Sta305-class7

February 1, 2016

Today's Class

- Calculating power via simulation
- Case study on power poses study: study replication and power
- Introduction to causal inference

Calculating Power by Simulation

- If the test statistic and distribution of the test statistic are known then the power of the test can be calculated via simulation.
- Consider a two-sample t-test with 30 subjects per group and the standard deviation of the clinical outcome is known to be 1.
- What is the power of the test $H_0 : \mu_1 - \mu_2 = 0$ versus $H_0 : \mu_1 - \mu_2 = 0.5$, at the 5% significance level?
- The power is the proportion of times that the test correctly rejects the null hypothesis in repeated sampling.

Calculating Power by Simulation

generates a random sample from $N(\mu, \sigma^2)$

We can simulate a single study using the `rnorm()` command. Let's assume that $n_1 = n_2 = 30, \mu_1 = 3.5, \mu_2 = 3, \sigma = 1, \alpha = 0.05$.

```
set.seed(2301)
t.test(rnorm(30,mean=3.5,sd=1),rnorm(30,mean=3,sd=1),var.equal = T)
```

n, $\mu=3.5$, $\sigma=1$, $\mu=3$ $\sigma=1$

```
    Two Sample t-test

data:  rnorm(30, mean = 3.5, sd = 1) and rnorm(30, mean = 3, sd = 1)
t = 2.1462, df = 58, p-value = 0.03605
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.03458122 0.99248595
sample estimates:
mean of x mean of y
 3.339362  2.825828
```

< 0.05

Should you reject H_0? Yes, reject

Calculating Power by Simulation

- Suppose that 10 studies are simulated.
- What proportion of these 10 studies will reject the null hypothesis at the 5% level?
- To investigate how many times the two-sample t-test will reject at the 5% level the `replicate()` command will be used to generate 10 studies and calculate the p-value in each study.
- It will still be assumed that $n_1 = n_2 = 30, \mu_1 = 3.5, \mu_2 = 3, \sigma = 1, \alpha = 0.05$.

```
set.seed(2301)
pvals <- replicate(10,t.test(rnorm(30,mean=3.5,sd=1),
                             rnorm(30,mean=3,sd=1),
                             var.equal = T)$p.value)
pvals # print out 10 p-values
```

# of studies

```
 [1] 0.03604893 0.15477655 0.01777959 0.40851999 0.34580930 0.11131007
 [7] 0.14788381 0.00317709 0.09452230 0.39173723
```

```
#power is the number of times the test rejects at the 5% level
sum(pvals<=0.05)/10
```

```
[1] 0.3
```

— power (replication not large. so possibly have large SE)

But, since we only simulated 10 studies the estimate of power will have a large standard error. So let's try simulating 10,000 studies so that we can obtain a more precise estimate of power.

```
set.seed(2301)
pvals <- replicate(10000,t.test(rnorm(30,mean=3.5,sd=1),
                                rnorm(30,mean=3,sd=1),
                                var.equal = T)$p.value)
sum(pvals<=0.05)/10000
```

power via simulation

```
[1] 0.4881
```

$H_0: \mu_1 - \mu_2 = 0$ v.s. $H_1: \mu_1 - \mu_2 = 3.5$

Calculating Power by Simulation

This is much closer to the theoretical power obtained from `power.t.test()`.

```
power.t.test(n = 30,delta = 0.5,sd = 1,sig.level = 0.05)
```

```
     Two-sample t test power calculation

              n = 30
          delta = 0.5
             sd = 1
      sig.level = 0.05
          power = 0.477841
    alternative = two.sided

NOTE: n is number in *each* group
```

Calculating Power by Simulation

- The built-in R functions `power.t.test()` and `power.prop.test()` don't have an option for calculating power where the there is unequal allocation of subjects between groups.
- One option is to simulate power for the scenarios that are of interest. Another option is to write your own function using the formula derived above.

Calculating Power by Simulation

- Suppose the standard treatment for a disease has a response rate of 20%, and an experimental treatment is anticipated to have a response rate of 28%.
- The researchers want both arms to have an equal number of subjects.
- A power calculation above revealed that the study will require 1094 patients for 80% power.
- What would happen to the power if the researchers put 1500 patients in the experimental arm and 500 patients in the control arm?

Calculating Power by Simulation

- The number of subjects in the experimental arm that have a positive response to treatment will be an observation from a $Bin(1500, 0.28)$.
- The number of subjects that have a positive response to the standard treatment will be an observation from a $Bin(500, 0.2)$.
- We can obtain simulated responses from these distributions using the `rbinom()` command in R.

```
set.seed(2301)
rbinom(1,1500,0.28)
```

```
[1] 403
```

403 subjects had positive response in experiment arm

```
rbinom(1,500,0.20)
```

```
[1] 89
```

89 subjects has positive response in control arm

Calculating Power by Simulation

- The p-value for this simulated study can be obtained using `prop.test()`.

```
set.seed(2301)
prop.test(x=c(rbinom(1,1500,0.28),rbinom(1,500,0.20)),
          n=c(1500,500),correct = F)
```

```
    2-sample test for equality of proportions without continuity
    correction

data:  c(rbinom(1, 1500, 0.28), rbinom(1, 500, 0.2)) out of c(1500, 500
X-squared = 16.62, df = 1, p-value = 4.568e-05
alternative hypothesis: two.sided
95 percent confidence interval:
 0.05032654 0.13100680
sample estimates:
   prop 1    prop 2
0.2686667 0.1780000
```

This is an example of simulating one study

- A power simulation repeats this process a large number of times.
- In the example below we simulate 10,000 hypothetical studies to calculate power.

```
set.seed(2301)
pvals <- replicate(10000,
prop.test(x=c(rbinom(n = 1,size = 1500,prob = 0.25),
              rbinom(n=1,size=500,prob=0.20)),
              n=c(1500,500),correct=F)$p.value)
sum(pvals<=0.05)/10000
```

```
[1] 0.6231
```

power decreased from 83% with equal allocation

If the researchers decide to have a 3:1 allocation ratio of patients in the treatment to control arm then the power will be ~~62%~~ ? to → with 3:1

Can power poses significantly change outcomes in your life?

Professor Amy Cuddy 

Research Report

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
21(10) 1363–1368

DOI: 10.1177/0956797610383437
http://pss.sagepub.com
SAGE

Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance

Dana R. Carney[1], Amy J.C. Cuddy[2], and Andy J. Yap[1]
[1]Columbia University and [2]Harvard University

Abstract

Humans and other animals express power through open, expansive postures, and they express powerlessness through closed, contractive postures. But can these postures actually cause power? The results of this study confirmed our prediction that posing in high-power nonverbal displays (as opposed to low-power nonverbal displays) would cause neuroendocrine and behavioral changes for both male and female participants: High-power posers experienced elevations in testosterone, decreases in cortisol, and increased feelings of power and tolerance for risk; low-power posers exhibited the opposite pattern. In short, posing in displays of power caused advantaged and adaptive psychological, physiological, and behavioral changes, and these findings suggest that embodiment extends beyond mere thinking and feeling, to physiology and subsequent behavioral choices. That a person can, by assuming two simple 1-min poses, embody power and instantly become more powerful has real-world, actionable implications.

Keywords

cortisol, embodiment, hormones, neuroendocrinology, nonverbal behavior, power, risk taking, testosterone

Received 1/20/10; Revision accepted 4/8/10

The proud peacock fans his tail feathers in pursuit of a mate. By galloping sideways, the cat manipulates an intruder's perception of her size. The chimpanzee, asserting his hierarchical rank, holds his breath until his chest bulges. The executive in the boardroom crests the table with his feet, fingers interlaced behind his neck, elbows pointing outward. Humans and other animals display power and dominance through expansive nonverbal displays, and these power poses are deeply intertwined with the evolutionary selection of what is "alpha" (Darwin, 1872/2009; de Waal, 1998).

But is power embodied? What happens when displays of power are posed? Can posed displays cause a person to feel more powerful? Do people's mental and physiological systems prepare them to be more powerful? The goal of our research was to test whether high-power poses (as opposed to low-power poses) actually produce power. To perform this test, we looked at the effects of high-power and low-power poses on some fundamental features of having power: feelings of power, elevation of the dominance hormone testosterone, lowering of the stress hormone cortisol, and an increased tolerance for risk.

Power determines greater access to resources (de Waal, 1998; Keltner, Gruenfeld, & Anderson, 2003); higher levels of agency and control over a person's own body, mind, and positive feelings (Keltner et al., 2003); and enhanced cognitive function (Smith, Jostmann, Galinsky, & van Dijk, 2008). Powerful individuals (compared with powerless individuals) demonstrate greater willingness to engage in action (Galinsky, Gruenfeld, & Magee, 2003; Keltner et al., 2003) and often show increased risk-taking behavior[1] (e.g., Anderson & Galinsky, 2006).

The neuroendocrine profiles of the powerful differentiate them from the powerless, on two key hormones—testosterone and cortisol. In humans and other animals, testosterone levels both reflect and reinforce dispositional and situational status and dominance; internal and external cues cause testosterone to rise, increasing dominant behaviors, and these behaviors can elevate testosterone even further (Archer, 2006; Mazur &

Corresponding Authors:
Dana R. Carney, Columbia University, Graduate School of Business, 717 Uris Hall, 3022 Broadway, New York, NY 10027-6902
E-mail: dcarney@columbia.edu
Amy J.C. Cuddy, Harvard Business School, Baker Library 449, Boston, MA 02163
E-mail: acuddy@hbs.edu

Fig. 1. The two high-power poses used in the study. Participants in the high-power-pose condition were posed in expansive positions with open limbs.

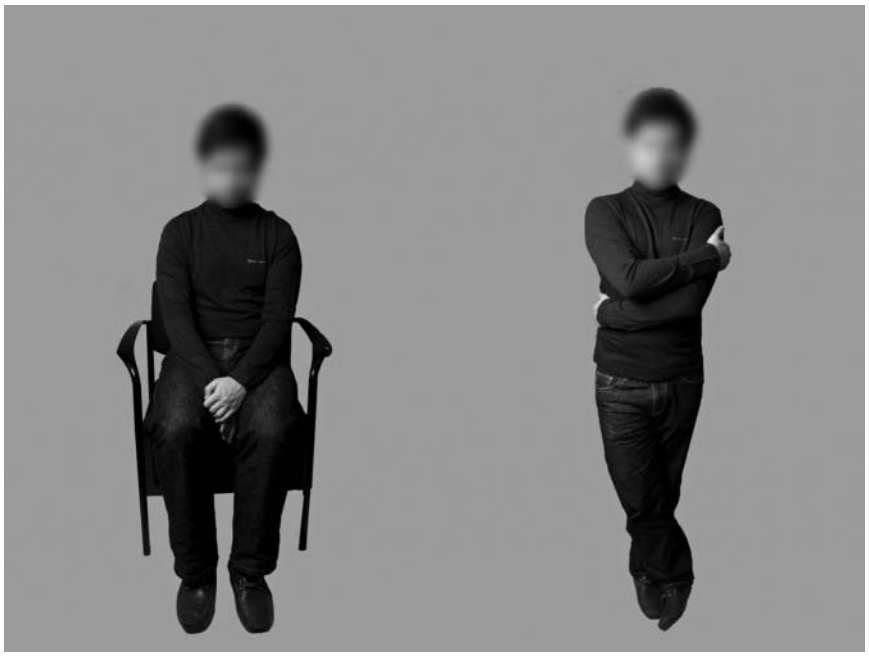


Fig. 2. The two low-power poses used in the study. Participants in the low-power-pose condition were posed in contractive positions with closed limbs.

To configure the test participants into the poses, the experimenter placed an electrocardiography lead on the back of each participant's calf and underbelly of the left arm and explained, "To test accuracy of physiological responses as a function of sensor placement relative to your heart, you are being put into a certain physical position." The experimenter then manually configured participants' bodies by lightly touching their arms and legs. As needed, the experimenter provided verbal

Can power poses significantly change outcomes in your life?

Cuddy's study methods:

- Randomly assigned 42 participants to the high-power pose or the low-power-pose condition.
- Participants believed that the study was about the science of physiological recordings and was focused on how placement of electrocardiography electrodes above and below the heart could influence data collection.
- Participants' bodies were posed by an experimenter into high-power or low-power poses. Each participant held two poses for 1 min each.
- Participants' risk taking was measured with a gambling task; feelings of power were measured with self-reports.
- Saliva samples, which were used to test cortisol and testosterone levels, were taken before and approximately 17 min after the power-pose manipulation.

(Carney, Cuddy, Yap, 2010)

Can power poses significantly change outcomes in your life?

Cuddy's study results:

> *As hypothesized, high-power poses caused an increase in testosterone compared with low-power poses, which caused a decrease in testosterone, $F(1, 39) = 4.29$, $p < .05$; $r = .34$. Also as hypothesized, high-power poses caused a decrease in cortisol compared with low-power poses, which caused an increase in cortisol, $F(1, 38) = 7.45$, $p < .02$; $r = .43$*

Can power poses significantly change outcomes in your life?

- The study was replicated by Ranehill et al. (2015)
- An initial power analysis based on the effect sizes in Carney et al. (power = 0.8, $\alpha = .05$) indicated that a sample size of 100 participants would be suitable.

```
library(pwr)
pwr.t.test(d=0.6,power = 0.8)
```

Effect size from cuddy study $\mu_1 - \mu_2 = ?$

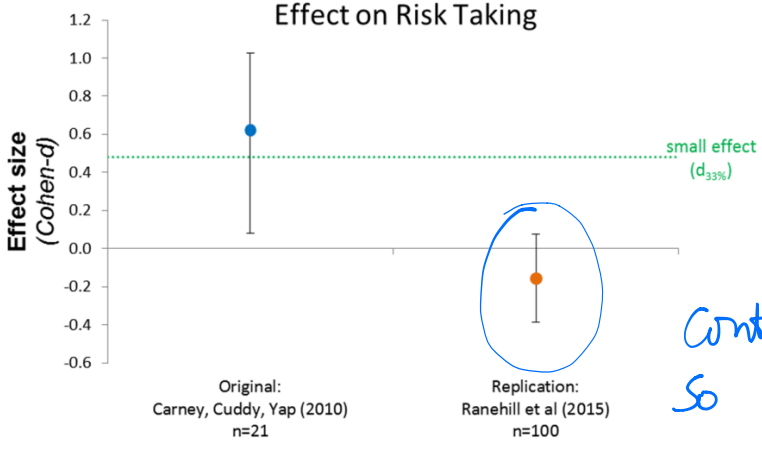
```
     Two-sample t test power calculation

              n = 44.58579
              d = 0.6
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

CS

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



Contains 0
So no effect!

Can power poses significantly change outcomes in your life?

- Ranehill et al. study used a sample of 200 participants to increase reliability.
- This study found none of the significant differences found in Cuddy's study.
- The replication study obtained very precise estimates of the effects.
- What happened?

Slate article by Gelman

Can power poses significantly change outcomes in your life?

- Sampling theory predicts that the variation between samples is proportional to $\frac{1}{\sqrt{n}}$.
- In small samples we can expect variability.
- Many researchers often expect that these samples will be more similar than sampling theory predicts.

Kahneman & Tversky 0.85 chance find significant difference

Suppose that you have run an experiment on 20 subjects, and have obtained a significant result from a two-sided z-test ($H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$) which confirms your theory ($z = 2.23$, $p < 0.05$, two-tailed). The researcher is planning to run the same experiment on an additional 10 subjects. What is the probability that the results will be significant at the 5% level by a one-tailed test ($H_1 : \mu > 0$), seperately for this group?

confirmatory study

but smaller sample

Correct answer is 0.47

END

Expt #1

$H_0: \mu = 0$ vs $H_1: \mu \neq 0$

$Z = 2.23 = \frac{\bar{x} - \mu}{\sigma/\sqrt{20}}$

$\bar{x} = 2.23\sigma/\sqrt{20}$

Expt #2

$H_0: \mu = 0$ vs $H_1: \mu > 0$

but what is power to detect mean for

Expt #1 $\mu = 2.23\sigma/\sqrt{20}$

power (μ) = 0.47 details to the class

- Suppose Bob, at a particular point in time, is contemplating whether or not to take an aspirin for a headache.
- There are two treatment levels, taking an aspirin, and not taking an aspirin.
- If Bob takes the aspirin, his headache may be gone, or it may remain, say, an hour later; we denote this outcome, which can be either "Headache" or "No Headache," by $Y(\text{Aspirin})$.
- Similarly, if Bob does not take the aspirin, his headache may remain an hour later, or it may not; we denote this potential outcome by $Y(\text{No Aspirin})$, which also can be either "Headache," or "No Headache."
- There are therefore two potential outcomes, $Y(\text{Aspirin})$ and $Y(\text{No Aspirin})$, one for each level of the treatment. The causal effect of the treatment involves the comparison of these two potential outcomes.

Because in this example each potential outcome can take on only two values, the unit- level causal effect – the comparison of these two outcomes for the same unit – involves one of four (two by two) possibilities:

1. Headache gone only with aspirin: Y(Aspirin) = No Headache, Y(No Aspirin) = Headache
2. No effect of aspirin, with a headache in both cases: Y(Aspirin) = Headache, Y(No Aspirin) = Headache
3. No effect of aspirin, with the headache gone in both cases: Y(Aspirin) = No Headache, Y(No Aspirin) = No Headache
4. Headache gone only without aspirin: Y(Aspirin) = Headache, Y(No Aspirin) = No Headache

There are two important aspects of this definition of a causal effect.

1. The definition of the causal effect depends on the potential outcomes, but it does not depend on which outcome is actually observed.
2. The causal effect is the comparison of potential outcomes, for the same unit, at the same moment in time post-treatment.

- The causal effect is not defined in terms of comparisons of outcomes at different times, as in a before-and-after comparison of my headache before and after deciding to take or not to take the aspirin.

The fundemental problem of causal inference

"The fundamental problem of causal inference" (Holland, 1986, p. 947) is the problem that at most one of the potential outcomes can be realized and thus observed.

- If the action you take is Aspirin, you observe $Y(\text{Aspirin})$ and will never know the value of $Y(\text{No Aspirin})$ because you cannot go back in time.
- Similarly, if your action is No Aspirin, you observe $Y(\text{No Aspirin})$ but cannot know the value of $Y(\text{Aspirin})$.
- In general, therefore, even though the unit-level causal effect (the comparison of the two potential outcomes) may be well defined, by definition we cannot learn its value from just the single realized potential outcome.

The outcomes that would be observed under control and treatment conditions are often called **counterfactuals** or **potential outcomes**.

- If Bob took asprin for his headache then he would be assigned to the treatment condition so $T_i = 1$.
- Then $Y(\text{Aspirin})$ is observed and $Y(\text{No Aspirin})$ is the unobserved counterfactual outcome—it represents what would have happened to Bob if he had no taken aspirin.
- Conversely, if Bob had not taken aspirin then $Y(\text{No Aspirin})$ is observed and $Y(\text{Aspirin})$ is counterfactual.
- In either case, a simple treatment effect for Bob can be defined as

$$\text{treatment effect for Bob } = Y(\text{Aspirin}) - Y(\text{No Aspirin}).$$

The assignment mechanism

- Suppose that a doctor prescribes surgery (labeled 1) or drug (labeled 0) for a certain condition.
- The doctor knows enough about the potential outcomes of the patients so assigns each patient the treatment that is more beneficial to that patient.

unit	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
patient #1	1	7	6
patient #2	6	5	-1
patient #3	1	5	4
patient #4	8	7	-1
Average	4	6	2

Y is years of post-treatment survival.

The assignment mechanism

- Patients 1 and 3 will receive surgery and patients 2 and 4 will receive drug treatment.
- The observed treatments and outcomes are in this table.

unit	T_i	Y_i^{obs}
patient #1	1	7
patient #2	0	6
patient #3	1	5
patient #4	0	8
Average Drug		**7**
Average Surg		**6**

- This shows that we can reach invalid conclusions if we look at the observed values of potential outcomes without considering how the treatments were assigned.