Homework #7

Carson Crenshaw (cgc8gdt)

Insert packages and set the seed.

```
# Set Seed
set.seed(10062002)
#Library
library(ggplot2)
library(BSDA)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'BSDA'
```

```
## The following object is masked from 'package:datasets':
##
##     Orange
```

```
library(UsingR)
```

```
## Loading required package: MASS
```

```
## Loading required package: HistData
```

```
##
## Attaching package: 'HistData'
```

```
## The following object is masked from 'package:BSDA':
##
##     Wheat
```

```
## Loading required package: Hmisc
```

```
##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(OpenMx)
```

```
## OpenMx may run faster if it is compiled to take advantage of multiple cores.
```

```
library(pwr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:Hmisc':
##
##     src, summarize

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Problem 1

Based on these assumptions, the sample size should be 27.

```
# Left-tailed one-sample t-test, moderate effect size, 80% power.
pwr.t.test(d=-0.5, sig.level=0.05, power=0.8, type="one.sample", alternative="less")
```

```
##
##      One-sample t test power calculation
##
##               n = 26.13753
##               d = -0.5
##       sig.level = 0.05
##           power = 0.8
##     alternative = less
```

Problem 2

Part a

The practitioner did not achieve their 80% goal.

```
q2afunc <- function(x){
  samp <- rnorm(x, 72.5, 3.8)
  ttest <- t.test(samp, mu = 73, alternative = "less")
  #to find empirical power the values should not match
  ttest$p.value <= 0.05
}
sum(replicate(10000, q2afunc(27)))/10000
```

```
## [1] 0.1622
```

Part b

The true effect size of the practitioner's test is 0.13.

```
## Determine the effect size: alt - null / sd
efs <- (73-72.5)/3.8
efs
```

```
## [1] 0.1315789
```

Part c

Based on the true effect size, the new sample size is 359.

```
pwr.t.test(d=-0.1316, sig.level=0.05, power=0.8, type="one.sample", alternative="less")
```

```
##
##      One-sample t test power calculation
##
##               n = 358.3469
##               d = -0.1316
##       sig.level = 0.05
##           power = 0.8
##     alternative = less
```

Part d

With the new sample size of 359, the practitioner did achieve their 80% goal.

```
q2dfunc <- function(x){
  samp <- rnorm(x, 72.5, 3.8)
  ttest <- t.test(samp, mu = 73, alternative = "less")
  #to find empirical power the values should not match
  ttest$p.value <= 0.05
}
sum(replicate(10000, q2dfunc(359)))/10000
```

```
## [1] 0.8054
```

Problem 3

Part a

The practitioner did not achieve their 80% goal using the original sample size of 27.

```
#chi-squared distribution with 4 degrees of freedom, mean of 4, sd of 2.83.
#exploring if the population mean is less than 4.25
#use sample size from question 1

q3afunc <- function(x){
  samp <- rchisq(x, df=4)
  ttest <- t.test(samp, mu = 4.25, alternative = "less")
  ttest$p.value <= 0.05
}
sum(replicate(10000, q3afunc(27)))/10000
```

```
## [1] 0.1722
```

Part b

The true effect size of the practitioner's test is 0.09.

```
## Determine the true effect size: alt - null / sd
efs2 <- (4.25-4)/2.83
efs2
```

```
## [1] 0.08833922
```

Part c

Using the true effect size found in part(b), the sample size required for 80% power is 795.

```
pwr.t.test(d=-0.0883, sig.level=0.05, power=0.8, type="one.sample", alternative="less")
```

```
##
##      One-sample t test power calculation
##
##               n = 794.3062
##               d = -0.0883
##       sig.level = 0.05
##           power = 0.8
##     alternative = less
```

Part d

With the new sample size of 795, the practitioner did achieve their 80% goal for the chi-squared distribution.

```
q3dfunc <- function(x){
  samp <- rchisq(x, df=4)
  ttest <- t.test(samp, mu = 4.25, alternative = "less")
  ttest$p.value <= 0.05
}
sum(replicate(10000, q3dfunc(795)))/10000
```

```
## [1] 0.7939
```

Problem 4

Following the results from Problems 2 and 3, one can conclude that the effect size is an important consideration regardless as to whether the study lacks or contains a population mean/standard deviation.

For both the normal distribution and the chi-square distribution, the code resulted in sample sizes that were much larger than the original sample using the assumed effect size (27 compared to 359 and 795, respectively). A smaller effect size and larger sample size are essentially a consequence of the slight variation between the null and alternative values in the problems which run Monte Carlo simulations. The small difference between the null and alternatives (72.5 vs. 73 and 4 vs. 4.25) in Problem 2 and 3 results in a small effect size, indicating that the power is going to be low. Identifying a smaller effect size would then result in lower power, then necessitating a much larger sample size so that the power can reach 80%.

To give this conclusion more context, if one pictures the null distribution in comparison to the alternative distribution, the power probability is represented under the curve. In that case, the further a part the two means are, the larger the power will be. As long as the sample size is fixed and the significance level is fixed, a higher effect size is going to result in higher power because it's essentially making the null hypothesis more and more false. For Problem 2 and 3, however, the null and alternative distributions heavily overlap and are very similar, therefore the concept of power as the probability that we correctly reject the null will be smaller.

For future studies, therefore, one must do what was accomplished within this homework exercise: When researchers have very low effect sizes, one must attempt to make that area under the aforementioned curve increase. One way to do that, for example, is to increase the sample size.