A table of **Standard Normal Probabilities** for this distribution is included in this book and in any basic statistics text. We used these tables when doing some normal curve problems in Chapter 6. Standard normal probabilities, and other normal probabilities, are also accessible on many calculators. We will use a table of standard normal probabilities as well as technology to solve probability problems, which are very similar to the problems we did in Chapter 6 involving the normal distribution.

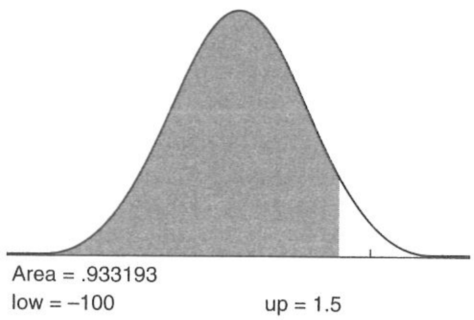
Using the tables, we can determine that the percentages in the empirical rule (the 68–95–99.7 rule) are, more precisely, 68.27%, 95.45%, 99.73%. The TI-83/84 syntax is for the standard normal is `normalcdf(lower bound, upper bound)`. Thus, the area between $z = -1$ and $z = 1$ in a standard normal distribution is `normalcdf(-1,1)=0.6826894809`.

Normal Probabilities

When we know a distribution is approximately normal, we can solve many types of problems.

example: In a standard normal distribution, what is the probability that $z < 1.5$? (Note that because z is a CRV, $P(X = a) = 0$, so this problem could have been equivalently stated "what is the probability that $z \leq 1.5$?")

solution: The standard normal table gives areas to the left of a specified z-score. From the table, we determine that the area to the left of $z = 1.5$ is 0.9332. That is, $P(z < 1.5) = 0.9332$. This can be visualized as follows:



Calculator Tip: The above image was constructed on a TI-83/84 graphing calculator using the `ShadeNorm` function in the `DISTR DRAW` menu. The syntax is `ShadeNorm (lower bound, upper bound, [mean, standard deviation])`—only the first two parameters need be included if we want standard normal probabilities. In this case we have `ShadeNorm(−100,1.5)` and press `ENTER` (not `GRAPH`). The lower bound is given as – 100 (any large negative number will do—there are very few values more than three or four standard deviations from the mean). You *will* need to set the WINDOW to match the mean and standard deviation of the normal curve being drawn. The WINDOW for the previous graph is [–3.5, 3.5, 1, –0.15, 0.5, 0.1, 1].

example: It is known that the heights (X) of students at Downtown College are approximately normally distributed with a mean of 68 inches and a standard deviation of 3 inches. That is, X has $N(68,3)$. Determine

(a) $P(X < 65)$.

solution: $P(X < 65) = P\left(z < \frac{65-68}{3} = -1\right) = 0.1587$ (the area to the left of $z = -1$

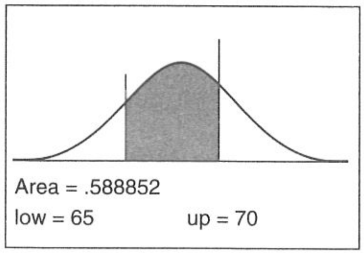
from Table A). On the TI-83/84, the corresponding calculation is `normalcdf(-100,-1) = normalcdf(-1000,65,68,30) = 0.1586552596`.

(b) $P(X > 65)$.

solution: From part (a) of the example, we have $P(X < 65) = 0.1587$. Hence, $P(X > 65) = 1 - P(X < 65) = 1 - 0.1587 = 0.8413$. On the TI-83/84, the corresponding calculation is `normalcdf(-1,100) = normalcdf(65,1000,68,3) = 0.8413447404`.

(c) $P(65 < X < 70)$.

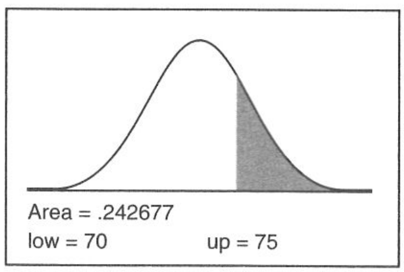
solution: $P(65 < X < 70) = P\left(\frac{65-68}{3} < z < \frac{70-68}{3}\right) = P(-1 < z < 0.667) =$ $0.7486 - 0.1587 = 0.5899$ (from Table A, the geometry of the situation dictates that the area to the left of $z = -1$ must be subtracted from the area to the left of $z = 0.667$). Using the TI-83/84, the calculation is `normalcdf(−1,0.67) = normalcdf (65,70,68,3)=0.5889`. This situation is pictured below.



Note that there is some rounding error when using Table A (see Appendix). In part (c), $z = 0.66667$, but we must use 0.67 to use the table.

(d) $P(70 < X < 75)$

solution: Now we need the area between 70 and 75. The geometry of the situation dictates that we subtract the area to the left of 70 from the area to the left of 75. This is pictured below.



We saw from part (c) that the area to the left of 70 is 0.7486. In a similar fashion, we find that the area to the left of 75 is 0.9901 (based on $z = 2.33$). Thus $P(70 < X < 75) = 0.9901 - 0.7486 = 0.2415$. The calculation on the TI-83/84 is: `normalcdf(70,75,68,3) = 0.2427`. The difference in the answers is due to rounding.

example: SAT scores are approximately normally distributed with a mean of about 500 and a standard deviation of 100. Laurie needs to be in the top 15% on the SAT in order to ensure her acceptance by Giant U. What is the minimum score she must earn to be able to start packing her bags for college?

solution: This is somewhat different from the previous examples. Up until now, we have been given, or have figured out, a *z*-score, and have needed to determine an area. Now, we are given an area and are asked to determine a particular score. If we are using the table of normal probabilities, it is a situation in which we must read from inside the table out to the *z*-scores rather than from the outside in. If the particular value of X we are looking for is the lower bound for the top 15% of scores, then there are 85% of the scores to the left of x. We look through the table and find the closest entry to 0.8500 and determine it to be 0.8508. This corresponds to a *z*-score of 1.04. Another way to write the *z*-score of the desired value of X is

$$z = \frac{x - 500}{100}.$$

Thus, $z = \dfrac{x-500}{100} = 1.04.$

Solving for x, we get $x = 500 + 1.04(100) = 604$. So, Laurie must achieve an SAT score of at least 604. This problem can be done on the calculator as follows: `invNorm(0.85,500,100)`.

Most problems of this type can be solved in the same way: express the *z*-score of the desired value in two different ways (from the definition; finding the actual value from Table A or by using the invNorm function on the calculator), then equate the expressions and solve for x.

Simulation and Random Number Generation

Sometimes probability situations do not lend themselves easily to analytical solutions. In some situations, an acceptable approach might be to run a **simulation.** A simulation utilizes some random process to conduct numerous trials of the situation and then counts the number of successful outcomes to arrive at an estimated probability. In general, the more trials, the more confidence we can have that the relative frequency of successes accurately approximates the desired probability. The **law of large numbers** states that the proportion of successes in the simulation should become, over time, close to the true proportion in the population.

One interesting example of the use of simulation has been in the development of certain "systems" for playing Blackjack. The number of possible situations in Blackjack is large but finite. A computer was used to conduct thousands of simulations of each possible playing decision for each of the possible hands. In this way, certain situations favorable to the player were identified and formed the basis for the published systems.

example: Suppose there is a small Pacific Island society that places a high value on families having a baby girl. Suppose further that *every* family in the society decides to keep having children until they have a girl and then they stop. If the first child is a girl, they are a one-child family, but it may take several tries before they succeed. Assume that when this policy was decided on that the proportion of girls in the population was 0.5 and the probability of having a girl is 0.5 for each birth. Would this behavior change the proportion of girls in the population? Design a simulation to answer this question.

solution: Use a random number generator, say a fair coin, to simulate a birth. Let heads = "have a girl" and tails = "have a boy." Flip the coin and note whether it falls heads or tails. If it falls heads, the trial ends. If it falls tails, flip again because this represents having a boy. The outcome of interest is the number of trials (births) necessary until a girl is born (if the third flip gives the first head, then $x = 3$). Repeat this many times and determine how many girls and how many boys have been born.

If flipping a coin many times seems a bit tedious, you can also use your calculator to simulate flipping a coin. Let 1 be a head and let 2 be a tail. Then enter `MATH PRB randInt(1,2)` and press `ENTER` to generate a random 1 or 2. Continue to press `ENTER` to generate additional random integers 1 or 2. Enter `randInt(1,2,n)` to generate n random integers, each of which is a 1 or a 2. Enter `randInt (a,b,n)` to generate n random integers X such that $a \le X \le b$.

The following represents a few trials of this simulation (actually done using the random number generator on the TI-83/84 calculator):

Trial #	Trial Results (H = "girl")	# Flips until first girl	Total # of girls after trial is finished	Total # of boys after trial is finished
1	TH	2	1	1
2	H	1	2	1
3	TTTH	4	3	4
4	H	1	4	4
5	TH	2	5	5
6	H	1	6	5
7	H	1	7	5
8	H	1	8	5
9	TH	2	9	6
10	H	1	10	6
11	TTTH	4	11	9
12	H	1	12	9
13	H	1	13	9
14	TTTTH	5	14	13
15	TTH	3	15	15

This limited simulation shows that the number of boys and girls in the population are equal. In fairness, it should be pointed out that you usually won't get exact results in a simulation such as this, especially with only 15 trials, but this time the simulation gave the correct answer: the behavior would not change the proportion of girls in the population.

TIP

Exam Tip: If you are asked to do a simulation on the AP Statistics exam (and there have been such questions), use a table of random numbers rather than the random number generator on your calculator. This is to make your solution understandable to the person reading your solution. A table of random numbers is simply a list of the whole numbers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 appearing in a random order. This means that each digit should appear approximately an equal number of times in a large list and the next digit should appear with probability 1/10 no matter what sequence of digits has preceded it.

The following gives 200 outcomes of a typical random number generator separated into groups of 5 digits:

79692	51707	73274	12548	91497	11135	81218	79572	06484	87440
41957	21607	51248	54772	19481	90392	35268	36234	90244	02146
07094	31750	69426	62510	90127	43365	61167	53938	03694	76923
59365	43671	12704	87941	51620	45102	22785	07729	40985	92589

example: A coin is known to be biased in such a way that the probability of getting a head is 0.4. If the coin is flipped 50 times, how many heads would you expect to get?

solution: Let 0, 1, 2, 3 be a head and 4, 5, 6, 7, 8, 9 be a tail. If we look at 50 digits beginning with the first row, we see that there are 18 heads (bold-faced below), so the proportion of heads is 18/50 = 0.36. This is close to the expected value of 0.4.

7969**2** 5**1**7**0**7 7**32**74 **12**548 9**1**497 **1113**5 8**121**8 7957**2** **0**6484 8744**0**

Sometimes the simulation will be a **wait-time simulation.** In the example above, we could have asked how long it would take, on average, until we get five heads. In this case, using the same definitions for the various digits, we would proceed through the table until we noted five even numbers. We would then write down how many digits we had to look at. Three trials of that simulation might look like this (individual trials are separated by \\):

7969**2** 5**1**7**0**7 7**32**\\74 **12**548 9**1**497 **11****13**5 8**121**\\.

So, it took 13, 14, and 7 trials to get our five heads, or an average of 11.3 trials (the theoretical expected number of trials is 12.5).

Calculator Tip: There are several random generating functions built into your calculator, all in the `MATH PRB` menu: `rand`, `randInt`, `randNorm`, and `randBin`. `rand(k)` will return *k* random numbers between 0 and 1; `randInt(lower bound,upper bound,k)` will return *k* random integers between *lower bound* and *upper bound* inclusive; `randNorm(mean, standard deviation, k)` will return *k* values from a normal distribution with mean *mean* and standard deviation *standard deviation*; `randBin(n,p,k)` returns *k* values from a binomial random variable having *n* trials each with probability of success *p*.

Remember that you will not be able to use these functions to do a required simulation on the AP exam, although you can use them to do a simulation of your own design.

TIP

Exam Tip: You may see probability questions on the AP exam that you choose to do by a simulation rather than by traditional probability methods. As long as you explain your simulation carefully and provide the results for a few trials, this approach is usually acceptable. If you do design a simulation for a problem where a simulation is not REQUIRED, you *can* use the random number generating functions on your calculator. Just explain clearly what you have done—clearly enough that the reader could replicate your simulation if needed.

Transforming and Combining Random Variables

If X is a random variable, we can transform the data by adding a constant to each value of X, multiplying each value by a constant, or some linear combination of the two. We may do this to make numbers more manageable. For example, if values in our dataset ranged from 8500 to 9000, we could subtract, say, 8500 from each value to get a dataset that ranged from 0 to 500. We would then be interested in the mean and standard deviation of the new dataset as compared to the old dataset.

Some facts from algebra can help us out here. Let μ_x and σ_x be the mean and standard deviation of the random variable X. Each of the following statements can be algebraically verified if we add or subtract the same constant, a, to each term in a dataset ($X \pm a$), or multiply each term by the same constant b (bX), or some combination of these ($a \pm bX$):

- $\mu_{a \pm bX} = a \pm b\mu_x$.
- $\sigma_{a \pm bX} = b\sigma_X$ ($\sigma^2_{a \pm bx} = b^2\sigma^2_X$).

example: Consider a distribution with $\mu_X = 14$, $\sigma_X = 2$. Multiply each value of X by 4 and then add 3 to each. Then $\mu_{3+4X} = 3 + 4(14) = 59$, $\sigma_{3+4X} = 4(2) = 8$.

Rules for the Mean and Standard Deviation of Combined Random Variables

Sometimes we need to combine two random variables. For example, suppose one contractor can finish a particular job, on average, in 40 hours ($\mu_x = 40$). Another contractor can finish a similar job in 35 hours ($\mu_y = 35$). If they work on two separate jobs, how many hours, on average, will they bill for completing both jobs? It should be clear that the average of $X + Y$ is just the average of X plus the average for Y. That is,

- $\mu_{X \pm Y} = \mu_X \pm \mu_Y$.

The situation is somewhat less clear when we combine variances. In the contractor example above, suppose that

$$\sigma^2_X = 5 \text{ and } \sigma^2_Y = 4.$$

Does the variance of the sum equal the sum of the variances? Well, yes and no. Yes, if the random variables X and Y are independent (that is, one of them has no influence on the other, i.e., the correlation between X and Y is zero). No, if the random variables are not independent, but are dependent in some way. Furthermore, it doesn't matter if the random variables are added or subtracted, we are still combining the variances. That is,

- $\sigma^2_{X \pm Y} = \sigma^2_X + \sigma^2_Y$, if and only if X and Y are independent.
- $\sigma_{X \pm Y} = \sqrt{\sigma^2_X + \sigma^2_Y}$, if and only if X and Y are independent.

Digression: If X and Y are *not* independent, then $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$, where ρ is the correlation between X and Y. $\rho = 0$ if X and Y are independent. You do *not* need to know this for the AP exam.

TIP

> **Exam Tip:** The rules for means and variances when you combine random variables may seem a bit obscure, but there have been questions on more than one occasion that depend on your knowledge of how this is done.

The rules for means and variances generalize. That is, no matter how many random variables you have: $\mu_{X1 \pm X2 \pm \ldots \pm Xn} = \mu_{X1} \pm \mu_{X2 \pm \ldots +} \mu_{Xn}$ and, if $X_1, X_2, \ldots, X_n$ are all independent, $\sigma^2_{X_1 \pm X_2 \pm \ldots \pm Xn} = \sigma^2_{X_1} + \sigma^2_{X_2 + \ldots +} \sigma^2_{X_n}$.

example: A prestigious private school offers an admission test on the first Saturday of November and the first Saturday of December each year. In 2002, the mean score for hopeful students taking the test in November (X) was 156 with a standard deviation of 12. For those taking the test in December (Y), the mean score was 165 with a standard deviation of 11. What are the mean and standard deviation of the total score $X + Y$ of all students who took the test in 2002?

solution: We have no reason to think that scores of students who take the test in December are influenced by the scores of those students who took the test in November. Hence, it is reasonable to assume that X and Y are independent. Accordingly,

$$\mu_{X+Y} = \mu_X + \mu_Y = 156 + 165 = 321,$$

$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{12^2 + 11^2} = \sqrt{265} = 16.28.$$

> Rapid Review

1. A bag has eight green marbles and 12 red marbles. If you draw one marble from the bag, what is P(draw a green marble)?

   Answer: Let s = number of ways to draw a green marble.

   Let f = number of ways to draw a red marble.

   $$P(E) = \frac{s}{s+f} = \frac{8}{8+12} = \frac{8}{20} = \frac{2}{5}.$$

2. A married couple has three children. At least one of their children is a boy. What is the probability that the couple has exactly two boys?

   Answer: The sample space for having three children is {BBB, BBG, BGB, GBB, BGG, GBG, GGB, GGG}. Of these, there are seven outcomes that have at least one boy. Of these, three have two boys and one girl. Thus, P(the couple has exactly two boys they have at least one boy) = 3/7.

3. Does the following table represent the probability distribution for a discrete random variable?

X	1	2	3	4
$P(X)$	0.2	0.3	0.3	0.4

Answer: No, because

$$\sum P_i = 1.2.$$

4. In a standard normal distribution, what is $P(z > 0.5)$?

   Answer: From the table, we see that $P(z < 0.5) = 0.6915$. Hence, $P(z > 0.5) = 1 - 0.6915 = 0.3085$. By calculator, `normalcdf (0.5,100)` = 0.3085375322.

5. A random variable X has $N(13,0.45)$. Describe the distribution of $2 - 4X$ (that is, each datapoint in the distribution is multiplied by 4, and that value is subtracted from 2).

   Answer: We are given that the distribution of X is normal with $\mu_X = 13$ and $\sigma_X = 0.45$. Because $\mu_{a \pm bX} = a \pm b\mu_X$, $\mu_{2-4X} = 2 - 4\mu_X = 2 - 4(13) = -50$. Also, because $\sigma_{a \pm bX} = b\sigma_X$, $\sigma_{2-4X} = 4\sigma_X = 4(0.45) = 1.8$.

Practice Problems

Multiple Choice

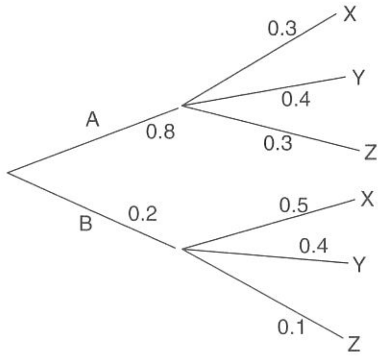
1.

	D	E	Total
A	15	12	27
B	15	23	38
C	32	28	60
Total	62	63	125

In the table above what are P(A and E) and $P(C \mid E)$?

(a) 12/125, 28/125
(b) 12/63, 28/60
(c) 12/125, 28/63
(d) 12/125, 28/60
(e) 12/63, 28/63

2.



For the tree diagram pictured above, what is $P(B \mid X)$?

(a) 1/4
(b) 5/17
(c) 2/5
(d) 1/3
(e) 4/5