

1

Notation

Random variables are capital letters such as X and Y .

A particular value of X is x .

Vectors are bold lowercase symbols (e.g., \mathbf{a} , $\boldsymbol{\theta}$).

Matrices are bold uppercase symbols (e.g., \mathbf{A} , $\boldsymbol{\Theta}$).

Unfortunately, if vectors are lowercase and random variables are uppercase, my notation comes in conflict when discussing a vector of random variables (i.e., the variables as theoretical entities, not their particular scores). Such a vector will be a bold uppercase letter, (e.g., \mathbf{X}). A vector of particular scores from the variables in \mathbf{X} will be shown as \mathbf{x} .

\mathbb{R}	= The set of all real numbers
\mathbb{Z}	= The set of all integers
\mathbb{N}_0	= The set of all non-negative integers: $\{0, 1, 2, \dots\}$
\mathbb{N}_1	= The set of all positive integers: $\{1, 2, 3, \dots\}$
$[a, b]$	= The set of real numbers in the interval between a and b
\in	= Is a member of
$\binom{n}{k}$	= The <i>binomial coefficient</i> . It is just a shortcut notation for $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Read aloud, $\binom{n}{k}$ is “ n choose k ” or the number of combinations that n things have when taken k at a time.
$f_X(x; \theta)$	= The probability density function or probability mass function of X with parameters θ
$F_X(x; \theta)$	= The cumulative distribution function of X with parameters θ
$E(X)$	= The expected value of X
μ_X, m_X	= The population and sample mean of X
σ_X, s_X	= The population and sample standard deviation of X
σ_X^2, s_X^2	= The population and sample variance of X
γ_1, g_1	= The population and sample skewness
γ_2, g_2	= The population and sample kurtosis
\mathbf{R}_X	= The correlation matrix of the variables in vector \mathbf{X}
\mathbf{R}_X	= The correlation matrix of the variables in vector \mathbf{X}
\mathbf{A}'	= The transpose of matrix \mathbf{A}
\mathbf{A}^{-1}	= The inverse of matrix \mathbf{A}
$\text{diag}(\mathbf{A})$	= The column vector formed from the diagonal of matrix \mathbf{A}
$\text{diag}(\mathbf{a})$	= The diagonal matrix formed from column vector \mathbf{a}
$\mathbf{1}$	= A column vector of ones. Its length can be inferred by context.
\mathbf{I}	= An identity matrix. Its dimensions can be inferred by context.

Introduction

Although great painters can make good art with cheap brushes, they need high quality tools to work at the upper limits of their craft. On the other hand, giving an untrained person an expensive set of brushes is unlikely to result in noticeably better art. So it is with these tools. They are of little use to unprepared hands—and in foolish hands, they might even be dangerous. But in hands caring and competent, they can make reasoning more rigorous, results more robust, and recommendations more relevant.

It takes a really long time to complete a comprehensive psychological assessment. Interviews, test administration, and behavioral observations take many hours. Scoring tests and integrating test information can take even more time. Then there is the considerable task of actually writing the report.¹ Finally, the results of the report are presented to the client and other decision-makers. If this much time is put into the process, it makes sense to make the most of it. Unfortunately, while making the most of it, it is easy to go too far—making reckless recommendations from iffy inferences and flights of fancy.

Humans are very very good at some things that are extraordinarily complex, such as pattern recognition. Humans are not so great at combining numeric data in their heads to come to valid conclusions.² Furthermore, certain kinds of formal logic, though once considered to be the pinnacle of human intellect, are actually fairly simple for computers. Thus, we should let computers do what they do best: calculate. We humans have the job of deciding which calculations the computers should perform, interpreting what the results mean, and deciding how the new information should be used.

Most introductory psychometrics textbooks are designed to help researchers create well constructed tests and therefore cover many details that are not useful to clinicians and fail to cover many practical issues that clinicians should know about. This book is intended to help you extract useful information from the data you already have in ways that you may not have known were possible. That my emphasis is on

¹ Putting together my first psychological evaluation report one summer in Texas, I labored, fretted, and sweated for more than 30 hours! Even now, at my fastest, it still takes me about 3 hours of uninterrupted work to write what I consider to be a good report.

² When Andy Clark (2004, p. 5) said that biological brains are “to put it bluntly, bad at logic, good at Frisbee,” it was no insult to Frisbee aficionados—robots are no match for humans at the sport.

the practical in no way implies that this book is dumbed down. My aim is to make psychometrics useful to clinicians. If some useful ideas are complex, I hope to make them accessible—but without resorting to superficial glossing. Some background knowledge of psychometrics is necessary to understand how these tools work and, more importantly, when their underlying assumptions have been violated.

This book probably never would have been written had I not several years ago stumbled across Ley's (1972) *Quantitative aspects of psychological assessment*. I admire the book's blend of clarity, practicality, and depth. Why did I write my own book instead of recommending that clinicians download and use Ley's book?³ Well, I do recommend reading Ley's book. In contrast to my approach, Ley often takes time to gently lay out mathematical proofs of many ideas. Thus Ley's book is a great introduction to the often considerably less accessible corpus of academic writings on psychometrics. I wanted to present much of the same material but with more of an eye to application. I also wanted to present many ideas not included in Ley's book. In addition, I chose to write this book because I believe that Ley had the right idea but that in an era in which no one had a home computer, few clinicians would have the knowledge, motivation, and stamina to use his equations on a regular basis. Now that computers are used by all clinicians, equations like those presented by Ley can be made easy to use. All of the "tools" in this toolkit have been made into computer programs which can be downloaded for free at my website:

<http://my.ilstu.edu/~wjschne/tests.html>

³ Although Ley's book is out of print, Ley has been made it freely available for download: psychassessment.com.au/

Part I

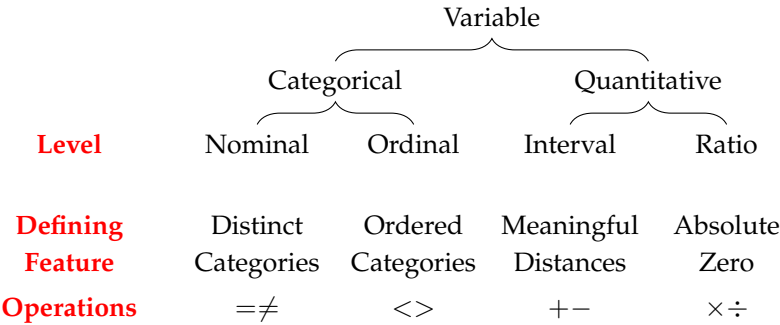
A Psychometrics Primer

3

Variables

Because there are so many different kinds of **variables**, it is helpful to classify them by type so that we can think clearly about them and not use them incorrectly. One reason that we must be aware of a variable’s classification is that some statistics only make sense with certain types of variables. To take a famously silly example from [Zacharias \(1975\)](#), you could calculate the average number in a telephone directory, but you wouldn’t try calling it. The mean has no meaning for this kind of variable. Unfortunately, it is not always so obvious which statistics can be applied to which variables. If we unknowingly apply the wrong kind of statistic to a variable, our statistical software will give us a result *but we will not know that it is nonsense*.

The classic **taxonomy** for variable types ([Stevens, 1946](#)) distinguishes among *nominal*, *ordinal*, *interval*, and *ratio* scales (Figure 3.1). Although it is clear that this taxonomy is incomplete ([Cicchetti et al., 2006](#)), there is not yet consensus about how exactly the list should be amended. If this book were about how to design psychological tests, we would explore this topic in depth. For our purposes, it is enough to say that the fundamentals of measurement and **scaling** is philosophically complex and that controversies abound ([Michell, 1997](#)).



A **variable** is an abstract category of information that can take on different values.

A **taxonomy** is a classification system.

Scaling refers to the procedures by which we decide how to assign numbers to the attributes we measure.

Figure 3.1: Stevens’ levels of measurement

3.1 Nominal scales

In a **nominal scale**, we note only that some things are different from others and that they belong to two or more **mutually exclusive** categories. If we say that a person has Down syndrome (trisomy 21), we are implicitly using a nominal scale in which there are people with Down syndrome and people without Down syndrome. In a true nominal scale, there are no cases that fall between categories. To be sure, we might have some difficulty figuring out and reliably agreeing upon the category to which something belongs—but there is no conceptual space between categories.

In the messy world of observable reality, few true nominal variables exist as defined here. Most so-called nominal variables in psychology are merely nominal-*ish*. With respect to Down syndrome, we could say that people either have three copies of the 21st chromosome or they do not. However, if we did say that—and meant it—we’d be wrong. In point of fact, there are many cases of partial trisomy. There are other cases of Down syndrome in which part of chromosome 21 is copied to another chromosome. However, because cases like this are sufficiently rare and because such distinctions are usually not of vital importance, Down Syndrome is treated as if it were a true nominal variable. Even though Down Syndrome might technically come in degrees (both phenotypically and in terms of the underlying chromosomal abnormalities), the distinction between not having the condition and not having it is not **arbitrary**.

In contrast, consider the diagnosis of *intellectual disability*, what has historically been called “mental retardation.” We might have only two categories in our coding scheme (Intellectual Disability: Yes or No) but it is widely recognized that the condition comes in degrees (e.g., “none,” “borderline,” “mild,” “moderate,” “severe,” and “profound”). Thus, intellectual disability is not even conceptually nominal. It is a continuum that has been divided at a convenient but mostly arbitrary point (Figure 3.2). Distinguishing between a **dichotomous** variable that is nominal by nature and one that has an underlying continuum matters because there are statistics that apply only to the latter type of variable (e.g., the tetrachoric correlation coefficient). Even so, in many procedures, the two types of dichotomies can be treated identically (e.g., comparing the means of two groups with an independent-samples t-test).

What if the categories in a nominal scale are not mutually exclusive? For example, suppose that we have a variable in which people can be classified as having either Down syndrome or Klinefelter syndrome (XXY). Obviously, this is a **false dichotomy**. Most people have neither condition. Well, we can expand the nominal variable to have three cat-

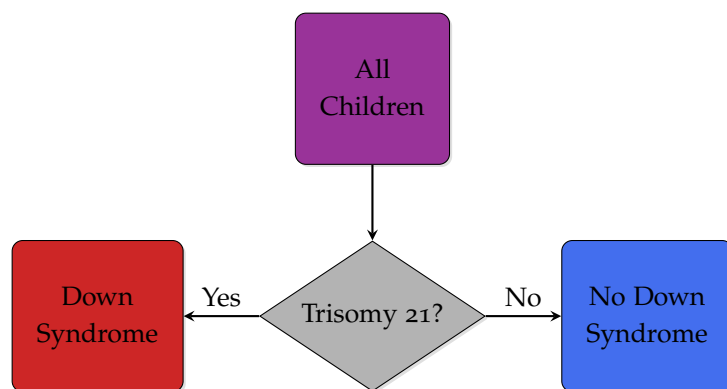
A **nominal scale** groups observations into unordered categories.

In **mutually exclusive** categories nothing belongs to more than one category (at the same time and in the same sense).

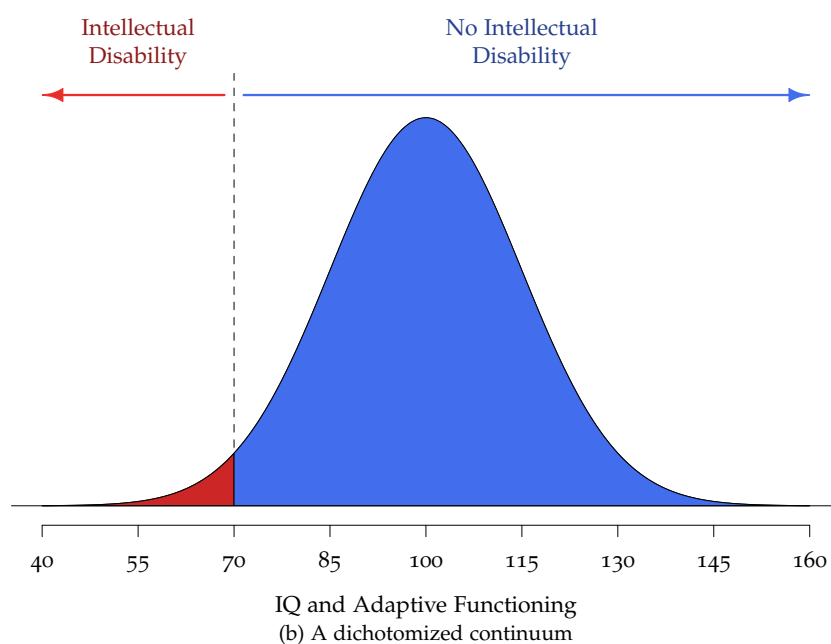
Arbitrary refers to things that are decided not by necessity but by preference, convenience, or whim.

A **dichotomy** is a division of something into two categories.

In a **false dichotomy**, two alternatives are presented as if they are the only alternatives when, in fact, there are others available.



(a) A (nearly) true dichotomy



(b) A dichotomized continuum

egories: Down syndrome, Klinefelter syndrome, and neither. What if a person has both Down syndrome *and* Klinefelter syndrome? Okay, we just add a fourth category: both. This combinatorial approach is not so much a problem for some purposes (e.g., the ABO blood group system), but for many variables, it quickly becomes unwieldy. If we wanted to describe all chromosomal abnormalities with a single nominal variable, the number of combinations increases exponentially with each new category added. This might be okay if having two or more chromosomal abnormalities is very rare. If, however, the categories are not mutually exclusive and combinations are common enough to matter, it is generally easiest to make the variable into two or more nominal variables (Down syndrome: Yes or no; Klinefelter syndrome: Yes or no; Edwards syndrome: Yes or no; and so forth). Some false

Figure 3.2: Two kinds of dichotomous variables

dichotomies are so commonly used that people know what you mean, even though they are incomplete (e.g., Democrat vs. Republican) or potentially insensitive to people who do not fit neatly into any of the typical categories (e.g., male vs. female).

When we list the categories of a nominal variable, the order in which we do so is mostly arbitrary. In the variable *college major*, no major intrinsically comes before any other. It is convenient to list the categories alphabetically but the order is different in different languages and will change as the names of college majors evolve. However, strict alphabetical order is not always logical or convenient.¹

3.2 Ordinal scales

In an **ordinal scale**, things are still classified by category but the categories have a particular order. Suppose that we are conducting behavioral observations of a child in school and we record when the behavior occurred. The precise time at which the behavior occurred (e.g., 10:38 AM) may be uninformative. If the class keeps a fairly regular schedule, it might be more helpful to divide the day into categories such as early morning, recess, late morning, lunch, and afternoon. This way it is easy to see if behavior problems are more likely to occur in some settings (or activities) than in others. It does not matter that these divisions are of unequal lengths or that they do not occur at precisely the same time each day. In a true ordinal variable, the distance between categories is either undefined, unspecified, or irrelevant.

Most measurement in psychological assessment involves ordinal scales, though in many cases ordinal scales might appear to be other types of scales. Questionnaires that use Likert-type scales are very clearly ordinal (e.g., Figure 3.3). Even though true/false items on questionnaires might seem like nominal scales, they are usually ordinal because the answer indicates whether a person has either more or less of an attribute. That is, *more* and *less* are inherently ordinal concepts. Likewise, ability test items are ordinal, even though *correct* vs. *incorrect* might seem like nominal categories. Ability tests are designed such that a correct response indicates more ability than an incorrect response. The ordinal nature of ability test items is especially clear in cases that allow for partial credit.

Some scales are only partially ordinal. For example, *educational attainment* is ordinal up to a certain point in most societies, but branches out as people acquire specialized training. For a career in psychology, the educational sequence is high school diploma, associate's degree, bachelor's degree, master's degree, and doctoral degree.² However, this is not the sequence for real estate agents, hair stylists, and pilots. If we wanted to compare educational degrees across professions, how

¹ For example, in a variable such as *ethnic identity*, the number of possible categories is very large so members of very small groups are given the option of writing in their answer next to the word "other." The "other" category is placed at the end of the list to avoid confusion.

An ordinal scale groups observations into ordered categories.

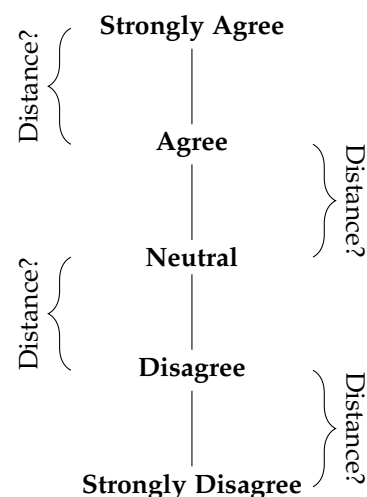


Figure 3.3: In a Likert-type scale, the distances between categories are undefined.

² Obviously, some of these degrees can be skipped and the endpoint is different for different careers in psychology. Furthermore, not all degrees fit neatly in this sequence (e.g., the school psychology specialist degree).

would we rank them? For example, how would we compare a law degree with a doctorate in geology? Are they the same? Does one degree indicate higher educational attainment than the other? The answers depend on the criteria that we care about—and different people care about different things. Thus, it is difficult to say that we have an ordinal scale when we compare educational attainment across professions.

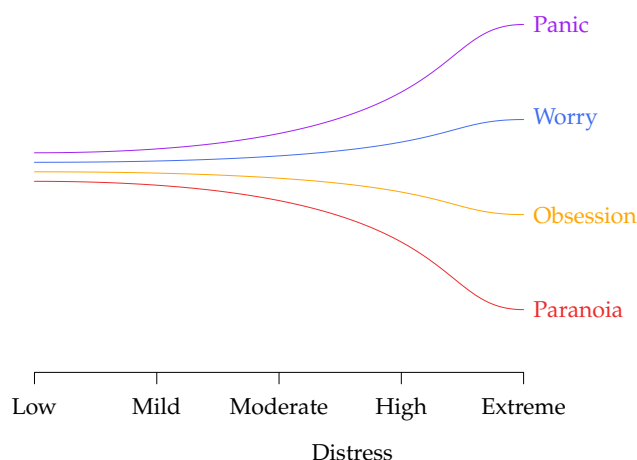


Figure 3.4: Anxiety is more differentiated at higher levels of distress

Like educational attainment, many psychological traits are more differentiated at some points in the continuum than at others. For example, as seen in Figure 3.4, it is sometimes convenient to lump the various flavors of trait anxiety together at the low and middle range of distress and then to distinguish among them at the high end. It is difficult to say who is more anxious, a person who is extremely paranoid³ or a person with a severe case of panic disorder. We can say that each is more anxious than the average person (an ordinal comparison) but each has a qualitatively different kind of anxiety. This problem is easily solved by simply talking about two different scales (paranoia and panic). However, there are different kinds of paranoia (e.g., different mixtures of hostility, fear, and psychosis) and different kinds of panic (e.g., panic vs. fear of panic). One can always divide psychological variables into ever narrower categories, making comparisons among and across related constructs problematic. At some point, we gloss over certain qualitative differences and treat them as if they were comparable, even though, strictly speaking, they are not.

³ Although *paranoia* is not traditionally considered a type of anxiety, it is clear that anxiety (about the possibly malevolent intentions of others) is a core feature of the trait.

3.3 Interval scales

With **interval scales**, not only are the numbers on the scale ordered, the distance between the numbers (i.e., intervals) is meaningful. A good example of an interval scale is the calendar year. The time elapsed from 1960 to 1970 is the same as the time elapsed from 1970 to 1980. In contrast, consider a standard Likert scale from a questionnaire. What is the distance between *disagree* and *agree*? Is it the same as the distance between *agree* and *strongly agree*? If it were, how would we know? In interval scales, all such mysteries disappear.

It is not always easy to distinguish between an interval scale and an ordinal scale. Therapists sometimes ask clients to rate their distress "on a scale from 0 to 10." Probably, in the mind of the therapist, the distance between each point of the scale is equal. In the mind of the client, however, it may not work that way. In Figure 3.5, a hypothetical client thinks of the distance between 9 and 10 as much greater than the distance between 0 and 1.

It is doubtful that any subjectively scaled measurement is a true interval scale. Even so, it is clear that some ordinal scales are more interval-like than others. Using item response theory (Embretson and Reise, 2000), it is possible to sum many ordinal-level items and scale the total score such that it approximates an interval scale. It is important to note, however, that item response theory does not accomplish magic. The application of item response theory in this way is justified only if the ordinal items are measuring an underlying construct that is by nature at the interval (or ratio) level. No amount of statistical wizardry can alter the nature of the underlying construct. Sure, you can apply fancy math to the numbers, but a construct that is ordinal by nature will remain ordinal no matter what you do or convince yourself that you have done.

Most of the tools used in psychological assessment make use of ordinal scales and transform them such that they are treated as if they were interval scales. Is this defensible? Yes, a defense is possible (e.g., Borsboom and Mellenbergh, 2004) but not all scholars will be convinced by it (Michell, 1997, 2008). As an act of faith, I will assume that most of the scales used in psychological assessment (measures of abilities, personality traits, attitudes, interests, motivation, and so forth) are close enough to interval scales that they can be treated as such. In many instances, my faith may be misplaced, but where exactly can only be determined by high quality evidence. While I await such evidence, I try to balance my faith with moderate caution.

In an interval scale the distance between numbers has a consistent meaning at every point on the scale.

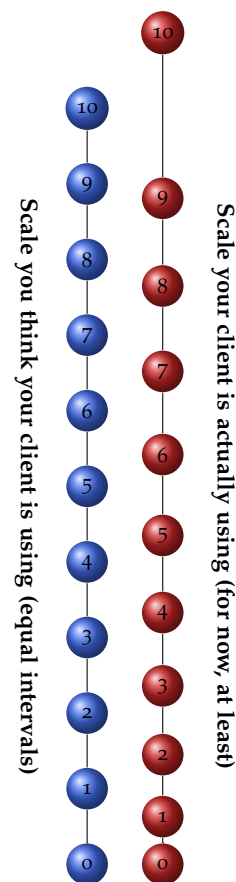


Figure 3.5: Subjective units of distress may not be of equal length

3.4 Ratio scales

In a **ratio scale**, zero represents something special: the absence of the quantity being measured. In an interval scale, there may be a zero, but the zero is just another number in the scale. For example, 0°C happens to be the freezing point of water at sea level but it does not represent the absence of heat.⁴ Ratio scales do not usually have negative numbers but there are exceptions.⁵

What does a true zero have to do with ratios? In interval scales, numbers can be added and subtracted but they cannot be sensibly divided. Why not? Because when you divide one number by another, you are creating a ratio. A ratio tells you how big one number is compared to another number. Well, how big is any number? The magnitude of a real number is its distance from zero (i.e., its absolute value). If zero is not a meaningful number on a particular scale, then ratios computed from numbers on that scale will not be meaningful. Therefore, because interval scales do not have a true zero, meaningful ratios are not possible. For example, although 20°C is twice as far from 0°C as 10°C , it does not mean that 20°C is twice as hot as 10°C . In contrast, these types of comparisons are possible on the Kelvin scale because 0°K is a true zero. That is, 20°K really is twice as hot as 10°C .

In psychological assessment, there are a few true ratio scales that are commonly used. Whenever anything is counted (e.g., counting how often a behavior occurs in a direct observation), it is a ratio scale. However, this is a bit tricky. If I observe how many times a child speaks out of turn in class and I use this as an index of impulsivity, it is no longer a ratio scale. Why? Well, the actual variable, *number of outbursts* is a true ratio variable because 0 outbursts means the absence of outbursts. However, if I use the number of outbursts as a proxy variable for *impulsivity*, then 0 outbursts probably does not indicate the absence of impulsivity. At best it indicates lower levels of impulsivity. This same problem exists for the measurement of reaction times. Reaction time is a true ratio because a reaction time of 0 means that no time has elapsed between the onset of the stimulus and the response. However, reaction time data used in clinical applications are often proxies for traits that are interval-level concepts, such as *inattention* on a continuous performance test. Why are psychological traits such as cognitive abilities, personality traits, and so forth interval-level concepts? Because we do not yet have any means of defining what, for example, zero intelligence or zero extraversion would look like. Attempts have been made (Jensen, 2006), but they have not yet proved persuasive.

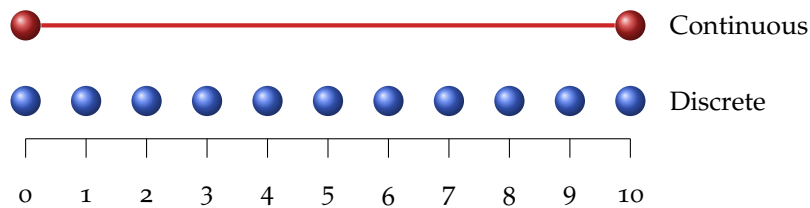
A ratio scale has a true zero, in addition to all the properties of an interval scale.

⁴ The absence of heat occurs at -273.15°C or 0°K .

⁵ For example, in a checking account balance, negative numbers indicate that the account is overdrawn. Still, a checking account balance is a true ratio scale because a zero indicates that there is no money in the account.

3.5 Discrete vs. Continuous Variables

Interval and ratio variables can be either discrete or continuous. **Discrete variables** can assume some values but not others. Once the list of acceptable values has been specified, there are no cases that fall between those values. For example, the number of bicycles a person owns is a discrete variable because the variable can assume only the non-negative integers. Fractions of bicycles are not considered. Discrete variables usually take on integer values but this is not necessarily the case.



A **discrete variable** can only take on exact values from a specified list.

Figure 3.6: Discrete variables have gaps whereas continuous variables have none.

When a variable can assume any value within a specified interval, the variable is said to be **continuous**. In theory, this means that fractions and decimals can be used to achieve any level of precision that we desire. In practice, we must round the numbers at some point, technically making the variable discrete. In truly discrete variables, however, we could not change our minds and decide to round the numbers such that they are more precise.

A **continuous variable** can take on any value within a specified range.

4

Probability Distributions

4.1 Random Variables

Because we first learn about variables in an algebra class, we tend to think of variables as having values that can be solved for—if we have enough information about them. Algebraic variables are “determined” by constraints—in the form of equations.¹

Random variables are not like algebraic variables. Random variables simply take on values because of some random process. If we say that the outcome of a throw of a six-sided die is a random variable, there is nothing to “solve for.” There is no equation that determines the value of the die. Instead, it is determined by chance and the physical constraints of the die. That is, the outcome must be one of the numbers printed on the die and the six numbers are equally likely to occur (if the die is “fair”). This illustrates an important point. The word *random* here does not mean “anything can happen.” Random variables have outcomes that are subject to random processes, but those random processes *do* have constraints on them such that some outcomes are more likely than others—and some outcomes never occur at all.

When we say that the throw of a six-sided die is a random variable, we are not talking about any particular throw of a particular die but, in a sense, *every* throw (that has ever happened or ever could happen) of *every* die (that has ever existed or could exist). Imagine an immense, roaring, never-ending, cascading flow of dice falling from the sky and a giant scoreboard nearby recording the relative frequencies of ones, twos, threes, fours, fives, and sixes as each die lands and then disappears. That’s a random variable.

4.2 Sample Spaces

The set of all possible outcomes of a random variable is the **sample space**. Continuing with our example, the sample space of a single throw of a six-sided die is the set $\{\square, \square, \square, \square, \square, \square\}$. *Sample space* is

¹ If I say that x is a variable and that $x + 6 = 8$, that equation constrains x so that it must be exactly 2, no more and no less. What about when there are two variables and one equation, such as $x + y = 10$? Isn’t x free to be any value? Yes, but even here, x is still constrained by y . Once one variable is determined, the other is also determined. If y is 3, then x must be 7. In other words, because of the constraint of the equation, there is only one degree of freedom, meaning that only one of the two variables is free to vary at a time.

Random variables have values that are determined by a random process.

A sample space is the set of all possible values that a random variable can assume.

a curious term. Why *sample* and why *space*? With random variables, **populations** are infinitely large, at least theoretically. Random variables just keep spitting out numbers forever! So any time we actually observe numbers generated by a random variable, we are always observing a **sample**; actual infinities cannot be observed in their entirety. A **space** is a set that has mathematical structure. Most random variables generate either integers or real numbers, both of which are structured in many ways (e.g., order).

Unlike distributions having to do with dice, many distributions have a sample space with an infinite number of elements. Actually there are two kinds of infinity we can consider. One distribution we will discuss later is the Poisson distribution. Its sample space is the set of whole numbers: $\{0, 1, 2, \dots\}$, which extends to positive infinity. The sample space of continuous variables is infinitely large for another reason. Between any two points in a continuous distribution, there is an infinite number of other points. For example, in the continuous uniform distribution, the sample space consists of all real numbers between two points. Many continuous distributions have sample spaces that involve both kinds of infinity. For example, the sample space of the normal distribution consists of all real numbers from negative infinity to positive infinity.

4.3 Probability Distributions

Each element of a random variable's sample space has some probability associated with it. When we list the probabilities of each possible outcome, we have specified the variable's **probability distribution**. In other words, if we know the probability distribution of a variable, we know how probable each outcome is. In the case of a throw of a single die, each outcome is equally likely (Figure 4.1).

There is an infinite variety of probability distributions, but a small subset of them have been given names. Now, one can manage one's affairs quite well without ever knowing what a Bernoulli distribution is, or what a χ^2 distribution is, or even what a normal distribution is. However, sometimes life is a little easier if we have names for useful things that occur often. Most of the distributions with names are not really single distributions, but families of distributions. The various members of a family are unique but they are united by the fact that their probability distributions are generated by a particular mathematical function (more on that later). In such cases, the probability distribution is often represented by a graph in which the sample space is on the X-axis and the associated probabilities are on the Y-axis. In Figure 4.2, 16 probability distributions that might be interesting and useful to clinicians are illustrated. Keep in mind that what are pic-

A **population** consists of all entities under consideration.

A **sample** is a subset of a population.

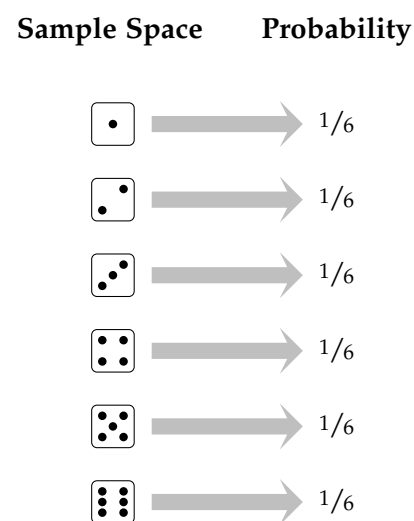


Figure 4.1: The probability distribution of a throw of a single die

In a **probability distribution**, there is an assignment of a probability to each possible element in a variable's sample space.

tured are only particular members of the families listed; some family members look quite different from what is shown in Figure 4.2.

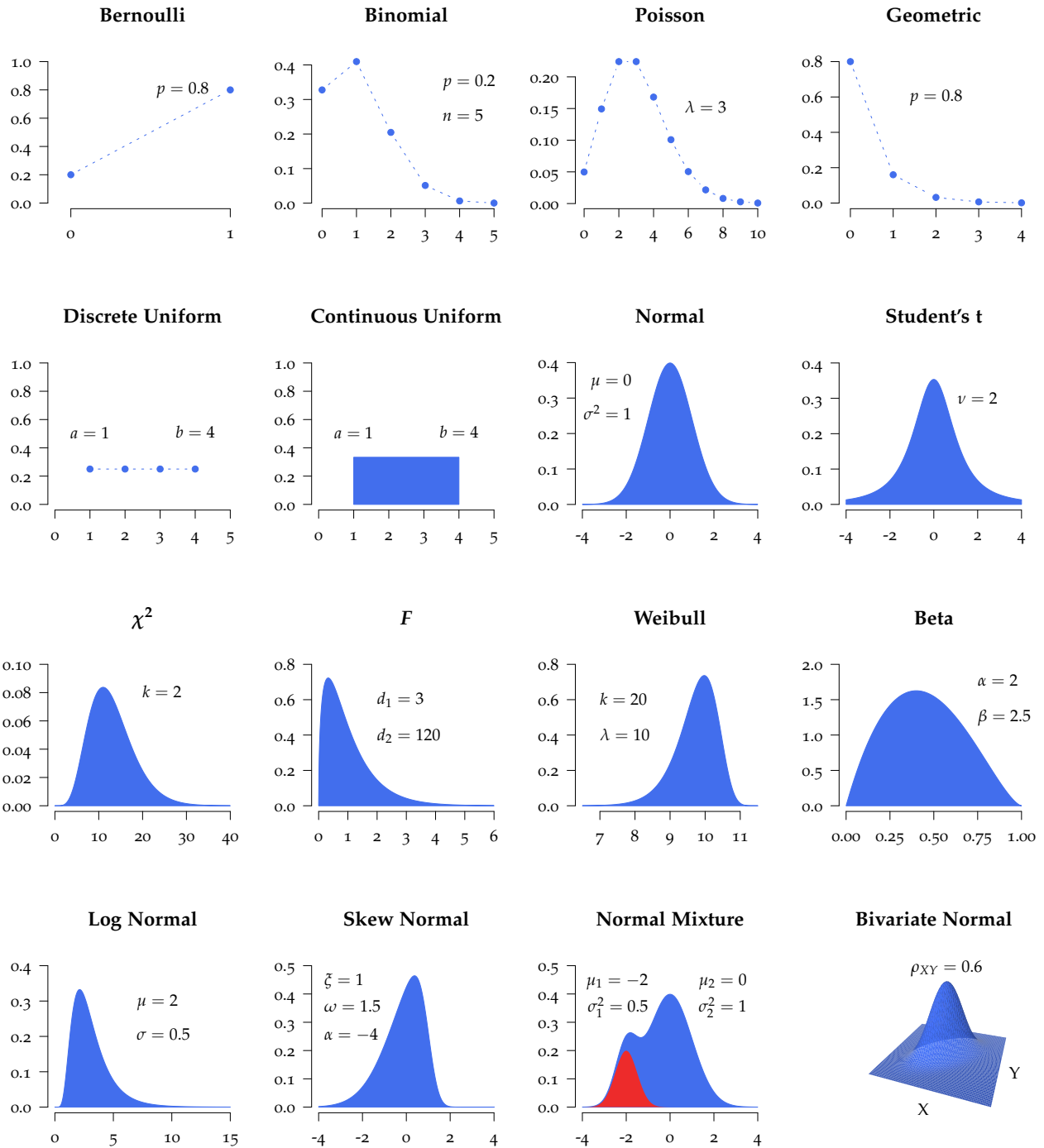


Figure 4.2: A gallery of useful distributions

4.4 Discrete Uniform Distributions

The throw of a single die is a member of a family of distributions called the **discrete uniform distribution**. It is “discrete” because the elements in the sample space are countable, with evenly spaced gaps between them. For example, there might be a sequence of 8, 9, 10, and 11 in the sample space but there are no numbers in between. It is “uniform” because all outcomes are equally likely. With dice, the numbers range from a lower bound of 1 to an upper bound of 6. In the family of discrete uniform distributions, the lower and upper bounds are typically integers, mostly likely starting with 1. However, any real number a can be the lower bound and the spacing k between numbers can be any positive real number. For the sake of simplicity and convenience, I will assume that the discrete uniform distribution refers to consecutive integers ranging from a lower bound of a and an upper bound of b .

This kind of discrete uniform distribution has a number of characteristics listed below. I will explain each of them in the sections that follow.

Lower Bound:	a	\in	\mathbb{Z}
Upper Bound:	b	\in	\mathbb{Z}
			$b > a$
Sample Space:	x	\in	$\{a, a + 1, \dots, b\}$
Number of points:	n	$=$	$b - a + 1$
Mean:	μ	$=$	$\frac{a+b}{2}$
Variance:	σ^2	$=$	$\frac{n^2-1}{12}$
Skewness:	γ_1	$=$	0
Kurtosis:	γ_2	$=$	$-\frac{6(n^2+1)}{5(n^2-1)}$
Probability Mass Function:	$f_X(x; a, b)$	$=$	$\frac{1}{n}$
Cumulative Distribution Function:	$F_X(x; a, b)$	$=$	$\frac{x-a+1}{n}$

4.5 Parameters of Random Variables

The lower bound a , the spacing between numbers k , and the number of points n are the discrete uniform distribution’s **parameters**. If we restrict ourselves to the more common case in which the sample space consists of consecutive integers, we can say that the parameters are simply the lower bound a and the upper bound b . The word *parameter* has many meanings but here it refers to a characteristic of a distribution family that helps us identify precisely which member of the family we are talking about. Most distribution families have one,

A **discrete uniform distribution** is a family of random variable distributions in which the sample space is an evenly spaced sequence of numbers, each of which is equally likely to occur.

As we go, I will also explain the mathematical notation. For example, $a \in \mathbb{Z}$ means that a is an integer because \in means *is a member of* and \mathbb{Z} is the set of all integers.

A **parameter** is a defining feature of a random variable’s probability distribution.

two, or three parameters.

If you have taken an algebra class, you have seen parameters before, though the word *parameter* may not have been used. Think about the formula of a line:

$$y = mx + b$$

Both x and y are variables, but what are m and b ? Well, you probably remember that m is the slope of the line and that b is the y -intercept. If we know the slope and the intercept of a line, we know exactly which line we are talking about. No additional information is needed to graph the line. Therefore, m and b are the line's *parameters*, because they uniquely identify the line.² All lines have a lot in common but there is an infinite variety of lines because the parameters, the slope and the intercept, can take on the value of any real number. Each unique combination of parameter values (slope and intercept) will produce a unique line. So it is with probability distribution families. All family members are alike in many ways but they also differ because of different parameter values.

The discrete uniform distribution (i.e., the typical variety consisting of consecutive integers) is defined by the lower and upper bound. Once we know the lower bound and the upper bound, we know exactly which distribution we are talking about.³ Not all distributions are defined by their lower and upper bounds. Indeed, many distribution families are unbounded on one or both sides. Therefore, other features are used to characterize the distributions, such as the population mean.

4.6 Probability Mass Functions

Many distribution families are united by the fact that their probability distributions are generated by a particular mathematical function. For discrete distributions, those functions are called **probability mass functions**. In general, a mathematical function is an expression that takes one or more constants (i.e., parameters) and one or more input variables, which are then transformed according to some sort of rule to yield a single number.

A probability mass function transforms a random variable's sample space elements into probabilities. In Figure 4.1, the probability mass function can be thought of as the arrows between the sample space and the probabilities. That is, the probability mass function is the thing that was done to the sample space elements to calculate the probabilities. In Figure 4.1, each outcome of a throw of the die was mapped onto a probability of $1/6$. Why $1/6$, and not some other number? The probability mass function of the discrete uniform distribution tells us the answer.

² What about other mathematical functions? Do they have parameters? Yes! All of them do! For example, in the equation for a parabola ($y = ax^2 + bx + c$), a , b , and c determine its precise shape.

³ If we allow the lower bound to be any real number and the spacing to be any positive real number, the discrete uniform distribution can be specified by three parameters: the lower bound a , the spacing between numbers k ($k > 0$), and the number of points n ($n > 1$). The upper bound b of such a distribution would be $b = a + k(n - 1)$.

A probability mass function (pmf) is a mathematical expression that gives the probability that a discrete random variable will equal a particular element of the variable's sample space.

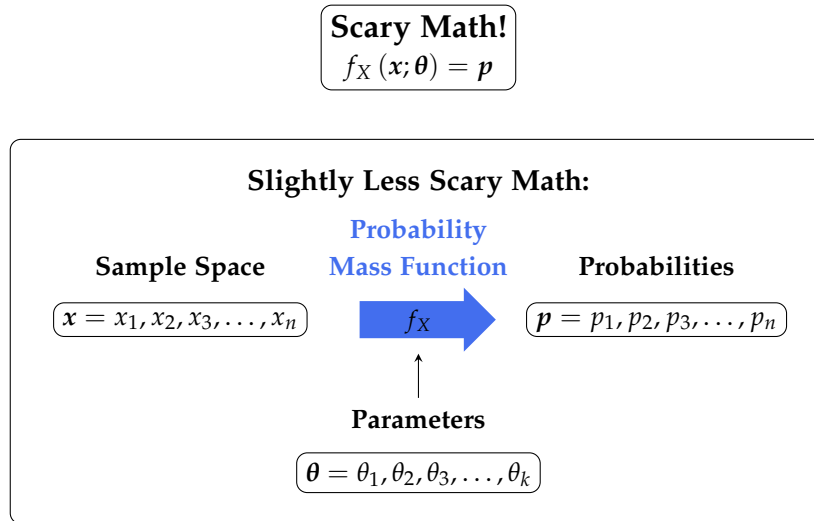


Figure 4.3: Though one may be scarier than the other, both boxes mean the same thing: Probability mass functions tell us how probable each sample space element is.

The probability mass function of the discrete uniform distribution is fairly simple but the notation can be intimidating at first (Figure 4.3). By convention, a single random variable is denoted by a capital letter X . Any particular value of X in its sample space is represented by a lowercase x . In other words, X represents the variable in its totality whereas x is merely one value that X can take on. Confusing? Yes, statisticians very work hard to confuse us—and most of the time they succeed!

The probability mass function of random variable X is denoted by $f_X(x)$. This looks strange at first. It means, "When random variable X generates a number, what is the probability that the outcome will be a particular value x ?" That is, $f_X(x) = P(X = x)$, where P means "What is the probability that..." Thus, $P(X = x)$ reads, "What is the probability that random variable X will generate a number equal to a particular value x ?" So, $f_X(7)$ reads, "When random variable X generates a number, what is the probability that the number will equal 7?"

Most probability mass functions also have parameters, which are listed after a semi-colon. In the case of the discrete uniform distribution consisting of consecutive integers, the lower and upper bounds a and b are included in the function's notation like so: $f_X(x; a, b)$. This reads, "For random variable X with parameters a and b , what is the probability that the outcome will be x ?" Some parameters can be derived from other parameters, as was the case with the number of points n in the sample space of a discrete uniform distribution: $n = b - a + 1$. The probability for each outcome in the sample space is the same and there are n possible outcomes. Therefore, the probability associated

with each outcome is $1/n$.

Putting all of this together, if a and b are integers and $a < b$, for all n integers x between a and b , inclusive:

$$f_X(x; a, b) = \frac{1}{b - a + 1} = \frac{1}{n} \quad (4.1)$$

Where

X = A random variable with a discrete uniform distribution

f_X = The probability mass function of X

x = Any particular member of the sample space of X


a = The lower bound of the sample space

b = The upper bound of the sample space

n = The number of points in the sample space

You might notice that x is not needed to calculate the probability. Why? Because this is a *uniform* distribution. No matter which sample space element x we are talking about, the probability associated with it is always the same. In all distributions that are not uniform, the position of x matters and thus influences the probability of its occurrence.

4.7 Cumulative Distribution Functions

The **cumulative distribution function** tells us where a sample space element ranks in a distribution. Whereas the probability mass function tells us the probability that a random variable will generate a particular number, the cumulative distribution function tells us the probability that a random variable will generate a particular number or less. The cumulative distribution function of the roll of a die (Figure 4.4) tells us that the probability of rolling at least a  is $4/6$ (i.e., $2/3$).

The cumulative distribution function is often distinguished from the probability mass function with a capital F instead of a lowercase f . In the case of a discrete uniform distribution consisting of n consecutive integers from a to b , the cumulative distribution function is:

$$F_X(x; a, b) = \frac{x - a + 1}{b - a + 1} = \frac{x - a + 1}{n} \quad (4.2)$$

Where

X = A random variable with a discrete uniform distribution

F_X = The cumulative distribution function of X

x = Any particular member of the sample space of X

a = The lower bound of the sample space

b = The upper bound of the sample space

n = The number of points in the sample space

A **cumulative distribution function (cdf)** is a mathematical expression that gives the probability that a random variable will equal a particular element of the variable's sample space or less.

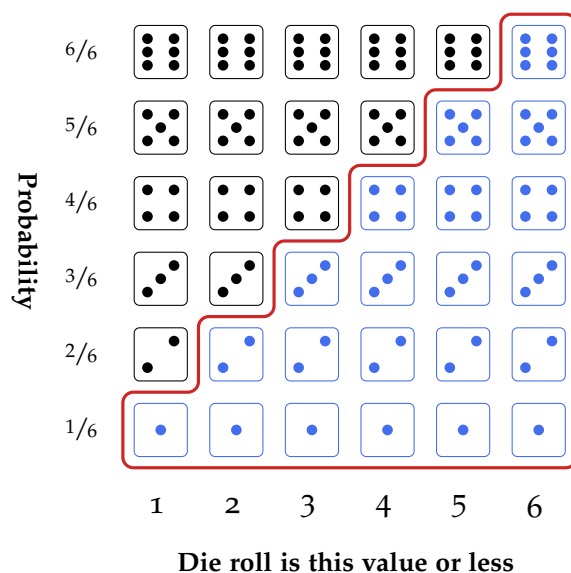


Figure 4.4: The cumulative distribution function of the roll of a die is $F_X(x) = \frac{x}{6}$

4.8 Quantile functions

The inverse of the cumulative distribution function is the **quantile function**. The cumulative distribution starts with a value X in the sample space and tells us p , the proportion of values in that distribution that are less than or equal to X . A quantile function starts with a proportion p and tells us the value X that splits the distribution such that the proportion p of the distribution is less than or equal to X . As seen in Figure 4.5, if you see a graph of a continuous distribution function, just flip the X and Y axes and you have a graph of a quantile function!

A **quantile function** tells us which value in the sample space of a random variable is greater than a particular proportion of the values the random variable generates.

4.9 Generating a Random Sample in Excel and R

To generate a number from the discrete uniform distribution in Excel, use the `RANDBETWEEN` function. For example, to generate a random number between 1 and 100, in any cell type:

```
=RANDBETWEEN(1,100)
```

This cell can be copied as many times as is needed to generate a random sample.

In R, the `runif` function generates numbers from the **continuous uniform distribution**. To make the distribution discrete, the ceiling function rounds the numbers up to the nearest integer.⁴

```
# n = the sample size
n <- 1000
# a = the lower bound
a <- 1
```

⁴ A common mistake is to use traditional rounding (up or down to the nearest integer), which makes the lower and upper bounds only half as likely to occur as all the numbers in between.

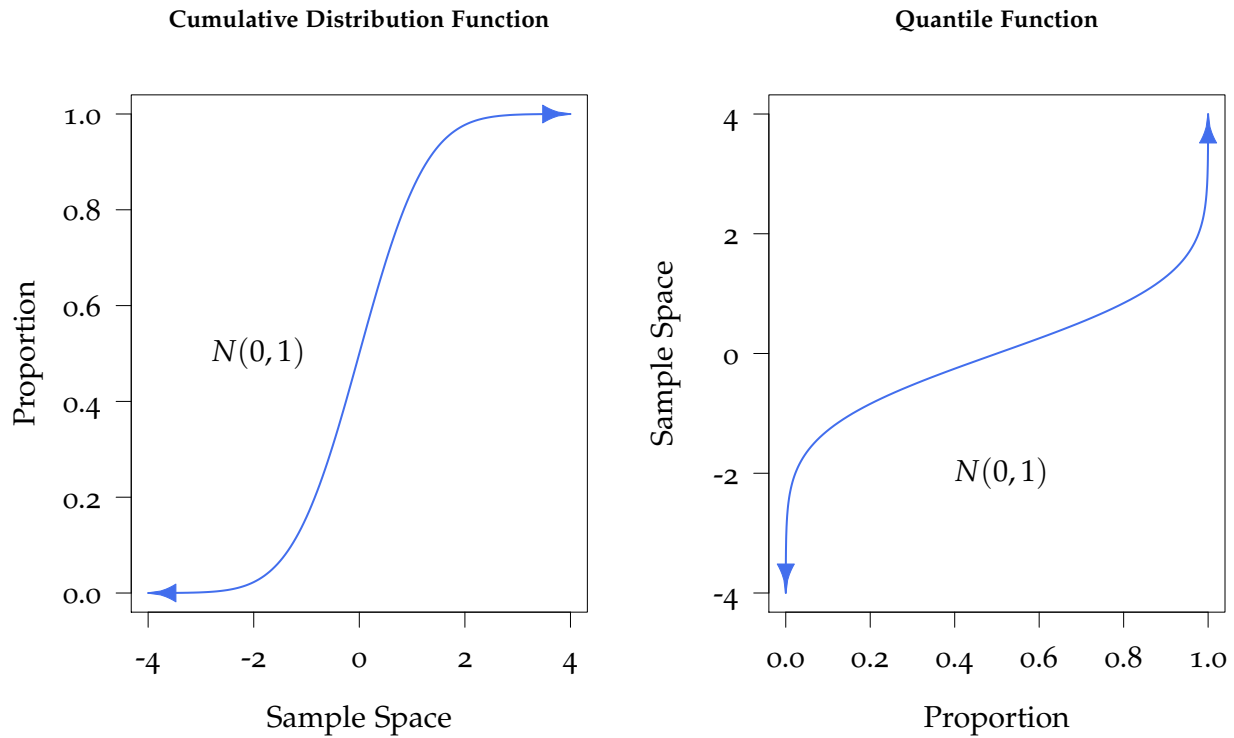


Figure 4.5: The quantile function is the inverse of the cumulative distribution function: Just flip the X and Y axes!

```
# b = the upper bound
b <- 100
# s = the sample with a discrete uniform
# distribution The runif function generates a
# number from the continuous uniform distribution.
# The ceiling function rounds up to the nearest
# integer.
s <- ceiling(runif(n, a - 1, b))
# Generate a bare-bones plot of the frequency
# distribution of the sample
hist(s)
```

4.10 Bernoulli Distributions

Sample Space:	x	\in	$\{0, 1\}$
Probability of success in each trial:	p	\in	$[0, 1]$
Mean:	μ	$=$	p
Variance:	σ^2	$=$	$p(1 - p)$
Skewness:	γ_1	$=$	$\frac{1 - 2p}{\sqrt{p(1 - p)}}$
Kurtosis:	γ_2	$=$	$\frac{1}{p(1 - p)} - 6$
Probability Mass Function:	$f_X(x; p)$	$=$	$p^x(1 - p)^{1 - x}$
Cumulative Distribution Function:	$F_X(x; p)$	$=$	$x + p(1 - x)$

The toss of a single coin has the simplest probability distribution that I can think of—there are only two outcomes and each outcome is equally probable (Figure 4.6). This is a special case of the **Bernoulli distribution**. The Bernoulli distribution can describe any random variable that has two outcomes, one of which has a probability p and the other has a probability $q = 1 - p$. In the case of a coin flip, $p = 0.5$. For other variables with a Bernoulli distribution, p can range from 0 to 1.

In psychological assessment, many of the variables we encounter have a Bernoulli distribution. In ability test items in which there is no partial credit, examinees either succeed or fail. The probability of success on an item (in the whole population) is p . In other words, p is the proportion of the entire population that correctly answers the question. Some ability test items are very easy and the probability of success is high. In such cases, p is close to 1. When p is close to 0, few people succeed and items are deemed hard. Thus, in the context of ability testing, p is called the **difficulty parameter**. This is confusing because when p is high, the item is easy, not difficult. Many people have suggested that it would make more sense to call it the “easiness parameter” but the idea has never caught on.

True/False and Yes/No items on questionnaires also have Bernoulli distributions. If an item is frequently endorsed as true (“I like ice cream.”), p is high. If an item is infrequently endorsed (“I like black licorice and mayonnaise in my ice cream.”), p is very low. Oddly, the language of ability tests prevails even here. Frequently endorsed questionnaire items are referred to as “easy” and infrequently endorsed items are referred to as “difficult,” even though there is nothing particularly easy or difficult about answering them either way.

Notation notes:

$[a, b]$ is the set of all real numbers in the interval between a and b .

In the **Bernoulli distribution**, there are only two outcomes: a “success” (1) and a “failure” (0). If a success has a probability p then a failure has a probability of $1 - p$.

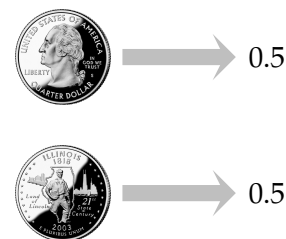
Sample Space Probability

Figure 4.6: The probability distribution of a coin toss

The **difficulty parameter** is the proportion of people who succeed on an item (or say ‘Yes’ or ‘True’ or otherwise score a 1 on a random variable with a Bernoulli distribution.).

Generating a sample from the Bernoulli distribution

In Excel, the RAND function generates a random real number between 0 and 1. The INT function rounds down to the nearest integer. To generate either a 0 or a 1, with 1 having a probability of p :

=INT(RAND()*p)

In R, the same idea is used but the names of functions are different. We have already used the runif, which generates real numbers between two values. By default, runif has a lower bound of 0 and an upper bound of 1. The floor function rounds round to the nearest integer. Using the runif and floor functions together produces the Bernoulli distribution.

```
# n = sample size
n <- 1000
# p = probability
p <- 0.8
# s = sample
s <- floor(runif(n) + p)
# bare-bones plot
barplot(table(s))
```



4.11 Binomial Distributions

Number of Trials:	n	\in	\mathbb{N}_1
Sample Space:	x	\in	$\{0, \dots, n\}$
Probability of success in each trial:	p	\in	$[0, 1]$
Probability of failure in each trial:	q	$=$	$1 - p$
Mean:	μ	$=$	np
Variance:	σ^2	$=$	npq
Skewness:	γ_1	$=$	$\frac{1-2p}{\sqrt{npq}}$
Kurtosis:	γ_2	$=$	$\frac{1}{npq} - \frac{6}{n}$
Probability Mass Function:	$f_X(x; n, p)$	$=$	$\binom{n}{x} p^x q^{n-x}$
Cumulative Distribution Function:	$F_X(x; n, p)$	$=$	$\sum_{i=0}^x \binom{n}{i} p^i q^{n-i}$

Let's extend the idea of coin tosses and see where it leads. Imagine that two coins are tossed at the same time and we count how many heads there are. The outcome we might observe will be zero, one, or two heads. Thus, the sample space for the outcome of the tossing of two coins is the set $\{0, 1, 2\}$ heads. There is only one way that we will observe no heads (both coins tails) and only one way that we will observe two heads (both coins heads). In contrast, as seen in Figure 4.7,


there are two ways that we can observe one head (heads-tails & tails-heads).

The probability distribution of the number of heads observed when two coins are tossed at the same time is a member of the **binomial distribution** family. The binomial distribution occurs when **independent** random variables with the same **Bernoulli distribution** are added together.

Imagine that a die is rolled 10 times and we count how often a  occurs. Each roll of the die is called a **trial**. The sample space of this random variable is $\{0, 1, 2, \dots, 10\}$. What is the probability that a  will occur 5 times? or 1 time? or not at all? Such questions are answered by the binomial distribution's **probability mass function**:


$$f_X(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x} \quad (4.3)$$

Applied to this example,

X = The random variable (i.e., the number of times that  occurs when the die is thrown 10 times)

x = Any particular member of the sample space (i.e., $x \in \{0, 1, 2, \dots, 10\}$.)

n = The number of times that the die is thrown (i.e., $n = 10$).

p = The probability that a  will occur on a single throw of the die (i.e., $p = 1/6$).

$\binom{n}{x}$ = The *binomial coefficient*. It is just a shortcut notation for $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. Read aloud, $\binom{n}{x}$ is “ n choose x ” or the number of combinations that n things have when taken x at a time.

Since $n = 10$ and $p = 1/6$, the probability mass function from Equation 4.3 simplifies to:

$$f_X(x) = \binom{10}{x} \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{10-x}$$

If we take each element x of the sample space from 0 to 10 and plug it into the equation above, the probability distribution will look like Figure 4.8.

Clinical Applications of the Binomial Distribution

When would a binomial distribution be used by a clinician? One particularly important use of the binomial distribution is in the detection of **malinger**ing. Sometimes people pretend to have memory loss or attention problems in order to win a lawsuit or collect insurance benefits. There are a number of ways to detect malingering but a common method is to give a very easy test of memory in which the person has

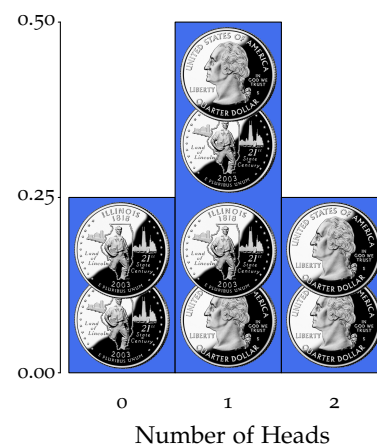


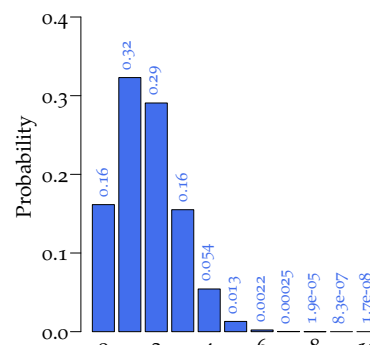


Figure 4.7: Probability distribution of the number of heads observed when two coins are tossed

Two random variable are said to be **independent** if the outcome of one variable does not alter the probability of any outcome in the other variable.

Wait! Hold on! I thought that throwing dice resulted in a (discrete) *uniform* distribution. Well, it still does. However, now we are asking a different question. We are only concerned with two outcomes each time the die is thrown:  and not . This is a Bernoulli distribution, not a uniform distribution, because the probability of the two events is unequal: $\{1/6, 5/6\}$

Every time a random variable generates a number, that instance of the variable is called a **trial**, which is also known as an **experiment**.



at least a 50% chance of getting each test item correct even if the person guesses randomly.

Suppose that there are 20 questions. Even if a person has the worst memory possible, that person is likely to get about half the questions correct. However, it is possible for someone with a legitimate memory problem to guess randomly and by bad luck answer fewer than half of the questions correctly. Suppose that a person gets 4 questions correct. How likely is it that that a person would, by random guessing, only answer 4 or fewer questions correctly?

We can use the binomial distribution's cumulative distribution function. However, doing so by hand is rather tedious. In Excel, the answer can be found quite easily:⁵

=BINOM.DIST(4,20,0.5,TRUE)

Using R, the answer is found with the pbinom function:

```
p <- pbinom(4, 20, 0.5)
```

Using either method, we can see that the probability of randomly guessing and getting 4 or fewer items correct out of 20 items total is approximately 0.006, which is so low that the hypothesis that the person is malingering seems plausible.⁶

Graphing the binomial distribution in Excel and R

In Excel, the BINOM.DIST function will calculate the binomial distribution's probability mass function and the cumulative distribution function. Let's say that $n = 10$ and $p = 0.8$.

First we need a series of integers from 0 to 10 (the number of Bernoulli trials) arranged in a column. We could enter each number one by one but there is a faster way. In cell A2, enter 0. In cell A3, enter 1. Now select both cells (Click down on cell A2, drag to cell A3, and release.). Position the mouse over the lower right corner of cell A3 and the mouse cursor icon will change from a thick white plus to thin black plus. Now, click and drag to cell A12 and release. Now you should see a column of integers from 0 to 10.

In cell B1, type *Probability Mass Function*. In cell B2, type:

=BINOM.DIST(A2,10,0.8,FALSE)

and press ENTER. It would be tedious to type the same formula ten more times. Fortunately, Excel has many shortcuts. Click cell B2. Position the mouse icon over the lower right corner of cell B2 (The cursor will change to a thin black plus again.) and double-click. Excel has extended the series all the way down to cell B12!

Now, in cell C1, type *Cumulative Distribution Function*. In cell C, type:

=BINOM.DIST(A2,10,0.8,TRUE)

⁵ In Excel, all the probability distribution functions have a final argument which determines whether the function is a cumulative distribution function or a probability mass function (or probability density function for continuous variables). If the final argument of the function is TRUE, then it is a cumulative distribution function. If FALSE, it is a probability mass function (or probability density function).

⁶ Note here that there is a big difference between these two questions:

1. If the person is guessing at random (i.e., not malingering), what is the probability of answering correctly 4 questions or fewer out of 20?
2. If the person answers 4 out of 20 questions correctly, what is the probability that the person is guessing at random (and therefore not malingering)?

Here we answer only the first question. It is an important one, but the answer to the second question is probably the one that we really want to know. We will answer it in another chapter when we discuss positive predictive power. For now, we should just remember that the questions are different and that the answers can be quite different from each other.

and press ENTER. This is the same formula as was in cell B2 except that the FALSE argument has been changed to TRUE. Extend the series by double-clicking the lower right corner of cell C2.

To create the graph, select the range A1:C12 (Click down on cell A1, drag to cell C12, and release.). The next step depends on which version of Excel you have but basically you insert a scatterplot. The default graph probably does not look very good but with some playing around, you can produce a graph to your liking. You can see a graph that looks good to me in Figure 4.9. I put the legend at the top, added major and minor gridlines on both axes, set the axis limits manually, increased the font sizes, changed the colors of the lines and markers, changed the markers to circles, set the dash type of lines to dashed, and resized the graph.

In R, graphing the binomial distribution is fairly simple if a bare-bones plot is needed. First, the sample space is generated (a sequence from 0 to 10.), using the seq function. The associated probability mass function probabilities are found using the dbinom function. The cumulative distribution function probabilities are found using the pbinom function.

```
# Make a sequence of numbers from 0 to 10
SampleSpace <- seq(0, 10)
# Probability mass distribution for binomial
# distribution with n = 10, p = 0.8
pmfBinomial <- dbinom(SampleSpace, size = 10, prob = 0.8)
# Cumulative distribution function for binomial
# distribution with n = 10, p = 0.8
cdfBinomial <- pbinom(SampleSpace, size = 10, prob = 0.8)
# Generate a bare-bones plot of the probability
# mass distribution
plot(pmfBinomial ~ SampleSpace)
# Generate a bare-bones plot of the cumulative
# distribution function
plot(cdfBinomial ~ SampleSpace)
```

However, making the graph look professional involves quite a bit of code that can look daunting at first. However, the results are often worth the effort. At first glance, it might not seem that much different from Excel graphs. However, a closer look reveals many subtle differences that make for a more aesthetically pleasing graph. Try running the code below to see the difference. Make sure to export the graph to .pdf to make it look truly presentation-worthy!

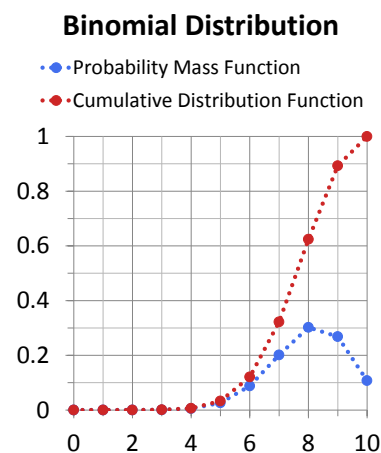


Figure 4.9: Excel graph of the Binomial Distribution's pmf and cdf ($n = 10$ and $p = 0.8$)

```

# Generate a graph for presentation or publication;
# The par function sets many different kinds of
# graphic parameters; Set the 4 margin sizes with
# the mar parameter
par(mar = c(5, 5, 4, 14) + 0.1)
# Make the plot; NA in the first position means
# that nothing will be plotted at first;
# xlim=c(0,10) means the x-axis limits are 0 and
# 10; ylim=c(0,1) means the y-axis limits are 0 and
# 1; main is the plot title; xlab is the x-axis
# title; ylab is the y-axis title; axes=FALSE means
# to not display the default axes; font.lab=2 means
# that the axis titles are bold
plot(NA, xlim = c(0, 10), ylim = c(0, 1), pch = 19,
     main = "Binomial Distribution", xlab = "Sample Space",
     ylab = "Probability", axes = FALSE, font.lab = 2)
# Make gridlines; v=seq(0,10) means make vertical
# lines at 0 through 10; h=seq(0,1,0.1) means make
# horizontal lines at 0 through 1 at 0.1 intervals;
# col='lightgray' means that the gridlines are
# light gray; lty=3 means that the gridlines have a
# dotted line type;
abline(v = seq(0, 10), col = "lightgray", lty = 3)
abline(h = seq(0, 1, 0.1), col = "lightgray", lty = 3)
# Add the probability mass function; type = 'b'
# means to plot both lines and points; lty=2 means
# line type is dashed; lwd=2 means line width is 2
# points; col='blue' means make the lines and
# points blue; pch=19 means that the points should
# be filled circles;
lines(x = SampleSpace, y = pmfBinomial, type = "b",
      lty = 2, lwd = 2, col = "blue", pch = 19)
# Add the cumulative distribution function; cex=0.7
# means make the dots at 70% the default size so
# that they do not completely cover the pmf series
lines(x = SampleSpace, y = cdfBinomial, type = "b",
      lty = 2, lwd = 2, col = "red", pch = 19, cex = 0.7)
# Add custom x-axis; at=SampleSpace means to label
# each point in the SampleSpace variable (0 to 10);
# cex.axis=0.8 means to size the axis labels at 80%
# of the default size;
axis(1, at = SampleSpace, cex.axis = 0.8)
# Add custom y-axis; at=seq(0,1,0.1) means to label

```

```

# each point from 0 to 1, at 0.1 intervals; las=1
# means to make the labels horizontal
axis(2, at = seq(0, 1, 0.1), las = 1, cex.axis = 0.8)
# xpd=NA allows for text to appear outside the plot
# region. Otherwise it is clipped.
par(xpd = NA)
# Add text at point (x,y); cex=0.8 means size the
# text at 80% of the default text size; adj=0 means
# left justify the text (0.5 means center and 1
# means right justify); The expression function
# allows for equations and text formatting
text(x = 10.4, y = dbinom(10, 10, 0.8), col = "blue",
      labels = "Probability Mass Function", adj = 0,
      cex = 0.8)
text(x = 10.4, y = pbinom(10, 10, 0.8), col = "red",
      labels = "Cumulative Distribution Function", adj = 0,
      cex = 0.8)
text(x = 5, y = 0.75, labels = expression(italic(n) ==
      10), cex = 0.8)
text(x = 5, y = 0.65, labels = expression(italic(p) ==
      0.8), cex = 0.8)
# Restore graphics parameter defaults
par(mar = c(5, 4, 4, 2) + 0.1, xpd = FALSE)

```

4.12 Poisson Distributions

Parameter:	λ	\in	\mathbb{R} $\lambda > 0$
Sample Space:	x	\in	\mathbb{N}_0
Mean:	μ	$=$	λ
Variance:	σ^2	$=$	λ
Skewness:	γ_1	$=$	$\frac{1}{\sqrt{\lambda}}$
Kurtosis:	γ_2	$=$	$\frac{1}{\lambda}$
Probability Mass Function:	$f_X(x; \lambda)$	$=$	$\frac{\lambda^x}{e^\lambda x!}$
Cumulative Distribution Function:	$F_X(x; \lambda)$	$=$	$\sum_{i=0}^x \frac{\lambda^i}{e^\lambda i!}$

If an event occurs at random, is equally likely to occur at any moment, and on average occurs a certain number of times per interval, then the probability distribution of the number of times that the event will occur during any particular interval will have a **Poisson distribution**. As seen in the box below, the Poisson distribution has a single

The **Poisson distribution** is a discrete distribution used to model how often an event will occur during a particular interval of time.

parameter λ , which is the mean (and, interestingly, also the variance).

A clinical application of the the Poisson distribution

Suppose that you begin treating an adult male client who has panic attacks that come at unpredictable times. Some weeks there are no panic attacks and some weeks there are many, but on average he has 2 panic attacks each week. The client knows this because he has kept detailed records in a spreadsheet for the last 5 years. The client had sought treatment once before, but terminated early and abruptly because, according to him, “It wasn’t working.” After sensitive querying, you discover that he expected that treatment should have quickly reduced the frequency of panic attacks to zero. When that did not happen, he became discouraged and stopped the treatment.

Because your client is well educated and quantitatively inclined, you decide to use the data he has collected as part of the intervention and to set a more realistic set of expectations.

You plot the frequency of how often he had 0 panic attacks in a week, 1 panic attack in a week, 2 panic attacks in a week, and so forth, as shown in blue in Figure 4.10. Because you have read this book, you immediately recognize that this is a Poisson distribution with $\lambda = 2$. When you graph an actual Poisson distribution and compare it with your client’s data, you see that it is almost a perfect match.⁷ Then you explain that although the goal is permanent cessation of the panic attacks, sometimes an intervention can be considered successful if the frequency of panic attacks is merely reduced. For example, suppose that in the early stages of treatment the frequency of panic attacks were reduced from twice per week to once every other week ($\lambda = 0.5$), on average. If such a reduction were achieved, there would still be weeks in which two or more panic attacks occur. According to Figure 4.10, this will occur about 9% of the time.

Excel, you can use the POISSON.DIST function to plot the Poisson probability mass function. For example, if the average number of events per time period is 2, then the probability that there will be 0 events is

`=POISSON.DIST(0,2,FALSE)`

Although the Poisson distribution extends to positive infinity, it often approaches zero probability fairly quickly. In this case, the client never had more than 7 panic attacks in a week. Thus, we need to repeat this calculation seven more times (i.e., `=POISSON.DIST(1,2,FALSE)`, `=POISSON.DIST(2,2,FALSE)`, ..., `=POISSON.DIST(7,2,FALSE)`) as seen in Figure 4.11. Veteran Excel users know shortcuts such that this process is not at all tedious and can be done in just a few seconds.⁸

⁷ Note that I am NOT claiming that all clients’ panic attack frequencies have this kind of distribution. It just so happens to apply in this instance.

⁸ I recommend the video tutorials by [Bill Jelen AKA Mr. Excel](#) to readers wishing to take their Excel skills to the next level.

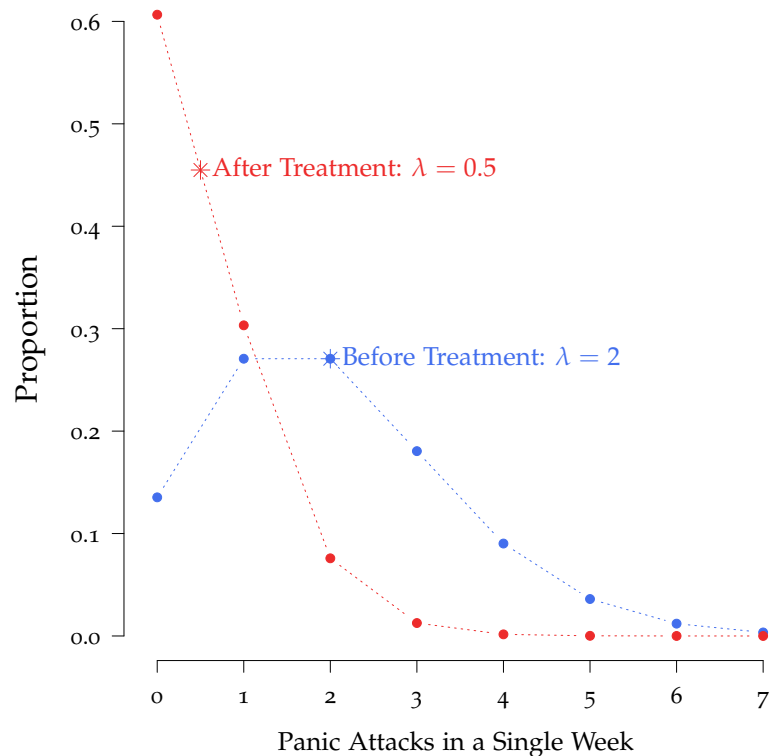


Figure 4.10: The variability of a hypothetical client's panic attack frequency

In R, we use the `dpois` function:

```
# Make a sequence of integers from 0 to 7
PanicAttacks <- seq(0, 7)

# Generate the probability mass function with
# lambda = 2
Probability <- dpois(PanicAttacks, 2)

# Bare-bones plot of the Poisson distribution's
# probability mass function
plot(Probability ~ PanicAttacks)
```

	A	B
1		Panic Attacks
2	0	0.135335283
3	1	0.270670566
4	2	0.270670566
5	3	0.180447044
6	4	0.090223522
7	5	0.036089409
8	6	0.012029803
9	7	=POISSON.DIST(A9,2,FALSE)

Figure 4.11: Using Excel to Calculate the Poisson Distribution

To calculate the cumulative distribution function of Poisson distribution in Excel, just change the last argument of the `POISSON.DIST` function to `TRUE`. For example, if we want to estimate the probability of having 4 panic attacks or more in a week if $\lambda = 2$, we must subtract the probability of having 3 panic attacks or less from 1, like so:

=1-POISSON.DIST(3,2)

Graphing the cumulative distribution function in R makes use of

the ppois function.

```
# Generate the cumulative distribution function
# with lambda = 2
CumulativeProbability <- ppois(PanicAttacks, 2)

# Bare-bones plot of the Poisson distribution's
# cumulative distribution function
plot(CumulativeProbability ~ PanicAttacks)
```

With some cosmetic changes and an additional series with $\lambda = 0.5$, the plot can look like Figure 4.12

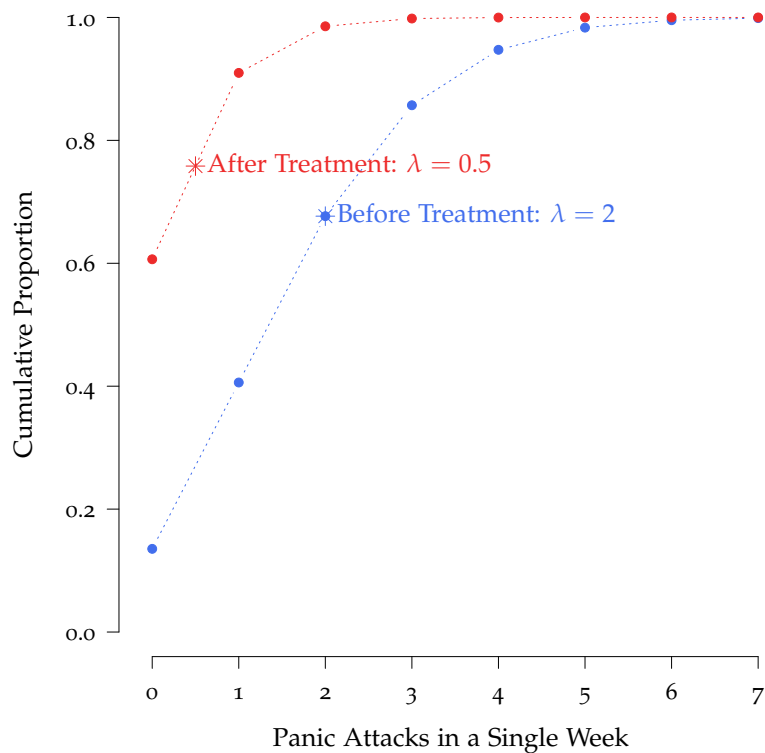


Figure 4.12: The cumulative distribution function of a hypothetical client's panic attack frequency

4.13 Geometric Distributions

Probability of success in each trial:	p	\in	$[0, 1]$
Sample Space:	x	\in	\mathbf{N}_1
Mean:	μ	$=$	$\frac{1}{p}$
Variance:	σ^2	$=$	$\frac{1-p}{p^2}$
Skewness:	γ_1	$=$	$\frac{2-p}{\sqrt{1-p}}$
Kurtosis:	γ_2	$=$	$6 + \frac{p^2}{1-p}$
Probability Mass Function:	$f_X(x; p)$	$=$	$(1-p)^{x-1}p^x$
Cumulative Distribution Function:	$F_X(x; p)$	$=$	$1 - (1-p)^x$

Atul Gawande (2007, p. 219–223) tells a marvelous anecdote about how a doctor used some statistics to help a young patient with cystic fibrosis to return to taking her medication more regularly. Because the story is masterfully told with perfect pacing, passion, and pathos, I will not repeat a clumsy version of it here. However, unlike Gawande, I *will* show how the doctor’s statistics were calculated.

According to the story, if a patient fails to take medication, the risk of a person with cystic fibrosis getting a bad lung illness on any particular day is 0.005. If medication is taken, the risk is 0.0005. Although these probabilities are both close to zero, over the the course of a year, they result in very different levels of risk. Off medication, the patient has about a 0.84% chance of getting sick within a year’s time. On medication, the patient’s risk falls to 0.17%. As seen in Figure 4.13, the cumulative risk over the course of 10 years is quite different. Without medication, the probability of becoming seriously ill within 10 years at least once is almost certain. With medication, however, a small but substantial percentage (~16%) of patients will go at least 10 years without becoming ill.

Such calculations make use of the **geometric distribution**. Consider a series of [Bernoulli trials](#) in which an event has a probability p of occurring on any particular trial. The probability mass function of the geometric distribution will tell us the probability that the x^{th} trial will be the first time the event occurs.

$$f_X(x; p) = (1 - p)^{x-1}p^x \quad (4.4)$$

Where

X = A random variable with a geometric distribution

f_X = The probability mass function of X

x = The number of Bernoulli trials on which the event first occurs

p = The probability of an event occurring on a single Bernoulli trial

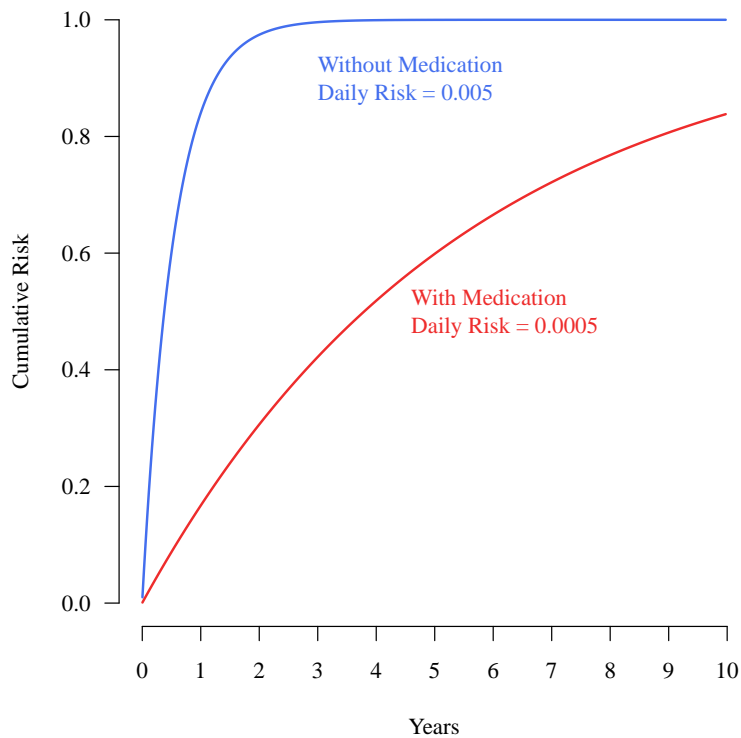


Figure 4.13: The cumulative risk of a getting sick with and without medication

Although Excel does not have specialized functions related to the geometric distribution, the probability mass function is easy to calculate. For example, if $p = 0.6$ and $x = 5$, then the probability that the first “success” will occur on the 5th trial is calculated like so:

$$=(1-0.6)^{(5-1)}(0.6^5)$$

In R, the probability mass function of the geometric distribution uses the `dgeom` function:

```
# Make a sequence of integers from 1 to 10
x <- seq(1, 10)

# Generate the probability mass function with p =
# 0.6
Probability <- dgeom(x, prob = 0.6)

# Bare-bones plot of the geometric distribution's
# probability mass function
plot(Probability ~ x)
```

The cumulative distribution function of the geometric distribution was used to create Figure 4.13. It tells us the probability that the event

will occur on the x^{th} trial or earlier:

$$F_X(x; p) = 1 - (1 - p)^x \quad (4.5)$$

Using Excel, we can calculate the probability that an event with a $p = 0.6$ probability will occur for the first time on the $x = 5^{\text{th}}$ trial or earlier like so:

=1-(1-0.6)^5

In R, the cumulative distribution function of the geometric distribution uses the `pgeom` function:

```
# Generate the cumulative distribution function
# with p = 0.6
CumulativeProbability <- pgeom(x, prob = 0.6)

# Bare-bones plot of the geometric distribution's
# cumulative distribution function
plot(CumulativeProbability ~ x)
```

4.14 Probability Density Functions

Although there are many more discrete distribution families, we will now consider some continuous distribution families. Most of what we have learned about discrete distributions applies to continuous distributions. However, there is a need of a name change for the probability mass function. In a discrete distribution, we can calculate an actual probability for a particular value in the sample space. In continuous distributions, doing so can be tricky. We can always calculate the probability that a score in a particular interval will occur. However, in continuous distributions, the intervals can become very small, approaching a width of 0. When that happens, the probability associated with that interval also approaches 0. Yet, some parts of the distribution are more probable than others. Therefore, we need a measure of probability that tells us the probability of a value *relative* to other values: the **probability density function**.

Considering the entire sample space of a discrete distribution, all of the associated probabilities from the probability mass function sum to 1. In a probability density function, it is the area under the curve that must sum to 1. That is, there is a 100% probability that a value generated by the random variable will be somewhere under the curve. There is nowhere else for it to go!

However, unlike probability mass functions, probability density functions do not generate probabilities. Remember, the probability of any value in the sample space of a continuous variable is infinitesimal. We

The **probability density function** (pdf) is function that can show relative likelihoods of sample space elements of a continuous random variable.

can only compare the probabilities to each other. To see this, compare the discrete uniform distribution and continuous uniform distribution in Figure 4.2. Both distributions range from 1 to 4. In the discrete distribution, there are 4 points, each with a probability of 0.25. It is easy to see that these probabilities sum to 1. Because of the scale of the figure, it is not easy to see exactly how high the probability density function is in the continuous distribution. It happens to be $1/3$. Why? First, it does not mean that each value has a $1/3$ probability. There are an infinite number of points between 1 and 4 and it would be absurd if each of them had a $1/3$ probability. The distance between 1 and 4 is 3. In order for the rectangle to have an area of 1, its height must be $1/3$. What does that $1/3$ mean, then? In the case of a single value in the sample space, it does not mean much at all. It is simply a value that we can compare to other values in the sample space. It could be scaled to any value but for the sake of convenience it is scaled such that the area under the curve is 1.

Note that some probability density functions can produce values greater than 1. If the range of a continuous uniform distribution is less than 1, at least some portions of the curve must be greater than 1 to make the area under the curve equal 1. For example, if the bounds of a continuous distribution are 0 and $1/3$, the average height of the probability density function would need to be 3 so that the total area is equal to 1.

4.15 Continuous Uniform Distributions

Lower Bound:	a	\in	\mathbb{R}
Upper Bound:	b	\in	\mathbb{R}
			$b > a$
Sample Space:	x	\in	$[a, b]$
Mean:	μ	$=$	$\frac{a+b}{2}$
Variance:	σ^2	$=$	$\frac{(b-a)^2-1}{12}$
Skewness:	γ_1	$=$	0
Kurtosis:	γ_2	$=$	$-\frac{6}{5}$
Probability Density Function:	$f_X(x; a, b)$	$=$	$\frac{1}{b-a}$
Cumulative Distribution Function:	$F_X(x; a, b)$	$=$	$\frac{x-a}{b-a}$

Notation notes:

\mathbb{R} is the set of all real numbers.

Unlike the [discrete uniform distribution](#), the uniform distribution is [continuous](#).⁹ In both distributions, there is an upper and lower bound and all members of the sample space are equally probable.

⁹ For the sake of clarity, the uniform distribution is often referred to as the *continuous uniform distribution*

Generating random samples from the continuous uniform distribution in Excel and R

In Excel, the RAND function generates a random number from the continuous uniform distribution with a lower bound of 0 and an upper bound of 1. To make the distribution have different lower and upper bounds, a and b , use this formula:

`=a+(b-a)*RAND()`

Thus, to generate a number from the continuous uniform distribution ranging from 1 to 99, use:

`=1+(99-1)*RAND()`

Using the continuous uniform distribution to generate random samples from other distributions

Uniform distributions can begin and end at any real number but one member of the uniform distribution family is particularly important. The cumulative distribution function of any continuous distribution converts into a continuous uniform distribution. A distribution's quantile function converts a continuous uniform distribution into that distribution. Most of the time, this process also works for discrete distributions. This process is particularly useful for generating random numbers with an unusual distribution. If the distribution's quantile function is known, a sample with a continuous uniform distribution can easily be generated and converted.

For example, the RAND function in Excel generates random numbers between 0 and 1 with a continuous uniform distribution. The BINOM.INV function is the binomial distribution's quantile function. Suppose that n (number of Bernoulli trials) is 5 and p (probability of success on each Bernoulli trial) is 0.6. A randomly generated number from the binomial distribution with $n = 5$ and $p = 0.6$ is generated like so:

`=BINOM.INV(5,0.6,RAND())`

Excel has quantile functions for many distributions (e.g., BETA.INV, BINOM.INV, CHISQ.INV, F.INV, GAMMA.INV, LOGNORM.INV, NORM.INV, T.INV). This method of combining RAND and a quantile function works reasonably well in Excel for quick-and-dirty projects but when high levels of accuracy are needed, random samples should be generated in a dedicated statistical package.

4.16 Normal Distributions

Sample Space:	x	\in	\mathbb{R}
Mean:	μ	$=$	$\frac{a+b}{2}$
Variance:	σ^2	$=$	$\frac{(b-a)^2-1}{12}$
Skewness:	γ_1	$=$	0
Kurtosis:	γ_2	$=$	$-\frac{6}{5}$
Probability Density Function:	$f_X(x; a, b)$	$=$	$\frac{1}{b-a}$
Cumulative Distribution Function:	$F_X(x; a, b)$	$=$	$\frac{x-a}{b-a}$

The normal distribution is in many ways the most important distribution in statistics and in psychological assessment. In the absence of other information, assuming that an individual difference variable is normally distributed is a good bet—not a sure bet, of course, but a good bet. Why? What is so special about the normal distribution? To get a sense of the answer to this question, consider what happens to the binomial distribution as the number of events (n) increases. To make the example more concrete, let's assume that we are tossing coins and counting the number of heads ($p = 0.5$). In Figure 4.14, the first plot shows the probability mass function for the number of heads when there is a single coin ($n = 1$). In the second plot, $n = 2$ coins. That is, if we flip 2 coins, there will be 0, 1, or 2 heads. In each subsequent plot, we double the number of coins that we flip simultaneously. Even with as few as 4 coins, the distribution begins to resemble the normal distribution, although the resemblance is very rough. With 128 coins, however, the resemblance is very close.

This resemblance to the normal distribution in the example is not coincidental to the fact that $p = 0.5$, making the binomial distribution symmetric. If p is extreme (close to 0 or 1), the binomial distribution is assymmetric. However, if n is large enough, the binomial distribution becomes very close to normal.

Many other distributions (e.g., Poisson, Student's T , F , and χ^2) have distinctive shapes under some conditions but approximate the normal distribution in others. Why? In the conditions in which non-normal distributions approximate the normal distribution, it is because, like in Figure 4.14, many independent events are summed.

In the distribution in Figure 4.7, because there is a 50% chance of heads, $p = 0.5$. Because there are two coins, $n = 2$. We can use the binomial distribution's probability mass function (Eq. 4.3) to see how the shape of the distribution changes as we increase the number of coins tossed (Figure 4.14).

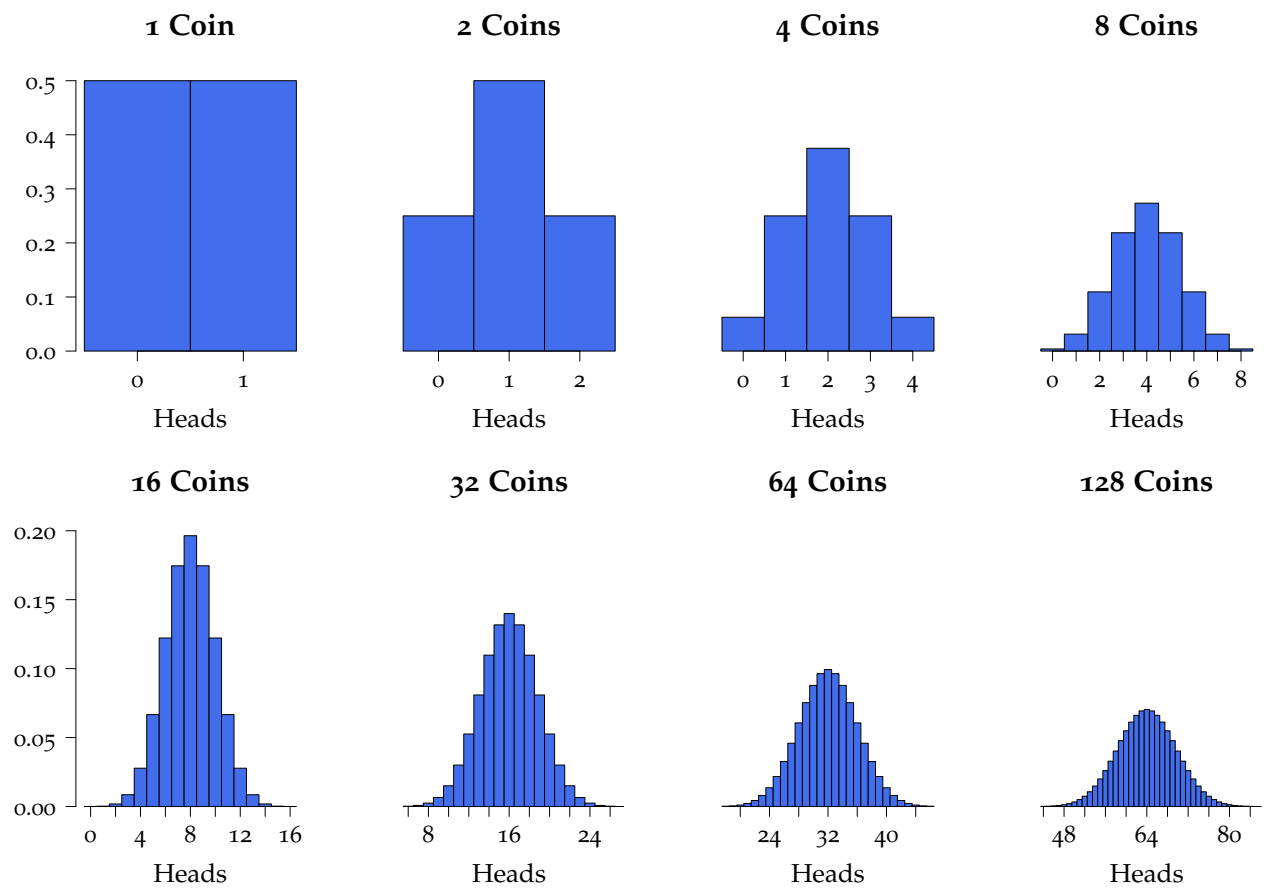


Figure 4.14: The binomial distribution begins to resemble the normal distribution when the number of events is large.

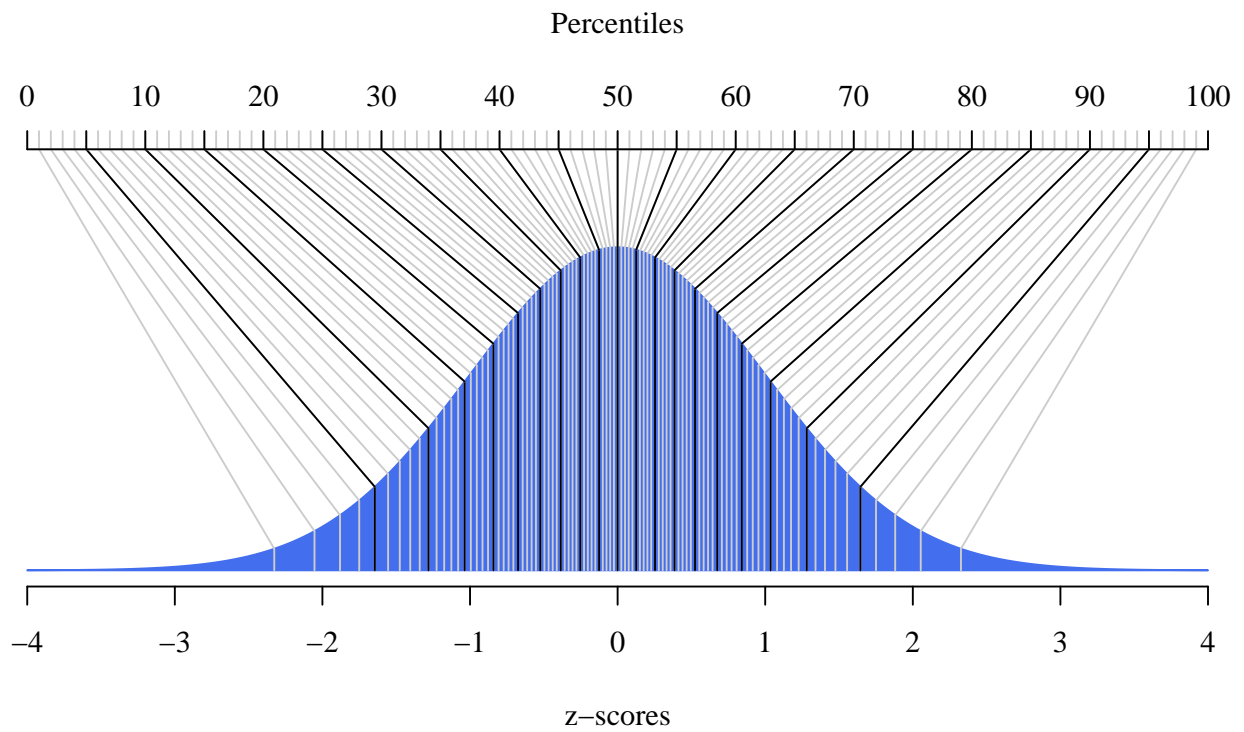


Figure 4.15: Percentiles convert a distribution into a uniform distribution

5

Describing Distributions with Expected Values & Moments

5.1 *Expected Values*

At one level, the concept of the **expected value** of a random variable is really simple; it is just the population mean of the variable. So why don't we just talk about population means and be done with this "expected value" business? It just complicates things! True. In this case, however, there is value in letting some simple things appear to become complicated for a while so that later we can show that some apparently complicated things are actually simple.

Why can't we just say that the expected value of a random variable is the population mean? You are familiar, of course, with the formula for a mean. You just add up the numbers and divide by the number of numbers n :

$$m_X = \frac{\sum_{i=1}^n x_i}{n}$$

Fine. Easy. Except...hmm...random variables generate an infinite number of numbers. Dividing by infinity is tricky. We'll have to approach this from a different angle...

The expected value of a random variable is a weighted mean. A mean of what? Everything in the sample space. How are the sample space elements weighted? Each element in the sample space is multiplied by its probability of occurring.

Suppose that the sample space of a random variable X is $\{2, 4, 8\}$ with respective probabilities of $\{0.3, 0.2, 0.5\}$, as shown in Figure 5.1. Each sample space element is multiplied by its probability and the resulting products are summed to calculate the expected value of X .

The **expected value** of a random variable is the population mean of the values that the random variable generates.

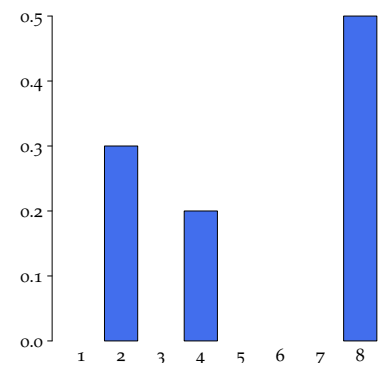


Figure 5.1: Probability distribution of a hypothetical random variable

The expected value operator is $E()$. Thus,

$$\begin{aligned} E(X) &= \sum_{i=1}^3 p_i x_i \\ &= p_1 x_1 + p_2 x_2 + p_3 x_3 \\ &= (0.3 \times 2) + (0.2 \times 4) + (0.5 \times 8) \\ &= 5.4 \end{aligned}$$

The term *expected value* is a little misleading. In this case, 5.4 is the expected value of X but X never once generates a value of 5.4. So the expected value is not “expected” in the sense that we expect to see it often (like we do with the **mode**). It is expected to be close to the mean of any sample of the variable that is sufficiently large:

$$E(X) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i$$

If a discrete random variable X with sample space S has a [probability mass function](#) $f_X(x)$, its expected value is

$$E(X) = \sum_{-\infty}^{\infty} x_i f_X(x_i) \quad (5.1)$$

With continuous variables we have a problem: the number of elements in a sample is infinite. Fortunately, calculus was designed to deal with this kind of infinity. The trick is to imagine that the continuous variable is sliced into bins and that the bins are sliced ever more thinly. If a continuous random variable has probability density function, the expected value is

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad (5.2)$$

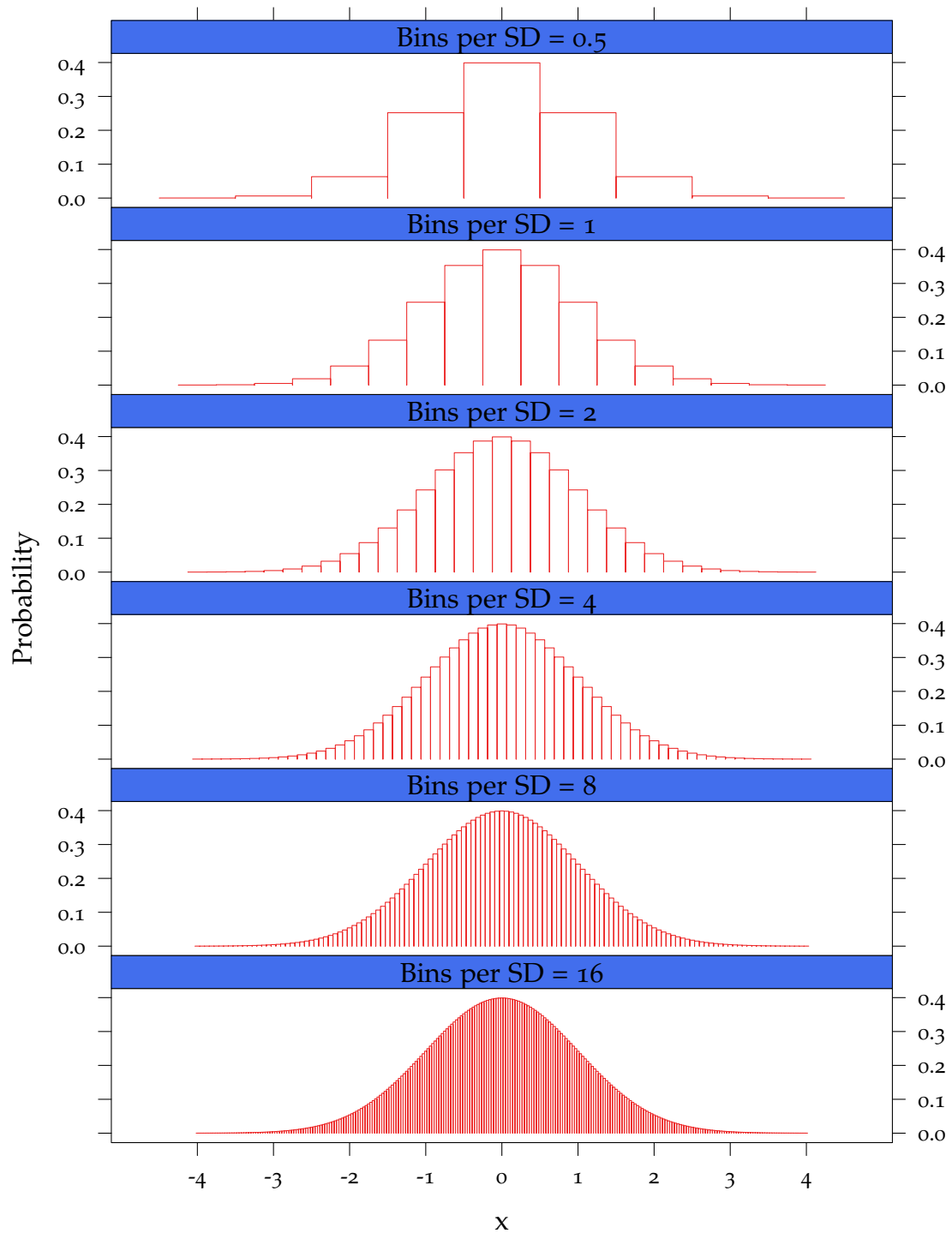


Figure 5.2: Slicing the standard normal distribution into ever thinner bins.

6

Composite Scores

Clinicians use **composite scores** whenever they can because they tend to be more reliable and valid than single scores. How much more reliable is a composite score than a single score? It, of course, depends. First, the reliability coefficients of all of the scores that make up the composite score matter. Second, the correlations among the components of the composite matter. The more reliable and correlated the components, the more reliable the composite.

As discussed previously, it is quite possible for different scores from the same test to have different reliabilities. Scores from the same test tend to have similar reliabilities, though it is quite possible for differences in reliability to be large. In a well designed test, such cases are rare. If we can assume that the test is well designed and that test score reliabilities are within a fairly narrow range, we can assume that the classical reliability coefficient is a good estimate of the reliability of all scores from a particular test. To the degree that such assumptions are incorrect, the following equations will be less accurate.

Before we can discuss the reliability of composites, we must cover more basic statistics as they apply to composite scores. I will present this material in three ways. First, I will show the equations in standard notation. Adjacent to this, I will show the same equations in matrix notation. Then I will walk through the calculations using R.

6.1 *The Mean of the Sum Is the Sum of the Means*

This section is going to make a simple idea look complicated. If you get lost, this is what I am trying to say: If we create a new random variable by adding a bunch of random variables together, the mean of that new variable is found by adding together the means of all the variables we started with. I'm sorry for what comes next but the work we put into it will pay off later. Okay, now let's make that simple idea formal and hard to understand:

We can start with the example of two variables: X_1 and X_2 . The

A **composite variable** is the (possibly weighted) sum of two or more variables.

Why bother with matrices? It does seem to be an unnecessary complication at first. However, many things become clearer and simpler with matrices, which I hope to illustrate. Furthermore, R was designed to work elegantly with vectors and matrices.

sum of these variables, which we will call X_S , has a mean, which is the sum of the means of X_1 and X_2 . That is, $\mu_S = \mu_1 + \mu_2$.

What if we have three variables? or four? or five? It would be tedious to illustrate each case one-at-a-time. We need a way of talking about the sum of a bunch of random variables but without worrying about how many variables there are. Here we go:

Calculating a Sum

So, k is a positive integer ($k \in \mathbb{N}_1$), presumably greater than 1. So if there are k random variables, the notation for the set of all them is $\{X_1, \dots, X_k\}$. However, it is even more compact to use matrix notation such that $\mathbf{X} = \{X_1, \dots, X_k\}$.

Now, \mathbf{X} is a set of random variables in their entirety, without referencing any particular values those variables might generate. A set of particular values of these variables would be shown as \mathbf{x} or $\{x_1, \dots, x_k\}$. In regular notation, the sum of these particular values would be:

$$x_S = \sum_{i=1}^k x_i \quad (6.1)$$

Where

k = The number of variables in $\{X_1, \dots, X_k\}$, ($k \in \mathbb{N}_1$)

x_S = The sum of all k scores in $\{x_1, \dots, x_k\}$

x_i = A particular score generated by variable X_i

In matrix notation, things are generally more compact:

$$x_S = \mathbf{1}'\mathbf{x}$$

Where

\mathbf{x} = A $k \times 1$ vector of scores $\{x_1, \dots, x_k\}$

$\mathbf{1}$ = A $k \times 1$ vector of ones

The $\mathbf{1}$ may be a bit confusing. It is a column of ones that has the same length (number of elements) as \mathbf{x} . Therefore, $\mathbf{1}'$ is **$\mathbf{1}$ transposed**, which is a row vector. A row vector multiplied by a column vector is the sum of the product of each analogous element in the pair of vectors. Suppose that \mathbf{x} has a length of three. Thus,

$$\mathbf{1}'\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}' \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 1 \times x_1 + 1 \times x_2 + 1 \times x_3 = x_S$$

Let's do some calculations in R with a particular example. Suppose that there are three variables: $\mathbf{X} = \{X_1, X_2, X_3\}$. In R, we will create a vector of the names of variables in \mathbf{X} :

To **transpose** means to make all columns of a matrix into rows (or all rows into columns). Transposition is noted with a prime symbol ($'$). If

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

then

$$\mathbf{A}' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

The `c` function combines numbers (or other objects) into a vector.

Now suppose that there are three particular scores: $x = \{100, 120, 118\}$. The sum of these three scores can be calculated in a variety of ways. Here is the easiest:

However, if we want to be matrix algebra masochists (and, apparently, at least one of us does!), we must first create a vector of ones that is the same length as x . If all we cared about was this particular example, we could do so by hard coding a vector of three ones: `c(1,1,1)`. A better shortcut is to use the `rep` function, which creates a vector with the same element repeated a certain number of times. So, `rep(1,3)` creates a vector of three ones.

However, we want to use code that will be conveniently applied to a vector that is arbitrarily long. It would be no fun to have to count the vector length each time we run the code. The `length` function tells us how many numbers are in a vector. That is, `length(x)` gives us the value of k , the number of variables. Using the `rep` and `length` functions together:

Premultiplying any column vector by a row of ones produces the sum of the vector. You can think of any vector in R as a column vector. To make a vector into a row vector, you transpose it with the `t` function. Matrix multiplication uses the matrix multiply operator `%*%` like so: Either way that we calculate it, $x_S = 348$.

Calculating the Mean of a Sum

The mean of X_S is:

$$\mu_S = \sum_{i=1}^k \mu_i = \mathbf{1}'\boldsymbol{\mu} \quad (6.2)$$

Where

μ_S = The mean of X_S

μ_i = The mean of X_i

$\boldsymbol{\mu}$ = A $k \times 1$ vector of means of the variables in X

Suppose that the means of X_1 , X_2 , and X_3 are all 100. Again, the mean of X_S (μ_S) can be calculated in two ways: Running this code, we can see that $\mu_S = 300$.

Calculating the Mean of a Weighted Sum

The mean of a weighted sum is the weighted sum of the means. That is, if

$$x_S = \sum_{i=1}^k w_i x_i = \mathbf{w}'\mathbf{x} \quad (6.3)$$

The `names` function can return or set the names of the elements of a vector. The `rep` function creates vectors of repeated elements. For example, `rep(5,3)` is the same as `c(5,5,5)`.

The `length` function returns the number of elements in a vector. For example, `length(c(3,3))` returns 2.

The `t` function transposes a vector or matrix.

To multiply matrices A and B in R (assuming they are compatible), the operator `%*%` is used. AB in R is `A %*% B`.

Where

w_i = The weight for X_i

w = A $k \times 1$ vector of weights for each of the variables in X

then

$$\mu_S = \sum_{i=1}^k w_i \mu_i = w' \mu \quad (6.4)$$

Suppose that $w = \{0.5, 1, 2\}$. That is, the weight for X_1 is 0.5, the weight for X_2 is 1, and the weight for X_3 is 2. We will continue to use the same values for x and μ as before:

Running the code shows that $x_S = 411$ and that $\mu_S = 350$.

6.2 The Variance of the Sum Is the Sum of the Covariance Matrix

If variables are uncorrelated, the variance of their sum is the sum of their variances. However, this is not true when variables are substantially correlated. The formula for the variance of a sum looks more complex than it is. It is just the sum of the covariance matrix.

$$\sigma_{X_S}^2 = \sum_{i=1}^k \sum_{j=1}^k \sigma_{ij} = \mathbf{1}' \Sigma \mathbf{1} \quad (6.5)$$

Where

$\sigma_{X_S}^2$ = The variance of X_S

σ_{ij} = The covariance between X_i and X_j ($\sigma_{ij} = \sigma_i^2$ if $i = j$)

Σ = The $k \times k$ covariance matrix of all the variables in X

Suppose that the standard deviations of X_1 , X_2 , and X_3 are all 15. Thus, $\sigma = \{15, 15, 15\}$. The correlations among the three variables are shown in matrix R :

$$R = \begin{bmatrix} 1 & 0.5 & 0.6 \\ 0.5 & 1 & 0.7 \\ 0.6 & 0.7 & 1 \end{bmatrix}$$

The covariance matrix Σ can be computed from the correlation matrix and the standard deviations like so:

$$\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$$

Where

σ_{ij} = The covariance between X_i and X_j

ρ_{ij} = The correlation between X_i and X_j

σ_i = The standard deviation of variable X_i

σ_j = The standard deviation of variable X_j

Note that the calculation of X_S and μ_S with matrix algebra is the same as it was with an equally weighted sum except that instead of premultiplying with a transposed vector of ones, we premultiply with a transposed vector of weights. In truth, an equally weighted sum is a special case of a weighted in which w consists entirely of ones.

In R, the multiplication operator (*) multiplies analogous elements of vectors and matrices. In the example, $w * s$ is $\{0.5 * 15, 1 * 15, 2 * 15\}$

Unfortunately, the notation for a covariance matrix is a bold capital sigma Σ , which is easily confused with the summation symbol, which is generally larger and not bold: Σ .

The symbol for a sample correlation is the Roman lowercase r and a matrix of such correlations is an uppercase R . Therefore, the population correlation coefficient is a Greek lowercase rho: ρ . This, unfortunately means that a matrix of correlations should be an uppercase rho: P . Somehow, statisticians are okay with ρ looking a lot like an italicized Roman letter p . However, using an uppercase rho for a correlation matrix is too crazy even for statisticians! You know, hobgoblins of little

Unfortunately, computing a covariance matrix like this in a computer program is inelegant because we have to make use of looping:

To calculate the covariance matrix in matrix notation, we can make use of the `diag` function. The `diag` function has two jobs: to turn a matrix into a vector and a vector into a matrix. Specifically, it takes the diagonal of a matrix and turns it into a vector. That is, if

$$A_{k,k} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \cdots & a_{k,k} \end{bmatrix}$$

Then $\text{diag}(A) = \{a_{1,1}, a_{2,2}, \dots, a_{k,k}\}$.

It also replaces the diagonal of a $k \times k$ matrix of zeros with a $k \times 1$ vector, like so:

$$\text{diag}(a) = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_k \end{bmatrix}$$

This is useful for calculating the covariance matrix (among many other statistics):

$$\begin{aligned} \Sigma &= \text{diag}(\sigma) R \text{diag}(\sigma) \\ &= \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{bmatrix} \begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,k} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k,1} & \rho_{k,2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho_{1,2} & \cdots & \sigma_1 \sigma_k \rho_{1,k} \\ \sigma_2 \sigma_1 \rho_{2,1} & \sigma_2^2 & \cdots & \sigma_2 \sigma_k \rho_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_k \sigma_1 \rho_{k,1} & \sigma_k \sigma_2 \rho_{k,2} & \cdots & \sigma_k^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,k} \\ \sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k,1} & \sigma_{k,2} & \cdots & \sigma_k^2 \end{bmatrix} \end{aligned}$$

Where

- Σ = The covariance matrix of variables in X
- σ = The standard deviations of variables in X
- R = The correlation matrix of variables in X

So this is where matrix algebra starts to shine. We do not need to initialize the variable that contains the covariance matrix, or calculate

k , or do any looping. R will even automatically apply the names from the correlation matrix to the covariance matrix.

Beautiful!

Running the code, we get:

$$\Sigma = \begin{bmatrix} 225 & 112.5 & 135 \\ 112.5 & 225 & 157.5 \\ 135 & 157.5 & 225 \end{bmatrix}$$

To calculate the variance of X_S if it is an unweighted sum we apply Equation 6.5:

Here we see that $\sigma_X^2 = 1485$

6.3 Calculating the Variance of a Weighted Sum

If X_S is a weighted sum, its variance is the weighted sum of the covariance matrix.

$$\sigma_S^2 = \sum_{i=1}^k \sum_{j=1}^k w_i w_j \sigma_{ij} = \mathbf{w}' \Sigma \mathbf{w} \quad (6.6)$$

Continuing with the same variables in our example, we see that things are starting to become clumsy and ugly without matrix algebra:

All three methods give the same answer: $\sigma_X^2 = 2193.75$

6.4 Calculating a Composite Score

We now have all of the information needed to make a composite score.

$$c = \frac{x_S - \mu_S}{\sigma_{X_S}} \sigma_C + \mu_C \quad (6.7)$$

Where

c = The particular score from composite variable C

x_S = The sum of all k scores in $\{x_1, \dots, x_k\}$

σ_{X_S} = The standard deviation of X_S

σ_C = The standard deviation of C

μ_C = The mean of C

Continuing with our example, suppose that $\mu_C = 100$ and $\sigma_C = 15$.

In clinical practice, the most common kind of composite score is an equally weighted composite consisting of scores with the same mean and standard deviation. Sometimes scores are weighted by the degree to which they correlate with the construct the composite is intended to measure. Other weighting schemes are also possible. In most cases, it makes sense to first convert all scores to z-scores and then weight the z-scores. Failing to convert the scores to a common metric such

as z-scores will result in an implicit weighting by standard deviations. That is, the score with the largest standard deviation will have the most weight in the composite score. Converting to z-scores first will equalize the each score's influence on the composite score. We can convert all the scores in x to z-scores like so:

$$z = (x - \mu) / \sigma$$

The nice thing about z-scores is that their means are zeros, their standard deviations are ones, and their covariances are correlations. Thus, the formula for a weighted composite score consisting of z-scores is fairly simple, especially if the composite score is a z-score itself:

$$c = \frac{w'z}{\sqrt{w'Rw}}$$

Continuing with our example, computing a composite score with $X_1 = 110$, $X_2 = 120$, and $X_3 = 118$:

Here we see, that the weighted composite score $c = 120$. If we had wanted an equally weighted composite score, the w vector would be set to equal ones. If we had done so, c would have been 117 instead of 120.

6.5 The reliability coefficient of a composite

Calculating the reliability coefficient of a composite score is much easier than it might appear at first. Remember that reliability is the ratio of a score's true score variance to its total variance:

$$r_{XX} = \frac{\sigma_T^2}{\sigma_X^2}$$

We know what the total variance of a composite score is: the sum of the covariance matrix of the composite's component scores. The variance of the composite's true scores is also the the sum of the covariance matrix of the component scores except that the diagonal of the covariance matrix is multiplied by the reliability coefficients of each of the component scores like so:

$$\sigma_T^2 = \mathbf{1}' \begin{bmatrix} r_{11}\sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,k} \\ \sigma_{2,1} & r_{22}\sigma_2^2 & \cdots & \sigma_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k,1} & \sigma_{k,2} & \cdots & r_{kk}\sigma_k^2 \end{bmatrix} \mathbf{1}$$

If element-wise division seems inelegant or somehow undignified, this little bit of trickery also works but seems like overkill to me:

$$z = \text{diag}(\text{diag}(\sigma)^{-1}(x - \mu))$$

The round function rounds to the nearest integer by default but you can round to any number of digits you wish. For example, rounding to 2 significant digits (i.e., the hundredths place), would be $\text{round}(x, 2)$.

6.6 Composite Scores and Their Correlations

If the population correlations among all of the components are known, it is possible to calculate the correlations among composite scores made from these components. Suppose that Composite A is calculated from the sum of two component tests, A_1 and A_2 . Composite B is calculated from the sum of B_1 and B_2 . Suppose that the correlation matrix for the four components is:

	A_1	A_2	B_1	B_2
A_1	1.0	0.3	0.35	0.40
A_2	0.3	1.0	0.42	0.48
B_1			1.00	0.56
B_2			0.56	1.00

The correlation between Composite A and Composite B is calculated by adding up the numbers in all three shaded regions of the correlation matrix and then dividing the sum of “between” correlations in purple by the geometric mean of the sums from the “within” correlations in the blue and red regions:

$$r_{AB} = \frac{\text{Sum of Correlations Between } A \text{ and } B}{\sqrt{\text{Sum of Correlations Within } A \times \text{Sum of Correlations Within } B}}$$

7

Bibliography

Borsboom, D. and Mellenbergh, G. J. (2004). Why psychometrics is not pathological: A comment on Michell. *Theory & Psychology*, 14(1):105–120.

Cicchetti, D., Bronen, R., Spencer, S., Haut, S., Berg, A., Oliver, P., and Tyrer, P. (2006). Rating scales, scales of measurement, issues of reliability: Resolving some critical issues for clinicians and researchers. *The Journal of nervous and mental disease*, 194(8):557–564.

Clark, A. (2004). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. Oxford University Press, USA, New York.

Embretson, S. E. and Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum, Mahwah, NJ.

Gawande, A. (2007). *Better: A Surgeon's Notes on Performance*. Metropolitan Books, New York, NY.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Elsevier Science Limited, Amsterdam.

Ley, P. (1972). *Quantitative aspects of psychological assessment: An introduction*. Duckworth, London.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88(3):355–383.

Michell, J. (2008). Is psychometrics pathological science? *Measurement*, 6(1-2):7–24.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684):677–680.

Zacharias, J. R. (1975). The trouble with IQ tests. *National Elementary Principal*, 54(4):23–29.