Chapter 5

The Standard Deviation as a Ruler and the Normal Model
5.1

Standardizing with z-Scores
Comparing Athletes

• Dobrynska took the gold in the Olympics with a long jump of 6.63 m for the women’s heptathlon, 0.5 m higher than average.

• Fountain won the 200 m run with a time of 23.21 s, 1.5 s faster than average.

• Whose performance was more impressive?
How Many Standard Deviations Above?

<table>
<thead>
<tr>
<th></th>
<th>Long Jump</th>
<th>200 m Run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6.11 m</td>
<td>24.71 s</td>
</tr>
<tr>
<td>SD</td>
<td>0.24 m</td>
<td>0.70 s</td>
</tr>
<tr>
<td>Individual</td>
<td>6.63 m</td>
<td>23.21 s</td>
</tr>
</tbody>
</table>

• The standard deviation helps us compare.

• Long Jump:
  • 1 SD above: $6.11 + 0.24 = 6.35$
  • 2 SD above: $6.11 + (2)(0.24) = 6.59$

• Just over 2 standard deviations above
The z-Score

• In general, to find the distance between the value and the mean in standard deviations:

  1. Subtract the mean from the value.
  2. Divide by the standard deviation.

\[ z = \frac{y - \bar{y}}{s} \]

• This is called the z-score.
The z-score

• The z-score measures the distance of the value from the mean in standard deviations.

• A positive z-score indicates the value is above the mean.

• A negative z-score indicates the value is below the mean.

• A small z-score indicates the value is close to the mean when compared to the rest of the data values.

• A large z-score indicates the value is far from the mean when compared to the rest of the data values.
How Many Standard Deviations Above?

<table>
<thead>
<tr>
<th></th>
<th>Long Jump</th>
<th>200 m Run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6.11 m</td>
<td>24.71 s</td>
</tr>
<tr>
<td>SD</td>
<td>0.24 m</td>
<td>0.70 s</td>
</tr>
<tr>
<td>Individual</td>
<td>6.63 m</td>
<td>23.21 s</td>
</tr>
</tbody>
</table>

• Standard Deviations from the Mean

Long Jump:

\[ z = \frac{6.63 - 6.11}{0.24} \approx 2.17 \]

200 m Run:

\[ z = \frac{23.23 - 24.71}{0.70} \approx -2.14 \]

• Dobrynska’s long jump was a little more impressive than Fountain’s 200 m run.
5.2 Shifting and Scaling
National Health and Examination Survey

- **Who?** 80 male participants between 19 and 24 who measured between 68 and 70 inches tall
- **What?** Their weights in kilograms
- **When?** 2001 – 2002
- **Where?** United States
- **Why?** To study nutrition and health issues and trends
- **How?** National survey
Shifting Weights

- Mean: 82.36 kg
- Maximum Healthy Weight: 74 kg

- How are shape, center, and spread affected when 74 is subtracted from all values?
  - Shape and spread are unaffected.
  - Center is shifted by 74.
Rules for Shifting

• If the same number is subtracted or added to all data values, then:

• The measures of the spread – standard deviation, range, and IQR – are all unaffected.

• The measures of position – mean, median, and mode – are all changed by that number.
Rescaling

• If we multiply all data values by the same number, what happens to the position and spread?

• To go from kg to lbs, multiply by 2.2.

• The mean and spread are also multiplied by 2.2.
How Rescaling Affects the Center and Spread

- When we multiply (or divide) all the data values by a constant, all measures of position and all measures of spread are multiplied (or divided) by that same constant.

<table>
<thead>
<tr>
<th></th>
<th>Weight (kg)</th>
<th>Weight (lb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>54.3</td>
<td>119.46</td>
</tr>
<tr>
<td>Q1</td>
<td>67.3</td>
<td>148.06</td>
</tr>
<tr>
<td>Median</td>
<td>76.85</td>
<td>169.07</td>
</tr>
<tr>
<td>Q3</td>
<td>92.3</td>
<td>203.06</td>
</tr>
<tr>
<td>Max</td>
<td>161.5</td>
<td>355.30</td>
</tr>
<tr>
<td>IQR</td>
<td>25</td>
<td>55</td>
</tr>
<tr>
<td>SD</td>
<td>22.27</td>
<td>48.99</td>
</tr>
</tbody>
</table>
Example: Rescaling Combined Times in the Olympics

• The mean and standard deviation in the men’s combined event at the Olympics were 168.93 seconds and 2.90 seconds, respectively.

• If the times are measured in minutes, what will be the new mean and standard deviation?
  • Mean: $168.93 / 60 = 2.816$ minutes
  • Standard Deviation: $2.90 / 60 = 0.048$ minute
Shifting, Scaling, and z-Scores

• Converting to z-scores:
  • Subtract the mean \( \bar{y} - \bar{y} = 0 \)
  • Divide by the standard deviation \( s/s = 1 \)
  • The shape of the distribution does not change.
  • Changes the center by making the mean 0
  • Changes the spread by making the standard deviation 1
Example: SAT and ACT Scores

• How high does a college-bound senior need to score on the ACT in order to make it into the top quarter of equivalent of SAT scores for a college with middle 50% between 1530 and 1850?

• SAT: Mean = 1500, Standard Deviation = 250

• ACT: Mean = 20.8, Standard Deviation = 4.8

• Think →

  • Plan: Want ACT score for upper quarter. Have $\bar{y}$ and $s$
  • Variables: Both are quantitative. Units are points.
Show → Mechanics: Standardize the Variable

- It is known that the middle 50% of SAT scores are between 1530 and 1850, $\bar{y} = 1500$, $s = 250$

- The top quarter starts at 1850.

- Find the $z$-score: $z = \frac{1850 - 1500}{250} = 1.40$

- For the ACT, 1.40 standard deviations above the mean:
  $20.8 + 1.40(4.8) = 27.52$
Conclusion

• To be in the top quarter of applicants in terms of combined SAT scores, a college-bound senior would need to have an ACT score of at least 27.52.
Normal Models
Models

• “All models are wrong, but some are useful.”
  George Box, statistician

• $-1 < z < 1$: Not uncommon

• $z = \pm 3$: Rare

• $z = 6$: Shouts out for attention!
The Normal Model

• Bell Shaped: unimodal, symmetric

• A Normal model for every mean and standard deviation.

  • $\mu$ (read “mew”) represents the population mean.
  
  • $\sigma$ (read “sigma”) represents the population standard deviation.
  
  • $N(\mu, \sigma)$ represents a Normal model with mean $\mu$ and standard deviation $\sigma$. 
Parameters and Statistics

- **Parameters:** Numbers that help specify the model
  - \( \mu, \sigma \)

- **Statistics:** Numbers that summarize the data
  - \( \bar{y}, s, \text{median, mode} \)

- \( N(0, 1) \) is called the standard Normal model, or the standard Normal distribution.

- The Normal model should only be used if the data is approximately symmetric and unimodal.
The 68-95-99.7 Rule

- 68% of the values fall within 1 standard deviation of the mean.
- 95% of the values fall within 2 standard deviations of the mean.
- 99.7% of the values fall within 3 standard deviations of the mean.
Example of the 68-95-99.7 Rule

- In the 2010 winter Olympics men’s slalom, Li Lei’s time was 120.86 sec, about 1 standard deviation slower than the mean. Given the Normal model, how many of the 48 skiers were slower?

- About 68% are within 1 standard deviation of the mean.
- 100% – 68% = 32% are outside.
- “Slower” is just the left side.
- 32% / 2 = 16% are slower.
- 16% of 48 is 7.7.
- About 7 are slower than Li Lei.
Three Rules For Using the Normal Model

1. Make a picture.
2. Make a picture.
3. Make a picture.

- When data is provided, first make a **histogram** to make sure that the distribution is **symmetric** and **unimodal**.
- Then sketch the Normal model.
Working With the 68-95-99.7 Rule

• Each part of the SAT has a mean of 500 and a standard deviation of 100. Assume the data is symmetric and unimodal. If you earned a 700 on one part of the SAT how do you stand among all others who took the SAT?

• Think →
• Plan: The variable is quantitative and the distribution is symmetric and unimodal. Use the Normal model \( N(500, 100) \).
Show and Tell

- **Show → Mechanics:**
  - Make a picture.
  - 700 is 2 standard deviations above the mean.

- **Tell → Conclusion:**
  - 95% lies within 2 standard deviations of the mean.
  - 100% - 95% = 5% are outside of 2 standard deviations of the mean.
  - Above 2 standard deviations is half of that.
    - 5% / 2 = 2.5%
  - Your score is higher than 2.5% of all scores on this test.
5.4

Finding Normal Percentiles
What if $z$ is not $-3, -2, -1, 0, 1, 2,$ or $3$?

• If the data value we are trying to find using the Normal model does not have such a nice $z$-score, we will use a computer.

• **Example:** Where do you stand if your SAT math score was 680? $\mu = 500$, $\sigma = 100$

• Note that the $z$-score is not an integer:

$$z = \frac{680 - 500}{100} = 1.8$$
Using StatCrunch for the Normal Model

• What percent of all SAT scores are below 680?
  • \( \mu = 500, \sigma = 100 \)

• Stat → Calculators → Normal

• Fill in info, hit Compute
Using StatCrunch for the Normal Model

- What percent of all SAT scores are below 680?
  - $\mu = 500$, $\sigma = 100$

- Stat → Calculators → Normal

- Fill in info, hit Compute

- 96.4% of SAT scores are below 680.
A Probability Involving “Between”

• What is the proportion of SAT scores that fall between 450 and 600?  \( \mu = 500, \sigma = 100 \)

• Think →
  • Plan: Probability that \( x \) is between 450 and 600
    \[ = \text{Probability that } x < 600 - \text{Probability that } x < 450 \]

• Variable: We are told that the Normal model works. \( N(500, 100) \)
A Probability Involving “Between”

• What is the proportion of SAT scores that fall between 450 and 600? \( \mu = 500, \sigma = 100 \)

• **Show → Mechanics:** Use StatCrunch to find each of the probabilities.

- Probability that \( x \) is between 450 and 600
  
  \[
  = \text{Probability that } x < 600 - \text{Probability that } x < 450 \\
  = 0.8413 - 0.3085 = 0.5328
  \]
A Probability Involving “Between”

• What is the proportion of SAT scores that fall between 450 and 600? \( \mu = 500, \sigma = 100 \)

• Probability that \( x \) is between 450 and 600
  \[ = \text{Probability that } x < 600 - \text{Probability that } x < 450 \]
  \[ = 0.8413 - 0.3085 \]
  \[ = 0.5328 \]

• Conclusion: The Normal model estimates that about 53.28% of SAT scores fall between 450 and 600.
• Suppose a college admits only people with SAT scores in the top 10%. How high a score does it take to be eligible? \( \mu = 500, \sigma = 100 \)

• Think →

• **Plan:** We are given the probability and want to go backwards to find \( x \).

• **Variable:** \( N(500, 100) \)
From Percentiles to Scores: \( z \) in Reverse

• Suppose a college admits only people with SAT scores in the top 10%. How high a score does it take to be eligible? \( \mu = 500, \sigma = 100 \)

• **Show → Mechanics:** Use StatCrunch putting in 0.9 for the probability.

• Probability \( x < 628 = 0.9 \)

• **Conclusion:** Because the school wants the SAT Verbal scores in the top 10%, the cutoff is 628.
Underweight Cereal Boxes

• Based on experience, a manufacturer makes cereal boxes that fit the Normal model with mean 16.3 ounces and standard deviation 0.2 ounces, but the label reads 16.0 ounces. What fraction will be underweight?

• Think →
  • Plan: Find Probability that $x < 16.0$
  • Variable: $N(16.3, 0.2)$
Underweight Cereal Boxes

• What fraction of the cereal boxes will be underweight (less than 16.0)?
  \( \mu = 16.3, \sigma = 0.2 \)

• Show → Mechanics: Use StatCrunch to find the probability.

• Probability \( x < 16.0 = 0.0668 \)
Underweight Cereal Boxes

• What fraction of the cereal boxes will be underweight (less than 16.0)? $\mu = 16.3, \sigma = 0.2$

• Probability $x < 16.0 = 0.0668$

• Conclusion: I estimate that approximately 6.7% of the boxes will contain less than 16.0 ounces of cereal.
Underweight Cereal Boxes Part II

• Lawyers say that 6.7% is too high and recommend that at most 4% be underweight. What should they set the mean at? $\sigma = 0.2$

• Think →
  • **Plan:** Find the mean such that $\text{Probability}(x < 16.0) = 0.04$.

• **Variable:** $N(?, 0.2)$

• **Reality Check:** Note that the mean must be less than 16.3 ounces.
Underweight Cereal Boxes Part II

- Lawyers say that 6.7% is too high and recommend that at most 4% be underweight. What should they set the mean at? $\sigma = 0.2$

- Mechanics: Sketch a picture.

- Use StatCrunch to find $z$ such that the area to the left of the standard Normal Model is 0.04.

  - $z = -1.75$
  - Find $16 + 1.75(0.02) = 16.35$ ounces
Underweight Cereal Boxes Part II

- Lawyers say that 6.7% is too high and recommend that at most 4% be underweight. What should they set the mean at? $\sigma = 0.2$

- $z = -1.75$

- Find $16 + 1.75(0.02) = 16.35$ ounces

- Conclusion: The company must set the machine to average 16.35 ounces per box.
Underweight Cereal Boxes Part III

• The CEO vetoes that plan and sticks with a mean of 16.2 ounces and 4% weighing under 16.0 ounces. She demands a machine with a lower standard deviation. What standard deviation must the machine achieve?

• Think →
  • Plan: Find $\sigma$ such that Probability $x < 16.0 = 0.04$.
  • Variable: $N(16.2, ?)$
Underweight Cereal Boxes Part III

• What standard deviation must the machine achieve? \( N(60.2, ?) \)

• Show \( \rightarrow \) Mechanics: From before, \( z = -1.75 \)
  
  \[ -1.75 = \frac{16.0 - 16.2}{\sigma} \]

• \( 1.75\sigma = 0.2, \quad \sigma = 0.114 \)

• Conclusion: The company must get the machine to box cereal with a standard deviation of no more than 0.114 ounces. The machine must be more consistent.
Section 5.5

Normal Probability Plots
Checking if the Normal Model Applies

• A histogram will work, but there is an alternative method.

• Instead use a Normal Probability Plot.
  • Plots each value against the z-score that would be expected had the distribution been perfectly normal.
  • If the plot shows a line or is nearly straight, then the Normal model works.
  • If the plot strays from being a line, then the Normal model is not a good model.
The Normal probability plot is nearly straight, so the Normal model applies. Note that the histogram is unimodal and somewhat symmetric.
The Normal Model Does Not Apply

- The Normal probability plot is not straight, so the Normal model does not apply applies. Note that the histogram is skewed right.
What Can Go Wrong

- Don’t use the Normal model when the distribution is not unimodal and symmetric.
  - Always look at the picture first.

- Don’t use the mean and standard deviation when outliers are present.
  - Check by making a picture.

- Don’t round your results in the middle of the calculation.
  - Always wait until the end to round.

- Don’t worry about minor differences in results.
  - Different rounding can produce slightly different results.