Basic statistics

- suppose that you perform multiple measurements of some phenomenon
- the “average” measurement
- the amount of variation in the measurements
- the order of the measurements

Average

- the arithmetic average is an estimate of the mean
- for N measurements x_i the *sample mean* is

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Average

- easy to implement using a for loop

```
function mu = average(x)

N = length(x);
sumN = 0;
for i = 1:N
   sumN = sumN + x(i);
end
mu = sumN / N;
```

Average

- you should use the function **mean** instead

```
>> mean([-1 0 1])
ans =
      0

>> X = [1  2  3;
        3  8  11];
>> mean(X)
ans =
       2  5  7
```

mean computes the average of each column for a matrix

Mean

- one problem with the mean is that it is sensitive to erroneous measurements

```
>> x = randn(1, 20);
>> x(21) = 100;
>> hist(x, 20);
>> mean(x)
```

Trimmed mean

- one solution is to use the trimmed mean
- to compute the trimmed mean
  - remove the smallest and largest $\alpha\%$ of the values then compute the mean

```
>> trimmean(x, 5)
```

Median

- the median is the middle value
- for N measurements x_i in *sorted order*
  - if N is odd, the median is the value of the element with index $(N + 1)/2$
  - if N is even, the median is the average of the elements with indices $N/2$ and $(N/2) + 1$

```
>> median(x)
```

Variance

- the variance is a measure of spread around the mean
- for N measurements x_i the *sample variance* is

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

```
>> x = randn(1, 100);          % "low" variance
>> hist(x, 20);
>> var(x)
>> x = 10 * randn(1, 100);     % "high" variance
>> hist(x, 20);
>> var(x)
```

Standard deviation

- the standard deviation is the square root of the variance
- for N measurements x_i the *sample standard deviation* is often calculated as

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

```
>> std(x)
```