Methods Homework 4

Tim Vigers

September 30, 2018

# 1. The Abuse of Power Summary

Hoenig and Heisey argue against a surprisingly large literature which advocates the use of post-hoc power testing, to assist with interpreting results that are not statistically significant. This approach to data analysis is generally used to reduce the chance of failing to reject the null hyopthesis of no change, when an effect actually does exist. While this approach usually comes from a genuine desire to improve public welfare, there are a multitude of flaws in its statistical reasoning. First, “advocates of observed power argue that there is evidence for the null hypothesis being true if statistical significance was not achieved despite the computed power being high at the observed effect size.” (Hoenig & Heisey, 2001) However, observed power is always a function of the p value, so it can never add to the interpretation of the results. The authors also provide several counter-examples to reinforce this point.

A second approach to using post-hoc power analyses is calculating a theoretical difference which would result in a desired power, based on the observed data. As a counter-example, imagine two hypothetical experiments with equal observed effect sizes and sample size, but with one experiment closer to significance. This approach would suggest that for a given level of power, the detectable effect size for the significant experiment would be smaller than the second. Advocates would conclude that the more significant experiment offers more support for the null hypothesis, which is “in direct contradiction to the standard interpretation of the experimental results (p values).” (Hoenig & Heisey, 2001)

# 2.

## a. Power Calculations

### i. Known SD (by hand):

### ii. Unknown SD (with R):

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

##   
## One-sample t test power calculation   
##   
## n = 5  
## delta = 100  
## sd = 75  
## sig.level = 0.05  
## power = 0.6141832  
## alternative = two.sided

power\_test$power

## [1] 0.6141832

## b. n Calculations

### i. Known SD (by hand):

### ii. Unknown SD (with R):

n\_test <- power.t.test(power=0.9, sd = 75, sig.level = 0.05, delta=100, type = "one.sample", alternative = "two.sided")  
n\_test

##   
## One-sample t test power calculation   
##   
## n = 8.072323  
## delta = 100  
## sd = 75  
## sig.level = 0.05  
## power = 0.9  
## alternative = two.sided

n\_test$n

## [1] 8.072323

## c. Smallest detectable change

### i. Known SD (by hand):

### ii. Unknown SD (with R):

# 90% power  
delta\_test\_90 <- power.t.test(n = 5,power = 0.9,sd = 75,sig.level = 0.05,type = "one.sample", alternative = "two.sided")  
delta\_test\_90

##   
## One-sample t test power calculation   
##   
## n = 5  
## delta = 147.4417  
## sd = 75  
## sig.level = 0.05  
## power = 0.9  
## alternative = two.sided

delta\_test\_90$delta

## [1] 147.4417

#80% power  
delta\_test\_80 <- power.t.test(n = 5,power = 0.8,sd = 75,sig.level = 0.05,type = "one.sample", alternative = "two.sided")  
delta\_test\_80

##   
## One-sample t test power calculation   
##   
## n = 5  
## delta = 126.1498  
## sd = 75  
## sig.level = 0.05  
## power = 0.8  
## alternative = two.sided

delta\_test\_80$delta

## [1] 126.1498

# 3.

## ia. Rejection Simulation (based on provided code):

set.seed(2345)  
# Set input values  
n <- 5  
mean <- 0  
sd <- 75  
numTrials <- 10000  
alpha <- 0.05  
# Set a counter to determine the number of rejected hypothesis tests  
count <- 0  
for(i in 1:numTrials){  
# Generate data  
y <- rnorm(n,mean,sd)  
# Perform test  
t <- t.test(y,alternative = "two.sided")  
count <- count + (t$p.value < alpha)  
}  
power <- count/numTrials  
power

## [1] 0.0498

The loop above generates 10,000 normal distributions with a mean of 0 and SD of 75. Then it performs a t-test on each distribution to see if the distribution’s mean is different from 0. If the p-value of the t-test is below our significance level of 0.05, the loop adds to the counter. Then, you divide the number of times that the mean was different from 0 by the number of simulations, to get the proportion of times the null hypothesis was rejected. The result is 0.0498, or about 5%, which is exactly what we’d expect.

## ib. Power Simulation (based on provided code):

set.seed(1796)  
# Set input values  
n <- 5  
mean <- 100  
sd <- 75  
numTrials <- 10000  
alpha <- 0.05  
# Set a counter to determine the number of rejected hypothesis tests  
count <- 0  
for(i in 1:numTrials){  
# Generate data  
y <- rnorm(n,mean,sd)  
# Perform test  
t <- t.test(y,alternative = "two.sided")  
count <- count + (t$p.value < alpha)  
}  
# Power = proportion of rejections  
power <- count/numTrials  
power

## [1] 0.613

The answer above is closest to 2.a.ii, which was calculating power with an unknown standard deviation. Power is the probability that we reject the null hypothesis given that the alternative hypothesis is true. In other words, the probability of finding a difference if it exists. So the result above is pretty much exactly what we would expect based on the power calculation in 2.a.ii.

## iii.

To estimate the required sample size for a given power, you could take the above for loop and run it for a range of values for n. Once you reached a value of n that resulted in a power of 0.9, you would have a decent estimate for your answer. You could also do the same thing for various values of mean (with a fixed n), in order to determine the smallest detectable difference.