4: Probability

b	binomial
μ	expected value [parameter]
n	number of trials [parameter]
N	normal
p	probability of success [parameter]
pdf	probability density function
pmf	probability mass function
RV	random variable
σ	standard deviation [parameter]
x	value for random variable X (e.g., observed number of successes for a binomial random variable)
X	random variable X

What is probability?

The **probability of an event** is its relative frequency (expected proportion) in the long run. If an event occurs x times out of n, then its probability will *converge* on $x \div n$ as n becomes infinitely large. For example, if we flip a coin many, many times, we expect to see half the flips turn up heads, but only *in the long run*.

When n is small, the observed relative frequency (proportion) of an event is not be a reliable reflection of its probability. However, as the number of observations n increases, the observed frequency becomes a more reliable reflection of the probability.

EXAMPLE. If a coin is flipped 10 times, there is no guarantee that it will turn up heads 50% of the time. (In fact, most of the time it will not show "5 of 10" heads.) However, if the coin is flipped 100,000 times, chances are pretty good that the proportion of "heads" will be pretty close to 50%.

Random variables (RVs)

A **random variable (RV)** is a quantity that takes of various values depending on chance. In broad mathematical terms, there are two types of random variables: discrete random variables and continuous random variables.

- **Discrete random variables** form a countable set of outcomes. We will study binomial random variables as an example of a type of discrete random variable.

- **Continuous random variables** form an continuum of possible outcomes. We will study normal (Gaussian) random variables as a way to familiarize ourselves with continuous random variables.

Binomial random variables

Definition

There are many types of discrete random variables. Here, we introduce the binomial family.
Binomial random variables are discrete RVs of “counts” that describe the number of “successes” (*X*) in *n* independent Bernoulli trials,[a] where each Bernoulli trial has probability of success designated as *p*.

Binomial random variables have two **parameters**, *n* and *p*.

n ≡ the number of independent “Bernoulli” trials
p ≡ the probability of success for each trial (which does not change from trial to trial)

EXAMPLE. Consider the number of successful treatments (random variable *X*) in 3 patients (*n* = 3) where the probability of success in each instance (*p*) is 0.25. *X* can take on the discrete values of 0, 1, 2, or 3.

Notation. Let “b” represent “binomial distribution” and “~” represent “distributed as.” Thus, *X*~b(*n, p*) is read as “random variable *X* is distributed as a binomial random variable with parameters *n* and *p*.”

EXAMPLE. *X*~b(3, .25) is read “*X* is distributed as a binomial random variable with parameters n=3 and p=.25.”

More notation.

- Let Pr(*X* = *x*) represent “the probability that random variable *X* takes on a value of *x*.”
- Let Pr(*X* ≤ *x*) represent “the probability random variable *X* takes on a value less than or equal to *x*.” This is the **cumulative probability** of the event.

DEFINITION. The **probability mass function (pmf)** assigns probabilities for all possible outcomes of a discrete random variable.

EXAMPLE. The pmf for X~b(3, .25) is shown in Table 1. Probabilities for each potential outcome are shown in the second column. Cumulative probabilities are shown in the third column.

TABLE 1. The pmf for X~b(3, .25).

X **Number of successes**	**Pr(*X* = *x*)** **Probability**	**Pr(*X* ≤ *x*)** **Cumulative Probability**
0 (event A)	0.4219	0.4219
1 (event B)	0.4219	0.8438
2 (event C)	0.1406	0.9844
3 (event D)	0.0156	1.0000

INTERPRETATION. How we calculated these probabilities is not currently the issue. Instead, let us focus on meaning. The above pmf states that for *X*~b(3, .25) we expect to see 0 successes 0.4219 of the time, 1 success 0.4219 of the time, 2 successes 0.1406 of the time, and 3 successes 0.0156 of the time.
Calculations. We will use the app http://www.di-mgt.com.au/binomial-calculator.html to calculate binomial pmfs, There is a link to this app on www.sjsu.edu/faculty/gerstman/StatPrimer .If you for some reason you *need* to calculate binomial probabilities by hand, use the formulas in Chapter 6 of *Basic Biostatistics for Public Health Practice* (Gerstman 2015, Jones & Bartlett, Burlington, MA).

[a] A **Bernoulli trial** is a random event that can take on one of two possible outcomes. One possible outcome is arbitrarily designated as a “success.” The other outcome is designated a “failure.” Outcomes are also designated as either 0 (“failure”) or 1 (“success”).

Rules for working with probabilities

Notation:

- $A \equiv$ event A
- $B \equiv$ event B
- $\Pr(A) \equiv$ the probability of event A
- $\bar{A} \equiv$ the *complement* of event A ≡ not A (i.e., anything other than A)
- $\cup \equiv$ union of events. For example, $A \cup B$ means that either A *or* B occur.
- $\cap \equiv$ intersection of events. For example, $A \cap B$ means that both A *and* B occur.

Rule 1: Probabilities can be no less than 0% and no more than 100%. An event with probability 0 can never occur. An event with probability 1 is certain or always occurs.

$$0 \leq \Pr(A) \leq 1$$

Note that an all the events in Table 1 obey this rule.

Rule 2: All possible outcomes taken together have probability exactly equal to 1.

$$\Pr(\text{all possible outcomes}) = 1$$

Note that in Table 1, Pr(all possible outcomes) = 0.4129 + 0.4129 + .1406 + 0.0156 = 1.

Rule 3: When two events are disjoint (cannot occur together), the probability of their union is the sum of their individual probabilities.

$$\Pr(A \cup B) = \Pr(A) + \Pr(B), \text{ if A and B are disjoint}$$

In Table 1 let $A \equiv$ 0 successes and $A \equiv$ 1 success. $\Pr(A \cup B) = 0.4219 + 0.4219 = 0.8438$.

Rule 4: The probability of a complement is equal to 1 minus the probability of the event.

$$\Pr(\bar{A}) = 1 - \Pr(A)$$

In Table 1, $\bar{A} \equiv$ (1, 2, or 3 successes) and $\Pr(\bar{A}) = 1 - 0.4219 = 0.5781$.

The area under the curve (AUC)

Probability mass functions (pmfs) can be drawn as pmf **histograms**. The area under the bars of pmf histograms correspond to probabilities. For example, the pmf histogram for the random variable in Table 1 is as follows:

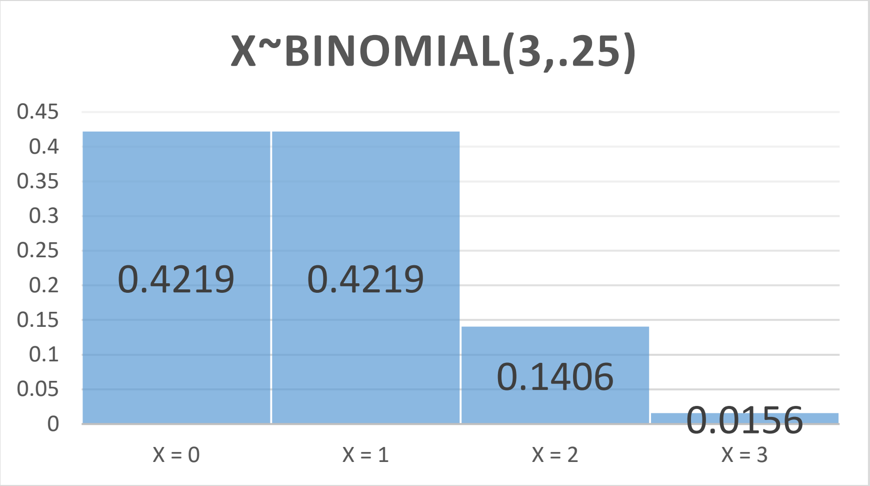


Figure 1. X~b(3, .25).

Area of the first bar: Pr(X = 0). The height of the bar = 0.4219. On the horizontal axis, the first bar stretches from 0 to 1. Therefore, this rectangle has base = 1. The **area** of this bar = height × base = 0.4219 × 1 = 0.4219. This is also the probability that zero events occur. Therefore, Pr(*X* = 0) = area of the bar = 0.4219.

The area under the bars of a pmf histogram corresponds to its probability.

Area of the second bar: Pr(X = 1). The second bar has height = 0.4219, base = 1 (from 1 to 2), and area (i.e., probability) = h × b = 0.4219 × 1 = 0.4219.

Area of the first two bars, i.e., Pr(X = 0) ∪ Pr(X = 1).The combined area of the first two bars = 0.4219 + 0.4219 = 0.8438, corresponding to the probability of 0 or 1 successes.

The area under the pmf histogram **("area under the curve")** between any two points is equal to the probability of the corresponding outcomes.

The "Rule of Complements"

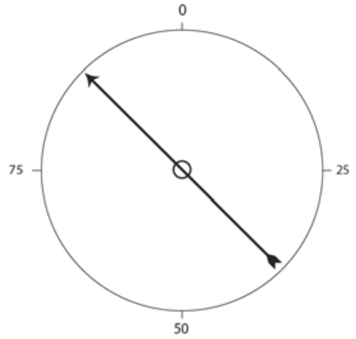
Recall that Ā ≡ the *complement* of event A, i.e., "not A," i.e., anything other than A. Rule 4 (prior page) says Pr(Ā) = 1 – Pr(A).

EXAMPLE. Consider the pmf in Table 1 (X~b(3, .25). Let A ≡ 0 successes. Therefore Ā ≡ 1, 2, or 3 successes. This corresponds to the AUC in the "right tail" of the pmf historgram. By the rule of complements, Pr(Ā) = 1 – 0.4219 = 0.5781.

Introduction to continuous random variables and pdfs

Recall that **continuous random variables** form a continuum of possible outcomes. There are many different types of continuous random variables. These random variable types occur in families (e.g., uniform random variables, normal random variables, chi-squared random variables, etc.).

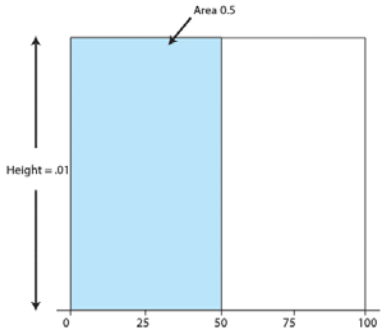
Consider the spinner below. This spinner will generate a continuous uniform random variable with values between 0 and 100. This is a continuous random variable with a range of 0 to 100. The spinner can land on any value between any two points. For example, between 27.5 and 28, it can land on 27.5, 27.75, 27.875, 27.27.9375, etc.



Continuous random number generator: Uniform(0, 100).

To understand continuous random variables, you must accept the thought experiment that the probability of landing on any specific number is 0 (or at least not determinable). For example, Pr(X = 50) = 0. However, the probability of landing between any two values is determinable. For example, the probability of the above random spinner landing on a value between 0 and 50 is .5, i.e., $\Pr(0 \leq X \leq 50) = .50$.

Probability density functions (pdf) assign probabilities for all possible outcomes for continuous random variables. pdfs cannot be shown in tabular form. They can, however, be represented with integral functions (calculus). They can also be drawn. For example, the pdf for the above random number spinner looks like this:



pdf for continuous random number between 0 and 100 showing Pr(0 <= X <= 50)

Importantly, that the **area under the curve (AUC) concept** introduced on the prior page also applies to pdf graphs. For example, the AUC between 0 and 50 (shaded above) = height × base = .01 × 50 = .50, or 50%. Therefore, $\Pr(0 \leq X \leq 50) = .50$ for this particular continuous random variable.

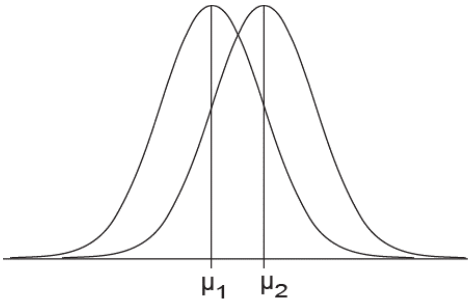
For additional instruction on pdfs see §5.4 in *Basic Biostatistics for Public Health Practice* (Gerstman 2015, Jones & Bartlett, Burlington, MA).

The Normal Distribution

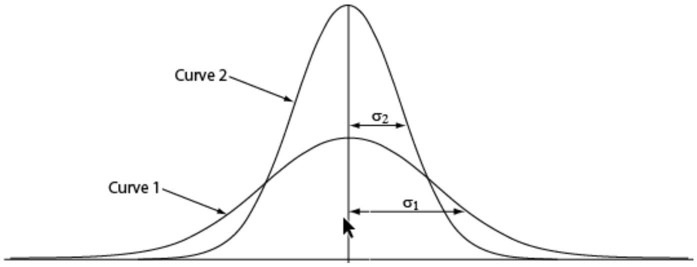
Normal random variables are a *family* of continuous random variables. Each family member is characterized by two **parameters**, μ ("mu") and σ ("sigma").

- μ ≡ the pdf's mean or expected value (indicating central location)
- σ ≡ the pdf's standard deviation (indicating spread)

When μ changes, the location of the pdf changes.



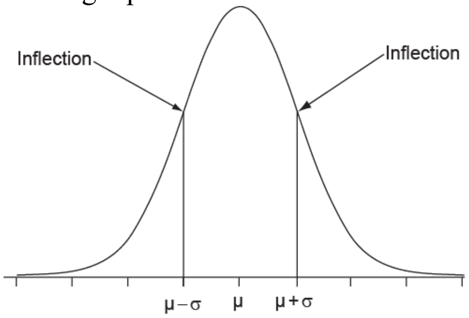
When σ changes, the spread of the pdf changes.



The parameters μ and σ are the analogues (but not the same as) the statistics $\bar{x}$ and *s*. However, you cannot calculate μ and σ. μ and σ are not from any data source.

You can visualize the size of σ on a normal pdf plot by identifying the curve's **points of inflection**. This is where the curve begins to change slope. Trace the slope of the normal curve with your finger. As you "ski" down the slope, the point of inflection is where the slope *begins* to flatten. The left inflection point marks the location μ – σ. This is one σ-unit below the mean.

The right point of inflection marks the location of μ – σ. This is one σ-unit below the mean.



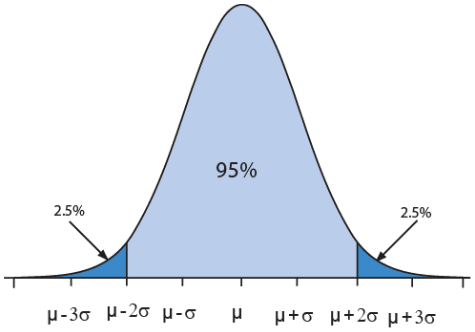
Normal probabilities

The 68-95-99.7 rule helps get a grip on normal probabilities.[b]

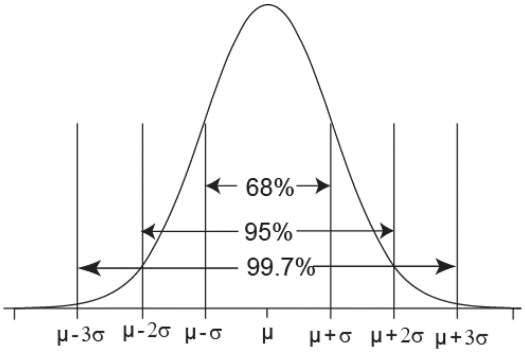
- 68% of the AUC for normal RVs lies in the region μ ± σ
- 95% of the AUC for normal RVs lies in the region μ ± 2σ
- 99.7% of the AUC for normal RVs lies in the region μ ± 3σ

These rules apply only to normal random variables.

Visually, the "95" part of the rule looks like this:



Think in terms of these landmarks:



Although μ and σ vary from one normal random variable to the next, you can apply the 68-95-99.7 rule to any normal random variable if you keep these facts in mind: (1) probability = AUC; (2) The total AUC for the pdf = 1; (3) Values for the random variable lie on the horizontal axis

EXAMPLE. The Wechsler Intelligence Scale is calibrated to produce a normal distribution with μ = 100 and σ = 15 within each age group.

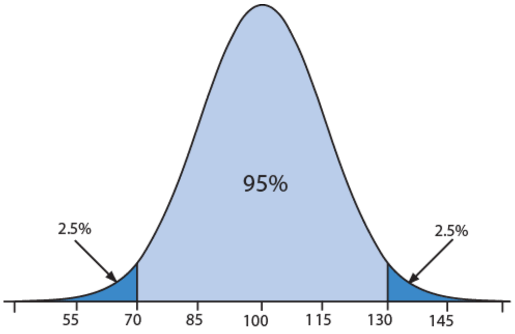
Notation. Let X~N(μ, σ) represent a normal random variable with mean μ and standard deviation σ. Using this notation, Wechsler Intelligence scale scores in a population *X*~N(100, 15). This is stated as "*X* is distributed as a normal random variable with mean 100 and standard deviation 15."

The 68–95–99.7 rule states that for *X*~N(100, 15):

[b] You must accept the fact that the area under the curve (AUC) represents probabilities.

- 68% of the AUC lies in the range μ ± σ = 100 ± 15 = 85 to 115
- 95% of the AUC lies in the range μ ± 2σ = 100 ± (2)(15) = 70 to 130
- 99.7% of the AUC lies in the range μ ± 3σ = 100 ± (3)(15) = 55 to 145

This next figure shows the AUC for X~N(100, 15). Notice the center of the curve is on μ. Also notice landmarks at ±1σ, ±2σ, ±3σ on the horizontal axis.

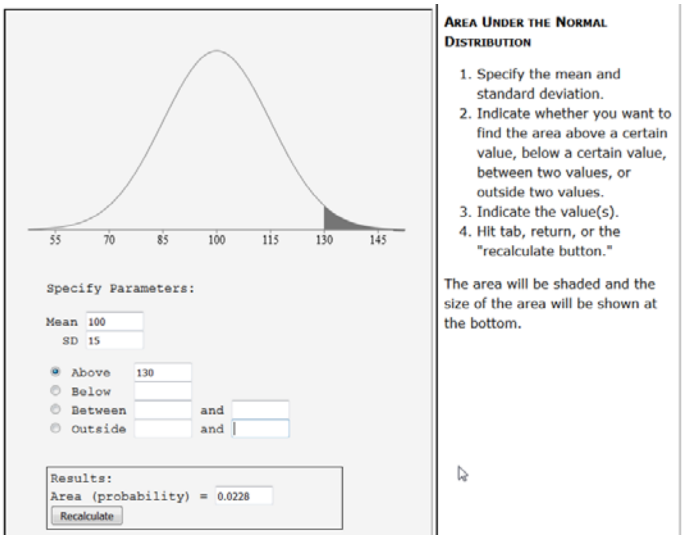


Finding AUCs with for normal random variable app

“In the old days, we found normal probabilities with a a tedious process that relied on tables. We can now use a app for the purpose. Either way, the key concept is the AUC between any two points corresponds to probability.

We can use this app to calculate AUCs between any two points for any X~N(μ, σ): http://onlinestatbook.com/2/calculators/normal_dist.html. There is a link to this app on www.sjsu.edu/faculty/gerstman/StatPrimer .

Example. Plug in values for X~(100,15). The AUC between 130 and ∞ corresponds to the right tail of the pdf. Note that this AUC (probability) is 0.0228 (roughly 2.5%).



Specify Parameters:

Mean 100
SD 15

● Above 130
○ Below
○ Between and
○ Outside and

Results:
Area (probability) = 0.0228

Area Under the Normal Distribution

1. Specify the mean and standard deviation.
2. Indicate whether you want to find the area above a certain value, below a certain value, between two values, or outside two values.
3. Indicate the value(s).
4. Hit tab, return, or the "recalculate button."

The area will be shaded and the size of the area will be shown at the bottom.