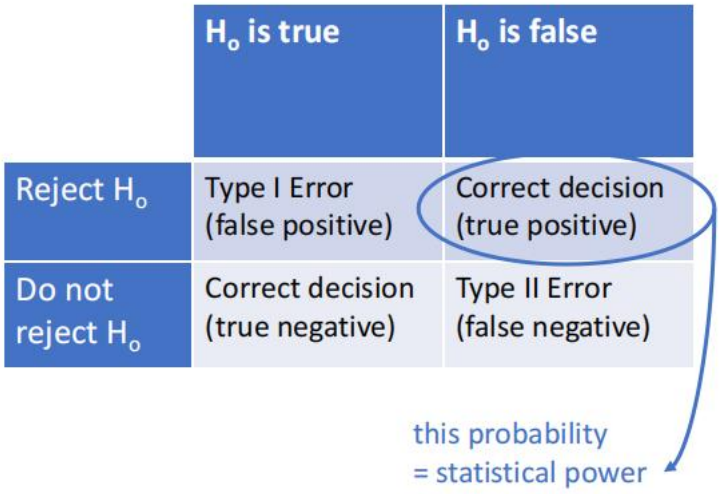
The lower the variability, the higher probability that we will reject the null hypothesis.
Statistical power is the probability that the statistical test will reject a false null hypothesis. (即，HO 是 false 的，而且我们也正确地把 HO 给 reject 了)
Or, in plain English, Statistical power is the likelihood that a study will detect an effect when there is an effect there to be detected.

文心一言
Statistical power is the probability of rejecting a null hypothesis when it is false.
a measure of the sensitivity of a statistical test to detect effects or differences when they truly exist.
A higher power value indicates a greater chance of detecting a true effect, while a lower power value indicates a greater chance of failing to detect a true effect (i.e., a type II error).
Factors: sample size, effect size, significance level, data distribution.

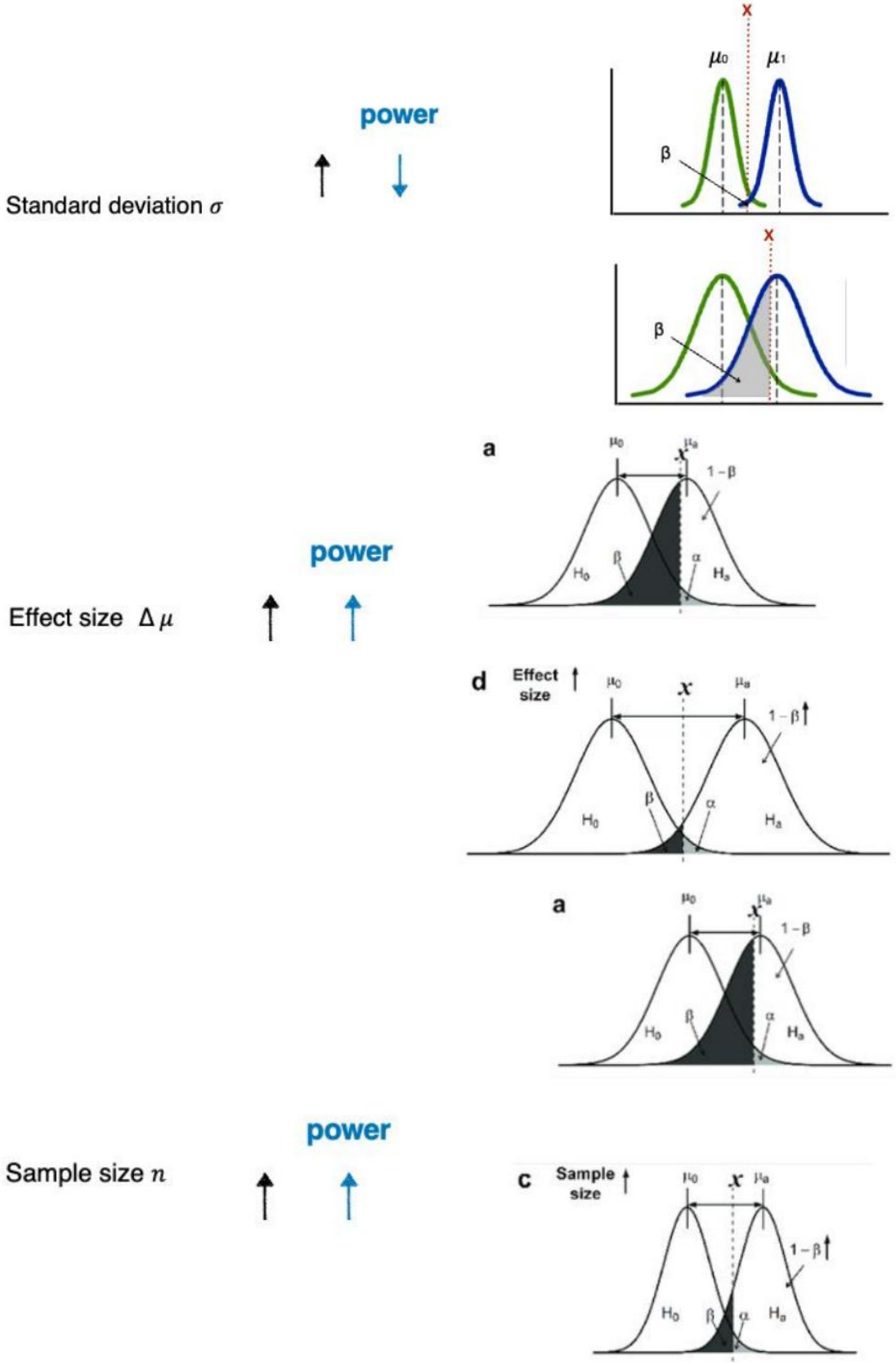



Type I error
• A Type I Error is rejecting the null hypothesis when it is true.
• Prob(Type I Error) = Significance level = P(reject Ho|Ho true)
Type II error
• A Type II error is not rejecting a null hypothesis when it is false.
• Prob(Type II Error) = β = P(accept Ho|H1 true)
• Value of β typically depends on which particular alternative hypothesis is true.

Power of a hypothesis test
• Power = 1 - β = P(reject Ho|H1 true)
• Probability of rejecting the null hypothesis if the alternative hypothesis is true

In clinical trials, Phase III: industry minimum power = 80%

Factors



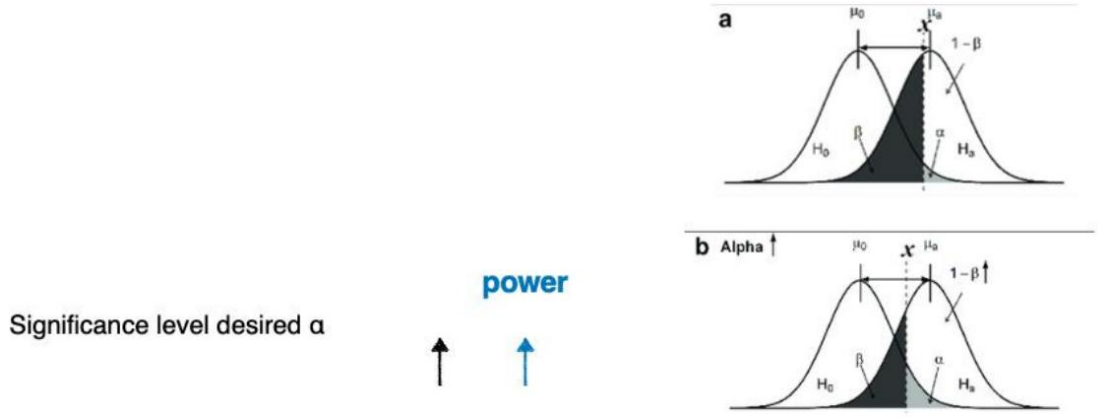

这个是有道理的。不是说 increase sample size 改变了 standard deviation，我之前这样理解是谬论。如果把样本重复几遍，均值方差标准差都是不变的，但是依然增大了 power。

而且还有一个惊人的发现，就是 increase sample size 之后居然更加 normal 了，因为 shapiro test 的 p-value 变小了!这应该就是上面文心一言回答我 power 与什么有关的时候告诉我与 data distribution 有关的原因。

文心一言

您提到的分母是 n-1 的情况通常是在计算样本方差时使用的，这是为了修正由于样本估计总体均值时带来的偏差。但是，当您明确知道整个数据集（即总体，而不是样本）的均值时，应该使用 n 作为分母来计算方差。

而我们这里画的是 population distribution，不应该单纯从 sample 意义上分析这个问题。




power.t.test()

做这个的时候比如说问你 sample size 要到达多少才行，就用 power.t.test()算出来一个小数之后，再用 power.t.test()验证一下

注意如果 type = "two.sample"，无论你给定函数的 n 还是函数算出来的 n 都是 number in each group,回答问题的时候要说明像这样：You require 46 students in each group (heavy vs non-heavy drinkers).

Problem set 比较坑，他是这么问的 What happens to our statistical power (increases or decreases) if we decrease our sample size to 10? 我们做的话是要理解成总样本量为 10，即两个组加起来样本量为 10，也就是说 two-sample 的话 5 observations in each group