

DISTRIBUTIONS OF RANDOM VARIABLES

Jose Toledo Luna

2026-10-13

RANDOM VARIABLES

A **random variable** is a function that maps each event in a sample space to a number

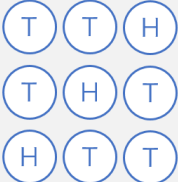
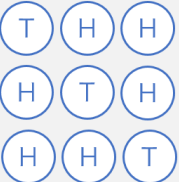
Random variables are useful when the focus is on **functions of outcomes** rather than on **individual outcomes** themselves

In notation, random variables are conventionally represented by uppercase letters, often from the end of the alphabet, such as X , Y , or Z

Example: Flip a fair coin 3 times, the sample space is then

$$S = \{HHH, HHT, HTT, TTT, TTH, THH, THT, HTH\}$$

Let $X = \#$ of heads

			
$X = 0$	$X = 1$	$X = 2$	$X = 3$

Here X can take on the values $x = 0, 1, 2, 3$. For example, $X(HHT) = 2$

Example: Flip a fair coin 3 times. What is the probability of only getting 2 heads ?

Let X = # of heads, then

$$P(X = 2) = \frac{|\{HHT, THH, HTH\}|}{8} = \frac{3}{8}$$

To better understand the behavior of a random variable, we can describe its possible values and the probabilities associated with each. That is, we construct its **probability distribution**

DISTRIBUTION OF A RANDOM VARIABLE

Flipping a fair coin 3 times and define $X = \#$ of heads. Then X can take on the values $x = 0, 1, 2, 3$.

Here X is known as a **discrete random variable**, its distribution is the collection of its values and the probabilities associated with those values

x_i	0	1	2	3
$P(X = x_i)$	1/8	3/8	3/8	1/8

The table represents the **Probability Mass Function (PMF)** of X , where

$$\sum_{i=0}^3 P(X = x_i) = 1.$$

GUIDED PRACTICE

Consider flipping a fair coin 3 times. Let $Y = (\text{\#of heads} - \text{\#of tails})$.
Create the probability distribution of Y .

The sample space $S = \{HHH, HHT, HTT, TTT, TTH, THH, THT, HTH\}$.

EXPECTATION OF A RANDOM VARIABLE

If X is a discrete random variable with possible values $x_1, \dots, x_n$ and corresponding probabilities $P(X = x_1), \dots, P(X = x_n)$, the **expected value** is given by

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^n x_i P(X = x_i) \\ &= x_1 P(X = x_1) + x_2 P(X = x_2) + \dots + x_n P(X = x_n)\end{aligned}$$

$\mathbb{E}(X)$ represents the **long-run average** of a random variable if the experiment it represents were repeated infinitely many times.

It is common to represent $\mathbb{E}(X)$ using the Greek letter μ .

GUIDED PRACTICE

Going back to flipping a fair coin 3 times, with $X = \#$ of heads. Obtain the expected value of X . On average how many number of heads do we expect to obtain ?

VARIANCE OF A RANDOM VARIABLE

Variance measures the spread or dispersion of a random variable's possible values around its **expected value (mean)**.

For a discrete random variable X , the variance is defined as

$$\begin{aligned}\text{Var}(X) &= \sum_{i=1}^n (x_i - \mu)^2 P(X = x_i) \\ &= (x_1 - \mu)^2 P(X = x_1) + (x_2 - \mu)^2 P(X = x_2) + \cdots (x_n - \mu)^2 P(X = x_n)\end{aligned}$$

where $\mu = \mathbb{E}(X)$ is the mean of X .

Specifically, it represents the average squared deviation from the mean, accounting for the probability of each possible outcome.

GUIDED PRACTICE

Going back to flipping a fair coin 3 times, with X = # of heads. Obtain the variance of X .

PROPERTIES OF EXPECTATIONS

Let X and Y be two random variables. Let a, b, c be any constants, then the following hold:

- $\mathbb{E}(cX) = c \mathbb{E}(X)$
- $\mathbb{E}(X + c) = \mathbb{E}(X) + c$
- $\mathbb{E}(X \pm Y) = \mathbb{E}(X) \pm \mathbb{E}(Y)$
- $\mathbb{E}(aX \pm bY) = a \mathbb{E}(X) \pm b \mathbb{E}(Y)$

PROPERTIES OF VARIANCE

- $\text{Var}(aX) = a^2 \text{Var}(X)$
- $\text{Var}(X + c) = \text{Var}(X)$

GUIDED PRACTICE

Let X and Y be two random variables with the following:

$$\mathbb{E}(X) = 5, \mathbb{E}(Y) = 3, \text{Var}(X) = 4, \text{Var}(Y) = 2$$

Now, use properties of expectation and variance to compute the following:

- $\mathbb{E}(2X)$
- $\mathbb{E}(X + 7)$
- $\mathbb{E}(2X - 3Y)$
- $\text{Var}(2X)$
- $\text{Var}(X + 7)$

BINOMIAL DISTRIBUTION

BINOMIAL DISTRIBUTION

The **Binomial distribution** is a discrete probability distribution that describes the **number of successes** in a fixed number of independent trials of a binary experiment.

The key characteristics of a **Binomial distribution** are as follows:

- The process consists of n independent trials
- The outcome of each trial is independent of the others
- Each trial results in either a “success” or “failure”
- The probability of success in each trial is constant and denoted by p

BINS

A good way to remember these assumptions is using BINS

- Binary: success/failures
- Independent: each outcome is independent from each other
- Number: n trials
- Success: each trial has the same probability of success

THE BINOMIAL COEFFICIENT

The binomial coefficient $\binom{n}{k}$ is the number of ways to choose k items from a set of size n , where the order of the choice is ignored.

Mathematically,

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

where,

- $n > 1$
- $k = 0, 1, 2, \dots, n$
- For any integer k , its factorial is

$$k! = k(k-1)(k-2) \dots 1$$

For example, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$

GUIDED PRACTICE

Suppose we flip a fair coin 3 times. How many ways can we obtain exactly 2 heads ?

BINOMIAL DISTRIBUTION FORMULA

Let X represent the number of successes in n trials, follows a Binomial distribution with parameters n and p , denoted $X \sim \text{Bin}(n, p)$

The probability of observing exactly k successes is given by the **probability mass function**:

$$\begin{aligned} P(X = k) &= \binom{n}{k} p^k (1 - p)^{n-k} \\ &= \frac{n!}{(n - k)! k!} p^k (1 - p)^{n-k} \end{aligned}$$

for $k = 0, 1, \dots, n$.

In R, we use the function `dbinom()` to calculate $P(X = k)$

```
> dbinom(x, n, p)
```

GUIDED PRACTICE

Flip a fair coin three times and record the number of heads obtained. Let $X = \#$ of heads, does X follow a Binomial distribution?

Obtain the PMF of X

GUIDED PRACTICE

What is the probability of getting exactly 2 heads ?

What is the probability of getting at least 1 head ?

EXPECTATIONS AND VARIANCE FOR BINOMIAL DISTRIBUTION

Let X follow a Binomial distribution with parameters n and p , it can be shown that

- $\mathbb{E}(X) = np$
- $\text{Var}(X) = np(1 - p)$

For standard deviation, simply take the square root of $\text{Var}(X)$.

NORMAL DISTRIBUTION

CONTINUOUS RANDOM VARIABLES

A **discrete random variable** takes on a finite number of values.

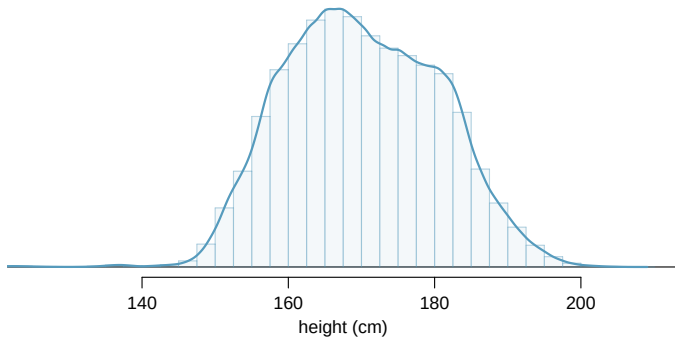
- Number of heads in a set of coin tosses
- Number of people who have had chicken pox in a random sample

A **continuous random variable** can take on any real value in an interval.

- Height in a population
- Blood pressure in a population

A general distinction to keep in mind: discrete random variables are **counted**, but continuous random variables are **measured**.

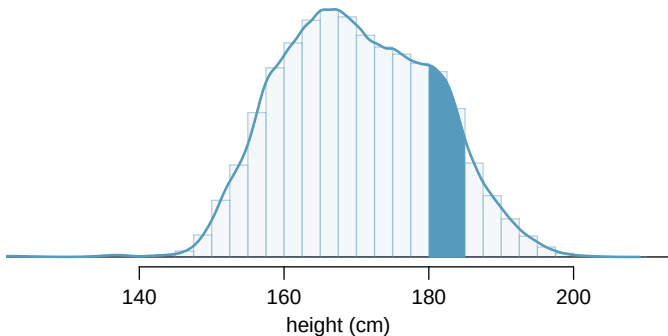
CONTINUOUS DISTRIBUTIONS



The total area under the **density curve** is 1

PROBABILITIES FOR CONTINUOUS DISTRIBUTIONS

The probability that a variable has a value within a specified interval is the area under the curve over that interval



We are calculating $P(a \leq X \leq b)$.

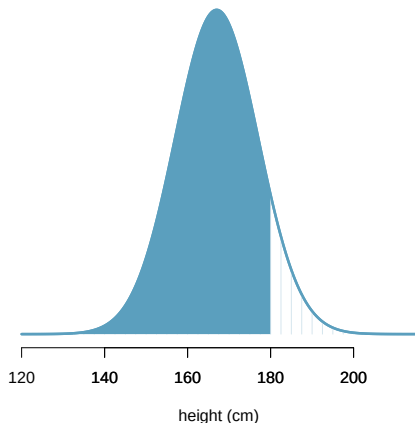
When working with continuous random variables, probability is found for **intervals of values** rather than **individual values**

- The probability that a continuous random variable X takes on any single individual value is 0. That is, $P(X = x) = 0$
- Therefore, $P(a \leq X \leq b)$ is equivalent to $P(a < X < b)$

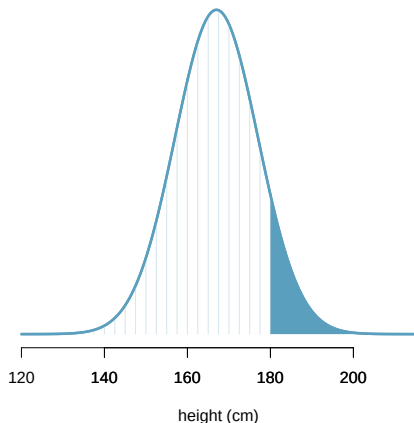
This is not the case for discrete random variables.

The following figures demonstrate how to compute $P(X < a)$ and $P(X > a)$, respectively.

$P(X < a)$



$P(X > a)$



Since the total area under the curve is 1. Then

$$P(X < a) + P(X > a) = 1$$

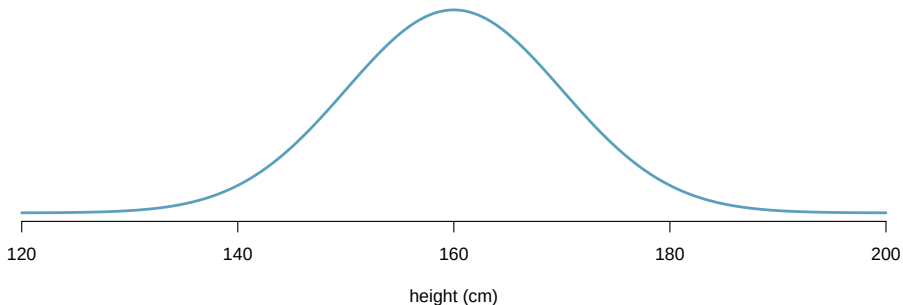
Therefore,

$$P(X < a) = 1 - P(X > a)$$

$$P(X > a) = 1 - P(X < a)$$

NORMAL DISTRIBUTION

The normal distribution is one of the most important and widely used distribution in statistics.



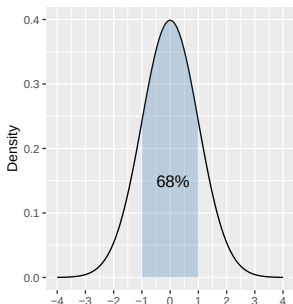
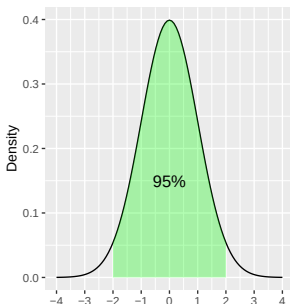
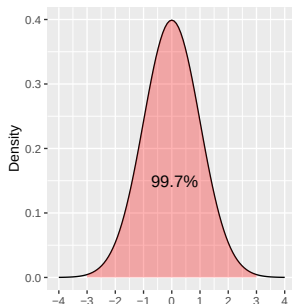
It is characterized by its bell shaped curved and is defined by two paramters: its mean (μ) and standard deviation (σ).

If X is a random variable following a normal distribution, then $X \sim N(\mu, \sigma)$

EMPIRICAL RULE

According to the Empirical Rule, for any normal distribution approximately

68–95–99.7 Rule



- 68% of the data are within 1 standard deviation of the mean
- 95% of the data are within 2 standard deviation of the mean
- 99.7% of the data are within 3 standard deviation of the mean

GUIDED PRACTICE

The distribution of studying for studying a stats 8 course in a week is approximately normal with mean 500 minutes and standard deviation 25 minutes.

What proportion of students study between 450 and 475 minutes a week ?

STANDARD NORMAL DISTRIBUTION

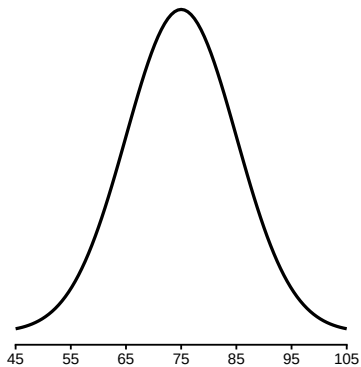
A **standard normal** distribution is defined as a normal distribution with $\mu = 0$ and $\sigma = 1$. It is denoted as $Z \sim N(0, 1)$

Any normal random variable $X \sim N(\mu, \sigma)$, can be transformed into a standard normal random variable Z using the transformation

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

Similarly, any standard normal random variable Z can be transformed back to $X \sim N(\mu, \sigma)$ using the transformation

$$X = \mu + Z\sigma$$



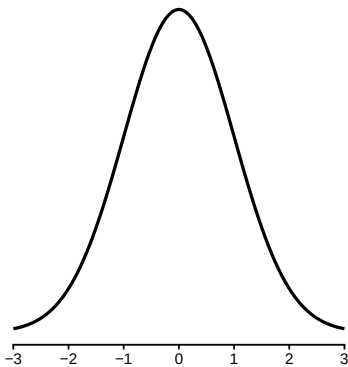
Population (X)

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

An arrow points from the Population (X) curve to the right, passing through this equation box.

$$X = z\sigma + \mu$$

An arrow points from the Standard normal (z) curve to the left, passing through this equation box.



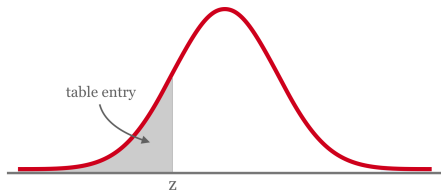
Standard normal (z)

GUIDED PRACTICE

Here are 10 exam scores from a random statistics course

[62, 93, 54, 76, 73, 98, 64, 55, 80, 71]

The scores came from data which followed a normal distribution with mean 72 and standard deviation 10. Standardize the above scores



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015

Table: z-Table for selected values, $P(Z < z)$ values reported

In R, we can compute $P(Z < z)$ using `pnorm()` by default $\mu = 1$, and $\sigma = 1$

```
> pnorm(1.23)
```

```
[1] 0.8906514
```

GUIDED PRACTICE

Grading Policy: Give letter grade A to the top 15% of scores based on a standard normal distribution. What is the cutoff to receive an A ?

QUANTILES OF THE NORMAL DISTRIBUTION

A **quantile** is a value x that has a specified proportion p of the distribution below it, known as **cummulative probability**.

For a normally distributed random variable X , we want to find the value x such that

$$P(X \leq x) = p$$

In R, the function **qnorm()** finds this value for a normal distribution with mean μ and standard deviation σ

Example: For a standard normal distribution, 50% of the observations fall below $x = 0$. Therefore

$$P(X \leq 0) = 0.5$$

We can obtain this value in R

```
qnorm(0.5)
```

```
[1] 0
```

which we can verify by calculating

```
pnorm(0)
```

```
[1] 0.5
```

`qnorm()` is the inverse of `pnorm()`

- `qnorm(p)`: Given a cumulative probability p , find the corresponding value x
- `pnorm(x)`: Given a value x , find the cumulative probability p

GUIDED PRACTICE

Grading Policy: On a standard normal distribution, students who are in the middle 50% of the data will receive a “passing grade” on the class. What are the cutoff values to “pass” the course ?