The Normal Curve and Standardization Notes[1]

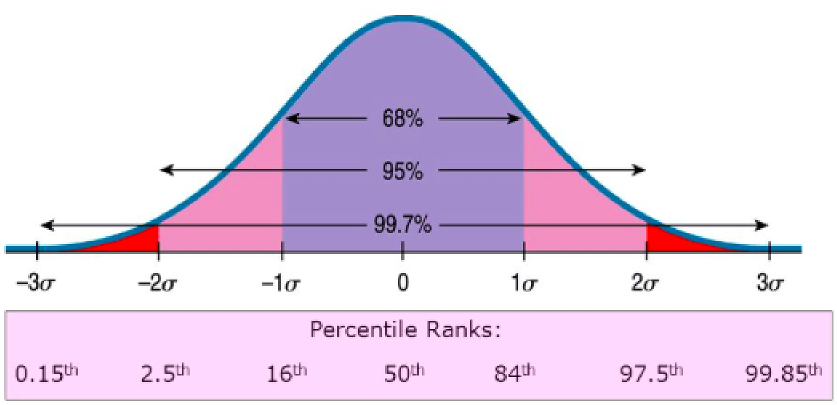
Stanford GSE Math Camp

Do Not Distribute Outside GSE

1 The Normal Curve

You've probably seen "bell curves" before. In Statistics, we call them normal models or normal curves. **Normal curves** are always unimodal and roughly symmetric. They also have other special characteristics.

- **Parameters** A normal model is specified by two parameters:
  - a mean, μ and a standard deviation, σ
  - The notation for the normal model looks like this:
    - $\sim$N(μ,σ)
- **The Empirical Rule** The empirical rule is a characteristics of the normal model and is also sometimes referred to as the 68-95-99.7 rule. These numbers relay what percentage of the data lies within one, two, and three standard deviations of the mean, respectively. We can also use the empirical rule to determine the percentile ranks for some values. *Note: A value outside of 2 sds of the mean is considered unusual.*



2 Standardized Scores

There are multiple types of standardized scores. We'll begin with "z-scores," as they are the standardized score associated with the normal model.

- A **z-score** measures the number of standard deviations a data value is away from the mean.
  - The formula to find a z-score is as follows:

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

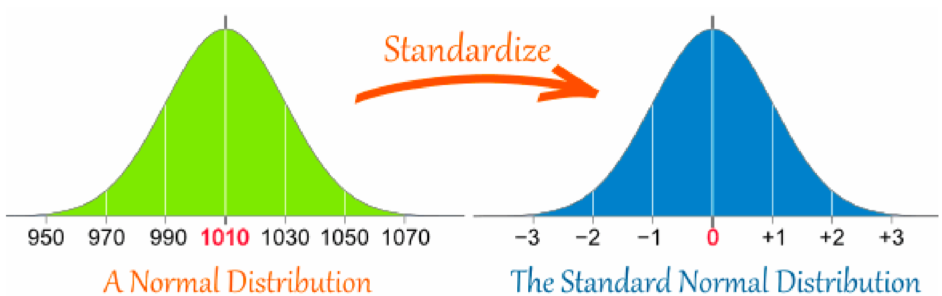
$$z = \frac{x - \bar{x}}{s}$$

[1]Contributor(s): Kelly Boles. If you find errors, please let us know so that we may correct them. Thanks!

- Utilizing standardized scores essentially allows us to compare apples to oranges. That is, comparing two seemingly unrelated observations, we can use their standardized scores to determine which is more unusual.

3 Standard Normal Distribution

- A **standard normal distribution** or a **standard normal model** will *always* have a mean of 0 and a standard deviation of 1.
- Standardizing shifts the distribution up or down, making the mean zero. (This is the numerator of the z-score formula). Then, it scales the spread, making the standard deviation one. (This is the denominator).
- Standardization does not change the shape of the distribution. Further, when we use the standard normal model, we must *start* with a normal model. That is, we should have a unimodal approximately symmetric distribution *before* we standardize scores. Otherwise, we are making unreasonable assumptions.
- When we check to make sure we have a normal distribution, this is called checking **conditions**. Specifically, when we check to make sure we can use a normal model to represent our data, we're checking the **Nearly Normal Condition**. An easy way to check this is to make a histogram or dotplot and look at the shape of the distribution. *Note: There are more requirements than just shape to fit a normal model to a distribution, but we'll investigate this more later in the week.*



4 Finding Normal Percentiles

- Earlier, we could use the empirical rule to find some percentiles, but only if a value happened to fall neatly at one, two, or three standard deviations from the mean. However, we can find the percentile for *any* value using its z-score and the standard normal distribution.
- To find the percentile, we'll need to use <u>technology</u> or a <u>z-score table</u>.

5 Reversing the Process: Going from Percentiles to Scores

- We can also use the z-table or technology to go from percentages to raw scores.
- To do this, we'll first utilize the table and then solve for the raw score within the z-score formula.

6 Student's or Gosset's t

- Tomorrow, we'll consider when the normal model fails to well-represent unimodal, approximately symmetric data distributions. For now, it's enough to know that another family of distributions, called t distributions exist. They are particularly helpful when working with small sample sizes and/or when we don't know the population standard deviation (which is often!).
- There are different tables we'll need to use when working with the family of t distributions.