§3.3–The Five-Number Summary
Boxplots

Tom Lewis

Fall Term 2009

Outline

Quartiles

We can extend the concept of a median in an obvious way. Roughly speaking, the quartiles of an ordered data set divide the set into four "equal" parts, called the first, second, third quartile.

Here are the steps to find the quartiles of a data set:

Quartiles

We can extend the concept of a median in an obvious way. Roughly speaking, the quartiles of an ordered data set divide the set into four "equal" parts, called the first, second, third quartile.

Here are the steps to find the quartiles of a data set:

- Arrange the data and determine the median.

Quartiles

We can extend the concept of a median in an obvious way. Roughly speaking, the quartiles of an ordered data set divide the set into four "equal" parts, called the first, second, third quartile.

Here are the steps to find the quartiles of a data set:

- Arrange the data and determine the median.
- The first quartile, denoted by Q_1, is the median of that part of the data set that lies **at or below** the median of the entire set.

Quartiles

We can extend the concept of a median in an obvious way. Roughly speaking, the quartiles of an ordered data set divide the set into four "equal" parts, called the first, second, third quartile.

Here are the steps to find the quartiles of a data set:

- Arrange the data and determine the median.
- The first quartile, denoted by Q_1, is the median of that part of the data set that lies **at or below** the median of the entire set.
- The second quartile, denoted by Q_2, is the median of the entire set.

Quartiles

We can extend the concept of a median in an obvious way. Roughly speaking, the quartiles of an ordered data set divide the set into four "equal" parts, called the first, second, third quartile.
Here are the steps to find the quartiles of a data set:

- Arrange the data and determine the median.
- The first quartile, denoted by Q_1, is the median of that part of the data set that lies **at or below** the median of the entire set.
- The second quartile, denoted by Q_2, is the median of the entire set.
- The third quartile, denoted by Q_3, is the median of that part of the data set that lies **at or above** the median of the entire set.

Quartiles

We can extend the concept of a median in an obvious way. Roughly speaking, the quartiles of an ordered data set divide the set into four "equal" parts, called the first, second, third quartile.

Here are the steps to find the quartiles of a data set:

- Arrange the data and determine the median.
- The first quartile, denoted by Q_1, is the median of that part of the data set that lies **at or below** the median of the entire set.
- The second quartile, denoted by Q_2, is the median of the entire set.
- The third quartile, denoted by Q_3, is the median of that part of the data set that lies **at or above** the median of the entire set.

Problem

Determine the quartiles of the ACT data set.

Warning!

There is no universally recognized definition of quartile. For example, the program **R** uses a different method to determine the quartiles. For large data sets, these difference are not likely to be of any consequence.

Warning!

There is no universally recognized definition of quartile. For example, the program **R** uses a different method to determine the quartiles. For large data sets, these difference are not likely to be of any consequence.

Problem

calculate the quartiles of the ACT data using **R**. *This can be done from the* **R Commander** *menus as follows:*

Statistics $\rightarrow$ *Summaries* $\rightarrow$ *Numerical Summaries...*

Warning!

There is no universally recognized definition of quartile. For example, the program **R** uses a different method to determine the quartiles. For large data sets, these difference are not likely to be of any consequence.

Problem

calculate the quartiles of the ACT data using **R**. *This can be done from the* **R Commander** *menus as follows:*

Statistics → *Summaries* → *Numerical Summaries...*

Quantiles

The notion of a quartile can be naturally extended. For example, deciles break the data set up into 10 equal blocks and percentiles break the data set up into 100 equal blocks. Quartiles, deciles, percentiles, etc. are collectively called quantiles.

Definition (Inner quartile range)

The inner quartile range (IQR) is the difference between the first and the third quartiles; thus,

$$\mathrm{IQR} = Q_3 - Q_1$$

Definition (Inner quartile range)

The inner quartile range (IQR) is the difference between the first and the third quartiles; thus,

$$\mathrm{IQR} = Q_3 - Q_1$$

Problem

Determine IQR for the ACT data.

Definition (Five-number summary)

The five number summary of a data set is

$$\text{Min}, Q_1, Q_2, Q_3, \text{Max}$$

where Min and Max are the minimum and maximum observations in the set.

Definition (Five-number summary)

The five number summary of a data set is

$$\text{Min}, Q_1, Q_2, Q_3, \text{Max}$$

where Min and Max are the minimum and maximum observations in the set.

Problem

Give the five-number summary of the ACT data.

Definition (Lower and upper limits)

The lower and upper limits of a data set are

$$\text{Lower limit} = Q_1 - 1.5 \cdot IQR$$
$$\text{Upper limit} = Q_3 + 1.5 \cdot IQR$$

Definition (Lower and upper limits)

The lower and upper limits of a data set are

$$\text{Lower limit} = Q_1 - 1.5 \cdot IQR$$
$$\text{Upper limit} = Q_3 + 1.5 \cdot IQR$$

Note

The multiplier, 1.5, is not universal.

Definition (Lower and upper limits)

The lower and upper limits of a data set are

$$\text{Lower limit} = Q_1 - 1.5 \cdot IQR$$
$$\text{Upper limit} = Q_3 + 1.5 \cdot IQR$$

Note

The multiplier, 1.5, is not universal.

Problem

Determine the upper and lower limits for the ACT data.

Definition (Adjacent values)

The adjacent values of a data set are the most extreme observations of the data set that still lie within the lower and upper limits of the data set.

Definition (Adjacent values)

The adjacent values of a data set are the most extreme observations of the data set that still lie within the lower and upper limits of the data set.

Problem

Determine the adjacent values of the ACT data.

Definition (Outlier)

Roughly speaking, an outlier is an observation that is distant from the rest of the data. We will identify potential outliers as those observations that fall below the lower limit or exceed the upper limit.

Definition (Outlier)

Roughly speaking, an outlier is an observation that is distant from the rest of the data. We will identify potential outliers as those observations that fall below the lower limit or exceed the upper limit.

Problem

Identify any potential outliers from the ACT data.

Definition (Outlier)

Roughly speaking, an outlier is an observation that is distant from the rest of the data. We will identify potential outliers as those observations that fall below the lower limit or exceed the upper limit.

Problem

Identify any potential outliers from the ACT data.

Problem

Make a boxplot (also called a box and whisker plot) of the ACT data.

Definition (Outlier)

Roughly speaking, an outlier is an observation that is distant from the rest of the data. We will identify potential outliers as those observations that fall below the lower limit or exceed the upper limit.

Problem

Identify any potential outliers from the ACT data.

Problem

Make a boxplot (also called a box and whisker plot) of the ACT data.

Problem

Make a boxplot of the SunRise run data.