PSY 3307

Z-Scores

Jonathan A. Pedroza, MS, MA

Cal Poly Pomona

2021-09-02

z-Scores

- **relative standing**
- **z-score** is a statistic that indicates the distance a score is from the mean when measured in standard deviation units
    - positive values are above the mean
    - negative values are below the mean
- z-score is the location on a distribution as well as the distance from the mean

z-score Example

- Get a score of 90 to the mean of 60
  - deviate of 30
- Standard deviation of 10
- Now means that the +30 score is 3 SD above the mean +30 score doesn't tell us much +3 sd tells us much more
  - also tells us that the z-score is 3

Calculating the z-score

- For a sample

$$z = \frac{X - \overline{X}}{S_X}$$

- For a population

$$z = \frac{X - \mu}{\sigma_X}$$

Calculating the raw score from a z-score

$$X = (z)(S_X) + \overline{X}$$

```r
z = 1.4
S_X = 1.14
mean = 3.1

(1.4)*(1.14) + 3.1
```

```
## [1] 4.696
```

z-distribution to interpret scores

- **z-distribution** is transforming all raw scores into z-scores
- a z-score of 0 is the mean
- the z-distribution also represents that everything has been standardized
  - every variable is now comparable and on the same scale

Some Reminders

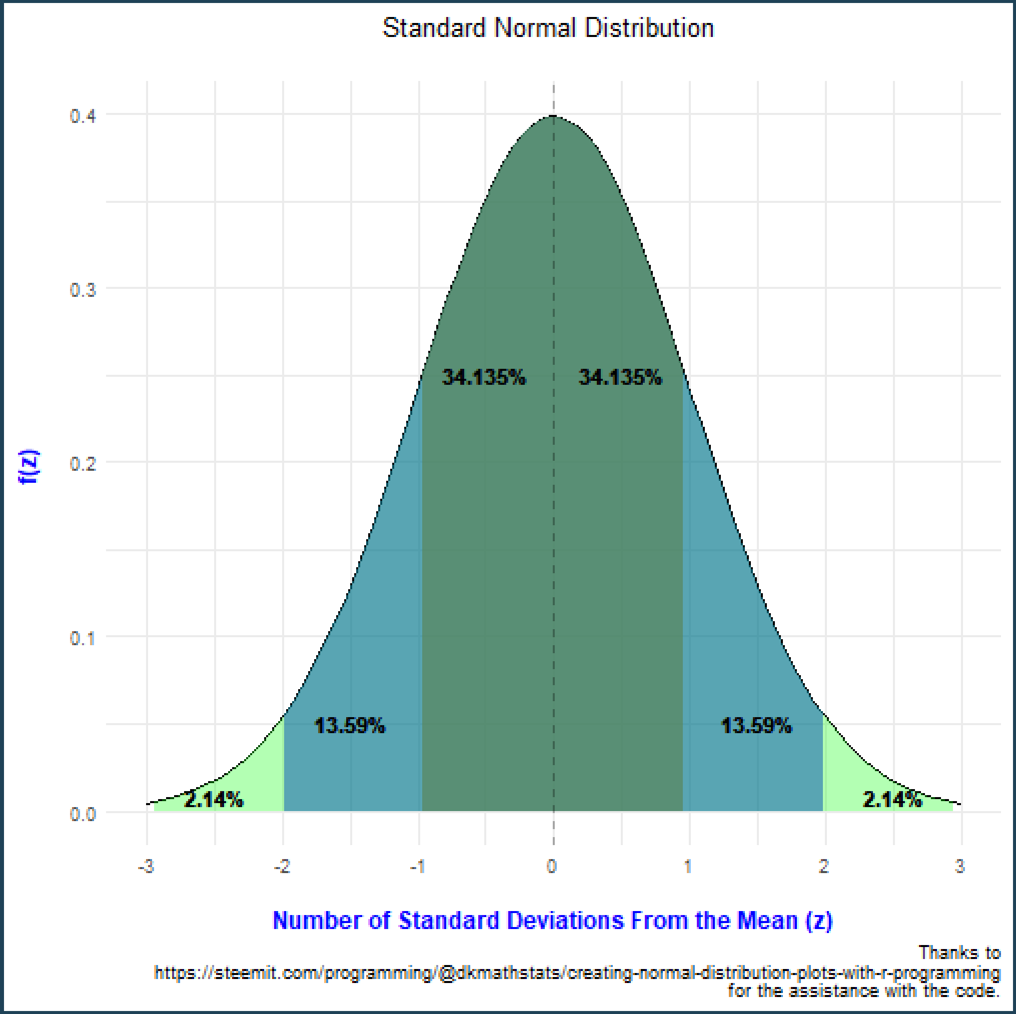
- A z-distribution will always have the same shape as the raw score distribution
- the mean is and will always be zero
- the standard deviation is 1, even when the standard deviation of the raw scores is a different value (e.g., 10, 15, 100)
- z-scores that are greater in either the positive or negative direction mean that values are less likely to occur

Comparing Apples to Oranges

- When using the z-distribution, every variable is put on the same scale
- often referred to as **standardized scores**

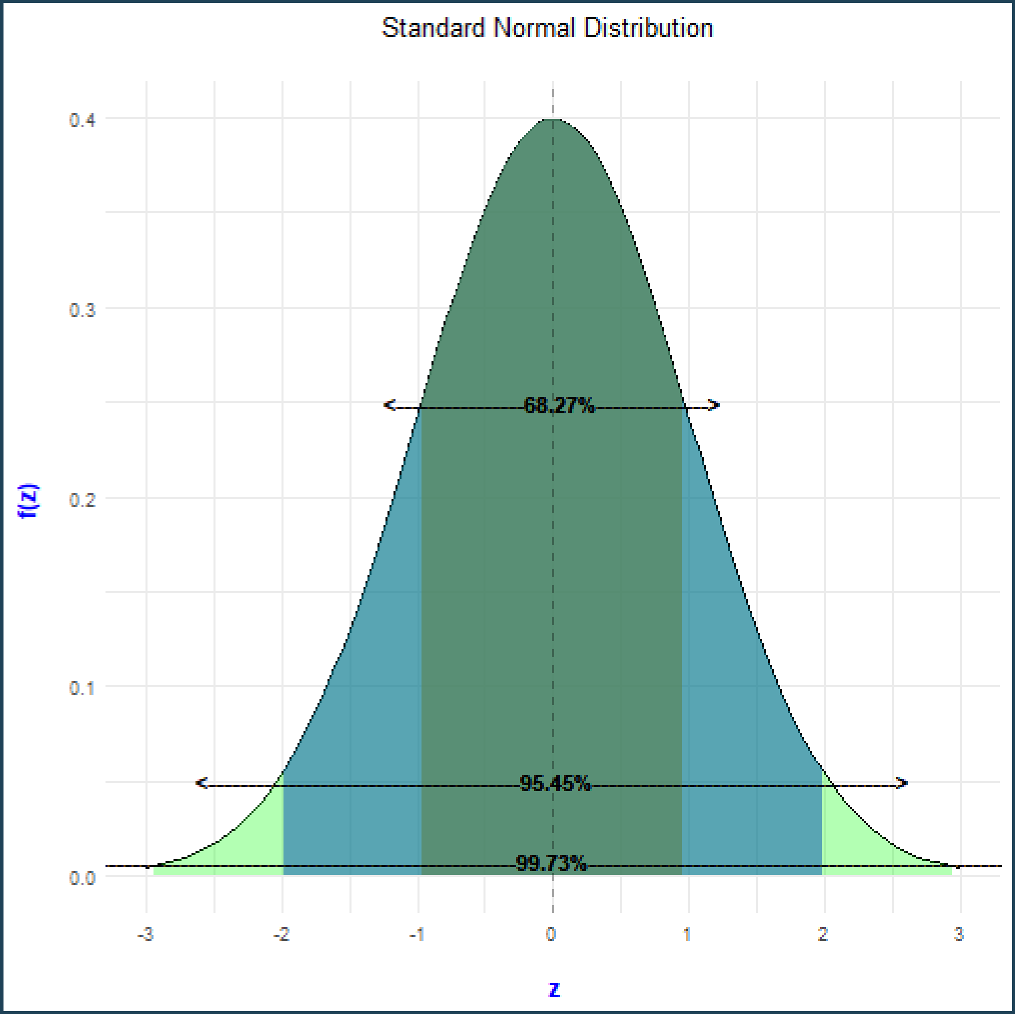
z-distributino for Relative Frequency

- relative frequency of z-scores will be the same on all normal z-distributions
- Knowing that +1SD is above the mean, whatever raw scores fall within this part of the distribution will occur 34% of the time


Standard Normal Distribution
0.4
0.3
0.2
0.1
0.0
f(z)
34.135%
34.135%
13.59%
13.59%
2.14%
2.14%
-3
-2
-1
0
1
2
3
Number of Standard Deviations From the Mean (z)
Thanks to
https://steemit.com/programming/@dkmathstats/creating-normal-distribution-plots-with-r-programming
for the assistance with the code.

Standard Normal Curve

- **Standard Normal Curve** is a perfect normal z-distribution that serves as the model of any approximately normal z-distribution
- Approach is most common with
  - large sample (or population)
  - interval or ratio scores
  - come close to forming a normal distribution
- While z-scores past +-3SD are possible, they occur .0013 of the time, which is why we often only look at +-3SD


Standard Normal Distribution
68.27%
95.45%
99.73%
f(z)
z
0.4
0.3
0.2
0.1
0.0
-3
-2
-1
0
1
2
3

Relative Frequency Using the z-Distribution

- Calculate the z-score for one observation
- Multiply by N
- Answer will be the number of observations between the mean and the z-score you calculated
- Then you can add up the other half to see all the observations that are at or below the score that the one observation had

Example

```r
z_score = -1

N = 100

#
.3413*100
```

```
## [1] 34.13
```

```r
50 - 34.13
```

```
## [1] 15.87
```

```r
#15.87% of scores were at or below the z-score we calculated
```

Using the z-Table

- In your Book at the end, you'll see the z-distribution
- There are two ways to show this
    - Area between mean and the z-score along with area beyond z in tail
    - z-scores with the tenths position on the column and the hundredths position on the top row

Using z-Scores to Describe Sample Means

- **Sampling Distribution of Means** is the frequency distribution showing all possible sample means when samples are drawn from a population
- Every time a population is sampled from, they will always be slightly different, with some means being higher and some being lower
- **Central Limit Theorem** statistical concept that defines the mean, standard deviation, and shape of a sampling distribution
  - Always follows some rules
  - A sampling distribution is always an approximately normal distribution
  - mean of sampling distribution equals mean of underlying raw score population used to create the sampling distribution
    - \mu is the mean of the means
  - standard deviation of the sampling distribution is related to the standard deviation of the raw score population

Central Limit Theorem

- The use of this concept is that we can assume that our sample is representative of the population without having to sample the whole population

The Standard Error of the Mean

$$ \sigma_{\overline{X}} = \frac{\sigma_{X}}{\sqrt{N}} $$

- **Standard Error of the Mean** is the standard deviation of the sampling distribution of means
- $\sigma_{\overline{X}}$ will be used to describe the true standard error of the mean
    - which is the population standard deviation of the population of sample means

Example

```r
# population standard deviation of raw scores is 15
# population mean/mu is 25
sigma_raw = 15

N = 100

15/sqrt(N)
```

```
## [1] 1.5
```

```r
# 1.5
```

- Out of the sampling distribution, the individual sample means differ from the population mean of 25 by an average of 1.5 points

z-Score from a Sample Mean

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

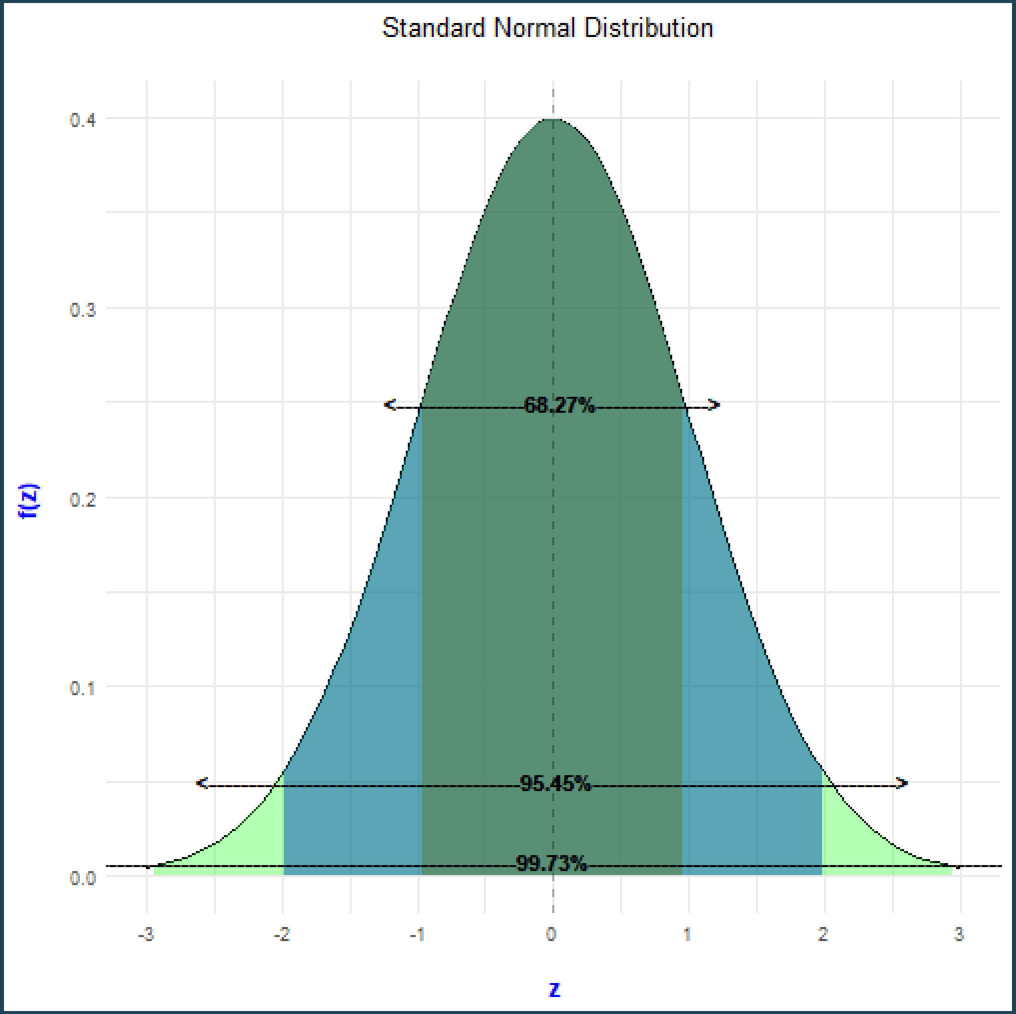
- Using information from the previous slide
  - \mu = 25
  - \overline{X} = 21
  - \sigma_{\overline{X}} = 1.5

```
(21 - 25)/1.5
```

```
## [1] -2.666667
```

Relative Frequency of Sample Means

- Because a sampling distribution is always an approximation of a normal distribution, transforming all sampling means into z-scores means it is a normal z-distribution


Standard Normal Distribution
0.4
0.3
0.2
0.1
0.0
f(z)
68.27%
95.45%
99.73%
-3
-2
-1
0
1
2
3
z