The Z score

The Z score of an observation is the number of standard deviations it falls above or below the mean. We compute the Z score for an observation x that follows a distribution with mean μ and standard deviation σ using

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

⊙ **Exercise 3.3** Use Tom's ACT score, 24, along with the ACT mean and standard deviation to compute his Z score.[3]

Observations above the mean always have positive Z scores while those below the mean have negative Z scores. If an observation is equal to the mean (e.g. SAT score of 1500), then the Z score is 0.

⊙ **Exercise 3.4** Let X represent a random variable from $N(\mu = 3, \sigma = 2)$, and suppose we observe $x = 5.19$. (a) Find the Z score of x. (b) Use the Z score to determine how many standard deviations above or below the mean x falls.[4]

⊙ **Exercise 3.5** Head lengths of brushtail possums follow a nearly normal distribution with mean 92.6 mm and standard deviation 3.6 mm. Compute the Z scores for possums with head lengths of 95.4 mm and 85.8 mm.[5]

We can use Z scores to roughly identify which observations are more unusual than others. One observation x_1 is said to be more unusual than another observation x_2 if the absolute value of its Z score is larger than the absolute value of the other observation's Z score: $|Z_1| > |Z_2|$. This technique is especially insightful when a distribution is symmetric.

⊙ **Exercise 3.6** Which of the observations in Exercise 3.5 is more unusual?[6]

3.1.3 Normal probability table

● **Example 3.7** Ann from Example 3.2 earned a score of 1800 on her SAT with a corresponding $Z = 1$. She would like to know what percentile she falls in among all SAT test-takers.

Ann's **percentile** is the percentage of people who earned a lower SAT score than Ann. We shade the area representing those individuals in Figure 3.6. The total area under the normal curve is always equal to 1, and the proportion of people who scored below Ann on the SAT is equal to the *area* shaded in Figure 3.6: 0.8413. In other words, Ann is in the 84^{th} percentile of SAT takers.

We can use the normal model to find percentiles. A **normal probability table**, which lists Z scores and corresponding percentiles, can be used to identify a percentile based on the Z score (and vice versa). Statistical software can also be used.

[3] $Z_{Tom} = \frac{x_{Tom} - \mu_{ACT}}{\sigma_{ACT}} = \frac{24-21}{5} = 0.6$

[4] (a) Its Z score is given by $Z = \frac{x-\mu}{\sigma} = \frac{5.19-3}{2} = 2.19/2 = 1.095$. (b) The observation x is 1.095 standard deviations *above* the mean. We know it must be above the mean since Z is positive.

[5] For $x_1 = 95.4$ mm: $Z_1 = \frac{x_1-\mu}{\sigma} = \frac{95.4-92.6}{3.6} = 0.78$. For $x_2 = 85.8$ mm: $Z_2 = \frac{85.8-92.6}{3.6} = -1.89$.

[6] Because the *absolute value* of Z score for the second observation is larger than that of the first, the second observation has a more unusual head length.

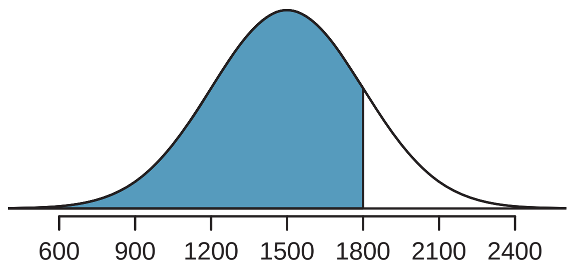

Figure 3.6: The normal model for SAT scores, shading the area of those individuals who scored below Ann.

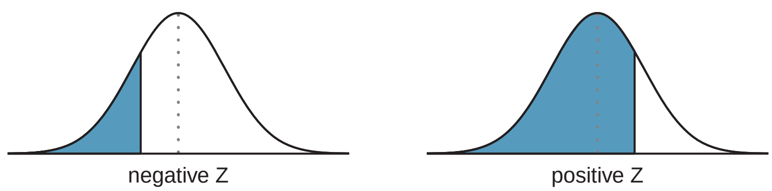

Figure 3.7: The area to the left of Z represents the percentile of the observation.

A normal probability table is given in Appendix B.1 on page 407 and abbreviated in Table 3.8. We use this table to identify the percentile corresponding to any particular Z score. For instance, the percentile of $Z = 0.43$ is shown in row 0.4 and column 0.03 in Table 3.8: 0.6664, or the 66.64^{th} percentile. Generally, we round Z to two decimals, identify the proper row in the normal probability table up through the first decimal, and then determine the column representing the second decimal value. The intersection of this row and column is the percentile of the observation.

We can also find the Z score associated with a percentile. For example, to identify Z for the 80^{th} percentile, we look for the value closest to 0.8000 in the middle portion of the table: 0.7995. We determine the Z score for the 80^{th} percentile by combining the row and column Z values: 0.84.

Exercise 3.8 Determine the proportion of SAT test takers who scored better than Ann on the SAT.[7]

3.1.4 Normal probability examples

Cumulative SAT scores are approximated well by a normal model, $N(\mu = 1500, \sigma = 300)$.

Example 3.9 Shannon is a randomly selected SAT taker, and nothing is known about Shannon's SAT aptitude. What is the probability Shannon scores at least 1630 on her SATs?

First, always draw and label a picture of the normal distribution. (Drawings need not be exact to be useful.) We are interested in the chance she scores above 1630, so we shade this upper tail:

[7] If 84% had lower scores than Ann, the number of people who had better scores must be 16%. (Generally ties are ignored when the normal model, or any other continuous distribution, is used.)