Z-SCORE

How a Statistic Used in Psychology Will Revolutionize Baseball

John G. Cottone, PhD
&
Jason Wirchin

Foreword by David Lennon, National Baseball Writer, *Newsday*

Story Bridge Books
Tucson, AZ

Z-score: How a Statistic Used in Psychology Will Revolutionize Baseball

Published by Story Bridge Books
Tucson, AZ

ISBN: 978-0-9903939-3-1
Library of Congress Control Number: 2015934569

Written by
John G. Cottone, PhD
&
Jason Wirchin

Edited by
Laura Markowitz

Foreword by
David Lennon

Cover Design & Artwork by
Rook Designs:
John Tsopanarias, Jr & Lisa A. Cottone, PhD

CHAPTER 1

WHAT THE *f* IS A Z-SCORE?

"Billy [Beane] was forever telling Paul [DePodesta] that when you try to explain probability theory to baseball guys, you just end up confusing them."

- Michael Lewis, Author,
Moneyball

Note: While an appreciation of the mathematics presented in this chapter is helpful for a deeper understanding of z-scores, for those who may not be interested, we have provided a simplified summary at the end of this chapter highlighting "the least you need to know" about z-scores.

Mr. Hardnose and Mr. Goodgrade

Imagine for a moment that you are back in high school, competing with your rival, Jack Silverspoon, for valedictorian. It's the last week of school and the final class rankings have just been posted. Your heart drops as you check the list and notice that you've been ranked second in the class, just behind Jack. Jack stole your girlfriend Heather at the Senior Prom, showing up in a new convertible that his parents got him, and now he just beat you out for valedictorian.

Dejected, you look for answers. In your search you discover that what nudged Jack ahead of you was that he finished chemistry with a final grade of **99**, while you only got an **83**. You think back to chemistry and how difficult it was, and how it seemed that your teacher, Mr. Hardnose, believed that for every 'A' he gave out a year would be taken off his life. How could Jack have gotten a **99** in chemistry while you only got an **83**? Everyone knows you are smarter than Jack, you work harder, and your mother is a freakin' chemistry professor! Despite your disdain for Jack, you know he's an honest guy, so it's not likely that he cheated.

You reflect a little longer and then it hits you: Jack didn't have to take chemistry with Mr. Hardnose, he had Mr. Goodgrade – a guy who decided to *mail it in* that year because he was about to retire to Arizona. He couldn't be bothered with actually grading test papers, so he just handed out 'A's like the military hands out uniforms. In fact, Mr. Goodgrade gave almost everyone in his class an 'A+' just for showing up (except, of course, for the kids who taped a "kick me" sign to his back on the first day of school). How unfair: you had to take chemistry with Mr. Hardnose, where the average grade in the class was a **68**, while Jack got to slide by with Mr. Goodgrade, where the average grade was a **93**.

You think to yourself that if there was justice in the world, your school would have taken into account that your chemistry class was much harder than Jack's and weighed your scores accordingly. This would have validated the fact that the **83** you got in Mr. Hardnose's class, which was the second highest grade in the class, was actually more impressive than the **99** that Jack got in Mr. Goodgrade's class. But how can you prove this? What evidence do you have? Enter the z-score!

But what the *f* is a z-score?

Z-scores are a type of standardized statistic that describes an individual's performance, *relative* to the performance of others in the same group or **distribution** of interest. When scores from any distribution are converted to z-scores, this process automatically controls for factors that might influence the whole distribution (like Mr. Goodgrade deciding to give almost everyone in the class an 'A+,' just for showing up) and distort these comparisons. This is because inherent within every z-score is information about two important components of the distribution from which it came: the **mean,**[1] or average score of the distribution (e.g., the average grade in Mr. Goodgrade's chemistry class) and the **standard deviation** (SD) of the distribution – a statistic that describes how far apart the scores of a distribution are from the mean. As illustrated below, z-scores are computed by subtracting the mean of a distribution from a particular score, and then dividing by the standard deviation of the data set.[2]

$$\textbf{Z-score} = \frac{(Score\ X - Mean)}{Standard\ Deviation}$$

[1] The mean is calculated by adding all scores in a distribution and then dividing by the number of scores in the distribution (*N*). The formula is:

$$\text{Mean} = \frac{(Sum\ of\ All\ Scores)}{N}$$

[2] Lockhart, R.S. (1998). *Introduction to Statistics and Data Analysis for the Behavioral Sciences*. W.H. Freeman, New York.

Using the equation above, the computation of a z-score starts with the specific score you are looking to convert – for example, the final grade you received in Mr. Hardnose's class, which was an **83**. From that you subtract the mean (i.e., the class average), which you recall from Mr. Hardnose's class was **68**. Now, this is the easy part: you subtract **68** from **83** and get **15** in the numerator of the fraction.

$$\textbf{Z-score} = \frac{83 - 68}{Standard\ Deviation}$$

$$\textbf{Z-score} = \frac{15}{Standard\ Deviation}$$

The denominator, however, which involves the computation of the standard deviation, is a bit more challenging to compute and explain. As noted above, the standard deviation of a distribution refers to how far apart scores are in a distribution relative to the mean. The formula for the standard deviation is a bit complicated, and not necessary to comprehend to understand the meaning of the z-scores in this book; however, we have provided it below for those who may be interested.

[3]**Standard Deviation (SD)** $= \sqrt{\frac{\Sigma\ (Score\ X - Mean)^2}{N}}$

---

[3]Presented here is the formula to compute the standard deviation of an entire population, rather than a specific sample. We presented this version of the formula here to help readers more easily calculate the standard deviation in the example on the following page. However, for our analyses, we used the formula for the standard deviation that corresponds to the standard deviation of a sample:

$$\sqrt{\frac{\Sigma\ (Score\ X - Mean)^2}{N-1}}$$

Though a comprehensive discussion of the standard deviation is unnecessary for the purposes of this book, one thing about the standard deviation that is potentially helpful to know to better understand z-scores is that two distributions can have the same mean, but very different standard deviations. For example, imagine that two classes with 10 students – Class A and Class B – take the same spelling test, as illustrated in Table 1.1 below. In Class A, five students get a score of 100 on the test, while the other five students get a score of 20. In Class B, however, five students get a score of 70, while the other five students get a score of 50. Interestingly, what you will notice in Table 1.1 is that while the average test score in both classes is 60, the standard deviation for the two classes is very different: for Class A it is 40, while for Class B it is 10!

TABLE 1.1: SPELLING TEST SCORES FOR CLASS A & CLASS B

CLASS A	**CLASS B**
100	70
100	70
100	70
100	70
100	70
20	50
20	50
20	50
20	50
20	50
Mean (Class A) = 60	***Mean (Class B) = 60***
Standard Deviation (Class A) = 40	***Standard Deviation (Class B) = 10***

NOTE: The actual computation of the standard deviation for Class A and Class B is presented in Appendix D.

As shown in Table 1.1 (and Appendix D), the scores in Class A are more widely dispersed (ranging from 20 to 100) than the scores in Class B (which merely range from 50 to 70), which is why the standard deviation for Class A is much greater than the standard deviation for Class B. This type of information, captured by the standard deviation, is extremely important to statisticians and researchers because it provides the context necessary to better understand the data they interpret. As it relates to z-scores, since the formula used to compute them is a fraction with the standard deviation as the denominator, the larger the standard deviation, the smaller the z-score (assuming all else is equal), and vice versa.

Getting back to you and Jack… now that we've discussed all of the components comprising a z-score, we can demonstrate how your **83** in Mr. Hardnose's class was actually more impressive than Jack's **99** in Mr. Goodgrade's class.

As shown in the computations of Tables 1.2 and 1.3 on the following page, while the **83** you got in Mr. Hardnose's class yielded you a solid z-score of **1.26**, Jack's final grade of **99** yielded him a paltry z-score of just **0.49**! Why? Mostly because the z-score for Jack's final grade (**99**), took into account that the average grade in Mr. Goodgrade's class was a **93,** while the z-score for your **83** took into account that the average grade in Mr. Hardnose's class was a **68**. Hence, *your* grade was much better relative to *your* classmates than was Jack's relative to *his* classmates. Your grade was 15 points higher than your class average, while Jack's grade was only 6 points higher than his class average.[4]

---

[4] Note: In this example, the standard deviation (or *spread* of the scores) was roughly the same for the two classes (11.88 vs. 12.17) and as such, had less impact on the z-scores than did the differences in the class means.

TABLE 1.2: FINAL GRADES – MR. HARDNOSE'S CHEM. CLASS

CLASS MEAN = 68
CLASS STANDARD DEVIATION (SD) = 11.88

STUDENT	FINAL GRADE	Z-SCORE = (SCORE X – MEAN)/SD	Z-SCORE
Margaret	95	(95-68)/11.88	2.27
***** YOU *****	**83**	**(83-68)/11.88**	**1.26**
Joe	75	(75-68)/11.88	0.59
Jane	68	(68-68)/11.88	0
Rachel	62	(62-68)/11.88	-0.51
Chris	60	(60-68)/11.88	-0.67
Melissa	60	(60-68)/11.88	-0.67
Doug	60	(60-68)/11.88	-0.67
Randy	59	(59-68)/11.88	-0.76
Alyssa	58	(58-68)/11.88	-0.84

TABLE 1.3: FINAL GRADES – MR. GOODGRADE'S CHEM. CLASS

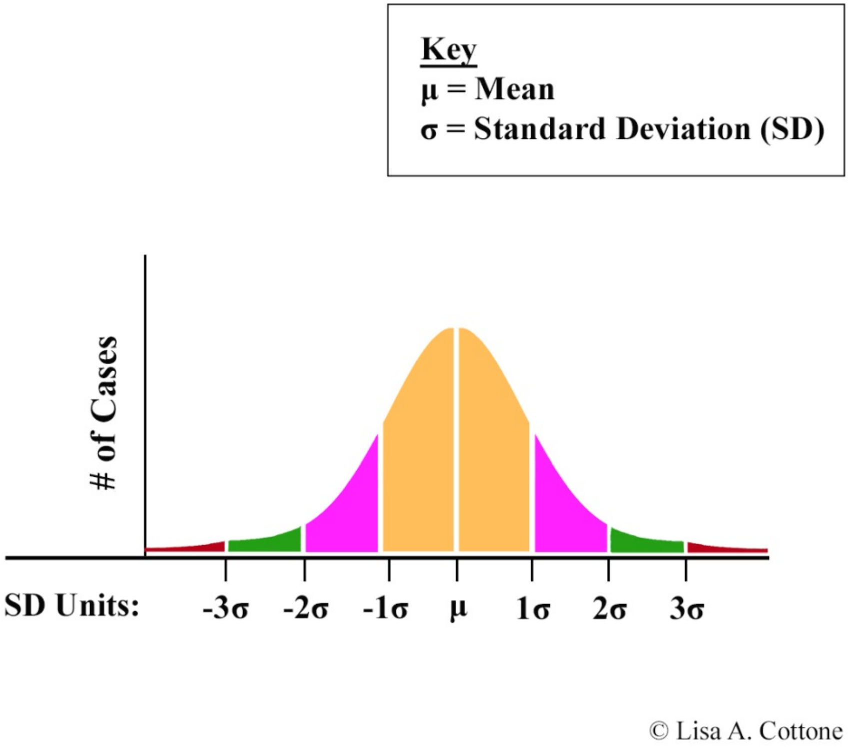
CLASS MEAN = 93
CLASS STANDARD DEVIATION (SD) = 12.17

STUDENT	FINAL GRADE	Z-SCORE = (SCORE X – MEAN)/SD	Z-SCORE
Julian	100	(100-93)/12.17	0.58
Francesca	100	(100-93)/12.17	0.58
Alison	100	(100-93)/12.17	0.58
Evelyn	100	(100-93)/12.17	0.58
Sarah	100	(100-93)/12.17	0.58
Joshua	100	(100-93)/12.17	0.58
***** JACK *****	**99**	**(99-93)/12.17**	**0.49**
Norm	93	(93-93)/12.17	0
Robbie	69	(69-93)/12.17	-1.98
Barbara	69	(69-93)/12.17	-1.98

Grading On a Curve

After any difficult exam, the most welcome words a student can hear from the teacher is: “This exam is being graded on a curve.” Hallelujah! But, what exactly does this mean? Well, when statisticians or teachers talk about “the curve,” what they are referring to is the **“bell curve**” or the curve of normally distributed scores (see Figure 1.1).

FIGURE 1.1: THE BELL CURVE (I.E., THE CURVE OF NORMALLY DISTRIBUTED SCORES)



The bell curve is a graphical representation of all large, hypothetical distributions of data. In theory, the tails on the left and right sides of the curve extend outward to infinity, as the theory from which the bell curve was derived assumes that all distributions are infinitely large. Interestingly, most large distributions – from human height, to the IQ score of 12-year olds in California, to the lifespan of dung beetles – tend to conform to a bell-shaped curve, with the largest percentage of scores near the middle of the distribution around the mean (symbolized in this graph by the Greek letter, μ, which is called Mu and pronounced "myoo"), and with diminishing percentages of scores extending outward from the middle (i.e., both above and below the mean).

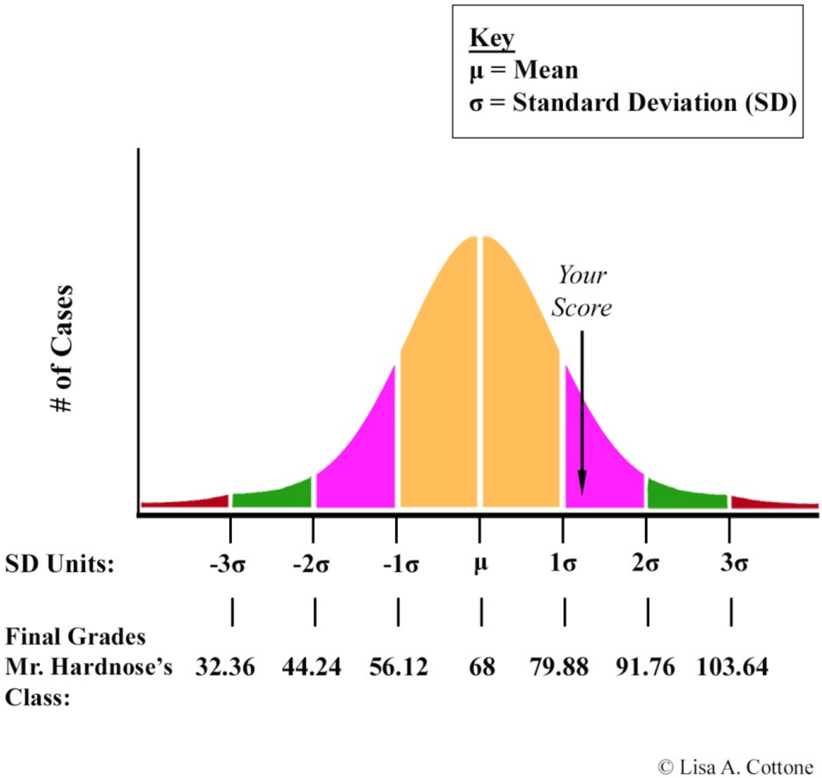
As shown in Figure 1.1 (on the previous page) and Figure 1.2 (on the following page), extending outward from the mean (μ), there are three negative values on the left (-1σ, -2σ and -3σ) and three positive values on the right (1σ, 2σ and 3σ). These values refer to the standard deviation units of the distribution, and the Greek letter σ (which is called sigma) is used to represent a standard deviation unit.

Specifically, 1σ (i.e., 1 sigma) refers to a value on the graph that is equal to the mean plus 1 standard deviation unit. In the example with Mr. Hardnose's class, where the mean was **68** and the standard deviation was **11.88**, the point on the graph corresponding to 1σ would be equal to a raw score of **79.88** (i.e., **68** + **11.88**). This is illustrated in Figure 1.2.

Similarly, using the data from Mr. Hardnose's class, 2σ would be equal to the mean plus 2 standard deviation units (i.e., **68** + **11.88** + **11.88**) which sums to **91.76**; and 3σ would be equal to the mean plus 3 standard deviation units (i.e., **68** + **11.88** + **11.88** + **11.88**) which sums to **103.64**.

On the left side of the graph, -1σ refers to a value that is equal to the mean minus 1 standard deviation unit. Hence, the point on the graph corresponding to -1σ would be equal to **56.12** (i.e., **68 - 11.88**). Going down the line, -2σ would be equal to the mean minus 2 standard deviation units (i.e., **68 – 11.88 – 11.88**) yielding a value of **44.24**; and -3σ would be equal to the mean minus 3 standard deviation units (i.e., **68 - 11.88 - 11.88 - 11.88**) yielding a value of **32.36**.

FIGURE 1.2: FINAL GRADES FOR MR. HARDNOSE'S CLASS, PLOTTED ON THE BELL CURVE



As noted above, the theory from which the bell curve was derived assumes that all distributions are infinitely large, and therefore, in theory, this graph could extend outward infinitely on both sides, to -∞ σ on the left, and +∞ σ on the right. However, since it is rare for scores in most real distributions of data to extend beyond -3σ and 3σ, these are often the highest values displayed on bell curve graphs.

Now, the reason why it is important for us to explain the values on the bell curve is so that you can understand exactly what a z-score refers to, which is the number of standard deviation units a given score is above or below the mean. Hence, a z-score of 1.0 is equal to 1σ (i.e., the mean plus one standard deviation unit); a z-score of 2.0 is equal to 2σ (i.e., the mean plus two standard deviation units); a z-score of -1.0 is equal to -1σ (i.e., the mean minus one standard deviation unit); and a z-score of -2.0 is equal to -2σ (i.e., the mean minus two standard deviation units); and so on, and so forth…

Most of the time, z-scores for a given distribution do not conform neatly to whole values, like 1.0, 2.0, or 3.0; however, if you understand the concept, interpreting actual z-scores is still quite easy and straightforward. For instance, if we consider the z-score for your final grade in chemistry (as shown in Table 1.2), which was **1.26**, this means that your grade was 1.26 standard deviation units above the mean of Mr. Hardnose's class. Interpreting z-scores is as easy as **π,** right?

Sibling Rivalry: Z-scores & Percentiles

As alluded to earlier, z-scores are just one type of standardized score that provides useful information about a particular data point relative to other data points in the distribution from which it came. Another type of standardized score – one with which many people are more familiar – is the percentile. In many instances percentiles can be used interchangeably with z-scores, and each can be converted to the other. Unlike z-scores, which, in theory, can range from -∞ to +∞, percentiles[5] have a restricted range of 0 to 100. In simplest terms, a given percentile represents the value within a distribution, below which a certain percentage of scores fall. Hence, if an individual achieves a percentile of 84, this means his score is higher than 84% of the other individuals in the same distribution.

Perhaps the context with which most people are familiar with percentiles is the SAT (Scholastic Assessment Test). Though the

---

[5] Percentiles are not limited to whole numbers, so, in theory, there are an infinite number of percentiles within this finite range of 0 to 100.

current version of the SAT exam, introduced in 2005, has three sections of performance (Mathematics, Critical Writing, and Critical Reading), with possible scores on the full exam ranging from 600 to 2400, the version of the exam that you may be most familiar with had two sections (Mathematics and Verbal), with possible scores on the full exam ranging from 400 to 1600. From year to year, changes in the difficulty of the exam, scoring procedures and various other anomalies meant that the mean of each cohort would vary, in some cases by more than 30 points. Hence, traditional scores on the SAT alone could not be trusted to give an accurate depiction of an individual's performance on the exam. As such, percentile scores are used because they provide information about how a student performed on the SAT exam relative to his or her peers taking that exact same exam, in a way that controls for factors like exam difficulty and other year-to-year differences. Hence, if you check your SAT transcript from high school, the percentile of your scores will tell you not only how you did on the exam relative to your peers, but how you might have done on the current exam with its new scoring system. If your SAT score yielded a percentile of 50, that means that your score was higher than 50% of your peers taking the same exam that year. Your performance would thus be equivalent to someone achieving a score at the 50th percentile on today's SAT exam, even though the raw scores on the respective exams may differ. For example, a score at the 50th percentile in 1996 for the full exam would have been 1013, whereas on the 2011 exam, a score at the 50th percentile would be 1500.[6] In Chapter 8, we provide an expanded discussion of the SAT exam and how z-scores could be used to make more sophisticated, and possibly fairer, comparisons between students taking the exam.

For the most part, percentiles and z-scores provide the same type of information, though on different continuums. Whereas the

---

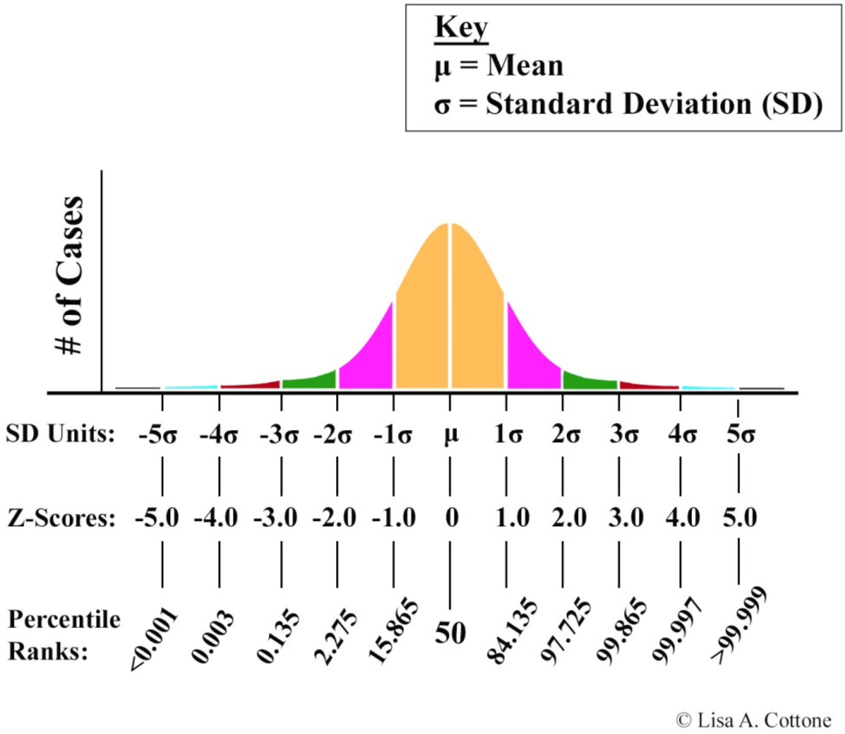
[6] National Center for Education Statistics (NCES; nces.ed.gov); SAT mean scores of college-bound seniors and percentage of graduates taking SAT, by state; extracted June 20, 2013.

mean of a distribution yields a z-score of 0, and the standard deviation points emanating from the mean extend outward in positive and negative directions at fixed units in multiples of 1.0 (i.e., 1.0, 2.0, 3.0, etc.), when it comes to percentiles, that same exact mean translates into a percentile of 50, and the standard deviation points emanating from that mean extend outward in positive and negative directions at variable units that decrease the further they are from the mean. On the following page, in Table 1.4, we list the percentiles at various standard deviation units emanating from the mean, along with their z-score equivalents. This is also illustrated graphically in Figure 1.3. As you can see, the mean of a distribution always corresponds to a z-score of 0, and a percentile of 50.

Table 1.4: Z-Scores & Percentiles Directly Compared

Standard Deviation Units From the Mean	Z-score	Percentiles
-5.0 σ	-5.0	< 0.001
-4.0 σ	-4.0	0.003
-3.0 σ	-3.0	0.135
-2.0 σ	-2.0	2.275
-1.0 σ	-1.0	15.865
0 σ (a.k.a. the mean)	0 (a.k.a. the mean)	50 (a.k.a. the mean)
1.0 σ	1.0	84.135
2.0 σ	2.0	97.725
3.0 σ	3.0	99.865
4.0 σ	4.0	99.997
5.0 σ	5.0	> 99.999

FIGURE 1.3: Z-SCORES & PERCENTILES PLOTTED ON THE BELL CURVE UP TO ±5σ



At this point, an obvious question might be: "If percentiles are statistically equivalent to z-scores, and people are more familiar with them, why not just use percentiles for the statistical comparisons of this book instead of z-scores?" The reason we chose to use z-scores instead of percentiles is because we believe that for many readers who may be unfamiliar with standardized scores, it is easier to recognize the statistical rarity of exceptional achievements on the z-score continuum (ranging, in theory, from -∞ to +∞) than on the percentile continuum (ranging from 0 to 100).

For instance, let's say we wanted to compare the standardized scores for height for a man who is 6'6" tall with a man who is 7'0" tall (see Table 1.5 for an expanded presentation of height and associated z-score values). A man who is 6’6” has a height that is 3σ (i.e., 3 standard deviation units above the mean for height, which in this case, is 5’10”), and a man who is 7’0” has a height that is 5σ (i.e., 5 standard deviation units above the mean for height). Converting these values to z-scores and percentiles, the man who is 6’6” has a z-score for height of 3.0 and the percentile for that height is 99.87 (i.e., this man is taller than 99.87% of other men); the man who is 7’0” has a z-score for height that is 5.0 and the percentile for that height is 99.99 (i.e., he is taller than 99.99% of other men). Now, if we are to compare the percentiles of the two men’s heights – 99.87 to 99.99 – it doesn’t appear that there is a tremendous difference (0.12 percentile units), because percentiles asymptote to 100 (i.e., they get closer and closer to 100 without ever reaching 100); however, if we compare the *z-scores* for the two men – 3.0 to 5.0 – this difference appears much more substantial (2.0 z-score units) because, in theory, z-scores extend outward to infinity. Hence, in order to make more salient the magnitude of the difference between two outstanding scores, we believe that z-scores are much better at doing this than percentiles. So, while the big brother of the standardized score family, the percentile, usually gets all the attention and fame, in our book, little brother z-score has exactly what we are looking for. Perhaps these two siblings have some family therapy in store for their future.

Z-scores in the Real World

The contexts with which z-scores are primarily used today are in the fields of clinical neuropsychological testing and research. Neuropsychological testing is often provided to kids when there is a suspicion that a learning disability is present. It is also provided to adults to check for the presence of any neurological or neuropsychological deficits, particularly following traumatic brain injury, onset of a neurological disease, or if dementia is suspected. A typical neuropsychological battery consists of dozens of tests, across myriad domains of functioning including: verbal and nonverbal intelligence; academic achievement; verbal and nonverbal memory; speech production; attention; gross motor skills; fine motor skills; reading fluency; reading comprehension; communication fluency; inhibition; abstract reasoning, and many others. Each of the tests given in a typical neuropsychological examination is scored differently. Some have a range of 40 to 160 points, with the mean being 100 points; others might have a range of zero to 60 points, with means varying by age and reference group; and still other tests may yield scores, not in the form of points, but in the form of time-to-completion, ranging from zero seconds to infinity. For example, on a typical neuropsychological profile, you may need to compare a patient's Full Scale IQ score of 105 (out of 160) on the SB5 (Stanford-Binet Intelligence Scales, Fifth Edition) to his score of 50 (out of 60) on the Boston Naming Test (a test of visual recognition and verbal fluency) and his score of 82 seconds (for completion) on the Grooved Pegboard Test (a test of fine motor skills).

In short, a neuropsychological exam requires a clinician to compare apples to oranges, not just in terms of the actual skills being assessed, but also in terms of the types of scores being compared. It is as if you need to compare a baseball player's batting average to the number of goals a hockey player scored, to the time it takes a sprinter to finish the 100-yard dash. How can this be done? Standardized scores, of which percentiles and z-scores are just a few, assist in this process. For reasons that are more complex than what is necessary to explain here, in neuropsychological testing, percentiles are more commonly used than z-scores to assist in these comparisons. However, in research (including research on neuropsychological test data), z-scores tend to be more commonly used. Having experience and training in both clinical and research settings, we believe that for the purposes of this book, z-scores will illuminate key distinctions between superb performances with greater clarity than will percentiles.

In an effort to help you understand how z-scores relate to familiar concepts in the real world, Table 1.5 lists a range of z-scores (from -5 to +5) for two variables with which most people have some experience: human height and IQ. The data for height in this table come from the American National Health and Nutrition Examination Survey, which reports American men as having a mean height of 70" (i.e., 5'10" tall) with a standard deviation of 2.8", and American women as having a mean height of 65" (i.e., 5' 5" tall) with a standard deviation of 2.5". For IQ, on most standardized, normed IQ tests, the mean is usually set at 100 and the standard deviation is set at 15. As a frame of reference for IQ, acceptance into Mensa International, the society for individuals with high IQs, requires an IQ score in at least the 98th percentile (or about 2σ)[7], which equates to an IQ score of about 130 on the most common intelligence tests used today.

As you can see from Table 1.5, an IQ score of 130, which will get you in to Mensa, yields a z-score of 2.0, which is pretty impressive. But just how impressive is a z-score of 2.0? What would that equate to in human height? Well, by looking at Table 1.5 you can determine that a z-score of 2.0 would be statistically equivalent to a man who is 6 feet, 3.6 inches tall, and a woman who is 5 feet, 10 inches tall. If you can understand this, pat yourself on the back: you have just passed our first exam and now you know how to compare *apples to oranges*!

---

[7] http://www.mensa.org/about-us

TABLE 1.5: Z-SCORES OF ADULT HUMAN HEIGHT & IQ

Z-SCORE	IQ[8]	ADULT MALE HEIGHT[9]	ADULT FEMALE HEIGHT[7]
-5.0	25	56.0” (4 ft. 8.0 in.)	52.5” (4 ft. 4.5 in.)
-4.0	40	58.8” (4 ft. 10.8 in.)	55.0” (4 ft. 7.0 in.)
-3.0	55	61.6” (5 ft. 1.6 in.)	57.5” (4 ft. 9.5 in.)
-2.0	70	64.4” (5 ft. 4.4 in.)	60.0” (5 ft. 0 in.)
-1.0	85	67.2” (5 ft. 7.2 in.)	62.5” (5 ft. 2.5 in.)
0	**100**	**70” (5 ft. 10.0 in.)**	**65.0” (5 ft. 5.0 in.)**
1.0	115	72.8” (6 ft. 0.8 in.)	67.5” (5 ft. 7.5 in.)
2.0	130	75.6” (6 ft. 3.6 in.)	70.0” (5 ft. 10 in.)
3.0	145	78.4” (6 ft. 6.4 in.)	72.5” (6 ft. 0.5 in.)
4.0	160	81.2” (6 ft. 9.2 in.)	75.0” (6 ft. 3.0 in.)
5.0	175	84.0” (7 ft. 0 in.)	77.5” (6 ft. 5.5 in.)

[8] Esther Strauss, Elisabeth M. S. Sherman, Otfried Spreen (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*.
[9] NHANES II Study (National Health and Nutrition Examination Survey of 1976-1980), as cited in *Seeing Through Statistics, 2nd Edition* by Jessica M. Utts (1999). Also found on the National Center for Health Statistics website: http://www.cdc.gov/nchs/nhanes.htm

From Apples and Oranges to Baseballs, Basketballs Pigskins and Pucks: How *Our* Z-scores Were Obtained

All of the sports data used in our analyses were acquired from sources we considered reputable. Our baseball data were obtained from Major League Baseball – specifically, from MLB.com, while the data for football, basketball and hockey, reported in Chapter 6, were obtained from Sports-Reference.com. We obtained our data from these sources, not only because we trusted them to have the most accurate data, but also because they each have a powerful sorting feature, allowing one to select player data that meet specific criteria of interest. For our analyses, we sorted and selected cases according to various criteria so that we could make the most refined and meaningful comparisons between players.

Before extracting any statistics, we debated how we would approach the analysis of such a large pool of data in a way that was both interesting and efficient. Rather than computing z-scores for every player, in every season, for every available statistical category (a monumental task, even for those with supercomputers and a lust for solitude), we decided to take a more streamlined approach. In our streamlined approach, we chose to compute z-scores only for the players who led certain statistical categories that we deemed most interesting, and only for selected seasons.

For baseball, we chose to examine a mix of traditional and sabermetric statistics in both hitting and pitching domains. Within the hitting domain, the traditional statistics we examined were batting average, home runs and RBI[10] and the sole sabermetric statistic we examined was OPS.[11] For pitching, the traditional statistics we examined were wins, ERA[12] and strikeouts, while the

[10] RBI = Runs Batted In.
[11] OPS = On-base [percentage] Plus Slugging [average]
[12] ERA = Earned Run Average

sole sabermetric statistic we examined was WHIP.[13] We then identified seven major eras of baseball history, demarcated by the popular reference book *The Baseball Timeline*[14] (an official publication of Major League Baseball). According to *The Baseball Timeline*, the seven eras of baseball history are:

1. 1876 - 1900 – Pre-Modern Era[15]
2. 1901 - 1919 – Deadball Era[16]
3. 1920 - 1960 – Home Run Era
4. 1961 - 1975 – Expansion Era
5. 1976 - 1985 – Free Agency Era
6. 1986 - 2007 – Steroid Era
7. 2008 – 2014[17] – Post-Steroid Era

For each of the seven eras listed above, we only calculated z-scores for the players who had the best single-season performance for the ***entire era***, within their ***respective league*** (i.e., American or National), for the ***selected statistics of interest*** mentioned above. For instance, in the Deadball Era, which spanned the years 1901-1919, we didn't calculate a z-score for home runs for *every player* across *every season* between 1901 and 1919: we only calculated z-scores for the home run totals of Babe Ruth in 1919 and Gavvy Cravath in 1915 because their single-season home run totals – 29 for Ruth, and 24 for Cravath – were higher than those of any other player in their respective leagues, for any season within the Deadball

[13] WHIP = Walks [plus] Hits [divided by] Innings Pitched

[14] Solomon, Burt (2001). *The Baseball Timeline.* DK Publishing, Inc. New York

[15-17] Though *The Baseball Timeline* set the years of the Pre-Modern Era as being between 1876-1902, and the years of the Deadball Era as being between 1903-1919, we decided to make 1900 the end of the Pre-Modern Era and 1901 the start of the Deadball Era for the purposes of our analyses, given that the American League was founded in 1901.

[17] 2014 is the year in which this book was completed. Hence the 2014 baseball season is the final season in which statistics were obtained for the data analyses of this book.

Era. Likewise, in all other eras, and for all other statistical categories, we only computed z-scores for the single-season category leaders (i.e., for batting average, home runs, RBI, OPS, wins, ERA, strikeouts, and WHIP) of that particular ***era***, within each league.

As briefly noted above, we examined the data for each league separately. The reason is because for much of baseball history, the leagues were very different, and still are to some extent. Each league has its own distinct ballparks and even to this day most of the teams in one league still do not play any games in the ballparks of the teams from the other league. In addition, the leagues have different rules regarding the designated hitter (since 1973), and for many years there were even different umpires and umpiring norms for each league (until umpires began working both leagues during the 2000 season). Furthermore, each league had a different timeline for integrating non-White players. Hence, in light of all of these differences, we decided to look at the data for the American League and the National League separately.

Additionally, it should be noted that when computing z-scores for each category leader, rather than using means and standard deviations derived from the entire pool of players within a respective league for a given season, we narrowed the pool (using the sorting feature on MLB.com) for each season of interest to include only the players who qualified for the batting title[18] (for

---

[18] Leaderboard Glossary – Baseball. www.Baseball-reference.com. Retrieved 2012-05-26:

Qualifications for the Batting Title

- Pre-1920: A player generally had to appear in 100 or more games when the schedule was 154 games; and 90 games when the schedule was 140 games.
- 1920–1949: A player generally had to appear in 100 games to qualify in the National League; the American League used 100 games from 1920–1935, and 400 at-bats from 1936–1949.
- 1950–1956: A player needed 2.6 at-bats per team game originally scheduled. (With the 154-game schedule of the time, that meant a rounded-off 400 at-bats.) From 1951–1954, if the player with the highest

hitting statistics) and the ERA title[19] (for pitching statistics) within that season for each respective league. We did this to avoid skewing

---

average in a league failed to meet the minimum at-bat requirement, the remaining at-bats until qualification (e.g., five, if the player finished the season with 395 at-bats) were hypothetically considered hitless at-bats; if his recalculated batting average still topped the league, he was awarded the title. This standard applied in the AL from 1936–1956.

- 1957 – The Present: A player needs 3.1 plate appearances per team game originally scheduled; In the 154-game schedule, the required number of plate appearances was 477, and since the advent of the 162-game schedule, the requisite number of plate appearances has been 502. (Adjustments to this 502 figure have been made during strike-shortened seasons, such as 1972, 1981, 1994, and 1995.)
- *From 1967 to the present, if the player with the highest average in a league fails to meet the minimum plate-appearance requirement, the remaining at-bats until qualification (e.g., five at-bats, if the player finished the season with 497 plate appearances) are hypothetically considered hitless at-bats; if his recalculated batting average still tops the league, he is awarded the title.*

[19] Article by Dan Levitt in Society for American Baseball Research's *The National Pastime: No. 25.*

Qualifications for the ERA Title

- In 1917, the National League adopted the standard that a pitcher had to have pitched a minimum of 10 complete games in order to qualify as an ERA leader. Before that, the National League used a variety of standards: in 1912, it was 15 games pitched; in 1913, it was five complete games; from 1914-1915, it was 15 games pitched; and in 1916, it was 12 games pitched.
- In 1946, the American League adopted the standard that a pitcher had to have pitched a minimum of 10 complete games. Before that, the league used various criteria. Up until 1919, the American League simply listed the top pitchers by ERA without identifying a standard. In 1919, the American League finally adopted a minimum standard of 45 innings pitched. The next year, 1920, it changed the standard to 10 complete games, but in 1921, it reverted to 45 innings pitched. The following year, 1922, it switched back to 10 complete games, where it remained for three seasons. In 1925, the American League once again reverted back to a standard of 45 innings pitched.
- In 1951, the standard of the Major League was changed to one inning pitched per team game, which is the current standard (i.e., 162 innings

the means and standard deviations for each season, which would have happened if we included, for example, a player who batted .500 in a given season, but only had 10 at-bats; or, alternately, a pitcher who may have had an ERA of 0.00 but pitched only one inning. Here it should be noted that relief pitchers were not included in the main set of our analyses for pitching (as they generally do not qualify for the ERA title, according to the norms of Major League Baseball). However, later in the book in Chapter 4 we provide z-scores for relief pitchers who won the MVP and in that chapter we describe in detail the methods we used to classify pitchers (as either a starter or reliever) and compute their z-scores.

For each of the other sports we examined – football, basketball and hockey – we made similar decisions to pare down the qualifying pool so as to avoid the inclusion of part-time players into our analyses. Since baseball tends to have more formal rules for inclusion into the pool of full-time players for various statistical distinctions, such as the batting title and ERA title, we used baseball as a guide to help us determine the appropriate inclusion criteria for the other sports. As such, the inclusion guidelines that we chose for the other sports generally corresponded to the least restrictive qualification standards for baseball's batting title in the history of the sport: namely, participation in at least 65% of the team's games.[20] Regarding the specific statistical categories of interest for the other three sports, we examined the following:

---

pitched for a standard 162-game season, without any rainouts or tie-breaker playoff games).

[20] Used between 1920 and 1949 in the National League, when the guideline for qualification for the batting title was playing in 100 of the team's 154 games (i.e., 65%).

- **Football**
    - Offense: *Rushing Yards; Touchdown Passes; Completion Percentage; Receiving Yards;* and *Touchdown Receptions.*
    - Defense: *Sacks;* and *Interceptions.*

- **Basketball**
    - Scoring: *Total Points; Points Per Game (PPG); Total Rebounds; Rebounds Per Game (RPG); Total Assists; Assists Per Game (APG);* and *Three-Point Field Goal Percentage.*

- **Hockey**
    - Scoring: *Total Goals; Total Assists;* and *Total Points.*
    - Goaltending: *Wins; Shutouts; Goals Against Average;* and *Save Percentage*.

As noted above, we didn't compute z-scores for the selected statistics for *every player*, in *every season*. Rather, as a means of making certain specific comparisons across sports, we computed z-scores for the single-season record holders in each category of interest, within each sport. We then pitted these z-scores in direct competition with those we derived from our baseball analyses for a **BATTLE ROYAL** of z-scores to identify the best achievements of the four major American team sports. Who emerges from this battle royal as the Heavyweight Champion of Z-scores? Read on to find out…

The Least You Need to Know

1. Z-scores are a standardized statistic, like percentiles, that indicate how a person's performance on a given measure compares to others in the same group.

2. Virtually any measure of performance, including most sports statistics, can be converted to z-scores.

3. The formula to compute a z-score is:

$$\textbf{Z-score} = \frac{(Score\ X - Mean)}{Standard\ Deviation}$$

4. For the purposes of this book, the better the performance, the higher the z-score in the positive direction (e.g., for home runs a z-score of +2 is better than a z-score of +1, 0, -1, -2, etc.).

NOTE: Technically, for variables in which lower values represent better performance (e.g., ERA and WHIP) the more *negative* the z-score, the better the performance. However, to simplify presentation in this book, negative z-scores for ERA and WHIP were flipped to positive z-scores to help readers make easier comparisons with z-scores from other statistical categories.

The Least You Need to Know

5. Though, in theory, z-scores extend outward to infinity in both positive and negative directions (i.e., they range from -∞ to +∞), for most variables in the real world it is rare to encounter z-scores extending beyond -5 and +5, and the vast majority of z-scores fall between -2 and +2. The average or "mean" score for any statistic will always have a z-score of 0.

6. As a familiar point of reference, an IQ score of 130, which is the minimum IQ necessary to get into Mensa (the society for people with high IQs), is equivalent to a z-score of2.0. A z-score of +5.0 in the realm of height among men yields a measurement of 7'0" (which is about the height of Shaquille O'Neal).

APPENDIX C

FORMULAS

Key:
N = Number of scores in a distribution
X = Score of interest
∑ = Summation of scores in a given set or distribution

Z-score $= \frac{(Score\ X - Mean)}{Standard\ Deviation}$

Mean $= \frac{(Sum\ of\ All\ Scores)}{N}$

Standard Deviation (Population) $= \sqrt{\frac{\Sigma\ (Score\ X - Mean)^2}{N}}$

Standard Deviation (Sample) $= \sqrt{\frac{\Sigma\ (Score\ X - Mean)^2}{N-1}}$

Market Power = (Median Household Income) **x** (Population)

APPENDIX D

COMPUTATION EXAMPLES

COMPUTATION OF THE STANDARD DEVIATION:
EXAMPLE - CLASS A AND CLASS B (CHAPTER 1, SEE PG. 7)

Standard Deviation (SD) = $\sqrt{\frac{\Sigma\,(Score\ X - Mean)^2}{N}}$

∑ = Sum
N = Number of Scores or Observations

Class A: Computation of the Standard Deviation

Raw Test Score	Mean	(Score X- Mean)	(Score X- Mean)^2
100	60	(100-60) = 40	(40x40) = 1,600
100	60	(100-60) = 40	(40x40) = 1,600
100	60	(100-60) = 40	(40x40) = 1,600
100	60	(100-60) = 40	(40x40) = 1,600
100	60	(100-60) = 40	(40x40) = 1,600
20	60	(20-60) = - 40	(-40x-40) = 1,600
20	60	(20-60) = - 40	(-40x-40) = 1,600
20	60	(20-60) = - 40	(-40x-40) = 1,600
20	60	(20-60) = - 40	(-40x-40) = 1,600
20	60	(20-60) = - 40	(-40x-40) = 1,600
Sum of Squares: Σ (Score X- Mean)^2			16,000

Standard Deviation (SD) = $\sqrt{\frac{16{,}000}{10}}$

Standard Deviation (SD) = $\sqrt{1{,}600}$

Standard Deviation (SD) = 40

Class B: Computation of the Standard Deviation

Raw Test Score	**Mean**	**(Score X- Mean)**	**(Score X- Mean)^2**
70	60	(70-60) = 10	(10x10) = 100
70	60	(70-60) = 10	(10x10) = 100
70	60	(70-60) = 10	(10x10) = 100
70	60	(70-60) = 10	(10x10) = 100
70	60	(70-60) = 10	(10x10) = 100
50	60	(50-60) = - 10	(-10x-10) = 100
50	60	(50-60) = - 10	(-10x-10) = 100
50	60	(50-60) = - 10	(-10x-10) = 100
50	60	(50-60) = - 10	(-10x-10) = 100
50	60	(50-60) = - 10	(-10x-10) = 100
Sum of Squares: Σ (Score X- Mean)^2			1,000

Standard Deviation (SD) = $\sqrt{\frac{1,000}{10}}$

Standard Deviation (SD) = $\sqrt{100}$

Standard Deviation (SD) = 10

ABOUT THE AUTHORS

Dr. John G. Cottone is a clinical psychologist in private practice in Stony Brook, NY. He received his Ph.D. in clinical psychology from St. John's University and previously earned a master's degree in biopsychology from Stony Brook University. He is the author of numerous peer-reviewed research publications, including articles and book chapters on topics in psychology, psychiatry and neuroscience. He is the author of a 2013 self-help book entitled *Who Are You? Essential Questions for Hitchhikers on the Road of Truth* (Story Bridge Books) and has had vocational and avocational contributions to *The New York Times* and *The Washington Post*. He is also an avid baseball fan, a member of Red Sox Nation, and was a four-year scholar-athlete at Dowling College, where he pitched on the school's baseball team.

Jason Wirchin earned his B.A. in Political Science, along with a Minor in Journalism, from Stony Brook University in 2010. He is currently a television news producer for News 12 The Bronx, and has also established himself as a digital content producer with the News 12 Networks. He has been published several times by both national and local papers, including *The New York Times*, *New York Daily News* and *Newsday;* and is a frequent caller to WFAN. Jason takes great pride in the history of our national pastime and is a religious fan of the New York Mets. The eldest of three, he grew up in Huntington Station, Long Island and lives with his wife, Ellen, in New York City.

Check out other great publications from **Story Bridge Books**:

http://www.storybridgebooks.com

Who Are You?
Essential Questions for Hitchhikers on the Road of Truth
by John G. Cottone, PhD © 2013

Who Are You? Essential Questions for Hitchhikers on the Road of Truth explores the questions we need to ask – about the psychology of human behavior, politics, science, metaphysics, and the mysteries of God – to live lives of meaning.
It inspires readers to gently spiral through deeper states of contemplation and self-inquiry using reflective questioning and Socratic dialogues. *Who Are You?* is a meditation companion, a catalyst for group discussion, a personal mirror for honest glimpses at the soul and an invitation for self-growth.

Book of the Sky God
by Laura Markowitz © 2013

The gods are coming back to Earth, and they're not particularly happy. Five unlikely teenagers and one immortal shape-shifter are the only ones who can save humanity. **Ram Rajathani** thinks he knows everything about his best friend, but he doesn't know that Henry has been turned into a part-time zombie, or that Henry's got a crush on a girl he recently met online. The girl, **ComixChik8**, is torturing Henry on purpose, but what she really wants to know is what she's supposed to do with the cursed blue iguana that mysteriously appeared in her school locker. **Henry Lipton** has no idea he's a part-time zombie. He thinks his biggest problem is girl-o-phobia. Ram's sister **Laila** has a dangerous Mayan artifact stashed in the back of her closet. She suspects it can save humanity from the gods, but she has no idea how. Her older sister, **Nina**, isn't afraid of gods, but she's desperate to make some friends. She's about to discover she has mad spy skills. In the meantime, the secretive **Brotherhood of the Prophecy** rescheduled the end of the world, and now they plan to cash in on it. These five teenagers must race against time to outwit the Brotherhood and save humanity from the gods. No biggie.