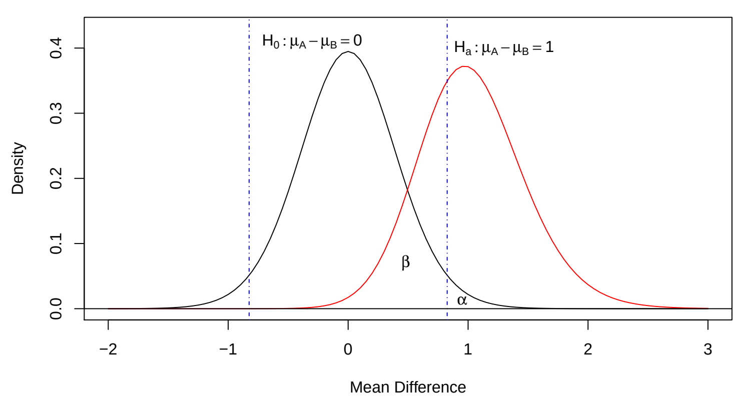
Homework 9 Solutions STAT 506 Spring 2015

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 for n subjects. Estimate its variance assuming B and A are equally effective, and that the two measurements on the same person have correlation: ρ.
   The difference in 2 dependent measurements has variance $\sigma^2 - \rho\sigma^2 = 4(1-\rho)$. We then take an average of n of these from (independent) different individuals, the variance of the mean difference, $\overline{x}_D$ is $4(1-\rho)/n$.

   (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)



   Based on $n = 25$, the plot shows the distribution of the sample mean under $H_0: \mu_D = 0$ and under the alternative: $H_0: \mu_D = 1$.

   (c) Use pt and qt functions in S to compute the power of the test if 25 subjects are used and the testing is done with a significance level of $\alpha = 0.05$ for a two-sided rejection region. (show code)

```
1-pt(qt(0.975,24), ncp=1/sqrt(4/n),24)

## [1] 0.6697014
```

   (d) Why might regulators insist on a smaller probability of type II error than the probability of type I error?
   In this case, a type II error, to decide the two are equivalent when they really are not, is more serious than a type I error, because we would be approving a non-effective or non-safe alternative. Therefore it makes sense to use $\beta \leq \alpha$.

   (e) What minimum sample size is needed to bring the power up to at least 0.95? (Use qt and pt in R/S or one of the power functions demonstrated in class.)
   We need at least 54 subjects to achieve 95% power.

   (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).

Variance of the difference within one individual is $2\sigma^2(1-\rho)$*. If we assume correlation* $\rho =$ 0 *(two observations on the same individual are far enough apart as to be uncorrelated) for simplicity, we now need 105 people in each group. If* $\rho = .5$*, then 53 in each group is adequate.*

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?
   2, $3n-8$

   (b) Using Cohen's derivation, compute the non-centrality parameter, λ, as a function of n.
   $(3n-5)\frac{.05}{.75}$

   (c) If the researchers use alpha = 0.05, how large will each group of pairs need to be to achieve power of 90%?
   66

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 $\alpha = .05$ as they suggest and target power of 80%.

   (a) State all assumptions.
   Each sampling unit is one school, at which we pick the two classrooms with lowest reading scores. We will assume, as in # 1 above, that classrooms in the same school are independent. ($\rho = 0$*). We assume classroom mean scores are normally distributed. We must also assume that we will be able to draw a random sample of schools from the population of interest, and that preliminary data from two cities is adequate for the larger population of schools.*

   (b) Simply compare the average of treated classes with the average for controls. (ignore the repeated measures aspect of the data.)
   First we need an estimate of variance in post test scores for second grade. Using the t test power function, I get a sample size of 79 needed in each group based on using $\hat{\sigma} = 11.1$ *as the standard deviation of post-treatment reading scores.*

   (c) Do the same with average gain scores.
   Gain scores have lower spread ($\hat{\sigma} = 6$*), so now we need only 24 per group. from* `power.t.test`

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

	Sample Size:	16	17	20
Power		0.80	0.81	0.87

   Simulating data under the null, I used $\hat{\beta}_1 = .79$ *and* $\hat{\sigma} = 5.48$ *based on fitting postScore to PreScore and treatment. Prescores were created as random normals with mean = 73.3 and SD = 12.5. To find sample size, I noted that with 34 schools, we did get a small p-value, so I theorized that fewer than 34 schools were needed. With 25 schools, I got power = 0.944, with 20 schools,* $1-\beta$ *= 0.871, (and I had to do 5000 sims to get a good estimate). Using 16 schools gave less than 80% power, so I settled on 17 schools for power = 0.812*

   (e) Write up the results to explain to a consulting client how large a sample size is needed.
   We should reiterate the assumptions: normal distributions and independence between classes in the same school.

Looking just at posttest scores is not a good idea. Gain scores are like regression with slope set to one, but a sample size of 24 schools is adequate to obtain power of 0.8 to detect a 5 unit shift with significance level α.

An alternative is to recognize that we will use a linear model to estimate effects in the end, and the slope on prescore need not be set to one. With regression assumptions, the sample size needed drops to 17 schools, but my recommendation would be to use at least 25 schools to attempt to cover unforeseen variation.

R Code

1 b.

```
n <- 25
curve( dt(x/sqrt(4/n) ,24),from = -2, 3, xlab="Mean Difference", ylab = "Density",ylim=c(0,.43))
 curve( dt(x/sqrt(4/n),ncp=1/sqrt(4/n) ,24),typ="l",col=2, add=TRUE)
 abline(v=c(-1,1)*qt(.025,24)*2/5,lty=4,col=4);   abline(h=0)
 text(x=c(.95,.48),y=c(.012,.07), expression(alpha, beta))
 text(x=c(-.3,1.3), y = c(0.41,0.40),expression(H[0]: mu[A] - mu[B]==0, H[a]: mu[A] - mu[B] == 1))
```

1 c.

```
 1-pt(qt(0.975,24), ncp=1/sqrt(4/n),24)

## [1] 0.6697014
```

1 e.

```
 powert <- function(alpha, n, delta)  1-pt(qt(1-alpha, n-1),n-1, ncp=delta)
 n <- seq(25,60,5)
  rbind(round(powert(.025,n, sqrt(n)/2),3),n)

##    [,1]   [,2]  [,3]   [,4]   [,5]   [,6]   [,7]   [,8]
##    0.67  0.754  0.82  0.869  0.907  0.934  0.954  0.968
## n 25.00 30.000 35.00 40.000 45.000 50.000 55.000 60.000

 n <- 51:55
  rbind(round(powert(.025,n, sqrt(n)/2),3),n)

##     [,1]   [,2]   [,3]  [,4]   [,5]
##    0.938  0.943  0.947  0.95  0.954
## n 51.000 52.000 53.000 54.00 55.000

 power.t.test(power=.95,sig.level=.05,delta=1,sd=2,type="one",alt="two")$n

## [1] 53.94062
```

1 f.

```
 power.t.test(power=.95,sig.level=.05,delta=1,sd=sqrt(8),type="two",alt="two")$n

## [1] 208.8807

 power.t.test(power=.95,sig.level=.05,delta=1,sd=2,type="two",alt="two")$n

## [1] 104.928

 power.t.test(power=.95,sig.level=.05,delta=1,sd=sqrt(2),type="two",alt="two")$n

## [1] 52.95899
```

2

```r
powerf <- function(alpha,u,v,r2initial, r2final){
    ## function to compute power as in Cohen (1988) using R^2
    ## alpha is significance level,
    ## u, v are df used by baseline vbles and added predictor,
    ## r2initial: proportion of variance explained by baseline
    ## r2final: proportion variation explained using all vbles
    1-pf(qf(1-alpha,u,v),u,v,(u+v+1)*(r2final - r2initial)/(1-r2final))
}
 n <- 61:69
 powerf(.05,2,3*n-8,.20,.25)

## [1] 0.8735943 0.8793787 0.8849299 0.8902552 0.8953618 0.9002567 0.9049472
## [8] 0.9094400 0.9137421
```

3 a.

```r
  electric <- read.table("../../data/electric.dat",head=TRUE)
  post <- with(subset(electric, Grade==2),cbind(treated.Posttest,control.Posttest))
  sd1 <- mean(apply(post,2,sd))
  power.t.test(power=.80,sig.level=.05,delta=5,sd=sd1,type="two",alt="two")$n

## [1] 78.15407
```

3 b.

```r
  diffrnce <- with(subset(electric, Grade==2),cbind(treated.Posttest-treated.Pretest,
                                      control.Posttest-control.Pretest))
  sd2 <-  mean(apply(diffrnce,2,sd))
power.t.test(power=.80,sig.level=.05,delta=5,sd=sd2,type="two",alt="two")$n
```

3 c.

```r
 school2 <- with(subset(electric, Grade==2),
                 data.frame( preScore = c(treated.Pretest, control.Pretest),
                             postScore = c(treated.Posttest, control.Posttest),
                             trt = rep(1:0, each = 34)))
#summary( fit1 <- lm(postScore ~ preScore + trt, school2))

fake.data <- function(shift, slope,mean1,  sigma1, int1, sigma2, n.group){
    preScore <- rnorm(2*n.group, mean1, sigma1)
    trt <- rep(0:1, each = n.group)
    postScore <-  int1 + trt * shift + slope * preScore + rnorm(2*n.group, 0, sigma2)
    data.frame(preScore, trt, postScore)
}

school.sim <- function(n.sim,...){
    hits <- rep(NA, n.sim)
    for(i in 1:n.sim){
        fit <- lm( postScore ~ preScore + trt, fake.data(...))
        hits[i] <- (summary(fit)$coef[3,4] < .10) & (coef(fit)[3] > 0)
    }
    sum(hits)/n.sim
}
 ## try 25 schools
 p25 <- school.sim(n.sim=1000, shift=5, slope=.79, mean1 = 73.3, sigma1 = 12.5, int1 =  37.4,
```

```
                  sigma2 = 5.48, n.group = 25)
 p17 <- school.sim(n.sim=5000, shift=5, slope=.79, mean1 = 73.3, sigma1 = 12.5, int1 =  37.4,
 p16 <- school.sim(n.sim=5000, shift=5, slope=.79, mean1 = 73.3, sigma1 = 12.5, int1 =  37.4,

p20 <- school.sim(n.sim=5000, shift=5, slope=.79, mean1 = 73.3, sigma1 = 12.5, int1 =  37.4,
                  sigma2 = 5.48, n.group = 20)
p21 <- school.sim(n.sim=5000, shift=5, slope=.79, mean1 = 73.3, sigma1 = 12.5, int1 =  37.4,
                  sigma2 = 5.48, n.group = 21)

results <- matrix( c(NA, p16, p17, p20),1,4)
dimnames(results) <- list( "Power",  c("Sample Size:", 16, 17, 20))
xtable::xtable(results)
```