Properties of Continuous Probability Distributions

The graph of a continuous probability distribution is a curve. Probability is represented by area under the curve.

The curve is called the **probability density function** (abbreviated as **pdf**). We use the symbol $f(x)$ to represent the curve. $f(x)$ is the function that corresponds to the graph; we use the density function $f(x)$ to draw the graph of the probability distribution.

Area under the curve is given by a different function called the **cumulative distribution function** (abbreviated as **cdf**). The cumulative distribution function is used to evaluate probability as area.

- The outcomes are measured, not counted.
- The entire area under the curve and above the x-axis is equal to one.
- Probability is found for intervals of x values rather than for individual x values.
- $P(c < x < d)$ is the probability that the random variable X is in the interval between the values c and d. $P(c < x < d)$ is the area under the curve, above the x-axis, to the right of c and the left of d.
- $P(x = c) = 0$ The probability that x takes on any single individual value is zero. The area below the curve, above the x-axis, and between $x = c$ and $x = c$ has no width, and therefore no area (area = 0). Since the probability is equal to the area, the probability is also zero.
- $P(c < x < d)$ is the same as $P(c \leq x \leq d)$ because probability is equal to area.

A probability distribution that plots all of its values in a symmetrical fashion and most of the results are situated around the probability's mean is called a **normal distribution**. Values are equally likely to plot either above or below the mean. Grouping takes place at values that are close to the mean and then tails off symmetrically away from the mean.

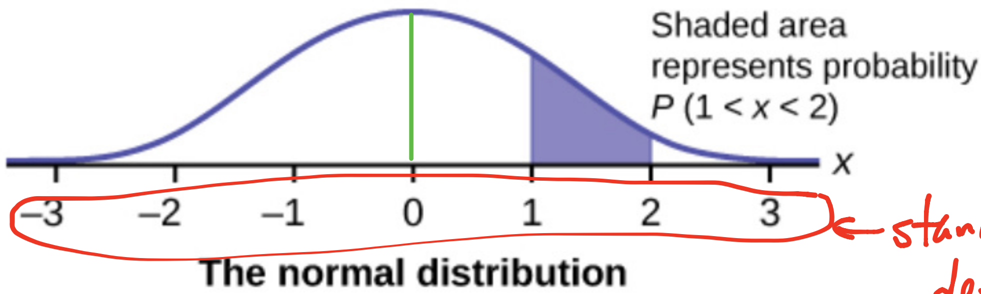
Some Properties of a Normal Distribution

1. The value in the middle of the distribution, which appears most often in the sample, is the mean.
2. The distribution is symmetric about the mean. This means that the graph has two halves that are mirror images on either side of the mean value.
3. This is the key fact: the area under any portion of the curve is the percentage (in decimal form) of data values that fall between the values that begin and end that region.
4. The total area under the entire curve is 1.

2. the mean is "axis of reflection" of normal distribution.

3. area = probability

$\bar{x}$ ← mean



← standard deviations

Figure 5.4 The graph shows the Standard Normal Distribution with the area between $x = 1$ and $x = 2$ shaded to represent the probability that the value of the random variable X is in the interval between one and two.

"bell curve" = normal distribution

0 standard deviations → mean

- outliers should exist on the edges.
- large set of data considered "average" near the center.

Z-Scores

If X is a normally distributed random variable and $X \sim N(\mu, \sigma)$, then the *z*-score is:

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

The *z*-score tells you how many standard deviations the value *x* is above (to the right of) or below (to the left of) the mean, μ. Values of x that are larger than the mean have positive *z*-scores, and values of x that are smaller than the mean have negative *z*-scores. If x equals the mean, then x has a *z*-score of zero.

Example 6.1

Suppose $X \sim N(5, 6)$. This says that X is a normally distributed random variable with mean $\mu = 5$ and standard deviation $\sigma = 6$. Suppose $x = 17$. Then:

$$z = \frac{x - \mu}{\sigma} \quad \text{or} \quad \frac{x - \bar{x}}{s}$$

x = any data value

μ = population mean

$\bar{x}$ = sample mean

σ = population SD

s = sample SD.

z = # of deviations away from mean in normal distribution.

$$z = \frac{x - \mu}{\sigma} = \frac{(17) - 5}{6} = \frac{12}{6} = +2$$

above

$\to x = 17$ is 2 SD to the right of the mean: 5

$x = 17$, $\mu = 5$, $\sigma = 6$ $\Rightarrow z = +2$

6.1 What is the *z*-score of *x*, when *x* = 1 and *X* ~ *N*(12,3)?

$\sigma = 3$

$\mu = 12$

$$z\text{-score}: \frac{x-\mu}{\sigma} = \frac{(1)-(12)}{3} = -\frac{11}{3} = -3.\overline{6}$$

$$\approx -3.67$$

$$\left.\begin{array}{l} x = 1 \\ \mu = 12 \\ \sigma = 3 \end{array}\right] \Rightarrow z = -3.67$$

→ 1 is 3.67 SD left of / below the mean 12.

Example 6.2

Some doctors believe that a person can lose five pounds, on the average, in a month by reducing his or her fat intake and by exercising consistently. Suppose weight loss has a normal distribution. Let *X* = the amount of weight lost (in pounds) by a person in a month. Use a standard deviation of two pounds. *X* ~ *N*(5, 2). Fill in the blanks.

μ σ

a. Suppose a person **lost** ten pounds in a month. The *z*-score when *x* = 10 pounds is *z* = 2.5 (verify). This *z*-score tells you that *x* = 10 is __2.5__ standard deviations to the __right__ (right or left) of the mean __5__ (What is the mean?).

$$z = \frac{x-\mu}{\sigma} = \frac{(10)-(5)}{(2)} = +2.5$$

Example 4 Using z Scores to Compare Standardized Test Scores

As you probably know, there are two main companies that offer standardized college entrance exams, ACT and SAT. Since each has a completely different scoring scale, it's really difficult to compare the scores of students that took different exams. One year the ACT had a mean score of 21.2 and a standard deviation of 5.1. That same year, the SAT had a mean score of 1498 and a standard deviation of 347. Suppose that a scholarship committee is considering two students, one who scored 26 on the ACT and another who scored 1800 on the SAT. Both are pretty good scores, but which one is better?

ACT: $\mu = 21.2$, $\sigma = 5.1$, $x = 26$

SAT: $\mu = 1498$, $\sigma = 347$, $x = 1800$

$$z_{ACT} = \frac{26 - 21.2}{5.1}$$

$$z_{SAT} = \frac{1800 - 1498}{347}$$

$$z_{ACT} \approx +0.94$$

$$z_{SAT} \approx +0.57$$

→ ACT student scored higher

b/c $z_{ACT} > z_{SAT}$

$+0.94 > +0.57$

The **standard normal distribution** is a normal distribution with mean 0 and standard deviation 1.

The standard normal distribution is shown in Figure 11-11. The values under the curve shown in Figure 11-11 indicate the proportion of area in each section. For example, the area between the mean and 1 standard deviation above or below the mean is about 0.341, or 34.1%.

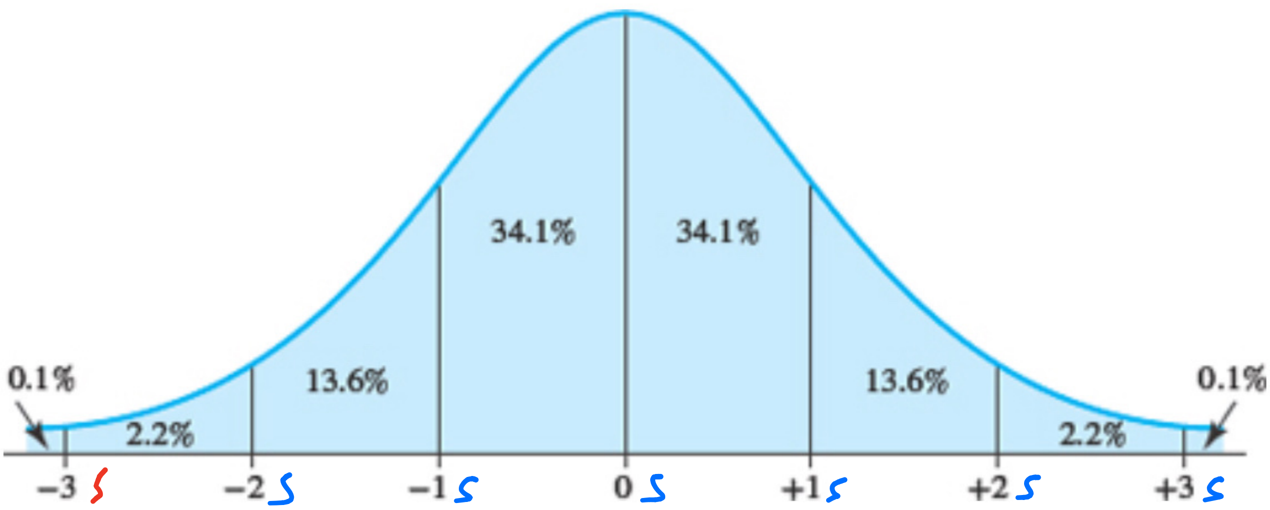


Figure 11-11

σ

Example 6.1

Suppose $X \sim N(5, 6)$. This says that X is a normally distributed random variable with mean $\mu = 5$ and standard deviation $\sigma = 6$. Suppose $x = 17$. Then:

$$z = \frac{x - \mu}{\sigma} = \frac{(17) - 5}{6} = \frac{12}{6} = +2$$

$\rightarrow x = 17$ is 2 SD to the right (above) of the mean: 5

What percentage of data values is below 17?

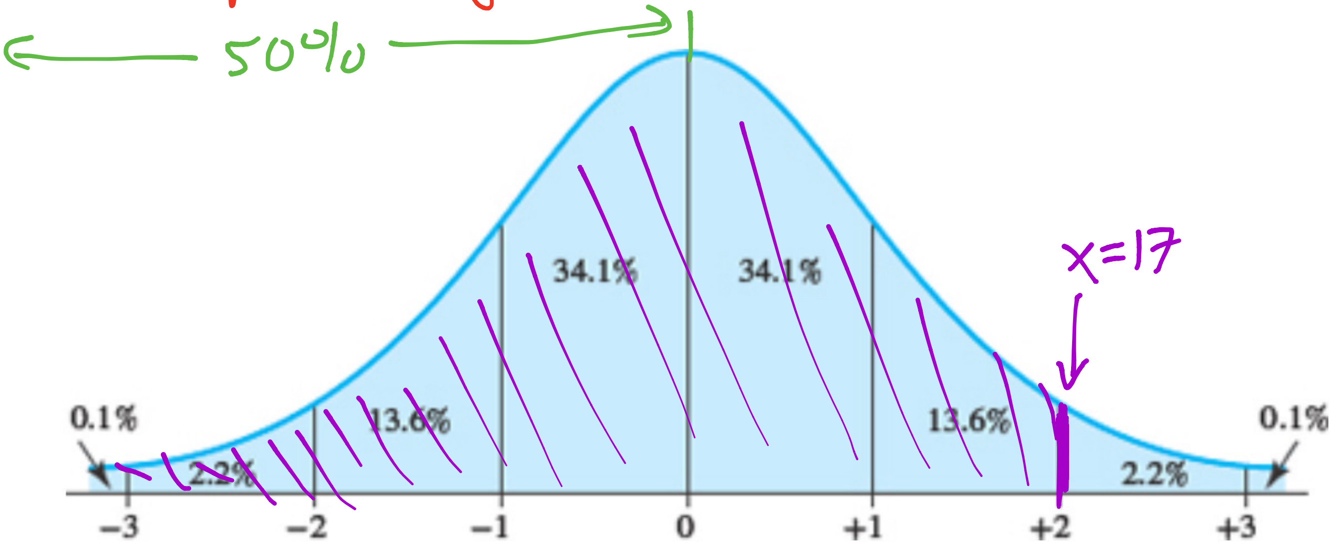


Figure 11-11

Recall that 17 is +2SD

→ add all areas to the left of +2SD.

$50\% + 34{,}1\% + 13{,}6\% = 97.7\%$

↑ one half of the normal distribution.

→ recall: area under normal distribution = 1

1 − right side area

$100\% - (2{,}2\% + 0{,}1\%) = 97{,}7\%$

$100\% - 2{,}3\% = 97.7\%$

What percentage of data values is above 17?

$2{,}2\% + 0{,}1\% = 2{,}3\%$