Some estimates of the height of the podium

24 36 40 40 40 41 42 44 46 48 50 53 65 98

5 number summary

Inter-quartile range (IQR)

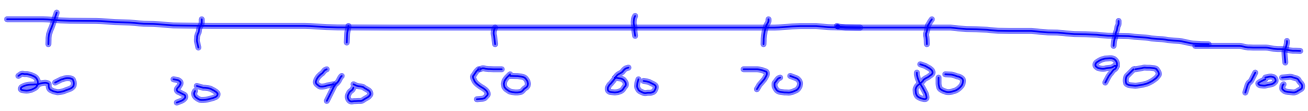
range = max-min

1.5 IQR outlier rule

24 is an outlier because
it is lower than $40 - 1.5(10) = 25$
$Q_1 - 1.5\,IQR$

98 is an outlier because
it's higher than $50 + 1.5(10) = 65$
$Q_3 + 1.5\,IQR$

make a boxplot



24 36 40 40 40 41 42 44 46 48 50 53 65 98

mean

add values and divide by the number of items

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \ldots x_n}{n}$$

$$\bar{x} = \frac{1}{n} \sum x_1$$

median M
midpoint of distribution
so that half the values are above and half are below
the 50th percentile value

1. Arrange values in order.
2. Odd # of values, median is center value.
3. Even # of values, median is mean of 2 center values.

To calculate the quartiles:

1. Arrange values in order, find median.
2. **first quartile Q1** median of smaller half of the values. (25th percentile value)
3. **third quartile Q3** median of larger half of the values. (75th percentile value)

There are about 12 different ways to find the quartiles (such as including the median or not, only when n is even, etc.). The important thing is that you follow A rule, not that there is a single rule.

The **interquartile range (IQR)** is the distance between the first and third quartiles, IQR = Q3 - Q1.

Call an observation an **outlier** if it falls more than 1.5 IQR above the third quartile or below the first quartile.

IQR is a number -
Many students write things like "The IQR goes from 15 to 32".

Every AP grader knows exactly what you mean, namely, "The box in my boxplot goes from 15 to 32.", but this statement is not correct. The IQR is defined as Q3 - Q1 which gives a single value. Writing the statement above is like saying "17 goes from 15 to 32." It just doesn't make sense.

five number summary

Min	smallest observation
Q1	first quartile
M	median
Q3	third quartile
Max	largest observation

A **modified boxplot** is a graph of the five number summary, with outliers plotted individually.

* A central box spans the quartiles.
* A line in the box marks the median.
* Observations more than 1.5 IQR outside the central box are plotted individually.
* Lines extend from the box out to the smallest and largest observations that are not outliers.

aspect	best graphs
Shape	dotplot or histogram
Outliers	boxplot
Center	dotplot, histogram, or boxplot
Spread	dotplot, histogram, or boxplot

Make a boxplot by hand:

Min Q_1 Med Q_3 Max

$IQR=Q_3-Q_1=$

$1.5(IQR)=$

Outlier rule:

$Q_1-1.5(IQR)=$

$Q_3+1.5(IQR)=$

The variance s^2 of a set of observations is an average of the squares of the deviations from their mean. In symbols, tl variance of n observations $x_1, x_2, x_3, \ldots, x_n$ is

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots (x_n - \bar{x})^2}{n-1}$$

or, more compactly,

$$s^2 = \frac{\sum (x_1 - \bar{x})^2}{n-1}$$

The standard deviation s is the square root of the variance s^2 :

$$s = \sqrt{\frac{\sum (x_1 - \bar{x})^2}{n-1}}$$

s or s_x = standard deviation $s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$

s^2 = variance

{10, 12, 17}

{13, 13, 13}

{10, 12, 17}

{-13, 13, -39}

{-113, 39, 113}

http://bcs.whfreeman.com/yates2e/default.asp?s=&n=&i=&v=&o=&ns=0&t=&uid=0&rau=0

5, 8, 9, 10, 12, 17, 19

	original	+5	-5	x2	x10
mean					
Q_1					
Med					
Q_3					
s					
IQR					

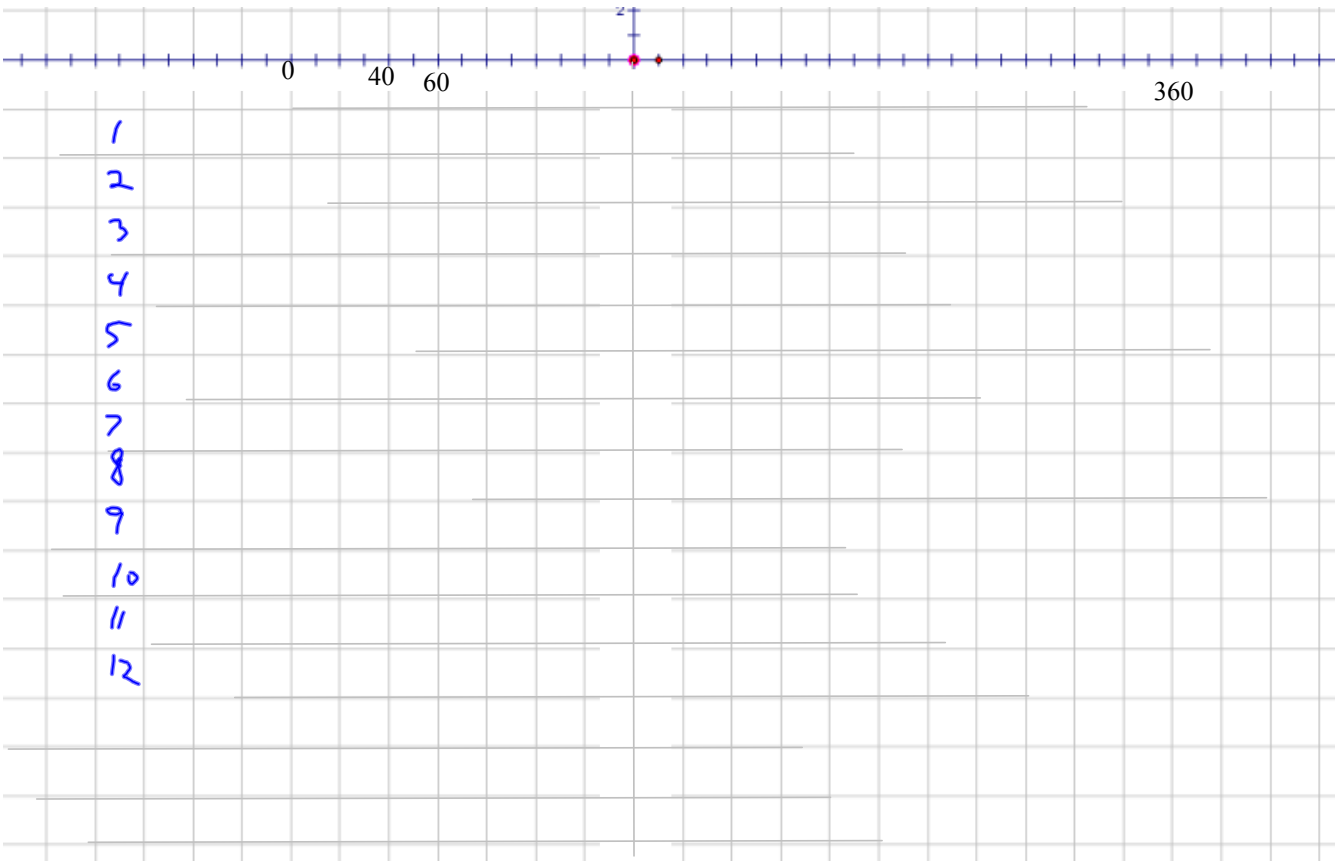
You can now do
p. 52-53: 41,43, p. 56: 45

Properties of the standard deviation, s

- s measures spread about $\bar{x}$
- s = 0 means there is no spread and all the values are the same.
- $s \geq 0$
- more spread means s will be larger
- s, like $\bar{x}$, is not resistant-- it is affected by skewness or outliers

Choosing summaries

	symmetric	skewed, outliers
center	mean	median
spread	standard deviation	IQR


0
40
60
360
1
2
3
4
5
6
7
8
9
10
11
12

Here's more info about the Vietnam Draft, in case you want to check out the pictures, data, or history involved.

http://www.sss.gov/lotter1.htm

http://www.niles-hs.k12.il.us/timmil/draft_project.html

Adding a constant to all values will add the same contant to the mean and the 5 number summary, but won't change s, range or IQR.

Multiplying all values by a positive #, multiply the mean, 5 number summary, s, range, and IQR by that same #.

The effect of changing units or using linear transformations

Change	Effect on summary statistic
add the same number c to each data value	add that same number c to each statistic (c + the mean, median, quartile, max, or min) However, s & the IQR remain the same.
multiply the same number c by each data value	multiply that same number c by each statistic (c times the mean, median, quartile, max, or min) However, s & the IQR are multiplied by the absolute value of c.