4: Probability

What is probability? The probability of an event is its relative frequency (proportion) in the population. An event that happens half the time (such as a head showing up on the flip of a fair coin) has probability 50%. A horse that wins 1 in 4 races has a 25% probability of winning. A treatment that works in 4 of 5 patients has an 80% probability of success.

Probability can occasionally be derived **logically** by counting the number of ways a thing can happen and determine its relative frequency. To use a familiar example, there are 52 cards in a deck, 4 of which are Kings. Therefore, the probability of randomly drawing a King = 4 ÷ 52 = .0769.

Probability can be *estimated* through **experience**. If an event occurs x times out of n, then its probability will *converge* on $X \div n$ as n becomes large. For example, if we flip a coin many times, we expect to see half the flips turn up heads. This experience is unreliable when n is small, but becomes increasingly reliable as n increases. For example, if a coin is flipped 10 times, there is no guarantee that we will observe exactly 5 heads. However, if the coin is flipped 1000 times, chances are better that the proportions of heads will be close to 0.50.

Probability can be used to quantify **subjective** opinion. If a doctor says "you have a 50% chance of recovery," the doctor *believes* that half of similar cases will recover in the long run. Presumably, this is based on knowledge, and not on a whim. The benefit of stating subjective probabilities is that they can be tested and modified according to experience.

Notes:

- **Range of possible probabilities:** Probabilities can be no less than 0% and no more than 100% (of course).

- N**otation:** Let A represent an event. Then, Pr(A) represents the probability of the event.

- **Complement:** Let $\bar{A}$ represent the complement of event A. The complement of an event is its "opposite," i.e., the event *not* happening. For example, if event A is recovery following treatment, then $\bar{A}$represents failure to recover.

- **Law of complements:** $\Pr(\bar{A}) = 1 \quad \Pr(A)$. For example, if $\Pr(A) = 0.75$, then $\Pr(\bar{A}) = 1 - 0.75 = 0.25$.

- **Random variable:** A random variable is a quantity that varies depending on chance. There are two types of random variables,

    - **Discrete random variables** can take on a finite number of possible outcomes. We study binomial random variables as examples of discrete random variables.

    - **Continuous random variables** form an unbroken chain of possible outcomes, and can take on an infinite number of possibilities. We study Normal random variables as examples of continuous random variables.

Binomial Distributions

Binomial Random Variables

Consider a random event that can take on only one of two possible outcomes. Each event is a **Bernoulli trial**. Arbitrarily, define one outcome a "success" and the other a "failure." Now, take a series of *n* independent Bernoulli trials. The random number of successes in n Bernoulli trials is a **binomial random variable.**

Illustrative Example. Suppose a treatment is successful 75% of the time. The treatment is used in 4 patients. *On average,* 3 of the 4 patients will respond to treatment. However, it would be foolish to think 3 of 4 patients will always respond. The number of patients responding will vary from trial to trial according to a **binomial distribution** (**binomial probability mass function)**. For the current example, the binomial probability mass function is:

Number of successes	Probability
0	0.0039
1	0.0469
2	0.2109
3	0.4219
4	0.3164

Binomial distributions are characterized by two parameters. These are:

$n \equiv$ the number of independent trials
$p \equiv$ the probability of success for each trial

We use the notation X~b(n, p) to denote a given binomial distribution. In words, this is "the random variable X is distributed as a binomial random variable with parameters n and p." The distribution in the table above is X~ b(4, 0.75).

Before calculating binomial probabilities we must first learn the **combinatorics ("choose") function**. The combinatorics function answers the question "How many different ways can I choose i items out of n."

Let ${}_nC_i$ denote the number of ways to choose i items out of n:

$${}_nC_i = \frac{n!}{i!(n-i)!} \qquad \textbf{(4.1)}$$

where **!** represents the **factorial function**, which is the product of the series of integers from n to 1. In symbols, $n! = (n)(n \quad 1)\ (n \quad 2)(n \quad 3) \ldots (1)$. For example, $3! = (3)(2)(1) = 6$. By definition, $1! = 1$ and $0! = 1$.

Illustrative examples (Combinatorics). Three examples are presented:

- "How many ways are there to choose 2 items out of 3?" By formula, ${}_3C_2 = \frac{3!}{(2!)(3-2)!} = \frac{3 \cdot 2 \cdot 1}{(2 \cdot 1)(1)} = \frac{6}{2} = 3$. By logic, consider 3 items labeled A, B, and C. There are three different sets of two: {A, B}, {A, C}, or {B, C}.
- "How many ways are there to choose 2 items out of 4?" ${}_4C_2 = \frac{4!}{(2!)(4-2)!} = \frac{4 \cdot 3 \cdot 2!}{(2 \cdot 1)(2!)} = \frac{4 \cdot 3}{2} = 6$; there are 6 ways to choose 2 items out of 4: {A, B}, {A, C}, {A, D}, {B, C}, {B, D}, and {C, D}.
- How many ways are there to choose 3 items out of 7?" ${}_7C_3 = \frac{7!}{(3!)(7-3)!} = \frac{7 \cdot 6 \cdot 5 \cdot 4!}{(3 \cdot 2 \cdot 1)(4!)} = 35$. There are 35 ways to choose 3 items out of 7.

Binomial Formula

We are now ready to calculate binomial probabilities with this formula:

$$\Pr(X = i) = {}_nC_i\, p^i q^{n-i} \tag{4.2}$$

where *X* represents the random number of successes,
i is the observed number of successes,
n and *p* are the binomial parameters, and
$q = 1 \quad p$.

Illustrative Example. Recall the example that considers a treatment that is successful 75% of the time ($p = 0.75$). The treatment is used in 4 patients ($n = 4$). "What is the probability of seeing 2 successes in 4 patients?"

Given: $n = 4$ $i = 2$, $p = 0.75$ $q = 1 - 0.75 = 0.25$.

Calculation: ` $\Pr(X = 2) = {}_4C_2\ (.75)^2\ (.25)^{4-2} = (6)(0.5625)(0.0625) = 0.2109$.

Probability Mass Function

The listing of probabilities for all possible outcomes for a discrete random variable is the binomial **probability mass function.**

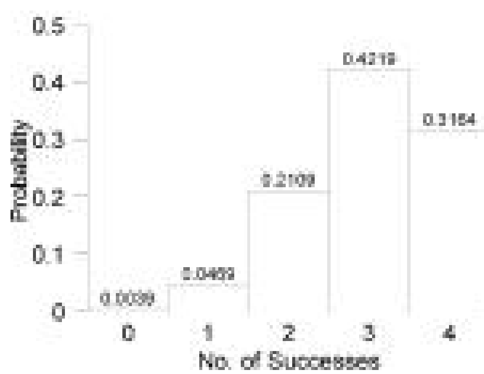
Illustrative Example. For a treatment that is successful 75% of the time used in 4 patients, the binomial probability mass function is:

The probability of 0 successes ≡ $Pr(X = 0) = {}_4C_0\ (0.75)^0\ (0.25^{4-0}) = (1)(1)(0.0039) = 0.0039$
The probability of 1 success ≡ $Pr(X = 1) = {}_4C_1\ (0.75)^1\ (0.25^{4-1}) = (4)(0.75)(0.0156) = 0.0469$
The probability of 2 successes ≡ $Pr(X = 2) = {}_4C_2\ (0.75)^2\ (0.25^{4-2}) = (6)(0.5625)(0.0625) = 0.2109$
The probability of 3 successes ≡ $Pr(X = 3) = {}_4C_3\ (0.75)^3\ (0.25^{4-3}) = (4)(0.4219)(0.25) = 0.4219$
The probability of 4 successes ≡ $Pr(X = 4) = {}_4C_4\ (0.75)^4\ (0.25^{4-4}) = (1)(0.3164)(1)= 0.3164$

In tabular form:

(No. of successes) *i*	Pr(*X* = *i*)
0	0.0039
1	0.0469
2	0.2109
3	0.4219
4	0.3164

In graphical form:



The **areas under the bars** in the histogram represent probabilities. For example, the bar corresponding to 2 out of 4 successes has a width of 1.0 and height of 0.2109. The area of this bar = height × width = 1× 0.2109 = 0.2109, which is equal to the probability of observing 2 successes.

Cumulative Probability

The cumulative probability is the probability of observing *less than or equal to* a given number of successes. For example, the cumulative probability of 2 successes is the probability of 2 or less successes, denoted Pr(X ≤ 2).

Illustrative example. For the illustrative example, Pr(X ≤ 2) = Pr(X = 0) + Pr(X = 1) + Pr(X = 2) = 0.0039 + 0.0469 + 0.2109 = 0.2617.

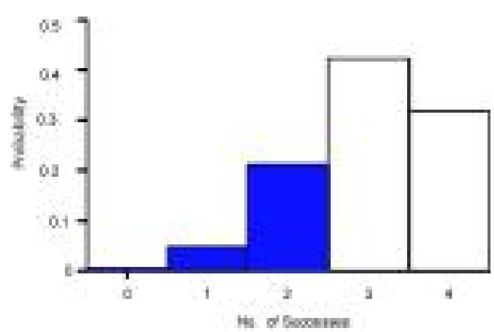
The **cumulative probability distribution** is the compilation of cumulative probabilities for all possible outcomes. The cumulative probability distribution for the illustrative example is:

Pr(X ≤ 0) = .0039
Pr(X ≤ 1) = [Pr(X = 0) + Pr(X = 1)] = 0.0039 + 0.0469 = 0.0508
Pr(X ≤ 2) = [Pr(X ≤ 1) + Pr(X = 2)] = 0.0508 + 0.2109 = 0.2617
Pr(X ≤ 3) = [Pr(X ≤ 2) + Pr(X = 3)] = 0.2617 + 0.4219 = 0.6836
Pr(X ≤ 4) = [Pr(X ≤ 3) + Pr(X = 4)] = 0.6836 + 0.3164 = 1.0000

In tabular form:

No. of successes i	Cumulative Probability Pr($X \le i$)
0	0.0039
1	0.0508
2	0.2617
3	0.6836
4	1.0000

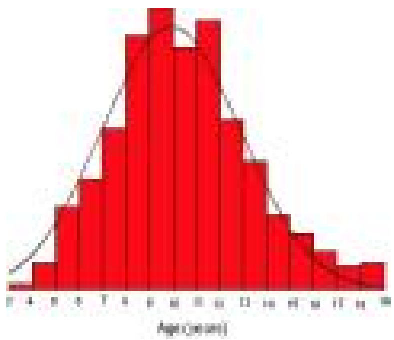
Cumulative probabilities corresponds to **areas under the bars to the LEFT** of points. The shaded region in the figure below corresponds to Pr(X ≤2) for the illustrative example.



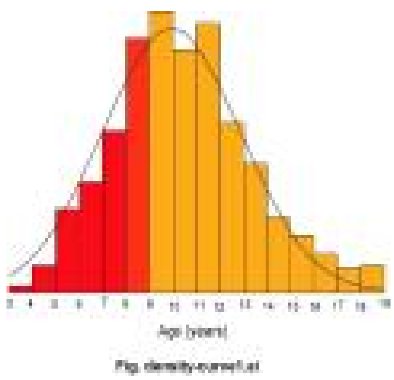
Normal Probability Distributions

The previous section used the binomial formula to calculate probabilities for binomial random variables. Outcomes were discrete, and probabilities were displayed with probability histograms. We need a different approach for modeling continuous random variables. This approach involves the use of **density curves**.

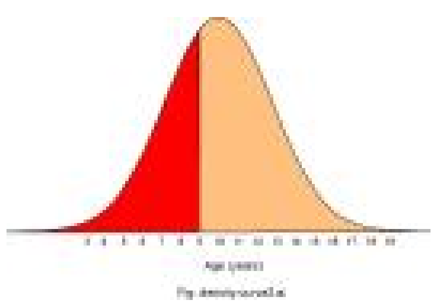
Density curves are smoothed probability histograms. A Normal density curve is superimposed on this age distribution:



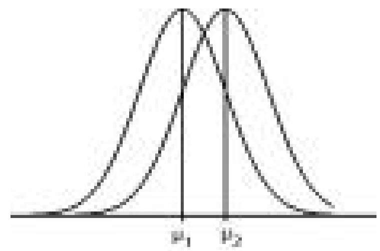
The next curve shows the same distribution with the six left-most bars shaded. This corresponds to individuals less than or equal to 9-years of age. There were 215 such individuals, making up about one-third of the entries. Therefore, Pr($X \le 8$) ≈ ⅓.



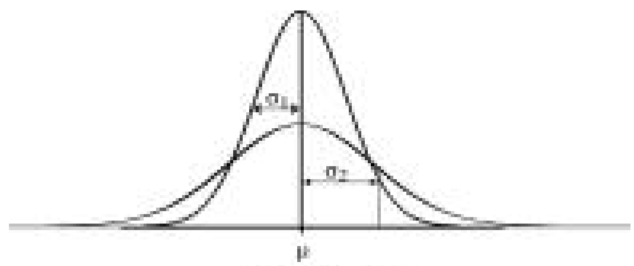
When working with Normal probabilities, we drop the histogram and look only at the curve.



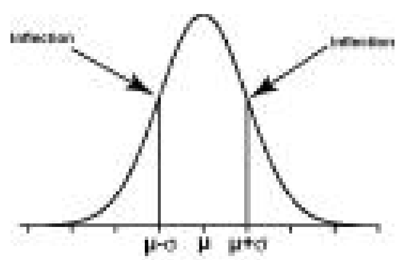
Normal distributions are characterized by two **parameters**: μ and σ. Mean μ locates the center of the distribution. Changing μ shifts the curve along its X axis.



Standard deviation determines the spread.



You can see the size of standard deviation σ by identifying "points of inflection" on the curve. Points of inflection are where the curve begins to turn.



This allows you to scale the axis. This is important because:

- 68% of the area under the curve lies within one standard deviation of the mean (μ ± σ)
- 95% lies within 2 standard deviations of the mean (μ ± 2σ)
- 99.7% lies within 3 standard deviations of the mean (μ ± 3σ)

This fact is known as the **68-95-99.7 rule** for Normal distributions.

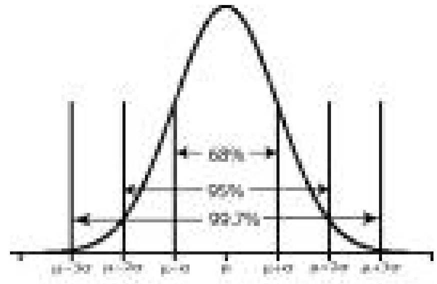


Fig. normal-68-95.ai

This characteristic of Normal curves is referred to as the **68–96–99.7 rule**.

Illustrative example: *Wechsler IQ Scores*. Wechsler scores measure intelligence by providing scores that vary according to a Normal distribution with μ = 100 and σ = 15. Let X represent Wechsler scores: X~N(100, 15). Based on the 68 – 95 – 99.7 rule, we know that 68% of the scores lie in the range 100 ± 15 = 85 to 115, 95% lie in the range 100 ± (2)(15) = 70 to 130, and 99.7% lie in the range 100 ± (3)(15) = 55 to 145.

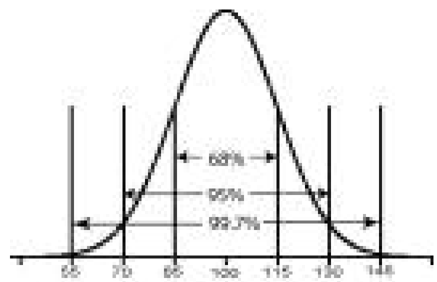


Fig. normal-IQ.ai

To determine probabilities for Normal random variables, we first **standardize** the scores. The standardized values are called **z-scores**:

$$z = \frac{x - \mu}{\sigma} \qquad \textbf{(4.3)}$$

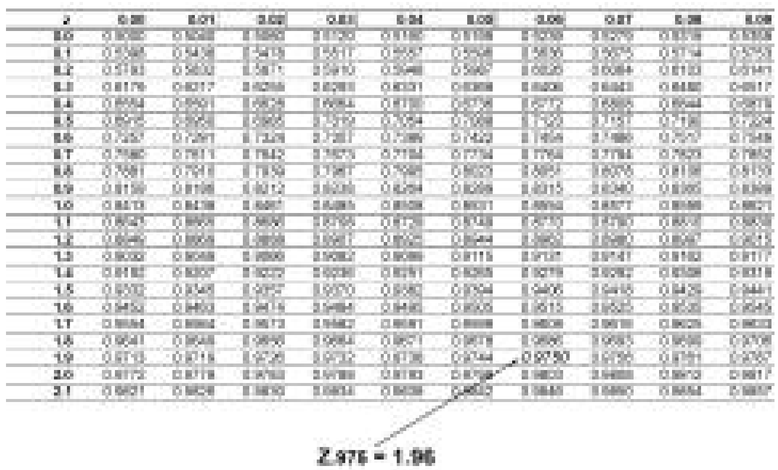
Standardization merely re-scales the variable so that it has μ = 0 and σ = 1. Z Data that are larger than the mean will have positive z scores. Data points that are smaller than the mean have negative z scores. A z score of 1 tells you that the value is one standard deviation *above* the mean. A z score −2 tells you the value is two standard deviations *below* the mean.

Illustrative example: *Weschler*. Weschler intelligence scores vary according to a Normal distribution with mean μ = 100 and σ = 15. An individual with a score of 115 has a z = (115−100) / 15 = 1.00 or 1.00 standard deviations above the mean.

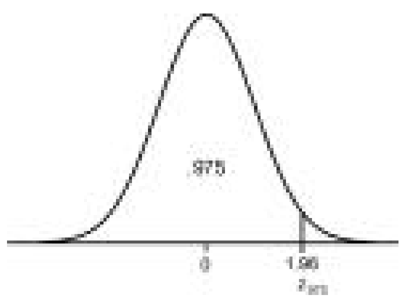
Illustrative example: *Pregnancy length*. Uncomplicated gestational lengths (from last menstrual to birth) vary according to a Normal distribution with mean μ = 39 weeks and standard deviation σ = 2 weeks. A woman whose pregnancy lasts 36 weeks has z = (36 − 39) / 2 = −1.5, or 1.5 standard deviations below average.

Calculating Normal probabilities

Once a data point is standardized, you can use a **Standard Normal table ("z table")** to look up its *cumulative* probability. Our *z* table (back of lecture notes and online) looks like this:



The Normal *z* score of 1.96 (left column 1.9, top row 0.06) is pointed-out. The table entry lets us know that it has cumulative probability 0.9750. Graphically:



Notation: Let $\mathbf{z}_p$ denote a Normal *z*-score with cumulative probability *p*. For example, $z_{.975}$ = 1.96.

We can find the cumulative probability for any value that comes from a a Normal distribution by following these steps:

1. **State** the problem
2. **Standardize** the values
3. Optional: **Draw** the curve (with landmarks)
4. Use the **z table** to determine the cumulative probability

Illustrative example: *Wechsler*. Recall that Wechsler intelligence scores vary according to a Normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. What proportion of Wechsler scores are less than 129.4?

1. **State:** Let X represent Wechsler scores: X~N(100, 15). We want to know $\Pr(X \le 129.4)$.
2. **Standardize:** The score of 129.4 has $z = (129.4 - 100) / 15 = 1.96$.
3. **Draw:** See curve, prior page.
4. ***z* Table:** $\Pr(X \le 129.4) = \Pr(Z \le 1.96) = 0.9750$.

Illustrative example: *Pregnancy length*. Uncomplicated human gestation varies according to a Normal distribution with $\mu = 39$ weeks and $\sigma = 2$ weeks. What proportion of pregnancies lasts less than 41 weeks?

1. **State:** Let X represent gestational length: X~N(39, 2). We want to know $\Pr(X \le 41)$.
2. **Standardize:** $z = (41 - 39) / 2 = 1$ (i.e., one standard deviation above average).
3. **Draw:** optional.
4. ***z* Table:** $\Pr(X \le 41) = \Pr(Z \le 1) = 0.8413$. About 84% of pregnancies last 41 weeks or less.

Probabilities above a certain value (right-tails). The z table includes cumulative probabilities ("left-tails"). When you need to probabilities greater than points (right-tails) use the fact:

(Area under the curve in the right-tail) = 1 − (Area under the curve in the left-tail)

For example, in the above pregnancy length illustration, the probability of a gestation greater than or equal to 41 weeks = 1 − (probability less than or equal to 41 weeks) = 1 − 0.8413 = 0.1587, or about 16%.

Probabilities for observations between certain values. You can calculate areas under the curve between any two points (call them a and b) by subtracting their cumulative probabilities according to the formula $\Pr(a \le Z \le b) = \Pr(Z \le b) - \Pr(Z \le a)$. For example, gestations less than 35 weeks are "premature." Those more than 40 weeks are "post-date." What proportion of births fall between these values?

1. **State:** X~N(39, 2). We are looking for $\Pr(35 \le X \le 40)$.
2. **Standardize:** For 35 weeks, $z = (35 - 39) / 2 = -2$. For of 40 weeks, $z = (40 - 39) / 2 = 0.5$.
3. **Draw** (optional, not shown).
4. **Use *z* table:** $\Pr(35 \le X \le 40) = \Pr(-2 \le Z \le 0.5) = \Pr(Z \le 0.5) - \Pr(Z \le -2) = 0.6915 - 0.0228 = 0.6687$ or about two-thirds of the pregnancies.