Chapter 3 – The Normal Distributions

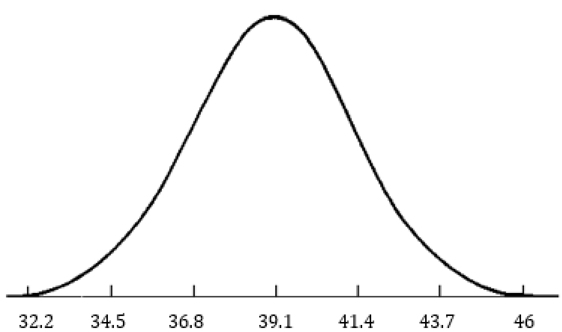
3.1 Sketches will vary. Use them to confirm that students understand the meaning of (a) symmetric and bimodal, and (b) skewed to the left and unimodal.

3.2 (a) It is on or above the horizontal axis everywhere, and because it forms a 1/8 × 8 rectangle, the area beneath the curve is 1. **(b)** Three-eighths of accidents occur in the first three miles: This is a 1/8 × 3 rectangle, so the proportion is 3/8, or 0.375. **(c)** The length of the path along the stream is (3 – 51.5) = 1.5 miles. So, this is a (1.5) × (1/8) rectangle; the proportion is 0.1875. **(d)** The part of the bike path more than a mile from either town is the 6-mile stretch from the 1-mile marker to the 7-mile marker. This is a (6) × (1/8) rectangle, so the proportion is 3/4, or 0.75.

3.3 (a) μ = 4, which is the obvious balance point of the rectangle. The median is also 4, because the distribution is symmetric (so that median equals mean) and half the area under the curve lies to the left and half to the right of 4. **(b)** The first quartile is 2, and the third quartile is 6.

3.4 (a) Mean is C; median is B (the right-skew pulls the mean to the right of the median). **(b)** Mean is B; median is B (this distribution is symmetric). **(c)** Mean is A; median is B (the left-skew pulls the mean to the left of the median).

3.5 Here is a sketch of the distribution of the Normal curve describing upper arm lengths of adult males. The tick marks are placed at the mean, and at one, two, and three standard deviations above and below the mean for scale.



3.6 Use the sketch from Exercise 3.5 and shade in the appropriate areas to answer these questions. **(a)** 99.7% of all upper arm lengths are within three standard deviations of the mean, or between 32.2 centimeters and 46 centimeters. **(b)** This is the area one or more standard deviations above the mean. Because the curve is symmetric and there is 100 – 68 = 32% of the area more than one standard deviation away from the mean, 32%/2 = 16% of upper arm lengths are greater than 41.4 centimeters.

3.7 (a) In the middle 95% of all years, monsoon rain levels are between 688 mm and 1016 mm—two standard deviations above and below the mean: 852 ± 2(82) = 688 mm to 1016 mm. **(b)** The driest 2.5% of monsoon rainfalls are less than 688 mm; this is more than two standard deviations below the mean.

3.8 Linda's standardized score is $z = \frac{680 - 511}{120} = 1.41$. Jack's standardized score is $z = \frac{26 - 20.8}{5.4} = 0.96$. Linda's score is relatively higher than Jack's.

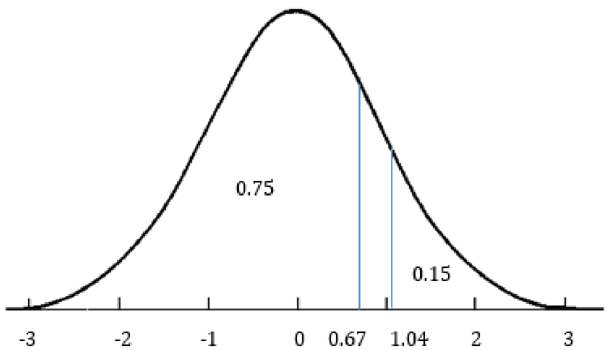
3.9 We need to use the same scale, so recall that 5.5 feet = 66 inches. A woman 5.5 feet tall has standardized score $z = \frac{66 - 64.2}{2.8} = 0.64$. A man 5.5 feet tall has standardized score $z = \frac{66 - 69.4}{3} = -1.13$. A woman 5.5 feet tall is 0.64 standard deviations taller than average for women. A man 5.5 feet tall is 1.13 standard deviations below average for men.

3.10 (a) 0.3372. **(b)** 0.9429. **(c)** 0.9830. **(d)** 0.9830 – 0.3372 = 0.6458.

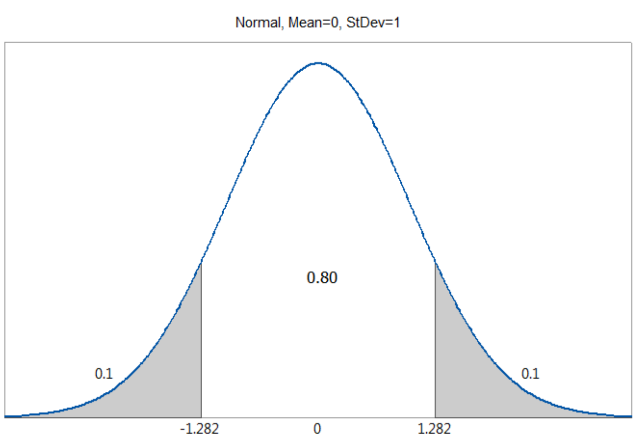
3.11 Let x be the monsoon rainfall in a given year. **(a)** $x \leq 697$ mm corresponds to $z \leq \frac{697-852}{82} = -1.89$, for which Table A gives 0.0294 = 2.94%. **(b)** $682 < x < 1022$ corresponds to $\frac{682-852}{82} < z < \frac{1022-852}{82}$, or $-2.07 < z < 2.07$. This proportion is 0.9808 – 0.0192 = 0.9616 = 96.16%.

3.12 (a) Let x be the MCAT score of a randomly selected student. Then $x > 510$ corresponds to $z > \frac{510 - 500}{10.6} = 0.94$, for which Table A gives 0.8264 as the area to the left. Hence, the answer is 1 – 0.8264 = 0.1736, or 17.36%. **(b)** $505 \leq x \leq 515$ corresponds to $\frac{505 - 500}{10.6} \leq z \leq \frac{515 - 500}{10.6}$, or $0.47 \leq z \leq 1.42$. Hence, using Table A, the area is 0.9222 – 0.6808 = 0.2414, or 24.14%.

3.13 (a) We want the value such that the proportion below (to the left) is 0.75. Using Table A and looking for an area as close as possible to 0.7500, we find this value has z = 0.67 (software would give the more precise z = 0.6744). **(b)** Now we want the value such that the proportion above is 0.15. This means that we want a proportion of 0.85 below. Using Table A and looking for an area as close to 0.8500 as possible, we find this value has z = 1.04 (software gives z = 1.036). **(c)** By symmetry, the z with 15% below it is the negative of the z with 15% above it. That is, $z = -1.04$.



3.14 (a) Because the Normal distribution is symmetric, its median and mean are the same. So the median MCAT score is 500.0. The first quartile has z = –0.67, because the area under the curve to the left of the first quartile is 0.2500 (software gives z = –0.6745). Similarly, the third quartile has z = 0.67, because the area under the curve to the left of the third quartile is 0.7500. The first quartile is 500 – 0.67(10.6) = 492.898, and the third quartile is 500 + 0.67(10.6) = 507.102. The IQR is $IQR = 507.102 - 492.898 = 14.204$. **(b)** If we are interested in the central 80%, there is 10% in each of the two tails. Software tells us the z-values corresponding to 10% in the tails are ±1.282 (from Table A, you'll find that z = ±1.28 has cumulative area 0.1003 in each tail). The MCAT scores are then $500 \pm 1.28(10.6)$, or 486.432 to 513.568.



3.15 (c) Lengths of 100 newborns in Connecticut. Economic variables such as income and prices of houses are usually right-skewed.

3.16 (a) the mean and the standard deviation. These tell you center and variability, which is all you need for a Normal distribution.

3.17 (b) 2. The curve is centered at 2.

3.18 (b) 3. Estimating a standard deviation is more difficult than estimating the mean, but among the three options, 2 is clearly too small and 5 is clearly too large, so 3 seems to be the most reasonable for the standard deviation.

3.19 (c) 4.5 and 8.5 hours per week night. 6.5 ± 2(1) = 4.5 to 8.5 hours.

3.20 (b) 6.7%. At least 7.5 hours would be 16% and at least 8.5 hours would be 2.5%, so at least 8 hours must be 6.7%.

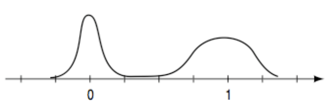
3.21 (b) 2.33. $z = \frac{135 - 100}{15} = 2.33$.

3.22 (c) 0.0375. 1 – 0.9625 = 0.0375.

3.23 (b) 0.135. Using the 68–95–99.7 rule, it is about 0.975 – 0.84 = 0.135.

3.24 (c) About 99%. In Exercise 3.21, we found Alysha's z-score to be 2.33. The proportion of adults who score below her is the proportion below 2.33. By Table A, that is 0.9901, or 99.01%.

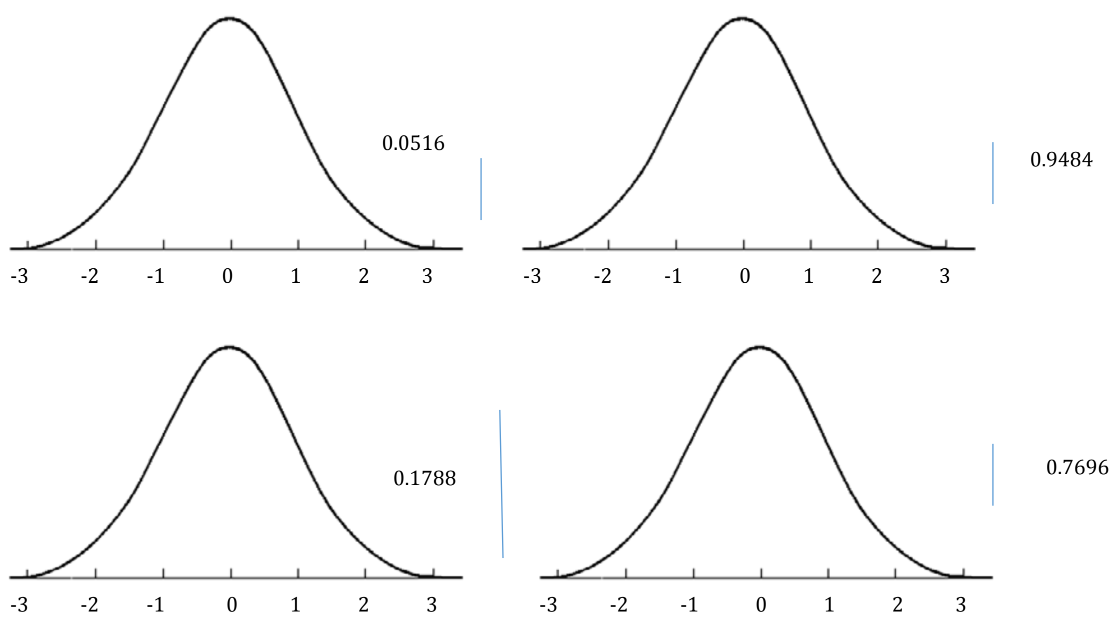
3.25 Sketches will vary, but should be some variation on the one shown here: the peak at 0 should be "tall and skinny," while near 1, the curve should be "short and fat."



3.26 For each distribution, take the mean plus or minus two standard deviations. For mildly obese people, this is 373 ± 2(67) = 239 minutes to 507 minutes. For lean people, this is 526 ± 2(107) = 312 minutes to 740 minutes.

3.27 By the 68–95–99.7 rule, 68% have LDL levels within one standard deviation of the mean, meaning (100 – 68)/2 = 16% have LDL levels one or more standard deviations above the mean.

3.28 (a) 0.0516. **(b)** 1 – 0.0516 = 0.9484. **(c)** 1 – 0.8212 = 0.1788. **(d)** 0.8212 – 0.0516 = 0.7696.



3.29 (a) We want the proportion less than *z* to be 0.20, so looking up a left-tail area of 0.2000 in the table, we find *z* = −0.84. (Software gives *z* = −0.8416.) **(b)** If 40% are more than *z*, then 60% are less than or equal to *z*. Hence, *z* = 0.25. (Software gives *z* = 0.2533.)

3.30 Let *x* be the length of a thorax for a randomly selected fruit fly. **(a)** *x* < 0.7 mm corresponds to $z < \frac{0.7 - 0.8}{0.078} = -1.28$. Hence, the area is 0.1003, or 10.03%. **(b)** *x* > 1 mm corresponds to $z > \frac{1 - 0.8}{0.078} = 2.56$. The area is 1 – 0.9948 = 0.0052, or 0.52%. **(c)** 0.7 mm < *x* < 1 mm corresponds to –1.28 < *z* < 2.56. Hence, the area is 0.9948 – 0.1003 = 0.8945, or 89.45%.

3.31 The proportion of rainy days with rainfall pH below 5.0 is about 0.2119: *x* < 5.0 corresponds to $z < \frac{5.0-5.43}{0.54}$ = −0.80, for which Table A gives 0.2119.

3.32 (a) $x > 140$ corresponds to $z > \frac{140 - 104}{12.5} = 2.88$. Table A gives 1 – 0.9980 = 0.0020, or 0.2%. **(b)** $x > 140$ for non-runners corresponds to $z > \frac{140 - 130}{17} = 0.59$. Table A gives 1 – 0.7224 = 0.2776, or 27.76% of non-runners have heart rates above 140 after the exercise.

3.33 (a) $x > 130$ corresponds to $z > \frac{130 - 100}{15} = 2$. Table A gives 1 – 0.9772 = 0.0228, or 2.28% had very superior scores in 1932. **(b)** $x > 130$ corresponds to $z > \frac{130 - 120}{15} = 0.67$, for which Table A gives 1 – 0.7486 = 0.2514, or 25.14%.

3.34 Let *x* be the BMI for a randomly selected young man aged 20 to 29. **(a)** Being underweight corresponds to $x < 18.5$. This gives $z < \frac{18.5-26.8}{5.2} = -1.60$. Hence, 0.0548, or 5.48% are underweight. **(b)** Being obese corresponds to $x > 30$. This gives $z > \frac{30-26.8}{5.2} = 0.62$. Hence, 1 – 0.7324 = 0.2676, or 26.76% are obese.

For Exercises 3.35 to 3.38, let x *denote the gas mileage of a randomly selected vehicle type from the population of 2016 model vehicles (excluding the high mileage outliers, as mentioned).*

3.35 Cars with better mileage than the Beetle correspond to $x > 28$, which corresponds to $z > \frac{28 - 23.0}{4.9} = 1.02$. Hence, this proportion is 1 – 0.8461 = 0.1539, or 15.39%.

3.36 We need the proportion above our vehicle's mileage to be 0.15; this means 85% (0.8500) have worse mileage. Table A finds $z = 1.04$ has 0.8508 to the left. So our vehicle would need mileage to be 23 + (1.04)(4.9) = 28.096 mpg. A car would need to have gas mileage of about 28.096 mpg or higher to be in the top 15% for all 2016 models.

3.37 The first and third quartiles have $z = -0.67$ and $z = 0.67$, respectively (use the symmetry of the Normal distribution to find one of these, for example, Q_1 with 0.2500 as the area to the left). Hence, the first quartile is 23 – (0.67)(4.9) = 19.72 mpg, and the third quartile is 23 + (0.67)(4.9) = 26.28 mpg.

3.38 The first quintile is the mileage so that 20% of models have a lower mileage. This has $z = -0.84$ (find the number closest to 0.2000 in Table A as a left-tail area). Similarly, the second, third, and fourth quintiles have $z = -0.25$, $z = 0.25$, and $z = 0.84$, respectively. The first quintile is then 23 – 0.84(4.9) = 18.884 mpg. Similarly, the second, third, and fourth quintiles are, respectively, 21.775 mpg, 24.225 mpg, and 27.116 mpg.

3.39 (a) Cecile's arm has $z = \frac{33.9 - 35.8}{2.1} = -0.90$. Using Table A, her percentile is 0.1841, or 18.41%. **(b)** Answers will vary due to variation in students' arm lengths.

3.40 A score of 1600 standardizes to $z = \frac{1600 - 1010}{218} = 2.71$. Therefore, the proportion above 1600 (which are reported as exactly 1600) is 1 – 0.9966 = 0.0034.

3.41 We want the proportion corresponding to $x > 64.2$. This corresponds to $z > \frac{64.2 - 69.4}{3} = -1.73$, which has proportion 1 – 0.0418 = 0.9582, or 95.82%.

3.42 The distribution of weights of women is right-skewed. First, the mean weight is larger than the median weight. Another clue comes from the greater distance between the median and the third quartile (181.2 – 149.4 = 31.8) than between the median and the first quartile (149.4 – 126.3 = 23.1).

3.43 (a) Let X be a randomly selected man's SAT math score. $X > 750$ corresponds to $z >$

$\frac{750 - 527}{124} = 1.80$. The proportion is 1 – 0.9641 = 0.0359, or 3.59%. **(b)** Let X be a randomly selected woman's SAT math score. $X > 750$ corresponds to $z > \frac{750 - 496}{115} = 2.21$. The proportion is 1 – 0.9864 = 0.0136, or 1.36%.

3.44 If the distribution is Normal, it must be symmetric about its mean—and in particular, the 10th and 90th percentiles must be equal distances below and above the mean—so the mean is 250 points. If 225 points below (above) the mean is the 10th (90th) percentile, this is 1.28 standard deviations below (above) the mean, so the distribution's standard deviation is 225/1.28 = 175.8 points.

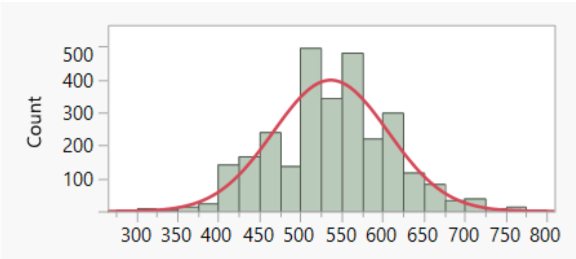
3.45 (a) About 0.6% of healthy young adults have osteoporosis (the cumulative probability below a standard score of −2.5 is 0.0062). **(b)** About 31% of this population of older women has osteoporosis. The BMD level that is 2.5 standard deviations below the young adult mean would standardize to −0.5 for these older women, and the cumulative probability for this standard score is 0.3085.

3.46 (a) 205584/1924436 = 0.1068, or 10.68% has scores greater than 28. **(b)** (205584 + 60551)/1924436 = 0.1383, or 13.83% had scores greater than or equal to 28. **(c)** $x > 28$ corresponds to $z > \frac{28 - 21}{5.5} = 1.27$, so the corresponding proportion is 1 – 0.8980 = 0.102, or 10.2%.

3.47 (a) If x is the return, then $x > 0$ corresponds to $z > \frac{0 - 0.84}{4.097} = -0.21$, so the corresponding proportion is 1 – 0.4168 = 0.5832, or 58.32% of returns are greater than 0%. $x > 4$ corresponds to $z > \frac{4 - 0.84}{4.097} = 0.77$, and the corresponding proportion is 1 – 0.7794 = 0.2206, or 22.06% have returns greater than 4%. **(b)** 62.6% and 21.95% of the actual returns are greater than 0% and 4%, respectively. These values are pretty close to the theoretical quantities assuming a Normal distribution, so it seems the $N(0.84, 4.097)$ distribution is a good approximation.

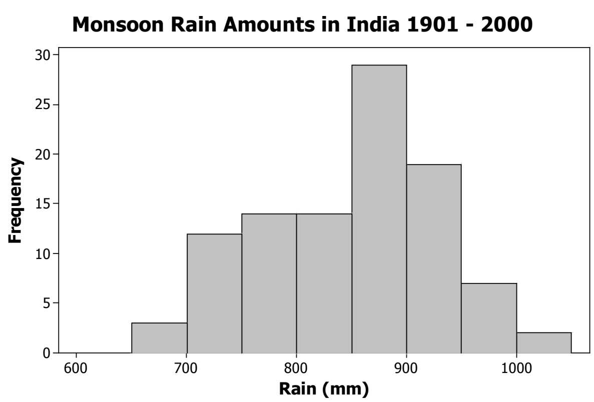
3.48 (a) The mean (5.43) is almost identical to the median (5.44), and the quartiles are similar distances from the median: $M - Q_1 = 0.39$ while $Q_3 - M = 0.35$. This suggests that the distribution is reasonably symmetric. **(b)** $x < 5.05$ corresponds to $z < \frac{5.05-5.43}{0.54} = -0.70$, and $x < 5.79$ corresponds to $z < \frac{5.79-5.43}{0.54} = 0.67$. Table A gives these proportions as 0.2420 and 0.7486. These are quite close to 0.25 and 0.75, which is what we would expect for the quartiles, so they are consistent with the idea that the distribution is close to Normal.

3.49 (a) A histogram is provided, and it appears to be roughly symmetric.



(b) Mean = 536.95, median = 540, standard deviation = 69.88, Q_1 = 490, and Q_3 = 580. The mean and median are close. The distance between Q_1 and the median is similar to the distance between the median and Q_3. This is consistent with a Normal distribution. **(c)** If x is the score of a randomly selected GSU entering student, then we are assuming x has the $N(536.95, 69.88)$ distribution. The proportion of GSU students scoring higher than the national average of 511 corresponds to the proportion of $z > \frac{511-536.95}{69.88} = -0.37$. Table A gives this proportion to be 1 – 0.3557 = 0.6443, or 64.43%. **(d)** 63.6% of GSU freshman scored higher than 511. This is very close to the probability computed in part (c).

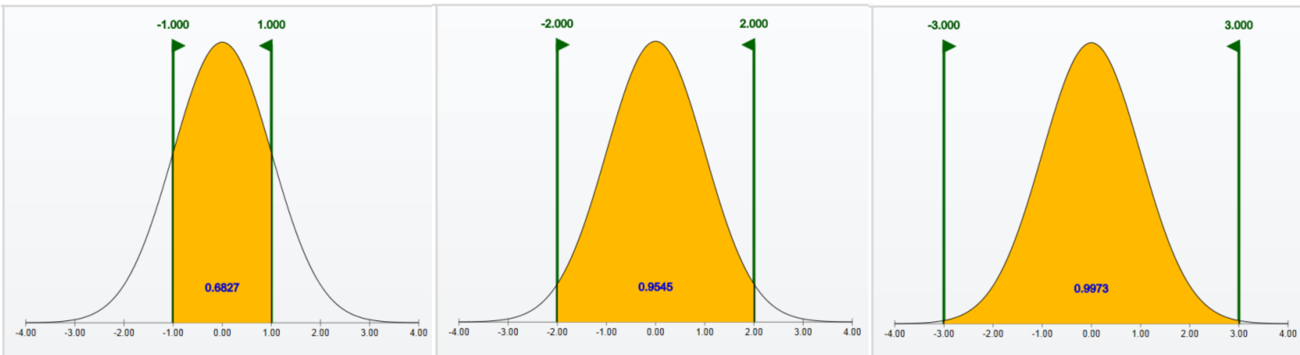
3.50 (a) One possible histogram is provided. The mean is 847.58 mm, and the median is 860.8 mm.



(b) The histogram shows a left-skew; this makes the mean lower than the median.

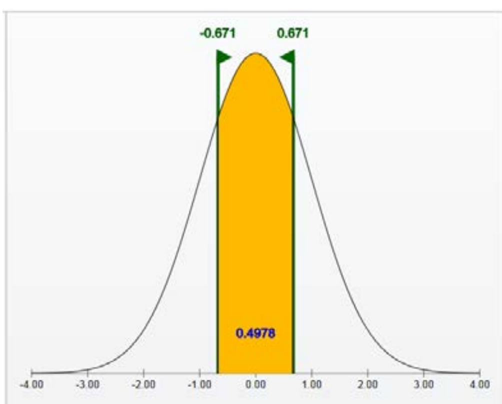
3.51 (a) 14/548 = 0.0255 (2.55%) weighed less than 100 pounds. $x < 100$ corresponds to $z < \frac{100-161.58}{48.96} = -1.26$. Using Table A, the area to the left of −1.26 is 0.1038 (10.38%). **(b)** 33/548 = 0.0602 (6.02%) weighed more than 250 pounds. $x > 250$ corresponds to $z > \frac{250-161.58}{48.96} = 1.81$. Using Table A, about 1 – 0.9649 = 0.0351 (3.51%) would be expected to weigh more than 250 pounds. **(c)** The Normal distribution model predicts 10.38% of women will weigh less than 100 pounds, while actually about 2.55% do. This is a substantial error since the Normal model also predicts 3.51% of values more than 250, where we actually observed 6.02% more than 250. In this application, the data seem to be far from Normal in distribution.

3.52 (a) The applet shows an area of 0.6827 between −1.000 and 1.000, while the 68–95–99.7 rule rounds this to 0.68. **(b)** Between −2.000 and 2.000, the applet reports 0.9545 (compared with the rounded 0.95 from the 68–95–99.7 rule). Between −3.000 and 3.000, the applet reports 0.9973 (compared with the rounded 0.997).

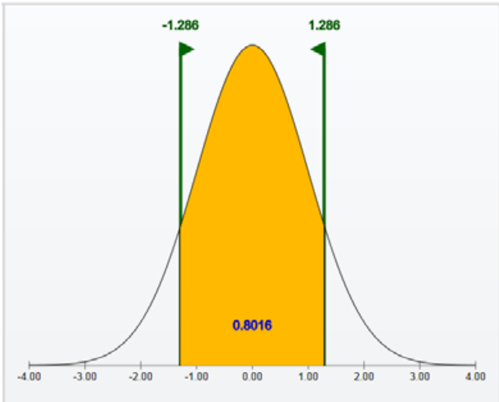


3.53 Because the quartiles of any distribution have 50% of observations between them, we seek to place the flags so that the reported area is 0.5. The closest the applet gets is an area of 0.4978, between −0.671 and 0.671. Thus the quartiles of any Normal distribution are about 0.67 standard deviations above and below the mean.

Note: *Table A places the quartiles at about 0.67; other statistical software gives ±0.6745.*



3.54 Placing the flags so that the area between them is as close as possible to 0.80, we find that the A/B cutoff is about 1.28 standard deviations above the mean and that the B/C cutoff is about 1.28 standard deviations below the mean.



3.55 and **3.56** are Web-based exercises.