Chapter 2: Modeling Distributions of Data

Section 2.2
Normal Distributions

The Practice of Statistics, 4th edition - For AP*
STARNES, YATES, MOORE
Chapter 2
Modeling Distributions of Data

- **2.1** Describing Location in a Distribution
- **2.2** Normal Distributions
Section 2.2
Normal Distributions

Learning Objectives

After this section, you should be able to…

- DESCRIBE and APPLY the 68-95-99.7 Rule
- DESCRIBE the standard Normal Distribution
- PERFORM Normal distribution calculations
- ASSESS Normality
Normal Distributions

One particularly important class of density curves are the **Normal curves**, which describe **Normal distributions**.

- All Normal curves are symmetric, single-peaked, and bell-shaped.
- A Specific Normal curve is described by giving its mean $\mu$ and standard deviation $\sigma$.

Two Normal curves, showing the mean $\mu$ and standard deviation $\sigma$. 
Normal Distributions

Definition:

A **Normal distribution** is described by a Normal density curve. Any particular Normal distribution is completely specified by two numbers: its mean $\mu$ and standard deviation $\sigma$.

- The mean of a Normal distribution is the center of the symmetric **Normal curve**.
- The standard deviation is the distance from the center to the change-of-curvature points on either side.
- We abbreviate the Normal distribution with mean $\mu$ and standard deviation $\sigma$ as $N(\mu,\sigma)$.

Normal distributions are good descriptions for some distributions of *real data*.

Normal distributions are good approximations of the results of many kinds of *chance outcomes*.

Many *statistical inference* procedures are based on Normal distributions.
The 68-95-99.7 Rule

Although there are many Normal curves, they all have properties in common.

Definition: The 68-95-99.7 Rule ("The Empirical Rule")
In the Normal distribution with mean \( \mu \) and standard deviation \( \sigma \):
- Approximately 68% of the observations fall within \( \sigma \) of \( \mu \).
- Approximately 95% of the observations fall within 2\( \sigma \) of \( \mu \).
- Approximately 99.7% of the observations fall within 3\( \sigma \) of \( \mu \).
The distribution of Iowa Test of Basic Skills (ITBS) vocabulary scores for 7th grade students in Gary, Indiana, is close to Normal. Suppose the distribution is $N(6.84, 1.55)$.

a) Sketch the Normal density curve for this distribution.

b) What percent of ITBS vocabulary scores are less than 3.74?

c) What percent of the scores are between 5.29 and 9.94?
In the previous alternate example about batting averages for Major League Baseball players in 2009, the mean of the 432 batting averages was 0.261 with a standard deviation of 0.034. Suppose that the distribution is exactly Normal with \( \mu = 0.261 \) and \( \sigma = 0.034 \).

(a) Sketch a Normal density curve for this distribution of batting averages. Label the points that are 1, 2, and 3 standard deviations from the mean.

(b) What percent of the batting averages are above 0.329? Show your work.

(c) What percent of the batting averages are between 0.193 and 0.295? Show your work.
The Standard Normal Distribution

All Normal distributions are the same if we measure in units of size $\sigma$ from the mean $\mu$ as center.

**Definition:**

The **standard Normal distribution** is the Normal distribution with mean 0 and standard deviation 1.

If a variable $x$ has any Normal distribution $N(\mu, \sigma)$ with mean $\mu$ and standard deviation $\sigma$, then the standardized variable

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

has the standard Normal distribution, $N(0,1)$. 
The Standard Normal Table

Because all Normal distributions are the same when we standardize, we can find areas under any Normal curve from a single table.

**Definition:** The Standard Normal Table

Table A is a table of areas under the standard Normal curve. The table entry for each value \( z \) is the area under the curve to the left of \( z \).

Suppose we want to find the proportion of observations from the standard Normal distribution that are less than 0.81. We can use Table A:

<table>
<thead>
<tr>
<th>( Z )</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>.7580</td>
<td>.7611</td>
<td>.7642</td>
</tr>
<tr>
<td>0.8</td>
<td>.7881</td>
<td>.7910</td>
<td>.7939</td>
</tr>
<tr>
<td>0.9</td>
<td>.8159</td>
<td>.8186</td>
<td>.8212</td>
</tr>
</tbody>
</table>

\[ P(z < 0.81) = 0.7910 \]
Finding Areas Under the Standard Normal Curve

Find the proportion of observations from the standard Normal distribution that are between -1.25 and 0.81.

Can you find the same proportion using a different approach?

\[
1 - (0.1056 + 0.2090) = 1 - 0.3146 = 0.6854
\]
Finding Area to the Right

Suppose we wanted to find the proportion of observations in a Normal distribution that were more than 1.53 standard deviations above the mean. That is, we want to know what proportion of observations in the standard Normal distribution are greater than \( z = 1.53 \). To find this proportion, locate the value 1.5 in the left-hand column of Table A, then locate the remaining digit 3 as .03 in the top row. The corresponding entry is 0.9370. This is the area to the left of \( z = 1.53 \). To find the area above \( z = 1.53 \), subtract 0.9370 from 1 to get 0.0630.
Finding Areas Under the Standard Normal Curve

Find the proportion of observations from the standard Normal distribution that are between -0.58 and 1.79.

Area to the left of $z = 1.79$ is 0.9633

Area to the left of $z = -0.58$ is 0.2810

Area between $z = -0.58$ and $z = 1.79$ is 0.6823
In a standard Normal distribution, 20% of the observations are above what value? Using Table A, we should look up an area of 0.8000 since the table always lists area to the left of a boundary. The closest area to 0.8000 is 0.7995 which corresponds to a \( z \)-score of \( z = 0.84 \). Thus, approximately 20% of the observations in a standard Normal distribution are above \( z = 0.84 \).
Normal Distribution Calculations

How to Solve Problems Involving Normal Distributions

State: Express the problem in terms of the observed variable $x$.

Plan: Draw a picture of the distribution and shade the area of interest under the curve.

Do: Perform calculations.
   • Standardize $x$ to restate the problem in terms of a standard Normal variable $z$.
   • Use Table A and the fact that the total area under the curve is 1 to find the required area under the standard Normal curve.

Conclude: Write your conclusion in the context of the problem.
Normal Distribution Calculations

When Tiger Woods hits his driver, the distance the ball travels can be described by $N(304, 8)$. What percent of Tiger’s drives travel between 305 and 325 yards?

When $x = 305$, $z = \frac{305 - 304}{8} = 0.13$

When $x = 325$, $z = \frac{325 - 304}{8} = 2.63$

Using Table A, we can find the area to the left of $z=2.63$ and the area to the left of $z=0.13$. $0.9957 - 0.5517 = 0.4440$. About 44% of Tiger’s drives travel between 305 and 325 yards.
Alternate Example

Serving Speed

In the 2008 Wimbledon tennis tournament, Rafael Nadal averaged 115 miles per hour (mph) on his first serves. Assume that the distribution of his first serve speeds is Normal with a mean of 115 mph and a standard deviation of 6 mph. About what proportion of his first serves would you expect to exceed 120 mph?

State: Let $x =$ the speed of Nadal’s first serve. The variable $x$ has a Normal distribution with $\mu =$115 and $\sigma =6$. We want the proportion of first serves with $x \geq 120$.

Plan: The figure below shows the distribution with the area of interest shaded. $x = 120 \ z = 0.83$

Do: Standardize: . Table A: Looking up a $z$-score of 0.83 shows us that the area less than $z = 0.83$ is 0.7967. This means that the area to the right of $z = 0.83$ is $1 - 0.7967 = 0.2033$.

Conclude: About 20% of Nadal’s first serves will travel more than 120 mph.
Assessing Normality

The Normal distributions provide good models for some distributions of real data. Many statistical inference procedures are based on the assumption that the population is approximately Normally distributed. Consequently, we need a strategy for assessing Normality.

☑ Plot the data.
  • Make a dotplot, stemplot, or histogram and see if the graph is approximately symmetric and bell-shaped.

☑ Check whether the data follow the 68-95-99.7 rule.
  • Count how many observations fall within one, two, and three standard deviations of the mean and check to see if these percents are close to the 68%, 95%, and 99.7% targets for a Normal distribution.
No Space in the Fridge?

The measurements listed below describe the usable capacity (in cubic feet) of a sample of 36 side-by-side refrigerators. Are the data close to Normal?

12.9  13.7  14.1  14.2  14.5  14.5  14.6  14.7  15.1  15.2  15.3  15.3 15.3  15.3 15.3  15.3  15.5  15.6  15.6  15.8  16.0  16.0  16.2  16.2  16.3  16.4 16.5  16.6  16.6  16.6 16.8  17.0  17.0  17.2  17.4  17.4  17.9  18.4

Here is a histogram of these data. It seems roughly symmetric and bell shaped.

The mean and standard deviation of these data are

\( \bar{x} = 15.825 \) and \( s_x = 1.217 \).

\( \bar{x} \pm 1s_x = (14.608, 17.042) \) 24 of 36 = 66.7%

\( \bar{x} \pm 2s_x = (13.391, 18.259) \) 34 of 36 = 94.4%

\( \bar{x} \pm 3s_x = (12.174, 19.467) \) 36 of 36 = 100%

These percents are quite close to what we would expect based on the 68-95-99.7 rule. Combined with the graph, this gives good evidence that this distribution is close to Normal.
Normal Probability Plots

Most software packages can construct Normal probability plots. These plots are constructed by plotting each observation in a data set against its corresponding percentile’s z-score.

Interpreting Normal Probability Plots

If the points on a Normal probability plot lie close to a straight line, the plot indicates that the data are Normal. Systematic deviations from a straight line indicate a non-Normal distribution. Outliers appear as points that are far away from the overall pattern of the plot.
Interpreting Normal Probability Plots

- **Usable Capacity**
  - Frequency of Usable Capacity
  - Histogram with bins at 12, 13, 14, 15, 16, 17, 18, 19, 20

- **Area (thousands)**
  - Frequency of Area
  - Histogram with bins at 0, 200, 400, 600, 800

- **FTpercent**
  - Frequency of FTpercent
  - Histogram with bins at 0.4, 0.6, 0.8, 1.0

- **Normal Quantile**
  - Normal quantile plot with data points plotted against theoretical quantiles

- **Normal Distributions**
Section 2.2
Normal Distributions

Summary

In this section, we learned that...

- The **Normal Distributions** are described by a special family of bell-shaped, symmetric density curves called **Normal curves**. The mean $\mu$ and standard deviation $\sigma$ completely specify a Normal distribution $N(\mu, \sigma)$. The mean is the center of the curve, and $\sigma$ is the distance from $\mu$ to the change-of-curvature points on either side.

- All Normal distributions obey the **68-95-99.7 Rule**, which describes what percent of observations lie within one, two, and three standard deviations of the mean.
Section 2.2
Normal Distributions

Summary
In this section, we learned that...

✓ All Normal distributions are the same when measurements are standardized. The standard Normal distribution has mean $\mu=0$ and standard deviation $\sigma=1$.

✓ Table A gives percentiles for the standard Normal curve. By standardizing, we can use Table A to determine the percentile for a given $z$-score or the $z$-score corresponding to a given percentile in any Normal distribution.

✓ To assess Normality for a given set of data, we first observe its shape. We then check how well the data fits the 68-95-99.7 rule. We can also construct and interpret a Normal probability plot.
Looking Ahead…

We’ll learn how to describe relationships between two quantitative variables

We’ll study

✓ Scatterplots and correlation
✓ Least-squares regression