7. Statistical Power

- Consider now that you plan an experiment with a treatment and control group
- But how many observations do you need in treatment and control group?
- In the past:
  - Experimental researchers often used simple rules of thumb
  - or decided based on the money available
- One problem:
  - If the treatment coefficient is not significant - is this due to the absense of an effect or a lack of sample size?
  - Common (but problematic) practice of collecting more observations ex-post
- Analysis of statistical power helps to plan ahead in a more structured manner

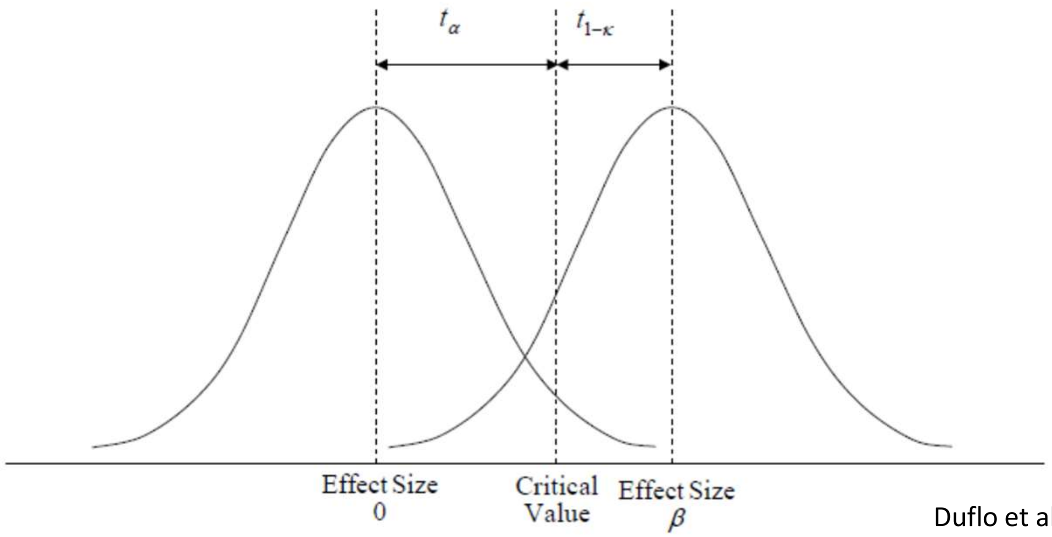
7.1 The Key Idea

Statistical Power

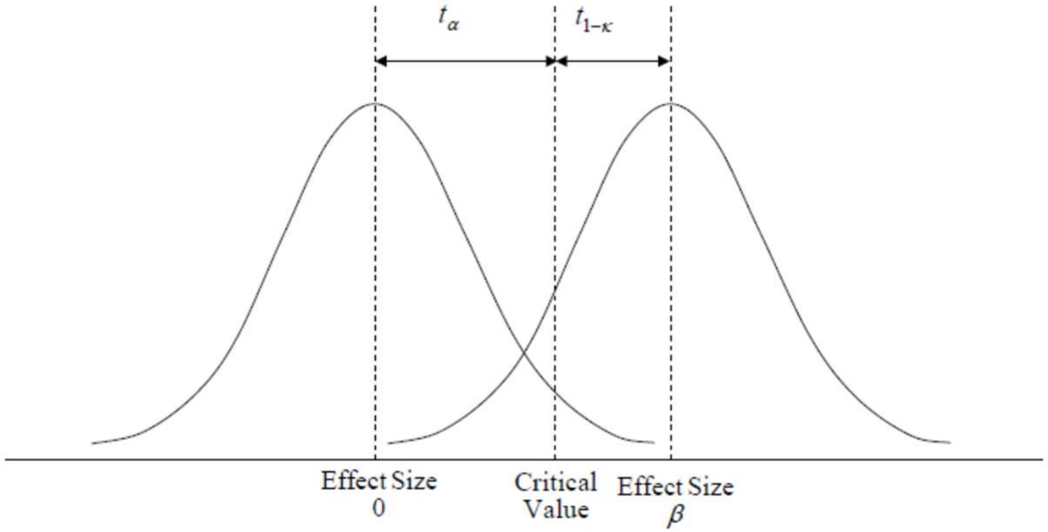
- Aim: find a measure with which you can assess the *ability of a statistical test to detect an effect, if the effect actually exists*
- Note you can thus determine the power only for a specific effect size
- Question: "*If the true effect is equal to a specific value of* $\beta > 0$, *what is the likelihood that the regression I run on my sample shows that the estimated* $\hat{\beta}$ *is statistically significant?*"
- The *statistical power* is the probability that,
  - for a given effect size and given statistical significance level,
  - we will be able to reject the hypothesis that $\beta = 0$, i.e. that the treatment has no effect
- $1 - power$ is the probability of falsely not rejecting the null hypothesis (type II error)

If we plan to test $\boldsymbol{H_0}$: $\boldsymbol{\beta = 0}$

- The significance level α is the probability of a type I error, i.e. the probability that we reject H_0 if it is in fact true
- For a given significance level, H_0 will be rejected if $\left|\frac{\hat{\beta}}{SE(\hat{\beta})}\right| > t_{\frac{\alpha}{2}}$ where $t_{\frac{\alpha}{2}}$ is the percentile in the t-distribution, i.e. $\Pr\left(|t| > t_{\frac{\alpha}{2}}\right) = \alpha$
- Suppose now that the true impact is $\beta > 0$



Duflo et al (2007)



- The distribution on the left is the distribution of $\hat{\beta}$ under H_0
- For a given α, H_0 will be rejected if $\left|\frac{\widehat{\beta}}{SE(\widehat{\beta})}\right| > t_{\frac{\alpha}{2}}$
- The distribution on the right is the distribution of $\hat{\beta}$ if the true impact is equal *to a specific* $\beta > 0$
- The power of the test for a true effect size of β is
  - the probability mass in the area under this curve
  - that falls to the right of the critical value $t_{\frac{\alpha}{2}}$
- It is the probability that we reject H_0 if it is false (& the true effect is β)

When do we have sufficient power to detect an effect of a specific size?

- To achieve a power of κ we thus must have that

$$\frac{\beta}{SE(\hat{\beta})} > t_{1-\kappa} + t_{\frac{\alpha}{2}}$$

- Only if this condition holds, an effect of size β will be significant at a level α with a probability larger than κ

Typical values used in the literature:

- For a power of 80%: $t_{0.2} \approx 0.84$
- with a significance level of 5% in a two-sided test: $t_{0.025} \approx 1.96$
- we must have that roughly

$$\frac{\beta}{SE(\hat{\beta})} > 2.8$$

7.2. Statistical Power in Experiments

- Suppose we are planning an experiment
- Treatment variable C_i indicates whether a subject is part of the treatment

$$Y_{C_i i} = \begin{cases} Y_{1i} & if \quad C_i = 1 \\ Y_{0i} & if \quad C_i = 0 \end{cases}$$

- Assume a population regression (i.e. a true treatment effect of β)

$$Y_i = \alpha + \beta C_i + \varepsilon_i$$

- When C is randomly assigned, the CIA holds and β is the causal effect of the treatment
- By testing H_0: $\beta = 0$ we can test whether the treatment has a statistically significant effect on Y

- When conducting experiments we can (often) decide on the
  - total sample size n and the
  - proportion of treated objects p
- Recall that when C is a dummy then $\hat{\beta}$ is just the difference between the sample means between treatment and control group

$$\hat{\beta} = \underbrace{\frac{1}{pn}\sum_{C_i=1} Y_i}_{estimate\ of\ E[Y|C_i=1]} - \underbrace{\frac{1}{(1-p)n}\sum_{C_i=0} Y_i}_{estimate\ of\ E[Y|C_i=0]}$$

- Intuitively: we want to design our experiment such that we mimimize the variance in $\hat{\beta}$

$$\hat{\beta} = \frac{1}{pn}\sum_{C_i=1} Y_i - \frac{1}{(1-p)n}\sum_{C_i=0} Y_i$$

Hence

$$V[\hat{\beta}] = \frac{1}{(pn)^2} V\left[\sum_{C_i=1} Y_i\right] + \frac{1}{\left((1-p)n\right)^2} V\left[\sum_{C_i=0} Y_i\right]$$

$$= \frac{1}{pn}\sigma^2 + \frac{\sigma^2}{(1-p)n} = \left(\frac{1}{p} + \frac{1}{1-p}\right)\frac{\sigma^2}{n}$$

$$= \frac{1}{p(1-p)}\frac{\sigma^2}{n}$$

The variance of the estimate decreases (it becomes more precise) if

- n is large
- $p = 0.5$ (as $\operatorname{argmax}_p p(1-p) = 0.5$) i.e. treament & control group of the same size

The Minimum Detectable Effect Size

- Now we can use that

$$V[\hat{\beta}] = \left(SE(\hat{\beta})\right)^2 = \frac{1}{p(1-p)}\frac{\sigma^2}{n}$$

- Hence the condition $\frac{\beta}{SE(\hat{\beta})} > t_{1-\kappa} + t_{\frac{\alpha}{2}}$ is equivalent to

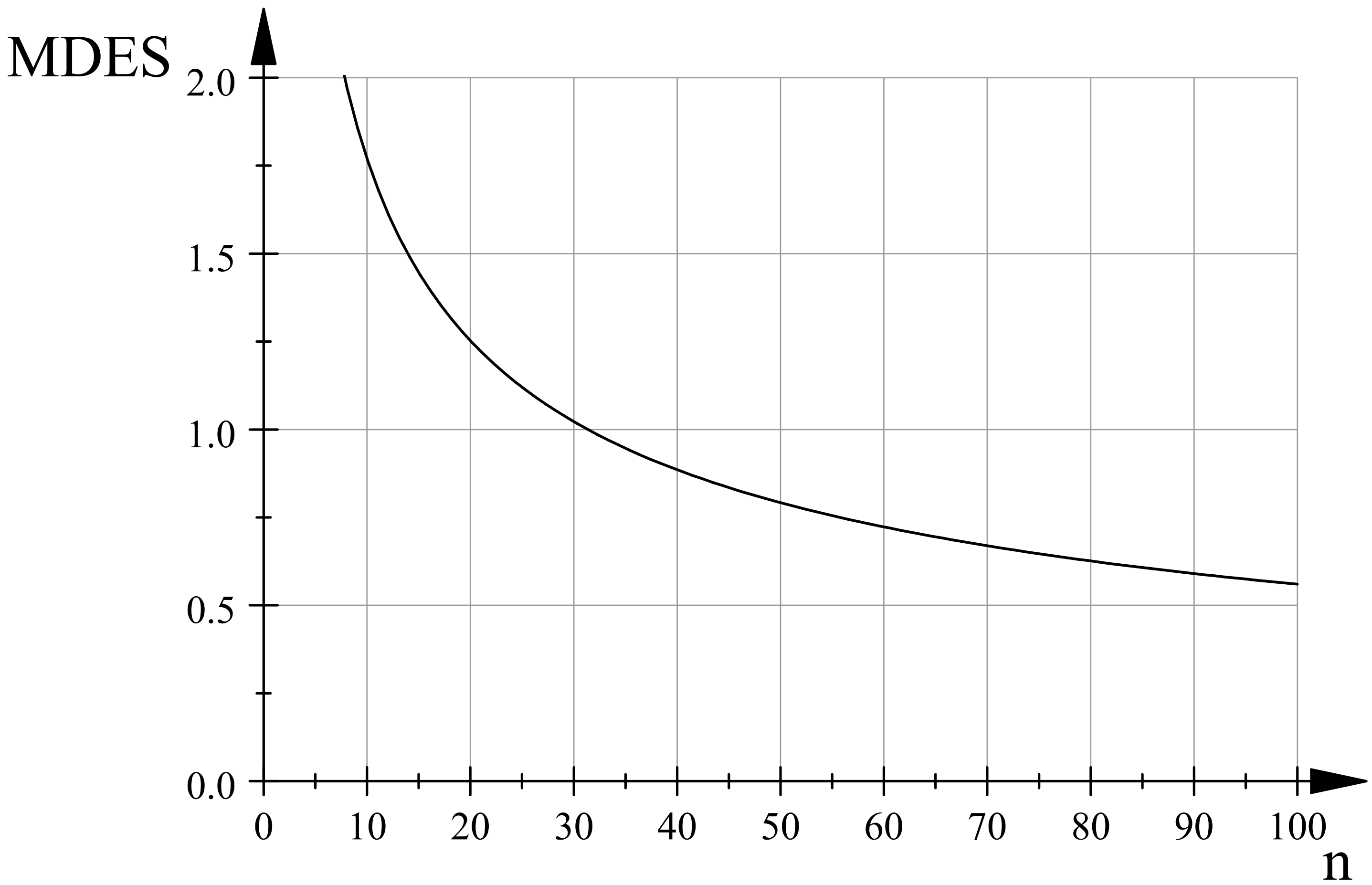
$$\beta > \left(t_{1-\kappa} + t_{\frac{\alpha}{2}}\right)\sqrt{\frac{1}{p(1-p)}\frac{\sigma^2}{n}}$$

- The value of β that satisfies this expression is the *minimum detectable effect size* for a given power (κ), significance level (α), sample size (n) and proportion of treated units p

- For $p = 0.5$, $\alpha = 0.05$ and $\kappa = 0.8$ and using the approximate values $t_{0.2} \approx 0.84$ and $t_{0.025} \approx 1.96$

$$\beta > 5.6\frac{\sigma}{\sqrt{n}}$$

With n observations & $\sigma = 1$, the *minimum detectable effect size* $5.6\frac{\sigma}{\sqrt{n}}$ is



Determining Sample Size

- An important question in any experiment:
  How many observations do we need to detect a specific effect size?
- We can rearrange

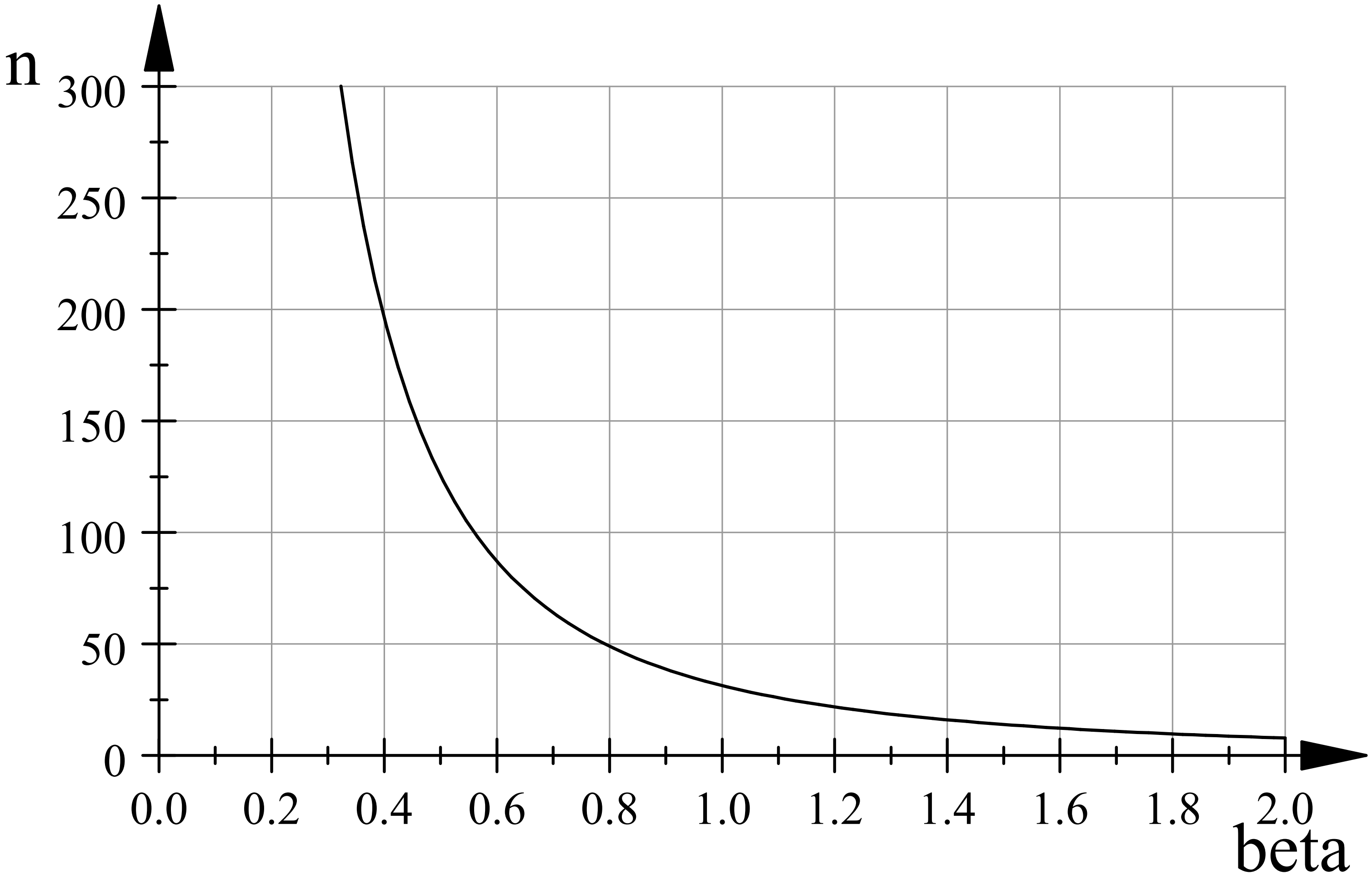
$$n > \left(t_{1-\kappa} + t_{\frac{\alpha}{2}}\right)^2 \frac{1}{p(1-p)} \frac{\sigma^2}{\beta^2}$$

- For $p = 0.5$, $\alpha = 0.05$ and $\kappa = 0.8$ and using the approximate values $t_{0.2} \approx 0.84$ and $t_{0.025} \approx 1.96$

$$n > 31.36 \left(\frac{\sigma}{\beta}\right)^2$$

- If n exceeds this value we will be able to detect a treatment difference of size β (i.e. obtain an effect that is statistically different from zero when the true effect size is β) with 80% probability

If we expect an effect size β & $\sigma = 1$, the we need $n = 31.36\frac{1}{\beta^2}$ observations



Python

Estimating Statistical Power in Python

- Statsmodels provides class TTestIndPower
- Import by from statsmodels.stats.power import TTestIndPower
- Perform power analysis for a t-test by:
  - analysis = TTestIndPower()
  - result = analysis.solve_power(effectSize, power=0.8, alpha=0.05) where you can substitute the respective effect size

- We can simulate different parameter constellations to plan a study
- Have to make assumptions/form beliefs prior to conducting experimental study
  - Variance of outcomes
  - Effect size: What magnitude do you expect?
  - In case of a fixed number of observations, given significance level and power: is the effect size sufficient?
  - Holding the number of observations, significance level and effect size constant, is power sufficiently high in order to conduct the experiment (cost-benefit analysis)?
- Information on these parameters can be derived from theory, prior empirical evidence, pilot studies