

1

Scores and Scales

Assessments yield information. The information is encoded in scores or in other types of data. It is important to consider the different types of data because the types of data restrict what options are available to analyze the data.

1.1 Getting Started

1.1.1 Load Libraries

First, we load the R packages used for this chapter:

```
library("petersenlab")
library("M0TE")
library("here")
library("tidyverse")
library("tinytex")
library("knitr")
library("kableExtra")
library("rmarkdown")
library("bookdown")
```

1.1.2 Prepare Data

1.1.2.1 Simulate Data

For this example, we simulate data below.

```
set.seed(52242)

rawData <- rnorm(
  n = 1000,
  mean = 200,
  sd = 30)

scores <- data.frame(
  rawData = rawData,
  rawDataNoMissing = rawData)
```

1.1.2.2 Add Missing Data

Adding missing data to dataframes helps make examples more realistic to real-life data and helps you get in the habit of programming to account for missing data.

```
scores$rawData[c(5,10)] <- NA
```

1.2 Data Types

There are four general data types: nominal, ordinal, interval, and ratio. Depending on the use of the variable, the data could fall into more than one category. The type of data influences what kinds of data analysis you can do. For instance, parametric statistical analysis (e.g., *t* test, analysis of variance [ANOVA], and linear regression) assumes that data are interval or ratio.

1.2.1 Nominal

Nominal data are distinct categories. They are categorical and unordered. Nominal data make no quantitative claims. Nominal data represent things that we can name (e.g., cat and dog). Nominal data can be represented with numbers. For example, zip codes are nominal. Numbers that represent a participant's sex, race, or ethnicity are also nominal. Categorical data are often dummy-coded into binary (i.e., dichotomous) variables, which represent nominal data. Higher numbers of nominal data do not reflect higher (or lower) levels of the construct because the numbers represent categories that do not have an order.

1.2.2 Ordinal

Ordinal data are ordered categories: they have a name and an order. They make no claim about the conceptual distance between the ranks, only that higher values represent higher (or lower) levels of the construct. For example, ranks following a race are ordinal—that is, the person with rank 1 finished before the person with rank 2, who finished before the person with rank 3 ($1 > 2 > 3 > 4$). Ordinal data make a limited claim because the conceptual distance between adjacent numbers is not the same. For instance, the person who finished the race first might have finished 10 minutes before the second-place finisher; whereas the third-place finisher might have finished 1 second after the second-place finisher.

That is, just because the numbers have the same *mathematical* distance does not mean that they represent the same *conceptual* distance on the construct. For example, if the respondent is asked how many drinks they had in the past day, and the options are 0 = 0 drinks; 1 = 1–2 drinks; 2 = 3 or more drinks, the scale is ordinal. Even though the numbers have the same mathematical distance (1, 2, 3), they do not represent the same conceptual distance. Most data in psychology are ordinal data even though they are often treated as if they were interval data.

1.2.3 Interval

Interval data are ordered and have meaningful distances (i.e., equal spacing between intervals). You can sum interval data (e.g., 2 is 2 away from 4), but you cannot multiply interval data

($2 \times 2 \neq 4$). Examples of interval data are temperatures in Fahrenheit and Celsius—100 degrees Fahrenheit is not twice as hot as 50 degrees Fahrenheit. A person's number of years of education is interval, whereas educational attainment (e.g., high school degree, college degree, graduate degree) is only ordinal. Although much data in psychology involve numbers that have the same mathematical distance between intervals, the intervals likely do not represent the same conceptual distance. For example, the difference in severity of two people who have two symptoms and four symptoms of depression, respectively, may not be the same difference in depression severity as two people who have four symptoms and six symptoms, respectively.

1.2.4 Ratio

Ratio data are ordered, have meaningful distances, and have a true (absolute) zero that represents absence of the construct. With ratio data, multiplicative relationships are true. An example of ratio data is temperature in Kelvin—100 degrees Kelvin is twice as hot as 50 degrees Kelvin. There is a dream of having ratio scales in psychology, but we still do not have a true zero with psychological constructs—what does total absence of depression mean (apart from a dead person)?

1.3 Score Transformation

There are a number of score transformations, depending on the goal. Some score transformations (e.g., log transform) seek to make data more normally distributed to meet assumptions of particular analysis approaches. Score transformations alter the original (raw) data. If you change the data, it can change the results. Score transformations are not neutral.

1.3.1 Raw Scores

Raw scores are the original data, or they may be aggregations (e.g., sums or means) of multiple items. Raw scores are the purest because they are closest to the original operation (e.g., behavior). A disadvantage of raw scores is that they are scale dependent, and therefore they may not be comparable across different measures with different scales. An example histogram of raw scores is in [Figure 1.1](#).

1.3.2 Norm-Referenced Scores

Norm-referenced scores are scores that are referenced to some norm. A norm is a standard of comparison. For instance, you may be interested in how well a participant performed relative to other children of the same sex, age, grade, or ethnicity. However, interpretation of norm-referenced scores depends on the measure and on the normative sample. A person's norm-referenced score can vary widely depending on which norms are used. Which reference group should you use? Age? Sex? Age and sex? Grade? Ethnicity? The optimal reference group depends on the purpose of the assessment. Pros and cons of group-based norms are discussed in [Section 15.5.2.2.2](#).

A standard normal distribution on various norm-referenced scales is depicted in [Figure 1.2](#), as adapted from [Bandalos \(2018\)](#).

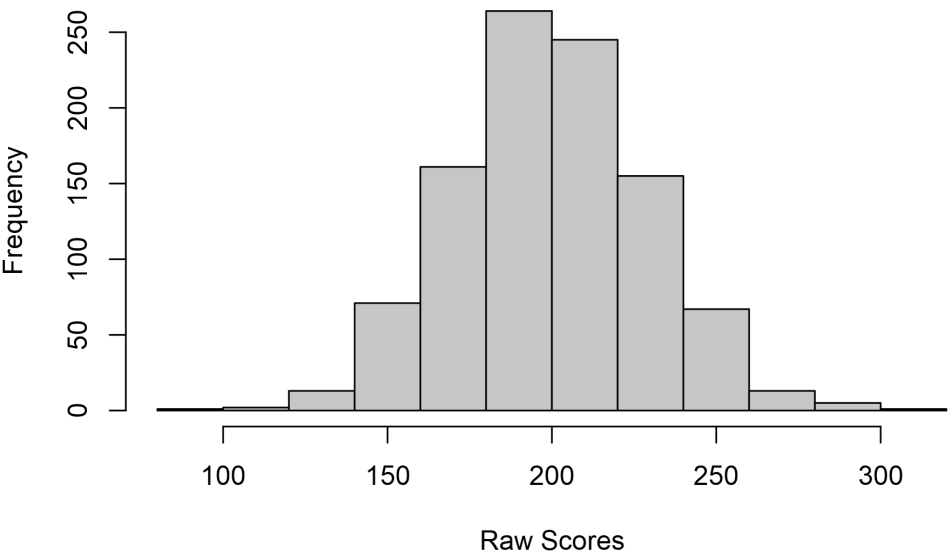


FIGURE 1.1 Histogram of Raw Scores.

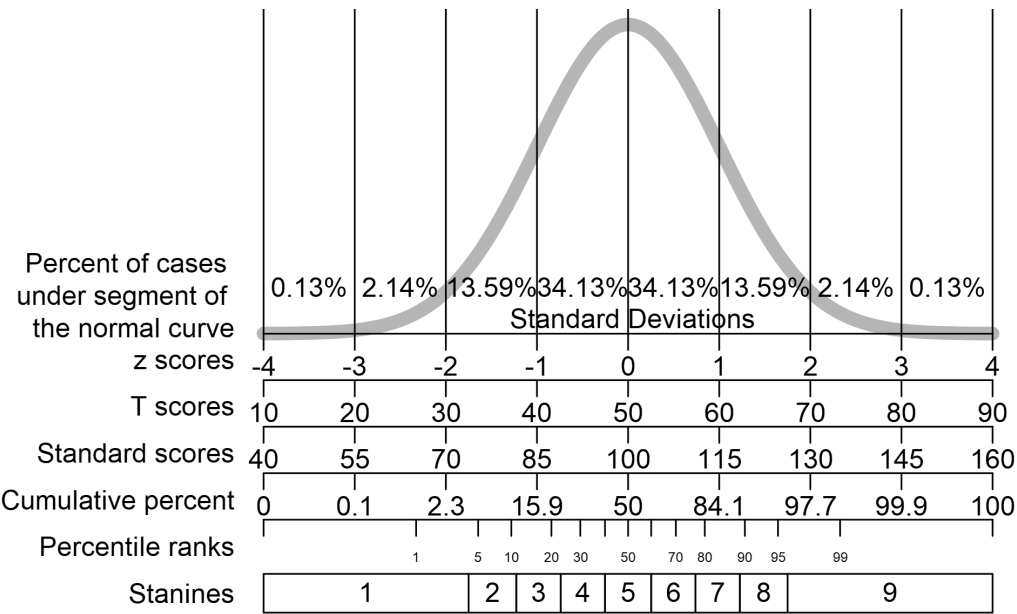


FIGURE 1.2 Various Norm-Referenced Scales.

1.3.2.1 Percentile Ranks

A percentile rank reflects what percent of people a person scored higher than, in a given group (i.e., norm). Percentile ranks are frequently used for tests of intellectual/cognitive ability, academic achievement, academic aptitude, and grant funding. They seem like interval data, but they are not intervals because the conceptual spacing between the numbers is not equal. The difference in ability for two people who scored at the 99th and 98th percentile, respectively, is not the same as the difference in ability for two people who scored at the 49th and 50th percentile, respectively. Percentile ranks are only judged against a baseline; there is no subtraction.

Percentile ranks have unusual effects. There are lots of people in the middle of a distribution, so a very small difference in raw scores gets expanded out in percentiles. For instance, a raw score of 20 may have a percentile rank of 50, but a raw score of 24 may have a percentile rank of 68. However, a larger raw score change at the ends of the distribution may have a smaller percentile change. For example, a raw score of 120 may have a percentile rank of 97, whereas a raw score of 140 may have a percentile rank of 99. Thus, percentile ranks stretch out differences for some people but constrict differences for others.

Here is an example of how to calculate percentile ranks using the `dplyr` package from the tidyverse ([Wickham et al., 2019](#); [Wickham, 2021](#)):

```
scores$percentileRank <-  
  percent_rank(scores$rawDataNoMissing) * 100
```

A histogram of percentile rank scores is in [Figure 1.3](#).

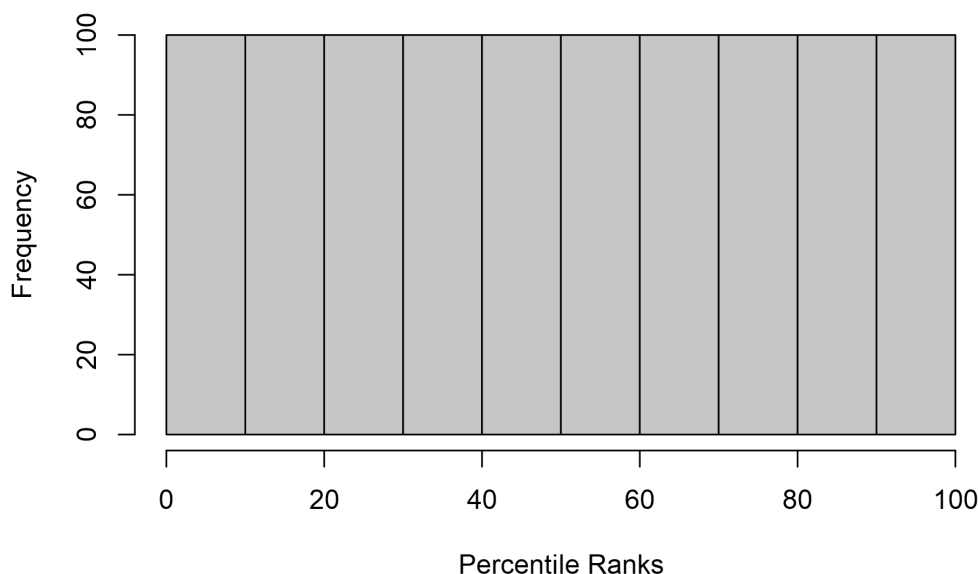


FIGURE 1.3 Histogram of Percentile Ranks.

1.3.2.2 Deviation (Standardized) Scores

Deviation or standardized scores are the transformation of raw scores to a normal distribution using some norm. The norm could be a comparison group, or it could be the sample itself. With deviation scores, you have similar challenges as with percentile ranks, including which reference group to use, but there are additional challenges. Deviation scores are more informative when the scores are normally distributed compared to when the scores are skewed. If scores are skewed, it can lead to two z scores on the opposite side of the mean having different probabilities.

Many constructs we study in psychology are not normally distributed. For example, the frequency of hallucinations among people would show a positively skewed distribution with a truncation at zero, representing a floor effect—i.e., most people do not show hallucinations. For instance, consider a hypothetical distribution of hallucinations. It might follow a folded distribution, as depicted in [Figure 1.4](#).

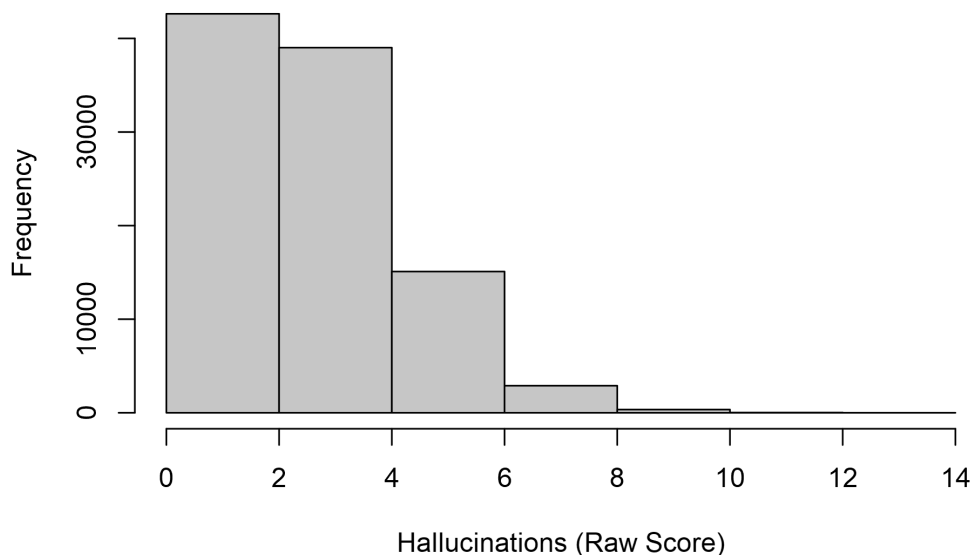


FIGURE 1.4 Histogram of Hallucinations (Raw Score).

Now consider the same distribution converted to a standardized (z) score, as depicted in [Figure 1.5](#).

Thus, you can compute a deviation score, but it may not be meaningful if the data and underlying construct are not normally distributed.

1.3.2.2.1 z scores

The z score is the most common standardized score, and it can help put scores from different measures with different scales onto the same playing field. z scores have a mean of zero and a standard deviation of one. To get a z score that uses the sample as its own norm, subtract

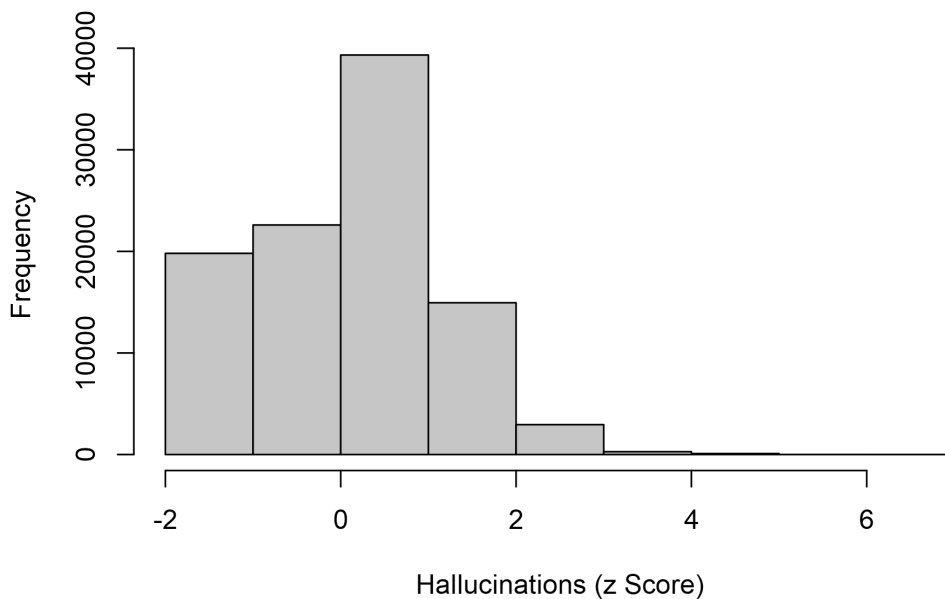


FIGURE 1.5 Histogram of Hallucinations (z Score).

the mean from all scores and divide by the standard deviation. Every z score represents how far that person's score is from the (normed) average, represented in standard deviation units. With a standard normal curve, 68% of scores fall within one standard deviation of the mean. 95% of scores fall within two standard deviations of the mean. 99.7% of scores fall within three standard deviations of the mean.

The area under a normal curve within one standard deviation of the mean is depicted in [Figure 1.6](#) and is calculated below using the `pnorm()` function, which calculates the cumulative density function for a normal curve.

```
stdDeviations <- 1
pnorm(stdDeviations) - pnorm(stdDeviations * -1)
```

```
[1] 0.6827
```

The area under a normal curve within two standard deviations of the mean is depicted in [Figure 1.7](#) and is calculated below:

```
stdDeviations <- 2
pnorm(stdDeviations) - pnorm(stdDeviations * -1)
```

```
[1] 0.9545
```

The area under a normal curve within three standard deviations of the mean is depicted in [Figure 1.8](#) and is calculated below:

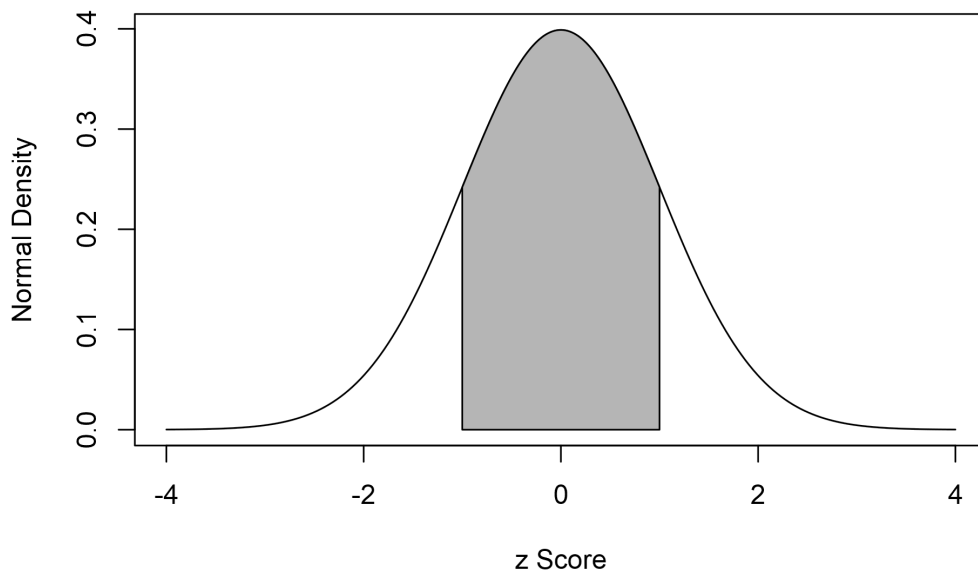


FIGURE 1.6 Density of Standard Normal Distribution. The gray region represents the area within one standard deviation of the mean.

```
stdDeviations <- 3
pnorm(stdDeviations) - pnorm(stdDeviations * -1)
```

```
[1] 0.9973
```

Alternatively, if you want to determine the z score associated with a particular percentile in a normal distribution, you can use the `qnorm()` function. For instance, the z score associated with the 37th percentile is:

```
qnorm(.37)
```

```
[1] -0.3319
```

z scores are calculated using Equation (1.1):

$$z = \frac{x - \mu}{\sigma} \quad (1.1)$$

where x is the observed score, μ is the mean observed score, and σ is the standard deviation of observed scores.

```
scores$zScore <- scale(scores$rawData)
```

An example histogram of z scores is in [Figure 1.9](#).

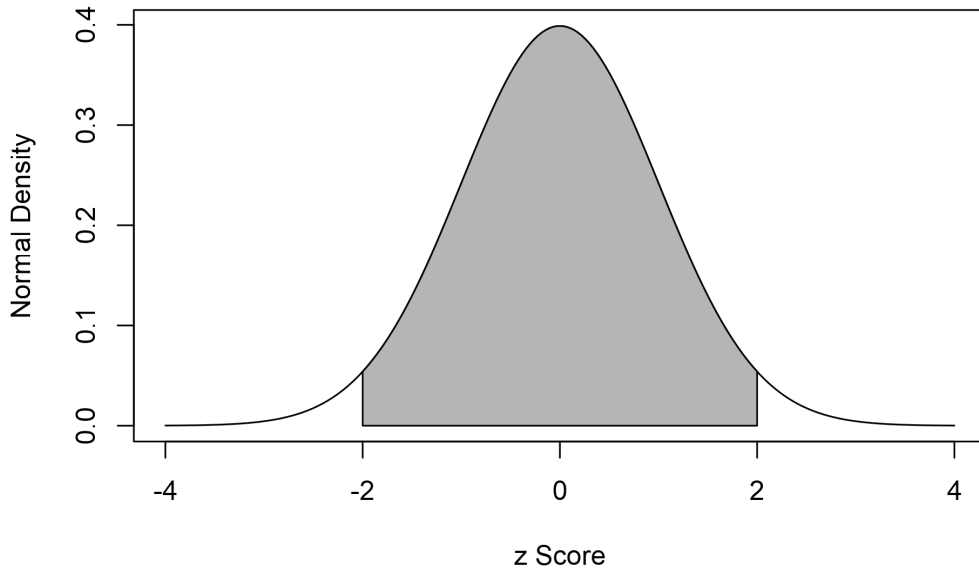


FIGURE 1.7 Density of Standard Normal Distribution. The gray region represents the area within two standard deviations of the mean.

1.3.2.2.2 T scores

T scores have a mean of 50 and a standard deviation of 10. T scores are frequently used with personality and symptom measures, where clinical cutoffs are often set at 70 (i.e., two standard deviations above the mean). For the Minnesota Multiphasic Personality Inventory (MMPI), you would examine peaks (elevations ≥ 70) and absences (≤ 30).

T scores are calculated using Equation (1.2):

$$T = 50 + 10z \quad (1.2)$$

where z are z scores.

```
scores$tScore <- 50 + 10*scale(scores$rawData)
```

An example histogram of T scores is in [Figure 1.10](#).

1.3.2.2.3 Standard scores

Standard scores have a mean of 100 and a standard deviation of 15. Standard scores are frequently used for tests of intellectual ability, academic achievement, and cognitive ability. Intellectual disability is generally considered an I.Q. standard score less than 70 (two standard deviations below the mean), whereas giftedness is at 130 (two standard deviations above the mean).

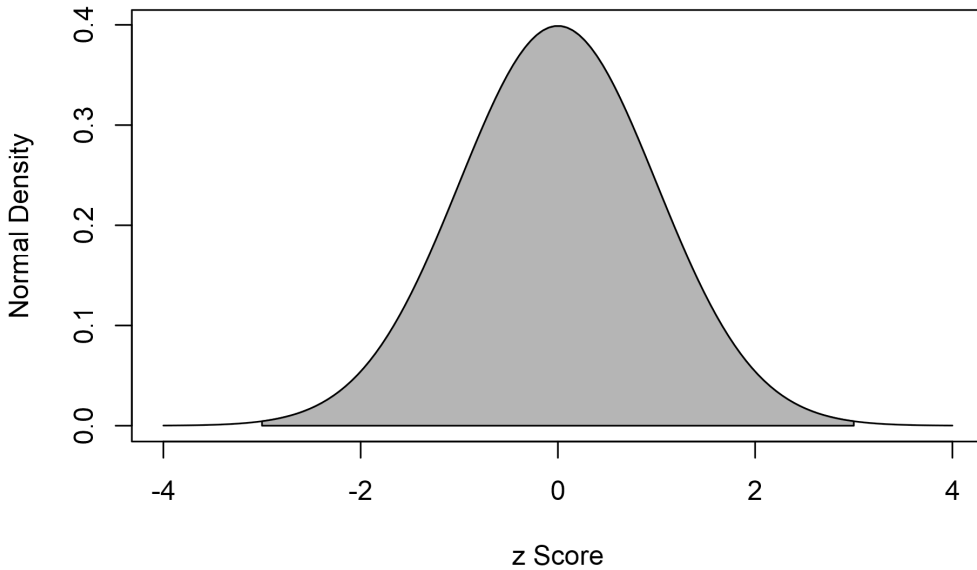


FIGURE 1.8 Density of Standard Normal Distribution. The gray region represents the area within three standard deviations of the mean.

Standard scores with a mean of 100 and standard deviation of 15 are calculated using Equation (1.3):

$$\text{standard score} = 100 + 15z \quad (1.3)$$

where z are z scores.

```
scores$standardScore <- 100 + 15*scale(scores$rawData)
```

An example histogram of standard scores is in [Figure 1.11](#).

1.3.2.2.4 Scaled scores

Scaled scores are raw scores that have been converted to a standardized metric. The particular metric of the scaled score depends on the measure. On tests of intellectual or cognitive ability, scaled scores commonly have a mean of 10 and a standard deviation of 3.

Using this metric, scaled scores can be calculated using Equation (1.4):

$$\text{scale score} = 10 + 3z \quad (1.4)$$

where z are z scores.

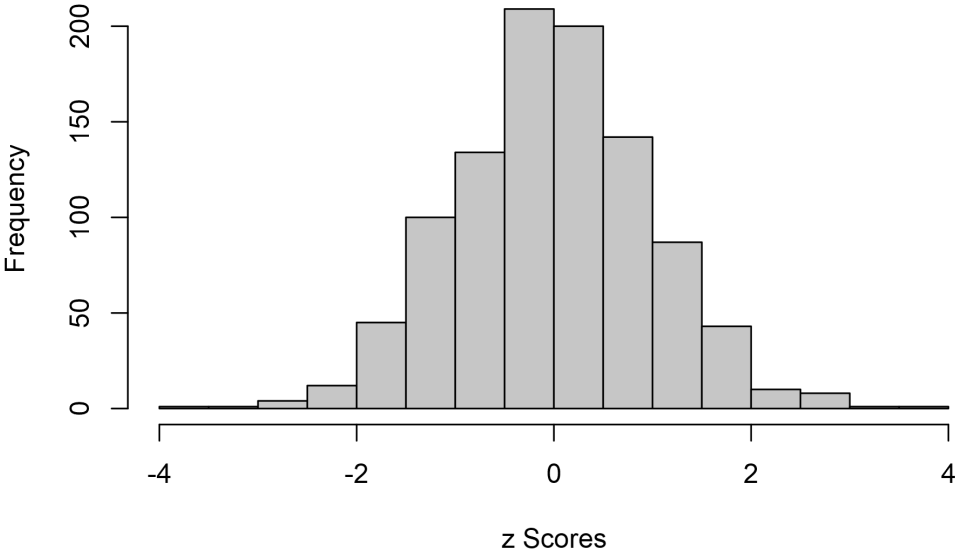


FIGURE 1.9 Histogram of z Scores.

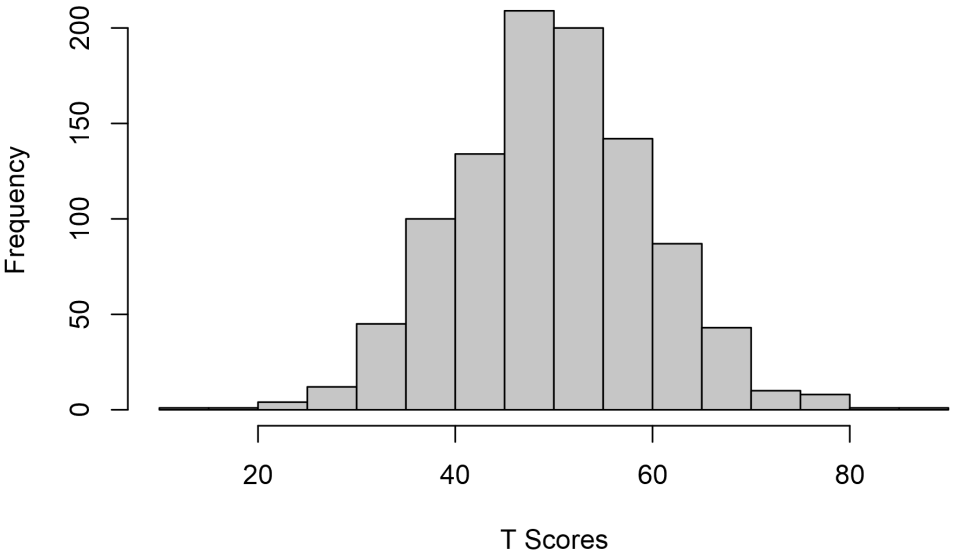


FIGURE 1.10 Histogram of T Scores.

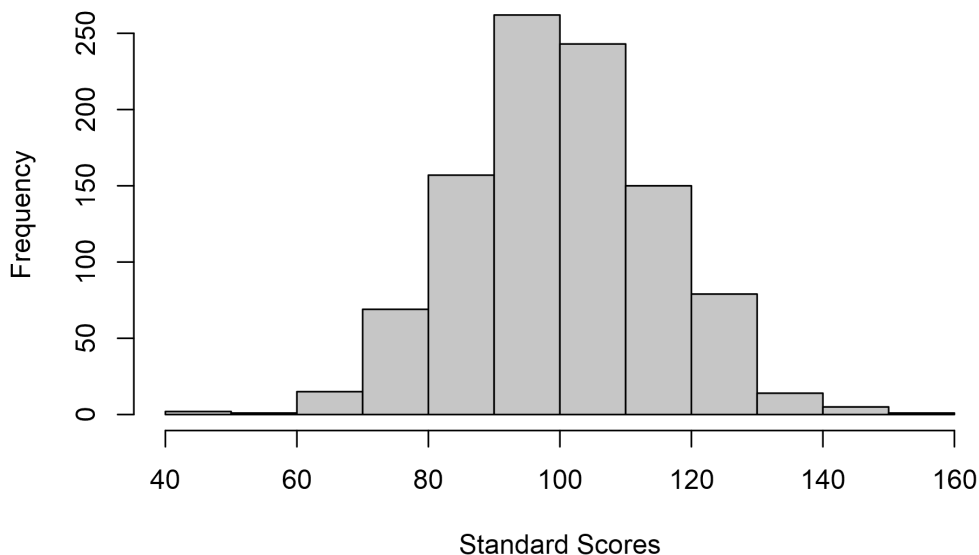


FIGURE 1.11 Histogram of Standard Scores.

TABLE 1.1 Calculating Stanine Scores

Stanine	Bracketed Percent	Cumulative Percent	z Score	Standard Score
1	4	4	below -1.75	below 74
2	7	11	-1.75 to -1.25	74 to 81
3	12	23	-1.25 to -0.75	81 to 89
4	17	40	-0.75 to -0.25	89 to 96
5	20	60	-0.25 to +0.25	96 to 104
6	17	77	+0.25 to +0.75	104 to 111
7	12	89	+0.75 to +1.25	111 to 119
8	7	96	+1.25 to +1.75	119 to 126
9	4	100	above +1.75	above 126

```
scores$scaledScore <- 10 + 3*scale(scores$rawData)
```

An example histogram of scaled scores is in [Figure 1.12](#).

1.3.2.2.5 Stanine scores

Stanine scores (short for STANdard Nine) have a mean of 5 and a standard deviation of 2. The scores range from 1–9. Stanine scores are calculated using the bracketed proportions in [Table 1.1](#). The lowest 4% receive a stanine score of 1, the next 7% receive a stanine score of 2, etc.

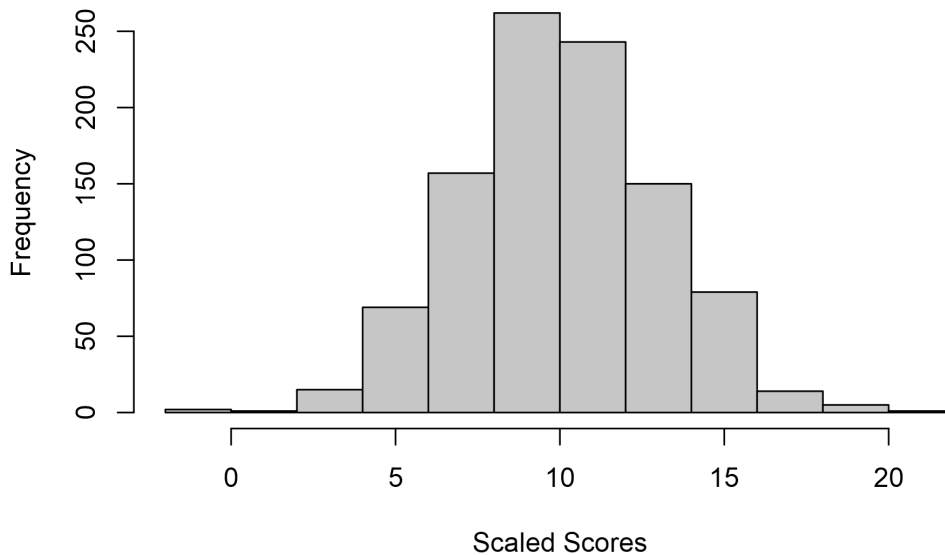


FIGURE 1.12 Histogram of Scaled Scores.

```
scores$stanineScore <- NA
scores$stanineScore[which(scores$percentileRank <= 4)] <- 1
scores$stanineScore[
  which(scores$percentileRank > 4 & scores$percentileRank <= 11)] <- 2
scores$stanineScore[
  which(scores$percentileRank > 11 & scores$percentileRank <= 23)] <- 3
scores$stanineScore[
  which(scores$percentileRank > 23 & scores$percentileRank <= 40)] <- 4
scores$stanineScore[
  which(scores$percentileRank > 40 & scores$percentileRank <= 60)] <- 5
scores$stanineScore[
  which(scores$percentileRank > 60 & scores$percentileRank <= 77)] <- 6
scores$stanineScore[
  which(scores$percentileRank > 77 & scores$percentileRank <= 89)] <- 7
scores$stanineScore[
  which(scores$percentileRank > 89 & scores$percentileRank <= 96)] <- 8
scores$stanineScore[which(scores$percentileRank > 96)] <- 9
```

A histogram of stanine scores is in [Figure 1.13](#).

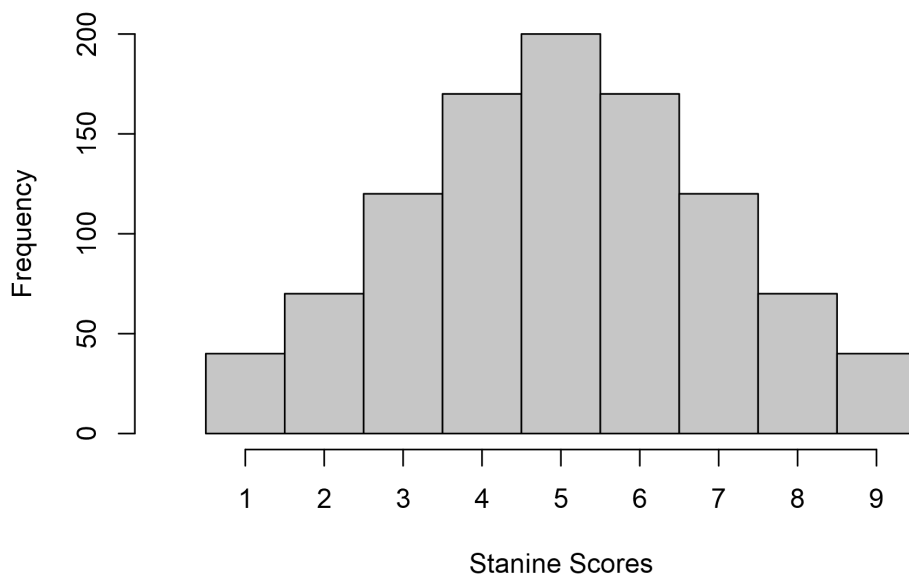


FIGURE 1.13 Histogram of Stanine Scores.

1.4 Conclusion

It is important to consider the types of data your data are because the types of data restrict what options are available to analyze the data. It is also important to consider whether the data were transformed because score transformations are not neutral—they can impact the results.

1.5 Suggested Readings

Childs et al. (2021): <https://dzchild.github.io/stats-for-bio/data-transformations.html>
(archived at <https://perma.cc/3YEB-QW2V>)