MATH 201: Lecture 4b Handout
Section 3.5 Continuous Disitrbutions and 4.1 Normal Distirbution

Name:_____ Date:_____

Learning goals for today

By the end of this lecture, you should be able to:

- Describe the shape and properties of a normal distribution.
- Identify the mean and standard deviation on a normal curve.
- Interpret the Empirical Rule (68–95–99.7 Rule).
- Standardize a value using a z-score.
- Interpret a z-score in context.

From Real Data to Models

So far in this course, we have worked with real datasets and used histograms to describe:

- Shape: modality (number of peaks) and symmetric or skewed
- Unusual features: outliers

Suppose we collected exam scores from a large introductory statistics class and made a histogram. Imagine the histogram looks:

- Roughly symmetric
- Unimodal
- No extreme outliers
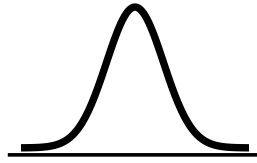- Most scores clustered near the middle

Sketch what you think this histogram might look like:

Real histograms are:

- Somewhat noisy
- Dependent on bin width
- Based on a finite sample

Instead of working with the imperfect histogram directly, we sometimes use a smooth mathematical model that approximates the distribution.

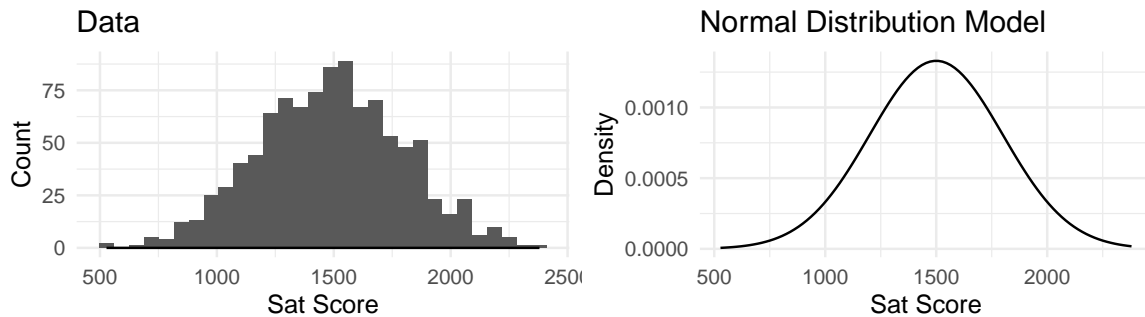One very important model is the **Normal Distribution**.

It is:

- Perfectly symmetric
- Bell-shaped

If our data is well represented by the normal curve, then we can use the curve describe our data (and more!).

Example: Consider SAT scores with a mean of 1500 and standard deviation of 300

The normal distribution does not describe every dataset. But it often provides a very good approximation when:

- The data are roughly symmetric
- There are no extreme outliers
- The distribution is unimodal (single peak)

For each variable below, do you think the normal distribution is a good approximation? (Hint: first try roughly drawing a histogram to get you thinking about the shape, then think if the three characteristics above describe it well.)
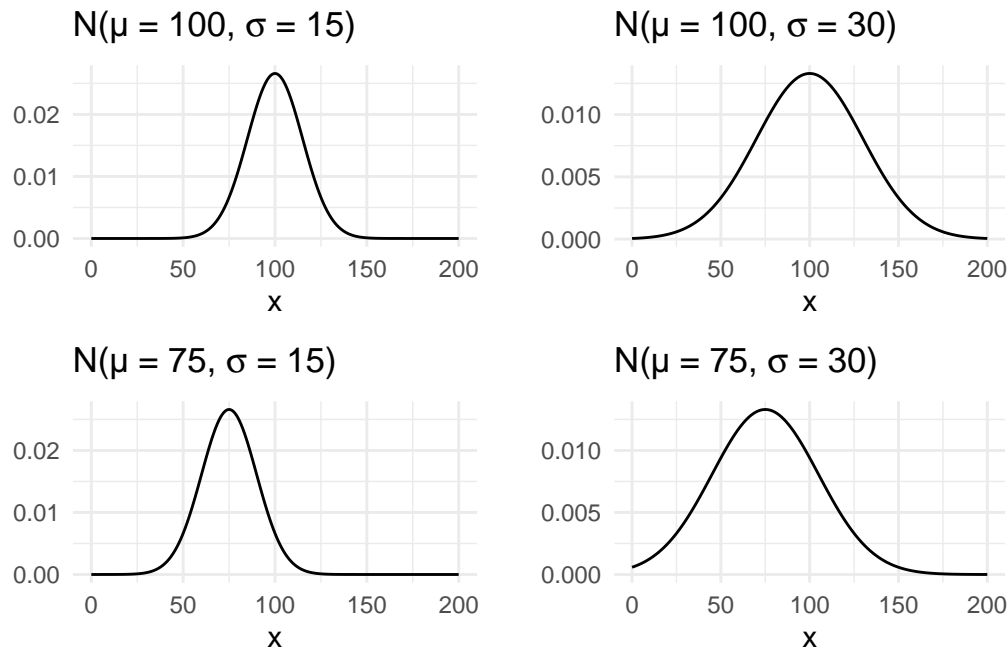
- Heights of adults in the US

- Number of children in a family

Features of the normal distirbution

The normal distribution is good for more than just describing a distribution. A normal distribution is a continuous, symmetric, bell-shaped distribution defined by:

- A mean (μ)
- A standard deviation (σ)

The mean μ is the center of the data and the standard deviation σ describes the spread of the data. We write $N(\mu, \sigma)$ to describe a normal distribution. We can get different normal distributions depending on the values of μ and σ.

Let's focus on the y-axis now. The y-axis of the normal distribution describes the density (where value are likely and unlikely). Think back to the normal distribution model for SAT scores, what SAT score seem more likely and which seem less likely?
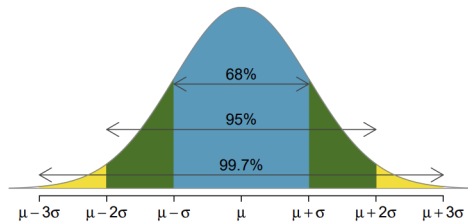
Key properties of the normal distribution:

- Symmetric around the mean
- Mean = Median
- Total area under curve = 1

This means we can use the normal distribution to calculate probabilities, if the normal distribution is a good fit for our variable. Look back to the normal distribution model for SAT scores. Approximately, what is probability of someone scoring:

- less than 1500?
- more than 500?
- between 1000 and 2000?

We can use the normal distribution to better approximate some probabilities using the **Empirical Rule**! For a normal distribution:

- About 68% of observations fall within 1 standard deviation of the mean.
- About 95% fall within 2 standard deviations.
- About 99.7% fall within 3 standard deviations.

Z-scores

The Empirical Rule gives us approximate probabilities, but what if we want something more precise? To answer more specific probability questions, we first convert our value into a z-score.

A **z-score** tells us: How many standard deviations a value is from the mean. We calculate a z-score using:

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

Where:

- x = observed value
- μ = mean
- σ = standard deviation

Interpreting z-scores

- If $z = 0$, the value is exactly at the mean.
- If $z > 0$, the value is above the mean.
- If $z < 0$, the value is below the mean.

Recall SAT scores are approximately normally distributed with $\mu = 1500$ and $\sigma = 300$. What is the z-score for a student who scored 1800? Would that student be considered average, below average, or above average?

What is the z-score for a student who scored 900? What does this tell us about the student?

Next lecture we will learn how to get probabilities from z-scores!