BMEG 802 – Advanced Biomedical Experimental Design and Analysis

Effect Size and Power

Joshua G. A. Cashaback, PhD

Recap

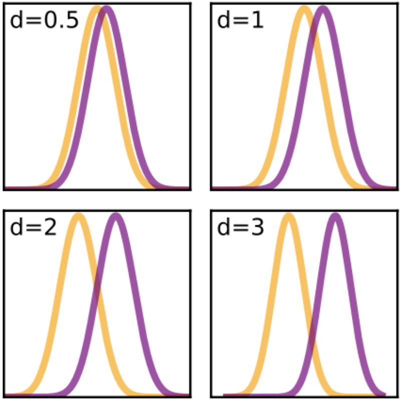
- Null Hypothesis testing
  - We can only falsify a theory!
    - p-value = likelihood of the observed data given the null hypothesis is true
  - single and two-sided tests
  - parametric vs. nonparametric
  - correcting for multiple comparisons

Learning Objectives

- Effect Size
  - Parametric
    - Cohen's D (1 sample, 2 sample, paired)
  - Nonparametric
    - Common Language Effect Size
- Power
  - parametric
  - sampling techniques

Effect Size

Effect size: Simple way to quantify the difference between two means / groups, by emphasizing the size of the difference rather than confounding with the sample size (like p-values).



Cohen's D: 1 Sample

$$d = \frac{\bar{X} - \mu_0}{s}$$

d = effect size

$\bar{X}$ = sample mean

μ_0 = theoretical mean against which the sample mean is compared

s = sample standard deviation

Cohen's D: 1 Sample

From last lecture: A cookie company claims that there are 15 chocolate chips per cookie, but you aren't convinced. You take 10 cookies and count the number of chocolate chips in each cookie. Here is what the data looks like: [13,14,15,17,18,19,21,20,19,20]. Are the number of chocolate chips significantly different from 15 (reminder: yes, $p = 0.015$)? Report the effect size.

```r
data1 = c(13,14,15,17,18,19,21,20,19,20)
X = mean(data1)
s = sd(data1)
mu = 15
d = (X - mu) / s
abs(d) # report positive value
```

```
## [1] 0.9431191
```

Cohen’s D: 1 Sample Cont’d

Using a built in function

install.packages(“effsize”)

```r
library(effsize)
cohen.d(data1, NA, mu = 15)
```

```
##
## Cohen's d (single sample)
##
## d estimate: 0.9431191 (large)
## Reference mu: 15
## 95 percent confidence interval:
##      lower      upper
## -0.5650353  2.4512735
```

Cohen's D: 2 Sample

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}}$$

d = effect size

$\bar{X}$ = sample mean

μ = theoretical mean against which the mean of our sample is compared

s = sample standard deviation

Cohen's D: 2 sample

From last lecture: A math test was given to 300 17 year old students in 1978 and again to another 350 17 year old students in 1992.

- Group 1: $X_1 = 300.4, s_1 = 34.9, n = 300$
- Group 2: $X_2 = 306.7, s_2 = 30.1, n = 350$

Report the effect size.

```
group1 <- rnorm(300, mean = 300.4, sd=34.9)
group2 <- rnorm(350, mean = 306.7, sd=30.1)
d2 = (mean(group1) - mean(group2)) / (sqrt(((length(group1) - 1) * sd(grou
abs(d2)
```

```
## [1] 0.2053413
```

Cohen's D: 2 sample

Using a built in function

```
cohen.d(group1,group2, var.equal = False) # var.equal = False performs a W
```

```
##
## Cohen's d
##
## d estimate: -0.2053413 (small)
## 95 percent confidence interval:
##       lower       upper
## -0.36024322 -0.05043937
```

Cohen's D: Paired

$$d_{rm} = \frac{\bar{X}_D - \mu_0}{\frac{\sqrt{s_1^2 + s_2^2 - 2 \cdot r \cdot s_1 \cdot s_2}}{\sqrt{2(1-r)}}}$$

d = effect size

$\bar{X}$ = sample mean of the paired differences

μ = theoretical mean that the paired mean differences are compared against

s_i = sample standard deviation at a particular time (i)

r = correlation coefficient (we'll talk more about this next lecture)

Other versions:

https://pure.tue.nl/ws/portalfiles/portal/3835042/1236489301722996.pdf

Cohen's D: Paired Samples

From last class: A manufacturer claims it has developed an additive that increases gas mileage. But you are not sure whether the additive will increase or decrease performance. They recruit 10 drivers. Each driver drives a car on a well-conditioned track. They record the gas mileage without any additive, then with additive. Report the effect size.

```
data2a = c(22,25,17,24,16,29,20,23,19,20)
data2b = c(18,21,16,22,19,24,17,21,23,18)
data2diff = data2a - data2b
d2 = (mean(data2diff) - 0) /
  (sqrt(sd(data2a)^2 + sd(data2b)^2
        - 2 * cor(data2a,data2b) * sd(data2a) * sd(data2b))/
     sqrt(2*(1-cor(data2a,data2b))))
d2
```

```
## [1] 0.4475411
```

Cohen's D: Paired Samples

```r
cohen.d(data2a,data2b, paired = TRUE)
```

```
##
## Cohen's d
##
## d estimate: 0.4475411 (small)
## 95 percent confidence interval:
##      lower      upper
## -0.1277464  1.0228286
```

Common Language Effect Size

- nonparamtric way to calculate effect size
- intuitive
  - the probability that a score sampled at random from one distribution will be greater than a score sampled from some other distribution
  - e.g., the probability that a male will be taller than a female is 0.92. In other words, the male will be taller than the female in 92 out of 100 blind dates among young adults.
- brute force

Common Language Effect Size: 2 Sample, Procedure

- Compare every possible value in group A to every other possible value in group B.
  - Add 1 each time the difference goes in the expected direction.
  - Add 0.5 if there is a tie
  - Add 0.0 if the difference goes in the unexpected direction
  - subtract 0.5, absolute, add 0.5, multiply by 100 (values should be between 50% and 100%)
- Lets try this out with a simple example first

Common Language Effect Size - 2 sample

```r
a = c(1,3,5)
b = c(2,4,6)
c <- array(dim=c(length(a),length(b)))
for (i in 1:length(a)) {
  for (j in 1:length(b)){
    if (a[i] > b[j]) {
    c[i,j] = 1
    } else if (a[i] == b[j]) {
    c[i,j] = 0.5
    } else {
    c[i,j] = 0.0
}}}
CLES = (abs(sum(c) / (length(a) * length(b)) - 0.5) + 0.5) * 100
CLES
```

```
## [1] 66.66667
```

$$\hat{\theta} = 66.7\%$$

Common Language Effect Size - 2 sample

Now lets do this on the 2 Sample example we did earlier

```
a = group1
b = group2
c <- array(dim=c(length(a),length(b)))
for (i in 1:length(a)) {
  for (j in 1:length(b)){
    if (a[i] > b[j]) {
    c[i,j] = 1
    } else if (a[i] == b[j]) {
    c[i,j] = 0.5
    } else {
    c[i,j] = 0.0
}}}
CLES = (abs(sum(c) / (length(a) * length(b)) - 0.5) + 0.5) * 100
CLES
```

```
## [1] 56.07524
```

Common Language Effect Size - 1 Sample

We can do this with a 1 Sample test (same example used above)

```
a = data1 # data
b = array(15,dim=c(length(a)))  # mean value you are comparing to (i.e., mu_0 = 15)
c <- array(dim=c(length(a),length(b)))
for (i in 1:length(a)) {
  for (j in 1:length(b)){
    if (a[i] > b[j]) {
    c[i,j] = 1
    } else if (a[i] == b[j]) {
    c[i,j] = 0.5
    } else {
    c[i,j] = 0.0
}}}
CLES = (abs(sum(c) / (length(a) * length(b)) - 0.5) + 0.5) * 100
CLES
```

```
## [1] 75
```

Common Language Effect Size - Paired

As well as paired test (using the paired differences example above)

```
a1 = data2a # e.g., pre intervention
a2 = data2b  # e.g., post intervention
a = a2 - a1
b = array(0,dim=c(length(a)))  # mean value you are comparing to (same as mu_0 = 0)
c <- array(dim=c(length(a),length(b)))
for (i in 1:length(a)) {
  for (j in 1:length(b)){
    if (a[i] > b[j]) {
    c[i,j] = 1
    } else if (a[i] == b[j]) {
    c[i,j] = 0.5
    } else {
    c[i,j] = 0.0
}}}
CLES = (abs(sum(c) / (length(a) * length(b)) - 0.5) + 0.5) * 100
CLES
```

```
## [1] 80
```

Effect Size Interpretation

report along p-value

- (p $=$ 0.01, d $=$ 0.5) or (p $=$ 0.01, $\hat{\theta} = 66.2\%$) Can state whether the effect is small ($d \approx 0.2$), medium ($d \approx 0.5$), or large ($d \approx 0.8$).
- For CLES ($\hat{\theta}$): small ($\hat{\theta} \approx 56\%$), medium ($\hat{\theta} \approx 64\%$), or large ($\hat{\theta} \approx 71\%$).
  - All of these definitions are somewhat arbitrary

Other Effect Sizes

- Other effect sizes: e.g., Glass's delta, Hedges' g
- Regression (r-value)
- ANOVA
  - eta squared
  - omega squared

STATISTICAL POWER

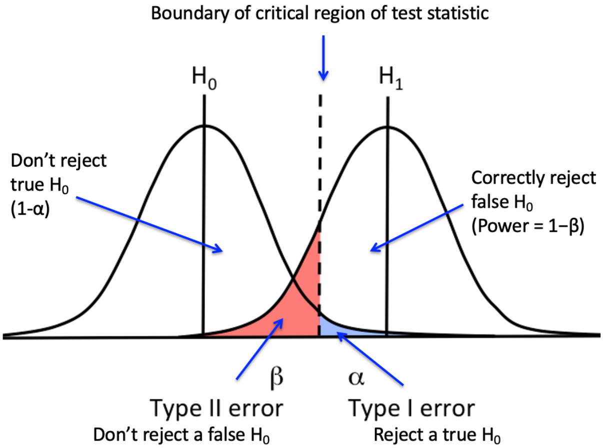
Power

Power: probability of a hypothesis test of finding an effect if there is an effect to be found.

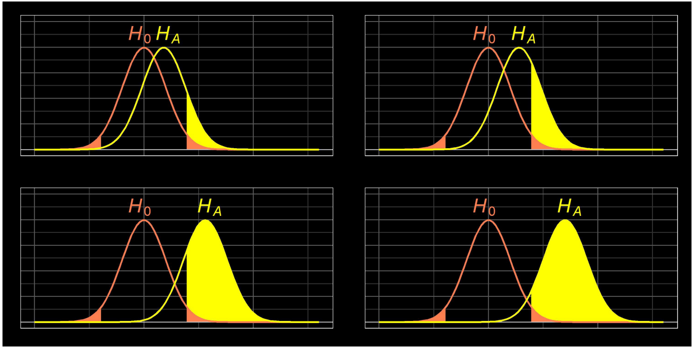
		Null Hypothesis - Reality	
		True	False
Research Action	Don't reject H_0	No error $1-\alpha$	Type II Error β
	Reject H_0	Type I Error α	No error $1-\beta$

Power

Research Action		Null Hypothesis - Reality: True	Null Hypothesis - Reality: False
	Don't reject H_0	No error $1-\alpha$	Type II Error β
	Reject H_0	Type I Error α	No error $1-\beta$



Power



Power depends on:

1. sample size
2. effect size
3. statistical significance criteria

Next Week

- linear regression (next week)
  - Pearson's r and Spearman