# Statistically Speaking: Power Analysis
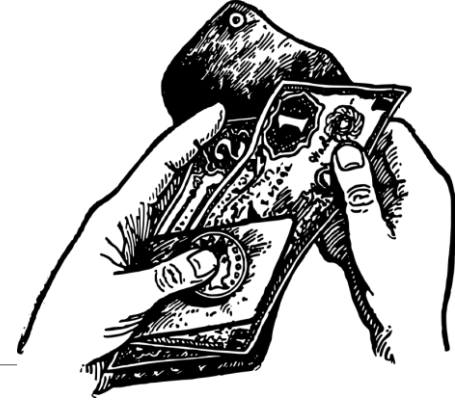
DAVID L. TABB, PH.D.

DECEMBER 7, 2017

# Overview

- Proposing a course of research is an essential skill for all work environments.

- Introducing the *four quantities* of statistical power calculations

- Meeting Jerzy Neyman, mathematical counterpoint to R.A. Fisher

# Contexts for research proposals



openclipart j4p4n

- You try to convince your advisor that your plan for experiments is well-considered.

- We request money from funding agencies for lab supplies, salaries, and bursaries.

- We seek to convince managers that our work will reward the company's investment.

- We explain to government that research in our agency contributes to the public good.

# Key questions asked of researchers

- Would testing your hypothesis through research make a difference in people's lives?

- Are you likely to yield the data you expect through this series of experiments?

- Would data produced by the experiment be capable of answering the question you pose?

- Is the experiment ethically sound?

- Do you know what you are doing?
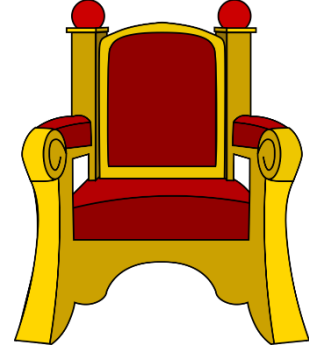
# Four legs to the throne of Power

- Sample Size: how many samples?

- Effect Size: minimum deviation from null hypothesis that you plan to detect
  - *Often difference in means over square of variance*

- Alpha or Significance Level: the probability of wrongly rejecting the null hypothesis
  - *Often 0.05*

- 1-Beta or Power: one minus the probability of accepting the null hypothesis when it is false
  - *Often 0.80*

# A common parallel

# Why do we perform three replicates?

- Two replicates have just one degree of freedom, giving a multiplier of 12.71 for 95% confidence interval.  Three replicates have two degrees of freedom, using 4.30 for CI.

- "Three replicates is the minimum number for Student's T-test." (*not true, actually*)

- "Three replicates are what we could afford."

# Effect size, or Cohen's *d*

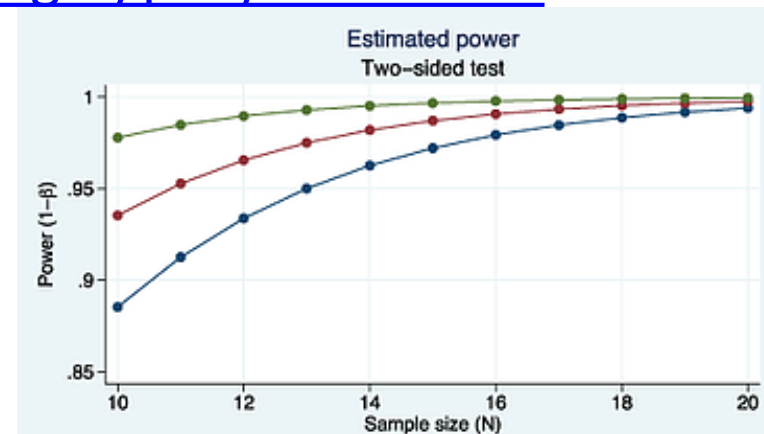- What is minimum biological difference we hope to find?

- $d = \dfrac{\overline{x_1} - \overline{x_2}}{s}$

- Difference between means divided by pooled standard deviation: 0.8 is "large," 0.5 is "moderate," and 0.2 is "small."

- *The smaller the difference you want to recognize, the lower the variability you can tolerate at a given sample size.*

# How we do power calculations

- Simulate data, and run the tests on those data to see how they would perform.

- Use a package like *pwr* in R.

- http://www.cyclismo.org/tutorial/R/power.html

- https://cran.r-project.org/web/packages/pwr/index.html

- Employ a web calculator.

# Jerzy Neyman (1894-1981)

- Born in what is now the Russian Federation, acquiring Ph.D. at University of Warsaw.

- A mathematician by inclination, he found statistics to be more marketable!

- Collaboration with Karl Pearson's son Egon led to Neyman-Pearson Theory of Hypothesis testing: two types of error, the idea of power, simple and composite hypotheses.

Sierpinski

# The attractions of ideas

"There were three summer houses on the estate, filled with young people, including girls whom I found beautiful and most attractive. However, the involvement with sets, measure and integration proved stronger than the charms of young ladies and most of my time was spent either in my room or on the adjacent balcony…"

# Takeaways

- Power is not just about estimating *n*: it expects researchers to be concrete about:
  - the magnitude of features to be detected,
  - the fraction of features they will recognize,
  - and their tolerance of false findings among their potential results.