BIOS 6611 Homework 4 Answer Key
Due Monday, October 1, 2018 by noon to Canvas Assignment Basket

1. Summarize briefly (one paragraph, half page max) Sections 1 and 2 of the paper by, Hoenig and Heisey, "*The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis*", which can be found on Canvas in the Paper Repository.

   This paper argues against proponents of post-experiment power calculations to explain failure to reject the null hypothesis. The first part was calculation of observed power when one fails to reject the null hypothesis. If the observed power is high enough, then one can conclude that the null hypothesis is true. Authors argue that this is a fallacy in that the p-value solely determines the observed power (after showing calculations on the article) thus the observed power adds no value to the result. The other part is the detectable effect size and biologically significant effect size. This is what the authors describe as the intriguing application of post experiment power calculation where one tries to find the hypothetical true difference given a particular power for example 0.9. After failing to reject the null hypothesis, one can calculate the detectable effect size needed to have a power of 0.9 based on the observed variability. Detectable effect size is assumed to be the upper bound limit because nature is unlikely to approach this upper limit due to the non-significant result. The closer the detectable size is to 0, the stronger the evidence for the null. Biologically significant effect size is another form or version of detectable effect size where one computes the power of the effect size that is thought to be biologically significant. This approach and its variant suffer from power approach paradox just like the observed power. If two experiments with nonrejected nulls were compared and the first experiment was found to have a higher computed power, it is incorrect to conclude that the first experiment provides more evidence in favor of the null. This is in direct contrast to the traditional interpretation of the results using p-values. Power should be based on pre-specified detectable difference and not based on what is seen in the study.

Commented [AK1]: 25 points, responses can vary and highlight different things than included in this sample answer, but should reflect the message of the paper

2. A clinical trial is planned to examine the bioavailability of twice per day beta carotene supplementation. Measurements are to be taken at baseline and 12 weeks later. Below, write out results by hand (i.e., show equations and work) for known s.d. and use R for unknown s.d. For R, you can refer to examples and code in Lecture 8. **Turn in your hand calculations/derivations, R code, and relevant output.**

   a. For the null hypothesis that the mean change in beta carotene is 0 mcg/dL and the alternative hypothesis that the mean change in beta carotene is 100 mcg/dL, σ_{change} = 75 mcg/dL, n=5, what is the *power* to detect that difference, using α = 0.05 (two-sided)? Assume two cases: known s.d. and unknown s.d.

- Two-sided test
- α = 0.05
- H_0: μ = 0 mcg/dL
- H_1: μ = 100 mcg/dL
- σ_{change} = 75 mcg/dL
- n=5

Known s.d.

$$1-\beta = \Phi\left[\frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} - Z_{1-\alpha/2}\right] = \Phi\left[\frac{|0-100|}{75/\sqrt{5}} - Z_{1-0.05/2}\right]$$
$$= \Phi\left[\frac{100}{33.54} - 1.96\right] = \Phi[1.0214] = 0.8465$$

The power to detect the difference of 100 mcg/dL greater than 0 mcg/dL is 84.65% when s.d. is known

Unknown s.d.

```
power.t.test(n=5,delta=100,sd=75,sig.level=0.05,
type="one.sample",alternative="two.sided")

     One-sample t test power calculation

              n = 5
          delta = 100
             sd = 75
      sig.level = 0.05
          power = 0.6141832
    alternative = two.sided
```

The power to detect the difference of 100 mcg/dL greater than 0 mcg/dL is 61.42% when s.d. is unknown

b. *How many subjects* would be needed in (a) above if 90% power were desired? Assume two cases: known s.d. and unknown s.d.

Known s.d.

$$n = \frac{\sigma^2\left(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}\right)^2}{(\mu_0 - \mu_1)^2} = \frac{75^2\left(Z_{1-0.1} + Z_{1-\frac{0.05}{2}}\right)^2}{(0-100)^2} = \frac{75^2(1.282+1.96)^2}{(100)^2} = 5.91219$$

A sample size of 6 would be needed to have 90% power with known s.d.

Unknown s.d.

```
power.t.test(delta=100,sd=75,sig.level=0.05,power=0.90,
type="one.sample",alternative="two.sided")

     One-sample t test power calculation

              n = 8.072323
          delta = 100
             sd = 75
      sig.level = 0.05
          power = 0.9
    alternative = two.sided
```

A sample size of 9 would be needed to have 90% power with unknown s.d.

c. What is the *smallest mean change* in beta carotene that could be detected as significantly different from a change of 0 mcg/dL, if σ_{change} = 75 mcg/dL, n = 5, and power is 90%? What is the *smallest mean change* for 80% power? Assume two cases: known s.d. and unknown s.d.

Commented [AK2]: 40 points (10 for each part)

Known s.d. and Power 90%

$$|\mu_0 - \mu_1| = \left(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}\right)\frac{\sigma}{\sqrt{n}} = (Z_{0.9} + Z_{0.975})\frac{75}{\sqrt{5}} = (1.282 + 1.96)33.541$$
$$|\mu_0 - \mu_1| = 108.74 \text{ mcg/dL}$$

With 90% power and a sample size of 5, we are able to detect a difference of at least 108.74 mcg/dL, given that it exists.

Known s.d. and Power 80%

$$|\mu_0 - \mu_1| = \left(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}\right)\frac{\sigma}{\sqrt{n}} = \ (Z_{0.8} + Z_{0.975})\frac{75}{\sqrt{5}} = \ \ (0.842 + 1.96)33.541$$
$$|\mu_0 - \mu_1| = 93.98 \text{ mcg/dL}$$

With 80% power and a sample size of 5, we are able to detect a difference of at least 93.98 mcg/dL, given that it exists.

Unknown s.d. and Power 90%

```
power.t.test(n=5,sd=75,sig.level=0.05,power=0.90,
type="one.sample",alternative="two.sided")

     One-sample t test power calculation

              n = 5
          delta = 147.4417
             sd = 75
      sig.level = 0.05
          power = 0.9
    alternative = two.sided
```

With a sample size of 5 and unknown s.d., there is 90% power to detect a difference of 147.44 mcg/dL, given that it exists.

Unknown s.d. and Power 80%

```
power.t.test(n=5,sd=75,sig.level=0.05,power=0.80,type="one.sample",
alternative="two.sided")

     One-sample t test power calculation

              n = 5
          delta = 126.1498
             sd = 75
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

With a sample size of 5 and unknown s.d., there is 80% power to detect a difference of 126.15 mcg/dL, given that it exists.

3. In some cases the test statistic you need to apply to a set of data is not of a standard form and so is not included as an option in R (or other sample size and power analysis software), or perhaps an exact formula does not exist for doing the power calculation. In these cases, simulation in R (or other software) can be used to estimate the power, as described below. As an example, consider using simulation in question 2 above. There, formulas exist so simulation would not be necessary, but simulation allows you to check your results. Assume normally distributed data for the change in beta carotene with σ_{change} = 75 mcg/dL, n=5, and significance level = 0.05, with a one-sample two-sided t-test. Answer questions i-iii based on (a) and (b):

   a. Use simulation to show that the one-sample t-test has the correct significance level, that is that under the null hypothesis the test rejects about 5% of the time. Use 10,000 iterations and the seed value 2345.

   b. Use simulation to obtain a power estimate for the situation in 1a for the alternative hypothesis that the mean change is 100 mcg/dL. Use 10,000 iterations and the seed value 1796. Which of your answers in question 2 matches most closely with your answer here? Explain briefly why.

      i. Using a loop in R, carry out (a) and (b) above. Summarize your results in a brief paragraph.

Commented [AK3]: 35 points

Part A

```
set.seed(2345)
n <- 5
mean <- 0
sd <- 75
numTrials <-10000
alpha <- 0.05

count<- 0
for(i in 1:numTrials){
  y <- rnorm(n,mean,sd)
  t <- t.test(y, mu = 0, alternative = "two.sided")
  if(t$p.value < 0.05) count <- count + 1 else count <- count
}

power <- count/numTrials
power
[1] 0.0498
```

Part B

```
set.seed(1796)
n <- 5
mean <- 100
sd <- 75
numTrials <-10000
alpha <- 0.05

count <- 0
for(i in 1:numTrials){
  y <- rnorm(n,mean,sd)
  t <- t.test(y, mu = 0, alternative = "two.sided")
  if(t$p.value < 0.05) count <- count + 1 else count <- count
}

power <- count/numTrials
power
[1] 0.613
```

The significance level for (a) is approximately 5% from our simulation, demonstrating th at we achieve the desired significance level. For (b), our estimated power of 61.3% matc hes most closely with question 2a when assuming unknown variance. This is because we are assuming the same scenario (i.e., the difference of 100 between H_0 and H_1, sd=75, n =5, etc.), so we would expect the results to be in agreement.

ii. EXTRA CREDIT: Using the same scenario and parameters as in (a) and (b), simulate power under the null and alternative hypotheses by writing a function in R. Provide your function and the results for applying your function for (a) and (b).

Commented [KAM4]: 3 extra credit points

```
compute_power = function(n, mean, sigma, numTrials, alpha){
  sample <- matrix(rnorm(n*numTrials, mean, sigma), ncol=numTrials)
  xbar <- apply(sample, 2, mean)
  variance <- apply(sample, 2, var)
  df.num = n-1
  test.stat <- (xbar-0)/sqrt(variance/n)
  return (mean(abs(test.stat) >= qt((1-(alpha/2)), df.num)))
}
```

Part A

```
> set.seed(2345)
> compute_power(n=5,mean=0,sigma=75,numTrials=10000,alpha=0.05)
[1] 0.0498
```

Part B

```
> set.seed(1796)
> compute_power(n=5,mean=100,sigma=75,numTrials=10000,alpha=0.05)
[1] 0.613
```

iii. It is not so easy to use simulation to estimate sample size or detectable difference directly. Explain in a sentence or two how you could use trial and error to estimate sample size as in 2b above. Do not do any further simulations or calculations (though they are not difficult to do).

Vary the sample size or detectable difference and see how that affects power, keep attempting different values until you get the desired power level.