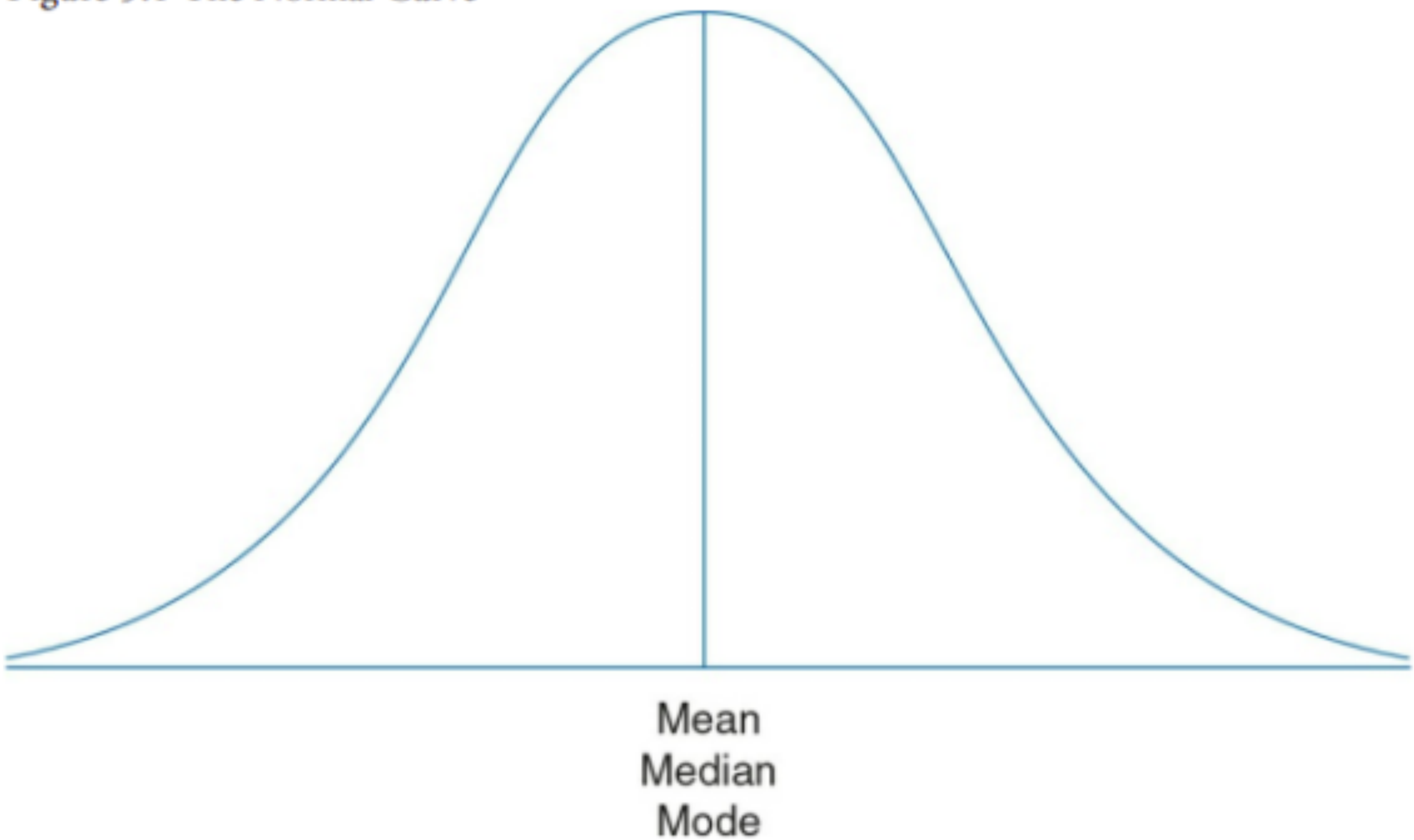
Chapter 5 - The Normal Distribution

Normal Distribution Definition

Definition of Normal Distribution:

• A bell-shaped & symmetrical theoretical distribution w/the mean, median & mode all at the peak & with the frequencies gradually decreasing at both ends of the curve.

Figure 5.1 The Normal Curve



•

• Notice that **most of the observations are clustered around the middle** with the **frequencies gradually decreasing at both ends of the distribution**.

Properties

Normal Distribution Properties:

• Perfectly symmetric (see page before this for a normal curve)

 ▪ Since it's perfectly symmetric, precisely half the observations fall on each side of the middle of the distribution.

• The mid point is the max frequency

 ▪ The peak is also where the **three** measures are:

  → the **mode**

  → the **median**

  → the **mean**

 ▪ The frequencies gradually decrease at both ends of the curve.

• The normal curve is pure theory, real-life distributions never match this model perfectly.

 ▪ so if we say a distribution is normal, we mean the distribution closely resembles this theoretical curve.

• Like an empirical distribution (which is based on real data), a theoretical distribution can be organized into frequency distributions, displayed using graphs, and described by its central tendency and variation using measures such as the mean & standard deviation

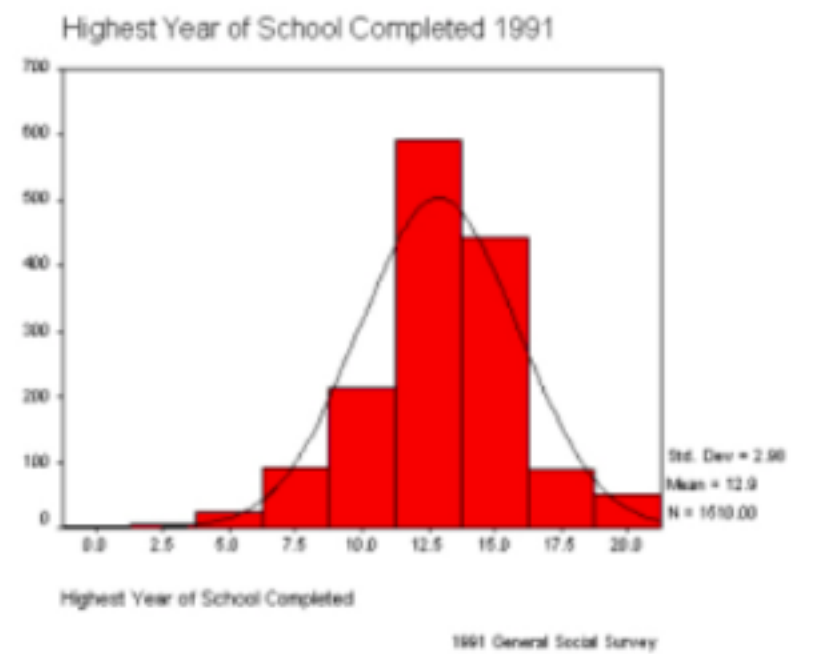
 ▪ But unlike an empirical distribution, a theoretical distribution is based on theory rather than real data.

Practice

Final Grades in Social Statistics of 44 Stude			
Midpoint Score	Frequency Bar Chart	Freq.	%
50	*****	5	11
60	**********	10	22
70	***************	15	33
80	**********	10	22
90	*****	5	11

- The above is an example of a 'normal' distribution
    - Look at the bar chart & calculate the **mean, median, & mode**:
        - → Calculate the **mean** by multiplying each score by the frequency & adding up all the values
        - → Calculate the **median** by looking for the middle value in the freqeuncy bar chart
        - → Calculate the **mode** by looking for the score that has the most frequencies.
- Mean - 70
- Median - 70
- Mode - 70
    - Since all three measures of central tendency are equal & this is reflected in the bar graph, we can assume that this data is "**normally distributed**"
    - Also, since the median = mean, we know that the distribution is **symmetrical**

Practice

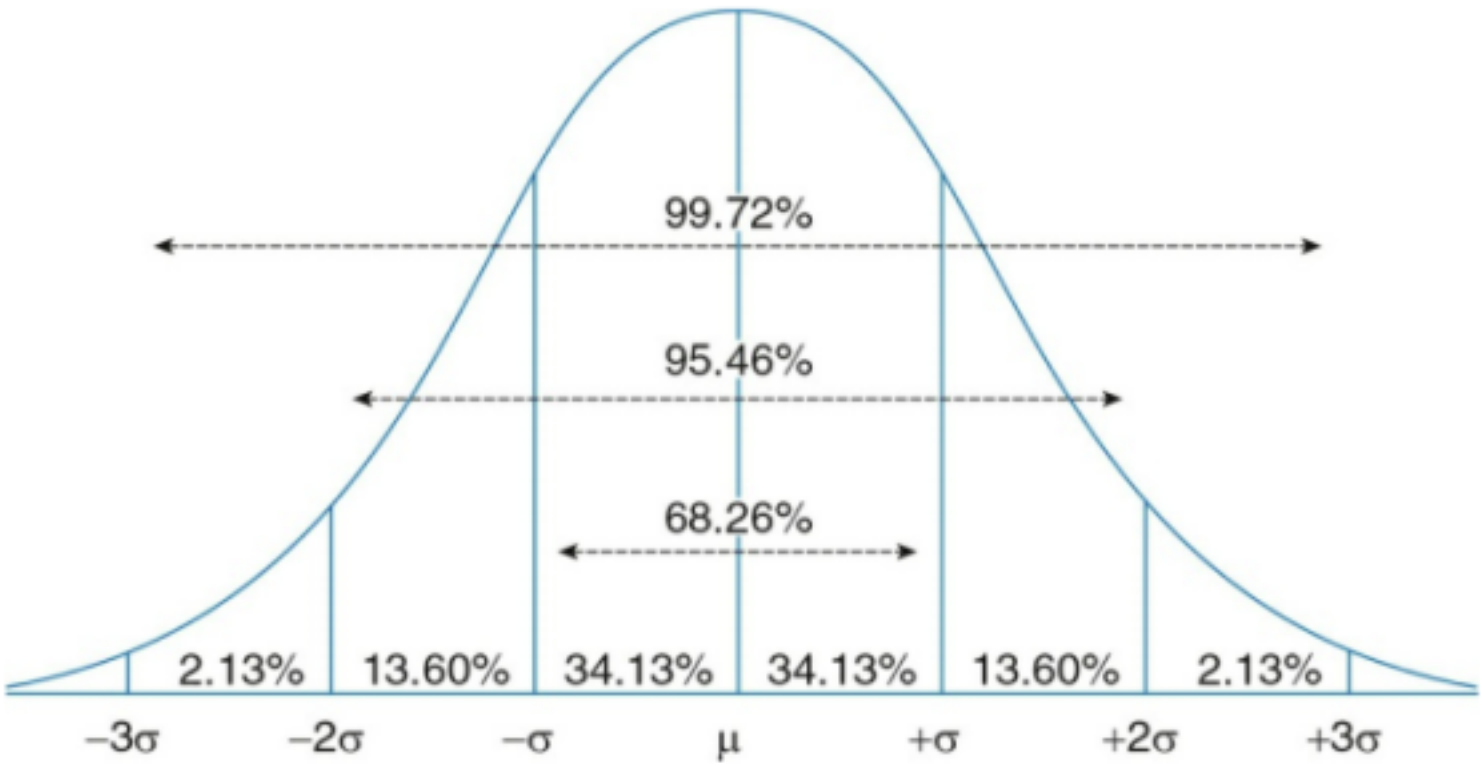


- Some shapes of normal curves in graphs are tall & thin or short & wide
- **All normal distributions are wider in the middle & symmetrical**

Areas Under Normal Curve

Areas Under Normal Curve:

• The area under the normal curve can be thought of as a proportion or percentage of the number of observations in the sample



- μ = Mean
- $-\sigma$ = 1 standard deviation below mean
- $+\sigma$ = 1 standard deviation above mean

■ So between the mean and above or below 1 standard deviation, 68.26% of all observations in the distribution occur

■ Between the mean and above or below 2 standard deviations, 95.46% of all observations in the distribution occur.

■ Between the mean and above or below 3 standard deviations, 99.72% of all observations in the distribution occur.

Interpreting Standard Deviation

Interpreting Standard Deviation:

• The relationship between the distance from the mean & the areas under the curve represents a property of the normal curve that has highly practical applications.

- ▪ As long as the distribution is normal & we know the mean and the standard deviation, we can **determine the proportion or percentage of cases that fall between any score & the mean.**
- ▪ For empirical distributions, when we know the mean & the standard deviation, we can determine the percentage or proportion of scores that are within any distance from that distribution's mean. (measured in standard deviation units)
- ▪ Note that this relationship between the distance from the mean & the areas under the curve only applies to normal or approximately normal distributions.

Transforming Raw Score into Z Score

Transformation:

- A raw score can be turned into a *Z score* (aka a *standard score*) in order to see how many standard deviations it is above or below the mean
  - **Standard (Z) score** - The number of standard deviations that a given raw score is above or below the mean.
- Steps to transform a raw score into a Z score:
  1. Raw score - Mean
  2. Divide the difference by the standard deviation
  3. You have your Z score
  - Formula: $$Z = \frac{Y - \bar{Y}}{s}$$
- The Z score tells us how far a given raw score is from the mean in stsandard deviation units.
  - A positive Z score says the score > mean
  - A negative Z score says the score < mean
  - The larger the Z score, the larger the difference between the score & the mean.

Transforming Z Score into Raw Score

Transformation:

- If you have a Z score but need to know the raw score, you can go backwards
- Formula:
  - $$Y = \bar{Y} + Z(S_y)$$
  - $\bar{Y}$ = mean
  - Z = Z score
  - (S_y) = standard deviation

Standard Normal Distribution

Standard Normal Distribution:

• When a normal distribution is represented in standard scores (Z scores), it's called the standard normal distribution

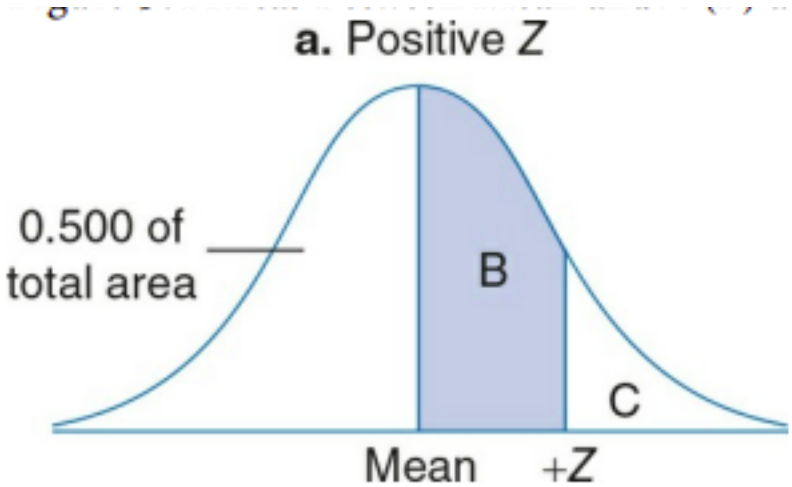
▪ Z scores are the #s that tell us the distance between an actual score & the mean in terms of standard deviation units.

• The standard normal distribution has a mean of 0.0 & a standard deviation of 1.0

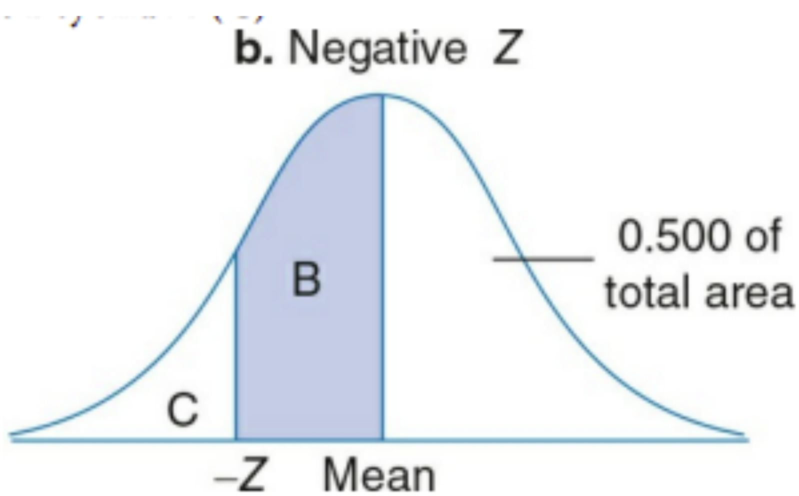
Standard Normal Table

Standard Normal Table:

• This is a table showing the area (as a proportion, which can be translated into a percentage) under the standard normal curve corresponding to any Z score or its fraction.
• Z scores can be used to determine the proportion of cases that are included between the mean & any Z score in a normal distribution.



•



•

▪ *Column B* shows the area included between the mean and the Z score listed in Column A.
→ When Z is positive, the area is located on the right side of the mean wheras for a negative Z score, the same area is located on the left side of the mean.
▪ *Column C* shows the proportion of the area that is beyond the Z score listed in Column A.
→ Areas corresponding to positive Z scores are on the right side of the curve whereas areas corresponding to negative Z scores are identical except they're on the left side of the curve.

Finding Area Between Mean & Positive or Negative Z Score

- The standard normal table can be used to find the area between the mean & specific Z scores
- The standard normal table is located in Appendix B of the book for reference
- To find the area between a mean of 475 and a raw score of 675, follow these steps:

  1> Convert 675 to a Z score

  2> You get 1.83

  3> Search for 1.83 in Appendix B

  4> Multiply 0.4664 by 100 to get 46.64%

  5> 46.64% of the total area lies between 475 and 675

  6> Say you wanted to find the actual number of students who scored between 475 & 675, multiply the proportion 0.4664 by the total number of students. So say 1,108,165 students, you'd multiply 0.4664 & 1,108,165 which would equal 516,848 students scoring between 475 & 675.

- Now say you wanted to find the area for a score lower than the mean, indicating the z score would be negative:
  - The score being 305, the mean being 475
  - You first find the Z score which is -**1.56**
  - Since the proportions that correspond to positive Z scores are identical to those corresponding to negatives, ignore the negative sign & look it up like normal & proceed as if it wasn't a negative Z score.

Finding Area Above a Positive Z Score or Below a Negative Z Score

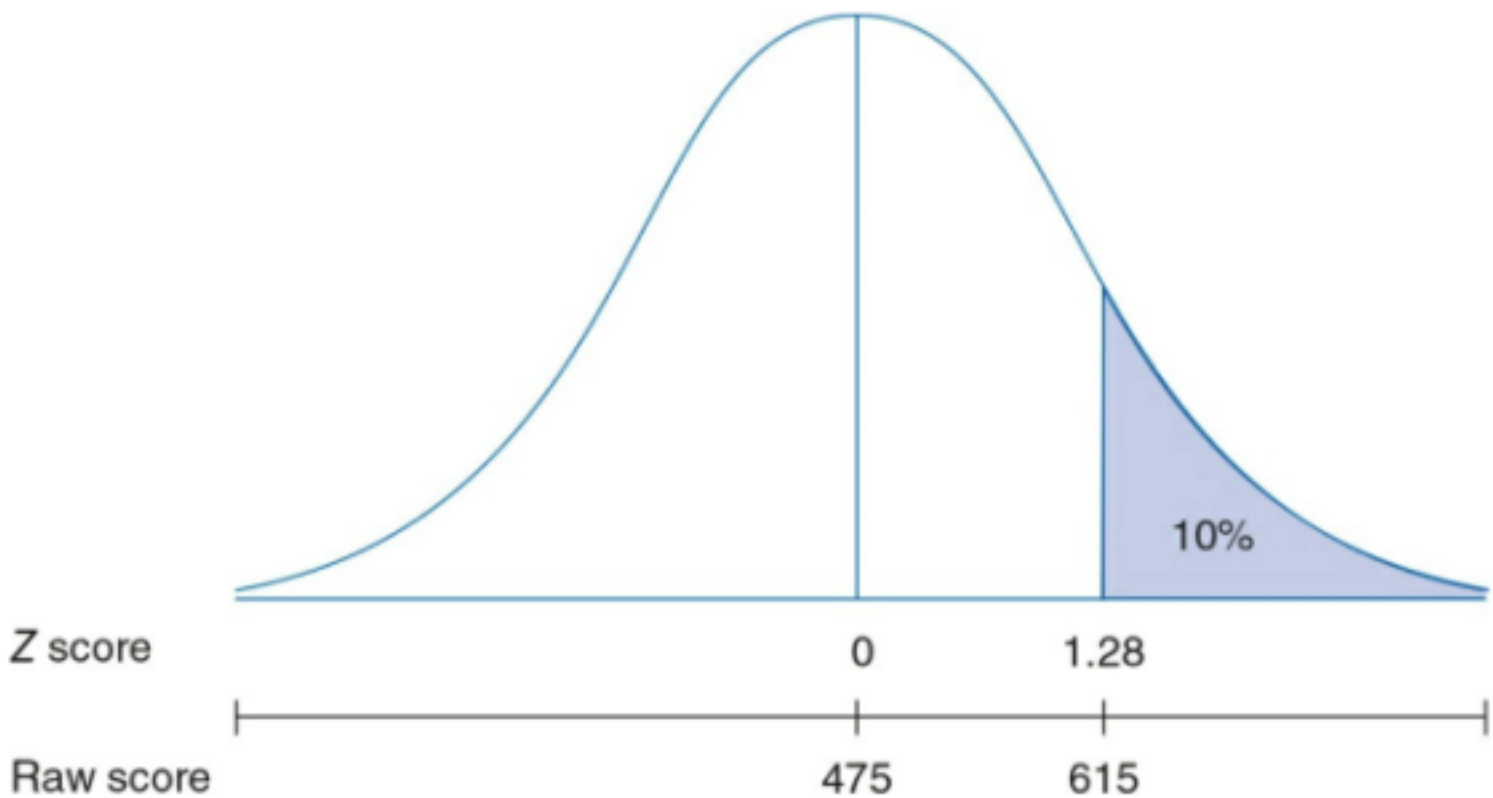
- The normal distribution table in appendix B could be used to find the area beyond a Z score too

Transforming Proportions & Percents into Z Scores

Turning percentile into raw score

Finding Z Score Which Bounds an Area Above It:

- Find the Z score for the percentile you're searching for (if it's the 20th percentile, you find the Z score corresponding to .20 in Appendix B column C or whatever is closest to .20)
- After you have the Z score, use this formula to find the raw scoe:
  - $Y = \bar{Y} + Z(s)$
  - Where Y is the raw score you're searching for
  - $\bar{Y}$ = mean
  - Z = Z score
  - S = standard deviation
- By finding the percentile, you're able to identify the cut off point
  - **EX**: Say you were attempting to find the cut off point for the top 10% of an SAT exam. You'd find that the Z score is 1.28 & multiply it by a mean of let's say 475 and a standard deviation of 109 & plug these values into the equation to find that the cut off point for the top 10% of the SAT exam takers is 615.



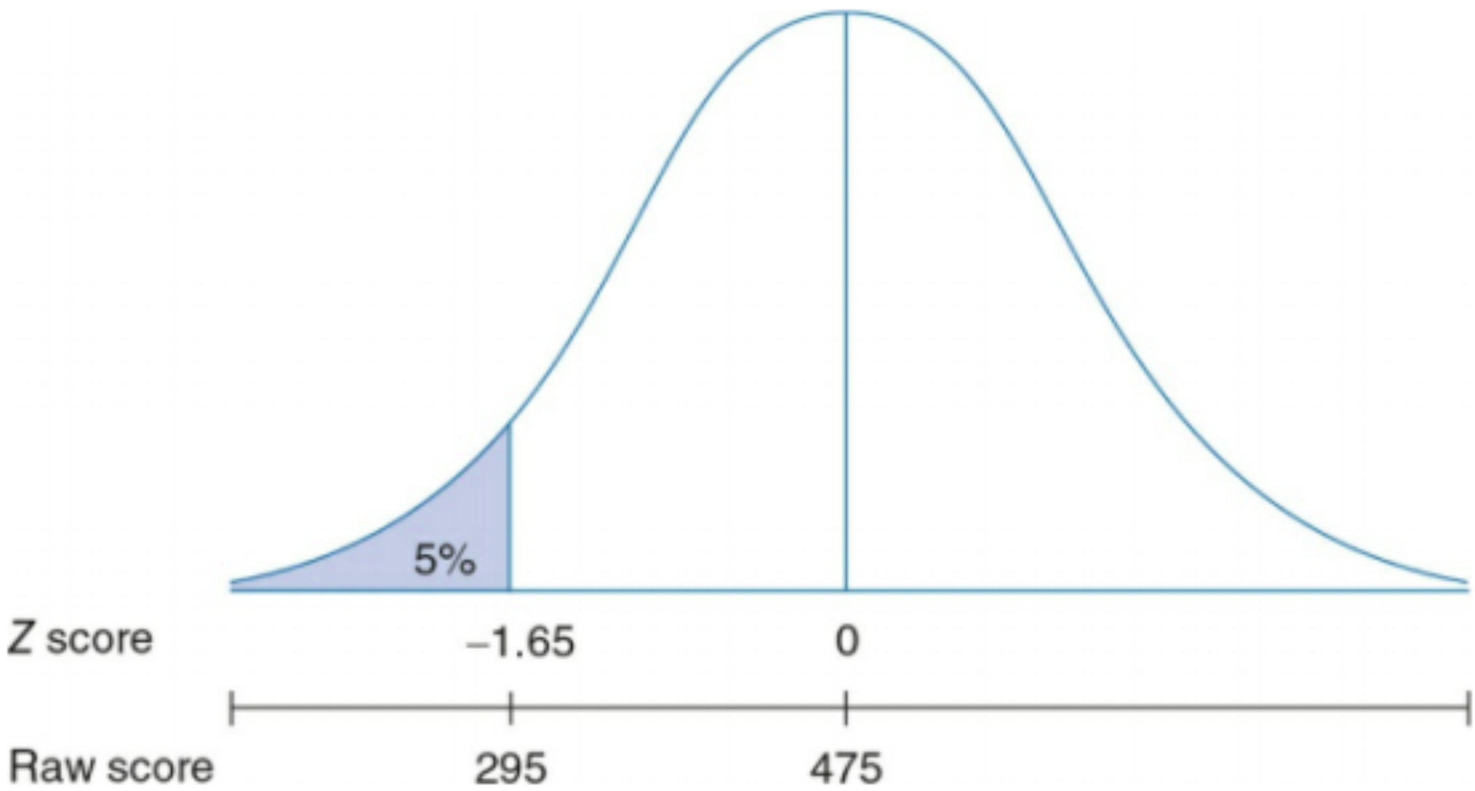
Finding a Z Score Which Bounds an Area Below It:

- Let's say you wanted to find the score that corresponds to the bottom 5% of test takers:

  1- Find the Z score that corresponds to 0.05 or the closest to it in Column C. Then locate the Z in Column A that corresponds to this proportion. In this case, it'd be 1.65 & **since the area we're looking for is on the left side of the curve (below the mean), the Z score is negative.** So the Z score associated with 0.05 is -1.65.

  2- Use the same equation as above to transform a Z score into a raw score.

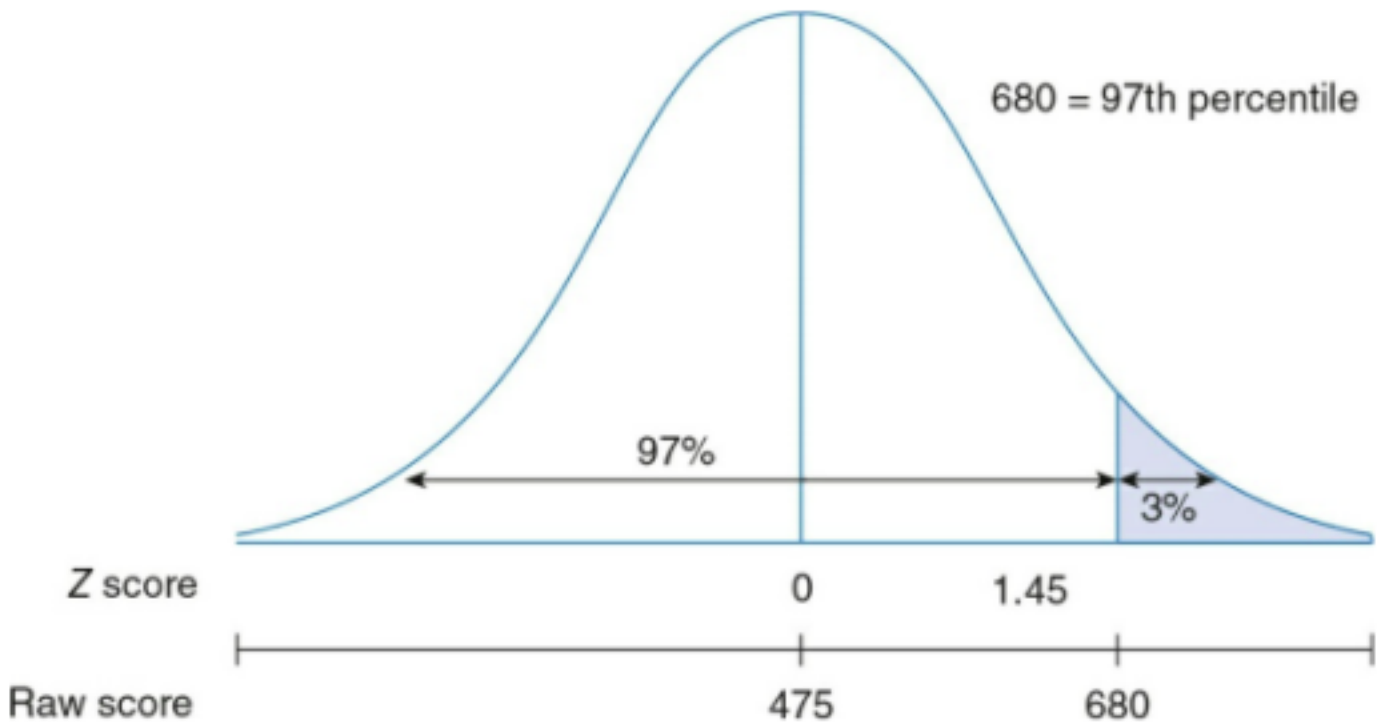
  3- If you use the values 475 for the mean, 109 for the standard deviation, and -1.65 then you'd find the raw score to be 295.
- The cutoff for the lowest 5% of SAT scores is 295.


5%
Z score
−1.65
0
Raw score
295
475

Working w/Percentiles in Normal Distribution

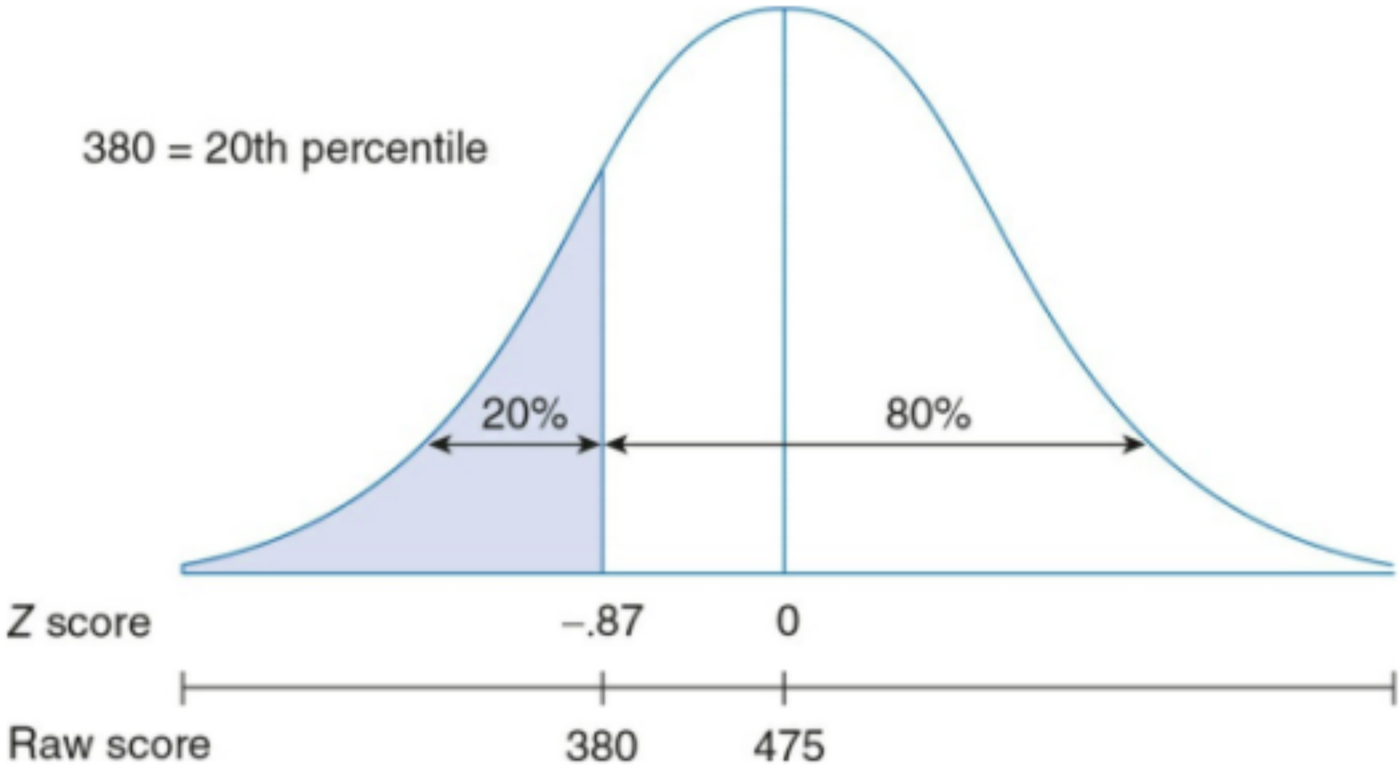
Finding Percentile Rank of a Score Higher Than the Mean:

• Say you took the SAT & scored a 680 but how did you do relative to other students who took the exam?
• To find the percentile rank of a score higher than the mean, follow:
  1- Convert the raw score to a Z score using the formula from earlier (or the script)
  2- Find the area beyond Z in Appendix B Column C
  3- Subtract the area from 1.00 & multiply by 100 to get the percentile rank
  ▪ In this case, you'd find that the Z score is 1.88 and the area is 0.0301 and by using the formula to find the percentile rank, you get 97% which tells you that 97% of all test takers scored lower than 680 & 3% scored higher than 680.
  ▪ Note, the mean & standard deviation used in this example are 475 & 109 respectively.



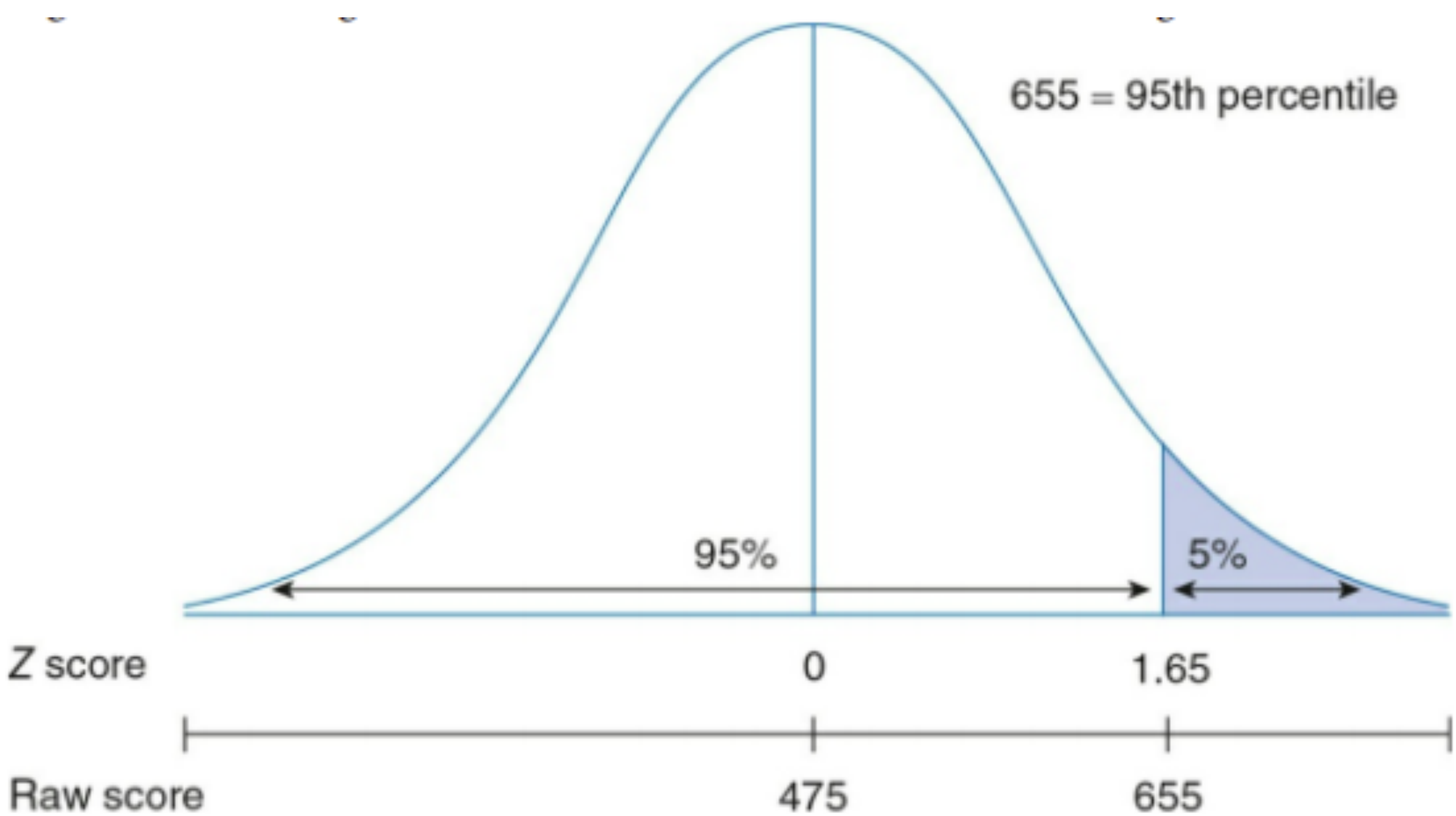
Finding Percentile Rank of a Score Lower Than the Mean:

• If your SAT score is 380 (which is below the mean of 475), what's your percentile rank?
1. Convert the raw score of 380 to a Z score
  ▪ The Z score would be -0.87
2. Find the area beyond Z in Appendix B Column C, the area beyond a Z score of -0.87 is 0.1992.
3. Multiply the area by 100 to obtain the percentile rank
• In this case, it's the 20th percentile rank which means that 20% of all test takers scored lower than you but 80% scored the same or higher.



Finding the Raw Score Associated w/a Percentile Higher Than 50:

- Let's assume a uni only admits students who score at or above the 95th percentile in an SAT exam. What is the cut off point required for acceptance?
- To find the score associated w/a percentile higher than 50, follow:
    1- Divide the percentile by 100 to find te area below the percentile rank (0.95 in this case)
    2- Subtract the area below the percentile rank from 1.00 to find the area above the percentile rate (0.05 in this case, meaning 0.05 get accepted)
    3- Find the Z score associated w/the area above the percentile rank. Refer to Appendix B in Column C. Then locate Column A which corresponds with the proportion (in this case, 1.65).
    4- Convert the Z score to a raw score:
      - Y = Mean + Z score(standard deviation)
    5. In this case, it'd be 654.85
      ▪ The final SAT associated with the 95th percentile is 654.85 which means the lowest score you can get to be admitted is 654.85



Finding Raw Score Associated w/a Percentile Lower Than 50:

1. Divide the percentile by 100 to find the area below the percentile rank
2. Find the Z score associated w/this area. Refer to Appendix B in Column C then locate the Z in column A that corresponds to this proportion.
3. Convert the Z score to a raw score using the same formula as in the previous sect.

• Recall that since it'd be lower than the mean, the Z score would be negative