Lecture Two

Section 2.1 – Basic Concepts of Probability and Addition Rule

Definition

An ***event*** is any collection of results or outcomes of a procedure

A ***simple event*** is an outcome or an event that cannot be further broken down into simpler components

The ***sample space*** for a procedure consists of all possible simple events; that is, the sample space consists of all outcomes that cannot be broken down any further

Example

We use "*f*" to denote a female baby and "*m*" to denote a male baby.

Procedure	***Example of Event***	***Complete Sample Space***
Single birth	1 female (simple event)	{*f*, *m*}
3 births	2 females and 1 male (ffm, fmf, mff) are simple events	{*fff, ffm, fmf, mff, mfm, mmf, mmm*}

Notation for Probabilities

P – denotes a probability

A, B, and C – denote specific events

$P(A)$ - denotes the probability of event A occurring.

Basic Rules for Computing Probability

***Rule* 1**: **Relative Frequency Approximation of Probability**

Conduct (or observe) a procedure, and count the number of times event A actually occurs. Based on these actual results, $P(A)$ is approximated as follows:

$$P(A) = \frac{\#\ of\ \ times\ A\ occured}{\#\ of\ \ times\ procedure\ was\ repeated}$$

***Rule* 2**: **Classical Approach to Probability** (Requires Equally Likely Outcomes)

Assume that a given procedure has n different simple events and that each of those simple events has an equal chance of occurring. If event A can occur in s of these n ways, then

$$P(A) = \frac{s}{n} = \frac{\#\ of\ \ ways\ A\ can\ occur}{\#\ of\ \ different\ simple\ events}$$

***Rule* 3**: **Subjective Probabilities**

$P(A)$, the probability of event A, is estimated by using knowledge of the relevant circumstances.

Example

Find the probability that a randomly selected car in U.S. will be in a crash this year. There were 6,511,100 cars that crashed among the 135,670,000 cars registered.

Solution

$$P(crash) = \frac{\#\ of\ cars\ that\ crashed}{Total\ \ number\ of\ cars} = \frac{6,511,100}{135,670,000}\ \underline{= 0.048|}$$

Example

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

Assuming that one of the 98 test results summarized in the table above is randomly selected, find the probability that it is a positive test result.

Solution

$$\begin{aligned} P(positive\ result) &= \frac{\#\ of\ positive\ results}{Total\ \ number\ of\ results} \\ &= \frac{42+15}{15+42+32+9} \\ &= \frac{57}{98} \\ &\underline{= 0.582|} \end{aligned}$$

Example

When studying the effect of heredity on height, we can express each individual genotype, AA, Aa, aA, and aa, on an index card and shuffle the four cards and randomly select one of them. What is the probability that we select a genotype in which the two components are different?

Solution

$$P(outcome\ with\ different\ components) = \frac{2}{4}\ \underline{= 0.5|}$$

Example

What is the probability that you will get stuck in the next elevator that you ride?

Solution

There are 2 possible outcomes (stuck or not becoming stuck). But that are not equally likely, so we cannot use the classical approach.

The leaves us with a subjective estimate, in this case, say 0.0001 (equivalent to 1 chance in ten thousand). This is likely to be in general ballpark of the true probability.

Example

Find the probability that when a couple has 3 children, they will have exactly 2 boys. Assume that boys and girls are equally likely and that the gender of any child is not influenced by the gender of any other child.

Solution

$$S = \{bbb,\ bbg,\ bgb,\ gbb,\ bgg,\ gbg,\ ggb,\ ggg\}$$

$$P(2\ boys) = \frac{3}{8}\ \underline{= 0.375}$$

Example

If a year is selected at random, find the probability that Thanksgiving Day will be

a) On a Wednesday

b) On a Thursday

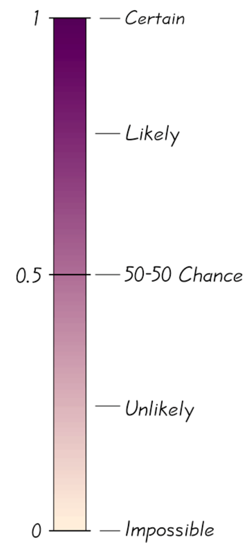
Solution

a) $P(On\ a\ Wednesday)\ \underline{= 0}$

b) It is certain that Thanksgiving Day will be on a Thursday. When an event certain to occur:

$$P(On\ a\ Thursday)\ \underline{= 1}$$

Probability Limits



Always express a probability as a fraction or decimal number between 0 and 1.

- ✓ The probability of an impossible event is 0.
- ✓ The probability of an event that is certain to occur is 1.
- ✓ For any event A, the probability of A is between 0 and 1 inclusive.

That is, $0 \le P(A) \le 1$.

Complementary Events

The complement of event A, denoted by $\bar{A}$, consists of all outcomes in which the event A ***does not*** occur.

Example

A typical question on a SAT test requires the test taker to select one of five possible choices: A, B, C, D, or E. Because only one answer is correct, if you make a random guess, your probability of being correct is $\frac{1}{5}$ *or* 0.2. Find the probability of making a random guess and not being correct (or being incorrect).

Solution

$$P(\text{not guessing the correct answer}) = P\left(\overline{\text{correct}}\right) = \frac{4}{5} \ \underline{= 0.8|}$$

or $$P\left(\overline{\text{correct}}\right) = 1 - P(\text{correct}) = 1 - \frac{1}{5} = \frac{4}{5} \ \underline{= 0.8|}$$

Rounding Off Probabilities

When expressing the value of a probability, either give the ***exact*** *fraction* or decimal or round off final decimal results to three significant digits. (*Suggestion*: When a probability is not a simple fraction such as $\frac{2}{3}$ or $\frac{5}{9}$, expresses it as a decimal so that the number can be better understood.)

- ✓ The probability of $\frac{1}{3}$ can be left as a fraction, or rounded to 0.333. (Do ***not*** round to 0.3)
- ✓ The probability of $\frac{2}{4}$ can be expressed as $\frac{1}{2}$ or 0.5; because is exact, there is no need to express as 0.500.
- ✓ $\frac{1941}{3405} \underline{= 0.570|}$

Odds

Definition

The ***actual odds against*** event A occurring are the ratio $\dfrac{P(\bar{A})}{P(A)}$, usually expressed in the form of a:b (or "a to b"), where a and b are integers having no common factors.

The ***actual odds in favor*** of event A occurring are the ratio $\dfrac{P(A)}{P(\bar{A})}$, which is the reciprocal of the actual odds against the event. If the odds against A are a:b, then the odds in favor of A are b:a.

The ***payoff odds*** against event A occurring are the ratio of the net profit (if you win) to the amount bet.

$$payoff\ \ odds\ against\ event\ A = \frac{net\ profit}{amount\ bet}$$

Example

If you bet \$5 on the number 13 in roulette, your probability of winning is $\frac{1}{38}$ and the payoff odds are given by the casino as 35:1.

a) Find the actual odds against the outcome of 13.

b) How much net profit would you make if you win by betting on 13?

c) If the casino was not operating for profit, and the payoff odds were changed to match the actual odds against 13, how much would you win if the outcome were 13?

Solution

a) With odds: $P(13) = \frac{1}{38}$ and $P(not\ 13) = \frac{37}{38}$

Actual odds against 13 $\dfrac{P(not\ 13)}{P(13)} = \dfrac{37/38}{1/38} = \dfrac{37}{1}$ or 37:1

b) Because the payoffs odds against 13 are 35:1, we have:

35:1 = (*net profit*) : (*amount bet*)

So there is a \$35 profit for each \$1 bet. For \$5 bet, the net profit is $5 \times 35 = \$175$.

The winning bettor would collect \$175 plus the original \$5 bet. That is, the total amount collected would be \$180, for the net profit of \$175.

c) If the casino were not operating for profit, the payoff odds would be equal to the actual odds against the outcome of 13, or 37:1. So there is a net profit of \$37 for each \$1 bet.

For a \$5 bet the net profit would be \$185. (The casino makes its profit by paying only \$175 instead of the \$185 that would be paid with a roulette game that is fair instead of favoring the casino.)

Definition

A ***compound event*** is any event combining 2 or more simple events

Notation for Addition Rule

$P(A \text{ or } B) = P(\text{in a single trial, event } A \text{ occurs or event } B \text{ occurs or they both occur})$

Example

If 1 subject is randomly selected from the 98 subjects given the polygraph, find the probability of selecting a subject who had a positive test result or lied.

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

<u>Solution</u>

There are 66 subjects who had a positive test result or lied.

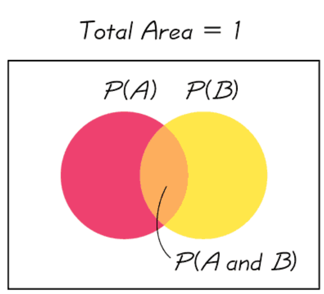
$$P(\text{positive test result or lied}) = \frac{66}{98} = 0.673$$

- When finding the probability that event *A* occurs or event *B* occurs, find the total number of ways *A* can occur and the number of ways *B* can occur, but *find that total in such a way that no outcome is counted more than once*.

Formal Addition Rule

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

where *P*(*A* and *B*) denotes the probability that *A* and *B* both occur at the same time as an outcome in a trial of a procedure.



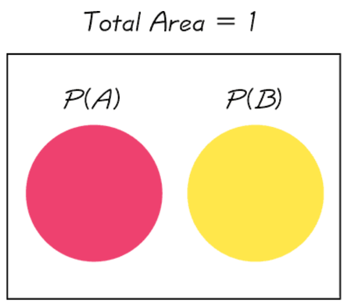
Intuitive Addition Rule

To find $P(A \text{ or } B)$, find the sum of the number of ways event A can occur and the number of ways event B can occur, *adding in such a way that every outcome is counted only once*. $P(A \text{ or } B)$ is equal to that sum, divided by the total number of outcomes in the sample space.

Disjoint or Mutually Exclusive

Definition

Events A and B are ***disjoint*** (or ***mutually exclusive***) if they cannot occur at the same time. (That is, disjoint events do not overlap.)



Example

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

a) Consider the procedure of randomly selecting 1 of the 98 subjects. Determine whether the following event are disjoint:
  A: Getting a subject with a negative test result.
  B: Getting a subject who did not lie.

b) Assuming that 1 subject is randomly selected from the 98 that were tested, find the probability of selecting a subject who had a negative test result or did not lie.

<u>Solution</u>

a) There are 41 subjects with negative test results and are 47 subjects who did lie. The event of getting a subject with a negative test result and getting a subject who did not lie can occur at the same time, there are 32 subjects. Therefore, the events are not disjoint.

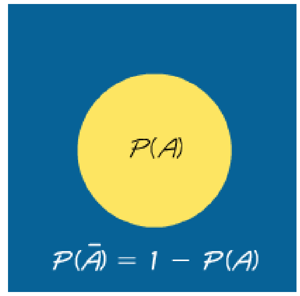
b) $P(\textit{negative test result or did not lie}) = \frac{56}{98}$

$$= 0.571$$

Complementary Events

$P(A)$ and $P\left(\overline{A}\right)$ are disjoint

It is impossible for an event and its complement to occur at the same time.



Rule of Complementary Events

$$P(A)+P\left(\overline{A}\right)=1$$

$$P\left(\overline{A}\right)=1-P(A)$$

$$P(A)=1-P\left(\overline{A}\right)$$

Example

FBI data show that 62.4% of murders are cleared by arrests. We can express the probability of a murder being cleared by an arrest as $P(cleared)=0.624$. For a randomly selected murder, find $P\left(\overline{cleared}\right)$

Solution

$$\begin{aligned} P\left(\overline{cleared}\right) &= 1-P(cleared) \\ &= 1-.624 \\ &= \underline{0.376} \end{aligned}$$

Exercises *Section* 2.1 – Basic Concepts of Probability – Addition Rule

1. Based on recent results, the probability of someone in the U.S. being injured while using sports or recreation equipment is $\frac{1}{500}$ (based on data from Statistical Abstract of the U.S.). What does it mean when we say that the probability is $\frac{1}{500}$? Is such an injury unusual?

2. When a baby is born, there is approximately a 50–50 chance that the baby is a girl. Indicate the degree of likelihood as a probability value between 0 and 1.

3. When a rolling a single die, there are 6 chances in 36 that the outcome is a 7. Indicate the degree of likelihood as a probability value between 0 and 1.

4. Identify probability values
   a) What is the probability of an event that is certain to occur?
   b) What is the probability of an impossible event?
   c) A sample space consists of 10 separate events that are equally likely. What is the probability of each?
   d) On a true/false test, what is the probability of answering a question correctly if you make a random guess?
   e) On a multiple-choice test with five possible answers for each question, what is the probability of answering a question correctly if you make a random guess?

5. When a couple has 3 children, find the probability of each event.
   a) There is exactly one girl.
   b) There are exactly 2 girls.
   c) All are girls

6. The 110^{th} Congress of the U.S. included 84 male Senators and 16 female Senators. If one of these Senators is randomly selected, what is the probability that a woman is selected? Does this probability agree with a claim that men and women have the same chance of being elected as Senators?

7. When Mendel conducted his famous genetics experiments with peas, one sample of offspring consisted of 428 green peas and 152 yellow peas. Based on those results, estimate the probability of getting an offspring pea that is green, Is the result reasonably close to the expected value of $\frac{3}{4}$, as claimed by Mendel?

8. A single fair die is rolled. Find the probability of each event
   a) Getting a 2
   b) Getting an odd number
   c) Getting a number less than 5
   d) Getting a number greater than 2
   e) Getting any number except 3

9. A jar contains 3 white, 4 orange, 5 yellow, and 8 black marbles. If a marble is drawn at random, find the probability that it is the following.
   a) White
   b) Orange
   c) Yellow
   d) Black
   e) Not black

10. The student sitting next to you in class concludes that the probability of the ceiling falling down on both of you before class ends is 1/2, because there are two possible outcomes - the ceiling will fall or not fall. What is wrong with this reasoning?

11. Let consider rolling 2 dice. Find the probabilities of the following events
   a) E = Sum of 5 turns up
   b) F = a sum that is a prime number greater than 7 turns up

12. A poll was conducted preceding an election to determine the relationship between voter persuasion concerning a controversial issue and the area of the city in which the voter lives. Five hundred registered voters were interviewed from three areas of the city. The data are shown below. Compute the probability of having no opinion on the issue or living in the inner city.

Area of city	***Favor***	***Oppose***	***No Opinion***
East	30	40	55
North	25	45	50
Inner	95	65	85

13. Suppose a single fair die is rolled. Use the sample space $S = \{1,2,3,4,5,6\}$ and give the probability of each event.
   a) E: the die shows an even number
   b) F: the die show a number less than 10
   c) G: the die shows an 8

14. A solitaire game was played 500 times. Among the 500 trials, the game was won 77 times. (The results are from the solitaire game, and the Vegas rules of "draw 3" with \$52 bet and a return of \$5 per card are used). Based on these results, find the odds against winning.

15. A roulette wheel has 38 slots. One slot is 0, another is 00, and the others are numbered 1 through 36, respectively. You place a bet that the outcome is an odd number.
   a) What is your probability of winning?
   b) What are the actual odds against winning?
   c) When you bet that the outcome is an odd number, the payoff odds are 1:1. How much profit do you make if you bet \$18 and win?
   d) How much profit would you make on the \$18 bet if you could somehow convince the casino to change its payoff odds so that they are the same as the actual odds against winning?

16. Women have a 0.25% rate of red/green color blindness. If a woman is randomly selected, what is the probability that she does not have red/green color blindness?

17. A pew Research center poll showed that 79% of Americans believe that it is morally wrong to not report all income on tax returns. What is the probability that an American does not have that belief?

18. When the author observed a sobriety checkpoint conducted by the Dutchess County Sheriff Department, he saw that 676 drivers were screened and 6 were arrested for driving while intoxicated. Based on those results, we can estimate that $P(I) = 0.00888$, where I denotes the event of screening a driver and getting someone who is intoxicated. What does $P(\bar{I})$ denote and what is its value?

19. Use the polygraph test data

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

a) If one of the test subjects is randomly selected, find the probability that the subject had a positive test result or did not lie

b) If one of the test subjects is randomly selected, find the probability that the subject did not lie

c) If one of the test subjects is randomly selected, find the probability that the subject had a true negative test result

d) If one of the test subjects is randomly selected, find the probability that the subject had a negative test result or lied.

20. Use the data

Was the challenge to the call successful?		
	Yes	***No***
Men	201	288
Women	126	224

a) If S denotes the event of selecting a successful challenge, find $P(\bar{S})$

b) If M denotes the event of selecting a challenge made by a man, find $P(\bar{M})$

c) Find the probability that the selected challenge was made by a man or was successful.

d) Find the probability that the selected challenge was made by a woman or was successful.

e) Find $P(\textit{challenge was made by a man or was not successful})$

f) Find $P(\textit{challenge was made by a woman or was not successful})$

21. Refer to the table below

	Age					
	18 – 21	22 – 29	30 – 39	40 – 49	50 – 59	60 and over
Responded	73	255	245	136	138	202
Refused	11	20	33	16	27	49

a) What is the probability that the selected person refused to answer? Does that probability value suggest that refusals are a problem for pollsters? Why or why not?
b) A pharmaceutical company is interested in opinions of the elderly, because they are either receiving Medicare or will receive it soon. What is the probability that the selected subject is someone 60 and over who responded?
c) What is the probability that the selected person responded or is in the 18–21 age bracket?
d) What is the probability that the selected person refused or is over 59 years of age?
e) A market researcher is interested in responses, especially from those between the ages of 22 and 39, because they are the people more likely to make purchases. Find the probability that a selected responds or is aged between the ages of 22 and 39.
f) A market researcher is not interested in refusals or subjects below 22 years of age or over 59. Find the probability that the selected person refused to answer or is below 22 or is older than 59.

22. Two dice are rolled. Find the probabilities of the following events.
a) The first die is 3 or the sum is 8
b) The second die is 5 or the sum is 10.

23. One card is drawn from an ordinary of 52 cards. Find the probabilities of drawing the following cards
a) A 9 or 10
b) A red card or a 3
c) A 9 or a black 10
d) A heart or a black card
e) A face card or a diamond

24. One card is drawn from an ordinary of 52 cards. Find the probabilities of drawing the following cards
a) Less than a 4 (count aces as ones)
b) A diamond or a 7
c) A black card or an ace
d) A heart or a jack
e) A red card or a face card

25. Pam invites relatives to a party: her mother, 2 aunts, 3 uncles, 2 brothers, 1 male cousin, and 4 female cousins. If the chances of any one guest first equally likely, find the probabilities that the first guest to arrive is as follows.
a) A brother or an uncle
b) A brother or a cousin
c) A brother or her mother
d) An uncle or a cousin
e) A male or a cousin
f) A female or a cousin

26. Suppose $P(E)=0.26$, $P(F)=0.41$, and $P(E\cap F)=0.16$. Find the following

a) $P(E\bigcup F)$

b) $P(E'\bigcap F)$

c) $P(E\bigcap F')$

d) $P(E'\bigcup F')$

27. Suppose $P(E)=0.42$, $P(F)=0.35$, and $P(E\cup F)=0.59$. Find the following

a) $P(E'\cap F')$ *b)* $P(E'\cup F')$ *c)* $P(E'\cup F)$ *d)* $P(E\cap F')$

28. From survey involving 1,000 people in the certain city, it was found that 500 people had tried a certain brand of diet cola, 600 had tried a certain brand of regular cola, and 200 had tried both types of cola. If a resident of the city is selected at random, what is the empirical probability that

a) The resident has not tried either cola? What are the empirical odds for this event?

b) The resident has tried the diet or has not tried the regular cola? What are the empirical odds against this event?

Section 2.2 – Multiplication Rule: Basics, Complements and Conditional

Notation

$$P(A \text{ and } B) = P(\text{event } A \text{ occurs in the first trial and event } B \text{ occurs in a second trial})$$

$$P(A \text{ and } B) = P(A) \bullet P(B)$$

Example

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

If two of the subjects are randomly selected without replacement, find the probability that the first selected person had a positive test result and the second selected person had a negative test result.

Solution

$$P(\text{positive test result}) = \frac{57}{98}$$

After the first selection of a subject, there are 97 subjects remaining

$$P(\text{negative test result}) = \frac{41}{97}$$

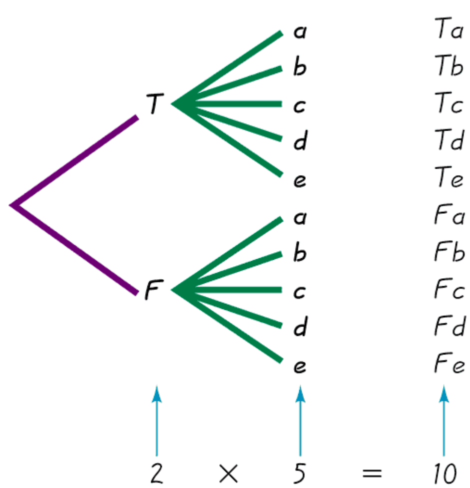
$$P(1st \text{ positive and } 2nd \text{ negative test result}) = \frac{57}{98} \cdot \frac{41}{97} \underline{= 0.246}$$

Tree Diagrams

A tree diagram is a picture of the possible outcomes of a procedure, shown as line segments emanating from one starting point. These diagrams are sometimes helpful in determining the number of possible outcomes in a sample space, if the number of possibilities is not too large.

This figure summarizes the possible outcomes for a true/false question followed by a multiple choice question.

Note *that there are 10 possible combinations.*



Conditional Probability *Key Point*

We must adjust the probability of the second event to reflect the outcome of the first event.

The probability for the second event *B* should take into account the fact that the first event *A* has already occurred.

Notation for Conditional Probability

$P(B|A)$ represents the probability of event *B* occurring after it is assumed that event *A* has already occurred (read $B|A$ as "*B* given *A*.")

Definition

Two events *A* and *B* are ***independent*** if the occurrence of one does not affect the *probability* of the occurrence of the other. (Several events are similarly independent if the occurrence of any does not affect the probabilities of the occurrence of the others.)
If *A* and *B* are not independent, they are said to be ***dependent***.

Two events are ***dependent*** if the occurrence of one of them affects the *probability* of the occurrence of the other, but this does not necessarily mean that one of the events is a ***cause*** of the other.

Formal Multiplication Rule

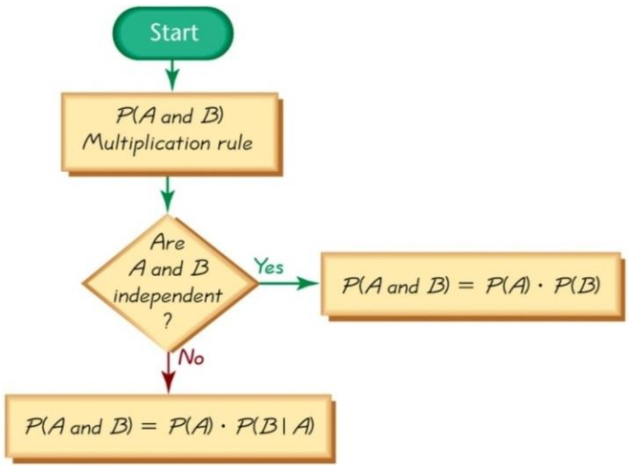
$$P(A \text{ and } B) = P(A) \cdot P(B|A)$$

if *A* and *B* are independent events, $P(B|A)$ is really the same as $P(B)$.

Intuitive Multiplication Rule

When finding the probability that event *A* occurs in one trial and event *B* occurs in the next trial, multiply the probability of event *A* by the probability of event *B*, but be sure that the probability of event *B* takes into account the previous occurrence of event *A*.

✓ *When applying the multiplication rule, always consider whether the events are independent or dependent, and adjust the calculations accordingly.*



Example

Assume that we have a batch of 100,000 heart pacemakers, including 99,950 that are good (G) and 50 that are defective (D).

a) If two of those 100,000 pacemakers are randomly selected without replacement, find the probability that they are both good.

b) If 20 of those 100,000 pacemakers are randomly selected without replacement, find the probability that they are both good.

Solution

a) $P(1st\ good) = \frac{99,950}{100,000}$ $\qquad$ $P(2nd\ good) = \frac{99,949}{100,000}$

$$P(1st\ good\ and\ 2nd\ good) = \frac{99,950}{100,000} \cdot \frac{99,949}{100,000}$$

$$\underline{= 0.999}$$

b) $P(all\ 20\ pacemakers\ are\ good) = \frac{99,950}{100,000} \cdot \frac{99,949}{100,000} \cdot \frac{99,948}{100,000} \cdot \ldots \cdot \frac{99,931}{100,000}$

Example

Assume that two people are randomly selected and also assume that birthdays occur on the same days of the week with equal frequencies.

a) Find the probability that the two people are born on the same day of the week.

b) Find the probability that the two people are born on Monday.

Solution

a) Because no particular day of the week is specified, the first person can be born on any one of the seven week days.

The probability that the second person is born on the same day as the first person is $\frac{1}{7}$.

Probability that 2 people are born on the same day of the week is $\frac{1}{7}$

b) $P(1st\ born\ on\ Monday) = \frac{1}{7}$

$$P(2nd\ born\ on\ Monday) = \frac{1}{7}$$

$$P(both\ born\ on\ Monday) = \frac{1}{7} \cdot \frac{1}{7}$$

$$\underline{= \frac{1}{49}}$$

Example

A geneticist developed a procedure for increasing the likelihood of female babies. In an initial test, 20 couples use the method and the results consist of 20 females among 20 babies. Assuming that the gender-selection procedure has no effect, find the probability of getting 20 females among 20 babies by chance. Does the resulting provide strong evidence to support the geneticist's claim that the procedure is effective in increasing the likelihood that babies will be females?

Solution

$$\begin{aligned} P(\textit{all } 20 \textit{ are female}) &= P(1st \textit{ is female and } 2nd \textit{ female } \cdots \textit{ and } 20th \textit{ is female}) \\ &= P(\textit{female}) \bullet P(\textit{female}) \bullet \cdots \bullet P(\textit{female}) \\ &= (0.5) \bullet (0.5) \bullet \cdots \bullet (0.5) \\ &= (0.5)^{20} \\ &= 0.000000954 \end{aligned}$$

The low probability of 0.000000954 indicates that instead of getting 20 females by chance, a more reasonable explanation is that females appear to be more likely with the gender-selection procedure. Because there is such a small probability of getting 20 females in 20 births, we do have to support the geneticist's claim that the gender-selection procedure is effective in increasing the likelihood that babies will be female.

Example

Modern aircraft engines are now highly reliable. One design feature contributing to that reliability is the use of redundancy, whereby critical components are duplicated so that if one fails, the other will work. For example, single-engine aircraft now have two independent electrical systems so that is one electrical system fails, the other can continue to work so that the engine does not fail. For the purposes of this example, we will assume that the probability of an electrical system failure is 0.001.

a) If the engine in an aircraft has one electrical system, what is the probability that it will work?

b) If the engine in an aircraft has 2 independent electrical systems, what is the probability that the engine can function with a working electrical system?

Solution

a) $P(\textit{electrical system failure}) = 0.001 \qquad P(\textit{doe not fail}) = 1 - 0.001 = 0.999$

$$\begin{aligned} P(\textit{working electrical system}) &= P(\textit{electrical system does not fail}) \\ &= 0.999 \end{aligned}$$

b) $$\begin{aligned} P(\textit{both electrical system fail}) &= P(1st \textit{ electrical system fails and } 2nd \textit{ electrical system fails}) \\ &= (.001)(.001) \\ &= 0.000001 \end{aligned}$$

There is a 0.000001 probability of both electrical systems failing, so the probability that the engine can function with a working electrical system is 1 – 0.000001 = 0.999999

Complements: The Probability of "*At Least One*"

- "At least one" is equivalent to "one or more."
- The complement of getting at least one item of a particular type is that you get no items of that type.

Finding the Probability of "*At Least One*"

To find the probability of at least one of something, calculate the probability of none, then subtract that result from 1. That is,

$$P(\text{at least one}) = 1 - P(\text{none})$$

Example

Find the probability of a couple having at least 1 girl among 3 children. Assume that boys and girls are equally likely and that the gender of a child is independent of any other child.

Solution

Let A = at least 1 of the 3 children is a girl.

$\overline{A}$ = not getting at least 1 girl among 3 children
= all 3 children are boys
= boy and boy and boy

$$\begin{aligned} P(\overline{A}) &= P(\text{boy and boy and boy}) \\ &= \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} \\ &= \frac{1}{8} \end{aligned}$$

$$\begin{aligned} P(A) &= 1 - P(\overline{A}) \\ &= 1 - \frac{1}{8} \\ &= \frac{7}{8} \end{aligned}$$

There is $\frac{7}{8}$ probability that if a couple has 3 children; at least 1 of them is a girl.

Example

Assume that the probability of a defective Firestone tire is 0.0003 (based on data from Westgard (QC). If the retail outlet Car Stuff buys 100 Firestone tires, find the probability that they get at least 1 that is defective. If that probability is high enough, plans must be made to handle defective tires returned by consumers. Should they make those plans?

Solution

Let A = at least 1 of the 100 tires is defective.

$\bar{A}$ = not getting at least 1 defective among 100 tires
= all 100 tires are good

$$\begin{aligned}P\left(\bar{A}\right) &= (1-.0003)(1-.0003)\ldots(1-.0003)\\ &= (0.9997)(0.9997)\ldots(0.9997)\\ &= (0.9997)^{100}\\ &= .9704\end{aligned}$$

$$\begin{aligned}P(A) &= 1-P\left(\bar{A}\right)\\ &= 1-.9704\\ &= .0296\end{aligned}$$

There is 0.0296 probability of at least 1 defective tire among the 100 tires. Because the probability is so low, it is not necessary to make plans for dealing with defective tires returned by consumers.

Conditional Probability

A ***conditional probability*** of an event is a probability obtained with the additional information that some other event has already occurred. $P(B|A)$ denotes the conditional probability of event B occurring, given that event A has already occurred, and it can be found by dividing the probability of events A and B both occurring by the probability of event A:

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

The conditional probability of B given A can be found by assuming that event A has occurred, and then calculating the probability that event B will occur.

Example

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

a) If 1 of the 98 test subjects is randomly selected, find the probability that the subject had a positive test result, given that the subject actually lied. That is, find *P*(positive test result | subject lied)

b) If 1 of the 98 test subjects is randomly selected, find the probability that the subject actually lied, given that he or she had a positive test result. That is, find *P* (subject lied | positive test result)

Solution

a) There are (42 + 9 =) 51 subjects that they lied, 42 has a positive test results.

$$P(\text{positive test result} | \text{lied}) = \frac{42}{51} \underline{= 0.824|}$$

Or

$$P(\text{positive test result} | \text{lied}) = \frac{P(\text{lied and had a positive test result})}{P(\text{subject lied})}$$

$$= \frac{\frac{42}{98}}{\frac{51}{98}}$$

$$\underline{= 0.824|}$$

This indicates that a subject who lies has a 0.824 probability of getting a positive test result.

b) There are (42 + 15 =) 57 with a positive test results subjects among that 42 lied.

$$P(\text{subject lied} | \text{positive test result}) = \frac{42}{57} \underline{= 0.737|}$$

This indicates that for a subject who gets a positive test result, there is a 0.737 probability that this subject actually lied.

Confusion of the Inverse

To incorrectly believe that $P(A|B)$ and $P(B|A)$ are the same, or to incorrectly use one value for the other, is often called confusion of the inverse.

Bayes' Theorem

Rare Event Rule for Inferential Statistics

If, under a given assumption, the probability of a particular observed event is extremely small, we conclude that the assumption is probably not correct.

Bayes' Formula

$$
\begin{aligned}
P\left(U_1 \mid E\right) &= \frac{P\left(U_1 \cap E\right)}{P(E)} \\
&= \frac{P\left(U_1 \cap E\right)}{P\left(U_1 \cap E\right) + P\left(U_2 \cap E\right) + \cdots} \\
&= \frac{P\left(E \mid U_1\right) P\left(U_1\right)}{P\left(E \mid U_1\right) P\left(U_1\right) + P\left(E \mid U_2\right) P\left(U_2\right) + \cdots}
\end{aligned}
$$

Exercises Section 2.2 – Multiplication Rule: Basics, Complements and Conditional

1. Use the data below:

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

a) If 2 of the 98 test subjects are randomly selected without replacement find the probability that they both had false positive results. Is it unusual to randomly select 2 subjects without replacement and get 2 results that are both false positive results? Explain.

b) If 3 of the 98 test subjects are randomly selected without replacement, find the probability that all had false positive results. Is it unusual to randomly select 3 subjects without replacement and get 3 results that are all false positive results? Explain.

c) If 4 of the test subjects are randomly selected without replacement find the probability that, in each case, the polygraph indicated that the subject lied. Is such an event unusual?

d) If 4 of the test subjects are randomly selected without replacement find the probability that they all had incorrect test result (either false positive or false negative). Is such an event Likely?

e) Assume that 1 of the 98 test subjects is randomly selected. Find the probability of selecting a subject with a negative test result, given that the subject lied. What does this result suggest about the polygraph test?

f) Find $P(\textit{negative test result} \mid \textit{subject did not lie})$

g) Find $P(\textit{subject did not lie} \mid \textit{negative test result})$

2. Use the data in the table below

	Group			
Type	***O***	***A***	***B***	***AB***
Rh^+	39	35	8	4
Rh^-	6	5	2	1

a) If 2 of the 100 subjects are randomly selected, find the probability that they are both group O and type Rh^+

i. Assume that the selections are made with replacement.
ii. Assume that the selections are made without replacement.

b) If 3 of the 100 subjects are randomly selected, find the probability that they are both group B and type Rh^-

i. Assume that the selections are made with replacement.
ii. Assume that the selections are made without replacement.

c) People with blood that is group O and type Rh^{-} are considered to be universal donors, because they can give blood to anyone. If 4 of the 100 subjects are randomly selected, find the probability that they are all universal donors.

   i. Assume that the selections are made with replacement.
   ii. Assume that the selections are made without replacement.

d) People with blood that is group AB and type Rh^{+} are considered to be universal donors, because they can give blood to anyone. If 3 of the 100 subjects are randomly selected, find the probability that they are all universal recipients.

   i. Assume that the selections are made with replacement.
   ii. Assume that the selections are made without replacement.

3. With one method of a procedure called acceptance sampling, a sample of items is randomly selected without replacement and the entire batch is accepted if every item in the sample is okay. The Telektronics Company manufactured a batch of 400 backup power supply units for computers, and 8 of them are defective. If 3 of the units are randomly selected for testing, what is the probability that the entire batch will be accepted?

4. It is common for public opinion polls to have a "confidence level" of 95% meaning that there is a 095 probability that the poll results are accurate within the claimed margins of error. If each of the following organizations conducts an independent poll, find the probability that all of them are accurate within the claim margins of error: Gallup, Roper, Yankelovich, Harris, CNN, ABC, CBS, and NBC, New York Times. Does the result suggest that with a confidence level of 95%, we can expect that almost all polls will be within the claimed margin of error?

5. The principle of redundancy is used when system reliability is improved through redundant or back up components. Assume that your alarm clock has a 0.9 probability of working on any given morning.

   a) What is the probability that your alarm clock will not work on the morning of an important final exam?
   b) If you have 2 such alarm clocks, what is the probability that they both fail on the morning of an important final exam?
   c) With one alarm clock, you have a 0.9 probability of being awakened. What is the probability of being awakened if you use two alarm clocks?
   d) Does a second alarm clock result in greatly improved reliability?

6. The wheeling Tire Company produced a batch of 5,000 tires that includes exactly 200 that are defective.

   a) If 4 tires are randomly selected for installation on a car, what is the probability that they are all good?
   b) If 100 tires are randomly selected for shipment to an outlet, what is the probability that they are all good? Should this outlet plan to deal with defective tires returned by consumers?

7. When the 15 players on the LA Lakers basketball team are tested for steroids, at least one of them tests positive. Provide a written description of the complement of this event.
8. If a couple plans to have 6 children, what is the probability that they will have at least one girl? Is that probability high enough for the couple to be very confident that they will get at least one girl in six children?

9. If a couple plans to have 8 children (it could happen), what is the probability that they will have at least one girl? Is the couple eventually has 8 children and they are all boys, what can the couple conclude?

10. If you make guesses for 4 multiple-choice test questions (each with 5 possible answers), what is the probability of getting at least one correct? If a very lenient instructor says that passing test occurs if there is at least one correct answer, can you reasonably expect to pass by guessing?

11. Find the probability of a couple having a baby girl when their fourth child is born, given that the first 3 children were all girls. Is the result the same as the probability of getting 4 girls among 4 children?

12. In China, the probability of a baby being a boy is 0.5845. Couples are allowed to have only one child. If relatives give birth to 5 babies, what is the probability that there is at least one girl? Can that system continue to work indefinitely?

13. An experiment with fruit flies involves one parent with normal wings and one parent with vestigial wings. When these parents have an offspring, there is a $\frac{3}{4}$ probability that the offspring has normal wings and a $\frac{1}{4}$ probability of vestigial wings. If the parents give birth to 10 offspring, what is the probability that at least 1 of the offspring has vestigial wings? If researchers need at least one offspring with vestigial wings, can they be reasonably confident of getting one?

14. According to FBI data, 24.9% of robberies are cleared with arrests. A new detective is assigned to 10 different robberies.
    a) What is the probability that at least 1 of them is cleared with an arrest?
    b) What is the probability that the detective clears all 10 robberies with arrests?
    c) What should we conclude if the detective clears all 10 robberies with arrests?

15. A statistics student wants to ensure that she is not late for an early statistics class because of a malfunctioning alarm clock. Instead of using one alarm clock, she decides to use three. What is the probability that at least one of her alarm clocks works correctly if each individual alarm clock has a 90% chance of working correctly? Does the student really gain much by using three alarm clocks instead on only one? How are the results affected if all of the alarm clocks run on electricity instead of batteries?

16. In a batch of 8,000 clock radios 8% are defective. A sample of 5 clock radios is randomly selected without replacement from the 8,000 and tested. The entire batch will be rejected if at least one of those tested is defective. Find the probability that the entire batch will be rejected.

17. In a blood testing procedure, blood samples from 3 people are combined into one mixture. The mixture will only test negative if all the individual samples are negative. If the probability that an individual sample tests positive is 0.1, find the probability that the mixture will test positive.

18. A sample of 4 different calculators is randomly selected from a group containing 16 that are defective and 36 that have no effects. Find the probability that at least one of the calculator is defective.

19. Among the contestants in a competition are 46 women and 29 men. If 5 winners are randomly selected, find the probability that they are all men?

20. A bin contains 60 lights bubs of which 7 are defective. If 4 light bulbs are randomly selected from the bin with replacement, find the probability that all the bulbs selected are good ones.

21. You are dealt two cards successively (without replacement) from a shuffled deck of 52 playing cards. Find the probability that both cards are black. Express your answer as a simplified fraction.

Section 2.3 – Counting

Fundamental Counting Rule

For a sequence of two events in which the first event can occur *m* ways and the second event can occur *n* ways, the events together can occur a total of $m \bullet n$ ways.

Example

It's wise not to disclose social security numbers, because they are often used by criminals attempting identity theft. Assume that a criminal is found using your social security number and claims that all of the digits were randomly generated. What is the probability of getting your social security number when randomly generated 9 digits? Is the criminal's claim that your number was randomly generated likely to be true?

Solution

Each of the 9 digits has 10 possible outcomes: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

$$10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 1,000,000,000$$

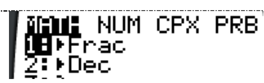
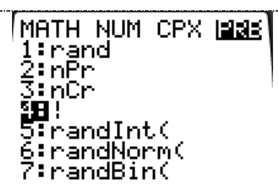
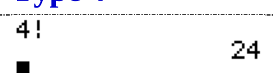
Notation

The ***factorial symbol*** (!) denotes the product of decreasing positive whole numbers.

For example,

$0! = 1$

$4! = 4 \bullet 3 \bullet 2 \bullet 1 = 24$

4	Math	PRB	Type 4	
				

Factorial Rule

A collection of *n* different items can be arranged in order *n*! different ways. (This factorial rule reflects the fact that the first item may be selected in *n* different ways, the second item may be selected in $n - 1$ ways, and so on.)

Example

During the summer, you are planning to visit these 6 national parks: Glacier, Yellowstone, Yosemite, Arches, Zion, and Grand Canyon. You would like to plan the most efficient route and you decide to list all of the possible routes. How many different routes are possible?

Solution

There 6 different parks can be arranged in order 6! different ways.

$$6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720 \ \ \text{ways}$$

Permutations Rule (*when items are all different*)

Requirements:

1. There are ***n* different** items available. (This rule does not apply if some of the items are identical to others.)
2. We select ***r*** of the ***n*** items (without replacement).
3. We consider rearrangements of the same items to be different sequences. (The permutation of ABC is different from CBA and is counted separately.)

If the preceding requirements are satisfied, the number of permutations (or sequences) of *r* items selected from *n* available items (without replacement) is

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

Example

In horse racing, a bet on an exacta in a race is won by correctly selecting the horses that finish first and second, and you must select those 2 horses in the correct order. The 132nd running of the Kentucky Derby has a field of 20 horses. If a bettor randomly selects 2 of those horses for an exacta bet, what is the probability of winning?

Solution

We have n = 20 horses and we must select r = 2 of them without replacement.

$$\begin{aligned} {}_{20}P_2 &= \frac{n!}{(n-r)!} \\ &= \frac{20!}{(20-2)!} \\ &= 380 \end{aligned}$$

Permutations Rule (*when some items are identical to others*)

Requirements:

1. There are *n* items available, and some items are identical to others.
2. We select all of the *n* items (without replacement).
3. We consider rearrangements of distinct items to be different sequences.

If the preceding requirements are satisfied, and if there are n_1 alike, n_2 alike, . . . , n_k alike, the number of permutations (or sequences) of all items selected without replacement is

$$\frac{n!}{n_1! \cdot n_2! \cdots n_k!}$$

Example

In a preliminary test of the MicroSort gender selection method developed by the Genetics and IVF Institute, 14 couples tried to have baby girls. Analysis of the effectiveness of the MicroSort method is based on a probability value, which in turn is based on numbers of permutations. Let's consider this simple problem: How many ways can 11 girls and 3 boys be arranged in sequence? That is, find the number of permutations of 11 girls and 3 boys.

<u>Solution</u>

$n = 14; \quad n_1 = 11; \quad n_2 = 3$

$$\frac{n!}{n_1! \cdot n_2!} = \frac{14!}{11! \cdot 3!}$$

$$= \frac{12 \cdot 13 \cdot 14}{1 \cdot 2 \cdot 3}$$

$$= 364$$

There are 364 different ways to arrange 11 girls and 3 boys.

Example

In how many ways can the letters in the word *Mississippi* be arranged?

<u>Solution</u>

$$\frac{11!}{1!4!4!2!} = 34,650 \ \ ways$$

m	*i*	*s*	*p*
1	4	4	2

Combinations Rule

Requirements:

1. There are *n* different items available.
2. We select *r* of the *n* items (without replacement).
3. We consider rearrangements of the same items to be the same. (The combination of ABC is the same as CBA.)

If the preceding requirements are satisfied, the number of combinations of *r* items selected from *n* different items is

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

Permutations versus Combinations

When different orderings of the same items are to be counted separately, we have a permutation problem, but when different orderings are not to be counted separately, we have a combination problem.

Permutation: *order matter*.

Combination: *Order doesn't matter*.

Example

A clinical rest on humans of a new drug is normally done in 3 phases. Phase I is conducted with a relatively small number of healthy volunteers. Let's assume that we want to treat 8 healthy humans with a new drug, and we have 10 suitable volunteers available.

a) If the subjects are selected and treated in sequence, so that the trial is discontinued if anyone displays adverse effects, how many different sequential arrangements are possible if 8 people are selected from the 10 that are available?

b) If 8 subjects are selected from the 10 that are available, and the 8 selected subjects are all treated at the same time, how many different treatment groups are possible?

Solution

a) Because order does count, we want the number of permutations of *r* = 8 people selected from the $n = 10$.

$$_{10}P_8 = \frac{10!}{(10-8)!} \underline{= 1,814,400|}$$

b) Because order does ***not*** count, we want the number of combinations of *r* = 8 people selected from the $n = 10$.

$$_{10}C_8 = \frac{10!}{(10-8)!8!} \underline{= 45|}$$

Example

The Florida Lotto game is typical of state lotteries. You must select 6 different numbers between 1 and 53. You win the jackpot if the same 6 numbers are drawn in any order. Find the probability of winning the jackpot.

Solution

Because order does ***not*** count, we want the number of combinations of $r = 6$ people selected from the $n = 53$

$${}_{53}C_6 = \frac{53!}{(53-6)!\cdot 6!} \ \boxed{= 22,957,480}$$

With 1 winning combination and 22,957,480 different possible combinations, the probability of winning the jackpot is

$$P = \boxed{\frac{1}{22,957,480}}$$

Exercises Section 2.3 – Counting

1. Find the number of different ways that five test questions can be arranged in order by evaluating 5!

2. In the game of blackjack played with one deck, a player is initially dealt 2 cards. Find the number of different two-card initial hands by evaluating ${}_{52}C_2$

3. A political strategist must visit state capitols, but she has time to visit only 3 of them. Find the number of different possible routes by evaluating ${}_{50}P_3$

4. Select the six winning numbers from 1, 2, …, 54. Find the probability of winning lottery by buying one ticket. $\left(\textit{of winning this lottery } \dfrac{1}{575,757} \right)$

5. Select the five winning numbers from 1, 2, …, 36. Find the probability of winning lottery by buying one ticket. $\left(\textit{of winning this lottery } \dfrac{1}{575,757} \right)$

6. In a club with 9 male and 11 female members, how many 5-member committees can be chosen that have
   a) All men?
   b) All women?
   c) 3 men and 2 women?

7. In a club with 9 male and 11 female members, how many 5-member committees can be selected that have
   a) At least 4 women?
   b) No more than 2 men?

8. In how many ways can 5 out of 9 plants be arranged in a row on a windowsill?

9. From a pool of 8 secretaries, 3 are selected to be assigned to 3 managers, one per manager. In how many ways can they be selected and assigned?

10. A group of 9 workers decides to send a delegation of 3 to their supervisor to discuss their grievances.
    a) How many delegations are possible?
    b) If it is decided that a particular worker must be in the delegation, how many different delegations are possible?
    c) If there are 4 women and 5 men in the group, how many delegations would include at least 1 woman?

11. Hamburger Hut sells regular hamburgers as well as a larger burger. Either type can include cheese, relish, lettuce, tomato, mustard, or catsup.
    a) How many different hamburgers can be ordered with exactly three extras?
    b) How many different regular hamburgers can be ordered with exactly three extras?
    c) How many different regular hamburgers can be ordered with at least five extras?

12. In an experiment on plant hardiness, a researcher gathers 6 wheat plants, 3 barley plants, and 2 rye plants. She wishes to select 4 plants at random.
    a) In how many ways can this be done?
    b) In how many ways can this be done if exactly 2 wheat plants must be included?

13. A legislative committee consists of 5 Democrats and 4 Republicans. A delegation of 3 is to be selected to visit a small Pacific island republic.
    a) How many different delegations are possible?
    b) How many delegations would have all Democrats?
    c) How many delegations would have 2 Democrats and 1 Republican?
    d) How many delegations would have at least 1 Republican?

14. Five cards are chosen from an ordinary deck to form a hand in poker. In how many ways is it possible to get the following results?
    a) 4 queens
    b) No face card
    c) Exactly 2 face cards
    d) At least 2 face cards
    e) 1 heart, 2 diamonds, and 2 clubs

15. The student sitting next to you in class concludes that the probability of the ceiling falling down on both of you before class ends is 1/2, because there are two possible outcomes - the ceiling will fall or not fall. What is wrong with this reasoning?

16. Identity theft often begins by someone discovering your 9-digit social security number. Answer each of the following. Express probabilities as fractions.
    a) What is the probability of randomly generating 9 digits and getting your social security number?
    b) In the past, many teachers posted grades along with the last 4 digits of your social security number, what is the probability that if they randomly generated the order digits, they would match yours? Is that something to worry about?

17. Credit card numbers typically have 16 digits, but not all of them are random. Answer the following and express probabilities as fractions.
    a) What is the probability of randomly generating 16 digits and getting your MasterCard number?
    b) Receipts often show the last 4 digits of a credit card number. If those last 4 digits are known, what is the probability of randomly generating the order digits of your MasterCard number?
    c) Discover cards begin with the digits 6011. If you also know the last 4 digits, what is the probability of randomly generating the other digits and getting all of them correct? Is this something to worry about?

18. When testing for current in a cable with five color-coded wires, the author used a meter to test two wires at a time. How many different tests are required for every possible pairing of two wires?

19. The starting 4 players for the Boston Celtics basketball team have agreed to make charity appearances tomorrow night. If you must send three players to the United Way event and the other 2 to a Heart Fund event, how many different ways can you make the assignments?

20. In phase I of a clinical trial with gene therapy used for treating HIV, 5 subjects were treated (based on data from Medical News Today). If 20 people were eligible for the Phase I treatment and a simple random of 5 is selected, how many different simple random samples are possible? What is the probability of each simple random sample?

21. Many newspapers carry "Jumble" a puzzle in which the reader must unscramble letters to form words. The letters BUJOM were included in newspapers. How many ways can the letters if BUJOM be arranged? Identify the correct unscrambling, then determine the probability of getting that result by randomly selecting one arrangement of the given letters.

22. There are 11 members on the board of directors for the Coca Cola Company.
    a) If they must select a chairperson, first vice chairperson, second vice chairperson, and secretary, how many different slates of candidates are possible?
    b) If they must form an ethics subcommittee of 4 members, how many different subcommittees are possible?

23. The author owns a safe in which he stores his book. The safe combination consists of 4 numbers between 0 and 99. If another author breaks in and tries to steal this book, what is the probability that he or she will get the correct combination on the first attempt? Assume that the numbers are randomly selected. Given the number of possibilities, does it seem feasible to try opening the safe by making random guesses for the combination?

24. In a preliminary test of the MicroSort gender selection method, 14 babies were born and 13 of them were girls
    a) Find the number of different possible sequences of genders that are possible when 14 babies are born.
    b) How many ways can 13 girls and 1 boy be arranged in a sequence?
    c) If 14 babies are randomly selected, what is the probability that they consist of 13 girls and 1 boy?
    d) Does the gender-selection method appear to yield a result that is significantly different from a result that might be expected by random chance?

25. You become suspicious when a genetics researcher randomly selects groups of 20 newborn babies and seems to consistently get 10 girls and 10 boys. The researchers claims that it is common to get 10 girls and 10 boys in such cases,
    a) If 20 newborn babies are randomly selected, how many different gender sequences are possible.
    b) How many different ways can 10 girls and 10 boys be arranged in sequence?
    c) What is the probability of getting 10 girls and 10 boys when 10 babies are born?
    d) Based on the preceding results, do you agree with the researcher's explanation that ir is common to get 10 girls and 10 boys when 20 babies are randomly selected?

26. The Powerball lottery is run in 29 states. Winning the jackpot requires that you select the correct five numbers between 1 and 55 and, in a separate drawing, you must also select the correct single number between 1 and 42. Find the probability of winning the jackpot.

27. The Mega Millions lottery is run in 12 states. Winning the jackpot requires that you select the correct 5 numbers between 1 and 56 and, in a separate drawing, you must also select the correct single number between 1 and 46. Find the probability of winning the jackpot.

28. A state lottery involves the random selection of six different numbers between 1 and 31. If you select one six number combination, what is the probability that it will be the winning combination?

29. How many ways can 6 people be chosen and arranged in a straight line if there are 8 people to choose from?

30. 12 wrestlers compete in a competition. If each wrestler wrestles one match with each other wrestler, what are the total numbers of matches?

Section 2.4 – Random Variables

Defintion

A ***random variable*** is a variable (typically represented by *x*) that has a single numerical value, determined by chance, for each outcome of a procedure

A ***Probability distribution*** is a description that gives the probability for each value of the random variable; often expressed in the format of a graph, table, or formula

Example

Consider the offsrping of peas from parents both having the green/yellow combination of pod genes. Under these conditions, the probability that the off spring has a green pod is $\frac{3}{4}$ $\left(P(green)=0.75\right)$. If 5 such offspring are obtained, and if we let

x = number of peas with green pods among 5 offspring peas

Then *x* is a random variable because its value depends on chance.

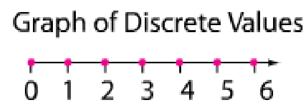
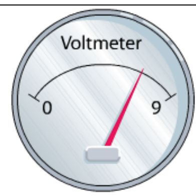
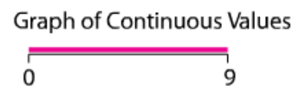
x (# of peas with Green Pods)	$P(x)$
0	0.001
1	0.015
2	0.088
3	0.264
4	0.396
5	0.237

The table is a probability distribution because it gives the probability for each value of the random variable *x*.

Defintion

A ***discrete random*** variable hass either a finite number of values or countable number of values, where "countable" refers to the fact that there might be infinitely many values, but they result from a counting process

A ***continuous random*** variable has infinitely many values, and those values can be associated with measurements on a continuous scale without gaps or interruptions

<table>
<tr><td>
</td><td>
</td></tr>
<tr><td>Discrete Random Variable: Count of the number of movies patrons.</td><td></td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Continuous Random Variable: The measured voltage of a smoke detector battery</td><td></td></tr>
</table>

Example

The following are example of discrete and continuous radom variables.

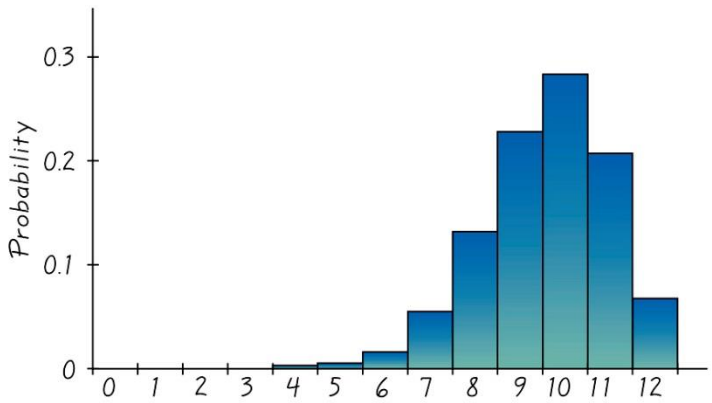
1. ***Discrete***: Let x = the number of eggs that a hen lays in a day. This is a *discrete* random variable because its only possible values are 0, or 1, or 2, and so on. No han can lay 2.343115 eggs, which would have been possible if the data had come from a continuous scale.
2. ***Discrete***: The count of the number of statistics students present in class on a given day is a whole number and is therefore a discrete random variable. The counting device is capable of indicating only a finite number of values, so it is used to obtain values for a *discrete* random variable.
3. ***Continuous***: Let x = the amount of milk a cow produces in one day. This is a *continuous* random variable because it can have any value over a continuous span. During a single day, a cow might yield an amount of milk that can be any value between 0 gallon and 5 gallons. It owould be 4.123 gallons, because the coe is not restricted to the discrete amounts of 0, 1, 2, 3, 4, or 5 gallons.
4. ***Continuous***: The measure of voltage for a particular smoke detector battery can be any value between 0 volt and 9 volts. It is therefore a *continuous* random variable.

Graphs

The probability histogram is very similar to a relative frequency histogram, but the vertical scale shows probabilities.

Requirements for Probability Distribution

- ✓ $\sum P(x) = 1$ where x assumes all possible values. The sum of all probailities must be 1. (but sunch 0.999 ot 1.001 are acceptable)
- ✓ $0 \le P(x) \le 1$ for every individual value of x



Probability Histogram for Number of Mexican-American Jurors Among 12

Example

Based on a survey conduced shown in the table:
The probabilities for the number of cell phones in user per household.Does describe a probability distribution.

x	$P(x)$
0	0.19
1	0.26
2	0.33
3	0.13

Solution

$$\sum P(x) = P(0) + P(1) + P(2) + P(3)$$
$$= 0.19 + 0.26 + 0.33 + 0.13$$
$$= 0.91 \quad \neq 1$$

Because the requirement is not satisfied. We conclude that the table does not describe a probability distribution.

Example

Does $P(x) = \frac{x}{10}$ (where x can be 0, 1, 2, 3, or 4) determine a probability distribution?

Solution

$$P(0) = \frac{0}{10} = 0; \quad P(1) = \frac{1}{10}; \quad P(2) = \frac{2}{10}; \quad P(3) = \frac{3}{10}; \quad and \quad P(4) = \frac{4}{10}$$

$$\sum P(x) = P(0)+P(1)+P(2)+P(3)+P(4)$$
$$= \frac{0}{10}+\frac{1}{10}+\frac{2}{10}+\frac{3}{10}+\frac{4}{10}$$
$$= \frac{10}{10}$$
$$\underline{=1}$$

Each value of the $P(x)$ is beteween 0 and 1.

Because both requirements are satisfied, the formula given is a probability distribution.

Mean, Variance and Standard Deviation of a Probability Distribution

$\mu = \sum\left[x \cdot P(x)\right]$	Mean of Discrete random variable
$\sigma^2 = \sum\left[(x-\mu)^2 \cdot P(x)\right]$	Variance for a probability distribution (easier to understand)
$\sigma^2 = \left[\sum x^2 \cdot P(x)\right] - \mu^2$	Variance for a probability distribution (easier computations)
$\sigma = \sqrt{\left[\sum x^2 \cdot P(x)\right] - \mu^2}$	Standard deviation for a probability distribution

Example

Find the mean, variance, and standard deviation for the probaiblity distribution described in the table.

x (# of peas)	$P(x)$
0	0.001
1	0.015
2	0.088
3	0.264
4	0.396
5	0.237

Solution

Mean: $\mu = \sum\left[x \cdot P(x)\right]$
$$= 0(0.001)+1(0.015)+2(0.088)+3(0.264)+4(0.396)+5(0.237)$$
$$= 3.752$$
$$\underline{= 3.8}$$

Variance: $\sigma^2 = \sum\left[(x-\mu)^2 \cdot P(x)\right]$
$$= (0-3.8)^2(0.001)+(1-3.8)^2(0.015)+(2-3.8)^2(0.088)+(3-3.8)^2(0.264)$$
$$+(4-3.8)^2(0.396)+(5-3.8)^2(0.237)$$
$$= 0.940574$$
$$\underline{= 0.9}$$

Standard deviation: $\sigma = \sqrt{0.940574} = 0.969832 \ \underline{= 1.0}$

Round-off Rule for μ, σ, and σ^2

Round results by carrying one more decimal place than the number of decimal places used for the random variable *x*. If the values of *x* are integers, round μ, σ, and σ^2 to one decimal place.

Identifying *Unusual* Results Range Rule of Thumb

According to the range rule of thumb, most values should lie within 2 standard deviations of the mean. We can therefore identify "unusual" values by determining if they lie outside these limits:

$$\textbf{Maximum usual value} = \mu + 2\sigma$$

$$\textbf{Minimum usual value} = \mu - 2\sigma$$

Identifying Unusual Results Probabilities

Rare Event Rule for Inferential Statistics

If, under a given assumption (such as the assumption that a coin is fair), the probability of a particular observed event (such as 992 heads in 1000 tosses of a coin) is extremely small, we conclude that the assumption is probably not correct.

Using Probabilities to Determine When Results Are Unusual

- Unusually high: *x* successes among *n* trials is an unusually high number of successes if $P(x \text{ or more}) \le 0.05$.
- Unusually low: *x* successes among *n* trials is an unusually low number of successes if $P(x \text{ or fewer}) \le 0.05$.

Example

We found that for groups of 5 offspring, the mean number of peas with Green pods is 3.87, and the standard deviation is 1.0. Use those results and the range rule of thumb to find maximum and minimum usual values. Based on the results, determine whether it is unusual to generate 5 offspring peas and find that only 1 of them has a green pod.

Solution

Maximum usual value = $\mu + 2\sigma = 3.8 + 2(1.0) = 5.8$

Minimum usual value = $\mu - 2\sigma = 3.8 - 2(1.0) = 1.8$

Based on these result, we conclude that for groups of 5 offsping peas, the number of offspring peas with green pods should usually fall between 1.8 and 5.8. If 5 offspring peas are generated, it would be unsual to get only 1 with green pod (because the value of 1 is outside of this range of usual values: 1.8 to 5.8). In this case, the maximim value is actually 5, because thatis the largest possible number of peas with grren pods.

Example

Use probabilities to determine whether 1 is an unusually low number of peas with green pods when 5 offspring are generated from parents both having the green/yellow pair of genes.

Solution

$$P(1\ or\ fewer) = P(1\ or\ \ 0) = 0.015 + 0.001 \underline{= 0.016}$$

Because the probability 0.016 is less than 0.05, we conclude that the result of 1 pea with a green pod is unusually low. There is a very samll lkelihood (0.016) of getting 1 or fewer peas with green pods.

Expected Value

The expected value of a discrete random variable is denoted by *E*, and it represents the mean value of the outcomes. It is obtained by finding the value of $\sum\left[x\cdot P(x)\right]$

$$E = \sum\left[x\cdot P(x)\right]$$

Example

You areconsidering placing a bet either on the number 7 in roulette or on the "pass line" in the dice game of craps at the casino.

a) If you bet \$5 on the number 7 in roulette, the probability of losing \$5 is $\frac{37}{38}$ and the probability of making a net gain of \$175 is $\frac{1}{38}$. (The prize id \$180, including your \$5 bet, so the net gain is \$175.) Find your expected value if you bet \$5 on the number 7 in roulette.

b) If you bet \$5 on the pass line in the dice game of craps, the probability of losing \$5 is $\frac{251}{495}$ and the probability of making a net gain of \$5 is $\frac{244}{495}$. (If you bet \$5 on the Pass Line and win, you are given \$10 that includes your bet, so the net gain is \$5.) Find your expected value if you bet \$5 on the pass line in the dice game? Why?.

Solution

a)

Event	x	$P(x)$	$x\cdot P(x)$
Lose	–\$5	$\frac{37}{38}$	–\$4.87
Gain (net)	\$175	$\frac{1}{38}$	\$4.61
Total			–\$0.26 ***Or*** –26¢

You can expect to lose an average of 26¢.

b)

Event	x	$P(x)$	$x \cdot P(x)$
Lose	–\$5	$\frac{251}{495}$	–\$2.54
Gain (net)	\$5	$\frac{244}{495}$	\$2.46
Total			–\$0.08 ***Or*** –8¢

You can expect to lose an average of 8¢.

Exercises *Section* 2.4 – Random Variables

1. Determine whether or not a probability distribution is given. If a probability is given, find its mean and standard deviation. If the probability is not given, identify the requirements that are not satisfied.

a)

x	$P(x)$
0	0.125
1	0.375
2	0.375
3	0.125

b)

x	$P(x)$
0	0.22
1	0.16
2	0.21
3	0.16

c)

x	$P(x)$
0	0.528
1	0.360
2	0.098
3	0.013
4	0.001
5	0^+

0^+ denotes a positive probability value that is very small.

d)

x	$P(x)$
0	0.02
1	0.15
2	0.29
3	0.26
4	0.16
5	0.12

2. Based on past results found in the *Information Please Almanac*, there is a 0.1919 probability that a baseball World Series context will last 4 games, is a 0.2121 probability that it will last 5 games, a 0.2222 probability that it will last 6 games, a 0.3737 probability that us will last 7 games.
   a) Does the given information describe a probability distribution?
   b) Assuming that the given information describes a probability distribution, find the mean and standard deviation for the numbers of games in World Series contests.
   c) Is it unusual for a team to "sweep" by winning in four games? Why or why not?

3. Based on information from MRI Network, some job applicants are required to have several interviews before a decision is made. The number of required interviews and the corresponding probabilities are: 1 (0.09); 2 (0.31); 3 (0.37); 4 (0.12); 5 (0.05); 6 (0.05).
   a) Does the given information describe a probability distribution?
   b) Assuming that a probability distribution is described, find its mean and standard deviation.
   c) Use the range rule of thumb to identify the range of values for usual numbers of interviews.
   d) Is it unusual to have a decision after just one interview? Explain?

4. Based on information from Car dealer, when a car is randomly selected the number of bumper stickers and the corresponding probabilities are: 0 (0.824); 1 (0.083); 2 (0.039); 3 (0.014); 4 (0.012); 5 (0.008); 6 (0.008); 7 (0.004); 8 (0.004); 9 (0.004).
   a) Does the given information describe a probability distribution?

b) Assuming that a probability distribution is described, find its mean and standard deviation.
c) Use the range rule of thumb to identify the range of values for usual numbers of bumper stickers.
d) Is it unusual for a car to have more than one bumper sticker? Explain?

5. A Company hired 8 employees from a large pool of applicants with an equal numbers of males and females. If the hiring is done without regard to sex, the numbers of females hired and the corresponding probabilities are: 0 (0.004); 1 (0.0313); 2 (0.109); 3 (0.219); 4 (0.273); 5 (0.219); 6 (0.109); 7 (0.031); 8 (0.004).
   a) Does the given information describe a probability distribution?
   b) Assuming that a probability distribution is described, find its mean and standard deviation.
   c) Use the range rule of thumb to identify the range of values for usual numbers of females hired in such groups of 8.
   d) If the most recent group of 8 newly hired employees does not include any females, does there appear to be discrimination based on sex? Explain?

6. Let the random variable *x* represent the number of girls in a family of 4 children. Construct a table describing the probability distribution; then find the mean and the standard deviation. (Hint: List the different possible outcomes.) Is it unusual for a family of 3 children to consist of 3 girls?

7. In 4 lottery game, you pay 50¢ to select a sequence of 4 digits, such 1332. If you select the same sequence of 4 digits that are drawn, you win and collect $2788.
   a) How many different selections are possible?
   b) What is the probability of winning?
   c) If you win, what is your net profit?
   d) Find the expected value.

8. When playing roulette at casino, a gambler is trying to decide whether to bet $5 on the number 13 or bet $5 that the outcomes any one of these 5 possibilities: 0 or 00 or 1 or 2 or 3.the expected value of the $5 bet for a single number is –26¢. For the $5 bet that the outcome 0 or 00 or 1 or 2 or 3, there is a probability of $\frac{5}{38}$ of making a net profit of $30 and a $\frac{33}{38}$ probability of losing $5.
   a) Find the expected value for the $5 bet that the outcome is 0 or 00 or 1 or 2 or 3.
   b) Which bet is better: A $5 bet on the number 13 or a $5 bet the outcome is 0 or 00 or 1 or 2 or 3? Why?

9. There is a 0.9986 probability that a randomly selected 30-year-old male lives through the year. As insurance company charges $161 for insuring that the male will live through the year. If the male does not survive the year, the policy pays out $100,000 as a death benefit.
   a) From the perspective of the 30-year-old male, what are the values corresponding to the 2 events of surviving the year and not surviving?
   b) If a 30-year-old male purchases the policy, what is his expected value?
   c) Can the insurance company expect to make a profit from many such policies? Why?

Section 2.5 – Binomial Distributions

Definition

A binomial probability distribution results from a procedure that meets all the following ***requirements***:

1. The procedure has a fixed number of trials.
2. The trials must be independent. (The outcome of any individual trial doesn't affect the probabilities in the other trials.)
3. Each trial must have all outcomes classified into two categories (commonly referred to as success and failure).
4. The probability of a success remains the same in all trials.

Notation for Binomial Probability Distributions

S and *F* (*success* and *failure*) denote the two possible categories of all outcomes; p and q will denote the probabilities of *S* and *F*, respectively, so

$P(S) = p$ (p = probability of success)

$P(F) = 1 - p = q$ (q = probability of failure)

n denotes the fixed number of trials.

x denotes a specific number of successes in n trials, so x can be any whole number between 0 and n, inclusive.

p denotes the probability of success in *one* of the n trials.

q denotes the probability of failure in one of the n trials.

***P*(*x*)** denotes the probability of getting exactly x successes among the n trials.

Example

Consider an experiment in which 5 offspring peas are generated each having the green/yellow combination of genes for pod color. The probability of an offspring pea with a green pod is 0.75. That is, $P(\text{green pod}) = 0.75$. Suppose we want to find the probability that exactly 3 of the 5 offspring peas have a green pod.

a) Does this procedure result in a binomial distribution?

b) If this procedure result in a binomial distribution, identify the values of n, x, p, and q.

Solution

a) 1. The number of trials (5) is fixed.

2. The 5 trials are independent, because the probability of any offspring pea having a green pod is not affected by the outcome of any other offspring pea.

3. Each of the 5 trials has 2 categories of outcomes: The pea has a green pod or it does not.

4. For each offspring pea, the probability that it has a green pod is 0.75, and that probability remains the same for each of the 5 peas.

b) Having concluded that the given procedure does result in a binomial distribution, we now proceed to identify the values of *n*, *x*, *p*, and *q*.

1. With 5 offspring peas, we have $n = 5$.
2. We want the probability of exactly 3 peas with green pods, so $x = 3$.
3. The probability of success (getting a pea with a green pod) for one selection is 0.75, so $p = 0.75$.
4. The probability of failure (not getting a green pod) is 0.25, so $q = 0.25$..

Important Hints

- Be sure that *x* and *p* both refer to the same category being called a success.
- When sampling without replacement, consider events to be independent if $n < 0.05N$.

Method 1: Using the Binomial Probability Formula

$$P(x) = \frac{n!}{(n-x)!\ x!} \cdot p^x \cdot q^{n-x} \quad for \quad x = 0,\ 1,\ 2,\ \ldots,\ n$$

$$P(x\ success) = C_{n,x}\, p^x q^{n-x}$$

where

n = number of trials
x = number of successes among *n* trials
p = probability of success in any one trial
q = probability of failure in any one trial ($q = 1 - p$)

Example

The probability of an offspring pea with a green pod is 0.75, use the binomial probability formula to find the probability that exactly 3 peas with green pods when of the 5 offspring peas have a green pod when 5 offspring peas are generated. That is, find $P(3)$ given that $n = 5$, $x = 3$, $p = 0.75$, and $q = 0.75$.

Solution

$$P(x=3) = \frac{5!}{(5-3)!\ 3!} \cdot (0.75)^3 \cdot (0.25)^{5-3}$$

$$= \frac{5!}{2!\ 3!} \cdot (0.75)^3 \cdot (0.25)^2 \qquad 5\left(math \to PROB \to 4\right)!*0.75\wedge*30.25\wedge 2/\left(2!*3!\right)$$

$$\underline{= 0.263671875}$$

The probability of getting exactly 3 peas with green pods among 5 offspring peas is 0.264,

Example

The fast food chain McDonald's has a brand name recognition rate of 95% around the world. Assuming that we randomly select 5 people, use the table to find the following

a) The probability that exactly 3 of the 5 people recognize McDonald's

b) The probability that the number of people who recognize McDonald's is 3 or fewer

n	***x***	***P(x) =.95***
5	0	0+
4	1	0+
3	2	0.001
2	3	0.021
1	4	0.204
0	5	0.774

Solution

a) $P(x=3)=.021$

b) $P(3 \text{ or fewer}) = P(3)+P(2)+P(1)+P(0)$

$$= 0.021 + 0.001 + 0 + 0$$

$$= 0.022$$

TI-83/84 PLUS Press **2nd VARS** (to get **DISTR,** which denotes "distributions"), then select the option identified as **binompdf(.** Complete the entry of **binompdf(n, p, x)** with specific values for *n*, *p*, and *x*, then press **ENTER.** The result will be the probability of getting *x* successes among *n* trials.

You could also enter **binompdf(n, p)** to get a list of *all* of the probabilities corresponding to $x = 0, 1, 2, \ldots, n$. You could store this list in L2 by pressing **STO** $\rightarrow$ **L2.** You could then manually enter the values of $0, 1, 2, \ldots, n$ in list L1, which would allow you to calculate statistics (by entering **STAT, CALC,** then **L1, L2**) or view the distribution in a table format (by pressing **STAT,** then **EDIT**).

The command **binomcdf** yields *cumulative* probabilities from a binomial distribution. The command **binomcdf(n, p, x)** provides the sum of all probabilities from $x = 0$ through the specific value entered for *x*.

Exercises *Section* 2.5 – Binomial Distributions

1. 20 different Senators are randomly selected from the 100 Senators in the current Congress, and each was asked whether he or she is in favor of abolishing estate taxes. Does this procedure result in a binomial distribution, if it is not binomial, identify at least one requirement that is not satisfied?

2. 15 different Governors are randomly selected from the 50 Governors in the currently office and the sex of each Governor is recorded. Does this procedure result in a binomial distribution, if it is not binomial, identify at least one requirement that is not satisfied?

3. 200 statistics students are randomly selected and each asked if he or she owns a TI-84 Plus calculator. Does this procedure result in a binomial distribution, if it is not binomial, identify at least one requirement that is not satisfied?

4. Multiple choice questions on the SAT test have 5 possible answers (*a, b, c, d, e*), 1 of which is correct. Assume that you guess the answers to 3 such questions.
   a) Use the multiplication rule to find the probability that the first 2 guesses are wrong and the third is correct. That is, find *P*(*WWC*), where *C* denotes a correct answer and *W* denotes a wrong answer.
   b) Beginning with *WWC*, make a complete list of the different possible arrangements of 2 wrong answers and 1 correct answer, then find the probability for each entry in the list.
   c) Based on the proceeding results, what is the probability of getting exactly 1 correct answer when 3 guesses are made?

5. A psychology test consists of multiple choice questions, each having 4 possible answers (*a, b, c, d*), 1 of which is correct. Assume that you guess the answers to 6 such questions.
   a) Use the multiplication rule to find the probability that the first 2 guesses are wrong and the last 4 guesses are correct. That is, find *P*(*WWCCCC*), where *C* denotes a correct answer and *W* denotes a wrong answer.
   b) Beginning with *WWCCCC*, make a complete list of the different possible arrangements of 2 wrong answers and 4 correct answers, then find the probability for each entry in the list.
   c) Based on the proceeding results, what is the probability of getting exactly 4 correct answers when 6 guesses are made?

6. Use the Binomial Probability Table to find the probability of *x* success given the probability *p* of success on a single trial
   $a)$ $n=2, \quad x=1, \quad p=.30$ $\qquad$ $b)$ $n=5, \quad x=1, \quad p=0.95$
   $c)$ $n=15, \quad x=11, \quad p=0.99$ $\qquad$ $d)$ $n=14, \quad x=4, \quad p=0.60$
   $e)$ $n=10, \quad x=2, \quad p=0.05$ $\qquad$ $f)$ $n=12, \quad x=12, \quad p=0.70$

7. Use the Binomial Probability Formula to find the probability of x success given the probability p of success on a single trial

$a)\quad n=12, \quad x=10, \quad p=\frac{3}{4}$ $\qquad$ $b)\quad n=9, \quad x=2, \quad p=0.35$

$c)\quad n=20, \quad x=4, \quad p=0.15$ $\qquad$ $d)\quad n=15, \quad x=13, \quad p=\frac{1}{3}$

8. In the US, 40% of the population have brown eyes. If 14 people are randomly selected, find the probability that at least 12 of them have brown eyes. Is it unusual to randomly select 14 people and find that at least 12 of them have brown eyes? Why or why not?

9. When blood donors were randomly selected, 45% of them had blood that is Group *O*. The display shows that the probabilities obtained by entering the values of $n = 5$ and $p = 0.45$.

x	***P*(*x*)**
0	0.050328
1	0.205889
2	0.336909
3	0.275653
4	0.112767
5	0.018453

a) Find the probability that at least 1 of the 5 donors has Group *O* blood. If at least 1 Group *O* donor is needed, is it reasonable to expect that at least 1 will be obtained?
b) Find the probability that at least 3 of the 5 donors have Group *O* blood. If at least 3 Group *O* donors are needed, is it very likely to expect that at least 3 will be obtained?
c) Find the probability that all donors have Group *O* blood. Is it unusual to get 5 Group *O* donors from 5 randomly selected donors? Why or Why not?
d) Find the probability that at most 2 of the 5 donors have Group *O* blood.

10. There is 1% delinquency rate for consumers with FICO credit rating scores above 800. If a bank provides large loans 12 people with FICL scores above 800, what is the probability that at least one of them becomes delinquent? Based on that probability, should the bank plan on dealing with a delinquency?

11. Ten peas are generated from parents having the green/yellow pair of genes, so there is a 0.75 probability that an individual pea will have a green pod. Find the probability that among the 10 offspring peas, at least 9 have green pods. Is it unusual to get at least 9 peas with green pods when 10 offspring peas are generated? Why or why not?

12. You purchased a slot machine configured so that there is a $\frac{1}{2{,}000}$ probability of winning the jackpot on any individual trial. Although no one would seriously consider tricking the author, suppose that a guest claims that she played the slot machine 5 times and hit the jackpot twice
a) Find the probability of exactly 2 jackpots in 5 trials.
b) Find the probability of at least 2 jackpots in 5 trials.
c) Does the guest's claim of hitting 2 jackpots in 5 trials seem valid? Explain.

13. In a survey of 320 college graduates, 36% reported that they stayed on their first full-time job less than one year.
a) If 15 of those survey subjects are randomly selected without replacement for a follow-up survey, find the probability that 5 of them stayed on their first full-time job less than one year.

b) If part (a) is changed so that 20 different survey subjects are selected, explain why the binomial probability formula *cannot* be used.

14. In a survey of 150 senior executives, 47% said that the most common job interview mistake is to have little or no knowledge of the company.
    a) If 6 of those surveyed executives are randomly selected without replacement for a follow-up survey, find the probability that 3 of them said that the most common job interview mistake is to have little or no knowledge of the company.
    b) If part (a) is changed so that 9 different surveyed executives are selected, explain why the binomial probability formula *cannot* be used.

Section 2.6 – Mean, Variance, and Standard Deviation for the Binomial Distributions

We consider important characteristics of a binomial distribution including center, variation and distribution. That is, given a particular binomial probability distribution we can find its mean, variance and standard deviation.
A strong emphasis is placed on *interpreting* and *understanding* those values.

	Discrete Probability Distribution	***Binomial Distribution***
Mean	$\mu = \sum [x \cdot P(x)]$	$\mu = np$
Variance	$\sigma^2 = \left[\sum x^2 \cdot P(x)\right] - \mu^2$	$\sigma^2 = npq$
Std. Dev.	$\sigma = \sqrt{\left[\sum x^2 \cdot P(x)\right] - \mu^2}$	$\sigma = \sqrt{npq}$

Where
n = number of fixed trials
p = probability of ***success*** in one of the ***n*** trials
q = probability of ***failure*** in one of the ***n*** trials

Interpretation of Results

It is especially important to interpret results. The range rule of thumb suggests that values are unusual if they lie outside of these limits:

Maximum usual values = $\boldsymbol{\mu + 2\sigma}$
Minimum usual values = $\boldsymbol{\mu - 2\sigma}$

Example

Find the mean and standard deviation for the numbers of peas with green pods when groups of 5 offspring peas are generated. Assume that there is 0.75 probability that an offspring pea has a green pod.

Solution

Given: $n = 5; \quad p = 0.75; \quad q = 1 - p = 1 - 0.75 = 0.25$

Mean: $\mu = np = 5(0.75) = 3.75 \ \underline{\approx 3.8}$

Standard deviation: $\sigma = \sqrt{npq} = \sqrt{(5)(0.75)(0.25)} = 0.9375 \ \underline{\approx 1}$

Example

Mendel generated 580 offspring peas. He claimed that 75% or 4.35, of them would have green pods. The actual experiment resulted in 428 peas with green pods.

a) Assuming that groups of 580 offspring peas are generated find the mean and standard deviation for the numbers of peas with green pods.

b) Use the range rule of thumb to find the minimum usual number and the maximum usual number of peas with green pods. Based on those numbers, can we conclude that Mendel's actual result of 428 peas with green pods is unusual? Does this suggest that Mendel's value of 75% wrong?

Solution

a) ***Given***: $n = 580;\quad p = 0.75;\quad q = 0.25$

Mean: $\mu = np = (580)(0.75)\ \boxed{= 435}$

Standard deviation: $\sigma = \sqrt{npq} = \sqrt{(580)(0.75)(0.25)}\ \boxed{\approx 10.4}$

b) *Maximum usual values* $= \mu + 2\sigma = 435 + 2(10.4)\ \boxed{= 455.8}$

Minimum usual values $= \mu - 2\sigma = 435 - 2(10.4)\ \boxed{= 414.2}$

If Mendel generated groups of 580 offspring peas and if his 75% rate is correct, the numbers of peas with green pods should usually fall between 414.2 and 455.8.
Mendel actually got 428 peas with green pods, and that value does fall within the range of usual values, so the experimental results are consistent with the 75% rate.
The results do not suggest that Mendel's claimed rate of 75% is wrong.

Exercises Section 2.6 – Mean, Variance, and Standard Deviation for the Binomial Distributions

1. In a Gallup poll of 1236 adults, it was found that 5% of those polled said that bad luck occurs after breaking a mirror. Based on these results, such randomly selected groups of 1236 adults will have a mean of 61.8 people with that belief, and a standard deviation of 7.7 people. What is the variance?

2. Random guesses are made for 50 SAT multiple choice questions, so $n = 50$ and $p = 0.2$.
   a) Find the mean μ and standard deviation σ.
   b) Use the range rule of thumb to find the minimum usual number and the maximum usual number.

3. In an analysis of test result from the YSORT gender selection method, 152 babies are born and it is assumed that boys and girls are equally likely, so $n = 152$ and $p = 0.5$.
   a) Find the mean μ and standard deviation σ.
   b) Use the range rule of thumb to find the minimum usual number and the maximum usual number.

4. In a Gallup poll of 1236 adults, it showed that 145% believe that bad luck follows if your path is crossed by a black car, so $n = 1236$ and $p = 0.14$.
   a) Find the mean μ and standard deviation σ.
   b) Use the range rule of thumb to find the minimum usual number and the maximum usual number.

5. The midterm exam in a nursing course consists of 75 true/false questions. Assume that an unprepared student makes random guesses for each of the answers.
   a) Find the mean and standard deviation for the number of correct answers for such students.
   b) Would it be unusual for a student to pass this exam by guessing and getting at least 45 correct answers? Why or why not?

6. The final exam in a nursing course consists of 100 multiple-choice questions. Each question has 5 possible answers, and only 1 of them is correct. An unprepared student makes random guesses for all of the answers.
   a) Find the mean and standard deviation for the number of correct answers for such students.
   b) Would it be unusual for a student to pass this exam by guessing and getting at least 60 correct answers? Why or why not?

7. In a test of the XSORT method of gender selection, 574 babies are born to couples trying to have baby girls, and 525 of those babies are girls.
   a) If the gender-selection method has no effect and boys and girls are equally likely, find the mean and standard deviation for the numbers of girls born in groups of 574.
   b) Is the result of 525 girls unusual? Does it suggest that the gender-selection method appears to be effective?

8. In a test of the YSORT method of gender selection, 152 babies are born to couples trying to have baby boys, and 127 of those babies are boys.
   a) If the gender-selection method has no effect and boys and girls are equally likely, find the mean and standard deviation for the numbers of boys born in groups of 152.
   b) Is the result of 127 boys unusual? Does it suggest that the gender-selection method appears to be effective?

9. A headline in USA Today states that "most stay at first job less than 2 years." That headline is based on a poll of 320 college graduates. Among those polled, 78% stayed at their full-time job less than 2 years.
   a) Assuming that 50% is the true percentage of graduates who stay at their first job less than 2 years, find the mean and the standard deviation of the numbers of such graduates in randomly selected groups of 320 graduates.
   b) Assuming that the 50% rate in part (*a*) is correct; find the range of usual values for the numbers of graduates among 320 who stay at their first job less than 2 years.
   c) Find the actual number of surveyed who stayed at their first job less 2 years. Use the range of values from part (*b*) to determine whether that number is unusual. Does the result suggest that the headline is not justified?
   d) This statement was given as part of the description of the survey methods used: "Alumni who opted-in to receive communications from Experience were invited to participate in the online poll, and 320 of them completed the survey." What does that statement suggest about the result?

10. In a study of 420,095 cell phone users in Denmark, it was found that 135 developed cancer of the brain or nervous system. If we assume that the use of cell phones has no effect on developing such cancer, then the probability of a person having such cancer is 0.000340.
    a) Assuming that the cell phones have no effect on developing cancer, find the mean and the standard deviation of the numbers of people in groups of 420,095 that can be expected to have cancer of the brain or nervous system.
    b) Based on the result from part (*a*), is it unusual to find that among 420,095 people, there are 135 cases of cancer of the brain or nervous system? Why or why not?
    c) What do these results suggest about the publicized concern that cell phones are a health danger because they increase the risk of cancer of the brain or nervous system?

11. Mario's Pizza Parlor has just opened. Due to a lack of employee training, there is only a 0.8 probability that a pizza will be edible. An order for 5 pizzas has just been placed. What is the minimum number of pizzas that must be made in order to be at least 99% sure that there will be 5 that are edible?

Section 2.7 – Standard Normal Distributions

The *standard normal distribution* has three properties:

1. It is graph is bell-shaped.
2. It is mean is equal to 0 ($\mu = 0$).
3. It is standard deviation is equal to 1 ($\sigma = 1$).

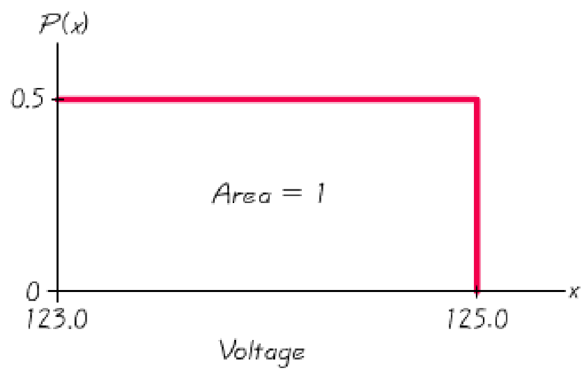
Uniform Distribution

Definition

A continuous random variable has a ***uniform distribution*** if its values are spread evenly over the range of probabilities. The graph of a uniform distribution results in a rectangular shape.

Example

The Power and Light Company provides electricity with voltage levels that are uniformly distributed between 123.0 volts and 125.0 volts. That is, any voltage amount between 123.0 volts and 125.0 volts is possible, and all of the possible values are equally likely. If we randomly select one of the voltage levels and represent its value by the random variable *x*, than *x* has a distribution that can me graphed as below



Uniform Distribution of Voltage Levels

Density Curve

A density curve is the graph of a continuous probability distribution. It must satisfy the following properties:

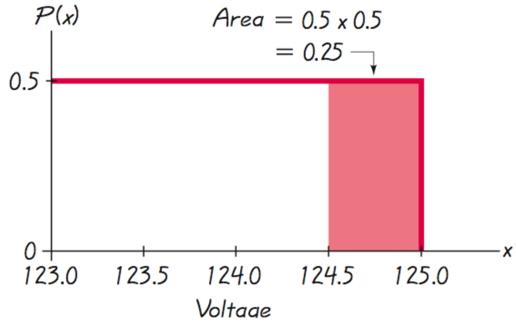
1. The total area under the curve must equal 1.
2. Every point on the curve must have a vertical height that is 0 or greater. (That is, the curve cannot fall below the *x*-axis.)

Area and Probability

Because the total area under the density curve is equal to 1, there is a correspondence between ***area*** and ***probability***.

Example

Given the uniform distribution illustrated in the figure below, find the probability that a randomly voltage level is greater than 124.5 volts.



Shaded area represents voltage levels greater than 124.5 volts.

$$P(\textit{voltage greater than } 124.5 \textit{ volts}) = \textit{area of the shaded region}$$

$$= 0.5 \times 0.5$$

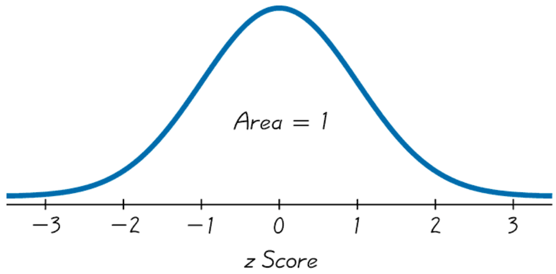
$$= 0.25$$

The probability of randomly selecting a voltage level greater than 124.5 volts is 0.25.

Standard Normal Distribution

Definition

The standard normal distribution is a normal probability distribution with $\mu = 0$ and $\sigma = 1$. The total area under its density curve is equal to 1.



Finding Probabilities: When Given z-scores

- Standard Normal distribution Table.

  Using Table (Normal Distribution Table):

  1. It is designed only for the *standard* normal distribution, which has a mean of 0 and a standard deviation of 1.
  2. It is on two pages, with one page for *negative* z-scores and the other page for *positive* z-scores.
  3. Each value in the body of the table is a *cumulative area from the left* up to a vertical boundary above a specific z-score.
  4. When working with a graph, avoid confusion between z-scores and areas.

     z Score
     Distance along horizontal scale of the standard normal distribution; refer to the leftmost column and top row of Normal Distribution Table.

     Area
     Region under the curve; refer to the values in the body of Normal Distribution Table.

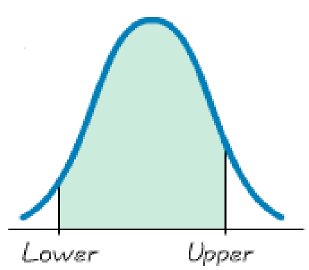
  5. The part of the z-score denoting hundredths is found across the top.

- Formulas and Tables insert card
- Find areas for many different regions

Methods for Finding Normal Distribution Areas

TI-83/84 Plus Calculator

Gives area bounded on the left and bounded on the right by vertical lines above any specific values.



TI-83/84 Press 2ND VARS **[2: normal cdf ( ]**, then enter the two z scores separated by a comma, as in (left z score, right z score).

Example

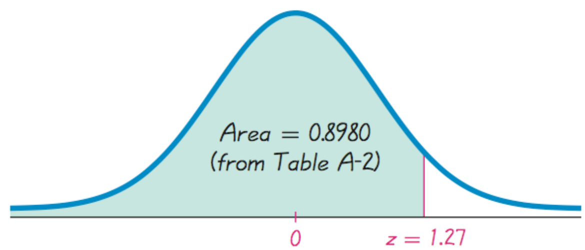
The Precision Scientific Instrument Company manufactures thermometers that are supposed to give readings of 0ºC at the freezing point of water. Tests on a large sample of these instruments reveal that at the freezing point of water, some thermometers give readings below 0º (denoted by negative numbers) and some give readings above 0º (denoted by positive numbers). Assume that the mean reading is 0ºC and the standard deviation of the readings is 1.00ºC. Also assume that the readings are normally distributed. If one thermometer is randomly selected, find the probability that, at the freezing point of water, the reading is less than 1.27º.

<u>*Solution*</u>

z	.00	.01	.02	.03	.04	.05	.06	.07
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292

$P(z<1.27) = 0.8980$

The *probability* of randomly selecting a thermometer with a reading less than 1.27º is 0.8980. Or 89.80% will have readings below 1.27º.



Example

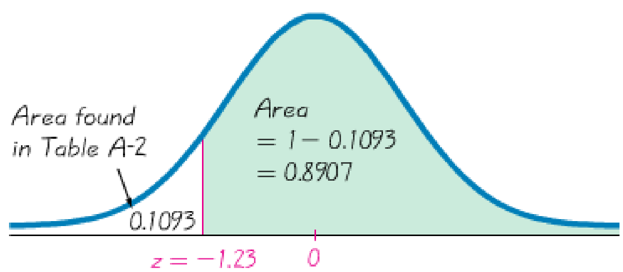
If thermometers have an average (mean) reading of 0 degrees and a standard deviation of 1 degree for freezing water, and if one thermometer is randomly selected, find the probability that it reads (at the freezing point of water) above –1.23 degrees.

<u>*Solution*</u>

$P(z>-1.23) = 1-0.1093 = 0.8907$

Probability of randomly selecting a thermometer with a reading above –1.23º is 0.8907.

89.07% of the thermometers have readings above –1.23 degrees.



Example

A thermometer is randomly selected. Find the probability that it reads (at the freezing point of water) between –2.00 and 1.50 degrees.

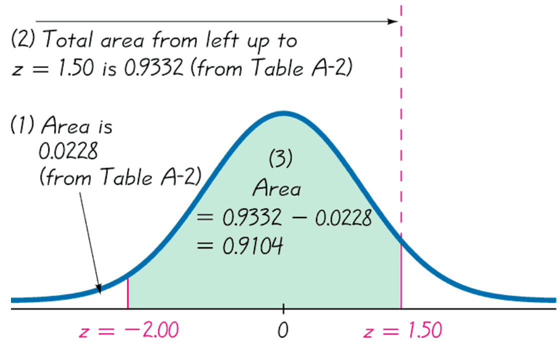
<u>.Solution</u>

$P(z < -2.00) = 0.0228$

$P(z < 1.50) = 0.9332$

$P(-2.00 < z < 1.50) = 0.9332 - 0.0228$

$= 0.9104$



The probability that the chosen thermometer has a reading between – 2.00 and 1.50 degrees is 0.9104.

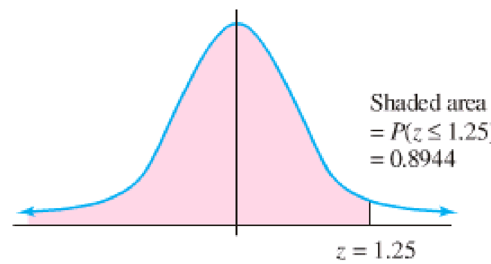
If many thermometers are selected and tested at the freezing point of water, then 91.04% of them will read between –2.00 and 1.50 degrees.

Example

Find the areas under the standard normal curve

a) The area to the ***left*** of $z = 1.25$

$A = P(z \le 1.25) \ \underline{= 0.8944}$ **(*Using left curve table*)**

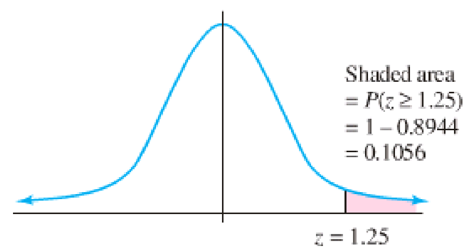


b) The area to the ***right*** of $z = 1.25$

$A = P(z \ge 1.25)$ **(*Using left curve table*)**

$= 1 - 0.8944$

$\underline{= 0.1056}$

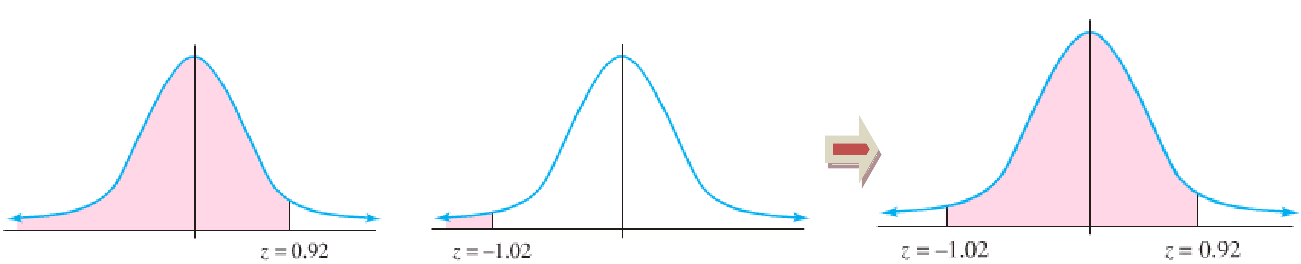


c) The area **between** $z = -1.02$ and

$A = P(-1.02 \le z \le 0.92)$ **(*Using left curve table*)**

$= 0.8212 - 0.1539$

$\underline{= 0.6673}$



Notation

$P(a < z < b)$ denotes the probability that the z score is between a and b.

$P(z > a)$ denotes the probability that the z score is greater than a.

$P(z < a)$ denotes the probability that the z score is less than a.

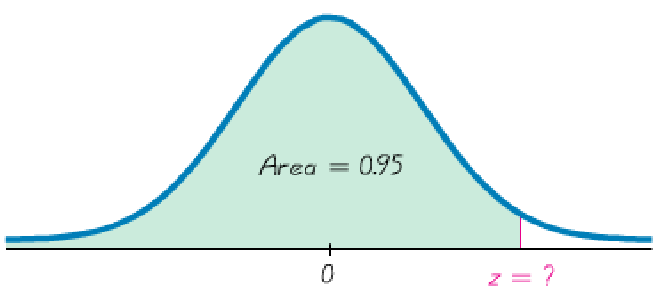
Finding a *z*–Score When Given a Probability Using Table

1. Draw a bell-shaped curve and identify the region under the curve that corresponds to the given probability. If that region is not a cumulative region from the left, work instead with a known region that is a cumulative region from the left.
2. Using the cumulative area from the left, locate the closest probability in the body of Table and identify the corresponding z score.

Example

With temperature readings at the freezing point of water that are normally distributed with a mean 0°C and a standard deviation of 1.00°C. Find the temperature corresponding to P_{95} , the 95th percentile. That is, find the temperature separating the bottom 95 from the top 5%.

<u>Solution</u>



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51[illegible]95	.51[illegible]94	.52392	.52790	.53188	.53586
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449

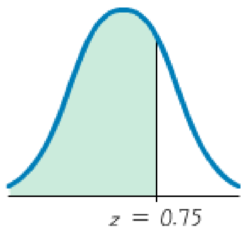
From the table, we find the areas of 0.9495 and 0.9505. The area 0.95 corresponds to a z-score of 1.645.

When tested at freezing, 95% of the readings will be less than or equal to 1.645°C, and 5% of them will be greater than or equal to 1.645°C

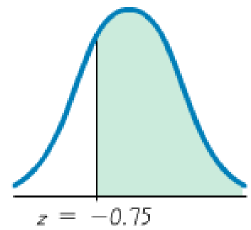
Exercises Section 2.7 – Standard Normal Distributions

1. Find the area shaded region. The graph depicts the standard distribution with mean 0 and standard deviation 1.

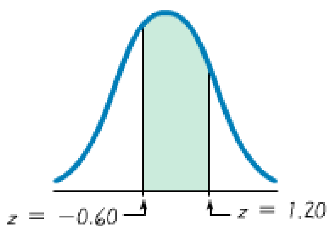
   a)



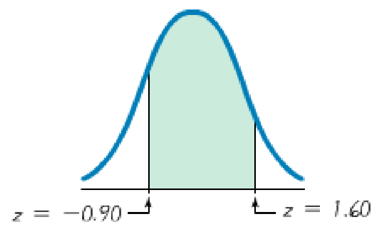
   b)



   c)

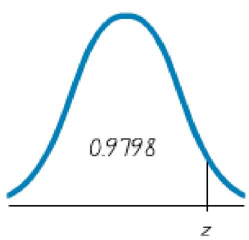


   d)

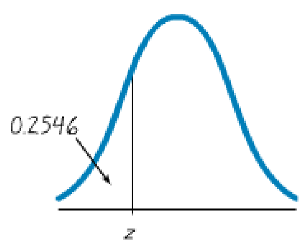


2. Find the indicated *z*-score. The graph depicts the standard distribution with mean 0 and standard deviation 1.

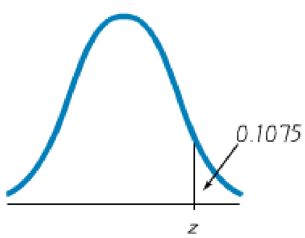
   a)



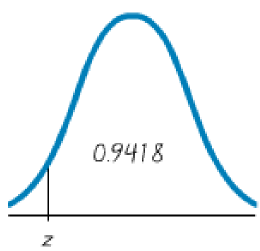
   b)



   c)



   d)



3. Assume that thermometer readings are normally distributed with a mean of 0ºC and the standard deviation of the readings is 1.00ºC. A thermometer is randomly selected and tested. In each case, draw a sketch, and find the probability of each reading.

   a) Less than –1.50
   b) Less than –2.75
   c) Less than 1.23
   d) Greater than 2.22
   e) Greater than 2.33
   f) Greater than –1.75
   g) Between 0.50 and 1.00
   h) Between –3.00 and –1.00
   i) Between –1.20 and 1.95
   j) Between –2.50 and 5.00
   k) Greater than 0
   l) Less than 0

4. Assume that thermometer readings are normally distributed with a mean of 0ºC and the standard deviation of the readings is 1.00º*C*. A thermometer is randomly selected and tested. In each case, draw a sketch, and find the temperature reading corresponding to the given information.

   a) Find P_{95}, the 95th percentile. This is the temperature separating the bottom 95% from the top 5%.

   b) Find P_1, the 1st percentile. This is the temperature separating the bottom 1% from the top 99%.

   c) If 2.5% of the thermometers are rejected because they have readings that are too high and another 2.5% are rejected because they have readings that are too low, find the 2 readings that are cutoff values separating the rejected thermometers from the others.

   d) If 0.5% of the thermometers are rejected because they have readings that are too high and another 0.5% are rejected because they have readings that are too low, find the 2 readings that are cutoff values separating the rejected thermometers from the others.

5. For a standard normal distribution, find the percentage of data that are

   a) Within 2 standard deviations of the mean.

   b) More than 1 standard deviation away from the mean.

   c) More than 1.96 standard deviations away from the mean.

   d) Between $\mu - 3\sigma$ and $\mu + 3\sigma$.

   e) More than 3 standard deviations away from the mean.

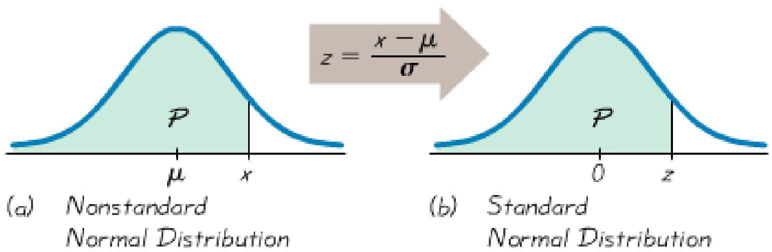
Section 2.8 – Applications Normal Distributions

Working with normal distributions that are not standard, that is, the mean is not 0 or the standard deviation is not 1, or both.
The key concept is that we can use a simple conversion that allows us to standardize any normal distribution so that the same methods of the previous section can be used.

Conversion Formula

$$z = \frac{x - \mu}{\sigma}$$

Round z scores to 2 decimal places



Finding Areas with a nonstandard normal distribution

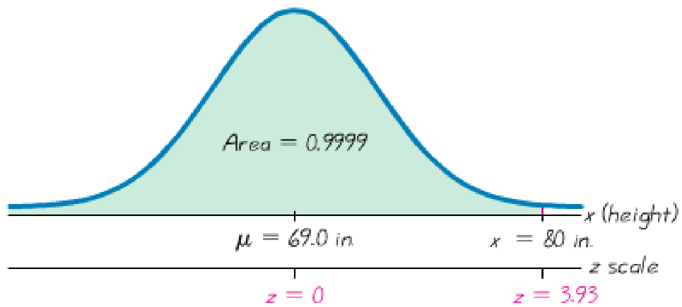
1. Sketch a normal curve, label the mean and the specific *x* values, then *shade* the region representing the desired probability.
2. For each relevant value x that is a boundary for the shaded region, use the formula to convert that value to the equivalent z-score.
3. Refer the Normal Distribution Table or use the calculator to find the area of the shaded region. This area is the desired probability.

Example

The typical home doorway has a height of 6 ft. 8 in., or 80 in. Because men tend to be taller than women, we will consider only men as we investigate the limitations of that standard doorway height. Given that heights of men are normally distributed with a mean of 69.0 in. and a standard deviation of 2.8 in., find the percentage of men who can fit through the standard doorway without bending or bumping their head. Is that percentage high enough to continue using 80 in. as the standard height? Will a doorway height of 80. Be sufficient in future years?

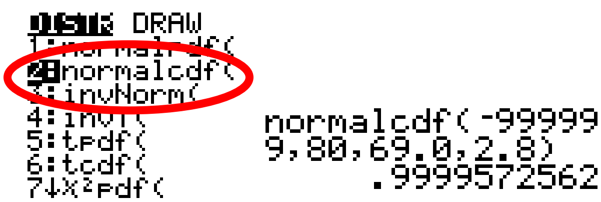
<u>Solution</u>

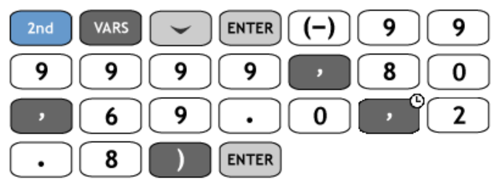
Men have heights that are normally distributed with a mean of 69.0 in. and a standard deviation of 2.8 in. The shaded region represents the men who can fir through a doorway that has a height 80 in.



The z-score: $z = \dfrac{x-\mu}{\sigma} = \dfrac{80-69}{2.8} = 3.93$ $\qquad (80-69)/2.8$

Referring to the table, the z-score values are less than 3.5, therefore; if we use calculator:





TI-83/84 PLUS

- To find the area between two values, press **2nd, VARS, 2** (for normalcdf), then proceed to enter the two values, the mean, and the standard deviation, all separated by commas, as in (left value, right value, mean, standard deviation). *Hint:* If there is no left value, enter the left value as −999999, and if there is no right value, enter the right value as 999999. In Example 1 we want the area to the left of $x = 80$ in., so use the command **normalcdf (−999999, 80, 69.0, 2.8)** as shown in the accompanying screen display.

$\boxed{normalcdf\,(-999999,\ x,\ \mu,\ \sigma)}$

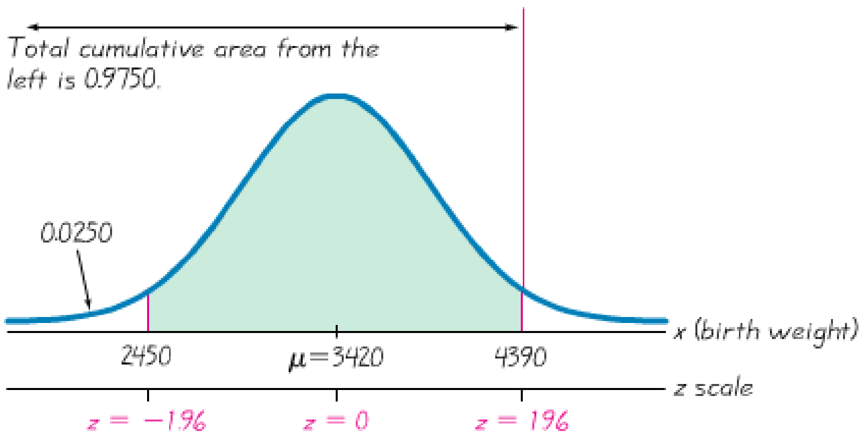
✓ The proportion of men who can fit through the standard doorway height of 80 in. is 0.9999, or 99.99%. Very few men will not be able to fit through the doorway without bending their head. This percentage id high enough to justify the use of 80 in. as the standard doorway height. However, heights of men and women have been increasing gradually but steady over the past decades.

Example

Birth weights in the U.S. are normally distributed with a mean of 3420 g and a standard deviation of 495 g. A hospital requires special treatment for babies that are less than 2450 g (unusual light) or more than 4390 g (unusually heavy). What is the percentage of babies who do not require special treatment because they have birth weights between 2450 g and 4390 g? Under these conditions, do many babies require special treatment?

Solution

Given: $\mu = 3420, \quad \sigma = 495$



The area to the ***left*** of $x = 2450 \quad \Rightarrow \quad \underline{z = \dfrac{2450 - 3420}{495} = -1.96}$

$$Area = A_1 \ (< 2450) \ \underline{= 0.0250}$$

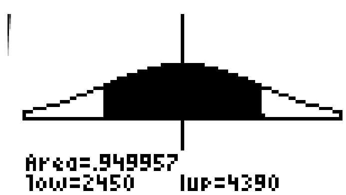
The area to the ***left*** of $x = 4390 \quad \Rightarrow \quad \underline{z = \dfrac{4390 - 3420}{495} = 1.96}$

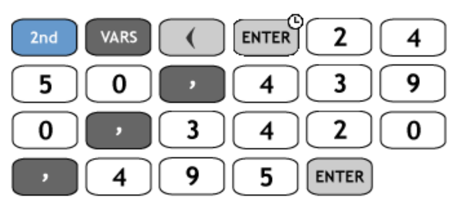
$$Area = A_2 \ (< 4390) \ \underline{= 0.9750}$$

The total shaded area:

$$Area = A_2 - A_1 = 0.9750 - 0.025 \ \underline{= 0.950}$$

[2ND] [VARS]





$$\boxed{ShadeNorm(low\ x,\ up\ x,\ \sigma)}$$

✓ We conclude that 95.00% of the babies do not require special treatment because they have birth weights between 2450 g and 4390 g. It follows that 5.00 % of the babies do require special treatment.

Helpful *Hints*

1. Don't confuse z scores and areas. z-scores are distances along the horizontal scale, but areas are regions under the normal curve. Table lists z-scores in the left column and across the top row, but areas are found in the body of the table.
2. Choose the correct (right/left) side of the graph.
3. A z-score must be negative whenever it is located in the left half of the normal distribution.
4. Areas (or probabilities) are positive or zero values, but they are never negative.

Procedure for Finding Values

1. Sketch a normal distribution curve, enter the given probability or percentage in the appropriate region of the graph, and identify the x value(s) being sought.
2. Use Table to find the z-score corresponding to the cumulative left area bounded by x. Refer to the body of Table to find the closest area, then identify the corresponding z score.
3. Using Formula, enter the values for μ, σ, and the z score found in step 2, then solve for x.

   $x = \mu + (z \bullet \sigma)$ (Another form of Formula)

   (If z is located to the left of the mean, be sure that it is a negative number.)
4. Refer to the sketch of the curve to verify that the solution makes sense in the context of the graph and the context of the problem.

Example

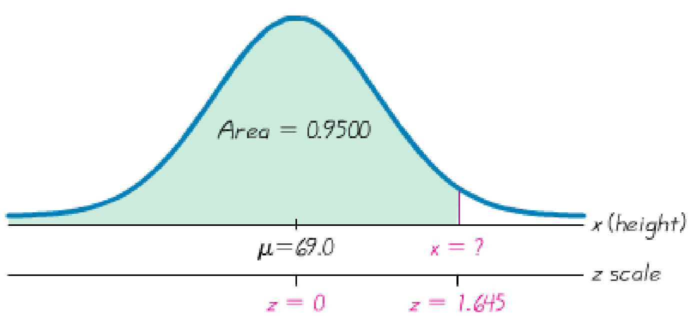
When designing an environment, one common criterion is to use a design that accommodates 95% of the population. How high should doorways be if 95% of men will fit through without bending or bumping their head? That is, find the 95th percentile of heights of men. Heights of men normally distributed with a mean of 69.0 in. and a standard deviation of 2.8 in.

Solution

Given: $\mu = 69.0, \quad \sigma = 2.8$

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449

From the table, we find the areas of 0.9495 and 0.9505. The area 0.95 corresponds to a z-score of 1.645.



$x = \mu + z \cdot \sigma = 69.0 + (1.645)(2.8) = 73.606$

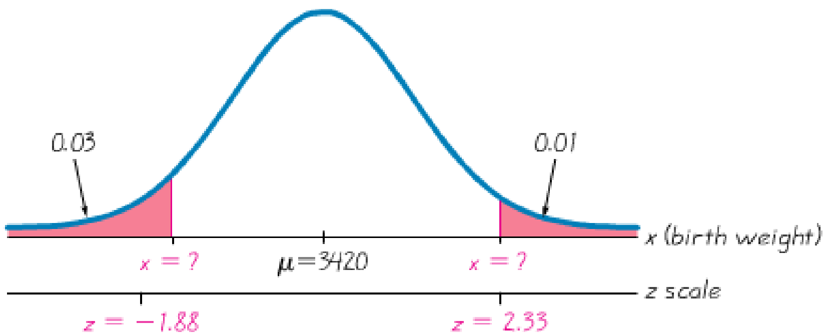
✓ A doorway height of 73.6 in. would allow 95% of men to fit without bending their head.

Example

Hospital wants to redefine the minimum and maximum birth weights that require special treatment because they are unusual low or unusual high. After considering relevant factors, a committee recommends special treatment for birth weights in the lowest 3% and the highest 1%. The committee members soon realize that specific birth weights need to be identified. Help this committee by finding the birth weights that separate the lowest 3% and the highest 1%. Birth weights in the U.S. are normally distributed with a mean of 3420 g and a standard deviation of 495 g.

<u>Solution</u>

Given: $\mu = 3420, \quad \sigma = 495$



For the leftmost value of *x*:

The cumulative area from the left is 0.03, from the table: $z = -1.88$

$x = \mu + z \cdot \sigma = 3420.0 + (-1.88)(495) = 2489.4$

For the rightmost value of *x*:

The cumulative area from the left is $1 - .01 = .99$, from the table: $z = 2.33$

$x = \mu + z \cdot \sigma = 3420.0 + (2.33)(495) = 4573.35$

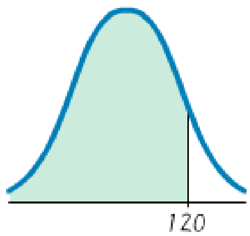
✓ The birth weight of 2489 g separates the lowest 3% of birth weights, and 4573 g separates the lowest 1% of birth weights. The hospital now has well-defined criteria for determining whether a newborn baby should be given special treatment for a birth weight that is unusual lo row high.

Exercises *Section* 2.8 – Applications Normal Distributions

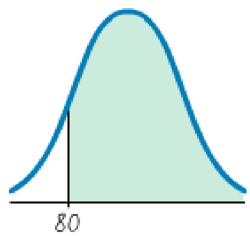
1. The distribution of IQ scores is a nonstandard normal distribution with mean of 100 and standard deviation of 15. What are the values of the mean and standard deviation after all IQ scores have been standardized by converting them to *z*-scores using $z = \dfrac{x-\mu}{\sigma}$?

2. Find the area of the shaded region. The graphs depict IQ scores adults, and those scores are normally distributed with mean of 100 and standard deviation of 15.

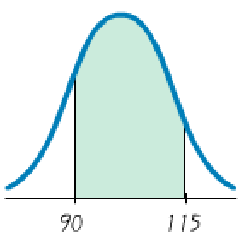
a)



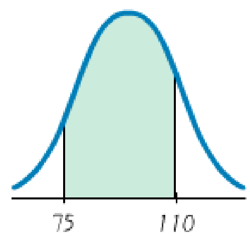
b)



c)

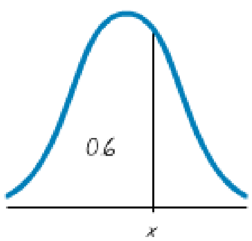


d)

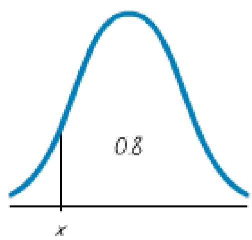


3. Find the Indicated IQ scores. The graphs depict IQ scores adults, and those scores are normally distributed with mean of 100 and standard deviation of 15.

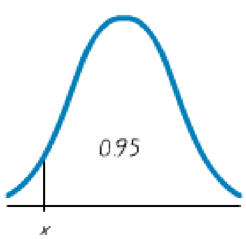
a)



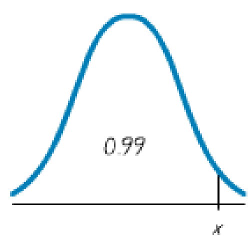
b)



c)



d)



4. Assume that adults have IQ scores that are normally distributed with mean of 100 and standard deviation of 15
   a) Find the probability that a randomly selected adult has an IQ that is less than 115.
   b) Find the probability that a randomly selected adult has an IQ that is greater than 131.5.
   c) Find the probability that a randomly selected adult has an IQ that is between 90 and 110.

d) Find the probability that a randomly selected adult has an IQ that is between 110 and 120.

e) Find P_{30} which is the IQ score separating the bottom 30% from the top 70%.

f) Find the first quartile Q_1 which is the IQ score separating the bottom 25% from the top 75%.

5. The Gulfstream 100 is an executive jet that seats six, and it has a doorway height of 51.6
   - ***Men's*** heights are normally distributed with mean 69.0 in. and standard deviation 2.8 in.
   - ***Women's*** heights are normally distributed with mean 63.6 in. and standard deviation 2.5 in.

   a) What percentage of adult men can fit through the door without bending?

   b) What percentage of adult women can fit through the door without bending?

   c) Does the door design with a height of 51.6 in. appear to be adequate? Why didn't the engineers design larder door?

   d) What doorway height would allow 60% of men to fit without bending?

6. Assume that human body temperatures are normally distributed with a mean of 98.20°F and a standard deviation of 0.62°F.

   a) A Hospital uses 100.6°F as the lowest temperature considered to be a fever. What percentage of normal and healthy persons would be considered to have fever? Does this percentage suggest that a cutoff of 100.6°F is appropriate?

   b) Physicians want to select a minimum temperature for requiring further medical tests. What should that temperature be, if we want only 5.0% of healthy people to exceed it? (Such a result is a false positive, meaning that the test result is positive, but the subject is not really sick.)

7. The lengths of pregnancies are normally distributed with a mean of 268 days and a standard deviation of 15 days.

   a) One classical use of the normal distribution is inspired by a letter to "Dear Abby" in which a wife claimed to have given birth 308 days after a brief visit from her husband. Given this information, find the probability of a pregnancy lasting 308 days or longer. What does the result suggest?

   b) If we stipulate that a baby is premature if the length of pregnancy is the lowest 4%, find the length that separates premature babies from those who are not premature. Premature babies often require special care, and this result could be helpful to hospital administrators in planning for that care.

8. A statistics professor gives a test and finds that the scores are normally distributed with a mean of 25 and a standard deviation of 5. She plans to curve the scores.

   a) If the curves by adding 50 to each grade, what is the new mean? What is the new standard deviation?

   b) Is it fair to curve by adding 50 to each grade? Why or why not?

   c) If the grades are curved according to the following scheme (instead of adding 50), find the numerical limits for each letter grade.

   A: Top 10%
   B: Scores above the bottom 70% and below the top 10%.
   C: Scores above the bottom 30% and below the top 30%.
   D: Scores above the bottom 10% and below the top 70%.

F: Bottom 10%.

d) Which method of curving the grades is fairer: Adding 50 to each grade or using the scheme given in part (*c*)? Explain.

Section 2.9 – Sampling Distributions and Estimators

The main objective of this section is to understand the concept of a sampling distribution of a statistic, which is the distribution of all values of that statistic when all possible samples of the same size are taken from the same population.

Definition

The ***sampling distribution of a statistic*** (such as the sample mean or sample proportion) is the distribution of all values of the statistic when all possible samples of the same size *n* are taken from the same population. (The sampling distribution of a statistic is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)

Definition

The ***sampling distribution of the mean*** is the distribution of sample means, with all samples having the same sample size *n* taken from the same population. (The sampling distribution of the mean is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)

Properties

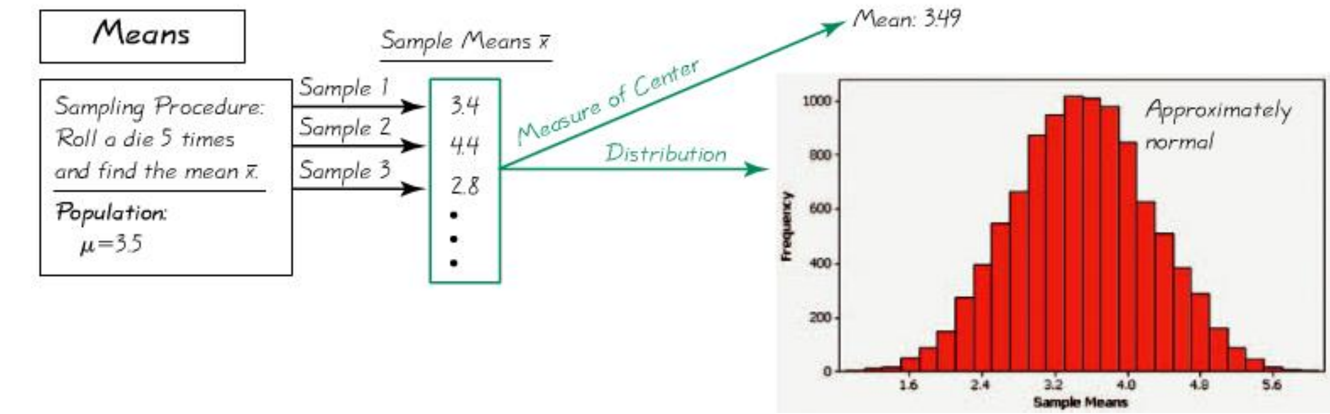
- Sample means target the value of the population mean. (That is, the mean of the sample means is the population mean. The expected value of the sample mean is equal to the population mean.)
- The distribution of the sample means tends to be a normal distribution.

Definition

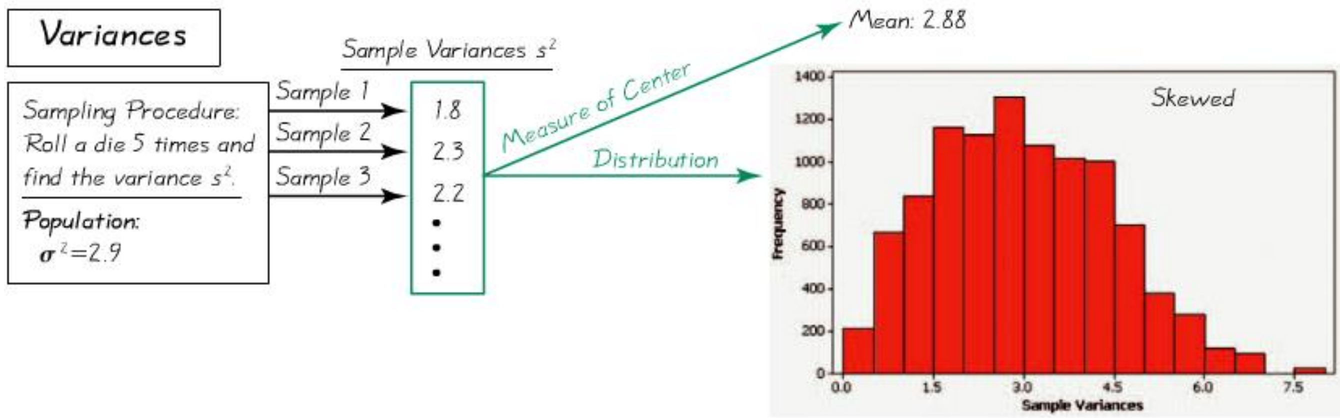
The ***sampling distribution of the variance*** is the distribution of sample variances, with all samples having the same sample size *n* taken from the same population. (The sampling distribution of the variance is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)

Example

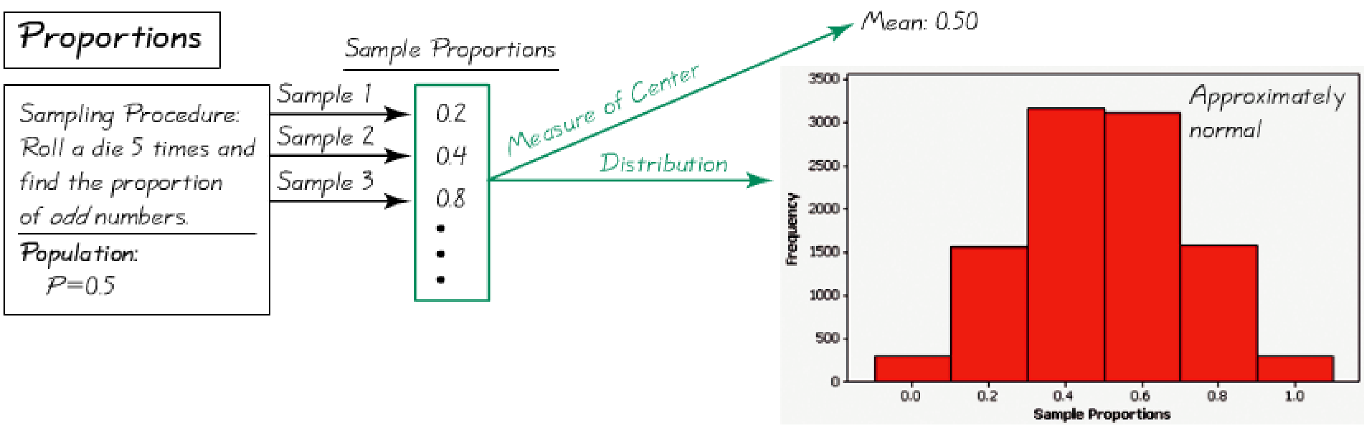
Consider repeating this process: Roll a die 5 times, find the mean , variance s^2, and the proportion of *odd* numbers of the results. What do we know about the behavior of all sample means that are generated as this process continues indefinitely? Specific Results from 10,000 trials



All outcomes are equally likely so the population mean is 3.5; the mean of the 10,000 trials is 3.49. If continued indefinitely, the sample mean will be 3.5. Also, notice the distribution is "normal."



All outcomes are equally likely so the population variance is 2.9; the mean of the 10,000 trials is 2.88. If continued indefinitely, the sample variance will be 2.9. Also, notice the distribution is "skewed to the right."



All outcomes are equally likely so the population proportion of odd numbers is 0.50; the proportion of the 10,000 trials is 0.50. If continued indefinitely, the mean of sample proportions will be 0.50. Also, notice the distribution is "approximately normal."

Example

Consider repeating this process: Roll a die 5 times and find the variances σ^2 of the results. What do we know about the behavior of all sample variances that are generated as this process continues indefinitely?

Solution

From the Variance table, it shows results from repeating this process 10,000 times, but the true sampling distribution variance involves repeating the process indefinitely. Because the values of 1, 2, 3, 4, 5, 6 are equally likely, the population has a variance of $\sigma^2 = 2.9$. The table shows that the sample variances have a mean of 2.88.
If the process is continued indefinitely, the mean of the sample variances will be 2.9.

Definition

The sampling distribution of the proportion is the distribution of sample proportions, with all samples having the same sample size *n* taken from the same population.

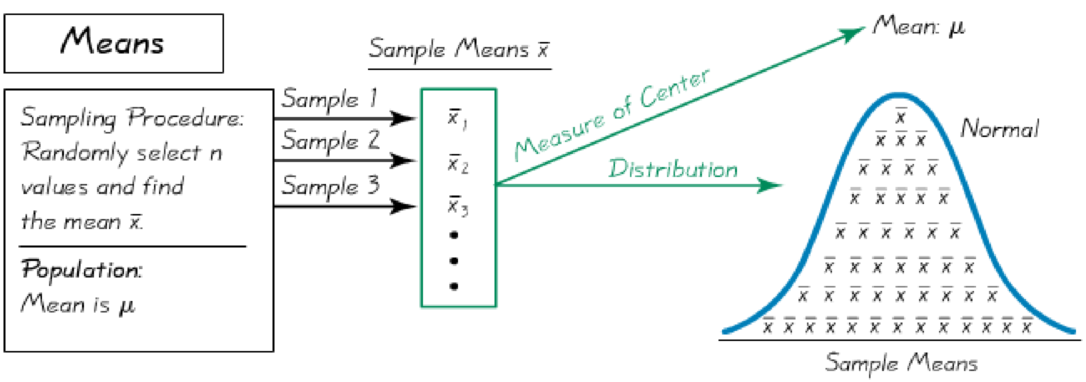
Notation for Proportions

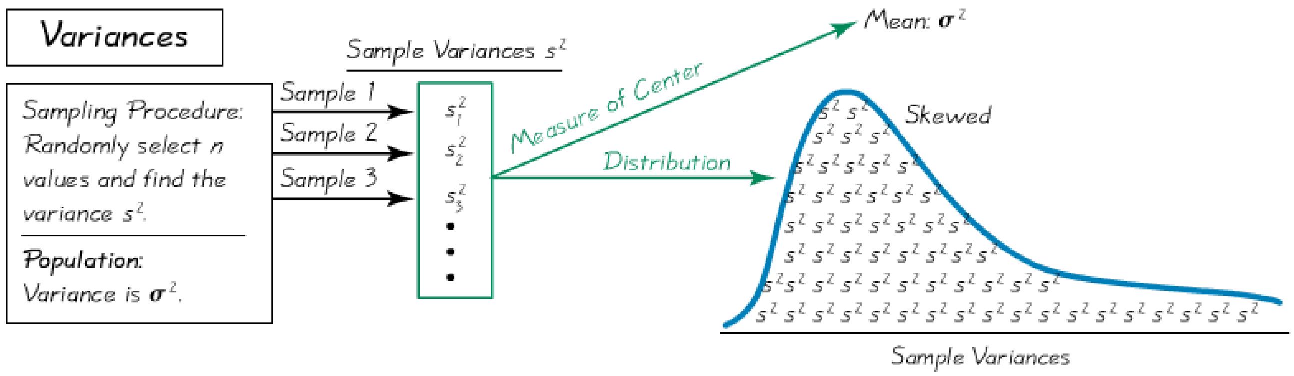
$p = \textit{population proportion}$

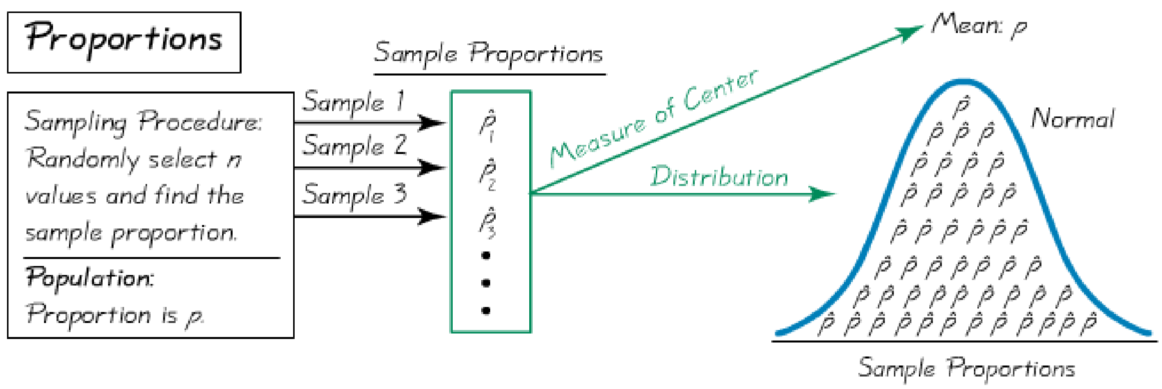
$\hat{p} = \textit{sample proportion}$

- Sample proportions target the value of the population proportion. (That is, the mean of the sample proportions is the population proportion. The expected value of the sample proportion is equal to the population proportion.)
- The distribution of the sample proportion tends to be a normal distribution.

General Behavior of Sampling Distributions







Unbiased Estimators

Sample means, variances and proportions are ***unbiased estimators***. That is they target the population parameter.

These statistics are better in estimating the population parameter

– *Mean* $\bar{x}$

– *Variance* s^2

– Pr*oportion* $\hat{p}$

Biased Estimators

Sample *medians*, *ranges* and *standard deviations* are biased estimators. That is they do NOT target the population parameter.

Note: the bias with the standard deviation is relatively small in large samples so *s* is often used to estimate.

Exercises *Section* 2.9 – Sampling Distributions and Estimators

1. You want to estimate the proportion of all U.S. college students who have the profound wisdom to take a statistics course. You obtain a simple random sample of students. Is the resulting sample proportion a good estimator of the population proportion? Why or why not?

2. Samples of size $n = 1000$ are randomly selected from the population of the last digits of telephone numbers. If the sample mean is found for each sample, what is the distribution of the sample means?

3. The ages (years) of the four U.S. presidents when they were assassinated in office are 56 (Lincoln), 49 (Garfield), 58 (McKinley), and 46 (Kennedy).
   a) Assuming that 2 of the ages are randomly selected with replacement, list the 16 different possible samples.
   b) Find the mean of each of the 16 samples; then summarize the sapling distribution of the means in the format of a table representing the probability distribution.
   c) Compare the population mean to the mean of the sample means.
   d) Do the sample means target the value of the population mean? In general, do sample means make good estimators of population means? Why or why not?

4. The ages (years) of the four U.S. presidents when they were assassinated in office are 56 (Lincoln), 49 (Garfield), 58 (McKinley), and 46 (Kennedy).
   a) Assuming that 2 of the ages are randomly selected with replacement, list the 16 different possible samples.
   b) Find the median of each of the 16 samples; then summarize the sapling distribution of the medians in the format of a table representing the probability distribution.
   c) Compare the population median to the median of the sample means.
   d) Do the sample medians target the value of the population mean? In general, do sample medians make good estimators of population medians? Why or why not?

5. After constructing a new manufacturing machine, 5 prototype integrated circuit chips are produced and it is found that 2 are defective (D) and 3 are acceptable (A). Assume that 2 of the chips are randomly selected *with replacement* from his population.
   a) After identify the 25 different possible samples, find the proportion of defects in each of them, then use a table to describe the sampling distribution of the proportions of defects.
   b) Find the mean of the sampling distribution.
   c) Is the mean of the sampling distribution (b) equal to the population proportion of defects? Does the mean of the sampling distribution of proportions always equal the population proportion?

6. Tell whether the following statistic is a biased or unbiased estimator of a population parameter
   a) Sample variance used to estimate a population variance.
   b) Sample mean used to estimate a population mean.
   c) Sample proportion used to estimate a population proportion.

Section 2.10 – Central Limit Theorem

The *Central Limit Theorem* tells us that for a population with *any* distribution, the distribution of the sample means approaches a normal distribution as the sample size increases.

1. For a population with any distribution, if n > 30, then the sample means have a distribution that can be approximated by a normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$
2. If $n \le 30$ and the original population has a normal distribution, then the sample means have a normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$
3. If $n \le 30$ and the original population does not have a normal distribution, then the methods of this section do not apply.

Central Limit Theorem

Given

1. The random variable *x* has a distribution (which may or may not be normal) with mean μ and standard deviation σ.
2. Simple random samples all of size *n* are selected from the population. (The samples are selected so that all possible samples of the same size *n* have the same chance of being selected.)

Conclusions

1. The distribution of sample means $\overline{x}$ will, as the sample size increases, approach a *normal* distribution.
2. The mean of the sample means is the population mean μ
3. The standard deviation of all sample means is $\frac{\sigma}{\sqrt{n}}$

Practical Rules Commonly Used

1. For samples of size *n* larger than 30, the distribution of the sample means can be approximated reasonably well by a normal distribution. The approximation gets closer to a normal distribution as the sample size *n* becomes larger.
2. If the original population is *normally distributed*, then for any sample size *n*, the sample means will be normally distributed (not just the values of *n* larger than 30).

Notation

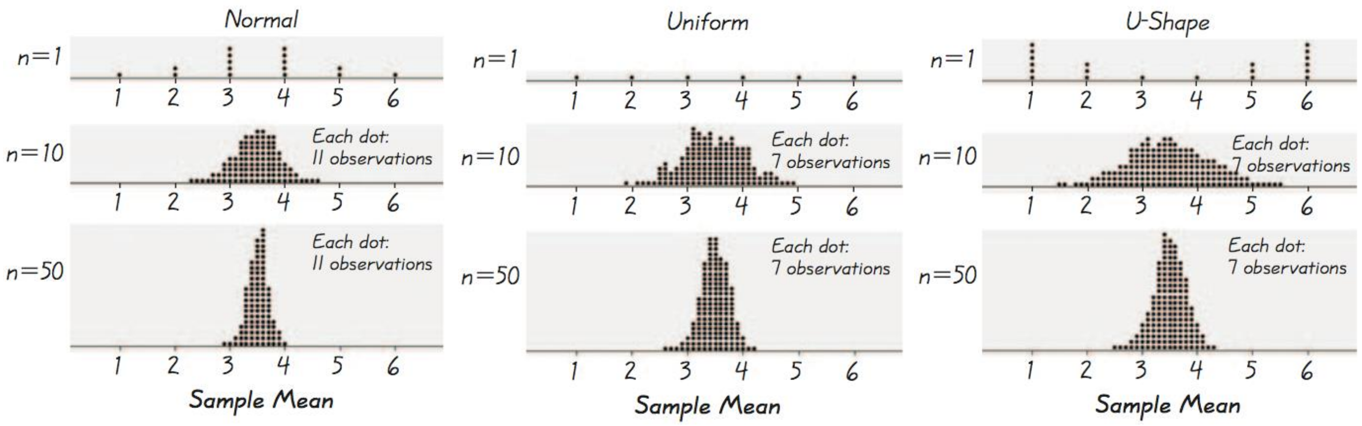
The mean of the sample means $\qquad \mu_{\overline{x}} = \mu$

The standard deviation of sample mean $\qquad \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$

($\sigma_{\overline{x}}$ is called the *standard error of the mean*)

Example

The top plots show an approximately normal distribution, a uniform distribution, and a distribution with a shape resembling the letter. In each column, the second dotplot shows the distribution of sample means where n = 10, and the bottom dotplot shows the distribution of sample means where n = 50
As we proceed from n = 1 to n = 50, we see that the distribution of sample means is approaching the shape of a normal distribution.



- As the sample size increases, the distribution of sample means tends to approach a normal distribution.
- The mean of the sample means is the same as the mean of the original population.
- As the sample size increases, the dot plots become narrower, showing that the standard deviation of the sample means becomes smaller.

Important Point

As the sample size increases, the sampling distribution of sample means approaches a normal distribution

Applying the Central Limit Theorem

Individual value: When working with individual value from a normally distributed population, use

$$z = \frac{x - \mu}{\sigma}$$

Sample value: When working with a mean for some sample (or group), use the value of $\frac{\sigma}{\sqrt{n}}$ for the standard deviation of the sample means. Use $z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

Example

Assume the population of weights of men is normally distributed with a mean of 172 lb. and a standard deviation of 29 lb.

a) Find the probability that if an *individual* man is randomly selected, his weight is greater than 175 lb.

b) b) Find the probability that *20 randomly selected men* will have a mean weight that is greater than 175 lb. (so that their total weight exceeds the safe capacity of 3500 pounds).

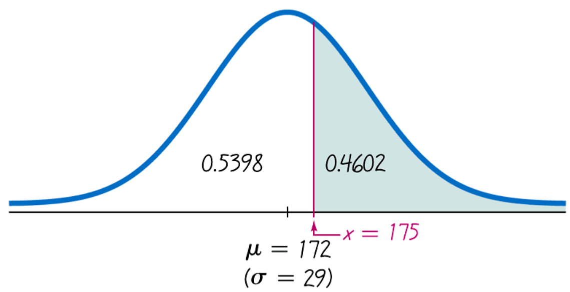
<u>Solution</u>

a) We are dealing with an individual value from a normally distributed population.

$$z = \frac{x-\mu}{\sigma} = \frac{175-172}{29} = 0.10$$

Using the Table (*Normal Distribution*), $A_1 = 0.5398$

Therefore, the shaded region area is: $A = 1 - 0.5398 \underline{= 0.4602}$



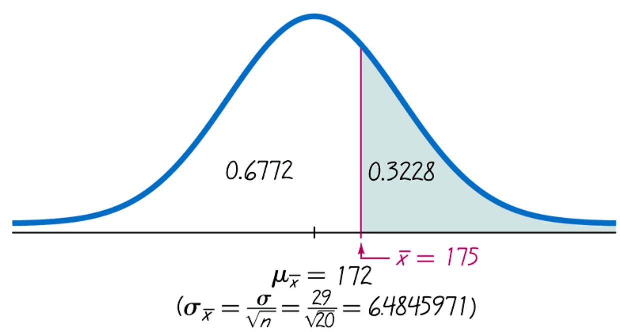
The probability of a randomly selected man weighing more than 175 lb. is 0.4602. (Calculator result is 0.4588)

b) Using the central limit theorem (not an individual man), although the sample size is not greater than 30. We use a normal distribution because the original population of men has a normal distribution.

$$\mu_{\bar{x}} = \mu = 172$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{29}{\sqrt{20}} = 6.4845971$$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{175-172}{6.4845971} = 0.46$$



$$A_2(<0.46) = 0.6772$$

Therefore, the shaded region area is: $A = 1 - 0.6772 \underline{= 0.3228}$

So, the probability that the 20 men have a mean weight greater than 175 lb. is 0.3228. (Calculator result is 0.3218)

✓ There is 0.4602 probability that an individual man will weigh more than 175 lb., and there is a 0.3228 probability that 20 men will have a mean weight of more than 175 lb. Given that the safe capacity of the water taxi is 3500 pounds, there is a fairly good chance (with probability 0.3228) that it will be overloaded with 20 randomly selected men. The capacity of 20 passengers is just not safe enough.

Example

Cans of regular Coke are labeled to indicate that they contain 12 *oz*. The corresponding sample statistics are n = 36 and $\bar{x} = 12.19\ oz$. If the Coke cans filled so that $\mu = 0.11\ oz$ and the population standard deviation is $\sigma = 0.11\ oz$ (based on the sample result), find the probability that a sample of 36 cans will have a mean of 12.19 *oz*. or greater. Do these results suggest that the Coke cans are filled with an amount greater than 12.00 *oz*.?

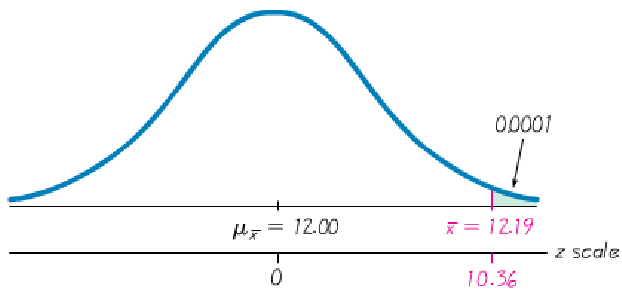
<u>Solution</u>

Because the sample size n = 36 exceeds 30, we apply the central limit theorem and conclude that the distribution of sample means is approximately a normal distribution with these parameters:

$$\mu_{\bar{x}} = \mu = 12$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.11}{\sqrt{36}} = 0.018333$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{12.19 - 12}{0.018333} = 10.36$$



Since the value of $z = 10.36$ is off the chart.
However, for values of z above 3.49, we use 0.9999 for the cumulative left area.
Therefore, we conclude that the shaded region is 0.0001.

✓ The result shows that there is an extremely small probability of getting a sample mean of 12.19 oz. or greater when 36 cans are randomly selected. It appears that the company has found a way to ensure that very few cans have less than 12 oz. while not wasting very much of their product.

Example

It is not totally unreasonable to think that screws labeled as being 3/4 in. in length would have a mean length that is somewhat close to 3/4 in. The lengths of a sample of 50 such screws have a mean length of 0.7468 in. Assume that the population of all such screws has a standard deviation described by $\sigma = 0.0123\ in$.

a) Assuming that the screws have mean length of 0.75 in. as labeled, find the probability that a sample of 50 screws has a mean length of 0.7468 in. or less.

b) The probability of getting a sample mean that is "at least as extreme as the given sample mean" is twice the probability found in part (*a*). Find this probability. (Note that the sample mean of 0.7468 in. misses the labeled mean of 0.75 in. by 0.0032 in., so any other mean is at least as extreme as the sample mean if it is below 0.75 in. by 0.0032 in. or more, or if it is above 0.75 in. by 0.0032 in. or more.)

c) Based on the result in part (b), does it appear that the sample mean misses the labeled mean of 0.75 in. by a significant amount? Explain.

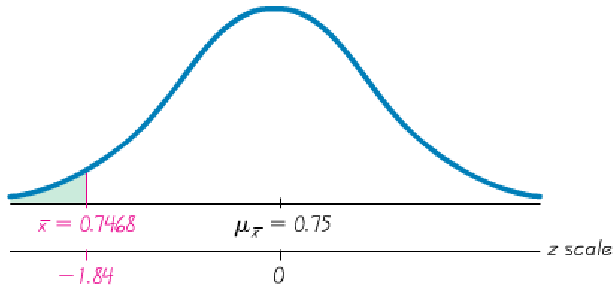
<u>Solution</u>

a) Because the sample size n = 50 exceeds 30, we apply the central limit theorem and conclude that the distribution of sample means is a normal distribution with these parameters:

$$\mu_{\bar{x}} = \mu = 0.75$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.0123}{\sqrt{50}} = 0.001739$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.7468 - 0.75}{0.001739} = -1.84$$



$$A(< -1.84) = 0.0329$$

The probability of getting a sample mean of 0.7468 in. or less is 0.0329.

b) $P(\text{at least as extreme as the given sample mean}) = 2P(\text{part } a)$

$$= 2 \times 0.0329$$

$$= 0.0658$$

c) The result from part (b) shows that there is a 0.0658 probability of getting a sample mean that is at least as extreme as the given sample. Using a 0.05 cutoff probability of 0.0658 exceeds 0.05, so the sample mean is not unusual. We conclude that the given sample mean does not miss the labeled mean of 0.75 in. by a substantial amount. The labeling of 0.75 in. appears to be justified.

Correction for a Finite Population

When sampling without replacement and the sample size n is greater than 5% of the finite population of size N (that is, $n > 0.05N$), adjust the standard deviation of sample means by multiplying it by the *finite population correction factor*:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \underbrace{\sqrt{\frac{N-n}{N-1}}}_{\text{finite population correction factor}}$$

Distribution of Sample Means

Population (with mean μ and standard deviation σ)	***Distribution of Sample Means***	***Mean of the Sample Means***	***Standard Deviation of the Sample Means***
Normal	Normal (for any sample size n)	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Not normal with $n > 30$	Normal (approximately)	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Not normal with $n \le 30$	Not normal	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Exercises *Section* 2.10 – Central Limit Theorem

1. Assume that SAT scores are normally distributed with mean $\mu = 1518$ and standard deviation $\sigma = 325$, use the Central Limit Theorem
   a) If 1 SAT score is randomly selected, find the probability that it is less than 1500.
   b) If 100 SAT scores are randomly selected, find the probability that they have a mean less than 1500.
   c) If 1 SAT score is randomly selected, find the probability that it is greater than 1600.
   d) If 64 SAT scores are randomly selected, find the probability that they have a mean greater than 1600.
   e) If 1 SAT score is randomly selected, find the probability that it is between 1550 and 1575.
   f) If 25 SAT scores are randomly selected, find the probability that they have a mean between 1550 and 1575.

2. Assume that weights of mean are normally distributed with a mean of 172 lb. and a standard deviation of 29 lb.
   a) Find the probability that if an individual man is randomly selected, his weight will be greater than 180 lb.
   b) Find the probability that 20 randomly selected men will have a mean weight that is greater than 180 lb.
   c) If 20 mean have a mean weight greater than 180 lb., the total weight exceeds the 3500 lb. safe capacity of a particular water taxi. Based on the proceeding results, is this safety concern? Why or why not?

3. Membership requires an IQ score above 131.5. Nine candidates take IQ tests, and their summary results indicated that their mean IQ score is 133, (IQ scores are normally distributed with mean of 100 and a standard deviation of 15.)
   a) If 1 person is randomly selected from the general population, find the probability of getting someone with an IQ score of at least 133.
   b) If 9 people are randomly selected, find the probability that their mean IQ score is at least 133.
   c) Although the summary results are available, the individual IQ test scores have been lost. Can it be concluded that all 9 candidates have IQ scores above 131.5 so that they are all eligible for membership?

4. For women aged 18-24, systolic blood pressures (in mm Hg) are normally distributed with a mean of 114.8 and a standard deviation of 13.1. Hypertension is commonly defined as a systolic blood pressure above 140.
   a) If a woman between the ages of 18 and 24 is randomly selected, find the probability that her systolic blood pressure is greater than 140.
   b) If 4 women in that age bracket are randomly selected, find the probability that their mean systolic blood pressure is greater than 140.
   c) Given that part (*b*) involves a sample size that is not larger than 30, why can the central limit theorem be used?

d) If a physician is given a report stating that 4 women have a mean systolic, blood pressure below 140, can she conclude that none of the women have hypertension (with a blood pressure greater than 140)?

5. Engineers must consider the breadths of male heads when designing motorcycle helmets. Men have head breaths that are normally distributed with a mean of 6.0 in. and a standard deviation of 1.0 in.
   a) If one male is randomly selected, find the probability that his head breadth is less than 6.2 in.
   b) The Safeguard Helmet company plans an initial production run of 100 helmets. Find the probability that 100 randomly selected men have a mean head breadth less than 6.2 in.
   c) The production manager sees the result from part (b) and reasons that all helmets should be made for men with head breadths less than 6.2 in., because they would fit all but a few men. What is wrong with that reasoning?

6. Currently, quarters have weighs that are normally distributed with a mean 5,670 g and a standard deviation of 0.062 g. A vending machine is configured to accept only those quarters with weights between 5.550 g and 5.790 g.
   a) If 280 different quarters are inserted into the vending machine, what is the expected number of rejected quarter?
   b) If 280 different quarters are inserted into the vending machine, what is the probability that the mean falls between the limits of 5.550 g and 5.790 g?
   c) If you own the vending machine, which result would concern you more? The result from part (*a*) or the result from part (*b*)? Why?

7. The annual precipitation amounts in a certain mountain range are normally distributed with a mean of 100 inches, and a standard deviation of 12 inches. What is the probability that the mean annual precipitation during 36 randomly picked years will be less than 112.8 inches?

8. The annual precipitation amounts in a certain mountain range are normally distributed with a mean of 72 inches, and a standard deviation of 14 inches. What is the probability that the mean annual precipitation during 49 randomly picked years will be less than 74.8 inches?

9. The weights of the fish in a certain lake are normally distributed with a mean of 13 lb. and a standard deviation of 6. If 4 fish are randomly selected, what is the probability that the mean weight will be between 10.6 and 16.6 lb.?

10. For women aged 18-24, systolic blood pressures (in mm Hg) are normally distributed with a mean of 114.8 and a standard deviation of 13.1. If 23 women aged 18-24 are randomly selected, find the probability that their mean systolic blood pressure is between 119 and 122.

11. A study of the amount of time it takes a mechanic to rebuild the transmission for 2005 Chevy shows that the mean is 8.4 hours and the standard deviation is 1.8 hours. If 40 mechanics are randomly selected, find the probability that their mean rebuild time
   a) Exceeds 8.7 hours.
   b) Exceeds 8.1 hours.

12. A final exam in Math 160 has a mean of 73 with standard deviation 7.8. If 24 students are randomly selected, find the probability that the mean of their test scores is greater than 71.

Section 2.11 – Normal Approximations to Binomial

If the conditions of $np \ge 5$ and $nq \ge 5$ are both satisfied, then probabilities from a binomial probability distribution can be approximated well by using a normal distribution with mean $\mu = np$ and standard deviation $\sigma = \sqrt{npq}$

Procedure for Using a Normal Distribution to Approximate a Binomial Distribution

1. Verify that both $np \ge 5$ and $nq \ge 5$. If not, you must use software, a calculator, a table or calculations using the binomial probability formula.
2. Find the values of the parameters μ and σ by calculating $\mu = np$ and $\sigma = \sqrt{npq}$.
3. Identify the discrete whole number x that is relevant to the binomial probability problem. Focus on this value temporarily.
4. Draw a normal distribution centered about μ, then draw a *vertical strip area* centered over x. Mark the left side of the strip with the number equal to $x – 0.5$, and mark the right side with the number equal to $x + 0.5$. *Consider the entire area of the entire strip to represent the probability of the discrete whole number itself.*
5. Determine whether the value of x itself is included in the probability. Determine whether you want the probability of at least x, at most x, more than x, fewer than x, or exactly x. Shade the area to the right or left of the strip; also shade the interior of the strip *if and only if x itself* is to be included. This total shaded region corresponds to the probability being sought.
6. Using $x – 0.5$ or $x + 0.5$ in place of x, find the area of the shaded region: find the z score; use that z score to find the area to the left of the adjusted value of x; use that cumulative area to identify the shaded area corresponding to the desired probability.

Example

The author was mailed a survey, and the survey included a request for an e-mail address. Assume that the survey was sent to 40,000 people and that for such surveys, the percentage of responses with an e-mail address is 3%. If the true goal of the survey was to acquire a bank of at least 1150 e-mail addresses, find the probability of getting at least 1150 responses with e-mail addresses.

<u>Solution</u>

It is a binomial distribution with a fixed number of trials (n = 40,000), which are independent.
There are 2 categories for each survey: a response is obtained with an e-mail address or it is not,
The probability of success $(p = 0.03)$ remains constant from trial to trial.

$q = 1 - p = 1 - .003 = 0.97$

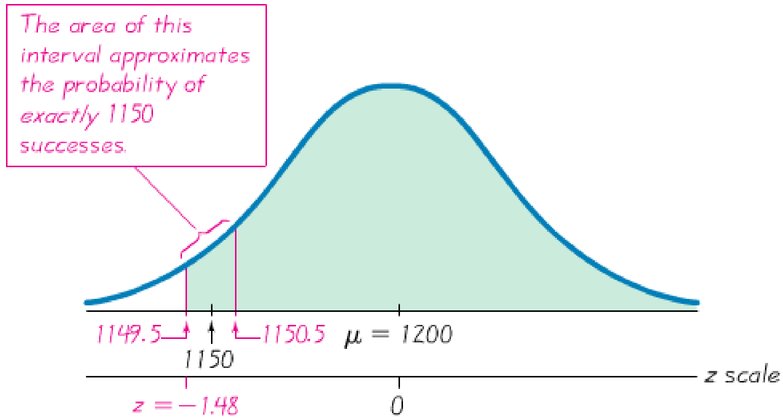
$np = 40,000 \times 0.03 = 1200 \qquad (Therefore \;\; np \ge 5)$

$nq = 40,000 \times 0.97 = 38,800 \qquad (Therefore \;\; np \ge 5)$

$\mu = np = 40,000 \cdot 0.03 \;\underline{= 1200}|$

$\sigma = \sqrt{npq} = \sqrt{40,000 \cdot 0.03 \cdot 0.97} \;\underline{= 34.11744}|$

We want the probability of at least 1150 responses with e-mail addresses so x = 1150 is the discrete whole number relevant.



The area of the interval approximates the probability of 1150: from 1149.5 to 1150.5.

For the boundary 1149.5:

$$z = \frac{x-\mu}{\sigma} = \frac{1149.5-1200}{34.117444} = -1.48$$

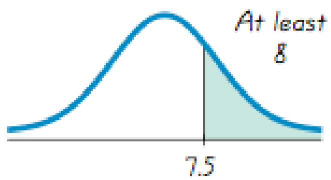
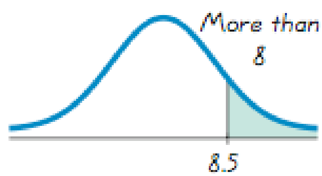
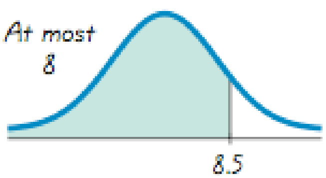
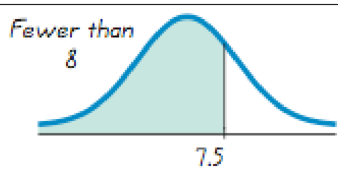
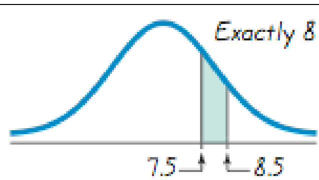
$$A(>1149.5) = 1 - A(z=-1.48) = 1 - 0.0694 = 0.9306$$

There is 0.9306 probability of getting at least 1150 responses with e-mail addresses among 40,000 surveys that were mailed. This probability is high enough to conclude that it is very likely that will be attained their goal of at least 1150 responses with e-mail addresses.

Continuity Correction

Definition

When we use the normal distribution (which is a ***continuous*** probability distribution) as an approximation to the binomial distribution (which is *discrete*), a continuity correction is made to a discrete whole number x in the binomial distribution by representing the discrete whole number x by the interval from x – 0.5 to x + 0.5 (that is, adding and subtracting 0.5).

	x = at least 8 (*includes* 8 *and above*)	To the right of 7.5
	x = more than 8 (*doesn't include* 8)	To the left of 8.5
	x = at most 8 (*includes* 8 *and below*)	To the right of 8.5
	x = fewer than 8 (*doesn't include* 8)	To the left of 7.5
	x = exactly 8	Between 7.5 and 8.5

Example

A recent survey showed that among 2822 randomly selected adults, 2060 (or 73%) stated that they are internet users. If the proportion of all adults using the Internet is actually 0.75, find the probability that a random sample of 2822 adults will result in exactly 2060 Internet users.

Solution

Given: $n = 2822, \quad x = 2060, \quad p = 0.75$

$q = 1 - p = 1 - .75 = 0.25$

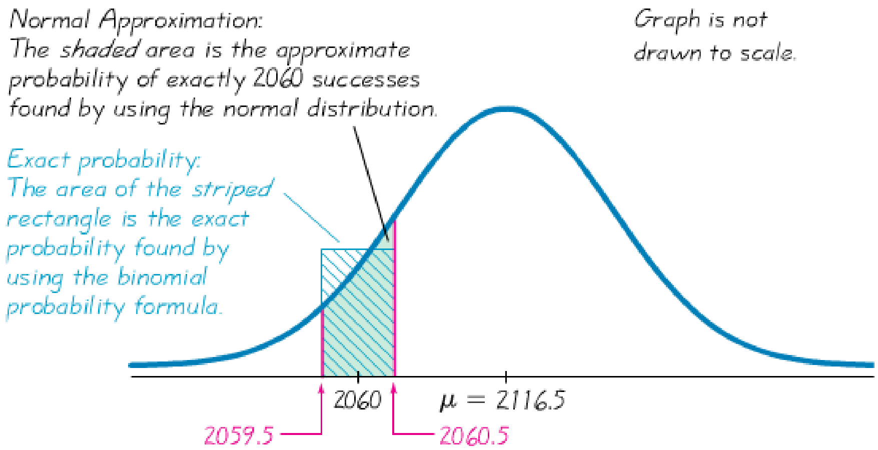
$np = 2822 \times 0.75 = 2116.5 \qquad (Therefore \ \ np \geq 5)$

$nq = 2822 \times 0.25 = 705.5 \qquad (Therefore \ \ np \geq 5)$

$\mu = np = 2822 \cdot 0.75 = 2116.5$

$$\sigma = \sqrt{npq} = \sqrt{2822 \cdot 0.75 \cdot 0.25} \; \underline{= 23.002717|}$$

We want the probability of at exactly 2060 internet users, so the discrete whole number relevant is 2060.



The area of the interval approximates the probability of 2060: from 2059.5 to 2060.5.

For the boundary 2059.5:

$$z = \frac{x-\mu}{\sigma} = \frac{2059.5 - 2116.5}{23.002717} \; \underline{= -2.48|}$$

$$A(z = -2.48) \; \underline{= 0.0066|}$$

For the boundary 2060.5:

$$z = \frac{x-\mu}{\sigma} = \frac{2060.5 - 2116.5}{23.002717} \; \underline{= -2.43|}$$

$$A(z = -2.43) \; \underline{= 0.0075|}$$

The shaded area is 0.0075 – 0.0066 = 0.0009.

✓ The probability of getting exactly 2060 Internet users among 2822 randomly selected adults is 0.0009. This probability tells us that if the percentage of Internet users in the adult population is 75% then it is highly unlikely that we will get exactly 2060 Internet users when we survey 2822 adults. The probability of any single number of Internet users will be very small.

Exercises *Section* 2.11 – Normal Approximations to Binomial

1. The Wechsler test is used to measure IQ scores. It is designed so that the mean IQ score 1—and the standard deviation is 15. It is known that IQ scores have a normal distribution. Assume that we want to find the probability that a randomly selected person has an IQ equal to 107. What is the continuity correction, and how would it be applied in finding that probability?

2. The Genetics & IVF Institute has developed methods for helping coupes determine the gender of their children. For comparison, a large sample of randomly selected families with four children is obtained, and the proportion of girls in each family is recorded. If the normal distribution a good approximation of the distribution of those proportions? Why or why not?

3. The given values are discrete. Use the continuity correction and describe the region of the normal distribution that corresponds to the indicated probability.
   The probability of "more than 30 defective items" corresponds to the area of the normal curve – The area to the right of 20.5.
   a) Probability of more than 8 Senators who are women.
   b) Probability of at least 2 traffic tickets this year.
   c) Probability of fewer than 5 passengers who do not show up for a flight.
   d) Probability that the number of students who are absent is exactly 4
   e) Probability that the number of defective computer power supplies is between 12 and 16 inclusive.
   f) Probability that exactly 24 felony indictments result in convictions.

4. Using Normal Approximation. Do the following
   i. Find the indicated *binomial probability* using the Table.
   ii. If $np \geq 5$ and $nq \geq 5$, also estimate the indicated probability by using the normal distribution as an approximation to the binomial distribution; if $np < 5 \quad or \quad nq < 5$, then state that the normal approximation is not suitable.
   a) $With\ n = 10\ \ and\ \ p = 0.5,\ \ find\ P(3)$
   b) $With\ n = 12\ \ and\ \ p = 0.8,\ \ find\ P(9)$
   c) $With\ n = 8\ \ and\ \ p = 0.9,\ \ find\ P(at\ least\ 6)$

5. In a test of the XSORT method to increase the probability of conceiving a girl. Among 574 women using that method, 525 had baby girls. Assuming that the method has no effect so that boys and girls are equally likely, find the probability of getting at least 525 girls among 574 babies. Does the result suggest that the XSORT method is effective? Why or why not?

6. In a test of the YSORT method to increase the probability of conceiving a boy. Among 152 women using that method, 127 had baby boys. Assuming that the method has no effect so that boys and girls are equally likely, find the probability of getting at least 127 boys among 152 babies. Does the result suggest that the YSORT method is effective? Why or why not?

7. When Mendel conducted his famous hybridization experiments, he used peas with green pods and yellow pods/ One experiment involved crossing peas in such a way that 25% (or 145) of the 580 offspring peas were expected to have yellow pods. Instead of getting 145 peas with yellow pods, he obtained 152. Assume that Mendel's 25% rate is correct
   a) Find the probability that among the 580 offspring peas, exactly 152 have yellow pods.
   b) Find the probability that among the 580 offspring peas, at least 152 have yellow pods.
   c) Which result is useful for determining whether Mendel's claimed rate of 25% is incorrect? ((*a*) or (*b*))
   d) Is there strong evidence to suggest that Mendel's rate of 25% is incorrect?

8. In a study of 420,095 cell phone users in Denmark, it was found that 135 developed cancer of the brain or nervous system. If we assume that the use of cell phones has no effect on developing such cancer, then the probability of a person having such cancer is 0.000340. We therefore expect about 143 cases of such cancer in a group of 420,095 randomly selected people. Estimate the probabitlity of 135 or fewer cases of such cancer in a group of 420,095 people. What do these results suggest about media reports that cell phones cause cancer of the brain or nervous system?

9. There is an 80% chance that a prospective employer will check the educational background of a job applicants, find the probability that exactly 85 have their educational backgrounds checked.

10. When working as investigator, you analyzed the leading digits of amounts on checks from companies that were suspected of fraud. Among 784 checks, 479 had amounts with leading digits of 5, but checks issued in the normal course of honest transactions were expected to have 7.9% of the checks with amounts having leading digits of 5. Is there strong evidence to indicate that the check amounts are significantly different from amounts that are normally expected? Explain?

11. The probability of flu symptoms for a person not receiving any treatment is 0.019. In a clinical trial of Lipitor, a drug commonly used to lower cholesterol, 863 patients were given a treatment of 10-mg atorvastatin tablets, and 19 of those patients experienced flu symptoms. Assuming that these tablets have no effect on flu symptoms, estimate the probability that at least 19 of 863 people experience flu symptoms. What do these results suggest about the flu symptoms as an adverse to the drug?

12. Assume that a baseball player hits .350, so his probability of a hit is 0.350. Also assume that his hitting attempts are independent of each other.
   a) Find the probability of at least 1 hit in 4 tries in a single game.
   b) Assuming that this batter gets up to bat 4 times each game, find the probability of getting a total of at least 56 hits in 56 games.
   c) Assuming that this batter gets up to bat 4 times each game, find the probability of getting a total of at least 1 hit in 56 consecutive games (Joe DiMaggio's 1941 record).

13. Estimate the probability of getting exactly 43 boys in 90 births.

14. The given values are discrete. Use the continuity correction and describe the region of the normal distribution that corresponds to the indicated probability.
    a) The probability of at least 44 boys.
    b) The probability of fewer than 33 democrats.
    c) The probability of exactly 51 green marbles.
    d) The probability that the number of correct answers is between 15 and 43 inclusive

15. For the binomial distribution with the given values for *n* and *p*, state whether or not it is suitable to use the normal distribution as an approximation: $n = 21 \quad and \quad p = 0.6$

16. A multiple choice test consists of 60 questions. Each question has 4 possible answers of which one is correct If all answers are random guesses, estimate the probability of getting at least 20% correct.