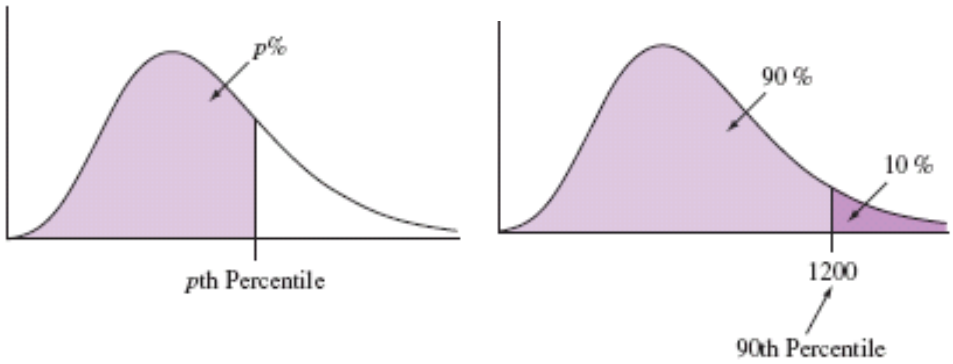
3.4/3.5: Measures of Position

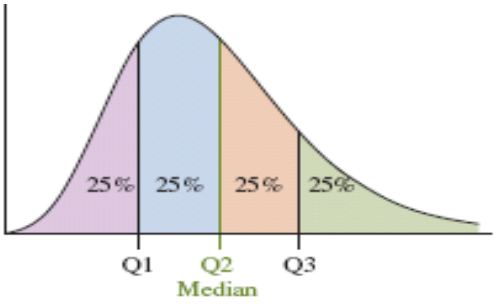
Percentile

- The p^{th} percentile is a value such that p% of the observations fall at or below that value



Finding Quartiles

- Quartiles split the data into four parts
  - Arrange the data in order
  - The median is the second quartile, Q_2
  - The first quartile, Q_1, is the median of the lower half of the observations
  - The third quartile, Q_3, is the median of the upper half of the observations



Quartiles divide a ranked data set into four equal parts.

- The **first quartile, Q_1,** is the value in the sample that has 25% of the data at or below it and 75% above
- The **second quartile** is the same as the **median** of a data set. 50% of the data are above the median and 50% are below
- The **third quartile, Q_3,** is the value in the sample that has 75% of the data at or below it and 25% above

Example: A medical researcher gathers the following data on the years that 25 patients survived after diagnosis with multiple myeloma. What are the quartiles

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0.6	1.2	1.5	1.6	1.9	2.1	2.3	2.3	2.5	2.8	2.9	3.3	3.4	3.6	3.7	3.8	3.9	4.1	4.2	4.5	4.7	4.9	5.3	5.6	8.1

Q1: **Q2:** **Q3:**

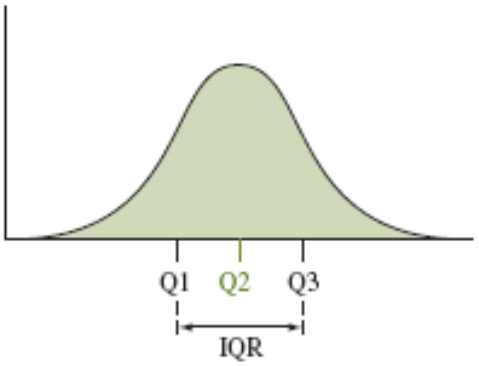
You Try It: Find the first and third quartiles, and the median.

Prices per share of 10 most actively traded stocks on NYSE (rounded to nearest $)

2 4 11 13 14 15 31 32 34 47

Calculating Inter Quartile Range

- **The inter quartile range (IQR) is the distance between the third quartile and first quartile:**
- *IQR* = ***Q3 – Q1***
- *IQR* gives spread of middle 50% of the data



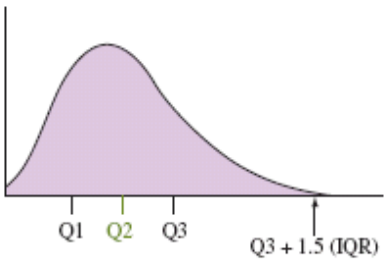
You Try It: Find the IQR for the stock prices

Prices per share of 10 most actively traded stocks on NYSE (rounded to nearest $)

2 4 11 13 14 15 31 32 34 47

Criteria for identifying an outlier using IQR

- An observation is a potential outlier if it falls more than *1.5 x IQR* below the first quartile or more than *1.5 x IQR* above the third quartile
- The upper and lower cutoff values are called "fences":
  Lower fence = $Q_1 - 1.5(IQR)$
  Upper fence = $Q_3 + 1.5(IQR)$



Consider This: Find the fences and any outliers in this data collected on multiple myeloma below?

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0.6	1.2	1.5	1.6	1.9	2.1	2.3	2.3	2.5	2.8	2.9	3.3	3.4	3.6	3.7	3.8	3.9	4.1	4.2	4.5	4.7	4.9	5.3	5.6	8.1

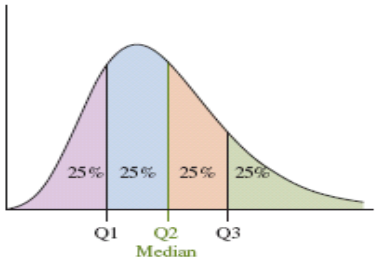
Upper Fence:

Lower Fence:

Are there any Outliers?

5 Number Summary

- The five-number summary of a dataset consists of the
  - "Minimum" value
  - First Quartile
  - Median
  - Third Quartile
  - "Maximum" value



- Example:
  Find the Five-number summary of the data below using your TI-84. (We'll use this more throughout this lesson.)
  Min: Q1: Q2: Q3: Max:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0.6	1.2	1.5	1.6	1.9	2.1	2.3	2.3	2.5	2.8	2.9	3.3	3.4	3.6	3.7	3.8	3.9	4.1	4.2	4.5	4.7	4.9	5.3	5.6	8.1

Z-Scores and outliers

- The z-score for an observation is the number of standard deviations that it falls from the mean

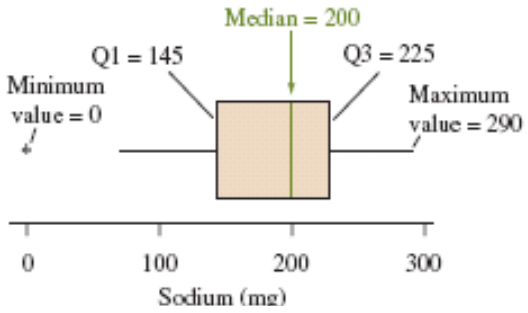
$$z = \frac{x - \bar{x}}{s} = \frac{\boldsymbol{observation - mean}}{\boldsymbol{standard\ deviation}}$$

- "2-Standard Deviation Rule": An observation from a bell-shaped distribution is often considered an outlier if its z-score < -2 or > +2

Try It: Find the z-score for the 8.1 year result in the Myeloma study. Is this an outlier ?

Box Plots

- A box goes from the Q1 to Q3
- A line is drawn inside the box at the median
- A line goes from the lower end of the box to the smallest observation, and from the upper end of the box to the largest observation (it is important that it be drawn to scale since it is meant to be a picture of the spread of the data)
- The potential outliers are shown separately

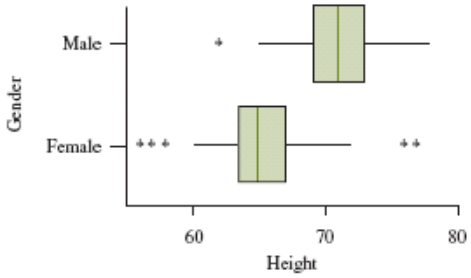


Try It: Use your TI-84 to Make a box plot of the data on Multiple Myeloma patients.

1	0.6
2	1.2
3	1.5
4	1.6
5	1.9
6	2.1
7	2.3
8	2.3
9	2.5
10	2.8
11	2.9
12	3.3
13	3.4
14	3.6
15	3.7
16	3.8
17	3.9
18	4.1
19	4.2
20	4.5
21	4.7
22	4.9
23	5.3
24	5.6
25	8.1

Comparing Distributions

Box Plots do not display the shape of the distribution as clearly as histograms, but are useful for making graphical comparisons of two
or more distributions



Box Plots and Histograms