UNIVERSITY OF WATERLOO

Faculty of Mathematics
Waterloo, Ontario N2L 3G1

Centre for Education in
Mathematics and Computing

Intermediate Math Circles

March 30, 2016

Presenting and Understanding Data

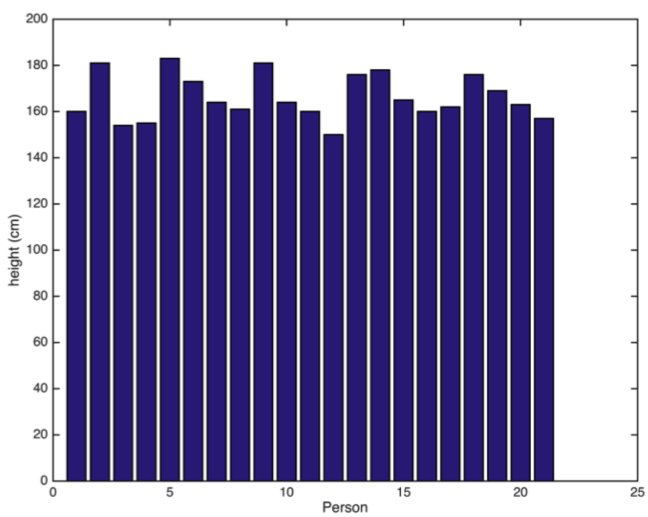
Presenting Data

Last week we collected data on your heights and reaction times using a ruler drop experiment. Here is the list of your heights.
Is this a good way to present the data?
What mathematical quantities did we learn about last week which we can use on this data?

Height (cm)				
160	181	154	155	183
173	164	161	181	164
160	150	176	178	165
160	162	176	169	163
157				

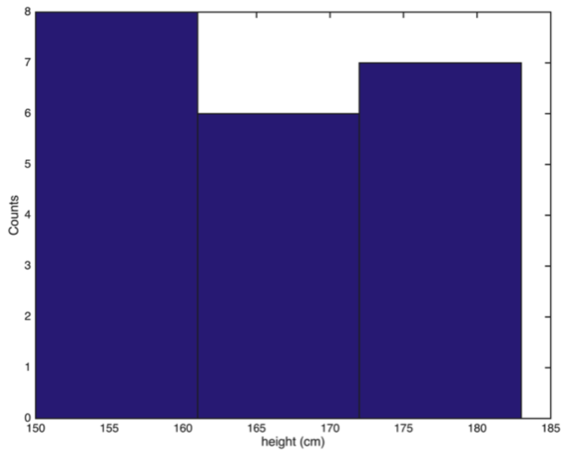
Mean: 166.3 cm
Standard Deviation: 9.77 cm
Max: 183 cm
Min: 150 cm
Can we make a plot or picture of this data?

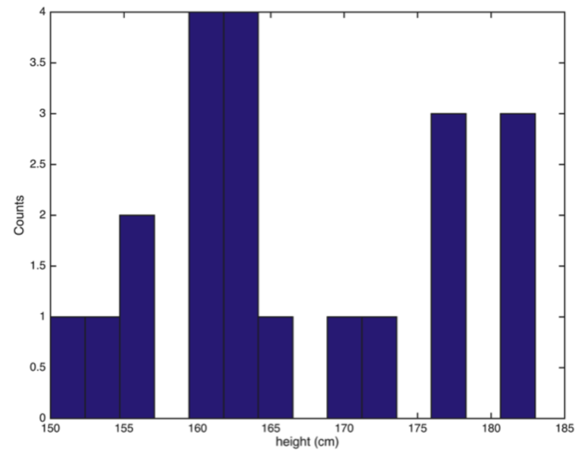
Here is a bar plot of each individuals height. Is this helpful?



The Histogram

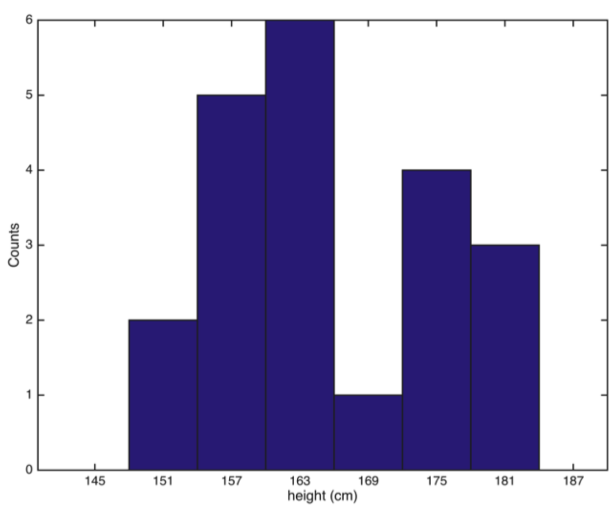
Instead of listing each height individually, why dont we group together similar heights.
How many people have heights in the range $[0 - 161]$? $[162 - 172]$? And $[173 - \infty)$?





Are either of these informative?

The number of bins is key to understanding the data. Too many bins and they are basically all the same. Not enough bins and there is no variation in the counts per bin. The right amount depends on the data.



What does the dip in the counts at 169 cm imply?
Is this reflective of your expectation of peoples heights?
What can we do to make the distribution more like what we expect for people heights?

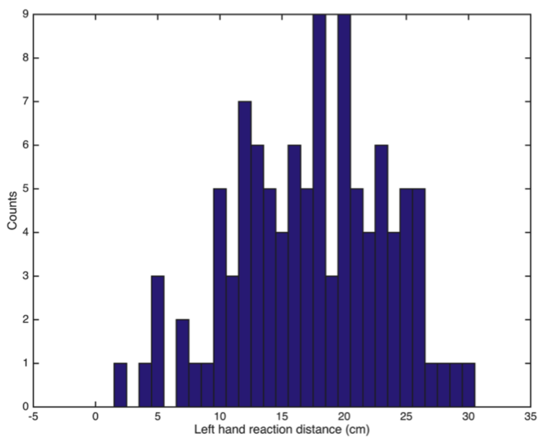
Reaction Distances

The other experiment from last week was to measure reaction times by dropping a ruler between students hands. We did both left and right hands. Lets make a hypothesis:

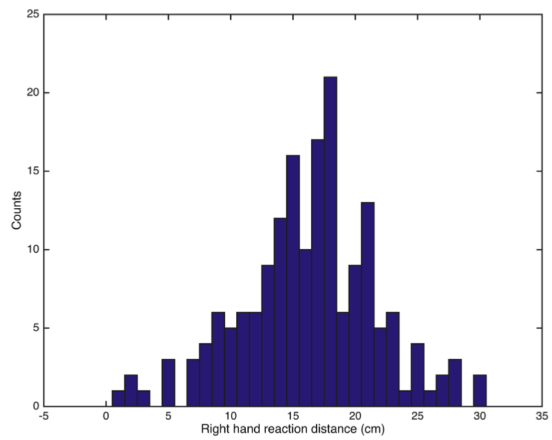
Which hand has faster reaction time the left hand or the right hand?

What are the possible answers?
To get an answer we must use math. Math (at least applied math) is concerned with using tools to understand and make convincing arguments about the world. But how we take the world and put it into math terms is not always easy.
Questions 1-3 on Measurement error worksheet.

Left hand
Mean: 17.3 cm
Standard Deviatioin: 6.07 cm



Right hand
Mean: 16.3 cm
Standard Deviatioin: 5.4 cm



Probability mass function (pmf)

Since there is more data points for the right hand, the right hand histogram has a larger maximum. To be able to compare the two we need to convert the histograms into probability mass function. The pmf is

$$f_X(x) = Pr(X = x) = Pr(\{s \in S : X(s) = x\})$$

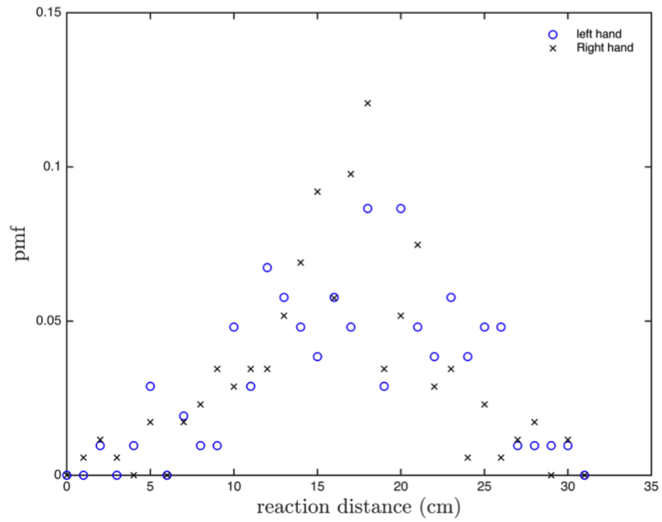
Where S is the sample space (all possible outcomes), and $X(s)$ is a relation between an element in the sample space and a real number. The range of $f(x)$ is [0, 1] so that the total probability of something occurring is 1:

$$\sum_x f_X(x) = 1$$

All this is to give us the probability of a certain event occurring.

Coin example.

pmf of both hands



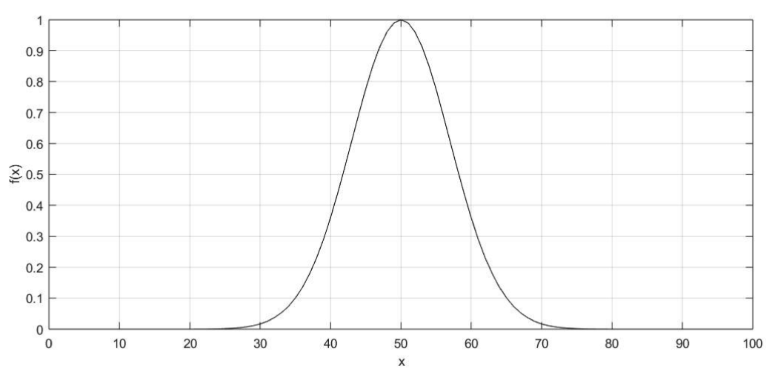
Is there a significant difference between the two?

Distributions worksheet.

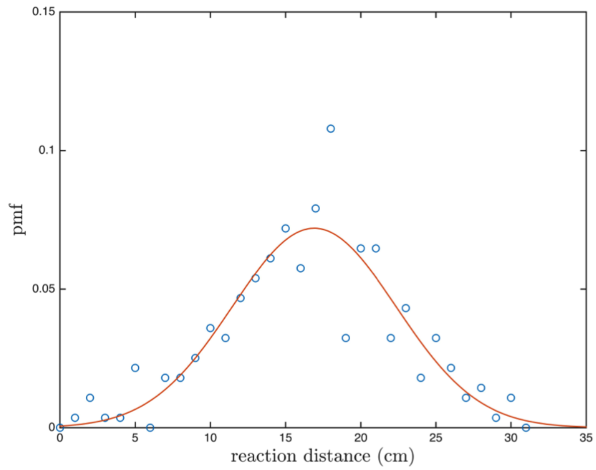
Gaussian Distribution

Many observations have a Gaussian shape (it looks like a bell).

$$f(x) = Ae^{-\frac{(x-x_0)^2}{2\sigma^2}}$$



pmf of Both Hands

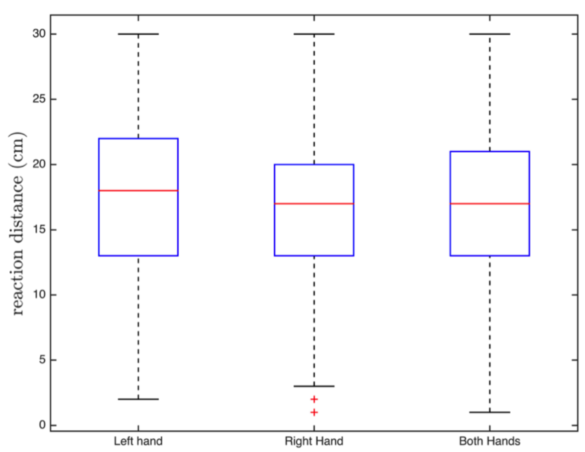


The pmf of reaction distance for both hands has a clear Gaussian distribution.

$$f(x) = Ae^{-\frac{(x-x_0)^2}{2\sigma^2}}$$

Curve fit parameters:	From definition:
$A = 0.072$	
$x_0 = 16.9$ cm	$x_0 = 16.7$ cm
$\sigma = 5.45$	$\sigma = 5.70$

Bar Plots



Another option is the bar plot to compare data from separate categories.
The red line is the median (50th percentile).
The lower blue line is the 25th percentile, and the upper blue line is the 75th percentile.
Is there a significant difference between the left and right hand?
What can we conclude about our hypothesis?