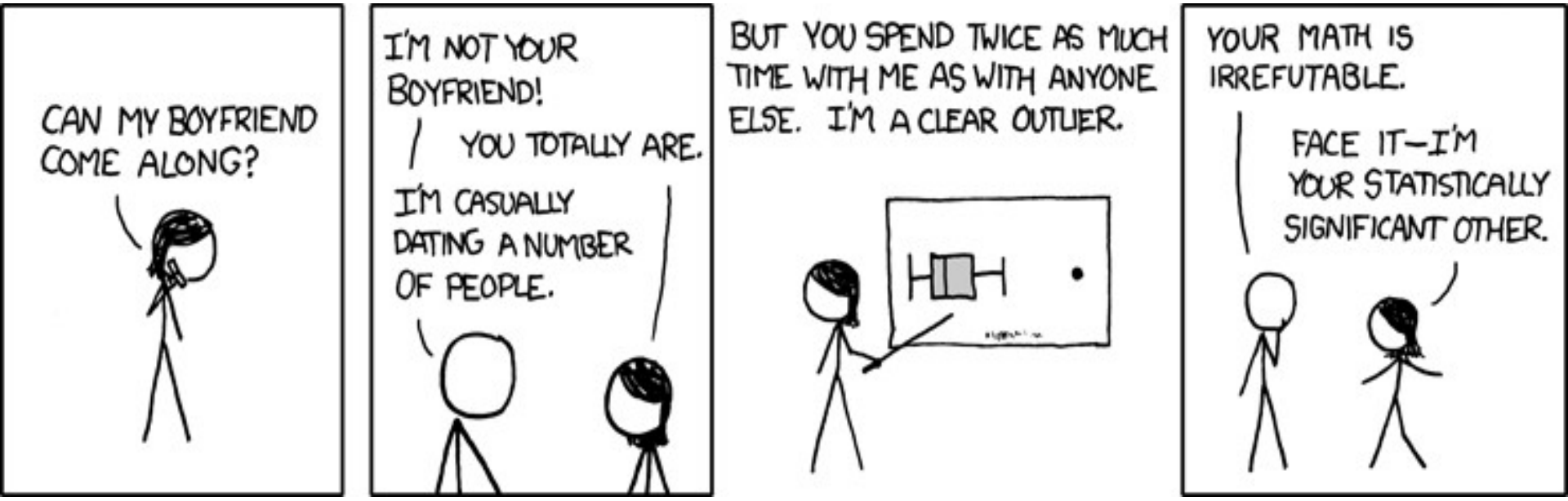
University Lecture Series for Research Methods on Sunday 10/9

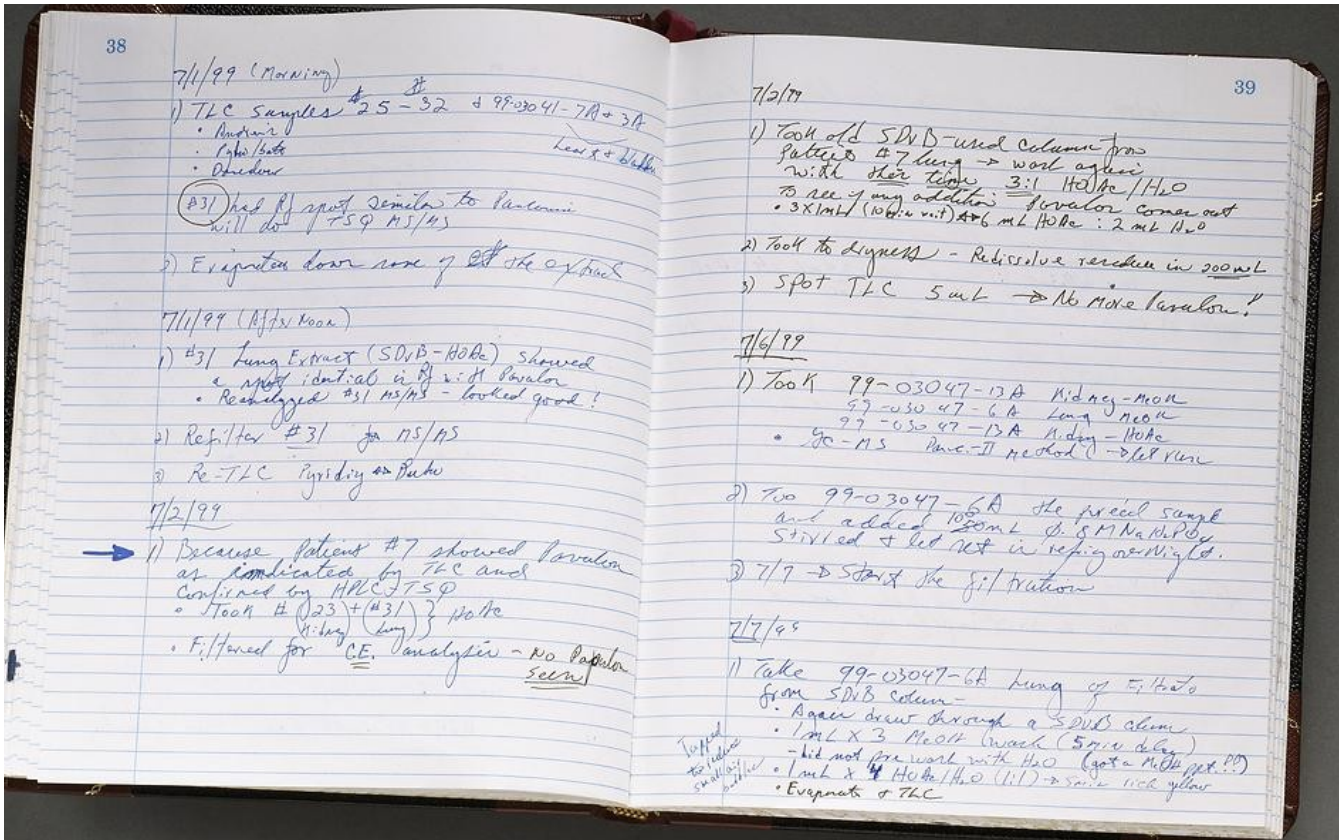
Today: comparing data sets



xkcd.com

- **In your lab notebook:** Write everything about your experiments. Each entry should have a date. Include notes (intro and conclusions), so when you, or someone else, go back to look at your notebook, the entries make sense.

Notebooks will be turned in as a HW later in the semester.



Watermelon seed spitting distances in meters:

Like Lady Gaga= 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 5, 5, 6, 7

Mean = 3.67 ±1.6 range = 2.07 to 5.27

Versus-

Not like Lady Gaga= 4, 4, 5, 5, 5, 6, 7, 7, 8, 8

Mean = 5.9 ± 1.52 range = 4.38 to 7.42

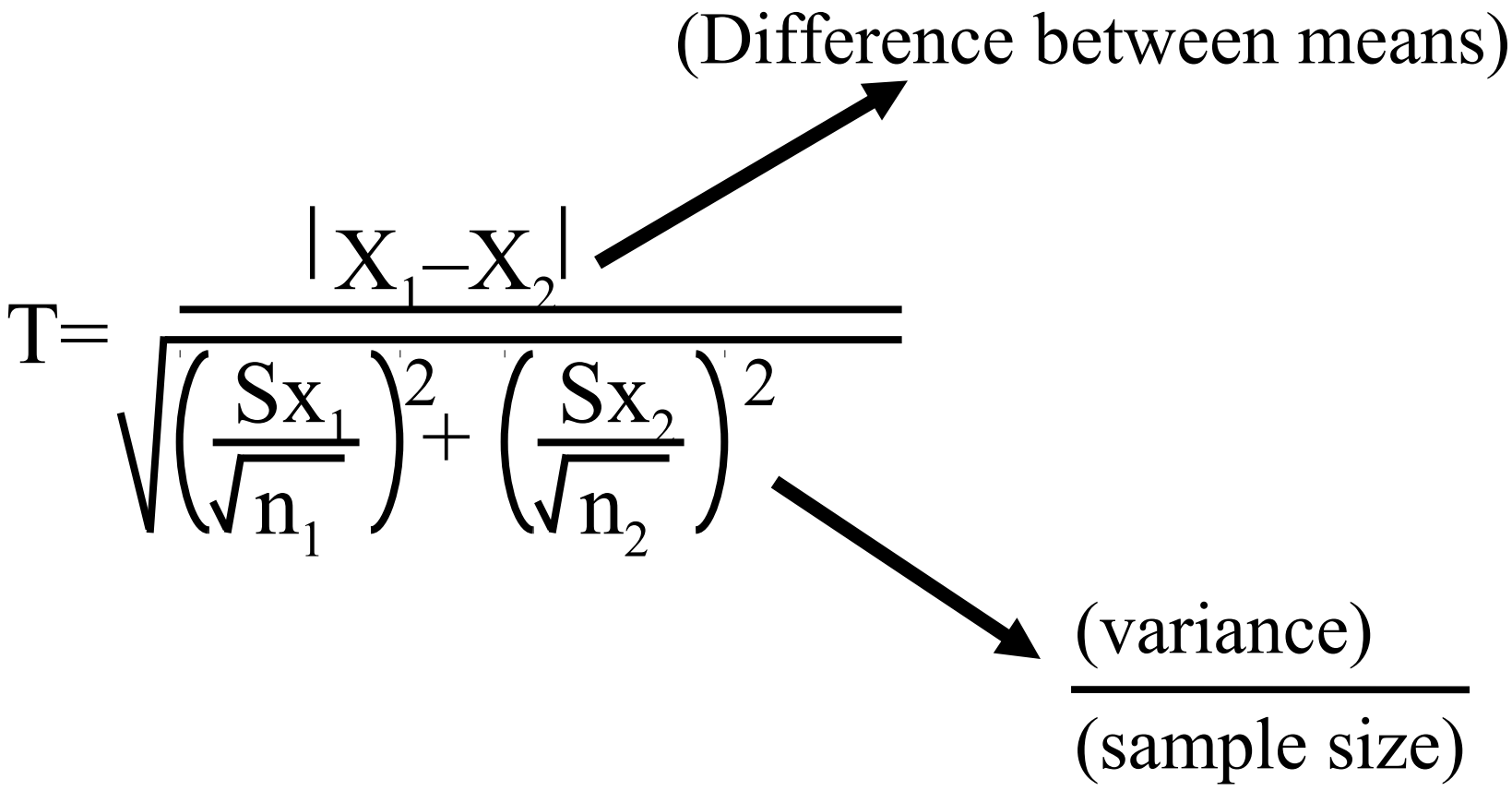
What information do we need to determine which group can spit watermelon seeds farther?

The ‘Students’ T-test is a method to assign a numerical value of statistical difference.

The ‘Students’ T-test is a method to assign a numerical value of statistical difference.

$$T = \frac{|X_1 - X_2|}{\sqrt{\left(\frac{Sx_1}{\sqrt{n_1}}\right)^2 + \left(\frac{Sx_2}{\sqrt{n_2}}\right)^2}}$$

The ‘Students’ T-test is a method to assign a numerical value of statistical difference.



The ‘Students’ T-test is a method to assign a numerical value of statistical difference.

$$T = \frac{|X_1 - X_2|}{\sqrt{\left(\frac{Sx_1}{\sqrt{n_1}}\right)^2 + \left(\frac{Sx_2}{\sqrt{n_2}}\right)^2}}$$

T is then used to look up the P-value from a table. Also need ‘degrees of freedom’ $= (n_1+n_2)-1$.

Partial table for determining P from T

	P-value		
Df	**0.05**	**0.02**	**0.01**
1	12.71	31.82	63.66
2	4.303	6.965	9.925

T

How significant of a difference is this?

Like Lady Gaga= 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 5, 5, 6, 7

Mean = 3.67 ±1.6 range = 2.07 to 5.27

Versus-

Not like Lady Gaga= 4, 4, 5, 5, 5, 6, 7, 7, 8, 8

Mean = 5.9 ± 1.52 range = 4.38 to 7.42

Spreadsheet T-test- Spreadsheets ask for 4 inputs: data set 1 and 2 then...
mode/tails- depends on your prediction about the direction of the difference between the groups. If you predicted group A would be lower than group B, pick 1 tail. If you predicted group B would lower than group A, pick 1 tail. If you didn't predict which would be lower/higher, use 2 tails. You can't change your mind after the data are gathered.

type/variance- There are three types of T-test that you can do. One type compares the same subjects in 2 different conditions. Like if you test whether heart rate increased after drinking a cup of hot sauce or whether plant growth would increase after adding fertilizer to pots of soil. In these cases you would be comparing the heart rate of the same people, or the growth of the same pot of plants, before and after the treatment. This requires a "paired" or "dependent" T test. Excel and Open Office call this a "type 1" test.

If you are comparing different subjects, this is an independent T-test. If you want to know whether nursing students consume more coffee than do biology students. You would then have two groups of test subjects rather than taking 2 measurements on each person. Now you would use an "unpaired" or "independent" T-test. Excel and Open Office call these "type 2" or "type 3" tests. Now the tricky part is to decide which of these to use. Are the standard deviations about the same for both groups, or are they different? If in doubt, go with "type 3" for unequal variances.

Information for mode and type adapted from Alverno College
(http://depts.alverno.edu/nsmt/stats.htm)

How significant of a difference is this? Using a speadsheet to get a P value = 0.002.

Like Lady Gaga= 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 5, 5, 6, 7

Mean = 3.67 ±1.6 range = 2.07 to 5.27

Versus-

Not like Lady Gaga= 4, 4, 5, 5, 5, 6, 7, 7, 8, 8

Mean = 5.9 ± 1.52 range = 4.38 to 7.42

How significant of a difference is this?

P value = 0.002. So the chance that these 2 sets of data are **not** significantly different is 0.002

Like Lady Gaga= 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 5, 5, 6, 7

Mean = 3.67 ±1.6 range = 2.07 to 5.27

Versus-

Not like Lady Gaga= 4, 4, 5, 5, 5, 6, 7, 7, 8, 8

Mean = 5.9 ± 1.52 range = 4.38 to 7.42

How significant of a difference is this?
P value = 0.002. So the chance that these 2 sets of data are significantly different is
1 – 0.002 or 0.998
We can be 99.8% certain that the difference is statistically significant.

Like Lady Gaga= 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 5, 5, 6, 7

Mean = 3.67 ±1.6 range = 2.07 to 5.27

Versus-

Not like Lady Gaga= 4, 4, 5, 5, 5, 6, 7, 7, 8, 8

Mean = 5.9 ± 1.52 range = 4.38 to 7.42

How certain do we need to be?

Generally a P-value of 0.05 or less is considered a statistically significant difference.

20% random difference : 80% confidence

10% random difference : 90% confidence

5% random difference : 95% confidence

1% random difference : 99% confidence

0.1% random difference : 99.9% confidence

T-test is one valid and accurate method for determining if 2 means have a statistically significant difference, or if the difference is merely by chance.

Samples vs populations

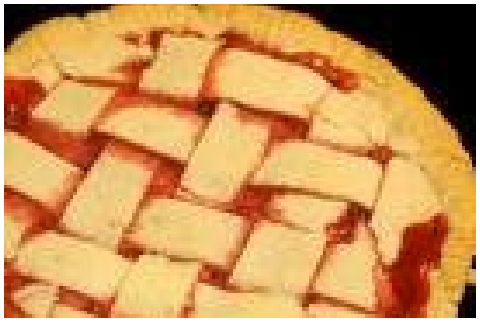








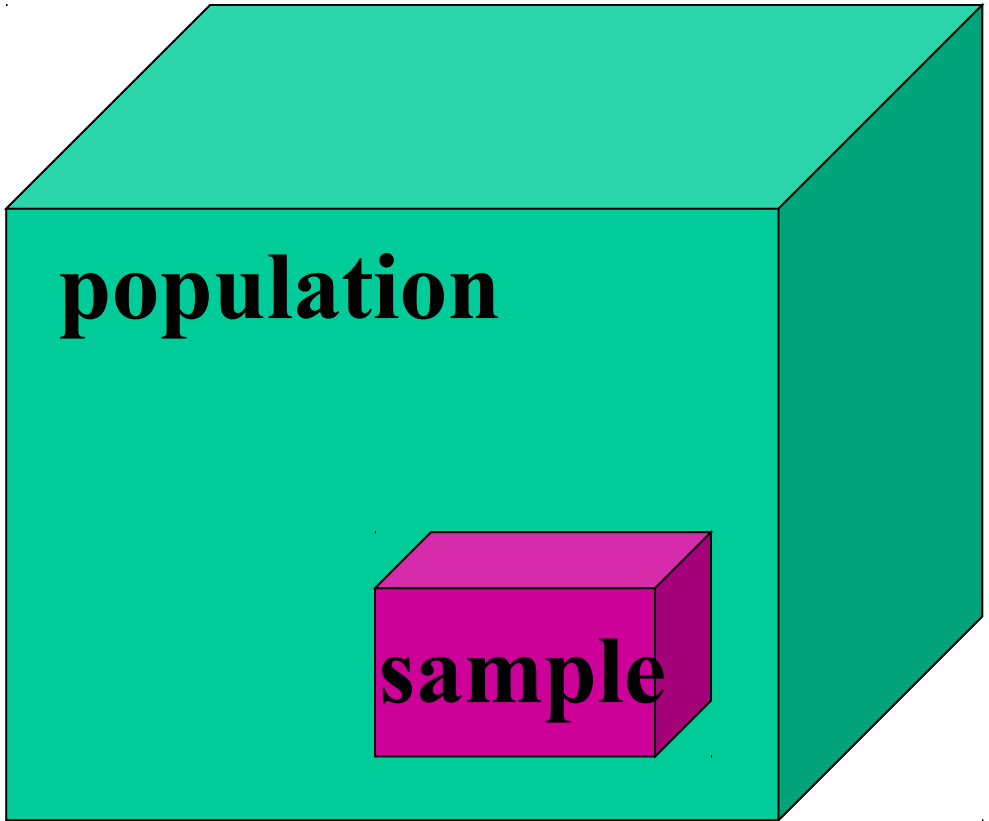





Samples vs populations

Population- everything or everyone about which information is sought

Sample- a subset of a population (that is hopefully representative of the population)



Population-

- U.S. census
- Dogs
- 1 – infinity

Sample-

- Travis county
- Poodles
- Prime numbers

Why use a sample instead of a population?

Why use a sample instead of a population?

- Logistics
- Cost
- Time

What sample size do you need?

It depends on the error you expect.

To determine an appropriate sample size, you need to estimate a few parameters.

- Means
- Standard Deviation

- Power:

The probability that an experiment will have a significant (positive) result, that is have a p-value of less than the specified significance level (usually 5%).

This calculator will help you determine the appropriate sample size:

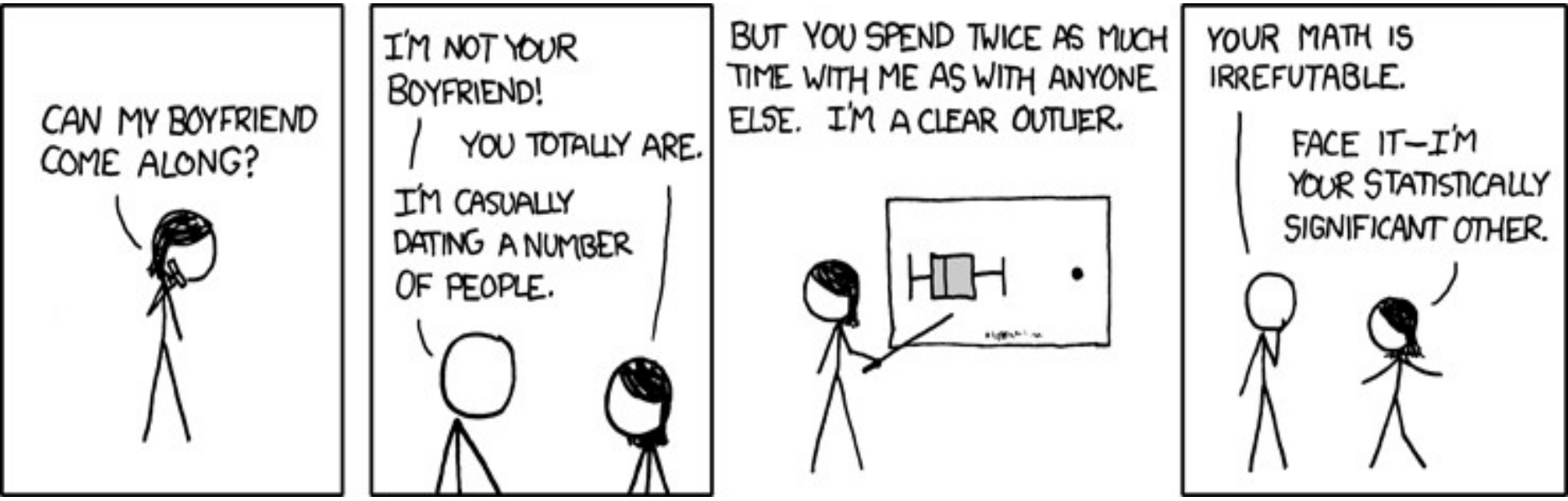
http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html

What sample size do you need?

It depends on the error you expect.

(So it is impossible to predict with 100% accuracy before the experiment is carried out.)

University Lecture Series for Research Methods on Sunday 10/9



xkcd.com