Statistical Testing

FOR ABMS

Disclaimer

The following set of slides introduce Statistics from the perspective of a computer science professor interested in running simulations: ABMS, Cognitive Modeling, etc.

Deeper statistical (gotcha) questions and/or philosophical discussions should be routed to a statistician.

We have a very good one in our campus: Dr. Mainak. He is always happy to help those (including me) who have doubts on statistics.

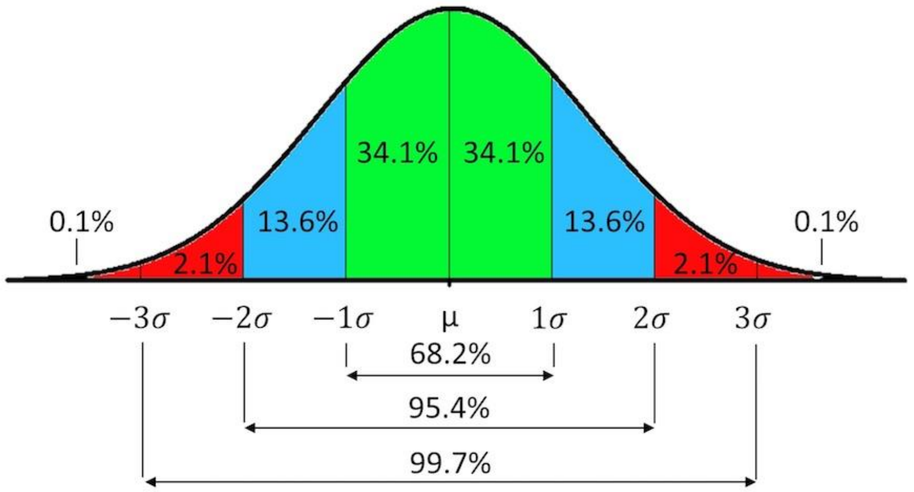
Nature or Man-made

Normal Distribution

if data tends to be around one central value then we say it is normally distributed

The spread, or distribution, has symmetry around the center (mean)

- 50% of the data falls to the left of the center
- 50% of the data falls to the right



Standard Normal Distribution

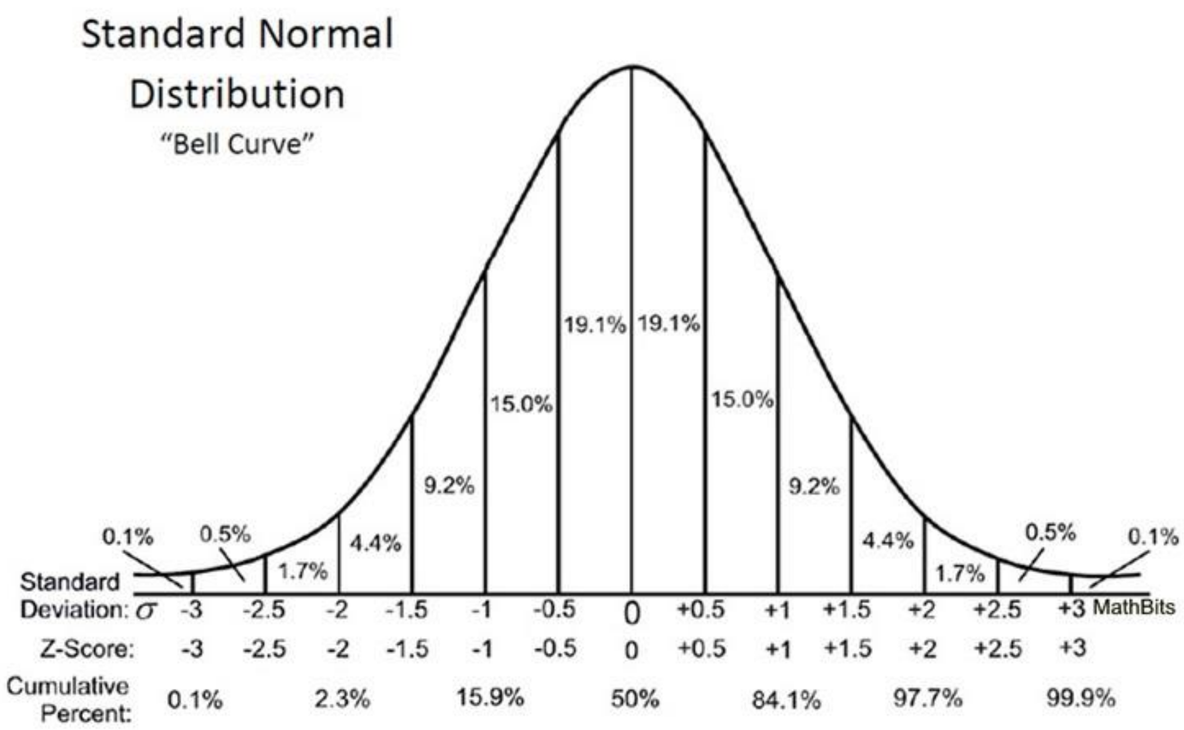
Normal Distribution with mean = 0 and standard-dev = Z

We can convert any normal distribution into a standard normal distribution

https://www.mathsisfun.com/data/standard-normal-distribution.html

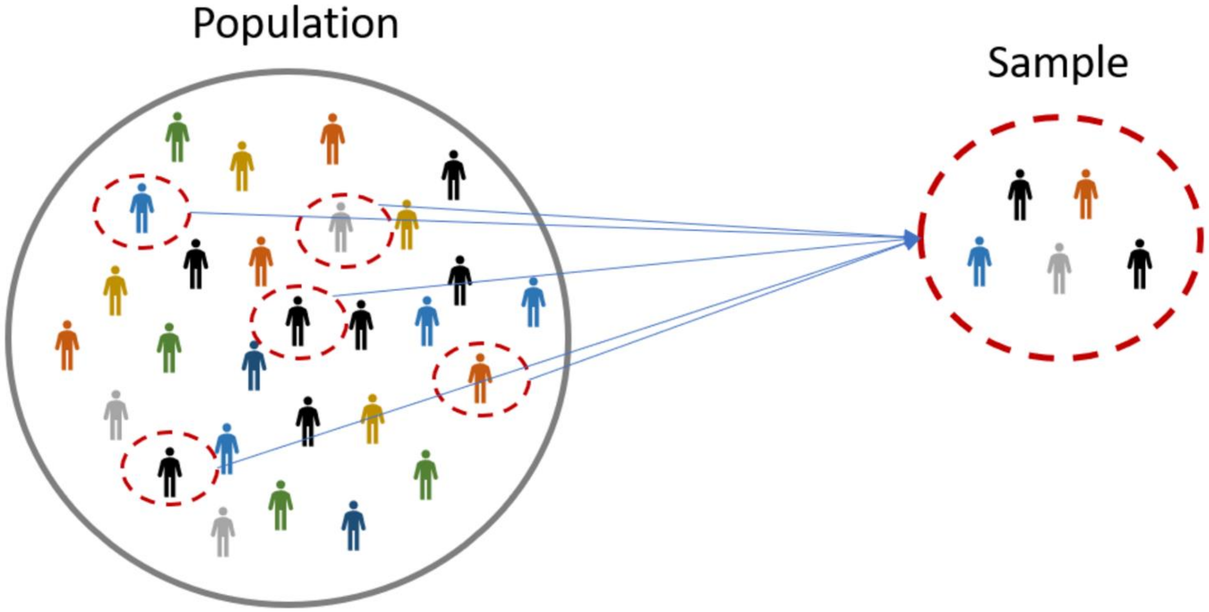
Once you are done, you have a easy job

https://www.mathsisfun.com/data/standard-normal-distribution-table.html



Sampling

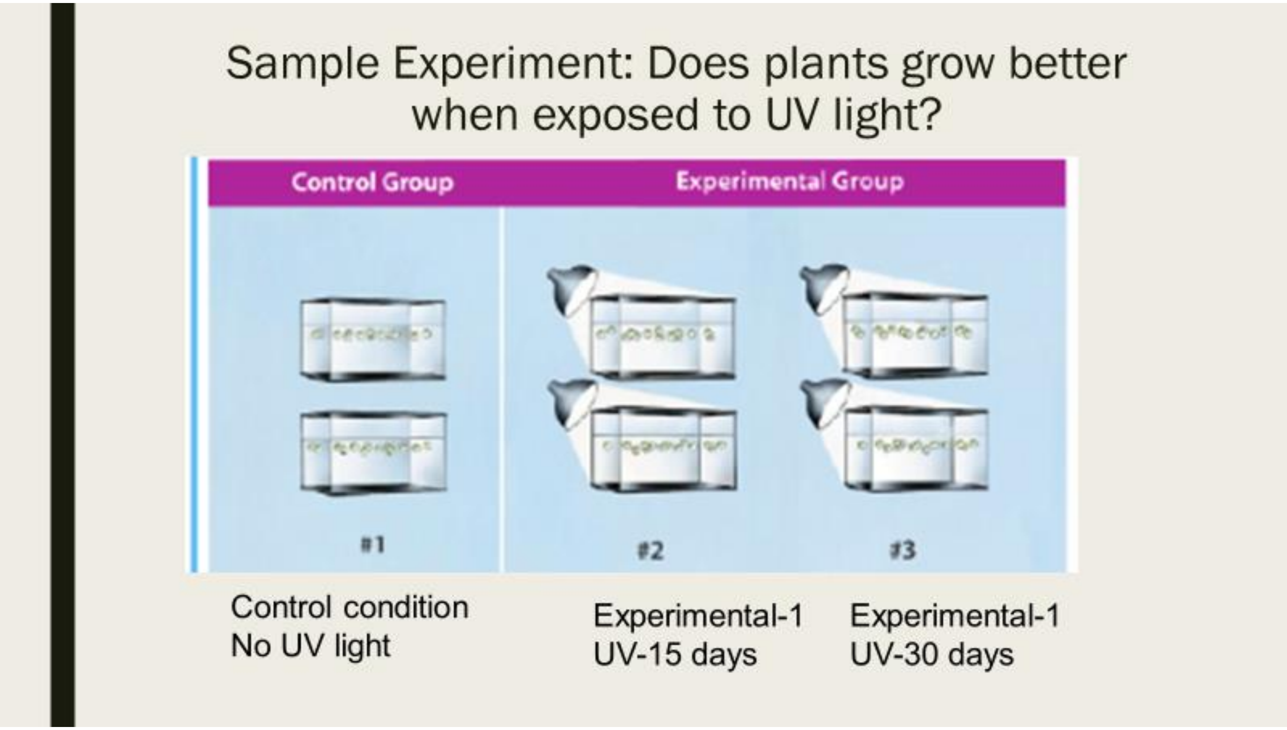
sampling is the selection of a subset of individuals from within a population to estimate characteristics of the whole population



Experimentation requires Sampling

Lets look at our UV light experiment to grow plants. We select a limited number of samples to test our hypothesis.

We can't test the whole population of plants



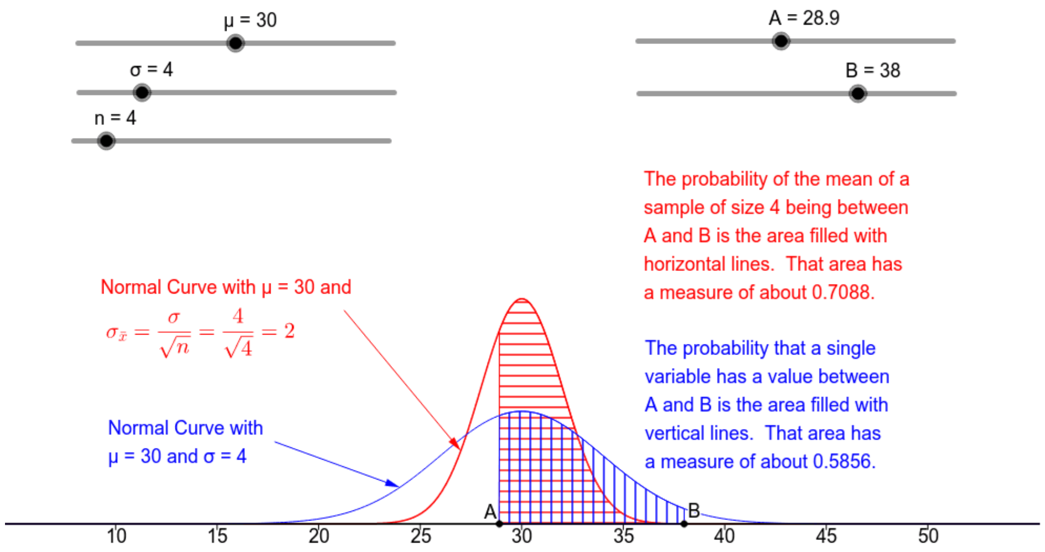
Central Limit Theorem

" Repeated samples of the same size taken from a population will have a **distribution of the sample means** that follows the Normal distribution "

https://www.geogebra.org/m/zshvnvuj

So what happens if you repeat the experimentation?

We repeat ***experimental runs***



Hypothesis testing exploits the central limit theorem

Since this theoretical **distribution of sample means** is a normal distribution, it has its own standard deviation. This standard deviation is called The **Standard Error of the Mean**.

Its important to note that the mean of sample-means is the same/close to the theoretical population mean

Remember the following

- Any normal distribution can be converted into the standard form
- We can compute the Z score of any data-point, which gives us the probability
- We have the following
  - Hypothesized population mean (which will the same as mean-of-means) [Control mean]
  - A sample-mean (the data point)
  - An estimate (Z-score) for the data point with which we can see how unlikely it is

Nothing changed!

Null hypothesis – Nothing changed (No effect)

The entire hypothesis testing revolves around whether to accept or reject the null hypothesis

Example: In the UV experiment the **null hypothesis** is the exposure to UV had no (significant) change to the growth of the plants

How its done

1. We will do experiment and assume the control-mean as representative of the population mean
2. We will take experimental-mean as the new sample-mean (data point)
3. We can get an estimate of the standard-error (std dev of distribution-of-means)
4. We can convert the data point to a Z-score and we can see the likelihood of the sample-mean. (is it in the extremes)
5. If its in the extreme (less than 5% region area) then we reject the null hypothesis: we say that the population-mean != the-population-mean-of-sample
   - It is less likely that the ‘sample’ is from the current population

Estimating the standard error from sample-mean

https://www.radford.edu/~biol-web/stats/standarderrorcalc.pdf

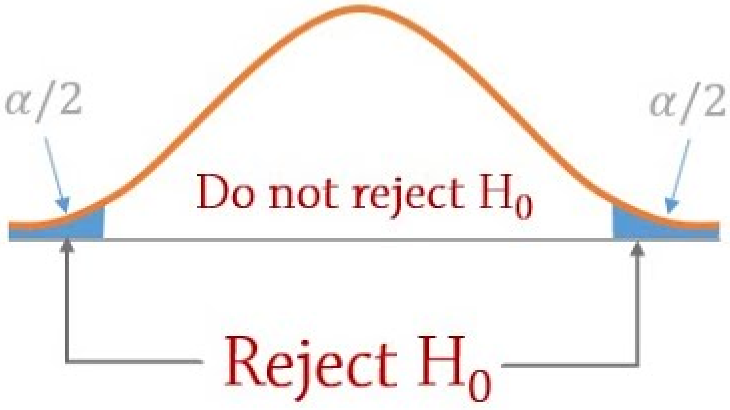
Divide the standard deviation by the square root of the sample size (n). That gives you the "standard error".

Hypothesis Testing

One-tailed

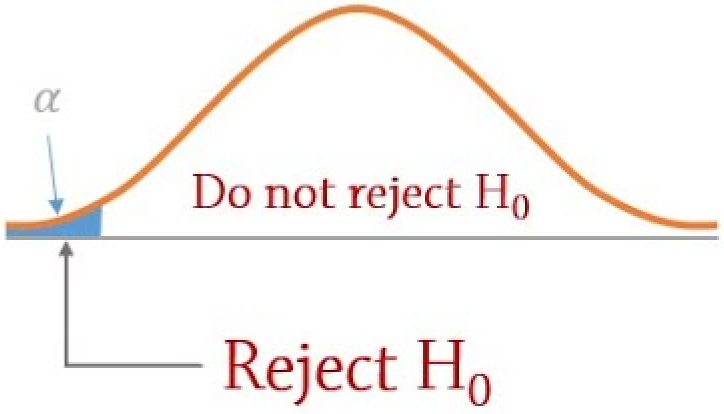
Two-tailed

$$H_0: \mu = 23$$
$$H_1: \mu \neq 23$$



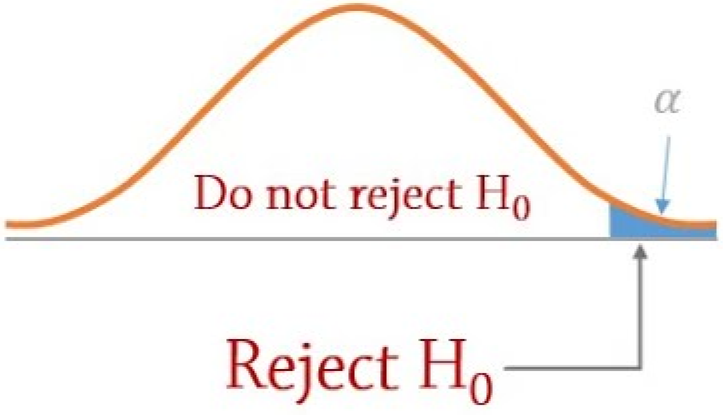
Left-tailed

$$H_0: \mu \geq 23$$
$$H_1: \mu < 23$$



Right-tailed

$$H_0: \mu \leq 23$$
$$H_1: \mu > 23$$



OK I get it!

There is a theoretical population of plants

- On average they grow 6 ft tall (we know from historical data)
- There are extremes some grow very small 1 ft and some up to 11 ft. But they are very rare
- The growth is normally distributed
- I can not test each and every plant to verify if UV helps, so I sample (50 plants for control and 50 for experimental-UV)
- (say) the control average is 5.5 ft
- Is my original estimate of 6 ft from historic data wrong?
  - It could just be a sampling error right?!
- (say) the experimental-UV average is 7 ft and std-dev X
- Did UV help? Or is it also a sampling error? (accidentally sampled very good plants)
- If only We can estimate the likelihood of me choosing a sample such as this, wouldn’t it be nice
  - That’s what hypothesis testing does

Tests to use

T-test is the sample is smaller than 35 (I recommend not less than 20)

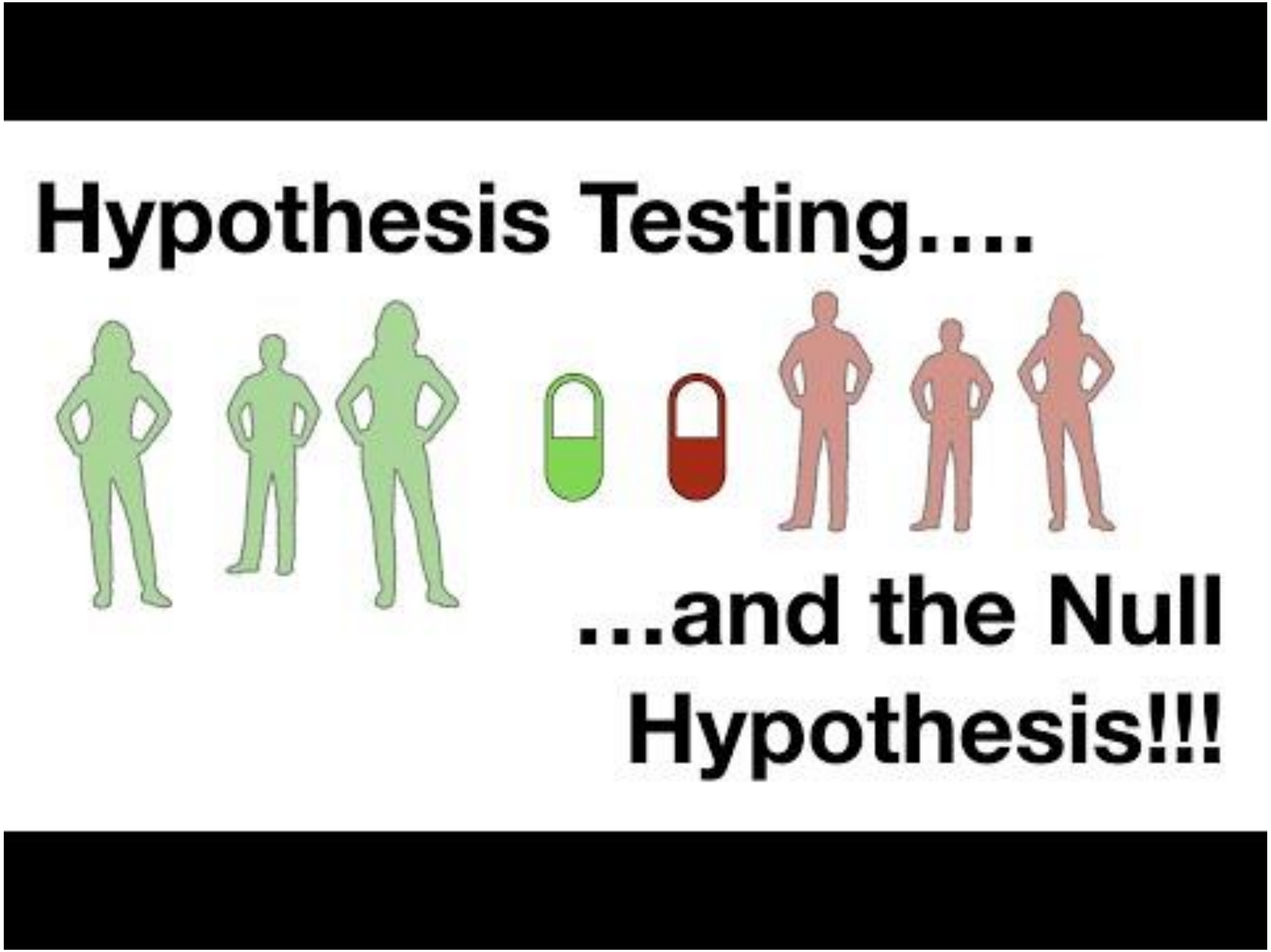
ANOVA if comparing more than 3 means control vs experiment-1 vs experiment-2

- We are not that lazy right can we do multiple t-tests?

Chi-square - never mind….!

And the winner is

T-test!!


Hypothesis Testing….
…and the Null
Hypothesis!!!