Lecture Notes

# Chapter 5: The Normal Distribution

## Learning Objectives

1. Explain the importance and use of the normal distribution in statistics.
2. Describe the properties of the normal distribution.
3. Transform a raw score into standard (*Z*) score and vice versa.
4. Transform a *Z* score into proportion (or percentage) and vice versa.
5. Calculate and interpret the percentile rank of a score.

## Chapter Outline

1. Introduction
   1. The distributions that we have described so far are all empirical distributions. They are all based on real data.
   2. The **normal distribution** is a bell-shaped and symmetrical theoretical distribution with the mean, median, and mode all coinciding at its peak and with the frequencies gradually decreasing at both ends of the curve.
   3. Similar to an empirical distribution in that it can be
      1. organized into frequency distributions.
      2. displayed using graphs.
      3. described by its central tendency and variation using measures such as the mean and the standard deviation.
   4. Many empirical distributions seem to approximate the value of the theoretical normal distribution.
   5. Learn a lot about the characteristics of empirical distributions based on our knowledge of the theoretical normal distribution.
2. Properties of the Normal Distribution
   1. The normal curve is bell shaped and shows perfect symmetry. This means that precisely half the observations fall on each side of the middle of the distribution.
   2. The midpoint of the normal curve is the point having the maximum frequency.
   3. This is also the point at which three measures coincide:
      1. The mode (the point of the highest frequency).
      2. The median (the point that divides the distribution into two equal halves).
      3. The mean (the average of all the scores).
   4. Most of the observations are clustered around the middle, with the frequencies gradually decreasing at both ends of the distribution.
   5. Empirical Distributions Approximating the Normal Distribution
      1. The normal curve is a theoretical ideal, and real-life distributions never match this model perfectly.
      2. Researchers study many variables that closely resemble this theoretical model.
      3. When a variable is “normally distributed,” the graphic display will reveal an approximately bell-shaped and symmetrical distribution closely resembling the idealized model.
      4. This property makes it possible for us to describe many empirical distributions based on our knowledge of the normal curve.
   6. Areas Under the Normal Curve
      1. In all normal curves, a constant proportion of the area under the curve lying between the mean and any given distance from the mean when measured in standard deviation units.
      2. The area under the normal curve may be conceptualized as a proportion or percentage of the number of observations in the sample.
      3. Thus, the entire area under the curve is equal to 1.00 or 100% (1.00 × 100) of the observations.
      4. Because the normal curve is perfectly symmetrical, exactly 0.50% or 50% of the observations lie above or to the right of the center, which is the mean of the distribution, and 50% lie below or to the left of the mean.
      5. The plus signs indicate standard deviations above the mean; the minus signs denote standard deviations below the mean.
   7. Interpreting the Standard Deviation
      1. The fixed relationship between the distance from the mean and the areas under the curve represents a property of the normal curve that has highly practical applications.
      2. As long as a distribution is normal and we know the mean and the standard deviation, one can determine the proportion or percentage of cases that fall between any score and the mean.
      3. This property provides an important interpretation for the standard deviation of empirical distributions that are approximately normal.
      4. For such distributions, knowing the mean and the standard deviation, one can determine the percentage or proportion of scores that are within any distance, measured in standard deviation units, from that distribution’s mean.
      5. Not every empirical distribution is normal.
3. An Application of the Normal Curve
   1. The current SAT includes two components:
      1. Evidence-Based Reading and Writing (ERW).
      2. Mathematics.
   2. The results of the SAT exam, combined or for each component, are assumed to be normally distributed.
   3. Throughout this chapter, one will use the normal (theoretical) curve to describe and better understand the characteristics of the SAT ERW empirical (real data) distribution.
   4. Transforming a Raw Score Into a *Z* Score
      1. The difference between any score in a distribution and the mean in terms of standard scores also known as *Z* scores.
      2. A **standard (*Z*) score** is the number of standard deviations that a given raw score (or the observed score) is above or below the mean.
      3. A raw score can be transformed into a *Z* score to find how many standard deviations it is above or below the mean.
      4. To transform a raw score into a *Z* score, divide the difference between the score and the mean by the standard deviation.
      5. This calculation gives us a method of standardization known as transforming a raw score into a *Z* score (also known as a standard score).
      6. The *Z*-score formula: .
      7. A *Z* score allows us to represent a raw score in terms of its relationship to the mean and to the standard deviation of the distribution.
      8. It represents how far a given raw score is from the mean in standard deviation units.
      9. A positive *Z* indicates that a score is larger than the mean, and a negative *Z* indicates that it is smaller than the mean.
      10. The larger the *Z* score, the larger the difference between the score and the mean.
4. The Standard Normal Distribution
   1. When a normal distribution is represented in standard scores (*Z* scores), we call it the **standard normal distribution**.
   2. Standard scores, or *Z* scores, are the numbers that tell us the distance between an actual score and the mean in terms of standard deviation units.
   3. The standard normal distribution has a mean of 0.0 and a standard deviation of 1.0.
   4. To understand the relationship between raw scores of a distribution and their respective standard *Z* scores, the SAT ERW scores that correspond to these standard scores are shown:
      1. For example, the mean for the SAT ERW distribution is 536 and the corresponding *Z* score, the mean of the standard normal distribution is 0.
      2. As calculated, the score of 638 is 1 standard deviation above the mean (536 + 102 = 638); therefore, its corresponding *Z* score is +1.
      3. The score of 434 is 1 standard deviation below the mean (536 – 102 = 434), and its *Z*-score equivalent is −1.
5. The Standard Normal Table
   1. The areas or proportions under the standard normal curve, corresponding to any *Z* score or its fraction, are organized into a special table called the **standard normal table**.
   2. The table consists of three columns:
      1. Column A lists positive *Z* scores:
         1. Because the normal curve is symmetrical, the proportions that correspond to positive *Z* scores are identical to the proportions corresponding to negative *Z* scores.
      2. Column B shows the area included between the mean and the *Z* score listed in Column A:
6. When *Z* is positive, the area is located on the right side of the mean.
7. For a negative *Z* score, the same area is located left of the mean.
   * 1. Column C shows the proportion of the area that is beyond the *Z* score listed in Column A:
8. Areas corresponding to positive *Z* scores are on the right side of the curve.
9. Areas corresponding to negative *Z* scores are identical except that they are on the left side of the curve.
   1. Four sections form examples of how to transform *Z* scores into proportions or percentages to describe different areas of the empirical distribution of SAT ERW scores.
   2. Finding the Area Between the Mean and a Positive or Negative *Z* Score
      1. We can use the standard normal table to find the area between the mean and specific *Z* scores.
      2. To find the actual number of students who scored between two scores, multiply the proportion by the total number of students.
      3. For a score lower than the mean, we can use the standard normal table and the following steps:
         1. Convert the score to a *Z* score.
         2. Because the proportions that correspond to positive *Z* scores are identical to the proportions corresponding to negative *Z* scores, ignore the negative sign of *Z* and look up in Column A.
         3. The area corresponding to a *Z* score indicates the area under the curve is included between the mean and a *Z* value.
         4. We convert this proportion to a percentage.
         5. Thus, percentage of the distribution lies between the scores.
   3. Finding the Area Above a Positive *Z* Score or Below a Negative *Z* Score
      1. The normal distribution table can also be used to find the area beyond a *Z* score, SAT scores that lie at the tip of the positive or negative sides of the distribution.
      2. A similar procedure can be applied to identify the number of students on the opposite end of the distribution.
         1. First convert a score to a *Z* score.
         2. The area beyond a *Z* includes all students who scored below target score.
         3. Locate the proportion of students in this area in Column C in the entry corresponding to a *Z*.
         4. Remember that the proportions corresponding to positive or negative *Z* scores are identical.
   4. Transforming Proportions and Percentages Into *Z* Scores
      1. Finding a *Z* Score Which Bounds an Area Above It
         1. To identify the score that corresponds to the top 10% of SAT test takers, we will need to identify the cutoff point for the top 10% of the class.
         2. This problem involves two steps:
            1. Find the *Z* score that bounds the top 10% or 0.1000 (0.1000\*100 = 10%) of all the students who took the ERW SAT.

Refer to the areas under the normal curve shown in Appendix B.

* + - * 1. Find the score associated with a *Z* of 1.28.
      1. To transform a *Z* score into a raw score we multiply the score by the standard deviation and add that product to the mean: 
    1. Finding a *Z* Score Which Bounds an Area Below It

1. To identify the score which corresponds to the bottom 5% of test takers, the problem involves two steps:
   * + - 1. Find the *Z* score that bounds the lowest 5% or 0.0500 of all the students who took the class.

Refer to the areas under the normal curve and look for an entry of 0.0500 (or the value closest to it) in Column C.

* + - * 1. To find the final ERW score associated with a *Z* of –1.65, convert the *Z* score to a raw score.
  1. Working with Percentiles in a Normal Distribution
     1. To determine the percentile rank of a raw score requires transforming *Z* scores into proportions or percentages.
     2. Finding the Percentile Rank of a Score Higher Than the Mean
        1. To find the percentile rank of a score higher than the mean, follow these steps:
           1. Convert the raw score to a *Z* score.
           2. Find the area beyond *Z* in Appendix B, Column C.
           3. Subtract the area from 1.00 and multiply by 100 to obtain the percentile rank.
        2. Being in the 92nd percentile means that 92% of all test takers scored lower than raw score and 8% scored higher than raw score.
     3. Finding the Percentile Rank of a Score Lower Than the Mean

1. To find the percentile rank of a score lower than the mean, follow these steps:
   * + - 1. Convert the raw score to a *Z* score.
         2. Find the area beyond *Z* in Appendix B, Column C.
         3. Multiply the area by 100 to obtain the percentile rank.
2. The 20th percentile rank means that 20% of all test takers scored lower than you.
   * 1. Finding a Raw Score Associated with a Percentile Higher Than 50
3. To find the score associated with a percentile higher than 50, follow these steps:
   * + - 1. Divide the percentile by 100 to find the area below the percentile rank.
         2. Subtract the area below the percentile rank from 1.00 to find the area above the percentile rank.
         3. Find the *Z* score associated with the area above the percentile rank.

Refer to the area under the normal curve shown in Appendix B.

* + - * 1. Convert the *Z* score to a raw score.
    1. Finding the Raw Score Associated with a Percentile Lower than 50

1. To find the percentile rank of a score lower than 50 follow these steps:
2. Divide the percentile by 100 to find the area below the percentile rank.
3. Find the *Z* score associated with this area.

Refer to the area under the normal curve shown in Appendix B.

1. Convert the *Z* score to a raw score.
2. Reading the Research Literature: Child Health and Academic Achievement
   1. A study by Margot Jackson found that the role of health in producing academic inequality depends on when, and for how long, children are in poor health.