Key Concept

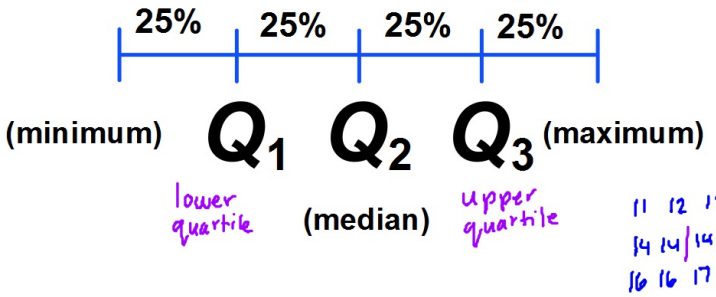
★ Measures of relative standing:

- They can be used to compare values from different data sets, or to compare values within the same data set.
- The most important concept is the *z* score.
- We will also discuss percentiles and quartiles, as well as a new statistical graph called the boxplot.

Quartiles

Q_1, Q_2, Q_3

Measures of loaction that divides sorted data values into four equal parts.



$Q_1 = 13$

11 12 12 12 13 13 | 13 13 14 14
14 14 | 14 14 15 15 15 15 | 16 16
16 16 17 21

$Q_3 = 15.5$

$Q_2 = \text{med} = 14$

Other Statistics

- Interquartile Range (or IQR): $Q_3 - Q_1$

IQR 15.5 − 13 = 2.5

Outliers

- An outlier is a value that lies very far away from the vast majority of the other values in a data set.

1.5(IQR)

$Q_1 - 1.5\ IQR$

$Q_3 + 1.5\ IQR$

(1.5)(2.5) = 3.75

$Q_1 - 3.75$ 13 − 3.75 = 9.

$Q_3 + 3.75$ 15.5 + 3.75 = 1

21 is an outlier

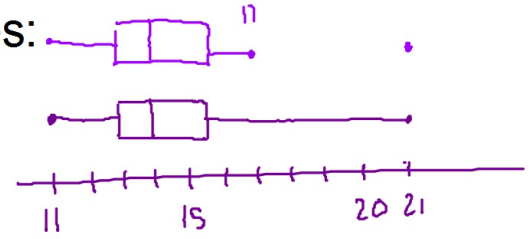
Important Principles

- An outlier can have a dramatic effect on the mean and the standard deviation.

- An outlier can have a dramatic effect on the scale of the histogram so that the true nature of the distribution is totally obscured.

5-Number Summary

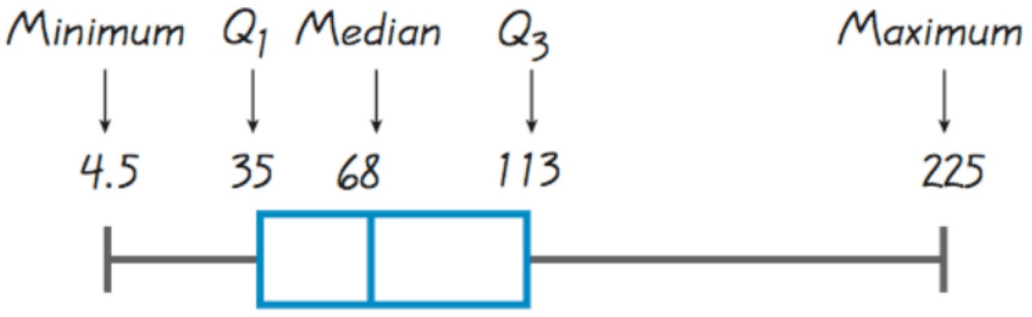
- For a set of data, the 5-number summary consists of these five values:



1. Minimum value 11
2. First quartile Q_1 13
3. Second quartile Q_2 (same as median) 14
4. Third quartile, Q_3 15.5
5. Maximum value 21

modified boxplot

Boxplots



Finding the Percentile of a Data Value

$$\text{Percentile of value } x = \frac{\text{number of values less than } x}{\text{total number of values}} \cdot 100$$

For the 40 Chips Ahoy cookies, find the percentile for a cookie with 23 chips.

$\frac{10}{40} \cdot 100 = 25^{th}$

19	19	20	20	20	20	22	22	22	22
23	23	23	23	23	23	23	24	24	24
24	24	25	25	25	25	25	25	25	26
26	26	26	26	26	27	27	28	28	30

Find P_{80} 80^{th} percentile

$\frac{x}{40} \cdot 100 = 80$

$x = 32$ cookie # 32 has 26 chips

z score

- *z* Score (or standardized value)
  the number of standard deviations that a given value *x* is above or below the mean

Sample

$$z = \frac{x - \bar{x}}{s}$$

Population

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

Round-Off Rule for *z* Scores:
Round *z* scores to two decimal places.

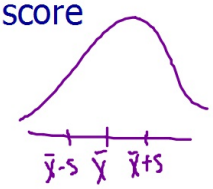
Properties of z Scores

1. A z score is the number of standard deviations that a given value x is above or below the mean.

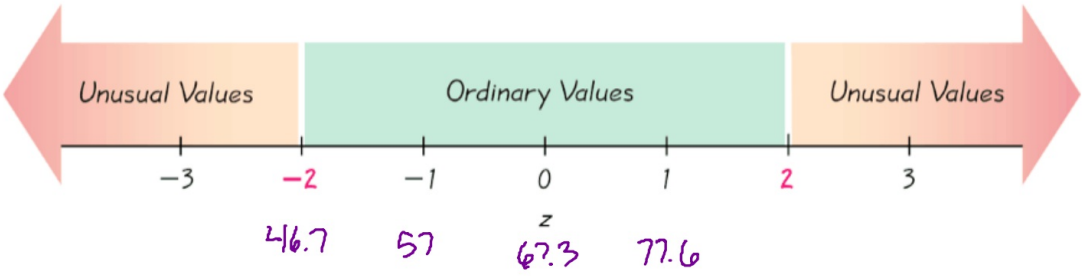
2. z scores are expressed as numbers with no units of measurement.

3. A data value is unusual if its z score is less than -2 or greater than +2.

4. If an individual data value is less than the mean, its z score is a negative number.



Interpreting Z Scores



The author of the text measured his pulse rate to be 48 beats per minute.

Is that pulse rate unusual if the mean adult male pulse rate is 67.3 beats per minute with a standard deviation of 10.3?

$$z = \frac{60-70}{8} = \frac{-10}{8} = -1.25$$

$$z = \frac{71-70}{8} = \frac{1}{8} = .125 \text{ or } .13$$

$$z = \frac{92-70}{8} = \frac{22}{8} = 2.75$$

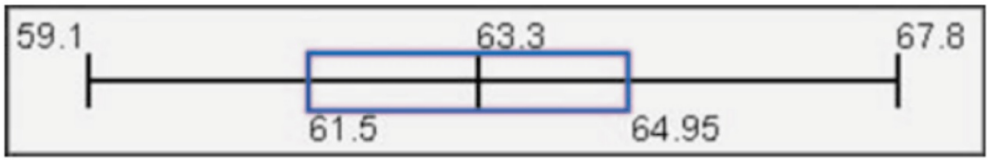
Example 8: We now consider a comparison of two *individual* data values with this question: Which of the following two data values is more extreme?

- The Chips Ahoy (regular) cookie with 30 chocolate chips (among 40 cookies with a mean of 24.0 chocolate chips and standard deviation of 2.6 chocolate chips) $z = \frac{30-24}{2.6} = 2.31$

- The can of regular Coke with a weight of 0.8295 lb (among 36 cans of regular Coke with a mean weight of 0.81682 lb and a standard deviation of 0.00751 lb). $\frac{.8295-.81682}{.00751} = 1.69$

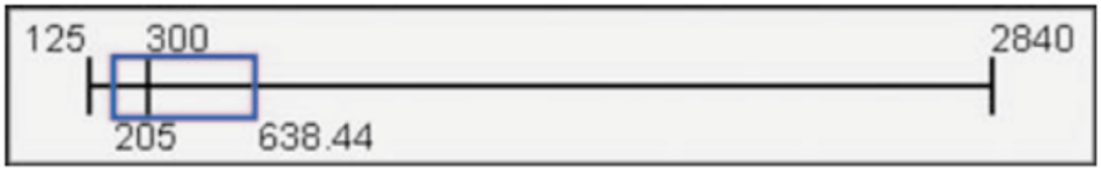
Both of the above data values are the largest values in their respective data sets, but which of them is more extreme relative to the data sets from which they came? The chips ahoy is more extreme

Boxplots - Normal Distribution



Normal Distribution:
Heights from a Simple Random Sample of Women

Boxplots - Skewed Distribution

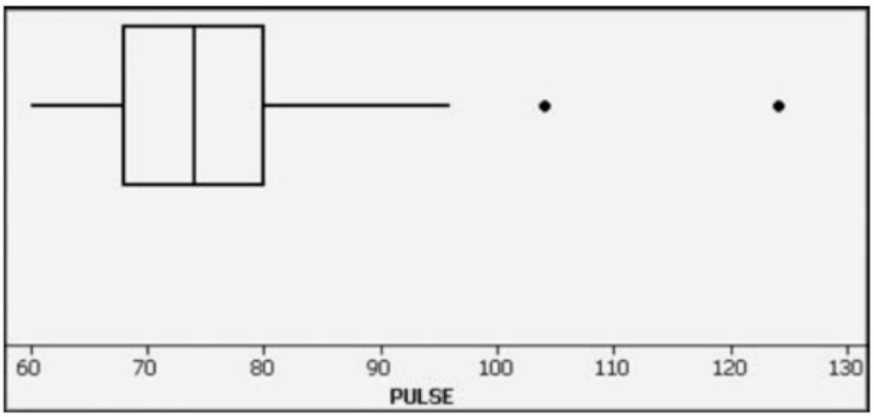


Skewed Distribution:
Salaries (in thousands of dollars) of NCAA Football Coaches

Modified Boxplots

Boxplots described earlier are called skeletal (or regular) boxplots.
Some statistical packages provide modified boxplots which represent outliers as special points.

Example



Example 9: Use the table below of Hannaford chocolate chip counts to construct a modified boxplot.

13	15	16	21	15	14	14	15	13	13	16	11
14	12	13	12	14	12	16	17	14	16	14	15

Putting It All Together

- So far, we have discussed several basic tools commonly used in statistics –
  - Context of data
  - Source of data
  - Sampling method
  - Measures of center and variation
  - Distribution and outliers
  - Changing patterns over time
  - Conclusions and practical implications
- This is an excellent checklist, but it should not replace ***thinking*** about any other relevant factors.