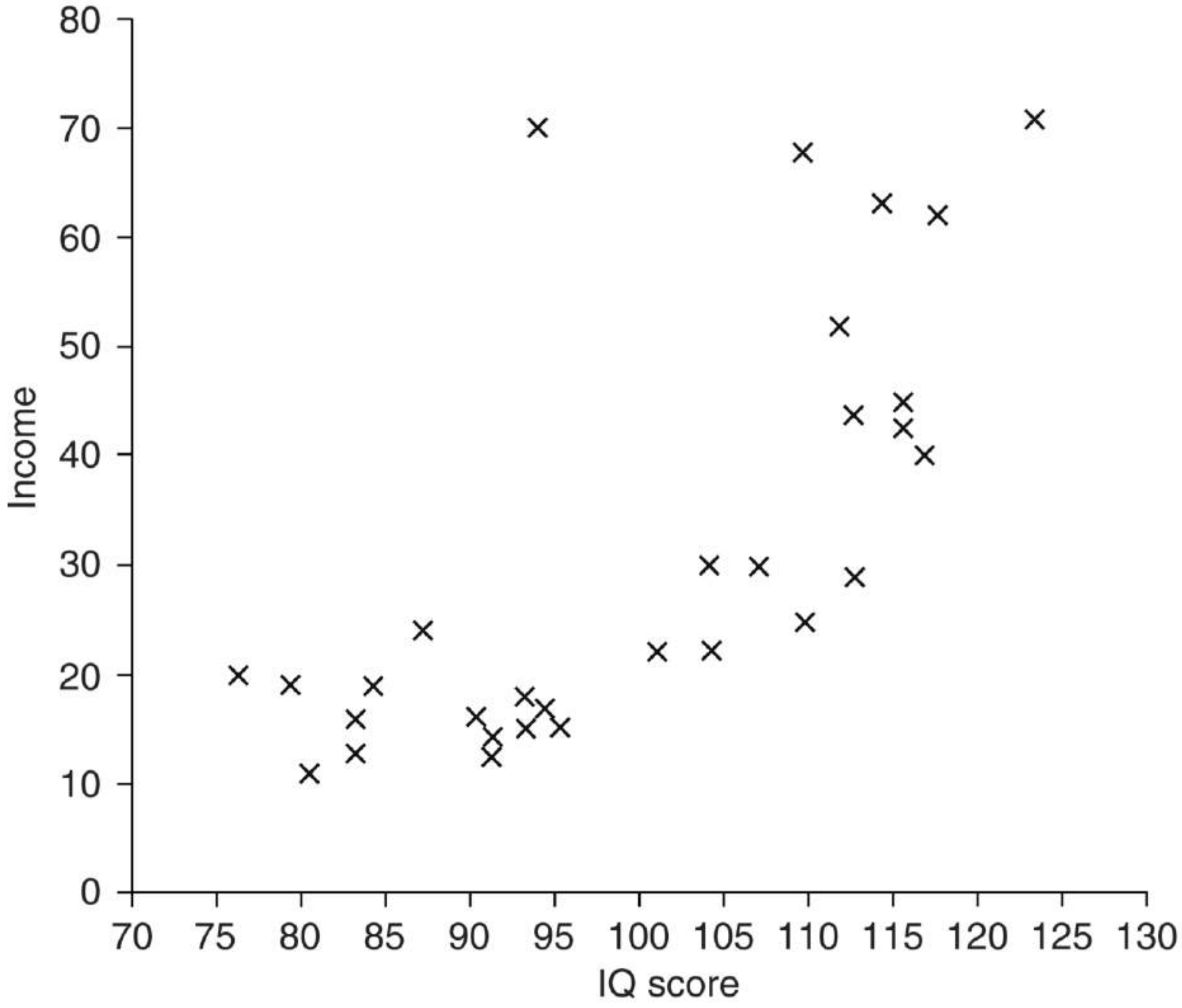
Table 2.12 Salaries versus IQ

Worker i	IQ score x_i	Annual salary y_i (in units of $1000)	Worker i	IQ score x_i	Annual salary y_i (in units of $1000)
1	110	68	16	84	19
2	107	30	17	83	16
3	83	13	18	112	52
4	87	24	19	80	11
5	117	40	20	91	13
6	104	22	21	113	29
7	110	25	22	124	71
8	118	62	23	79	19
9	116	45	24	116	43
10	94	70	25	113	44
11	93	15	26	94	17
12	101	22	27	95	15
13	93	18	28	104	30
14	76	20	29	115	63
15	91	14	30	90	16

A useful way of portraying a data set of paired values is to plot the data on a two dimensional graph, with the x-axis representing the x value of the data and the y-axis representing the y value. Such a plot is called a scatter diagram. Figure 2.13 presents a scatter diagram for the data of Table 2.12.

Scatter diagram of IQ versus income data

FIGURE 2.13



It is clear from Fig. 2.13 that higher incomes appear to go along with higher scores on the IQ test. Note that, not every worker with a high IQ score receives a larger salary than another worker with a lower score (compare worker 5 with worker 29).

The scatter diagram of Fig. 2.13 also appears to have some predictive uses. For instance, suppose we want to predict the salary of a worker whose IQ test score is 120. One way to do this is to “fit by eye” a line to the data set, as is done in Fig. 2.14. Since the y value on the line corresponding to the x value of 120 is about 45, this seems like a reasonable prediction for the annual salary of a worker whose IQ is 120.

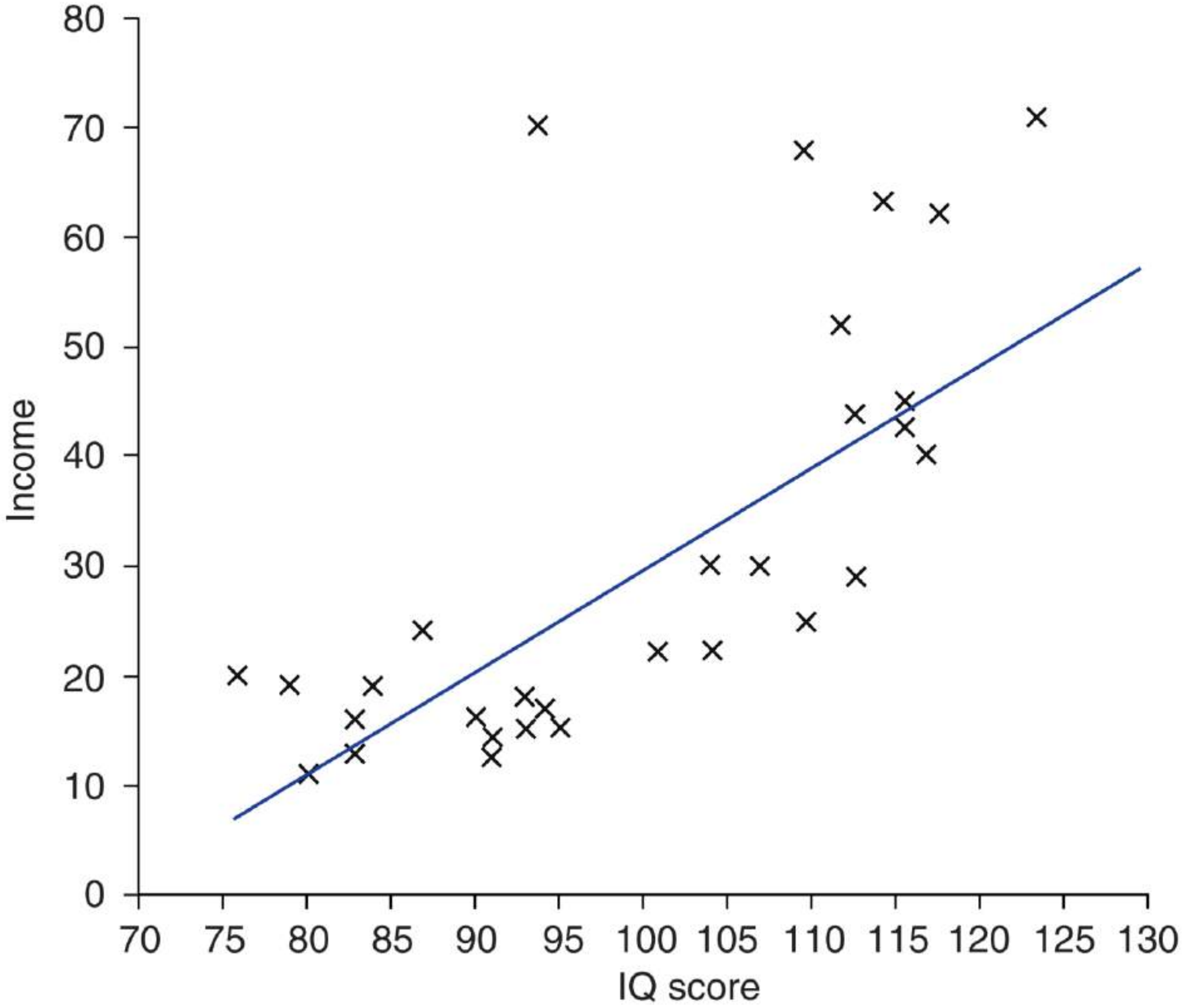


FIGURE 2.14

Scatter diagram for IQ versus income: fitting a straight line by eye.

Sample Correlation Coefficient

Let s_x and s_y denote, respectively, the sample standard deviations of the x values and the y values. The *sample correlation coefficient*, call it r, of the data pairs (x_i, y_i), $i = 1, \ldots, n$ is defined by

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

When $r > 0$ we say that the sample data pairs are *positively correlated*, and when $r < 0$ we say that they are *negatively correlated.*

The following are properties of the sample correlation coefficient.

1. $$-1 \leq r \leq 1$$
2. If for constants a and b, with $b > 0$,

$$y_i = a + bx_i, \qquad i = 1, \ldots, n$$

then $r = 1$.

3. If for constants a and b, with $b < 0$,

$$y_i = a + bx_i, \qquad i = 1, \ldots, n$$

then $r = -1$.

4. If r is the sample correlation coefficient for the data pairs $x_i, y_i, i = 1, \ldots, n$ then it is also the sample correlation coefficient for the data pairs

$$a + bx_i, \quad c + dy_i, \quad i = 1, \ldots, n$$

provided that b and d are both positive or both negative.

Property 1 says that the sample correlation coefficient r is always between −1 and +1.

Property 2 says that r will equal +1 when there is a straight line (also called a linear) relation between the paired data such that large y values are attached to large x values.

Property 3 says that r will equal −1 when the relation is linear and large y values are attached to small x values.

Property 4 states that the value of r is unchanged when a constant is added to each of the x variables (or to each of the y variables) or when each x variable (or each y variable) is multiplied by a positive constant. This property implies that r does not depend on the dimensions chosen to measure the data. This property implies that r does not depend on the dimensions chosen to measure the data. For instance, the sample correlation coefficient between a person's height and weight does not depend on whether the height is measured in feet or in inches or whether the weight is measured in pounds or in kilograms.

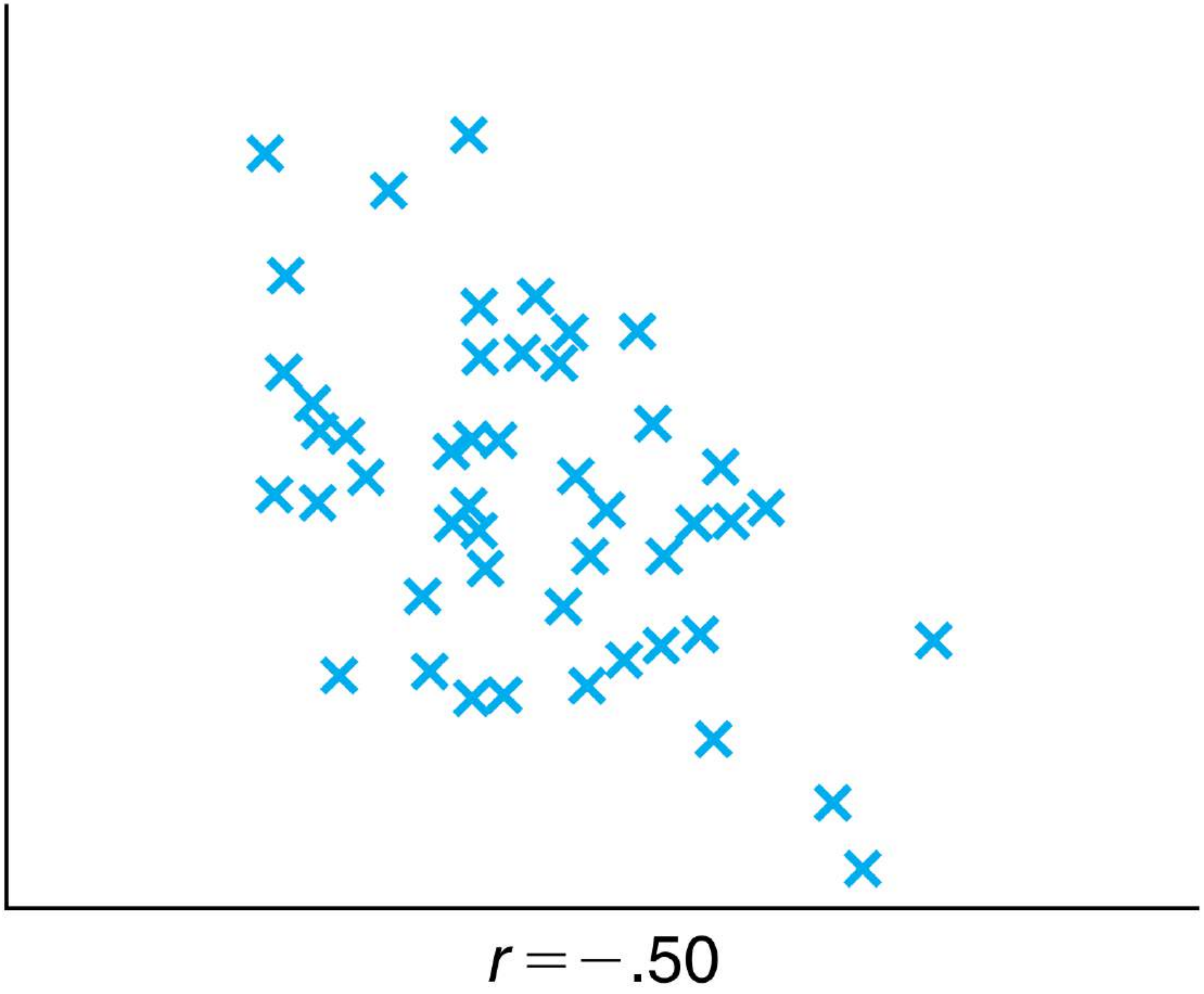
Also, if one of the values in the pair is temperature, then the sample correlation coefficient is the same whether it is measured in Fahrenheit or in Celsius.

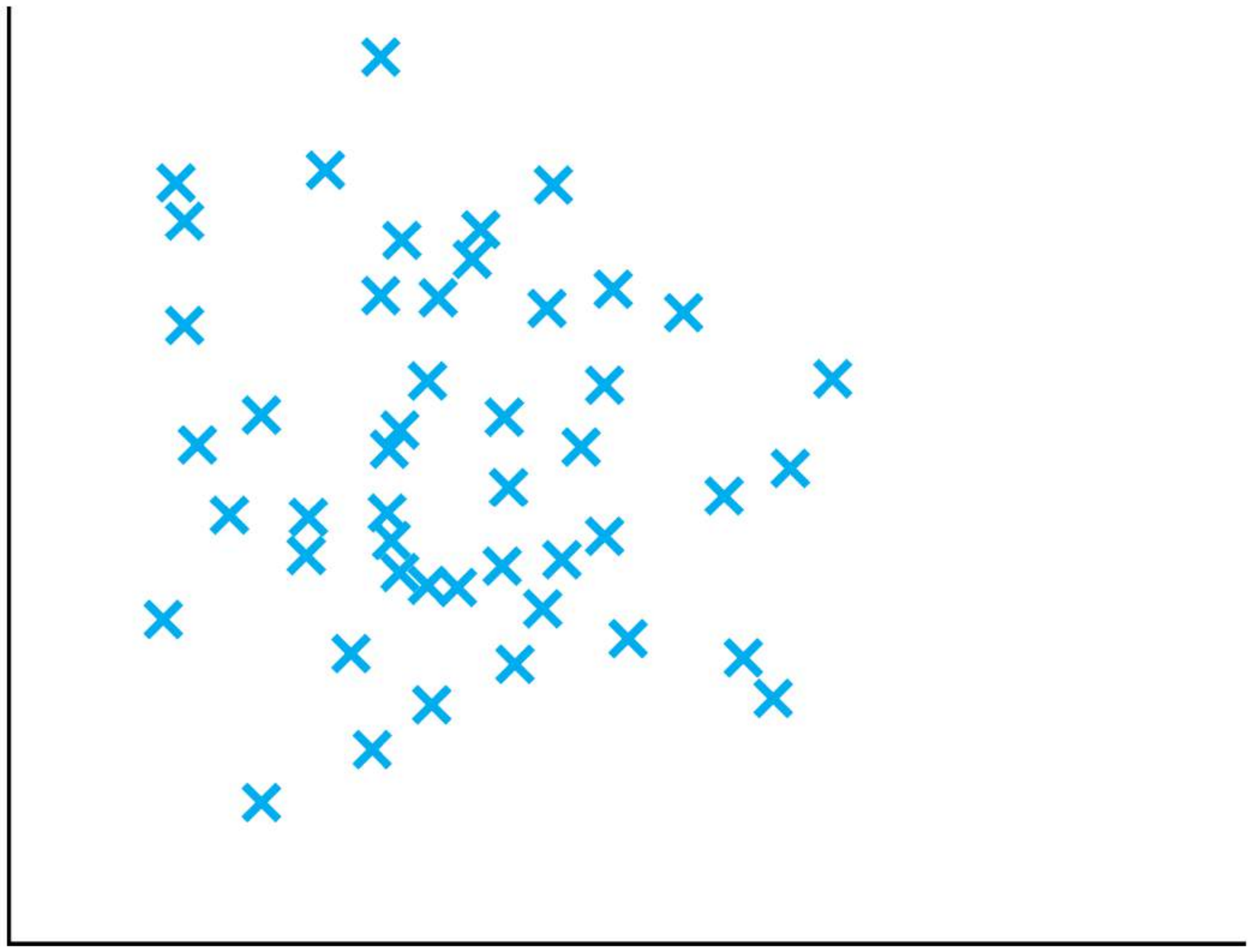
Absolute Value of sample Correlation

The absolute value of the sample correlation coefficient r (that is, $|r|$, its value without regard to its sign) is a measure of the strength of the linear relationship between the x and the y values of a data pair. A value of $|r|$ equal to 1 means that there is a perfect linear relation — that is, a straight line can pass through all the data points (x_i, y_i), $i = 1, \ldots, n$. A value of $|r|$ of around .8 means that the linear relation is relatively strong; although there is no straight line that passes through all of the data points, there is one that is "close" to them all. A value for $|r|$ of around .3 means that the linear relation is relatively weak.

Sign of “r”

The sign of r gives the direction of the relation. It is positive when the linear relation is such that smaller y values tend to go with smaller x values and larger y values with larger x values (and so a straight line approximation points upward), and it is negative when larger y values tend to go with smaller x values and smaller y values with larger x values (and so a straight line approximation points downward). Figure 2.14 displays scatter diagrams for data sets with various values of r.


$r = -.50$



$r = 0$

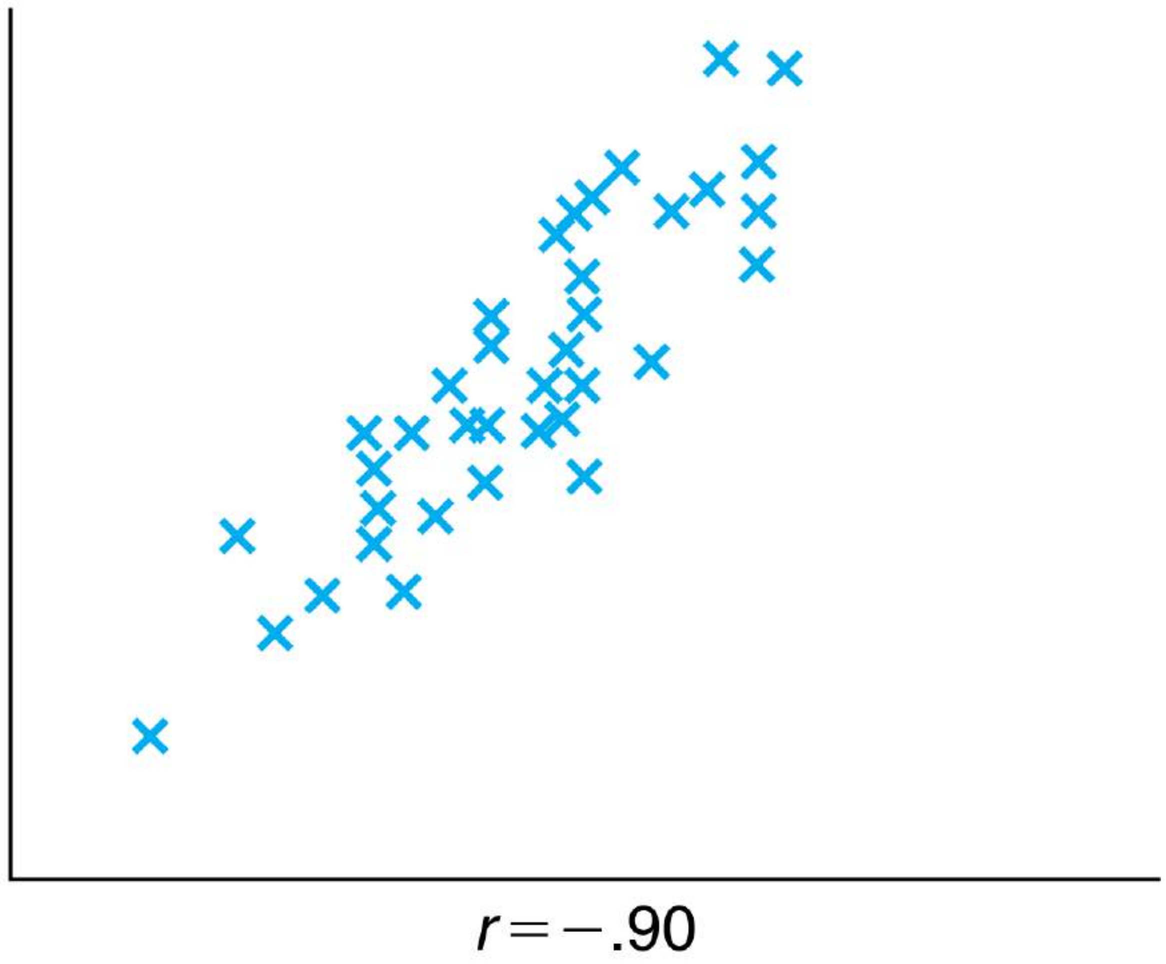


FIGURE 2.14 *Sample correlation coefficients.*

For computational purposes, the following is a convenient formula for the sample correlation coefficient.

Computational Formula for *r*

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}}$$

Problem

The following table gives the U.S. per capita consumption of whole milk (x) and of low-fat milk (y) in three different years.

	Per capita consumption (gallons)		
	1980	1984	1988
Whole milk (x)	17.1	14.7	12.8
Low-fat milk (y)	10.6	11.5	13.2

Source: U.S. Department of Agriculture, *Food Consumption, Prices, and Expenditures.*

Find the sample correlation coefficient r for the given data.

Solution

To make the computation easier, let us first subtract 12.8 from each of the x values and 10.6 from each of the y values. This gives the new set of data pairs:

	i		
	1	**2**	**3**
x_i	4.3	1.9	0
y_i	0	0.9	2.6

Now,

$$\bar{x} = \frac{4.3 + 1.9 + 0}{3} = 2.0667$$

$$\bar{y} = \frac{0 + 0.9 + 2.6}{3} = 1.1667$$

$$\sum_{i=1}^{3} x_i y_i = (1.9)(0.9) = 1.71 \qquad \sum_{i=1}^{3} x_i^2 = (4.3)^2 + (1.9)^2 = 22.10$$

$$\sum_{i=1}^{3} y_i^2 = (0.9)^2 + (2.6)^2 = 7.57$$

Thus,

$$r = \frac{1.71 - 3(2.0667)(1.1667)}{\sqrt{[22.10 - 3(2.0667)^2][7.57 - 3(1.1667)^2]}} = -0.97$$

Therefore, our three data pairs exhibit a very strong negative correlation between consumption of whole and of low-fat milk.

Practice Problems (Page # 41- 51 in Recommended Book)

1, 3, 6, 8, 10, 12, 24, 25

Practice Problem

The following are the number of traffic deaths in a sample of states, both for 2007 and 2008. Plot a scatter diagram and find the sample correlation coefficient for the data pairs.

2007 and 2008 Traffic Fatalities per State

State	2007	2008
WY	149	159
IL	1248	1044
MA	434	318
NJ	724	594
MD	615	560
OR	452	414
WA	568	504
FL	3221	2986
UT	291	271
NH	129	139

Elements of Probability

Sample Space and Events

An experiment is any process that produces an observation or **outcome**.

The **set of all possible outcomes** of an experiment is called the **sample space** and is denoted by S.

Examples:

(a) If the outcome of the experiment is the gender of a newborn child, then

$$S = \{g, b\}$$

where outcome g means that the child is a girl and b that it is a boy.

(b) If the experiment consists of flipping two coins and noting whether they land heads or tails, then

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

The outcome is (H, H) if both coins land heads, (H, T) if the first coin lands heads and the second tails, (T, H) if the first is tails and the second is heads, and (T, T) if both coins land tails.

(c) If the outcome of the experiment is the order of finish in a race among 7 horses having positions 1, 2, 3, 4, 5, 6, 7, then

$$S = \{\text{all orderings of } 1, 2, 3, 4, 5, 6, 7\}$$

The outcome (4, 1, 6, 7, 5, 3, 2) means, for instance, that the number 4 horse comes in first, the number 1 horse comes in second, and so on.

(d) Consider an experiment that consists of rolling two six-sided dice and noting the sides facing up. Calling one of the dice die 1 and the other die 2, we can represent the outcome of this experiment by the pair of upturned values on these dice. If we let (i, j) denote the outcome in which die 1 has value i and die 2 has value j, then the sample space of this experiment is

$$
\begin{aligned}
S = \{&(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4),\\
&(2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2),\\
&(4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),\\
&(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}
\end{aligned}
$$

Any set of outcomes of an experiment is called an **event**. That is, an event is a subset of the sample space. Events will be denoted by the capital letters A, B, C, and so on.

In example (b), if

$$A = \{(H, H), (H, T)\},$$

then A is the event that the first coin lands on heads.

In example (c) if

$$A = \{\text{all outcomes in } S \text{ starting with } 2\}$$

then A is the event that horse number 2 wins the race.

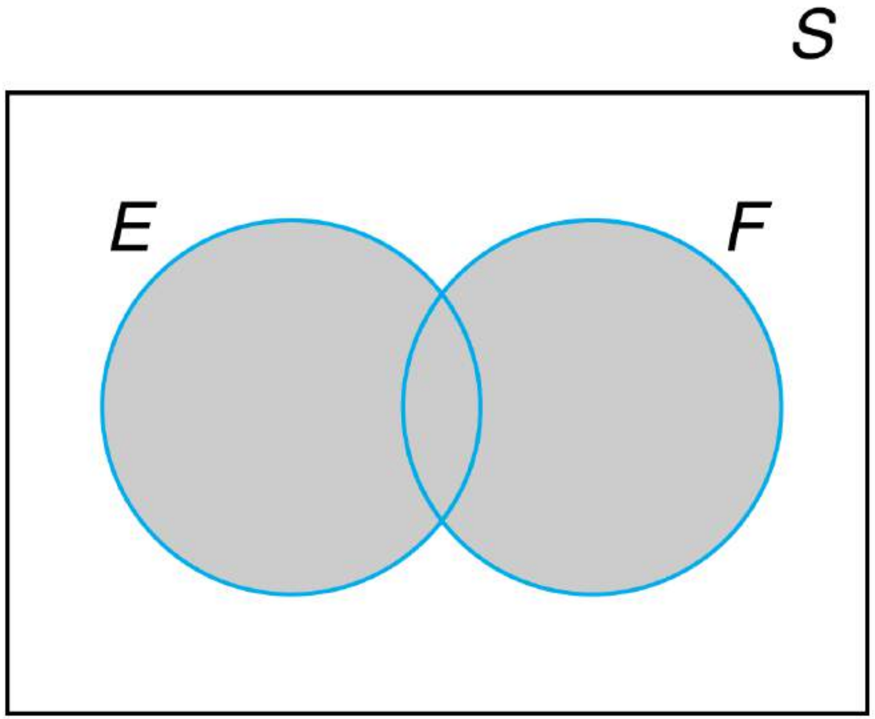
Simple & Compound Events

An event is simple if it consists of exactly one outcome and compound if it consists of more than one outcome.

Occurrence of an Event

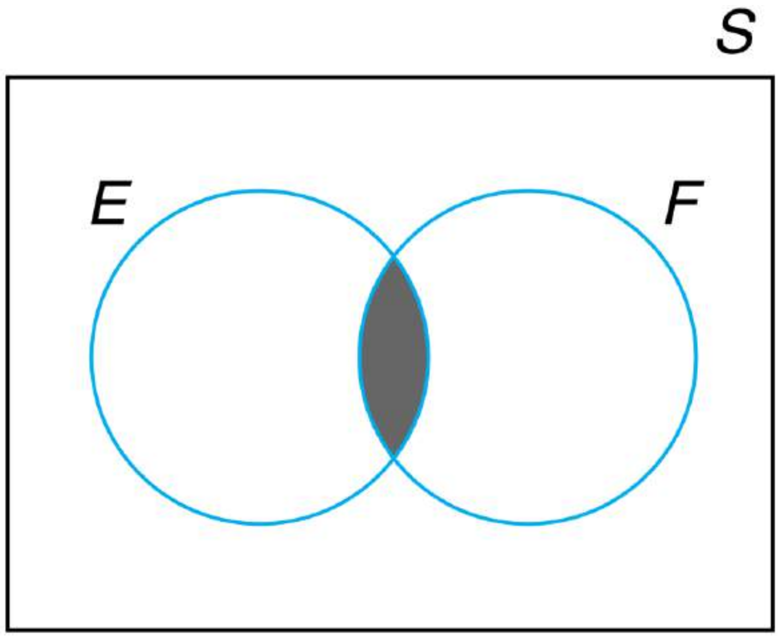
When an experiment is performed, a particular event A is said to occurred if the resulting experimental outcome is contained in A. In general, exactly one simple event will occur, but many compound events can occur simultaneously.

For any two events E and F of a sample space S, we define the new event $E \cup F$, called the union of the events E and F, to consist of all outcomes that are either in E or in F or in both E and F. That is, the event $E \cup F$ will occur if either E or F occurs.



Shaded region: $E \cup F$

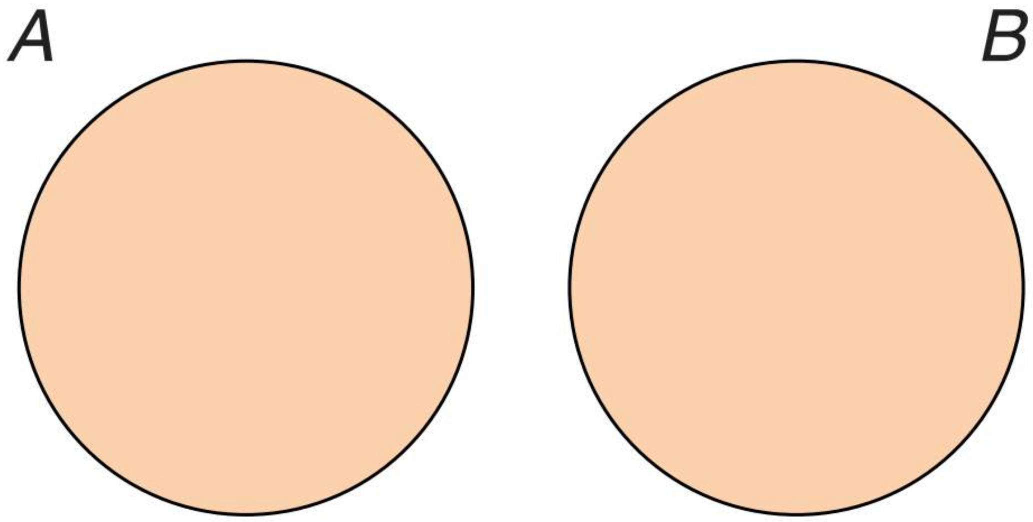
Similarly, for any two events E and F, we may also define the new event EF, sometimes written as E ∩ F, called the intersection of E and F, to consist of all outcomes that are in both E and F. That is, the event EF will occur only if both E and F occur.



shaded region: *EF*

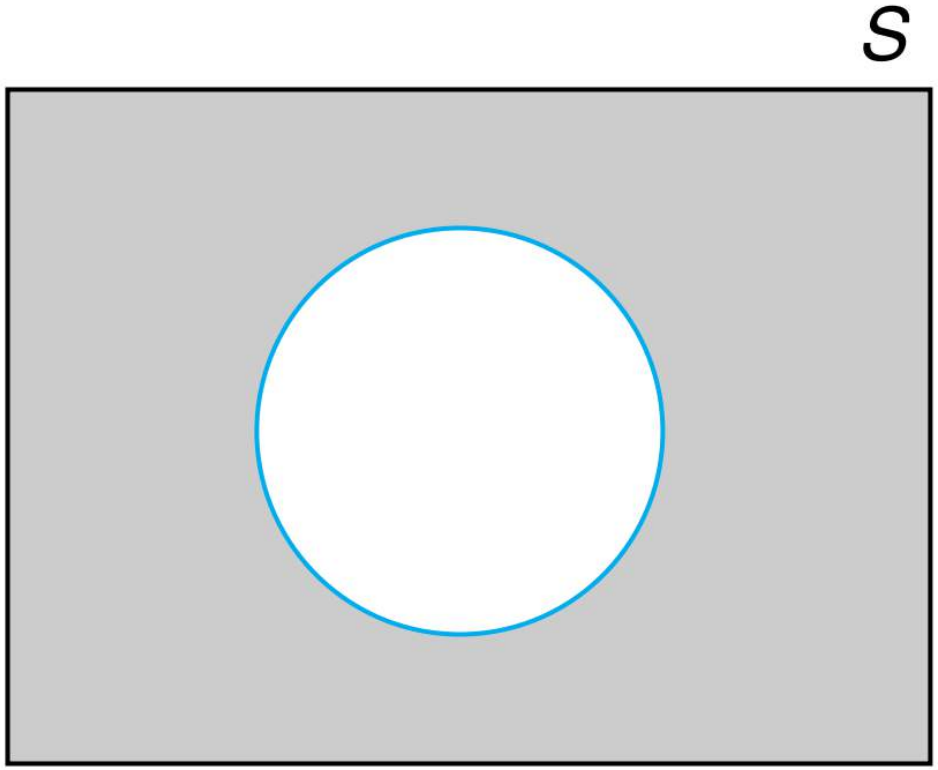
An event consisting of no outcome is called null event, and we denote it by $\emptyset$.

If the intersection of A and B is the null event, then since A and B cannot simultaneously occur, we say that A and B are disjoint, or mutually exclusive.



A and B are disjoint events.

For any event E we define the event E^C, called the complement of E, to consist of all outcomes in the sample space that are not in E. That is, E^C will occur if and only if E does not occur.



Shaded region: E^c

We say that events A, B, and C are disjoint if no two of them can simultaneously occur.

Commutative law	$E \cup F = F \cup E$	$EF = FE$
Associative law	$(E \cup F) \cup G = E \cup (F \cup G)$	$(EF)G = E(FG)$
Distributive law	$(E \cup F)G = EG \cup FG$	$EF \cup G = (E \cup G)(F \cup G)$

DeMorgan's laws

$$(E \cup F)^c = E^c F^c$$

$$(EF)^c = E^c \cup F^c$$

The word probability is a commonly used term that relates to the chance that a particular event will occur when some experiment is performed.

Axioms of Probability

Suppose that for each event E of an experiment having a sample space S there is a number, denoted by P (E), that is in accord with the following three axioms:

Axiom 1: $0 \le P(E) \le 1$

Axiom 2: $P(S) = 1$

Axiom 3: For any sequence of mutually exclusive events $E_1, E_2, \ldots$(that is, events for which $E_iE_j = \emptyset$ when $i \neq j$),

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i), \qquad n = 1, 2, \ldots, \infty$$

We call P(E) the probability of the event E.

These axioms will now be used to prove two simple propositions concerning probabilities. We first note that E and E^c are always mutually exclusive, and since $E \cup E^c = S$, we have by Axioms 2 and 3 that

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

Or equivalently, we have the following:

PROPOSITION 3.4.1

$$P(E^c) = 1 - P(E)$$

PROPOSITION 3.4.2

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

Proof

This proposition is most easily proven by the use of a Venn diagram as shown in Figure 3.4.

We can write $E \cup F = \mathrm{I} \cup \mathrm{II} \cup \mathrm{III}$

$\Rightarrow P(E \cup F) = P(\mathrm{I} \cup \mathrm{II} \cup \mathrm{III})$

As the regions I, II, and III are mutually exclusive, so by third axiom

$$P(E \cup F) = P(\mathrm{I}) + P(\mathrm{II}) + P(\mathrm{III}) \text{ --- (i)}$$

From Venn diagram and using axiom 3 we have

$$P(E) = P(\mathrm{I}) + P(\mathrm{II}) \quad \text{or} \quad P(\mathrm{I}) + P(\mathrm{II}) = P(E) \text{ —— (ii)}$$

$$P(F) = P(\mathrm{II}) + P(\mathrm{III}) \Rightarrow P(\mathrm{III}) = P(F) - P(\mathrm{II}) \text{ —— (iii)}$$

Substituting values from eq (ii) and (iii) in eq (i) we get

$$P(E \cup F) = P(E) + P(F) - P(\mathrm{II})$$

$$\Rightarrow \boxed{P(E \cup F) = P(E) + P(F) - P(EF)}$$

proved

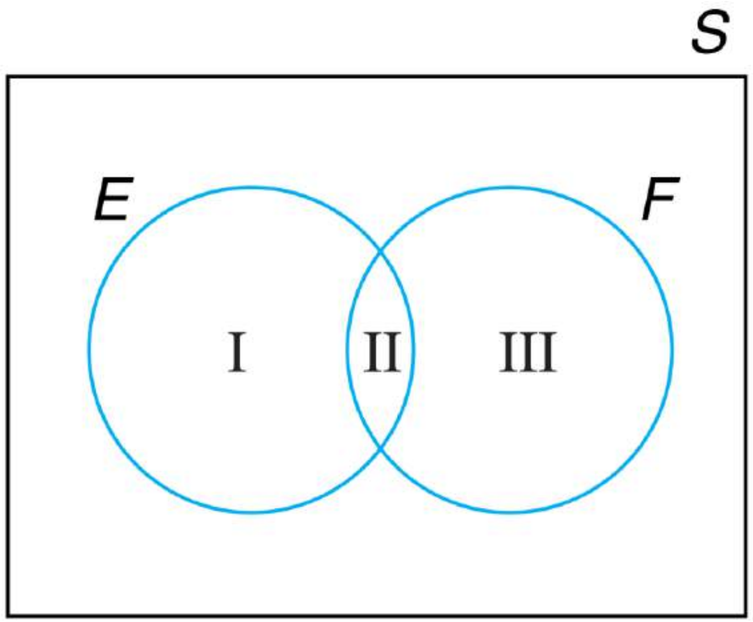


FIGURE 3.4 Venn diagram.

Odds of an Event

The *odds* of an event A is defined by

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

Thus the odds of an event A tells how much more likely it is that A occurs than that it does not occur. For instance, if $P(A) = 3/4$, then $P(A)/(1 - P(A)) = 3$, so the odds are 3. Consequently, it is 3 times as likely that A occurs as it is that it does not.

Sample spaces having equally likely outcomes

For a large number of experiments, it is natural to assume that each point in the sample space is equally likely to occur. That is, for many experiments whose sample space S is a finite set, say $S = \{1, 2, \ldots, N\}$, it is often natural to assume that

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{N\}) = p \quad \text{(say)}$$

As we can write $S = \{1\} \cup \{2\} \ldots \cup \{N\}$

$$\Rightarrow P(S) = P(\{1\} \cup \{2\} \ldots \cup \{N\}) \text{ —— (i)}$$

Since $\{1\}, \{2\}, \ldots, \{N\}$ are mutually exclusive. This implies

$$P(\{1\} \cup \{2\} \ldots \cup \{N\}) = P(\{1\}) + P(\{2\}) + \cdots + P(\{N\})$$

$$= P + P + \ldots + P$$

$$\Rightarrow \boxed{P(\{1\} \cup \{2\} \ldots \cup \{N\}) = NP} \quad \text{also} \quad P(S) = 1$$

Substituting values in (i) we get

$$1 = NP \Rightarrow P = \frac{1}{N}$$

$$\boxed{P(\{1\}) = P(\{2\}) = \dots = P(\{N\}) = \frac{1}{N}}$$

which shows that

$$\boxed{P(\{i\}) = p = 1/N}$$

Consider an event having n number of points denoted by 1, 2, ..., n. We can write $P(E) = P(\{1\} \cup P\{2\} \ldots \cup \{n\})$

$$\Rightarrow P(E) = P(\{1\}) + P(\{2\}) + \ldots + P\{n\})$$

$$= \frac{1}{N} + \frac{1}{N} + \ldots + \frac{1}{N} = \frac{n}{N} = \frac{\text{Number of Points in } E}{\text{Number of Point in } S}$$

Thus we have

$$P(E) = \frac{\text{Number of points in } E}{N}$$

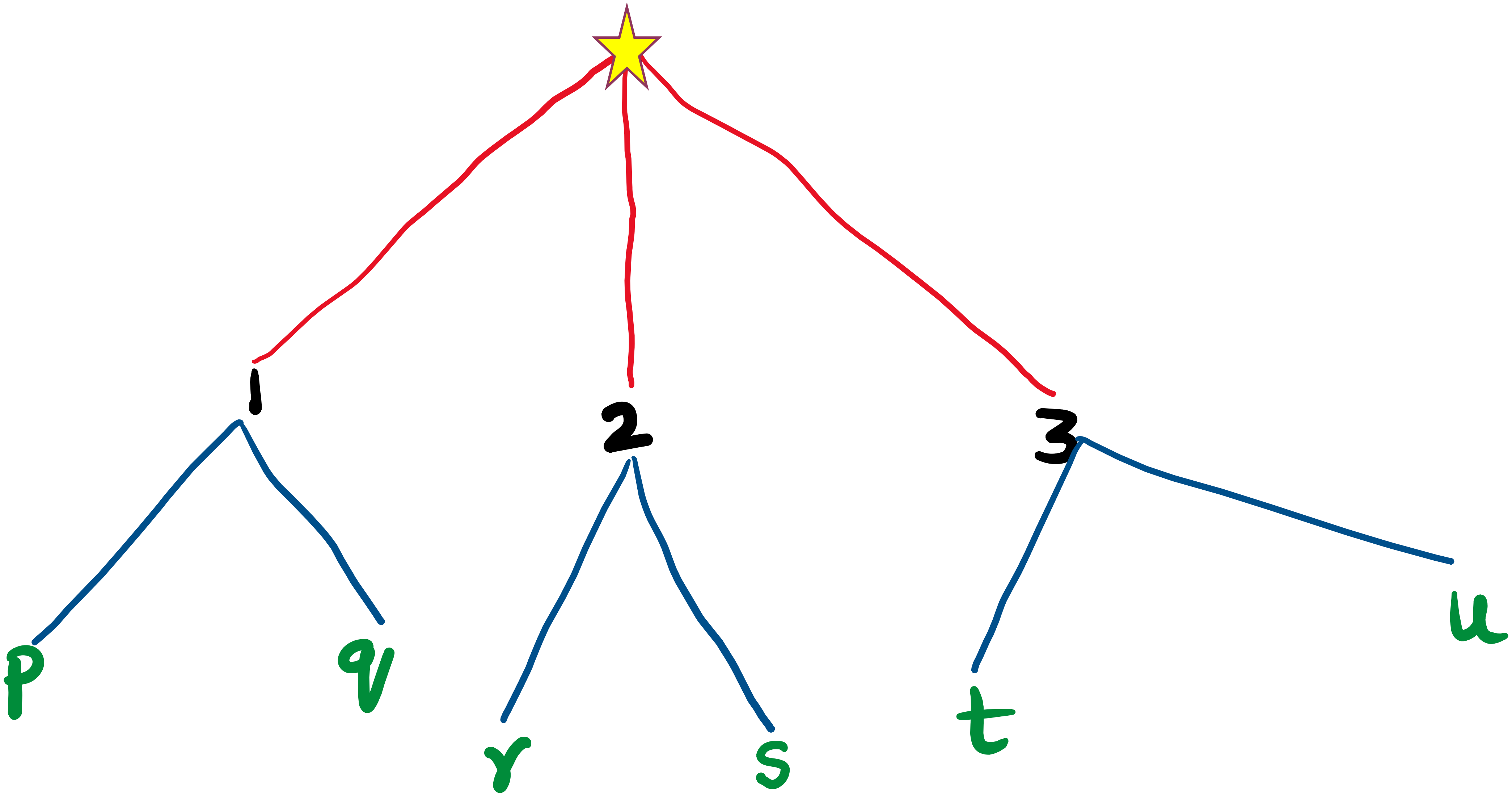
In words, if we assume that each outcome of an experiment is equally likely to occur, then the probability of any event E equals the proportion of points in the sample space that are contained in E.

Thus, to compute probabilities it is often necessary to be able to effectively count the number of different ways that a given event can occur.

The mathematical theory of counting is formally known as *combinatorial analysis*.

The basic principle of counting

Suppose that two experiments are to be performed. Then if experiment 1 can result in any one of m possible outcomes and if, for each outcome of experiment 1, there are n possible outcomes of experiment 2, then together there are mn possible outcomes of the two experiments.



$3 \times 2 = 6$ possible choices. These are 1p, 1q, 2r, 2s, 3t & 3u.

Problem

A football tournament consists of 14 teams, each of which has 11 players. If one team and one of its players are to be selected as team and player of the year, how many different choices are possible?

Solution

Selecting the team can be regarded as the outcome of the first experiment and the subsequent choice of one of its player as the outcome of the second experiment, so there are $14 * 11 = 154$ possibilities.

Problem

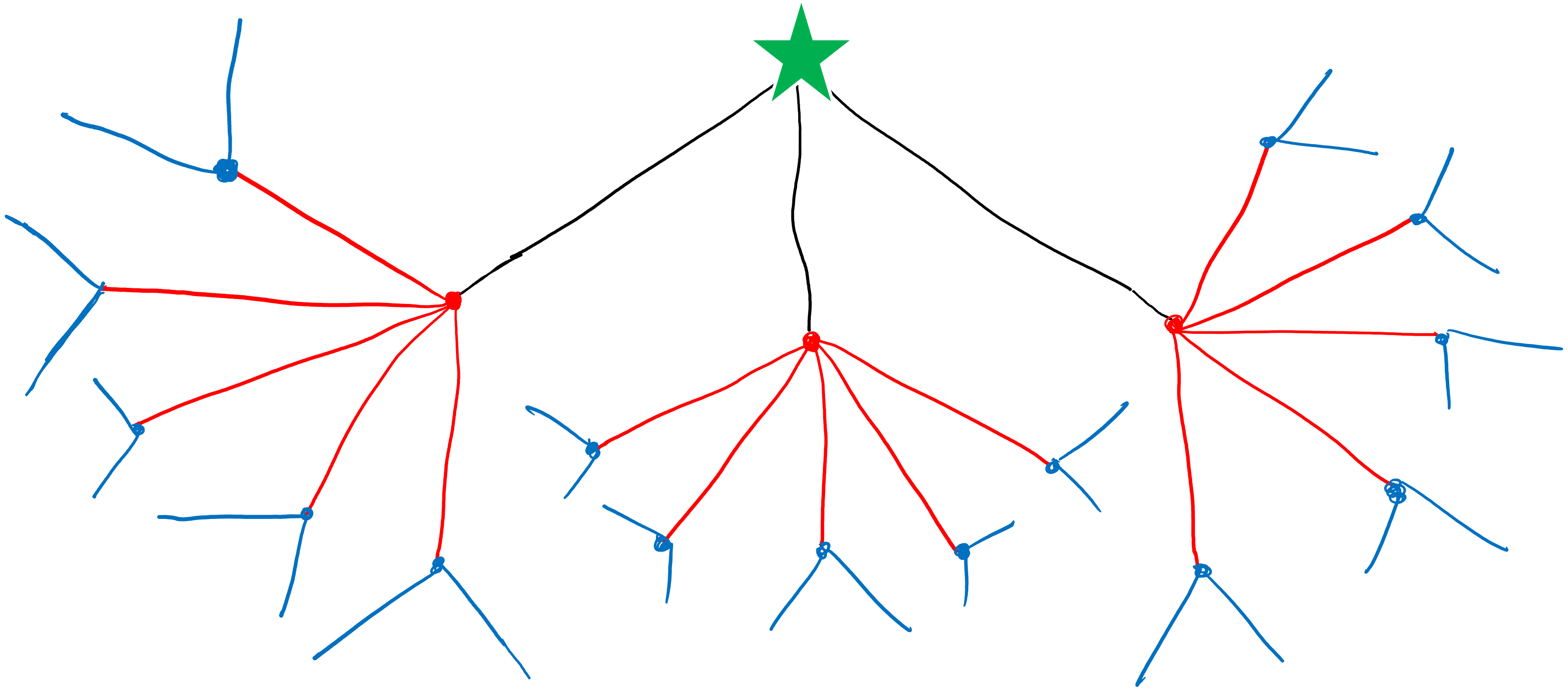
A small community consists of 10 women, each of whom has 3 children. If one woman and one of her children are to be chosen as mother and child of the year, how many different choices are possible?

Solution

By regarding the choice of the woman as the outcome of the first experiment and the subsequent choice of one of her children as the outcome of the second experiment, we see from the basic principle that there are $10 * 3 = 30$ possible choices.

The generalized basic principle of counting

If r experiments that are to be performed are such that the first one may result in any of n_1 possible outcomes; and if, for each of these n_1 possible outcomes, there are n_2 possible outcomes of the second experiment; and if, for each of the possible outcomes of the first two experiments, there are n_3 possible outcomes of the third experiment; and if ..., then there is a total of $n_1 \cdot n_2 \cdots n_r$ possible outcomes of the r experiments.



According to generalized basic rule of Counting we have $3 \times 5 \times 2 = 30$ possible outcomes.

Problem

A college planning committee consists of 3 freshmen, 4 sophomores, 5 juniors, and 2 seniors. A subcommittee of 4, consisting of 1 person from each class, is to be chosen. How many different subcommittees are possible?

Solution

We may regard the choice of a subcommittee as the combined outcome of the four separate experiments of choosing a single representative from each of the classes. It then follows from the generalized version of the basic principle that there are $3 * 4 * 5 * 2 = 120$ possible subcommittees.

Practice Problems

(a) How many different 7-place license plates are possible if the first 3 places are to be occupied by letters and the final 4 by numbers?

(b) How many license plates would be possible if repetition among letters or numbers were prohibited?

How many functions defined on n points are possible if each functional value is either 0 or 1?

Permutations

How many different ordered arrangements of the letters a, b, and c are possible? By direct enumeration we see that there are 6, namely, abc, acb, bac, bca, cab, and cba. Each arrangement is known as a *permutation*. Thus, there are 6 possible permutations of a set of 3 objects. This result could also have been obtained from the basic principle, since the first object in the permutation can be any of the 3, the second object in the permutation can then be chosen from any of the remaining 2, and the third object in the permutation is then the remaining 1. Thus, there are $3 \cdot 2 \cdot 1 = 6$ possible permutations.

General Result. Suppose you have n objects. The number of permutations of these n objects is given by

$$n\,(n-1)\,(n-2)\cdots 3\cdot 2\cdot 1 = n!$$

Problem

How many different batting orders are possible for a baseball team consisting of 9 players?

Solution

There are $9! = 362{,}880$ possible batting orders.

Problem

Ms. Jones has 10 books that she is going to put on her bookshelf. Of these, 4 are mathematics books, 3 are chemistry books, 2 are history books, and 1 is a language book. Ms. Jones wants to arrange her books so that all the books dealing with the same subject are together on the shelf. How many different arrangements are possible?

Solution

There are 4! 3! 2! 1! arrangements such that the mathematics books are first in line, then the chemistry books, then the history books, and then the language book. Furthermore, there are 4! possible orderings of the subjects, the desired answer is 4! 4! 3! 2! 1! = 6912.

General Result. Suppose you have n objects. The number of different permutations of these n objects of which n_1 are alike, n_2 are alike,..., n_r are alike is given by

$$\frac{n!}{n_1!n_2!\cdots n_r!}$$

Practice Problem

How many different letter arrangements can be formed from the letters EEPPPR?

Problem

A chess tournament has 10 competitors, of which 4 are Russian, 3 are from the United States, 2 are from Great Britain, and 1 is from Brazil. If the tournament result lists just the nationalities of the players in the order in which they placed, how many outcomes are possible?

Solution

There are $\dfrac{10!}{4!\,3!\,2!\,1!} = 12{,}600$ possible outcomes.

Problem

How many different signals, each consisting of 9 flags hung in a line, can be made from a set of 4 white flags, 3 red flags, and 2 blue flags if all flags of the same color are identical?

Solution

There are $\dfrac{9!}{4!\,3!\,2!} = 1260$ different signals.

Combinations

We are often interested in determining the number of different groups of r objects that could be formed from a total of n objects. For instance, how many different groups of **3** could be selected from the **5** items A, B, C, D, and E?

To answer this question, reason as follows: Since there are 5 ways to select the initial item, 4 ways to then select the next item, and 3 ways to select the final item, there are thus $5 \cdot 4 \cdot 3$ ways of selecting the group of 3 when the order in which the items are selected is relevant. However, since every group of 3—say, the group consisting of items A, B, and C—will be counted 6 times (that is, all of the permutations ABC, ACB, BAC, BCA, CAB, and CBA will be counted when the order of selection is relevant), it follows that the total number of groups that can be formed is

$$\frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$$

In general, as $n(n - 1) \cdots (n - r + 1)$ represents the number of different ways that a group of r items could be selected from n items when the order of selection is relevant, and as each group of r items will be counted $r!$ times in this count, it follows that the number of different groups of r items that could be formed from a set of n items is

$$\frac{n(n - 1) \cdots (n - r + 1)}{r!} = \frac{n!}{(n - r)!\, r!}$$

Summary

Notation and terminology

We define $\binom{n}{r}$, for $r \leq n$, by

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

and say that $\binom{n}{r}$ represents the number of possible combinations of n objects taken r at a time.†

†By convention, 0! is defined to be 1. Thus, $\binom{n}{0} = \binom{n}{n} = 1$. We also take $\binom{n}{i}$ to be equal to 0 when either $i < 0$ or $i > n$.

Thus, $\binom{n}{r}$ represents the number of different groups of size r that could be selected from a set of n objects when the order of selection is not considered relevant.

Equivalently, $\binom{n}{r}$ is the number of subsets of size r that can be chosen from a set of size n.

Problem

A committee of 3 is to be formed from a group of 20 people. How many different committees are possible?

Solution

There are $\binom{20}{3} = \dfrac{20 \cdot 19 \cdot 18}{3 \cdot 2 \cdot 1} = 1140$ possible committees.

The binomial theorem

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

Problem Expand $(x + y)^3$.

Solution

$$
\begin{aligned}
(x + y)^3 &= \binom{3}{0} x^0 y^3 + \binom{3}{1} x^1 y^2 + \binom{3}{2} x^2 y^1 + \binom{3}{3} x^3 y^0 \\
&= y^3 + 3xy^2 + 3x^2 y + x^3
\end{aligned}
$$

Problem Expand $(x+y)^5$.

Solution $$(x+y) = \sum_{k=0}^{5} \binom{5}{k} x^k y^{5-k}$$

$$\Rightarrow (x+y)^5 = \binom{5}{0} x^0 y^5 + \binom{5}{1} x^1 y^4 + \binom{5}{2} x^2 y^3 + \binom{5}{3} x^3 y^2 + \binom{5}{4} x^4 y^1 + \binom{5}{5} x^5 y^0$$

$$\Rightarrow \boxed{(x+y)^5 = y^5 + 5xy^4 + 10x^2y^3 + 10x^3y^2 + 5x^4y^1 + x^5}$$ Ans

Pascal's Triangle

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

1 6 15 20 15 6 1

Problem

Two balls are "randomly drawn" from a bowl containing 6 white and 5 black balls. What is the probability that one of the drawn balls is white and the other black?

Solution

If we regard the order in which the balls are selected as being significant, then as the first drawn ball may be any of the 11 and the second any of the remaining 10, it follows that the sample space consists of $11 \cdot 10 = 110$ points. Furthermore, there are $6 \cdot 5 = 30$ ways in which the first ball selected is white and the second black, and similarly there are $5 \cdot 6 = 30$ ways in which the first ball is black and the second white.

Thus the probability that one of the drawn balls is white and the other black is

$$\frac{30 + 30}{110} = \frac{6}{11}$$

Problem

A class in probability theory consists of 6 boys and 4 girls. An exam is given and the students are ranked according to their performance. If no two students obtain the same score then (a) how many different rankings are possible? (b) If all rankings are considered equally likely, what is the probability that boys receive the top 4 scores (c) what is the probability that girls receive the top 4 scores?

Solution

(a) Because each ranking corresponds to a particular ordered arrangement of the 10 people, we see the answer to this part is $10! = 3{,}628{,}800$.

(b) In the top 4 positions 4 boys can be place in 6.5.4.3 ways. In the remaining 6 positions 2 boys and 4 girls can be place in 6! ways, it follows from basic principle there are 6.5.4.3*6!= possible rankings in which the boys receive top 4 scores. Hence the desired probability is

$$\frac{6.5.4.3*6!}{10!} = 0.071$$

(c) Because there are 4! possible rankings of the girls among themselves and 6! possible rankings of the boys among themselves, it follows from the basic principle that there are (6!)(4!) = (720)(24) = 17, 280 possible rankings in which the girls receive the top 4 scores. Hence, the desired probability is

$$\frac{6!4!}{10!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{10 \cdot 9 \cdot 8 \cdot 7} = \frac{1}{210} = 0.005$$

Problem

A committee of size 5 is to be selected from a group of 6 men and 9 women. If the selection is made randomly, what is the probability that the committee consists of 3 men and 2 women?

SOLUTION Let us assume that "randomly selected" means that each of the $\binom{15}{5}$ possible combinations is equally likely to be selected. Hence, since there are $\binom{6}{3}$ possible choices of 3 men and $\binom{9}{2}$ possible choices of 2 women, it follows that the desired probability is given by

$$\frac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}} = \frac{240}{1001} \quad \blacksquare$$

Problem

If n people are present in a room, what is the probability that no two of them celebrate their birthday on the same day of the year?

Solution

Because each person can celebrate his or her birthday on any one of 365 days, there are a total of 365^n possible outcomes. Furthermore, there are $(365)(364)(363)\cdot(365-n+1)$ possible outcomes that result in no two of the people having the same birthday. This is so because the first person could have any one of 365 birthdays, the next person any of the remaining 364 days, the next any of the remaining 363, and so on. Hence, assuming that each outcome is equally likely, we see that the desired probability is

$$\frac{(365)(364)(363)\cdots(365-n+1)}{(365)^n}$$

Problem

The English alphabet has 5 vowels and 21 consonants. How many words with two different vowels and 2 different consonants can be formed from the alphabet?

Solution

Given : 5 vowels and 21 consonants.

select : 2 vowels and 2 consonants.

We can choose 2 vowels from 5 in $\binom{5}{2}$ways = 10 ways.

We can choose 2 consonants from 21 in $\binom{21}{2}$ ways = 210 ways.

Now these selected 4 letters can be arranged in 4! ways = 24 ways.

So by basic principle of counting,

Total number of ways(words) = 10.210.24 = 50400.

Conditional Probabilities

Suppose that one rolls a pair of dice. The sample space S of this experiment can be taken to be the following set of 36 outcomes.

$$S = \{(i, j), i = 1, 2, 3, 4, 5, 6, j = 1, 2, 3, 4, 5, 6\}$$

Suppose further that we observe that the first die lands on side 3. Then, given this information, what is the probability that the sum of the two dice equals 8?

So let E is the event that the sum of the dice is 8 and F is the event that the first die is a 3, then the probability of E given that F has occurred is called the conditional probability of E given that F has occurred, and is denoted by **P(E | F).**

$F = \{$first die is a 3$\}$

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$E = \{$Sum of the dice equals 8$\}$

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Note that above Equation is well defined only when $P(F) > 0$ and hence $P(E|F)$ is defined only when $P(F) > 0$.

Problem

A bin contains 5 defective (that immediately fail when put in use), 10 partially defective (that fail after a couple of hours of use), and 25 acceptable transistors. A transistor is chosen at random from the bin and put into use. If it does not immediately fail, what is the probability it is acceptable?

Solution

Since the transistor did not immediately fail, we know that it is not one of the 5 defectives and so the desired probability is:

$$P\{\text{acceptable} \mid \text{not defective}\}$$

$$= \frac{P\{\text{acceptable, not defective}\}}{P\{\text{not defective}\}}$$

$$= \frac{P\{\text{acceptable}\}}{P\{\text{not defective}\}}$$

where the last equality follows since the transistor will be both acceptable and not defective if it is acceptable.

$$P\{\text{acceptable}|\text{not defective}\} = \frac{25/40}{35/40} = 5/7$$

Problem

The organization that Jones works for is running a father–son dinner for those employees having at least one son. Each of these employees is invited to attend along with his youngest son. If Jones is known to have two children, what is the conditional probability that they are both boys given that he is invited to the dinner? Assume that the sample space S is given by S = {(b, b), (b, g), (g, b), (g, g)} and all outcomes are equally likely [(b, g) means, for instance, that the younger child is a boy and the older child is a girl].

SOLUTION The knowledge that Jones has been invited to the dinner is equivalent to knowing that he has at least one son. Hence, letting B denote the event that both children are boys, and A the event that at least one of them is a boy, we have that the desired probability $P(B|A)$ is given by

$$\begin{aligned} P(B|A) &= \frac{P(BA)}{P(A)} \\ &= \frac{P(\{(b, b)\})}{P(\{(b, b), (b, g), (g, b)\})} \\ &= \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \end{aligned}$$

Problem

Ninety percent of flights depart on time. Eighty percent of flights arrive on time. Seventy-five percent of flights depart on time and arrive on time.

(a) Jhon is meeting Ana's flight, which departed on time. What is the probability that Ana will arrive on time?

(b) Jhon has met Ana, and she arrived on time. What is the probability that her flight departed on time?

Solution. Denote the events,

$$\begin{aligned} A &= \{\text{arriving on time}\}, \\ D &= \{\text{departing on time}\}. \end{aligned}$$

We have:

$$\boldsymbol{P}\{A\} = 0.8, \quad \boldsymbol{P}\{D\} = 0.9, \quad \boldsymbol{P}\{A \cap D\} = 0.75.$$

(a) $\boldsymbol{P}\{A \mid D\} = \dfrac{\boldsymbol{P}\{A \cap D\}}{\boldsymbol{P}\{D\}} = \dfrac{0.75}{0.9} = \underline{0.8333}.$

(b) $\boldsymbol{P}\{D \mid A\} = \dfrac{\boldsymbol{P}\{A \cap D\}}{\boldsymbol{P}\{A\}} = \dfrac{0.75}{0.8} = \underline{0.9375}.$

Problem

The concept of conditional probability has countless uses in both industrial and biomedical applications. Consider an industrial process in the textile industry in which strips of a particular type of cloth are being produced. These strips can be defective in two ways, length and nature of texture. For the case of the latter, the process of identification is very complicated. It is known from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests. If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

Solution

Consider the events

$$L\text{: length defective,} \qquad T\text{: texture defective.}$$

Given that the strip is length defective, the probability that this strip is texture defective is given by

$$P(T|L) = \frac{P(T \cap L)}{P(L)} = \frac{0.008}{0.1} = 0.08.$$

$$P(A|B) = \frac{P(AB)}{P(B)} \Rightarrow \boxed{P(AB) = P(A|B)\,P(B)}$$

$$P(B|A) = \frac{P(BA)}{P(A)}$$

$$\Rightarrow P(B|A) = \frac{P(AB)}{P(A)} \Rightarrow \boxed{P(AB) = P(B|A)\,P(A)}$$

Note:

1) $P(AB) = P(A|B)\,P(B)$

2) $P(AB) = P(B|A)\,P(A)$

Problem

Suppose that two balls are to be selected at random, without replacement, from a box containing r red balls and b blue balls. What is the probability that the first ball will be red and the second ball will be blue?

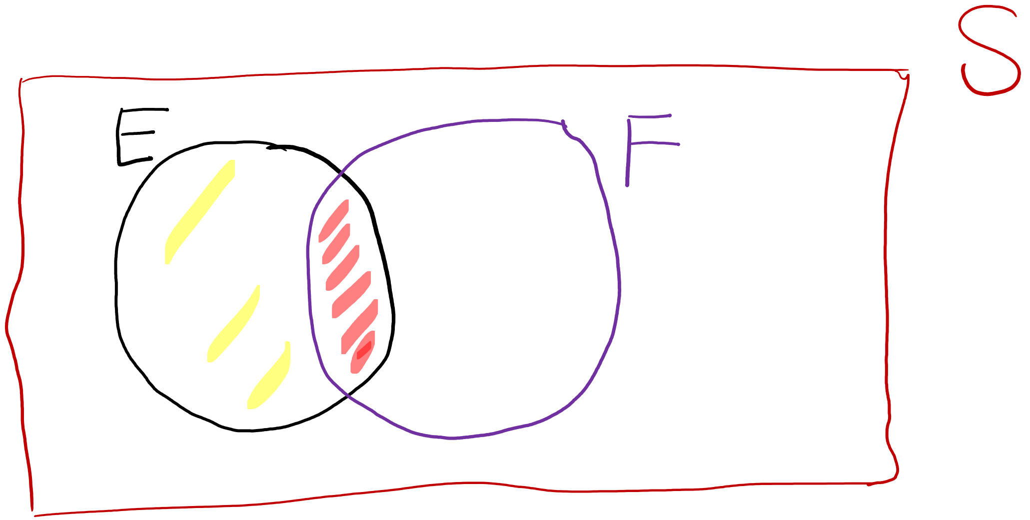
Solution

Let A be the event that the first ball is red, and let B be the event that the second ball is blue. Obviously, $P(A) = r/(r + b)$. Furthermore, if the event A has occurred, then one red ball has been removed from the box on the first draw. Therefore, the probability of obtaining a blue ball on the second draw will be

$$\mathrm{P}(B|A) = \frac{b}{r + b - 1}$$

It follows that

$$P(AB) = \frac{b}{r+b-1} \cdot \frac{r}{r+b}$$

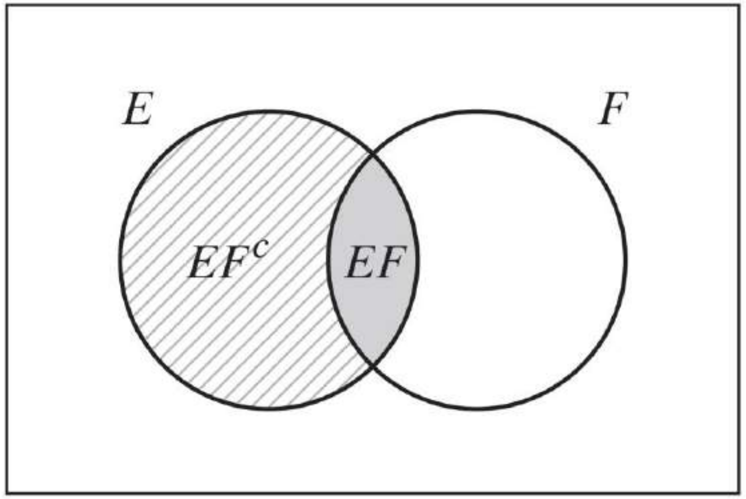


E = Yellow Region ∪ Red Region

$\Rightarrow E = EF^c \cup EF$

Let E and F be events. We may express E as

$$E = EF \cup EF^c$$



$E = EF \cup EF^c$. EF = Shaded Area; EF^c = Striped Area.

As EF and EF^c are clearly mutually exclusive, we have by Axiom 3 that

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \end{aligned}$$

$$\Rightarrow \boxed{P(E) = P(E|F)P(F) + P(E|F^c)[1 - P(F)]}$$

Problem

A laboratory blood test is 99 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a “false positive” result for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability .01, the test result will imply he or she has the disease.) If .5 percent of the population actually has the disease, what is the probability a person has the disease given that his test result is positive?

Solution

Let D be the event that the tested person has the disease and E the event that his test result is positive. The desired probability P(D|E) is obtained by

$$P(D|E) = \frac{P(DE)}{P(E)}$$

$$= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)}$$

$$P(D|E) = \frac{(.99)(.005)}{(.99)(.005) + (.01)(.995)}$$

$$= .3322$$

Thus, only 33 percent of those persons whose test results are positive actually have the disease.

Suppose that $F_1, F_2, F_3, \ldots, F_n$ are mutually exclusive events such that

$$\bigcup_{i=1}^{n} F_i = S$$

In other words, exactly one of the events $F_1, F_2, F_3, \ldots, F_n$ must occur. By writing

$$E = \bigcup_{i=1}^{n} EF_i$$

and using the fact that the events EF_i, $i = 1, \ldots, n$ are mutually exclusive, we obtain that

$$P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

OR

$$P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + \cdots + P(E|F_n)P(F_n)$$

The Law of Total Probability

Let $F_1, F_2, \ldots, F_n$ be mutually exclusive and exhaustive events. Then for any other event E,

$$P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + \cdots + P(E|F_n)P(F_n)$$
$$= \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

Events $F_1, F_2, \ldots, F_n$ are called exhaustive events if $F_1 \cup F_2 \cup \cdots \cup F_n = S$.

Problem

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

Solution

A: the product is defective,

B_1: the product is made by machine B_1,

B_2: the product is made by machine B_2,

B_3: the product is made by machine B_3.

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3).$$

$$P(B_1)P(A|B_1) = (0.3)(0.02) = 0.006,$$
$$P(B_2)P(A|B_2) = (0.45)(0.03) = 0.0135,$$
$$P(B_3)P(A|B_3) = (0.25)(0.02) = 0.005,$$

and hence

$$P(A) = 0.006 + 0.0135 + 0.005 = 0.0245.$$

Bayes' Formula

Suppose now that E has occurred and we are interested in determining which one of F_j also occurred. We have

$$P(F_j|E) = \frac{P(EF_j)}{P(E)}$$

$$\boxed{P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}}$$

In term of two events A and B we can work for Bayes Rule as follows :

$$P(A|\mathrm{B}) = \frac{P(AB)}{P(B)}$$

$$\Rightarrow P(A|\mathrm{B}) = \frac{P(B|A)P(A)}{P(B)}$$

Note:

Most of the time we don't have P(B) so we can use law of total probability to compute P(B).

Problem

Consider a test that can diagnose kidney cancer. The test correctly detects when a patient has cancer 90% of the time. Also, if a person does not have cancer, the test correctly indicates so 99.9% of the time. Finally, suppose it is known that 1 in every 10,000 individuals has kidney cancer. Find the probability that a patient has kidney cancer, given that the test indicates he does.

Solution

Let C denote the event that a patient has kidney cancer then C^c will denote the event that a patient has no kidney cancer.
Po denote the event that the patient tests positive for kidney cancer then $Po^c = N$ denote the event that the patient tests negative for kidney cancer.

Now we want $P(C|Po)$.

Given data is :

Test correctly detects when a patient has cancer 90% of the time : $P(Po|C) = \frac{90}{100} = 0.9$

If a person does not have cancer, the test correctly indicates so 99.9% of the time:

$$P(N|C^C) = \frac{99.9}{100} = 0.999 \Rightarrow P(Po|C^C) = 1 - 0.999 = 0.001$$

It is known that 1 in every 10,000 individuals has kidney cancer:

$P(C) = \frac{1}{10000} = 0.0001 \Rightarrow P(C^C) = 1 - P(C) = 0.9999.$

By Law of Total Probability

$$\begin{aligned} P(Po) &= P(Po|\mathrm{C})P(C) + P(Po|C^c)P(C^C) \\ &= (0.9 * 0.0001) + (0.001 * 0.9999) \\ &= 0.0010899 \end{aligned}$$

By Bayes Rule,

$$P(C|\mathrm{Po}) = \frac{P(Po|C)P(C)}{P(Po)} = \frac{0.9 * 0.0001}{0.0010899} \approx 0.08$$

This implies that only about 8% of patients that test positive under this particular test actually have kidney cancer.

Practice Problems

1. You ask your neighbor to water a sickly plant while you are on vacation. Without water it will die with probability 0.8; with water it will die with probability 0.15. You are 90 percent certain that your neighbor will remember to water the plant.

(a) What is the probability that the plant will be alive when you return?

(b) If it is dead, what is the probability your neighbor forgot to water it?

---

2. There is a 60 percent chance that the event A will occur. If A does not occur, then there is a 10 percent chance that B will occur. What is the probability that at least one of the events A or B occurs?

Independent Events

E is independent of F if knowledge that F has occurred does not change the probability that E occurs.

Since $P(E|F) = P(EF)/P(F)$, we see that E is independent of F if

$$P(EF) = P(E)P(F)$$

Since this equation is symmetric in E and F, it shows that whenever E is independent of F so is F of E.

Two events E and F are said to be independent if

$$P(EF) = P(E)P(F)$$

Two events E and F that are not independent are said to be dependent.

Problem

Suppose that we roll a pair of fair dice, so each of the 36 possible outcomes is equally likely. Let A denote the event that the first die lands on 3, let B be the event that the sum of the dice is 8, and let C be the event that the sum of the dice is 7.

(a) Are A and B independent?
(b) Are A and C independent?

Solution

(a) Since $A \cap B$ is the event that the first die lands on 3 and the second on 5, we see that

$$P(A \cap B) = P(\{(3,5)\}) = \frac{1}{36}$$

On the other hand,

$$P(A) = P(\{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}) = \frac{6}{36}$$

and

$$P(B) = P(\{(2,6),(3,5),(4,4),(5,3),(6,2)\}) = \frac{5}{36}$$

Therefore, since $1/36 \neq (6/36) \cdot (5/36)$, we see that

$$P(A \cap B) \neq P(A)P(B)$$

and so events A and B are not independent.

Similar solve part (b).

Problem

Toss two coins and observe the outcome. Define these events:

A: Head on the first coin

B: Tail on the second coin

Are events A and B independent?

Solution

$$P(A) = \frac{1}{2}, P(B) = \frac{1}{2}, \text{ and } P(A \cap B) = \frac{1}{4}.$$

Since $P(A)P(B) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$ and $P(A \cap B) = \frac{1}{4}$, we have $P(A)P(B) = P(A \cap B)$ and the two events must be independent.

Independency of Three or More Events

Three events E, F, and G are said to be independent if

$$P(EFG) = P(E)P(F)P(G)$$
$$P(EF) = P(E)P(F)$$
$$P(EG) = P(E)P(G)$$
$$P(FG) = P(F)P(G)$$

Note that if E, F, and G are independent, then E will be independent of any event formed from F and G.

Of course, we may also extend the definition of independence to more than three events. The events $E_1, E_2, \ldots, E_n$ are said to be independent if for every subset $E_{1'}, E_{2'}, \ldots, E_{r'}, r \leq n$ of these events,

$$P(E_{1'}E_{2'} \cdots E_{r'}) = P(E_{1'})P(E_{2'}) \cdots P(E_{r'})$$

Practice Problem

Of three cards, one is painted red on both sides; one is painted black on both sides; and one is painted red on one side and black on the other. A card is randomly chosen and placed on a table. If the side facing up is red, what is the probability that the other side is also red?

Practice Problem

Among employees of a certain firm, 70% know Java, 60% know Python, and 50% know both languages. What portion of programmers

(a) does not know Python?

(b) does not know Python and does not know Java?

(c) knows Java but not Python?

(d) knows Python but not Java?

(e) If someone knows Python, what is the probability that he/she knows Java too?

(f) If someone knows Java, what is the probability that he/she knows Python too?

Problem

(Diagnostics of computer codes): A new computer program consists of two modules. The first module contains an error with probability 0.2. The second module is more complex; it has a probability of 0.4 to contain an error, independently of the first module. An error in the first module alone causes the program to crash with probability 0.5. For the second module, this probability is 0.8. If there are errors in both modules, the program crashes with probability 0.9. Suppose the program crashed. What is the probability of errors in both modules?

Problem

A computer maker receives parts from three suppliers, S1, S2, and S3. Fifty percent come from S1, twenty percent from S2, and thirty percent from S3. Among all the parts supplied by S1, 5% are defective. For S2 and S3, the portion of defective parts is 3% and 6%, respectively.

(a) What is the probability that a random part is defective?

(b) A customer complains that a certain part in her recently purchased computer is defective. What is the probability that it was supplied by S1?

Problems of Quiz # 02

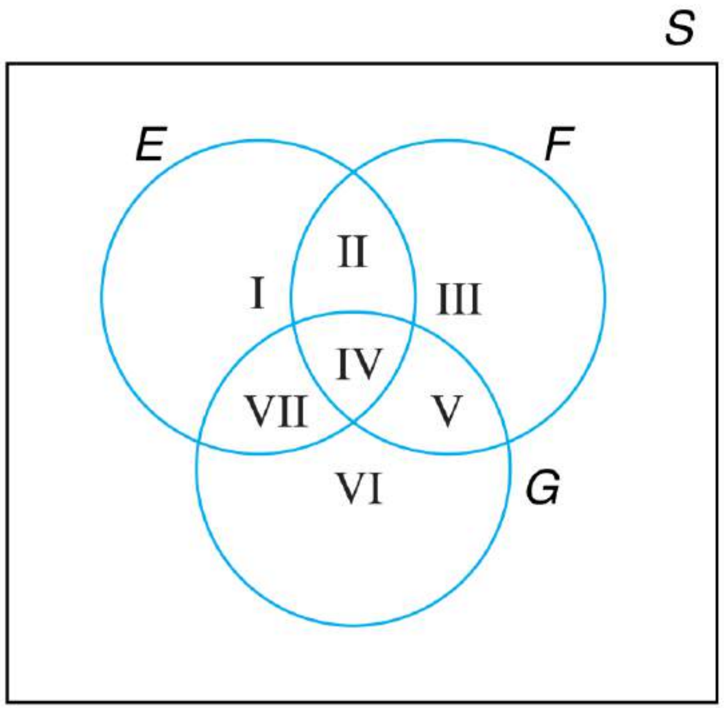
1. Prove that

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

2. A programming class is composed of 10 juniors, 30 seniors, and 10 graduate students. The final grades show that 5 of the juniors, 10 of the seniors, and 5 of the graduate students received an A for the course. If a student is chosen at random from this class and is found to have earned an A, what is the probability that he or she is a senior?

Practice Problem

For the following Venn diagram, describe in terms of E, F, and G the events denoted in the diagram by the Roman numerals I through VII.



Random Variables & Expectation

Random Variables

A **random variable** is a real-valued function whose domain is a sample space.

Random variables are typically denoted by uppercase letters, such as X, Y, and Z. The actual numerical values that a random variable can assume are denoted by lowercase letters, such as x, y, and z.

Mathematically,

$$X : \Omega \longrightarrow \mathbb{R}$$

where Ω represents sample space

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Examples :

Let X denote the random variable that is defined as the sum of two fair dice, then

$P\{X = 2\} = P\{(1, 1)\} = 1/36$

$P\{X = 3\} = P\{(1, 2), (2, 1)\} = 2/36$

$P\{X = 4\} = P\{(1, 3), (2, 2), (3, 1)\} = 3/36$

$P\{X = 5\} = P\{(1, 4), (2, 3), (3, 2), (4, 1)\} = 4/36$

$P\{X = 6\} = P\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} = 5/36$

$P\{X = 7\} = P\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} = 6/36$

$P\{X = 8\} = P\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} = 5/36$

$P\{X = 9\} = P\{(3, 6), (4, 5), (5, 4), (6, 3)\} = 4/36$

$P\{X = 10\} = P\{(4, 6), (5, 5), (6, 4)\} = 3/36$

$P\{X = 11\} = P\{(5, 6), (6, 5)\} = 2/36$

$P\{X = 12\} = P\{(6, 6)\} = 1/36$

In other words, the random variable X can take on any integral value between 2 and 12.

Note:

$$1 = P(S) = P\left(\bigcup_{i=2}^{12}\{X = i\}\right) = \sum_{i=2}^{12} P\{X = i\}$$

Another random variable of possible interest in this experiment is the value of the first die. Letting Y denote this random variable, then Y is equally likely to take on any of the values 1 through 6. That is,

$P\{Y = i\} = 1/6, i = 1, 2, 3, 4, 5, 6.$

Next Example :

Suppose that an individual purchases two electronic components, each of which may be either defective or acceptable. In addition, suppose that the four possible results — (*d*, *d*), (*d*, *a*), (*a*, *d*), (*a*, *a*) — have respective probabilities .09, .21, .21, .49 [where (*d*, *d*) means that both components are defective, (*d*, *a*) that the first component is defective and the second acceptable, and so on]. If we let X denote the number of acceptable components obtained in the purchase, then X is a random variable taking on one of the values 0, 1, 2 with respective probabilities

$$P\{X = 0\} = .09$$
$$P\{X = 1\} = .42$$
$$P\{X = 2\} = .49$$

If we were mainly concerned with whether there was at least one acceptable component, we could define the random variable I by

$$I = \begin{cases} 1 & \text{if } X = 1 \text{ or } 2 \\ 0 & \text{if } X = 0 \end{cases}$$

If A denotes the event that at least one acceptable component is obtained, then the random variable I is called the indicator random variable for the event A, since I will equal 1 or 0 depending upon whether A occurs. The probabilities attached to the possible values of I are

$$P\{I = 1\} = .91$$
$$P\{I = 0\} = .09$$

Discrete and Continuous Random Variables

In the two foregoing examples, the random variables of interest took on a finite number of possible values. Random variables whose set of possible values can be written either as a finite sequence $x_1, \ldots, x_n$, or as an infinite sequence $x_1, \ldots$ are said to be *discrete*. For instance, a random variable whose set of possible values is the set of nonnegative integers is a discrete random variable. However, there also exist random variables that take on a continuum of possible values. These are known as *continuous* random variables. In other words a random variable X is said to be continuous if it can take on the infinite number of possible values associated with intervals of real numbers. One example is the random variable denoting the lifetime of a car, when the car's lifetime is assumed to take on any value in some interval (a, b).

The Cumulative Distribution Function (CDF)

The cumulative distribution function, or more simply the distribution function, F of the random variable X is defined for any real number x by

$$F(x) = P\{X \leq x\}$$

That is, F (x) is the probability that the random variable X takes on a value that is less than or equal to x.

Notation: We will use the notation $X \sim F$ to signify that F is the distribution function of X .

Suppose we wanted to compute $P\{a < X \le b\}$. This can be accomplished by first noting that the event $\{X \le b\}$ can be expressed as the union of the two mutually exclusive events $\{X \le a\}$ and $\{a < X \le b\}$. Therefore, applying Axiom 3, we obtain that

$$P\{X \le b\} = P\{X \le a\} + P\{a < X \le b\}$$

$$\Rightarrow \quad P\{a < X \le b\} = P\{X \le b\} - P\{X \le a\}$$

or

$$P\{a < X \le b\} = F(b) - F(a)$$

Problem

Suppose the random variable X has distribution function

$$F(x) = \begin{cases} 0 & x \le 0 \\ 1 - \exp\{-x^2\} & x > 0 \end{cases}$$

What is the probability that X exceeds 1?

Solution The desired probability is computed as follows:

$$
\begin{aligned}
P\{X > 1\} &= 1 - P\{X \le 1\} \\
&= 1 - F(1) \\
&= e^{-1}
\end{aligned}
$$

$$\Rightarrow \quad P\{X > 1\} = .368$$

Probability Mass Function (PMF)

As was previously mentioned, a random variable whose set of possible values is a sequence is said to be discrete. For a discrete random variable X , we define the probability mass function p(a) of X by

$$p(a) = P\{X = a\}$$

If X assume one of the values $x_1, x_2, \ldots,$ then

$$p(x_i) > 0, \qquad i = 1, 2, \ldots$$

and $$p(x) = 0, \qquad \text{all other values of } x$$

Since X must take on one of the values x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Problem

Consider a random variable X that is equal to 1, 2, or 3. If we know that

$$p(1) = \frac{1}{2} \qquad \text{and} \qquad p(2) = \frac{1}{3}$$

Compute $p(3)$.

Solution

since $p(1) + p(2) + p(3) = 1$

$$\rightarrow \quad p(3) = \frac{1}{6}$$

A graph of $p(x)$ is presented in Figure 4.1.

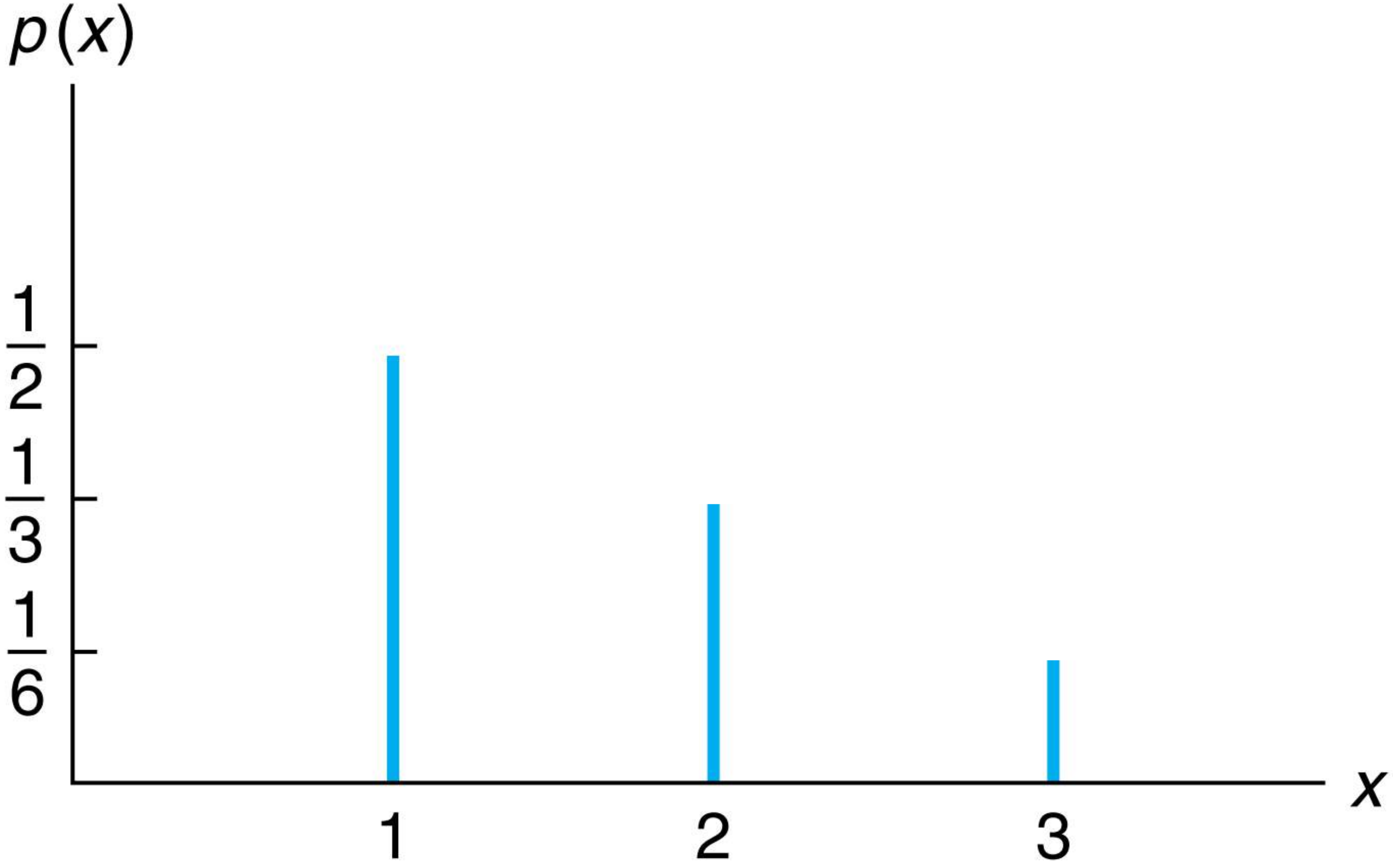


FIGURE 4.1 *Graph of* $p(x)$

Relation of CDF with PMF

The cumulative distribution function F can be expressed in terms of $p(x)$ by

$$F(a) = \sum_{\text{all } x \leq a} p(x)$$

If X is a discrete random variable whose set of possible values are $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \cdots$, then its distribution function F is a step function. That is, the value of F is constant in the intervals $[x_{i-1}, x_i)$ and then takes a step (or jump) of size $p(x_i)$ at x_i.

For instance, suppose X has a probability mass function given by

$$p(1) = \frac{1}{2}, \qquad p(2) = \frac{1}{3}, \qquad p(3) = \frac{1}{6}$$

Then the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{2} & 1 \le a < 2 \\ \frac{5}{6} & 2 \le a < 3 \\ 1 & 3 \le a \end{cases}$$

This is graphically presented in Figure 4.2.

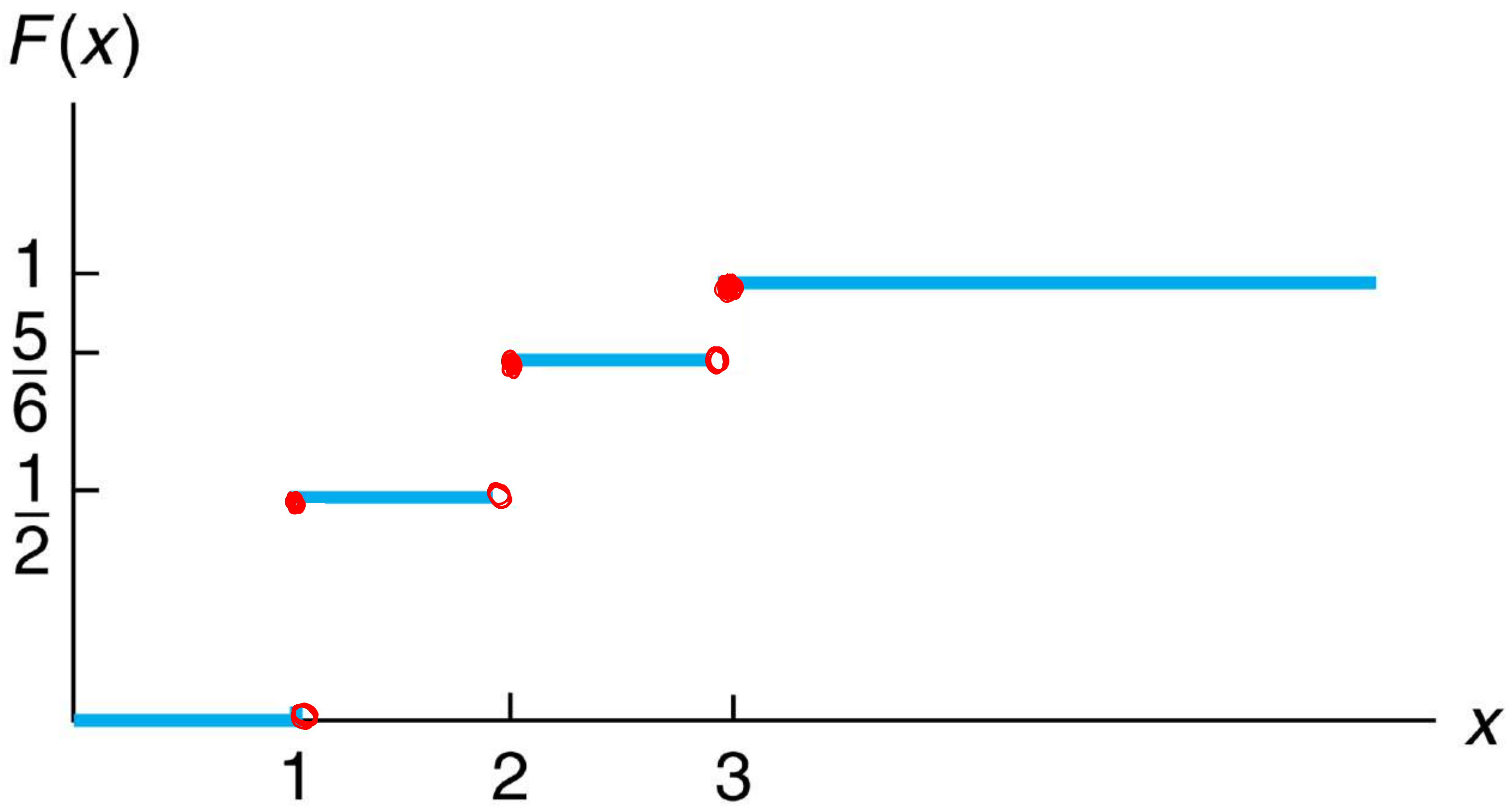


FIGURE 4.2 *Graph of* $F(x)$

Continuous Random Variables

We say that a random variable X has a *continuous distribution* or that X is a *continuous random variable* if there exists a nonnegative function f , defined on the real line, such that for every interval of real numbers (bounded or unbounded), the probability that X takes a value in the interval is the integral of f over the interval.

$$P\{a \leq X \leq b\} = \int_a^b f(x)\, dx$$

The function *f(x)* is called the *probability density function* of the random variable *X.*

Note that a probability density function f(x) must satisfy the following

$$f(x) \geq 0 \text{ for all } x$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

If we let $a = b$ in ★ then

$$P\{X = a\} = \int_a^a f(x)\,dx = 0$$

In words, this equation states that the probability that a continuous random variable will assume any particular value is zero. The fact that P(X = a) = 0 does not imply that X = a is impossible. If it did, all values of X would be impossible and X couldn't assume any value. What happens is that the probability in the distribution of X is spread so thinly that we can only see it on sets like nondegenerate intervals.

Relation b/w CDF and PDF

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)\, dx$$

Differentiating both sides yields

$$\frac{d}{da} F(a) = f(a)$$

Problem

Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of C?

(b) Find $P\{X > 1\}$.

Solution

(a) Since f is a probability density function, we must have that $\int_{-\infty}^{\infty} f(x)\, dx = 1$, implying that

$$C \int_0^2 (4x - 2x^2)\, dx = 1$$

$$\Rightarrow \quad C \left[2x^2 - \frac{2x^3}{3} \right] \Big|_{x=0}^{x=2} = 1$$

or

$$C = \frac{3}{8}$$

(b)

$$P\{X > 1\} = \int_1^\infty f(x)\,dx = \frac{3}{8}\int_1^2 (4x - 2x^2)\,dx = \frac{1}{2}$$

Problem

The amount of time in hours that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

What is the probability that

(a) a computer will function between 50 and 150 hours before breaking down?

(b) it will function for fewer than 100 hours?

Solution (a) Since

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \lambda \int_{0}^{\infty} e^{-x/100}\,dx$$

we obtain

$$1 = -\lambda(100)e^{-x/100}\Big|_{0}^{\infty} = 100\lambda \quad \text{or} \quad \lambda = \frac{1}{100}$$

Hence, the probability that a computer will function between 50 and 150 hours before breaking down is given by

$$P\{50 < X < 150\} = \int_{50}^{150} \frac{1}{100} e^{-x/100}\,dx = -e^{-x/100}\Big|_{50}^{150}$$
$$= e^{-1/2} - e^{-3/2} \approx .383$$

(b) Similarly,

$$P\{X < 100\} = \int_{0}^{100} \frac{1}{100} e^{-x/100}\,dx = -e^{-x/100}\Big|_{0}^{100} = 1 - e^{-1} \approx .632$$

In other words, approximately 63.2 percent of the time, a computer will fail before registering 100 hours of use. ■

Practice Problems

The distribution function of the random variable X is given

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{2} & 0 \le x < 1 \\ \dfrac{2}{3} & 1 \le x < 2 \\ \dfrac{11}{12} & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

(a) Plot this distribution function.

(b) What is $P\{X > \frac{1}{2}\}$?

(c) What is $P\{2 < X \leq 4\}$?

Five men and 5 women are ranked according to their scores on an examination. Assume that no two scores are alike and all 10! possible rankings are equally likely. Let X denote the highest ranking achieved by a woman (for instance, $X = 2$ if the top-ranked person was male and the next-ranked person was female). Find $P\{X = i\}, i = 1, 2, 3, \ldots, 8, 9, 10$.

Jointly Distributed Random Variables

To specify the relationship between two random variables, we define the joint cumulative probability distribution function of X and Y by

$$F(x, y) = P\{X \le x, Y \le y\}$$

A knowledge of the joint probability distribution function often enable us to compute the probability of statements concerning the values of X and Y . For instance, the distribution function of X — call it F_X — can be obtained from the joint distribution function F of X and Y as follows:

$$
\begin{aligned}
F_X(x) &= P\{X \le x\} \\
&= P\{X \le x, Y < \infty\} \\
&= F(x, \infty)
\end{aligned}
$$

Similarly, the cumulative distribution function of Y is given by

$$F_Y(y) = F(\infty, y)$$

In the case where X and Y are both discrete random variables whose possible values are, respectively, $x_1, x_2, \ldots$, and $y_1, y_2, \ldots$, we define the *joint probability mass function* of X and Y, $p(x_i, y_j)$, by

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}$$

The individual probability mass functions of X and Y are easily obtained from the joint probability mass function by the following reasoning. Since Y must take on some value y_j, it follows that the event $\{X = x_i\}$ can be written as the union, over all j, of the mutually exclusive events $\{X = x_i, Y = y_j\}$. That is,

$$\{X = x_i\} = \bigcup_j \{X = x_i, Y = y_j\}$$

and so, using Axiom 3 of the probability function, we see that

$$P\{X = x_i\} = P\left(\bigcup_j \{X = x_i, Y = y_j\}\right)$$

$$= \sum_j P\{X = x_i, Y = y_j\}$$

$$\Rightarrow \quad P\{X = x_i\} = \sum_j p(x_i, y_j)$$

Similarly, we can obtain $P\{Y = y_j\}$ by summing $p(x_i, y_j)$ over all possible values of x_i, that is,

$$P\{Y = y_j\} = \sum_i P\{X = x_i, Y = y_j\}$$

$$P\{Y = y_j\} = \sum_i p(x_i, y_j)$$

Hence, specifying the joint probability mass function always determines the individual mass functions. However, it should be noted that the reverse is not true. Namely, knowledge of $P\{X = x_i\}$ and $P\{Y = y_j\}$ does not determine the value of $P\{X = x_i, Y = y_j\}$.

Suppose that 3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries. If X and Y denote, respectively, the number of new and used but still working batteries that are chosen, then the joint probability mass function of X and Y , $p(i, j) = P\{X = i, Y = j\}$, is given by

$$p(0,0) = \binom{5}{3} \Big/ \binom{12}{3} = 10/220$$

$$p(0,1) = \binom{4}{1}\binom{5}{2} \Big/ \binom{12}{3} = 40/220$$

$$p(0,2) = \binom{4}{2}\binom{5}{1} \Big/ \binom{12}{3} = 30/220$$

$$p(0,3) = \binom{4}{3} \Big/ \binom{12}{3} = 4/220$$

$$p(1,0) = \binom{3}{1}\binom{5}{2} \Big/ \binom{12}{3} = 30/220$$

$$p(1,1) = \binom{3}{1}\binom{4}{1}\binom{5}{1} \Big/ \binom{12}{3} = 60/220$$

$$p(1,2) = \binom{3}{1}\binom{4}{2} \Big/ \binom{12}{3} = 18/220$$

$$p(2,0) = \binom{3}{2}\binom{5}{1} \Big/ \binom{12}{3} = 15/220$$

$$p(2,1) = \binom{3}{2}\binom{4}{1}\Big/\binom{12}{3} = 12/220$$

$$p(3,0) = \binom{3}{3}\Big/\binom{12}{3} = 1/220$$

These probabilities can most easily be expressed in tabular form as shown in Table 4.1.

TABLE 4.1 $P\{X = i, Y = j\}$

i \ j	0	1	2	3	Row Sum $= P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column Sums $=$ $P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

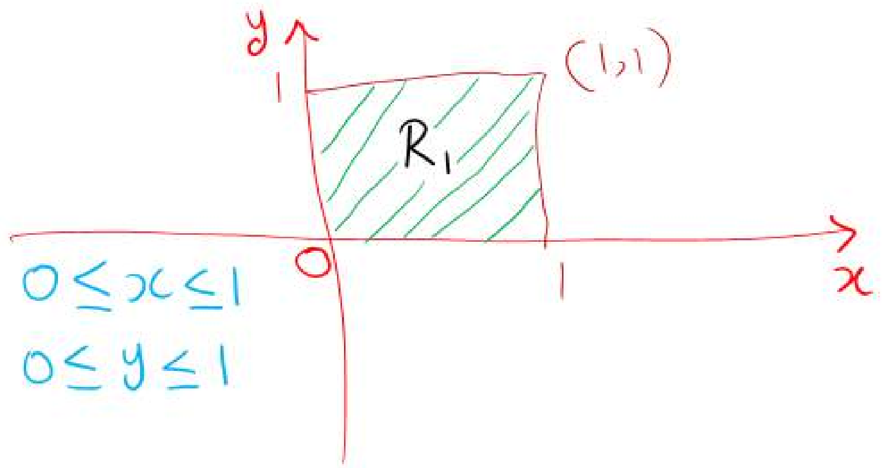
Because the individual probability mass functions of X and Y thus appear in the margin of such a table, they are often referred to as being the marginal probability mass functions of X and Y , respectively. It should be noted that to check the correctness of such a table we could sum the marginal row (or the marginal column) and verify that its sum is 1.

$\int_a^b f(x)\,dx$ is an integral defined over an interval $[a,b]$.

We also have integrals that are defined over a region R in xy-plane. e.g the double integral $\iint_R f(x,y)\,dA$ where R is some region in xy-plane.

The question is : How to evaluate a double integral?
Well! it depends upon the region R.

Consider region R_1 shown in the figure. In this case, we can evaluate the double integral quite easily. i.e.



$$\iint_{R_1} (x^2y + x)\, dA = \int_0^1 \int_0^1 (x^2 y + x)\, dx\, dy$$

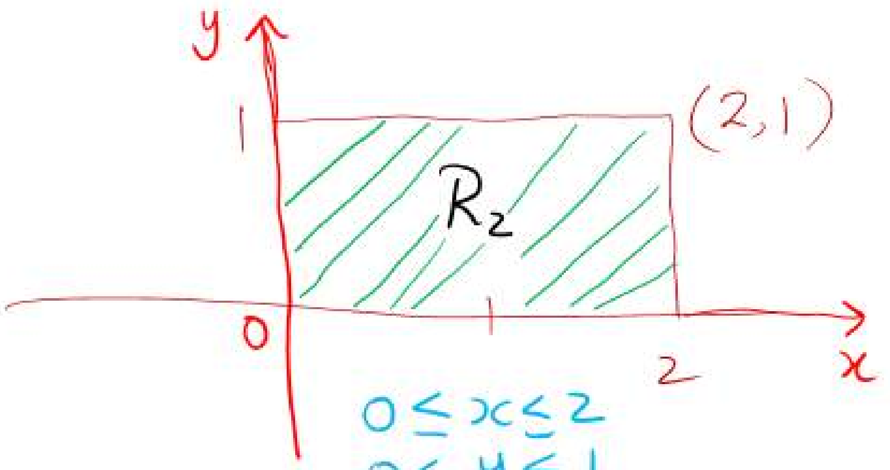
$$= \int_0^1 \left(\frac{x^3 y}{3} + \frac{x^2}{2} \right)\Big|_0^1 dy = \int_0^1 \left(\frac{y}{3} + \frac{1}{2} \right) dy$$

$$= \left(\frac{y^2}{6} + \frac{1}{2}x \right)\Big|_0^1 = \frac{1}{6} + \frac{1}{2} = \frac{1+3}{6} = \frac{4}{6}$$

$$\Rightarrow \boxed{\iint_{R_1} (x^2y + x)\, dA = \frac{2}{3}}$$

Note $\int_0^1 \int_0^1 (x^2y + x)\, dy\, dx$ also gives the same result.

Now consider the region shown in the figure.



$0 \le x \le 2$

$0 \le y \le 1$

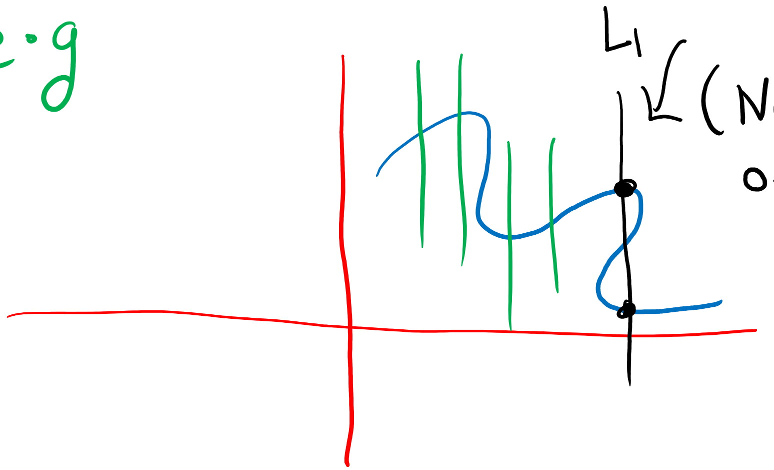
$$\iint_{R_2} (x^2 y + x)\, dA = \int_0^1 \int_0^2 (x^2 y + x)\, dx\, dy = \int_0^2 \int_0^1 (x^2 y + x)\, dy\, dx$$

Vertical Line Test

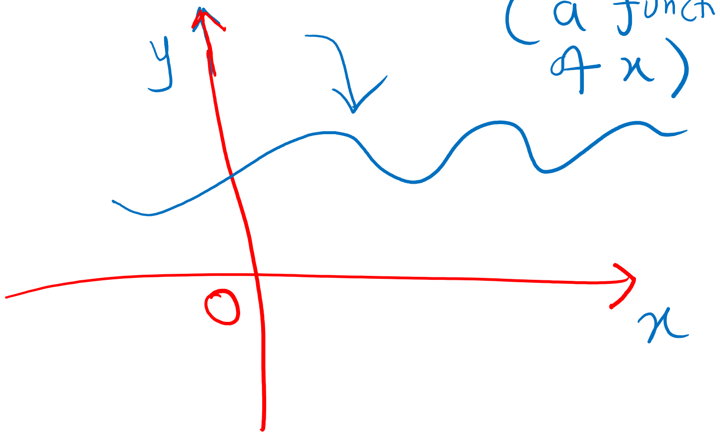
We use V.L.T to check whether a curve is a graph of a function of x.

A curve will be a graph of a function of 'x' if no vertical line intersect the curve more than once.

e.g



(Not a function of x because L_1 intersect the curve at two points)



(a function of x)

A plane region D is said to be of **type I** if it lies between the graphs of two continuous functions of x, that is,

$$D = \left\{(x, y) \mid a \leqslant x \leqslant b,\ g_1(x) \leqslant y \leqslant g_2(x)\right\}$$

where g_1 and g_2 are continuous on $[a, b]$. Some examples of type I regions are shown in Figure 5.

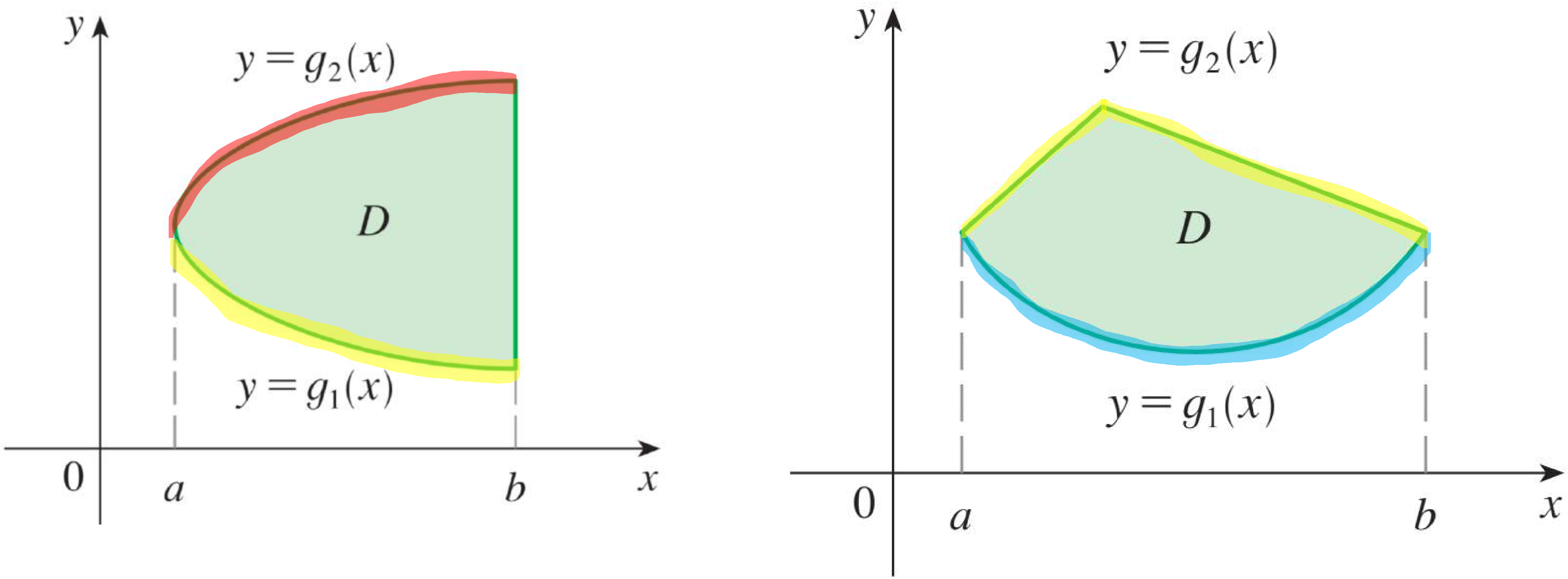


FIGURE 5

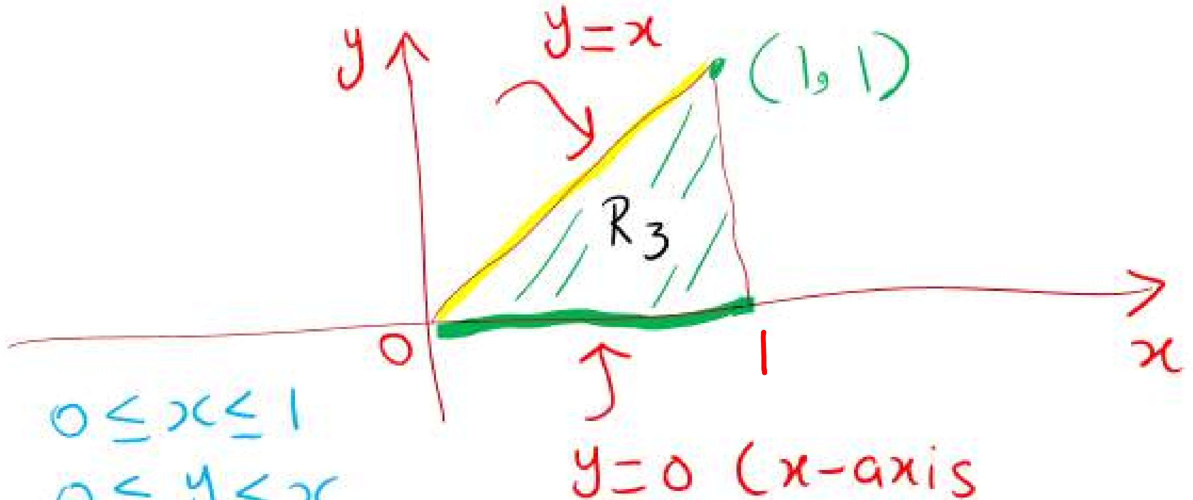
Some type I regions

$\boxed{3}$ If f is continuous on a type I region D described by

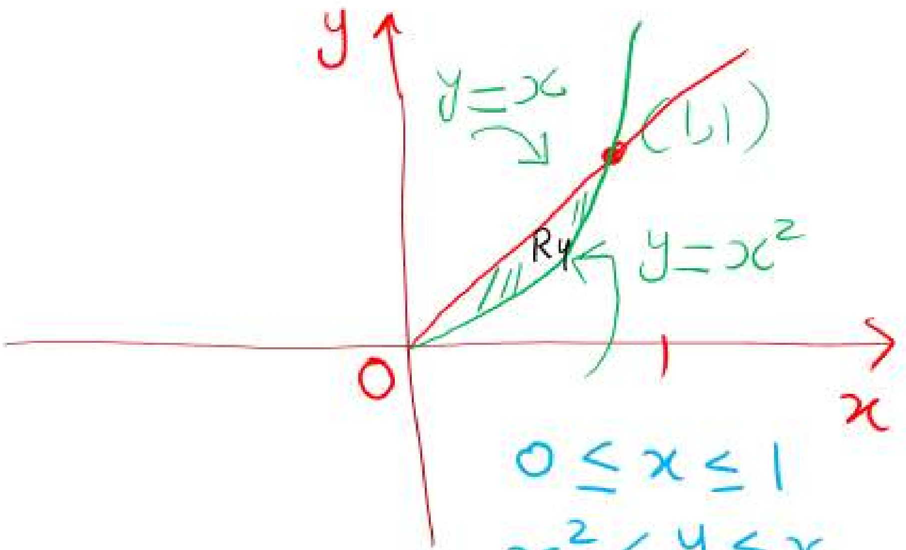
$$D = \{(x, y) \mid a \leqslant x \leqslant b,\ g_1(x) \leqslant y \leqslant g_2(x)\}$$

then
$$\iint_D f(x, y)\, dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x, y)\, dy\, dx$$

Let's evaluate $\iint_{R_3} (x^2y + x)\,dA$ & $\iint_{R_4} (x^2 + x)\,dA$, where R_3 & R_4 are shown in the figure. Both are type I region.



$0 \le x \le 1$
$0 \le y \le x$



$0 \le x \le 1$
$x^2 \le y \le x$

$$\iint_{R_3} (x^2y + x)\, dA = \int_0^1 \int_0^x (x^2 y + x)\, dy\, dx = \int_0^1 \left(\frac{x^2 y^2}{2} + xy \right) \Big|_0^x dx$$

$$= \int_0^1 \left(\frac{x^2 \cdot x^2}{2} + x \cdot x \right) dx = \int_0^1 \left(\frac{x^4}{2} + x^2 \right) dx$$

$$= \left(\frac{x^5}{10} + \frac{x^3}{3} \right) \Big|_0^1 = \frac{1}{10} + \frac{1}{3} = \frac{3+10}{30} = \frac{13}{30}$$

$$\Rightarrow \boxed{\iint_{R_3} (x^2y + x)\, dA = \frac{13}{30}}$$

Similarly $\iint_{R_4} (x^2 y + x)\, dx$

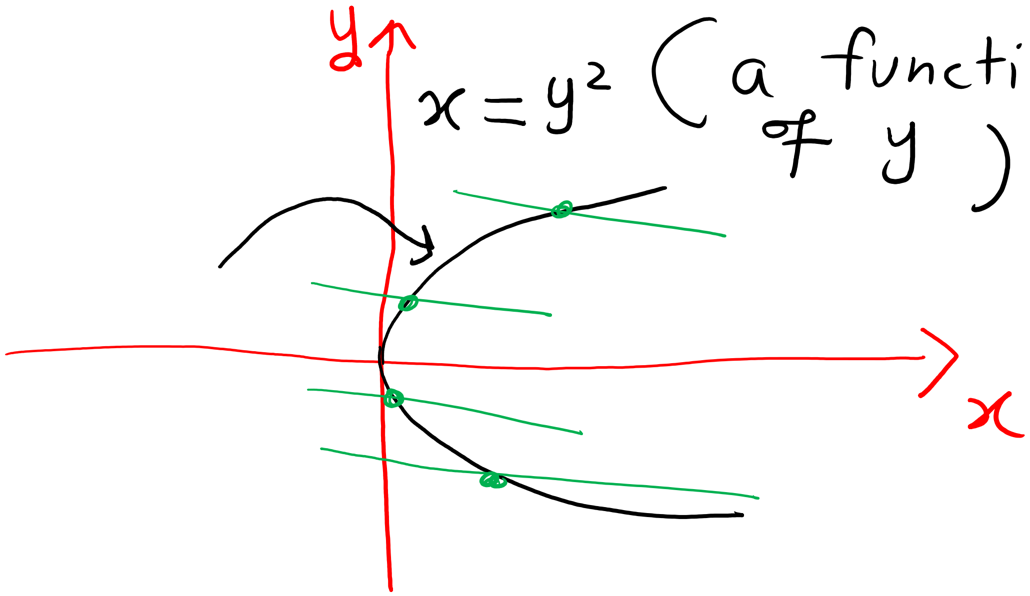
$$= \int_0^1 \int_{x^2}^{x} (x^2 y + x)\, dy\, dx$$

Next,

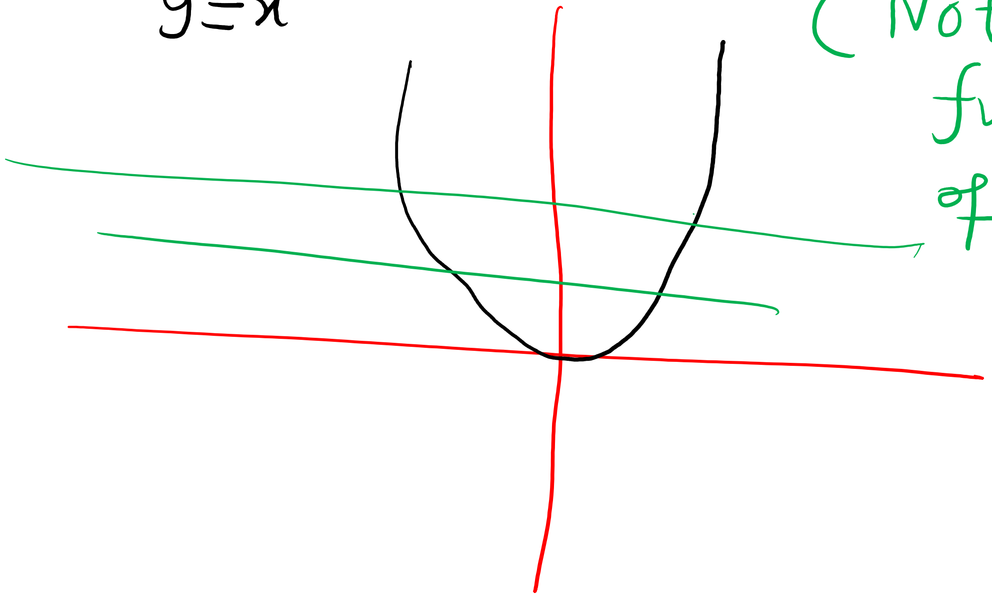
A curve will be a graph of a function of 'y' if no horizontal line intersect the curve more than once.

For example

$x = y^2$ (a function of y)



$y = x^2$



(Not a function of y)

We also consider plane regions of **type II**, which can be expressed as

$$D = \{(x, y) \mid c \leqslant y \leqslant d,\ h_1(y) \leqslant x \leqslant h_2(y)\}$$

where h_1 and h_2 are continuous. Three such regions are illustrated in Figure 7.

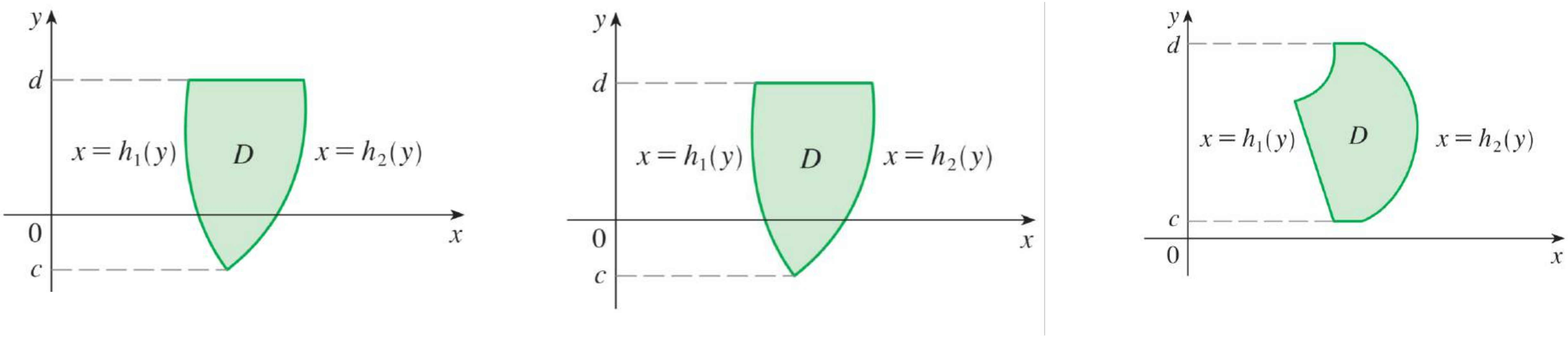


FIGURE 7

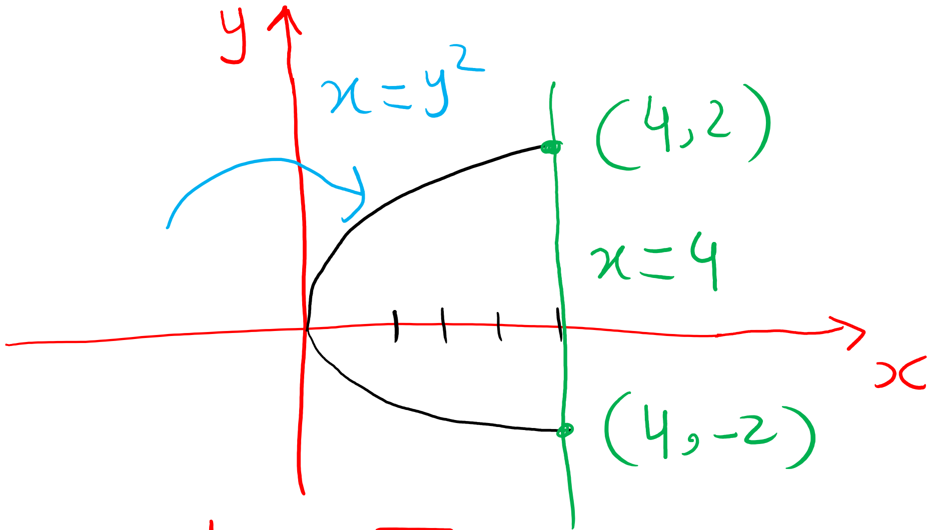
Some type II regions

$\boxed{4}$ If f is continuous on a type II region D described by

$$D = \left\{(x, y) \mid c \leqslant y \leqslant d,\ h_1(y) \leqslant x \leqslant h_2(y)\right\}$$

then
$$\iint_D f(x, y)\, dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y)\, dx\, dy$$

Let's evaluate $\iint_{R_5} (x^2y + x)\, dA$, where R_5 is shown in the figure.



Clearly R_5 is a type II region $-2 \le y \le 2$ and $y^2 \le x \le 4$

So

$$\iint_{R_5} (x^2y + x)\, dA = \int_{-2}^{2} \int_{y^2}^{4} (x^2y + x)\, dx\, dy$$

Recall

The Fundamental Theorem of Calculus, Part 1 If f is continuous on $[a, b]$, then the function g defined by

$$g(x) = \int_a^x f(t)\,dt \qquad a \leqslant x \leqslant b$$

is continuous on $[a, b]$ and differentiable on (a, b), and $g'(x) = f(x)$.

We say that X and Y are *jointly continuous* if there exists a function $f(x, y)$ defined for all real x and y, having the property that for every set C of pairs of real numbers (that is, C is a set in the two-dimensional plane)

$$P\{(X, Y) \in C\} = \iint_{(x,y)\in C} f(x, y)\, dx\, dy \tag{4.3.3}$$

The function $f(x, y)$ is called the *joint probability density function* of X and Y. If A and B are any sets of real numbers, then by defining $C = \{(x, y) : x \in A, y \in B\}$, we see from Equation 4.3.3 that

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y)\, dx\, dy \tag{4.3.4}$$

Because

$$F(a, b) = P\{X \in (-\infty, a], Y \in (-\infty, b]\}$$

$$= \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y)\, dx\, dy$$

it follows, upon differentiation, that

$$f(a, b) = \frac{\partial^2}{\partial a\, \partial b} F(a, b)$$

If X and Y are jointly continuous, they are individually continuous, and their probability density functions can be obtained as follows:

$$P\{X \in A\} = P\{X \in A, Y \in (-\infty, \infty)\} \tag{4.3.5}$$

$$= \int_A \int_{-\infty}^{\infty} f(x, y)\, dy\, dx$$

$$\Rightarrow \quad P\{X \in A\} = \int_A f_X(x)\, dx$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$ is thus the probability density function of X.

Similarly, the probability density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \tag{4.3.6}$$

Problem The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute **(a)** $P\{X > 1, Y < 1\}$; **(b)** $P\{X < Y\}$; and **(c)** $P\{X < a\}$.

Solution

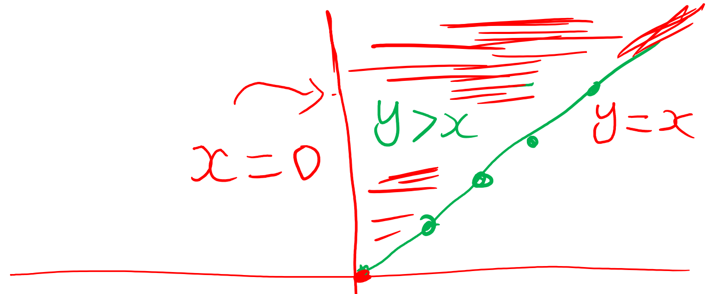
(a)

$$P\{X > 1, Y < 1\} = \int_0^1 \int_1^\infty 2e^{-x} e^{-2y}\, dx\, dy$$

$$= \int_0^1 2e^{-2y} (-e^{-x} |_1^\infty)\, dy$$

$$= e^{-1} \int_0^1 2e^{-2y}\, dy$$

$$\Rightarrow \quad P\{X > 1, Y < 1\} = e^{-1}(1 - e^{-2})$$



$$P\{X < Y\} = \iint_{(x,y):x<y} 2e^{-x}e^{-2y}\,dx\,dy$$

$$= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}\,dx\,dy = \int_0^\infty 2e^{-2y}(1 - e^{-y})\,dy$$

$$= \int_0^\infty 2e^{-2y}\,dy - \int_0^\infty 2e^{-3y}\,dy \quad = 1 - \frac{2}{3}$$

$$\Rightarrow \quad P\{X < Y\} = \frac{1}{3}$$

$$P\{X < a\} = \int_0^a \int_0^\infty 2e^{-2y} e^{-x}\, dy\, dx \ = \int_0^a e^{-x}\, dx$$

$$\Longrightarrow \quad P\{X < a\} \ = 1 - e^{-a}$$

Independent Random Variables

The random variables X and Y are said to be independent if for any two sets of real numbers A and B

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

When X and Y are discrete random variables, the condition of independence is equivalent to

$$p(x, y) = p_X(x)p_Y(y) \qquad \text{for all } x, y$$

where p_X and p_Y are the probability mass functions of X and Y.

In the jointly continuous case, the condition of independence is equivalent to

$$f(x, y) = f_X(x)f_Y(y) \qquad \text{for all } x, y$$

Loosely speaking, X and Y are independent if knowing the value of one does not change the distribution of the other. Random variables that are not independent are said to be dependent.

Expectation

If X is a discrete random variable taking on the possible values $x_1, x_2, \ldots$, then the expectation or expected value of X , denoted by E[X], is defined by

$$E[X] = \sum_i x_i P\{X = x_i\}$$

Problem

Find $E[X]$ where X is the outcome when we roll a fair die.

Solution

Since $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$, we obtain that

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

It is important to note that the expected value of X is not a value that X could possibly assume. (That is, in the previous problem, rolling a die cannot possibly lead to an outcome of 7/2.) Thus, even though we call $E[X]$ the expectation of X, it should not be interpreted as the value that we expect X to have but rather as the average value of X in a large number of repetitions of the experiment. That is, if we continually roll a fair die, then after a large number of rolls the average of all the outcomes will be approximately 7/2.

We can also define the expectation of a continuous random variable. Suppose that X is a continuous random variable with probability density function f then the expected value of X by

$$E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

Problem

Suppose that you are expecting a message at some time past 5 P.M. From experience you know that X, the number of hours after 5 P.M. until the message arrives, is a random variable with the following probability density function:

$$f(x) = \begin{cases} \dfrac{1}{1.5} & \text{if } 0 < x < 1.5 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the expected amount of time past 5 P.M. until the message arrives.

Solution

The expected amount of time past 5 P.M. until the message arrives is given by

$$E[X] = \int_0^{1.5} \frac{x}{1.5}\, dx = .75$$

Hence, on average, you would have to wait three-fourths of an hour.

Problem Let X denote a random variable that takes on any of the values -1, 0, and 1 with respective probabilities

$$P\{X = -1\} = .2 \quad P\{X = 0\} = .5 \quad P\{X = 1\} = .3$$

Compute $E[X^2]$.

Solution

Let $Y = X^2$. Then the probability mass function of Y is given by

$$P\{Y = 1\} = P\{X = -1\} + P\{X = 1\} = .5$$
$$P\{Y = 0\} = P\{X = 0\} = .5$$

$$E[X^2] = E[Y] = 1(.5) + 0(.5) = .5$$

Expectation of a Function of a Random variable

(a) If X is a discrete random variable with probability mass function $p(x)$, then for any real-valued function g,

$$E[g(X)] = \sum_{x} g(x)p(x)$$

(b) If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

If a and b are constants, then

$$\boxed{E[aX + b] = aE[X] + b}$$

If we take $a = 0$

$$E[b] = b$$

That is, the expected value of a constant is just its value. if we take $b = 0$, then we obtain

$$E[aX] = aE[X]$$

or, in words, the expected value of a constant multiplied by a random variable is just the constant times the expected value of the random variable.

The expected value of a random variable X , E[X], is also referred to as the mean or the first moment of X . The quantity $E[X^n]$, n ≥ 1, is called the nth moment of X .

$$E[X^n] = \begin{cases} \displaystyle\sum_x x^n p(x) & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} x^n f(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

In general, for any n,

$$E[X_1 + X_2 \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

Problem

A construction firm has recently sent in bids for 3 jobs worth (in profits) 10, 20, and 40 (thousand) dollars. If its probabilities of winning the jobs are respectively .2, .8, and .3, what is the firm's expected total profit?

Solution

Letting $X_i, i = 1, 2, 3$ denote the firm's profit from job i, then

$$\text{total profit} = X_1 + X_2 + X_3$$

and so

$$E[\text{total profit}] = E[X_1] + E[X_2] + E[X_3]$$

Now

$$E[X_1] = 10(.2) + 0(.8) = 2$$

$$E[X_2] = 20(.8) + 0(.2) = 16$$

$$E[X_3] = 40(.3) + 0(.7) = 12$$

and thus the firm's expected total profit is 30 thousand dollars.

Variance

If X is a random variable with mean μ, then the *variance* of X, denoted by Var(X), is defined by

$$\text{Var}(X) = E[(X - \mu)^2]$$

An alternative formula for Var(X) can be derived as follows:

$$\begin{aligned}\text{Var}(X) = E[(X - \mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - E[2\mu X] + E[\mu^2] = E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2\end{aligned}$$

That is,

$$\boxed{\mathrm{Var}(X) = E[X^2] - (E[X])^2}$$

or, in words, the variance of X is equal to the expected value of the square of X minus the square of the expected value of X. This is, in practice, often the easiest way to compute $\mathrm{Var}(X)$.

Problem Compute Var(X) when X represents the outcome when we roll a fair die.

Solution Since $P\{X = i\} = \frac{1}{6}, i = 1, 2, 3, 4, 5, 6$, we obtain

$$E[X^2] = \sum_{i-1}^{6} i^2 P\{X = i\} = 1^2\left(\tfrac{1}{6}\right) + 2^2\left(\tfrac{1}{6}\right) + 3^2\left(\tfrac{1}{6}\right) + 4^2\left(\tfrac{1}{6}\right) + 5^2\left(\tfrac{1}{6}\right) + 6^2\left(\tfrac{1}{6}\right)$$

$$= \frac{91}{6}$$

$$E[X] = \frac{7}{2}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

A useful identity concerning variances is that for any constants a and b,

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

Specifying particular values for a and b in above equation leads to some important result. For instance, by setting $a = 0$

$$\mathrm{Var}(b) = 0$$

That is, the variance of a constant is 0.

Similarly, by setting $a = 1$ we obtain

$$\mathrm{Var}(X + b) = \mathrm{Var}(X)$$

That is, the variance of a constant plus a random variable is equal to the variance of the random variable.

Finally, setting $b = 0$ yields

$$\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$$

The square root of the $\mathrm{Var}(X)$ is called the *standard deviation* of X, and we denote it by $\mathrm{SD}(X)$. That is,

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

Covariance

The *covariance* of two random variables X and Y, written $\text{Cov}(X, Y)$, is defined by

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

where μ_x and μ_y are the means of X and Y, respectively.

A useful expression for $\text{Cov}(X, Y)$ can be obtained by expanding the right side of the definition. This yields

$$
\begin{aligned}
\mathrm{Cov}(X, Y) &= E[XY - \mu_x Y - \mu_y X + \mu_x \mu_y] \\
&= E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x \mu_y \\
&= E[XY] - \mu_x \mu_y - \mu_y \mu_x + \mu_x \mu_y \\
&= E[XY] - E[X]E[Y]
\end{aligned}
$$

From its definition we see that covariance satisfies the following properties:

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$$

and

$$\mathrm{Cov}(X, X) = \mathrm{Var}(X)$$

In general, a positive value of Cov(X , Y) is an indication that Y tends to increase as X does, whereas a negative value indicates that Y tends to decrease as X increases. The strength of the relationship between X and Y is indicated by the correlation between X and Y , a dimensionless quantity obtained by dividing the covariance by the product of the standard deviations of X and Y . That is,

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

This quantity always has a value between −1 and +1.

Problem

A product is classified according to the number of defects it contains and the factory that produces it. Let X_1 and X_2 be the random variables that represent the number of defects per unit (taking on possible values of 0, 1, 2, or 3) and the factory number (taking on possible values 1 or 2), respectively. The entries in the table represent the joint possibility mass function of a randomly chosen product.

$X_1 \backslash X_2$	1	2
0	$\frac{1}{8}$	$\frac{1}{16}$
1	$\frac{1}{16}$	$\frac{1}{16}$
2	$\frac{3}{16}$	$\frac{1}{8}$
3	$\frac{1}{8}$	$\frac{1}{4}$

(a) Find the marginal probability distributions of X_1 and X_2.

(b) Find $E[(X_1)]$, $E[(X_2)]$, and $\mathrm{Cov}(X_1, X_2)$.

Solution

X_1 \ X_2	1	2	Row Sum $P\{X_1 = i\}$
0	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{8} + \frac{1}{16} = 3/16$
1	$\frac{1}{16}$	$\frac{1}{16}$	$1/8$
2	$\frac{3}{16}$	$\frac{1}{8}$	$5/16$
3	$\frac{1}{8}$	$\frac{1}{4}$	$3/8$
Column Sum $P\{X_2 = i\}$	$1/2$	$1/2$	

$$E[X_1] = \left(0 \times \frac{3}{16}\right) + \left(1 \times \frac{1}{8}\right) + \left(2 \times \frac{5}{16}\right) + \left(3 \times \frac{3}{8}\right)$$

$$E[X_1] = \frac{15}{8}$$

Similarly

$$E[X_2] = \left(1 \times \frac{1}{2}\right) + \left(2 \times \frac{1}{2}\right) = \frac{3}{2}$$

$$\mathrm{Cov}(X_1, X_2) = E[X_1 X_2] - E[X_1]E[X_2]$$

$$E[X_1X_2] = \sum_{x_1}\sum_{x_2} x_1x_2\,p(x_1,x_2) - \left(\frac{15}{8}\times\frac{3}{2}\right)$$

$$\sum_{x_1}\sum_{x_2} x_1x_2\,p(x_1,x_2)$$

$$= (0)(1)\left(\frac{1}{8}\right) + (0)(2)\left(\frac{1}{16}\right) + (1)(1)\left(\frac{1}{16}\right) +$$

$$(1)(2)\left(\frac{1}{16}\right) + (2)(1)\left(\frac{3}{16}\right) + (2)(2)\left(\frac{1}{8}\right) +$$

$$(3)(1)\left(\frac{1}{8}\right) + (3)(2)\left(\frac{1}{4}\right) - \frac{45}{16}$$

and so on.

Special Random Variables

Certain types of random variables occur over and over again in applications. Here we will study a variety of them.

The Bernoulli & Binomial Random Variables

Suppose that a trial, or an experiment, whose outcome can be classified as either a "success" or as a "failure" is performed. If we let X = 1 when the outcome is a success and X = 0 when it is a failure, then the probability mass function of X is given by

$$\boxed{\begin{aligned} P\{X = 0\} &= 1 - p \\ P\{X = 1\} &= p \end{aligned}} \qquad \text{(i)}$$

where $p, 0 \leq p \leq 1$, is the probability that the trial is a "success." A random variable X is said to be a Bernoulli random variable (after the Swiss mathematician James Bernoulli) if its probability mass function is given by Equations (i) for some $p \in (0, 1)$.

Its expected value is

$$E[X] = 1 \cdot P\{X = 1\} + 0 \cdot P\{X = 0\} = p$$

That is, the expectation of a Bernoulli random variable is the probability that the random variable equals 1.

Similarly by putting values in the formula of variance we get

$$Var(X) = p(1 - p)$$

Suppose now that n independent trials, each of which results in a "success" with probability p and in a "failure" with probability $1 - p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a *binomial* random variable with parameters (n, p).

The probability mass function of a binomial random variable with parameters n and p is given by

The probability mass function of a binomial random variable with parameters n and p is given by

$$\boxed{P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \ldots, n}$$ (ii)

where $\binom{n}{i} = n!/[i!(n-i)!]$

If X is a binomial random variable with parameters n and p, then

$$\boxed{\begin{aligned} E[X] &= np \\ \mathrm{Var}(X) &= np(1 - p) \end{aligned}}$$

Problem

Five fair coins are flipped. If the outcomes are assumed independent, find the probability mass function of the number of heads obtained.

Solution

If we let X equal the number of heads (successes) that appear, then X is a binomial random variable with parameters $\left(n = 5,\ p = \frac{1}{2}\right)$. Hence,

$$P\{X = 0\} = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

$$P\{X = 1\} = \binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 = \frac{5}{32}$$

$$P\{X = 2\} = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{10}{32}$$

$$P\{X = 3\} = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{10}{32}$$

$$P\{X = 4\} = \binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 = \frac{5}{32}$$

$$P\{X = 5\} = \binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 = \frac{1}{32}$$

Problem

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) at least 10 survive, (b) from 3 to 8 survive, and (c) exactly 5 survive?

Solution

Let X be the number of people who survive. For any patient there are two cases: Either the patient will survive (Success) or will die (Failure).

Number of people having the disease $= n = 15$.

$P\{X \geq 10\} = ?$, $P\{3 \leq X \leq 8\} = ?$, $P\{X = 5\} = ?$.

Clearly we can see that X is a Binomial Random Variable.

Here X can take the values $0, 1, 2, \ldots, 15$. Also it is given that the probability of patient recovery (success) is 0.4 So $p = 0.4$. Thus the PMF of X is

$$P\{X = i\} = \binom{15}{i}(0.4)^{i}(0.6)^{15-i}, \quad i = 0, 1, 2, \ldots, 15.$$

(a) Now

$$P\{X \geq 10\} = P\{X = 10\} + P\{X = 11\} + \cdots + P\{X = 15\}$$

$$= \binom{15}{10}(0.4)^{10}(0.6)^{5} + \binom{15}{11}(0.4)^{11}(0.6)^{4} + \binom{15}{12}(0.4)^{12}(0.6)^{3}$$

$$+ \binom{5}{13}(0.4)^{13}(0.6)^{2} + \binom{15}{14}(0.4)^{14}(0.6)^{1} + \binom{15}{15}(0.4)^{15}(0.6)^{0}$$

$$\Rightarrow \boxed{P\{X \geq 10\} = 0.0338} \quad \text{Ans}$$

(b)

$$P\{3 \leq X \leq 8\} = P\{X=3\} + P\{X=4\} + \cdots + P\{X=8\}$$

$$= \binom{15}{3}(0.4)^3(0.6)^{12} + \cdots + \binom{15}{8}(0.4)^8(0.6)^7$$

$$\Rightarrow \boxed{P\{3 \leq X \leq 8\} = 0.8779} \quad \text{Ans}$$

(c) Similarly

$$\boxed{P\{X=5\} = 0.1859} \quad \text{Ans}$$

Practice Problem

A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%.

(a) The inspector randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?

(b) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

The Poisson Random Variable

A random variable X, taking on one of the values $0, 1, 2, \ldots,$ is said to be a Poisson random variable with parameter λ, $\lambda > 0$, if its probability mass function is given by

$$\boxed{P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, \qquad i = 0, 1, \ldots}$$

The symbol e stands for a constant approximately equal to 2.7183. It is a famous constant in mathematics, named after the Swiss mathematician L. Euler, and it is also the base of the so-called natural logarithm.

The Poisson probability distribution was introduced by S. D. Poisson in a book he wrote dealing with the application of probability theory to lawsuits, criminal trials, and the like.

Both the mean and the variance of a Poisson random variable are equal to the parameter λ.

The Poisson random variable has a wide range of applications in a variety of areas because it can be used as an approximation for a binomial random variable with parameters (n, p) when n is large and p is small.

If n independent trials, each of which results in a "success" with probability p, are performed, then when n is large and p small, the number of successes occurring is approximately a Poisson random variable with mean $\lambda = np$.

Some examples of random variables that usually obey, to a good approximation, the Poisson probability law (that is, they usually obey equation for some value of λ) are:

1. The number of misprints on a page (or a group of pages) of a book.
2. The number of people in a community living to 100 years of age.
3. The number of wrong telephone numbers that are dialed in a day.
4. The number of transistors that fail on their first day of use.

Problem

Suppose that the average number of accidents occurring weekly on a particular stretch of a highway equals 3. Calculate the probability that there is at least one accident this week.

Solution

Let X denote the number of accidents occurring on the stretch of highway in question during this week. Because it is reasonable to suppose that there are a large number of cars passing along that stretch, each having a small probability of being involved in an accident, the number of such accidents should be approximately Poisson distributed. Hence,

$$
\begin{aligned}
P\{X \geq 1\} &= 1 - P\{X = 0\} \\
&= 1 - e^{-3}\frac{3^0}{0!} \\
&= 1 - e^{-3} \approx .9502
\end{aligned}
$$

Problem

Suppose that 1 person in 1000 makes a numerical error in preparing his or her income tax return. If 10,000 returns are selected at random and examined, find the probability that 6 of them contain an error.

Geometric Random Variable

In a series of Bernoulli trials (independent trials with constant probability p of success), let the random variable X denote the number of trials until the first success. Then, X is a geometric random variable with parameter p such that $0 < p < 1$ and the probability mass function of X is

$$P\{X = i\} = p(1 - p)^{i-1}, \qquad i = 1,2,3, \dots$$

The binomial and geometric distributions arise in very similar situations. The significant difference is that the number of trials in a binomial distribution is fixed from the start and the number of successes are counted, whereas, in a geometric distribution, trials are repeated as many times as necessary until the first success occurs.

Problem

Let X be a geometric random variable with $p = 0.25$. What is the probability that $X = 4$ (i.e. that the first success occurs on the 4th trial)?

Note: For X to be equal to 4, we must have had 3 failures, and then a success.

Solution

$$P\{X = 4\} = (0.25)(1 - 0.25)^{4-1} = 0.1055$$

Problem

For a certain manufacturing process, it is known that 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

Solution

Let X is the number of items inspected before we found a defective item.

p =probability of being defective item= $1/100$, then X is a geometric R.V. with PMF given by

$$P\{X = i\} = (1/100)(1 - 1/100)^{i-1}, i = 1,2,3, \dots$$

Now

$$P\{X = 5\} = (1/100)(1 - 1/100)^{5-1} = 0.0096$$

The mean and variance of geometric random variable are given by:

$$E[X] = \frac{1}{p} \qquad \& \qquad Var(X) = \frac{1-p}{p^2}$$

Practice Problem

The probability that a student pilot passes the written test for a private pilot's license is 0.7. Find the probability that a given student will pass the test

(a) on the third try;

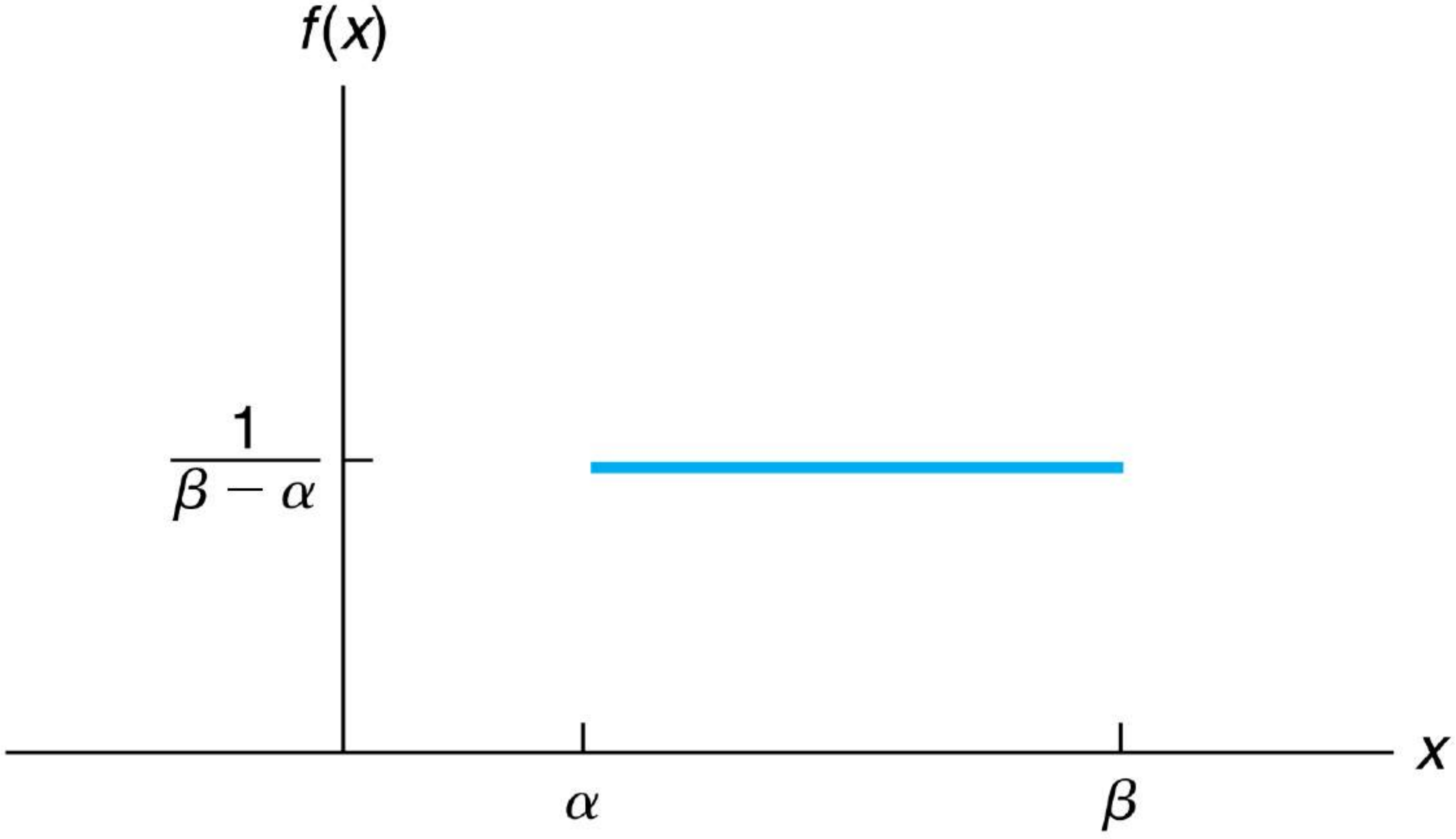
(b) before the fourth try.

The Uniform Random Variable

A random variable X is said to be uniformly distributed over the interval $[\alpha, \beta]$ if its probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{if } \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$

A graph of this function is given in Figure 1 on the next slide.



Graph of $f(x)$ *for a uniform* $[\alpha, \beta]$.

If we divide the domain of the uniform random variable in equal parts, then all parts will be equally likely. This mean that the probability that x lies in an interval of width Δx entirely contained in the interval from α to β is equal to $\Delta x/(\beta - \alpha)$, regardless of the exact location of the interval.

$$P\{a < X < b\} = \frac{1}{\beta - \alpha} \int_a^b dx$$
$$= \frac{b-a}{\beta - \alpha}$$

The probability that X lies in any subinterval of $[\alpha, \beta]$ is equal to the length of that subinterval divided by the length of the interval $[\alpha, \beta]$.

$$E[X] = \frac{\alpha + \beta}{2} \qquad \& \qquad \mathrm{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$

Problem

If X is uniformly distributed over the interval $[0, 10]$, compute the probability that (a) $2 < X < 9$, (b) $1 < X < 4$, (c) $X < 5$, (d) $X > 6$.

Solution

The respective answers are (a) $7/10$, (b) $3/10$, (c) $5/10$, (d) $4/10$.

Problem

Buses arrive at a specified stop at 15-minute intervals starting at 7 A.M. That is, they arrive at 7, 7:15, 7:30, 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability that he waits

(a) less than 5 minutes for a bus;

(b) at least 12 minutes for a bus

Solution

Let X denote the time in minutes past 7 A.M. that the passenger arrives at the stop. Since X is a uniform random variable over the interval (0, 30), it follows that the passenger will have to wait less than 5 minutes if he arrives between 7:10 and 7:15 or between 7:25 and 7:30. Hence, the desired probability for (a) is

$$P\{10 < X < 15\} + P\{25 < X < 30\} = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}$$

Similarly, he would have to wait at least 12 minutes if he arrives between 7 and 7:03 or between 7:15 and 7:18, and so the probability for (b) is

$$P\{0 < X < 3\} + P\{15 < X < 18\} = \frac{3}{30} + \frac{3}{30} = \frac{1}{5}$$

Problem

Beginning at 12:00 midnight, a computer center is up for 1 hour and down for 2 hours on a regular cycle. A person who doesn't know the schedule dials the center at a random time between 12:00 midnight and 5:00 AM. what is the probability that the center will be operating when he dials in?

Solution

Since person is equally likely to call between 12:00 midnight and 5:00 A.M. it is a uniform distribution. Let random variable Y be the time when the person calls.

Since the interval in which the person calls is 5 hours long, the interval will be $(0, 5)$.

Hence,

$$P(0 < Y < 1) + P(3 < Y < 4) = \int_0^1 f(y)dy + \int_3^4 f(y)dy$$

Since $f(y) = \dfrac{1}{5} \; \forall \; y \in (0,5)$, hence

$$\begin{aligned} P(0 < Y < 1) + P(3 < Y < 4) &= \int_0^1 \left(\frac{1}{5}\right) dy + \int_3^4 \left(\frac{1}{5}\right) dy \\ &= \left(\frac{1}{5}\right) y\Big|_0^1 + \left(\frac{1}{5}\right) y\Big|_3^4 \\ &= \frac{1-0}{5} + \frac{4-3}{5} \;\; = 0.4 \end{aligned}$$

Practice Problem

Waiting Times You arrive at a bus stop to wait for a bus that comes by once every 30 minutes. You don't know what time the last bus came by. The time x that you wait before the bus arrives is uniformly distributed on the interval from 0 to 30 minutes.

a. What is the probability that you will have to wait longer than 20 minutes?

b. What is the probability that you will have to wait less than 10 minutes?

Normal Random Variables

A random variable is said to be normally distributed with parameters μ and σ^2, and we write $X \sim N(\mu, \sigma^2)$, if its density is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

The normal density $f(x)$ is a bell-shaped curve that is symmetric about μ and that attains its maximum value of $\frac{1}{\sqrt{2\pi}\sigma} \approx \frac{0.399}{\sigma}$ at $x = \mu$ shown in figure 1 on the next slide.

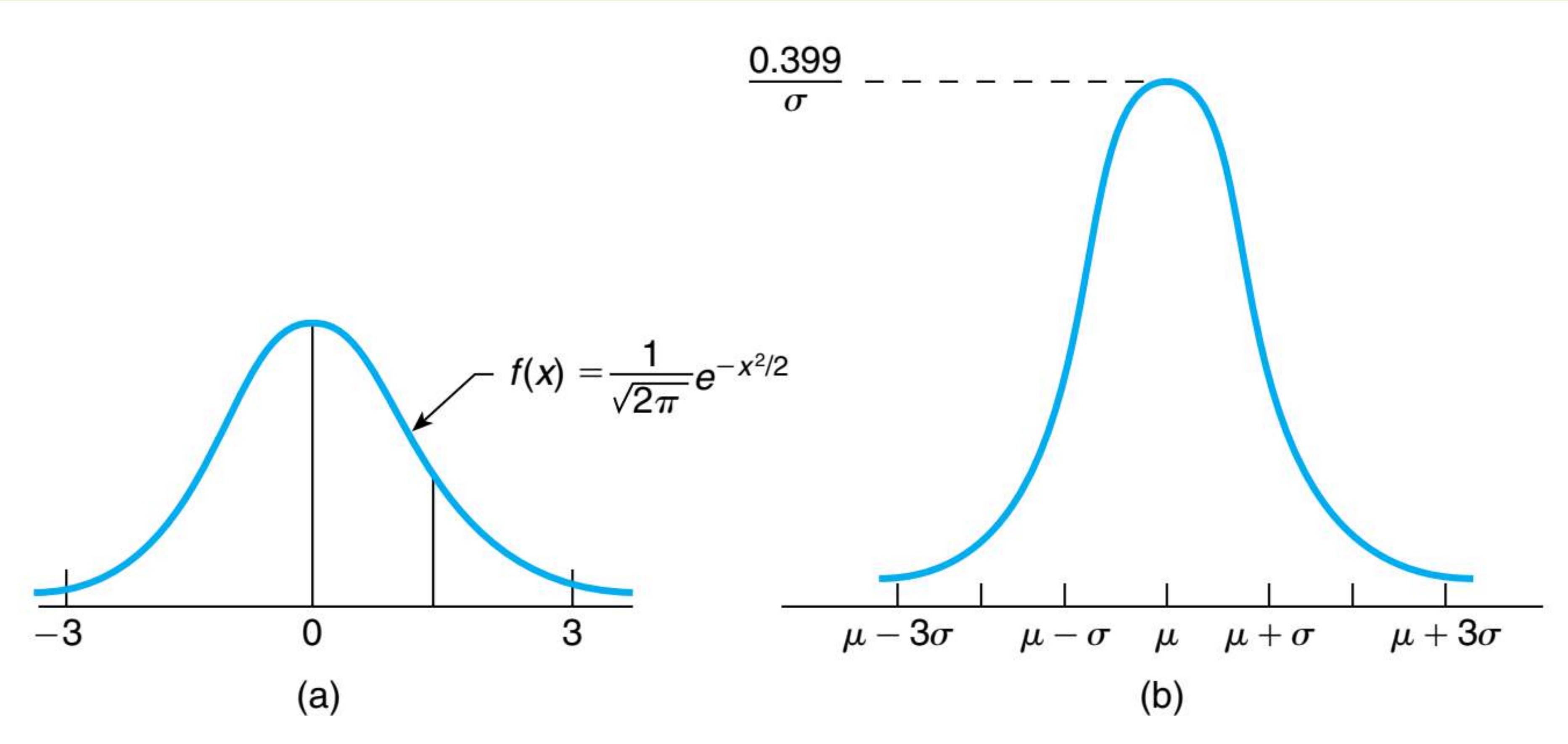
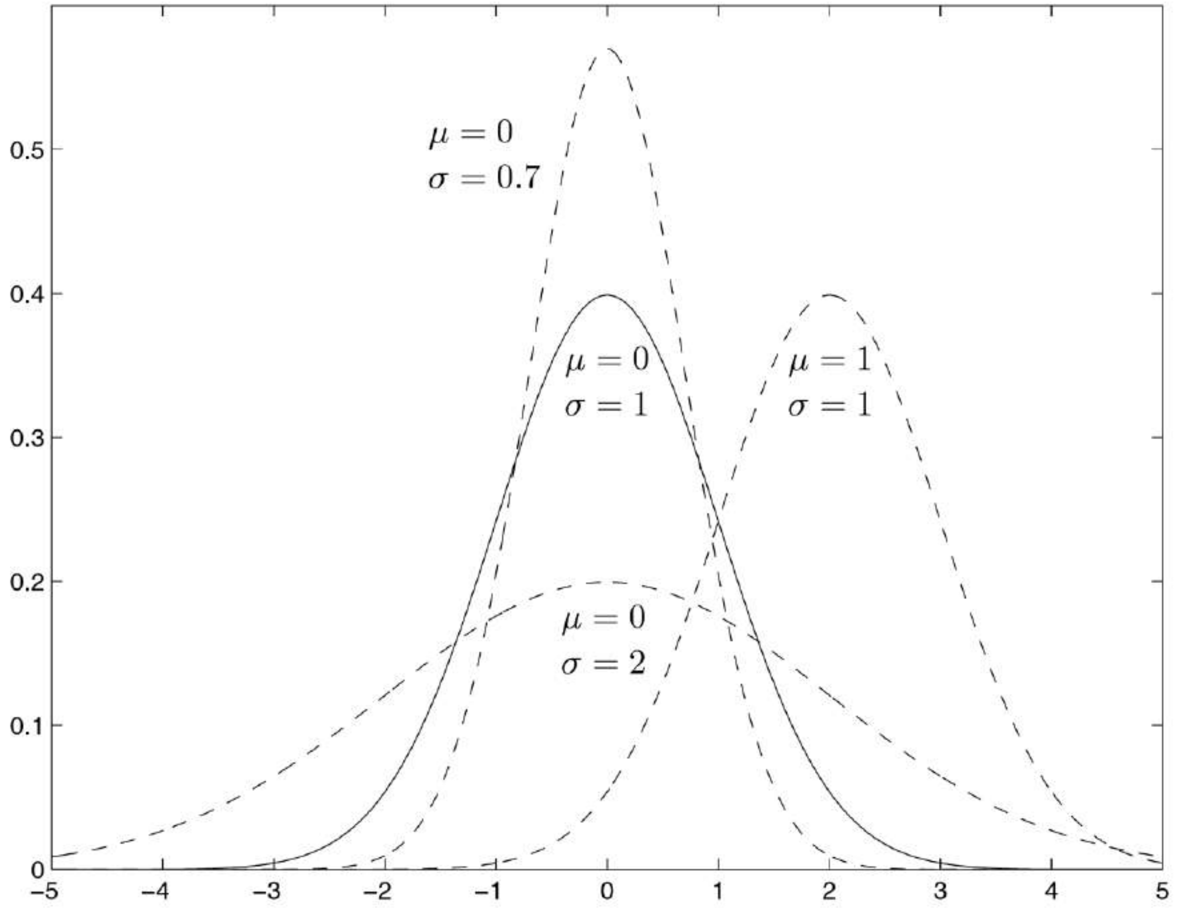


FIGURE *1 (a) with* $\mu = 0, \sigma = 1$ *and (b) with arbitrary* μ *and* σ^2.

The mean and variance of a normal random variable are given by:

$$E[X] = \mu \qquad \& \qquad Var(X) = \sigma^2$$

Changing μ shifts the curve to the left or to the right without affecting its shape, while changing σ makes it more concentrated or more flat. Often μ and σ are called location and scale parameters.



Normal densities with different location and scale parameters.

A normal random variable is called standard, or unit, normal random variable if it has mean 0 and variance 1 and it is denoted by $N(0,1)$. In case of standard r.v. we can calculate probabilities with the help of a table which is shown in the next slide.

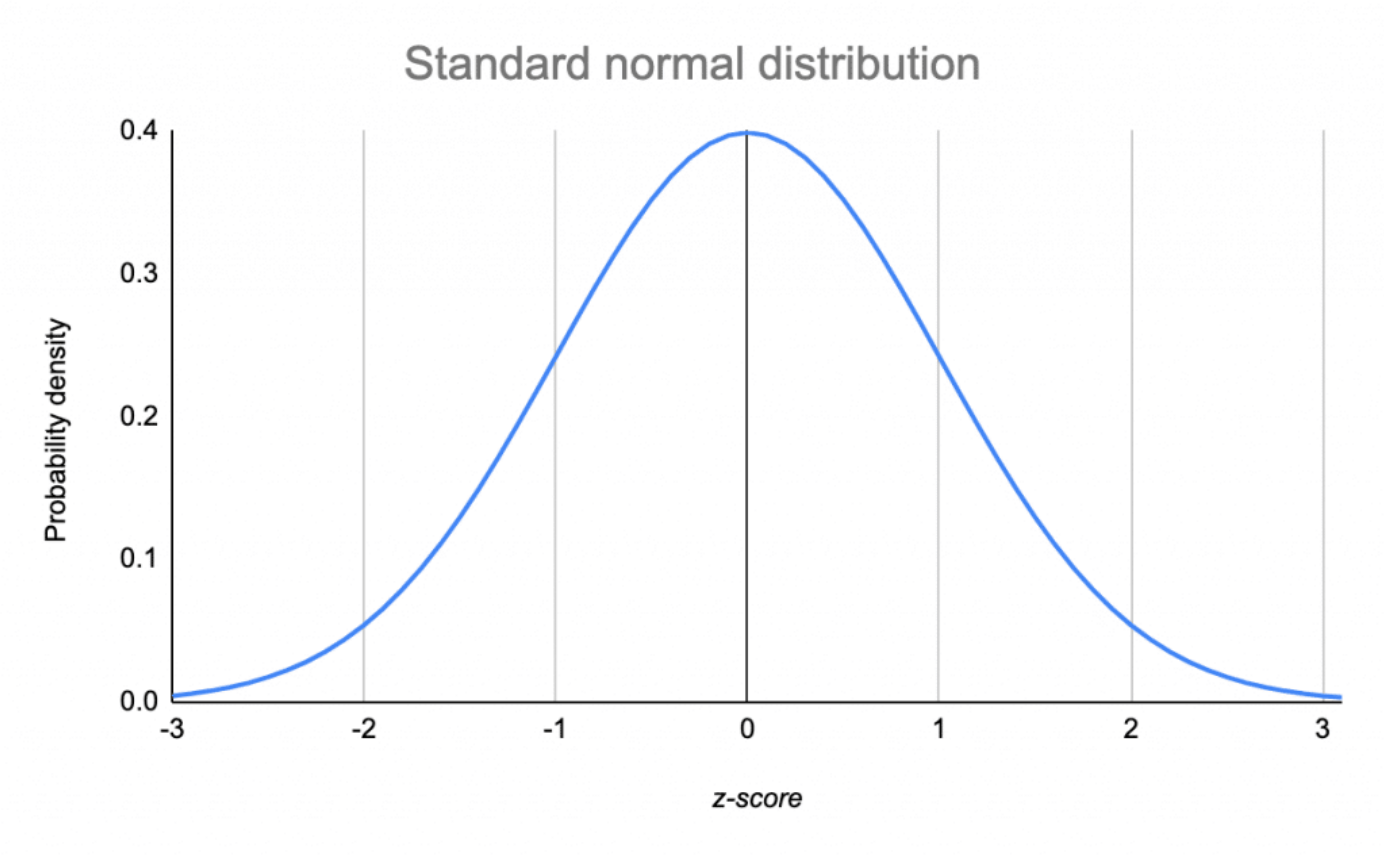


TABLE A1 *Standard Normal Distribution Function:* $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389

1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952

2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Problem

Given that Z has a normal distribution with $\mu = 0$ and $\sigma = 1$, Find (a) $P\{Z \leq 1.25\}$, (b) $P\{Z > 1.25\}$, (c) $P\{Z \leq -1.25\}$, (d) $P\{-0.38 \leq Z \leq 1.25\}$, and (e) $P\{Z \leq 5\}$.

Solution

(a) $P\{Z \leq 1.25\} = \varphi(1.25)$, a probability that is tabulated in standard normal table at the intersection of the row marked 1.2 and the column marked .05. The number there is .8944, so $P\{Z \leq 1.25\} = .8944$

(b) $P\{Z > 1.25\} = 1 - P\{Z \leq 1.25\} = 1 - \varphi(1.25) = 1 - 0.8944 = .1056$

(c) $P\{Z \leq -1.25\} = P\{Z \geq 1.25\}$ (By symmetry)

$= .1056$

(d) $P\{-0.38 \leq Z \leq 1.25\} = \varphi(1.25) - \varphi(-0.38) = 0.8944 - [1 - \varphi(0.38)] = 0.8944 - [1 - 0.6480]$

$= 0.5424$

(e) $P\{Z \leq 5\} \approx 1.$

A very important property of normal random variables is that if X is normal with mean μ and variance σ^2, then for any constants a and b, $b \neq 0$, the random variable $Y = a + bX$ is also a normal random variable with parameters

$a + b\mu$ and $b^2\sigma^2$

A Standard Normal variable, usually denoted by Z, can be obtained from a non-standard Normal(μ, σ) random variable X by *standardizing*, that is, subtracting the mean and dividing by the standard deviation,

$$Z = \frac{X - \mu}{\sigma}$$

Problem

If X is a normal random variable with mean $\mu = 3$ and variance $\sigma^2 = 16$, find

(a) $P\{X < 11\}$;

(b) $P\{X > -1\}$;

(c) $P\{2 < X < 7\}$.

Solution

(a)

$$P\{X < 11\} = P\left\{\frac{X-3}{4} < \frac{11-3}{4}\right\}$$

$$= \Phi(2)$$

$$= .9772$$

(b)

$$P\{X > -1\} = P\left\{\frac{X-3}{4} > \frac{-1-3}{4}\right\}$$

$$= P\{Z > -1\} = P\{Z < 1\}$$

$$= .8413$$

(c)

$$P\{2 < X < 7\} = P\left\{\frac{2-3}{4} < \frac{X-3}{4} < \frac{7-3}{4}\right\}$$

$$= \Phi(1) - \Phi(-1/4)$$

$$= \Phi(1) - (1 - \Phi(1/4))$$

$$= .8413 + .5987 - 1 = .4400 \quad \blacksquare$$

Atif Mehrban
Muhammad Zeeshan
Nasarumminallah
Hassaan Ahmad
Inam Ullah
MeerMurtaza
Aadil Shah
Mina Riaz
Hifza Majeed
Saqib Altaf
Noman Siddique
Ahmed Sattar
Jawad Hassan

Thank you all