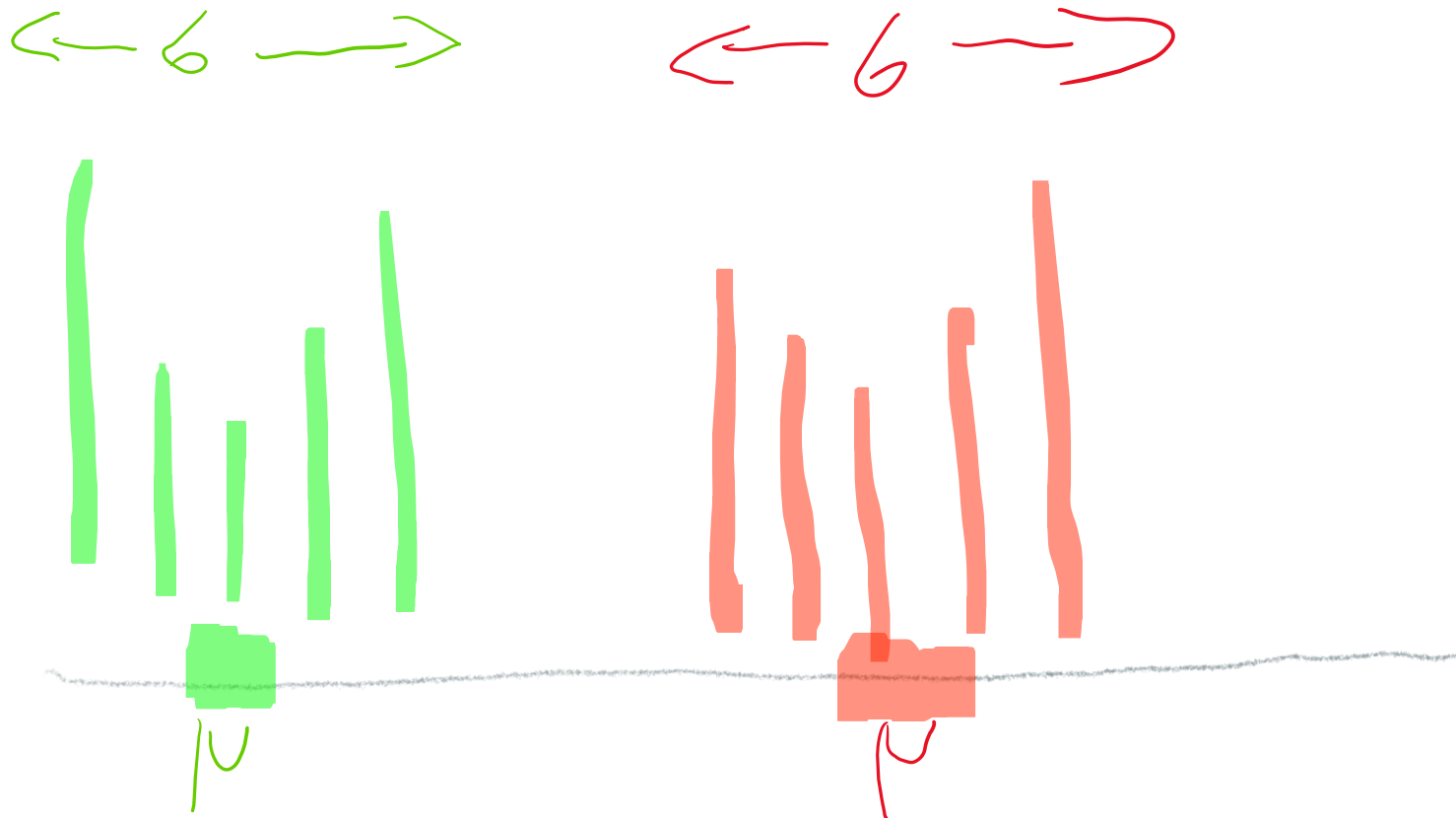# Power Analysis

Andres Namm

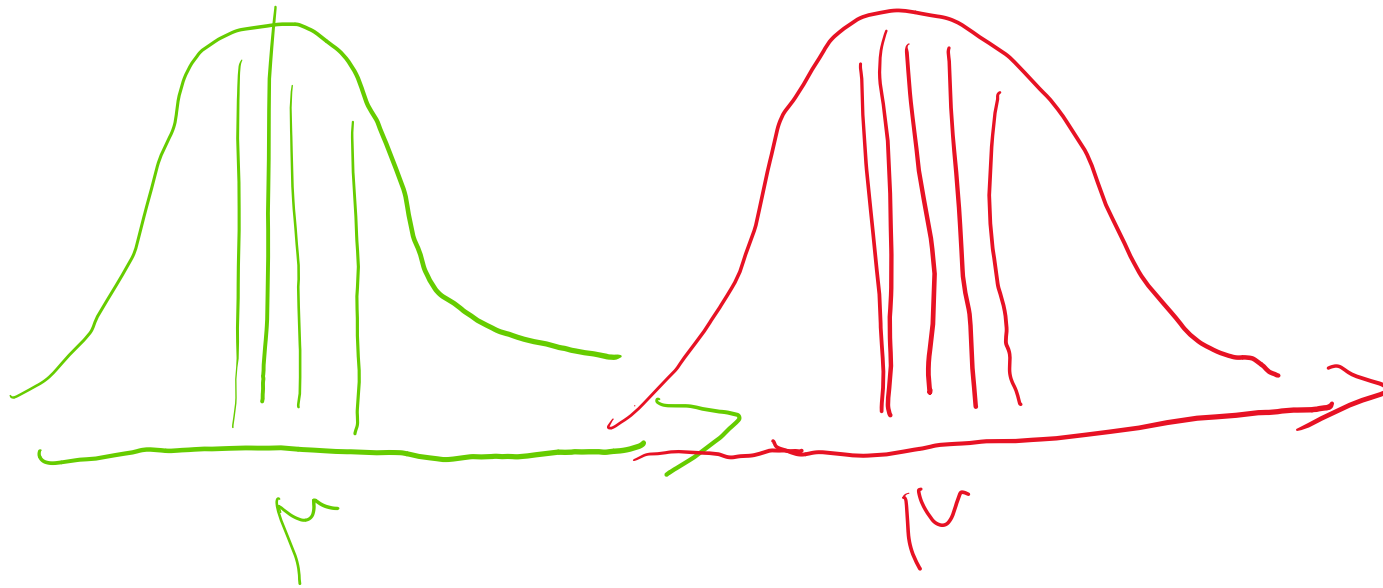# We have distributions over 2 groups of data – A & B

Both of them have their **GLOBAL** means and **GLOBAL** standard deviations.

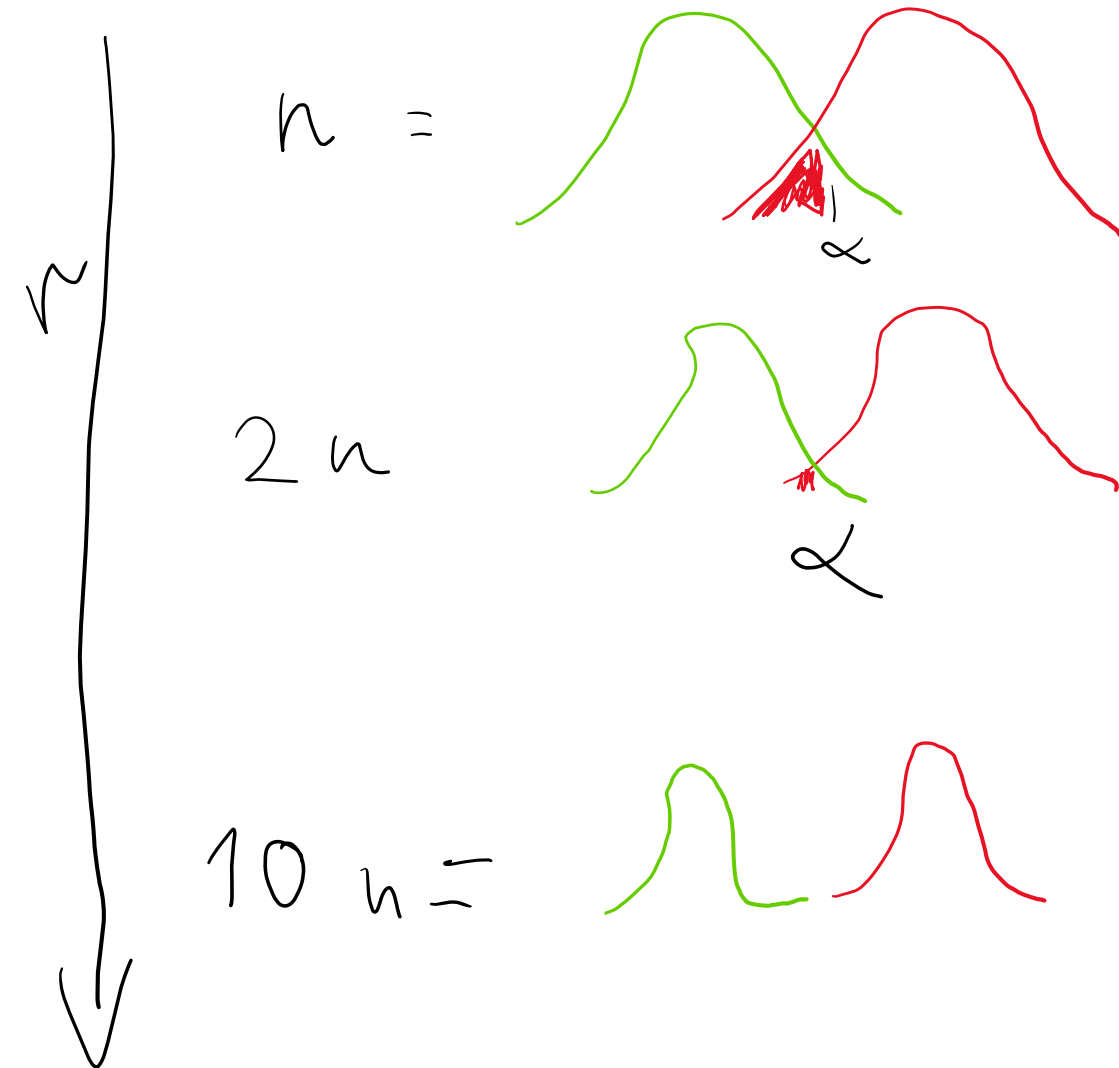# Based on CLT if we start to sample from these datasets

Their means of the samples would center around the global mean

$$STD = \frac{\sigma}{\sqrt{n}}$$

The only thing that would vary is the std of both distributions over sample means. Based on n = sample size

What this means?

The larger the n (sample size) , the tighter the distributions over samples.

When performing Hypothesis tests comparing 2 samples. Tighter distributions means less chance of accidentally taking data from red distribution which is inside the assumed possible are for green distribution.

Power analysis helps us to determine the n we need to use to get tight enough distribution over the sample means (or some other variable)

Visual
Example
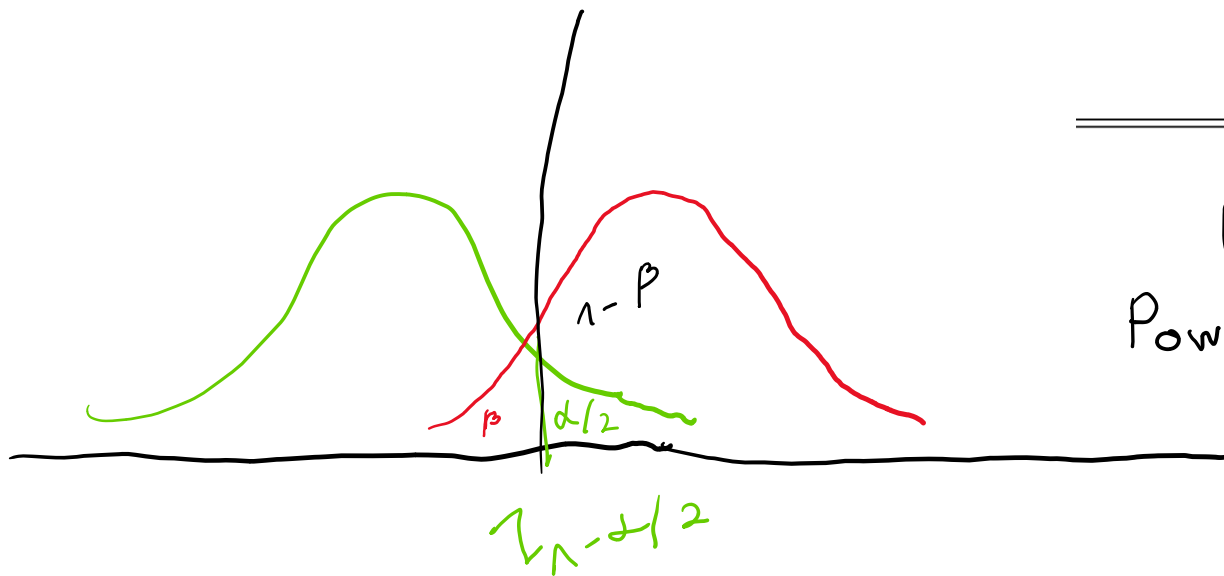
$$n_i = 2 \left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES} \right)^2$$

$$ES = \frac{|\mu_1 = \mu_2|}{\sigma}$$

Comes directly from data.
1) Can be pooled STD from 2 datasets
2) Can be some estimated number from total dataset with multiple groups.

$1-\beta$

β $\alpha/2$

$Z_{1-\alpha/2}$

Usually
Power $= 1 - \beta = 80\% \Rightarrow Z_{1-\beta} \approx 0.84$
$\alpha = 5\% \Rightarrow Z_{1-\alpha/2} \approx 1.96$