Chapter 6 Summary
The Standard Deviation as a Ruler and the Normal Model

What did you learn?

Data can be easier to understand after shifting or rescaling the data.

- Shifting data by adding or subtracting the same amount from each value affects measures of central and position but not measures of spread.
- Rescaling data by multiplying or dividing every value by a constant changes all of the summary statistics – center, position, and spread.

The power of standardizing data.

- Standardizing uses the standard deviation as a ruler to measure distance from the mean, creating *z*-scores.
- Using *z*-scores allows comparison of values from different distributions or values based on different units.
- A *z*-score can identify unusual or surprising values among data.

A Normal model can sometimes provide a useful way to understand data.

- We can decide whether a Normal model is appropriate by checking the Nearly Normal Condition with a histogram or Normal probability plot.
- Normal models follow the 68-95-99.7 Rules and we can use tables or technology for a more detailed analysis.

Using the standard deviation as a ruler allows the comparison of data sets with different units

Standardizing	We standardize the eliminate units. Standardized values can be compared and combined even if the original variables had different units and magnitudes.
Standardized value	A value found by subtracting the mean and dividing by the standard deviation.
z-score	A *z*-score tells how many standard deviations a value is from the mean; *z*-scores have a mean of zero and a standard deviation of one.

$$z = \frac{y - \bar{y}}{s}$$ (no units)

Changing center and spread	Changing the center and spread of a variable is equivalent to changing its *units.*
Shifting	Adding a constant to each data value adds the same constant to the mean, the median, and the quartiles, but does not change the standard deviation or IQR.
Rescaling	Multiplying each data value by a constant multiplies both the measures of position (mean, median, and quartiles) and the measures of spread (standard deviation and IQR) by that constant.

Standardizing into *z*-scores does not change the shape of the distribution, but it does change the center (mean = 0) and the spread (standard deviation = 1)

Normal model	A useful family of models for unimodal, symmetric distributions.
Parameter	A numerically valued attribute of a model. For example, the values of μ and σ in a N (μ, σ) model are parameters.
Statistic	A value calculated from data to summarize aspects of the data.
Standard normal model / distribution	A normal model, N (μ, σ), with mean μ = 0 and standard deviation σ = 1.

Using parameters, $z = \frac{y - \mu}{\sigma}$

Normality assumption	When using a normal model, we make the assumption that the distribution of the data is normal.
Nearly normal condition	The shape of the distribution of a data set is unimodal and symmetric.

68-95-99.7 Rule	In a normal model, about 68% of values fall within 1 standard deviation of the mean, about 95% fall within 2 standard deviations of the mean, and about 99.7% fall within 3 standard deviations of the mean.

Fewer than 1 out of a million values have a *z*-score of less than –5 or greater than 5

Remember to make a picture of the distribution for working with Normal models

Normal percentiles	The normal percentile corresponding to a *z*-score gives the percentage of values in a standard normal distribution found at that *z*-score or below.

Example: What proportion of SAT scores fall between 450 and 600?

Think Plan: State the problem
Variables: Name the variable, check conditions and specify Normal model
Show Mechanics: Make a picture of the Normal model. Locate values and shade.
Find *z*-scores for the cut points 450 and 600
Use technology to find the area (or use a table)
Tell Conclusion: Interpret your result in context

Normal probability plot	A display to help assess whether a distribution of data is approximately normal. If the plot is nearly straight, the data satisfy the nearly normal condition.

What can go wrong?

- Don't use a Normal model when the distribution is not unimodal and symmetric.
- Don't use the mean and standard deviation when outliers are present.
- Don't round off too soon.
- Don't round your results in the middle of a calculation.
- Don't worry about minor differences in results.