# z-scores and the empirical rule

*How far from typical is a value? A z-score counts the standard deviations, so values in different units can be compared, and for a normal model it tells you roughly what share of values lie beyond.*

## WORDS TO KNOW

**z-score:** the number of standard deviations a value lies above or below the mean.

**Empirical rule:** the 68-95-99.7 rule for normal models.

**Standard units:** another name for z-scores. They have no units.

**$\bar{x}$ and $s$:** the mean and standard deviation of a sample. They play the parts of $\mu$ and $\sigma$ for sample data.

## KEY IDEAS

### 1. A z-score counts standard deviations

The top of the fraction is the distance from the mean. The bottom is the size of one standard deviation. A value above the mean has a positive z-score, one below has a negative z-score, and the mean itself has $z = 0$.

$$z = \frac{x - \mu}{\sigma} \qquad \text{and back again} \qquad x = \mu + z\sigma$$

For a sample, use $z = \dfrac{x - \bar{x}}{s}$.

### 2. z-scores compare values in different units

The units cancel, so a z-score is a pure number. Two values from different distributions can be compared by their z-scores: the larger $|z|$ is further from typical in its own group. Changing units by $a + bx$ with $b > 0$ (such as minutes to seconds) leaves every z-score the same.

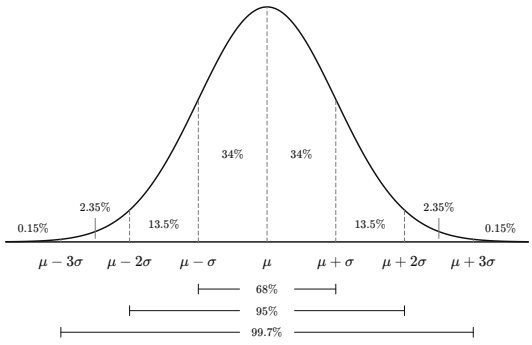
### 3. The empirical rule for a normal model

For a normal model, about 68% of values lie within one standard deviation of the mean, about 95% within two, and about 99.7% within three. Symmetry splits each band evenly between the two sides.

| Band of $z$ | 0 to 1 | 1 to 2 | 2 to 3 | beyond 3 |
|---|---|---|---|---|
| Share on each side | 34% | 13.5% | 2.35% | 0.15% |

Add neighboring bands to get the share between any two whole-number z-scores.

About 16% of values are more than one standard deviation above the mean, and about 2.5% more than two above.






Free sample from math.wiki for classroom use. Do not post. math.wiki

## WATCH OUT

- **Divide by the standard deviation, not the variance.** And a z-score is not the distance itself: 9 cm above the mean with $\sigma = 6$ cm is $z = 1.5$, not 9.
- **The empirical rule is for normal models only.** A z-score works for any distribution, but the 68-95-99.7 shares can be badly wrong for a skewed or two-peaked shape.
- **Whole numbers of standard deviations only.** The rule cannot give the share below $z = 1.3$. That needs a table or technology.
- **When lower is better, look at the sign.** For race times, a negative z-score is the good side.

## WORKED EXAMPLES

**Example 1. Bags of rice have masses modeled by $N(2.00, 0.04)$ in kilograms. Find the z-score of a 2.11 kg bag. Which mass has z-score −1.5?**

1. The bag is 2.11 − 2.00 = 0.11 kg above the mean, so $z = \dfrac{0.11}{0.04} = 2.75$.
2. Going back, $x = \mu + z\sigma = 2.00 + (-1.5)(0.04) = 2.00 - 0.06 = 1.94$.

**Answer:** $z = 2.75$, and a 1.94 kg bag has $z = -1.5$. **Check:** $(1.94 - 2.00) \div 0.04 = -0.06 \div 0.04 = -1.5$.

**Example 2. Rosa scored 82 on a math test (mean 70, standard deviation 10) and 91 on a reading test (mean 79, standard deviation 6). On which test did she do better compared with the class?**

1. Math: $z = \dfrac{82 - 70}{10} = \dfrac{12}{10} = 1.2$.
2. Reading: $z = \dfrac{91 - 79}{6} = \dfrac{12}{6} = 2$.
3. She beat both means by 12 points, but on the reading test 12 points is more standard deviations.

**Answer:** reading. She is 2 standard deviations above the mean there, and only 1.2 in math.

**Example 3. Boxes of cereal have masses modeled by $N(250, 12)$ in grams. Use the empirical rule to estimate the share heavier than 274 g and the share between 238 and 274 g. In a batch of 400 boxes, about how many are lighter than 238 g?**

1. Turn the boundaries into z-scores: 274 is $z = \dfrac{24}{12} = 2$, and 238 is $z = \dfrac{-12}{12} = -1$.
2. Above $z = 2$: the 5% outside $\pm 2$ splits into two tails, so about 2.5%.
3. From $z = -1$ to $z = 2$: 34% + 34% + 13.5% = 81.5%.
4. Below $z = -1$: the 32% outside $\pm 1$ splits into two tails of 16%. And 16% of 400 is 64.

**Answer:** about 2.5%, about 81.5%, and about 64 boxes. **Check:** the three regions cover everything: $2.5\% + 81.5\% + 16\% = 100\%$.


Free sample from math.wiki for classroom use. Do not post. math.wiki