

Stat 506 Assignment 9

Leslie Gains-Germain

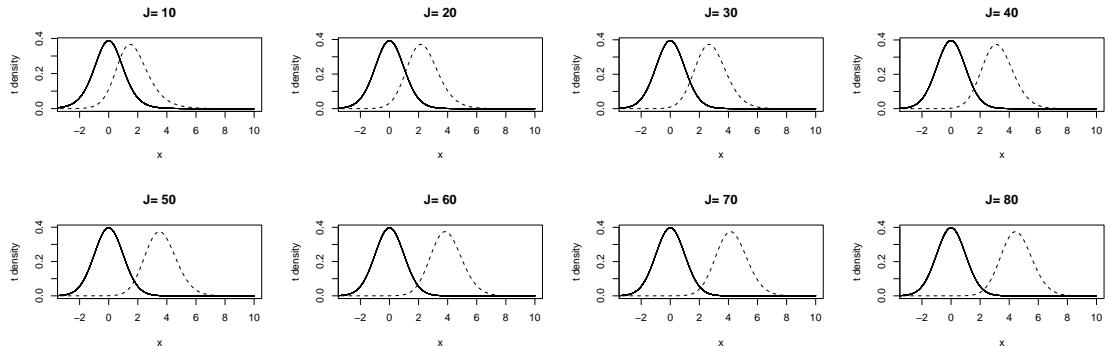
1. Two medications are designed to lower systolic blood pressure (SBP) for subjects with high blood pressure. Medication A has been shown to lower blood pressure by 30 units (on average) with subject-to-subject variance of 2 *units*², and has been approved by the FDA. The makers of medication B would like to show that it is as effective as Medication A, that is, to show that the mean difference between the medications is in the interval $(-1, 1)$ units. High blood pressure is a risk factor for stroke and heart attack so the subjects, who all suffer from high blood pressure if untreated, cannot be given a placebo. A cross-over design will be used with random allocation of the first medication (A or B) followed two weeks later by the other medication.
 - (a) The variable of interest will be the difference in mean blood pressure, SBP-A minus SBP-B. Write a formula for its variance assuming B and A are equally effective, and that the two measurements on the same person have correlation: ρ .

$$\begin{aligned} \text{Var}[\bar{x}_A - \bar{x}_B] &= \text{Var}[\bar{x}_A] + \text{Var}[\bar{x}_B] - 2\text{Cov}[\bar{x}_A, \bar{x}_B] \\ \text{Var}[\bar{x}] &= \frac{\sum \text{Var}[x_i]}{J^2} = \frac{\sigma_T^2}{J} \\ \text{Var}[\bar{x}_A - \bar{x}_B] &= \frac{\sigma_T^2}{J} + \frac{\sigma_T^2}{J} - 2\rho \frac{\sigma_T}{\sqrt{J}} \frac{\sigma_T}{\sqrt{J}} = \frac{4}{J} - \frac{4\rho}{J} \end{aligned}$$

where J is the number of subjects, and one measurement is taken on each person for each medication. σ_T^2 is assumed to be 2.

- (b) Use a stat package to draw a picture of the distributions of interest under the null hypothesis of no difference and under the alternative of interest: $\mu_A - \mu_B = 1$ unit. (Take $\rho = 0$ for simplicity)

I show the distributions of the t-statistics under the null and alternative hypotheses for sample sizes between 10 and 80.



- (c) Use pt and qt functions in R to compute the power of the test if 25 subjects are used and the testing is done with a significance level of $\alpha = .05$ for a two-sided rejection region. (show code)

```

pt(qt(0.025,24), 24, ncp=(5/2))+(1-pt(qt(0.975,24), 24, ncp=(5/2)))

## [1] 0.67

power.t.test(n=25, delta=1, sd=2, sig.level=0.05, power=NULL, type="paired")

##
##      Paired t test power calculation
##
##      n = 25
##      delta = 1
##      sd = 2
##      sig.level = 0.05
##      power = 0.67
##      alternative = two.sided
##
## NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs

```

The power is estimated to be 0.67.

- (d) Why might regulators insist on a smaller probability of type II error than the probability of type I error?

A type I error occurs if we conclude that medication B is not as effective as medication A, when in fact it is. If a type I error occurred, it would be unfortunate for the company that produces medication B because the product would be rejected. But, there would be no impact on consumers.

A type II error occurs if we conclude that medication B is as effective as medication A when in fact it is not. A type II error would have more of an impact because medication B would be approved for sale. In this case, consumers would be buying a less effective medication without knowing it.

Regulators will most likely insist on a smaller probability of type II error than type I error because a type II error would have a greater impact on consumers.

- (e) What minimum sample size is needed to bring the power up to at least 0.95?

```

##      J power
## 23 47 0.919
## 24 48 0.924
## 25 49 0.929
## 26 50 0.934
## 27 51 0.938
## 28 52 0.943
## 29 53 0.947
## 30 54 0.950
## 31 55 0.954
## 32 56 0.957

##
##      Paired t test power calculation
##
##      n = 53.9
##      delta = 1
##      sd = 2
##      sig.level = 0.05
##      power = 0.95
##      alternative = two.sided
##
## NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs

```

We need a sample size of at least 54 to bring the power up to 0.95.

- (f) How would things change if we used two independent groups of subjects: one getting SBP-A, the other (randomly assigned) SBP-B. Measurement is difference in SBP (before treatment - after treatment).

```
## Two-sample t test power calculation
##
##      n = 53
##      delta = 1
##      sd = 1.41
##      sig.level = 0.05
##      power = 0.95
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

We would need twice as many people, about 53 in each medication group (A and B) to achieve the same power as we did in the paired design.

2. Researchers are aware of many variables which are expected to affect y = the growth of calves in their first three months. These background variables use 5 degrees of freedom and are thought to account for 20% of the variation in y between calves. The researchers plan to study effects of dietary supplements fed to the mother cows, and have three formulations in mind. They come to you to ask about how many cow-calf pairs they should use in the study. [Aside: to keep the unit of the study distinct, cows will have to be randomly assigned treatments, not pens or herds of cows.] An important effect would be if the feeds explain an additional 5% of the variation in y .

- (a) After collecting the data, the analysis will compute an F test to assess the impact of feeds after adjusting for the background variables. Assuming we have n pairs in each treatment group, what are the numerator and denominator degrees of freedom for the F test?

The numerator degrees of freedom would be 2 and the denominator degrees of freedom would be $n - 8$.

- (b) Using Cohens derivation, compute the non-centrality parameter, λ , as a function of n .

$$\lambda = \frac{0.25-0.2}{1-0.25}(2 + (n - 8) + 1) = \frac{1}{15}(n - 5)$$

- (c) If the researchers use $\alpha = 0.05$, how large will each group of pairs need to be to achieve power of 90%?

```
n <- rep(0, 201)
power <- rep(0, 201)
for(i in 50:250){
  n[i-49] <- i
  power[i-49] <- 1-pf(qf(0.95, 2, i-8), 2, i-8, ncp=(1/15*(i-5)))
}
cbind.data.frame(n, power)[146:153,]

##      n power
## 146 195 0.895
## 147 196 0.897
## 148 197 0.899
## 149 198 0.900
## 150 199 0.902
## 151 200 0.903
```

```

## 152 201 0.905
## 153 202 0.906

require(pwr)
#pwr.anova.test(f=sqrt(1/15), k=3, power=0.9, sig.level=0.05)
#64.27981

```

My power analysis suggests that 198 calf cow pairs, 66 in each group, are required to achieve a power of 90% at a significance level of 0.05. The pwr.anova.test function says that a sample size of 65 pairs in each group will achieve a power of 90% (pretty close).

3. Exercise 5 in ARM Chapter 20 asks us to look back at the Electric Company data in Chapter 9. Remember that the treatment unit is a class of students. I think it would be of interest to detect a shift in mean of 5 points for the second graders. Use $\alpha = 0.05$ as they suggest and target power of 80%.

- (a) State all assumptions.

The sample standard deviations in the existing Electric Company data for the control and treated groups in the second grade are 12 and 10, respectively. For the power analysis, I'll assume a pooled standard deviation of 11, and I'll assume equal sample sizes in treated and control groups. We also assume that the population distributions of scores for both the control and treated groups are normally distributed, and we assume independence among class test scores. Lastly, we assume that the mean score of the treated group is five points higher than the non-treated group.

- (b) Simply compare the average of treated classes with the average for controls. (ignore the repeated measures aspect of the data.)

```

##      J power
## 125 149 0.787
## 126 150 0.790
## 127 151 0.792
## 128 152 0.795
## 129 153 0.798
## 130 154 0.800
## 131 155 0.803
## 132 156 0.805
## 133 157 0.808
## 134 158 0.810
## 135 159 0.813

```

```

##      Two-sample t test power calculation
##
##           n = 76.9
##           delta = 5
##             sd = 11
##    sig.level = 0.05
##        power = 0.8
##    alternative = two.sided
##
## NOTE: n is number in *each* group

```

Both my function and the power.t.test function suggest that a sample size of 154 total, 77 in each group, is needed to achieve a power of 80%.

- (c) Do the same with average gain scores.

The Electric Company data shows that the sample standard deviation of each group is approximately 6, and again I'll assume the same sample sizes in each group. I'll assume that we also want to detect a difference of 5 in the average gain scores as well. Both my function and the power.t.test function suggest that a sample of 48 total, 24 in each group, is needed to achieve a power of 80%.

```
##      J power
## 21 45 0.780
## 22 46 0.789
## 23 47 0.798
## 24 48 0.807
## 25 49 0.815
## 26 50 0.823
## 27 51 0.831
## 28 52 0.838
## 29 53 0.845

##
## Two-sample t test power calculation
##
##      n = 23.6
##      delta = 5
##      sd = 6
##      sig.level = 0.05
##      power = 0.8
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

- (d) Do the same in a regression context using pretest as predictor and simulating as in section 20.5.

I first fit a regression of posttest score on pretest and treatment for second grade classrooms. I then used the estimated regression model to simulate post test scores for simulated pretest scores in the treated/control groups.

```
electric <- read.csv("~/Documents/Stat506/Homework/HW9/electricComp.csv")
require(dplyr)
electric2 <- electric %>%
  filter(grade==2)
electric2$trt <- factor(electric2$trt, levels=c("untreated", "treated"))
lm.elec2 <- lm(posttest~pretest+trt, data=electric2)

electric.fake <- function(n){
  class <- 1:n
  treatment <- sample(rep(0:1, n/2))

  sigma <- 5.479

  #mean(electric2$pretest)
  #sd(electric2$pretest)
  pretest <- rnorm(n, 73.3, 12.47)

  posttest <- rnorm(n, coef(lm.elec2)[1]+coef(lm.elec2)[2]*pretest+
    coef(lm.elec2)[3]*treatment, sigma)
  return(data.frame(class, treatment, pretest, posttest))
}

electric.power <- function(n, n.sims=1000){
  signif <- rep(NA, n.sims)
  for(s in 1:n.sims){
    fake <- electric.fake(n)
    lm.power <- lm(posttest~pretest+treatment, data=fake)
    theta.hat <- coef(lm.power)[3]
```

```

    theta.se <- sqrt(vcov(lm.power)[3,3])
    signif[s] <- (theta.hat-2*theta.se) > 0
  }
  power <- mean(signif)
  return(power)
}

pow <- sapply(as.list(seq(48,64,by=2)), electric.power)
powd <- cbind.data.frame(n=seq(48,64,by=2), pow)
powd

##      n      pow
## 1 48 0.765
## 2 50 0.765
## 3 52 0.789
## 4 54 0.780
## 5 56 0.814
## 6 58 0.806
## 7 60 0.835
## 8 62 0.853
## 9 64 0.858

```

A sample size of 56, 28 in each group is required to achieve a power of 80%.

- (e) Write up the results to explain to a consulting client how large a sample size is needed.

If the goal is to detect a difference of 5 points in the mean posttest scores between treated and control groups, and you assume equal sample sizes between groups and a pooled standard deviation of 11, you will need 77 classes in each group to attain a power of 80%.

If the goal is to detect a difference of 5 points in the average gain scores (posttest-pretest) between treated and control groups, and you assume equal sample sizes between groups and a pooled standard deviation of 6, you will need 24 classes in each treatment group to attain a power of 80%. The above results can be found for different standard deviations and different effect sizes. Note that the sample size for gain scores is smaller because gain scores are estimated to be less variable than posttest scores.

In the simulation, we assume that the difference in posttest scores between treated and untreated groups in the existing sample is representative of the population. If you, as the researcher, have reason to believe that the current sample of second graders is not representative, then the results of this power analysis cannot be trusted. We first fit a regression model to the sample data, and then we create fake data by simulating pretest scores and treatment groups. We use the estimated regression model and the fake pretest scores and treatment groups to simulate the posttest scores. The estimated treatment effect from the regression model is the assumed effect size, 4.27 in this case, and the variability estimate also comes from the existing regression model. Under these assumptions, we found that a sample size of 28 classes in each treatment group achieves a power of 80%.

In the simulation, we account for pretest in the estimated regression model. So,

simulation is another method for determining the sample size needed to detect a difference in the average gain scores. In practice, if you do plan to account for pretest score in your analysis, I would use the estimated sample size of 28 per group (found by simulation) because it is more conservative than the estimate found with the parametric approach.

Scope of Inference: This power analysis is only meant to be used for further studies with the same population of second grade students and the same sampling design (with classes as the experimental units). We do not intend this analysis to be used in any other context because the variability estimates, effect sizes, and appropriate significance levels were estimated specifically for this Electric Company study.

R Code Appendix

```
nulldiff <- 0
altdiff <- 1
J <- seq(10,80, 10)
plot.fun <- function(J) {
  x <- seq(-6,10,.001)
  plot(x, dt(x, J-1), cex=.01, xlim=c(-3,10), main= paste("J=",J), ylab="t density")
  lines(x, dt(x, J-1, sqrt(J/4)), lty=2)
}
par(mfrow=c(2,4))
invisible(sapply(as.list(J), plot.fun))
```

```
pt(qt(0.025,24), 24, ncp=(5/2))+ (1-pt(qt(0.975,24), 24, ncp=(5/2)))
power.t.test(n=25, delta=1, sd=2, sig.level=0.05, power=NULL, type="paired")
```

```
J <- rep(0, 56)
power <- rep(0, 56)
for(i in 25:80){
  J[i-24] <- i
  power[i-24] <- pt(qt(0.025,i-1), i-1, ncp=(sqrt(i)/2))+ (1-pt(qt(0.975,i-1), i-1, ncp=(sqrt(i)/2)))
}
power.n <- cbind.data.frame(J, power)
require(xtable)
power.n[23:32,]
```

```
power.t.test(delta=1, sd=2, sig.level=0.05, power=0.95, type="paired")
```

```
power.t.test(delta=1, sd=sqrt(2), sig.level=0.05, power=0.95, type="two.sample")
```

```
n <- rep(0, 201)
power <- rep(0, 201)
for(i in 50:250){
  n[i-49] <- i
  power[i-49] <- 1-pf(qf(0.95,2,i-8),2, i-8, ncp=(1/15*(i-5)))
}
cbind.data.frame(n,power)[146:153,]
require(pwr)
#pwr.anova.test(f=sqrt(1/15), k=3, power=0.9, sig.level=0.05)
#64.27981
```

```

J <- rep(0, 176)
power <- rep(0, 176)
for(i in 25:200){
J[i-24]<-i
power[i-24] <- pt(qt(0.025,i-2), i-2, ncp=5*sqrt(i)/(11*sqrt(4)))+1-pt(qt(0.975,i-2), i-2, ncp=5*sqrt(i)/(11*sqrt(4)))
}
power.n <- cbind.data.frame(J, power)
require(xtable)
power.n[125:135,]
#pt(qt(0.025,143), 143, ncp=5*sqrt(145)/(11*sqrt(4)))+1-pt(qt(0.975,143), 143, ncp=5*sqrt(145)/(11*sqrt(4)))

```

```
power.t.test(delta=5, sd=11, sig.level=0.05, power=0.80, type="two.sample")
```

```

J <- rep(0, 176)
power <- rep(0, 176)
for(i in 25:200){
J[i-24]<-i
power[i-24] <- pt(qt(0.025,i-2), i-2, ncp=5*sqrt(i)/(6*sqrt(4)))+1-pt(qt(0.975,i-2), i-2, ncp=5*sqrt(i)/(6*sqrt(4)))
}
power.n <- cbind.data.frame(J, power)
require(xtable)
power.n[21:29,]
#pt(qt(0.025,143), 143, ncp=5*sqrt(145)/(11*sqrt(4)))+1-pt(qt(0.975,143), 143, ncp=5*sqrt(145)/(11*sqrt(4)))

```

```
power.t.test(delta=5, sd=6, sig.level=0.05, power=0.80, type="two.sample")
```

```

electric <- read.csv("~/Documents/Stat506/Homework/HW9/electricComp.csv")
require(dplyr)
electric2 <- electric %>%
  filter(grade==2)
electric2$trt <- factor(electric2$trt, levels=c("untreated", "treated"))
lm.elec2 <- lm(posttest~pretest+trt, data=electric2)

electric.fake <- function(n){
  class <- 1:n
  treatment <- sample(rep(0:1, n/2))

  sigma <- 5.479

  #mean(electric2$posttest)
  #sd(electric2$posttest)
  pretest <- rnorm(n, 73.3, 12.47)

  posttest <- rnorm(n, coef(lm.elec2)[1]+coef(lm.elec2)[2]*pretest+
    coef(lm.elec2)[3]*treatment, sigma)
  return(data.frame(class, treatment, pretest, posttest))
}

electric.power <- function(n, n.sims=1000){
  signif <- rep(NA, n.sims)
  for(s in 1:n.sims){
    fake <- electric.fake(n)
    lm.power <- lm(posttest~pretest+treatment, data=fake)
    theta.hat <- coef(lm.power)[3]
    theta.se <- sqrt(vcov(lm.power)[3,3])
    signif[s] <- (theta.hat-2*theta.se) > 0
  }
  power <- mean(signif)
  return(power)
}
```

```
pow <- sapply(as.list(seq(48,64,by=2)), electric.power)
powd <- cbind.data.frame(n=seq(48,64,by=2), pow)
powd
```