1 Power of the classical two-sample t-test

Q1.1: Power of a hypothesis test is equal to the probability of rejection of the null hypothesis when the alternative hypothesis is true. Power is always computed against a specific alternative which depends on a chosen effect size.

For this assignment, assume that the "null" distribution is standard normal, that is, $\mathcal{N}(0,1)$. Suppose we would like to detect a difference in means which is equal to two standard deviations (the effect size equals 2). So the alternative distribution can be chosen as $\mathcal{N}(2,1)$. Notice that the classical two-sample t-test assumes that both samples have the same variance. That is why, for the power simulations, one should choose an "alternative" distribution with the same variance as the null distribution.

Using R, implement the following algorithm to compute the achieved power :

- Specify: a number of simulations N (e.g. 10000), a sample size m (e.g. 8), a significance level (often, $\alpha = 0.05$).
- Simulate data under the null and alternative hypotheses: draw two independent samples of size m from the normal distributions described above.
- Use two-sample t-test to see whether it is able to reject the null hypothesis at the chosen significance level.
- Repeat the procedure N times.
- Compute the proportion of simulations in which the null hypothesis was rejected.
- HINT: **fix a random seed** at the beginning, if you want to have reproducible results.

```
sample.size <- 7;
mean.null <- 0;
alternative.mean <- 2;
sd.null <- 1;  ## it is assumed that the standard deviation is the same
Nsim <- 1000;
my.random.seed <- 20;
pvals <- numeric(Nsim);
set.seed(my.random.seed);
for(i in 1:Nsim)
{
  set.seed(my.random.seed + i);
  sample.1 <- rnorm(sample.size, mean.null,sd.null);
  sample.2 <- rnorm(sample.size,alternative.mean,sd.null);
  pvals[i] <- t.test(sample.1, sample.2,var.equal = TRUE)[[3]];
}
length(which(pvals < 0.05))/Nsim
```

Q1.2: How to determine a sample size which ensures a desired power for a given effect size?

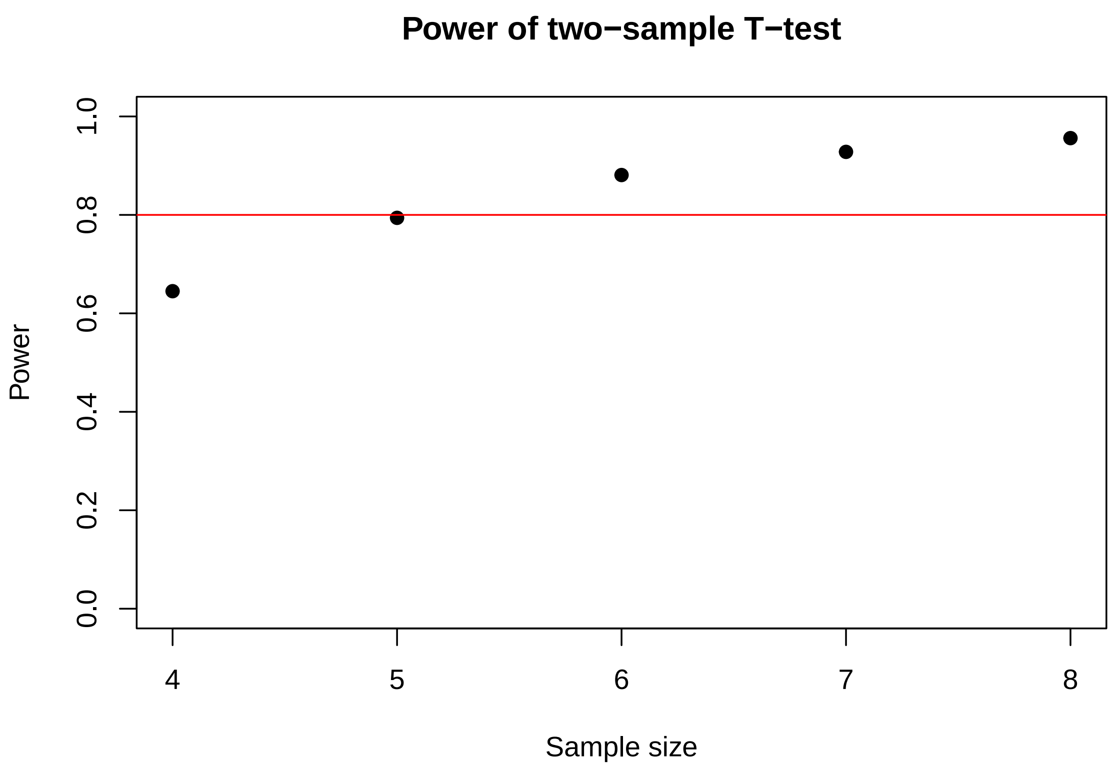
This procedure may be needed during the planning stage of your experiment, when you are to decide how large your samples should be to achieve a prespecified test power. The relationship between test power and sample size cannot be established without specifying an effect size, but if the experiment is performed for the first time, the expected effect size can only be guessed. A range of plausible values can sometimes be obtained from relevant literature.

1. Choose a range of sample sizes and, using the function in Q1.1 to find the minimal sample size which leads to the 80 percent power. Make a plot illustrating your explorations.

2. Use the R library *pwr to see how close your calculations to the R output. It could be the case that your answer is not exactly the same, but it should be relatively close.*

Q1.3: Re-use your R functions to explore different effect sizes which would lead to 80 percent power given a fixed sample size. Make a plot illustrating your explorations.

```
cms.course.power <- function(sample.size,mean.null,alternative.mean,
                             sd.null,Nsim,alpha,my.random.seed) {
   pvals <- numeric(Nsim);
   my.power <- 0;
   set.seed(my.random.seed);
   for(i in 1:Nsim)
   {
     set.seed(my.random.seed + i);
     sample.1 <- rnorm(sample.size, mean.null,sd.null);
     sample.2 <- rnorm(sample.size,alternative.mean,sd.null);
     pvals[i] <- t.test(sample.1, sample.2,var.equal = TRUE)[[3]];
   }

   my.power <- length(which(pvals < alpha))/Nsim;
   return(my.power);
}

cms.course.power(7, 0, 2, 1, 1000, 0.05, 20);

## [1] 0.928

SSIZE <- 4:8;
POWER <- numeric(length(SSIZE));
for(j in 1:length(SSIZE))
{
   POWER[j] <- cms.course.power(SSIZE[j], 0, 2, 1, 1000, 0.05, 20);
}

PWR.Table <- cbind(SSIZE, POWER);
PWR.Table  ## based on the selected random seed, the min sample size = 6.

##      SSIZE POWER
## [1,]     4 0.645
## [2,]     5 0.794
## [3,]     6 0.881
## [4,]     7 0.928
## [5,]     8 0.956

plot(SSIZE,POWER, main="Power of two-sample T-test",
     xlab="Sample size", ylab="Power", pch=20, cex=1.5, ylim = c(0,1))
abline(h=0.8, col="red")
```

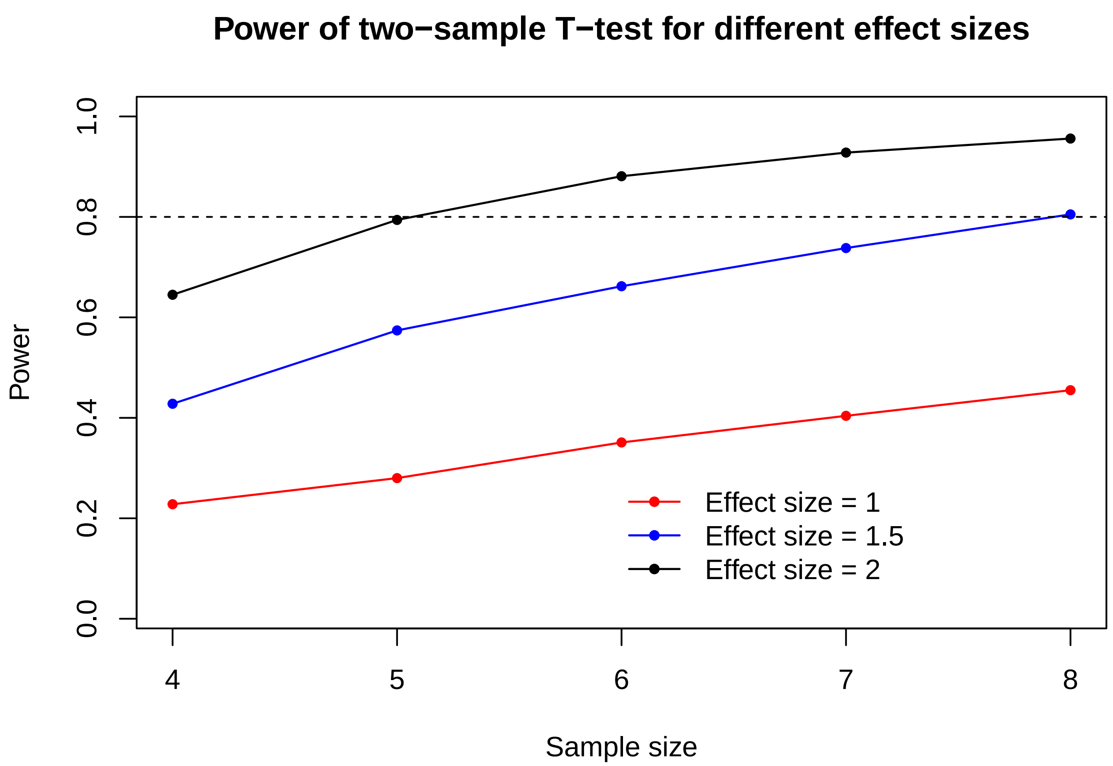



```
library(pwr);
mean.null <- 0;
alternative.mean <- 2;
sd.null <- 1;
ESize <- (alternative.mean - mean.null)/sd.null;  ## Effect size
pwr.t.test(d = ESize, power = 0.8, sig.level = 0.05, alternative = "two.sided", type = "two.sample");
```

```
##
##      Two-sample t test power calculation
##
##               n = 5.089995
##               d = 2
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
effect.size <- c(1,1.5,2);
sd.null <- 1;
POWER.all <- matrix(0, nrow =length(effect.size), ncol= length(SSIZE));

for( f in 1:length(effect.size))
{
    alt.mean <- effect.size[f]*sd.null;
```

```
    for(k in 1:length(SSIZE))
     {
        POWER.all[f,k] <- cms.course.power(SSIZE[k], 0, alt.mean, 1, 1000, 0.05, 20);
          #my.t.test.power(SSIZE[k],0, alt.mean,1, 1000, 0.05,2)
    }
}

POWER.all

##       [,1]  [,2]  [,3]  [,4]  [,5]
## [1,] 0.228 0.280 0.351 0.404 0.455
## [2,] 0.428 0.574 0.662 0.738 0.805
## [3,] 0.645 0.794 0.881 0.928 0.956

plot(SSIZE, POWER.all[1,],  col="red", pch=20, ylim=c(0.02, 1),
type="l",main="Power of two-sample T-test for different effect sizes",
     xlab="Sample size", ylab="Power", lwd=1.2, xaxt="n")
points(SSIZE, POWER.all[1,],  col="red", pch=20)
axis(1, at= SSIZE)
points(SSIZE, POWER.all[2,],  col="blue", pch=20, type="l",lwd=1.2)
points(SSIZE, POWER.all[2,],  col="blue", pch=20)
points(SSIZE, POWER.all[3,],  col="black", pch=20, type = "l",lwd=1.2)
points(SSIZE, POWER.all[3,],  col="black", pch=20)
abline(h=0.8, lty=2)
leg.txt = c("Effect size = 1", "Effect size = 1.5", "Effect size = 2")
legend(6,0.3,pch = c(20, 20, 20), lwd=c(1.2,1.2,1.2), col = c("red","blue","black"),
leg.txt,box.lwd = 0,box.col = "white",bg = "white")
```


Power of two−sample T−test for different effect sizes
Power
1.0
0.8
0.6
0.4
0.2
0.0
Effect size = 1
Effect size = 1.5
Effect size = 2
4
5
6
7
8
Sample size