

# Week 05 Good statistical practice

Open and reproducible science: dependable computations and statistics

## In-class tasks Solutions

### Task 1 Recalculate the sample size in “Avoiding overhead aversion in charity”

In this session we will further look at sample size calculations for “Avoiding overhead aversion in charity” using the R package `pwr`.

**Question 1** Assume again that an increase in donation frequency from 65% to 75% is meaningful for a charity to consider finding coverage for overhead costs of 5%. How big would a sample need to be to detect that difference with 80% power if you calculate with the function `pwr.2p.test`? Start by looking at `?pwr.2p.test`.

```
library(pwr)
#edit(vignette("pwr-vignette"))
pwr.2p.test(h = ES.h(p1 = 0.75, p2 = 0.65),
            sig.level = 0.05,
            power = 0.80,
            alternative = "greater")

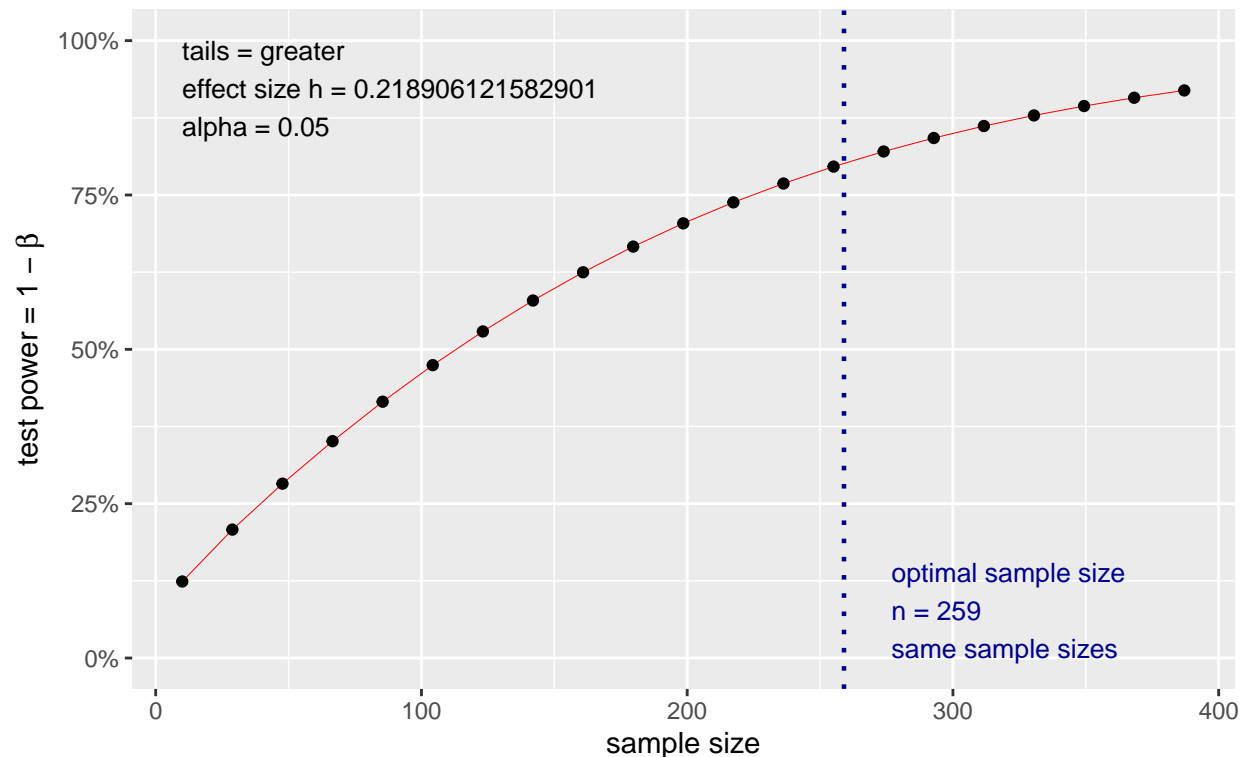
##
##      Difference of proportion power calculation for binomial distribution (arcsine transformation)
##
##              h = 0.2189061
##              n = 258.0372
##      sig.level = 0.05
##              power = 0.8
##      alternative = greater
##
## NOTE: same sample sizes
```

How large should you plan the sample size in total?

**Question 2** Use the `plot` function of the `pwr` package to investigate how power changes in this setting with sample size.

```
p.out <- pwr.2p.test(h = ES.h(p1 = 0.75, p2 = 0.65),
                    sig.level = 0.05,
                    power = 0.80,
                    alternative = "greater")
plot(p.out)
```

## Difference of proportion power calculation for binomial distribution (arcsine transformation)



Describe what you observe in the plot.

**Question 3** Do you expect the sample size to increase or decrease when you change the significance level to 1%? Check with `pwr.2p.test`.

```
pwr.2p.test(h = ES.h(p1 = 0.75, p2 = 0.65),
  sig.level = 0.01,
  power = 0.80,
  alternative = "greater")
```

```
##
##      Difference of proportion power calculation for binomial distribution (arcsine transformation)
##
##              h = 0.2189061
##              n = 418.8669
##      sig.level = 0.01
##      power = 0.8
##      alternative = greater
##
## NOTE: same sample sizes
```

**Question 4** Using a one-sided alternative is a strong assumption. Do you expect the sample size to increase or decrease when you change to a two sided alternative? Two sided alternatives are the default in the `pwr` package. In medical research they are also the default and one has to argue convincingly if a one sided version of a test needs to be used for some reason. Check with `pwr.2p.test`.

```
pwr.2p.test(h = ES.h(p1 = 0.75, p2 = 0.65),
            sig.level = 0.05,
            power = 0.80)
```

```
##
##      Difference of proportion power calculation for binomial distribution (arcsine transformation)
##
##              h = 0.2189061
##              n = 327.5826
##      sig.level = 0.05
##      power = 0.8
##      alternative = two.sided
##
## NOTE: same sample sizes
```

**Question 5** The implementation of `pwr.2p.test` is based on an arcsine transformation. This ensures that differences in the extremes are transformed to require larger sample sizes. How large a sample you need for an increase in donation frequency from 35% to 45%?

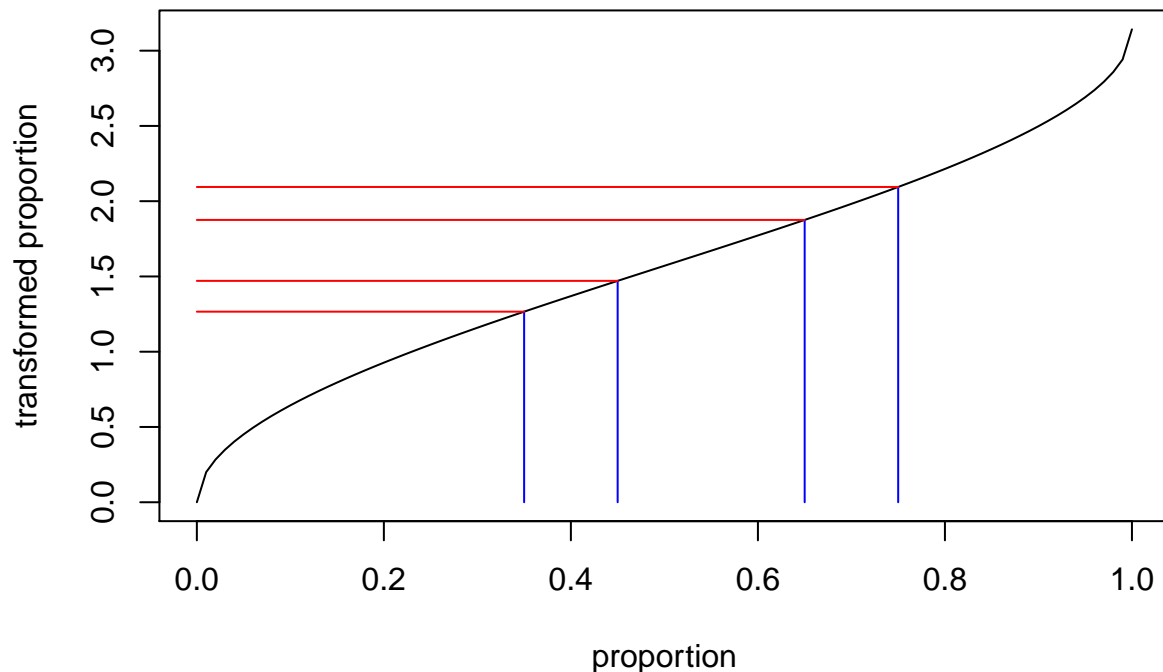
```
pwr.2p.test(h = ES.h(p1 = 0.45, p2 = 0.35),
            sig.level = 0.05,
            power = 0.80)
```

```
##
##      Difference of proportion power calculation for binomial distribution (arcsine transformation)
##
##              h = 0.2045252
##              n = 375.2691
##      sig.level = 0.05
##      power = 0.8
##      alternative = two.sided
##
## NOTE: same sample sizes
```

See the below picture adapted from the vignette of `pwr` illustrating that the transformed difference in proportion is smaller in the second case, the difference increases towards the extremes. The function `power.prop.test` that you used in the homework uses on raw proportions.

```
## ----fig.height=5, fig.width=5-----
addSegs <- function(p1, p2){
  tp1 <- ES.h(p1, 0); tp2 <- ES.h(p2, 0)
  segments(p1,0,p1,tp1, col="blue"); segments(p2,0,p2,tp2,col="blue")
  segments(0, tp1, p1, tp1, col="red"); segments(0, tp2, p2, tp2, col="red")
}

curve(expr = ES.h(p1 = x, p2 = 0), xlim = c(0,1),
      xlab = "proportion", ylab = "transformed proportion")
addSegs(p1 = 0.65, p2 = 0.75) # 50% vs 55%
addSegs(p1 = 0.35, p2 = 0.45) # 5% vs 10%
```



## Task 2 Sample size for continuous outcome

Consider a fictional example of a fitness intervention in schools. Does the inclusion of regular leisurly running outings during lessons over a semester improve the the distance covered in a six minute walk compared to a control of no changes in curriculum? There are two principle approaches to test this question, a before and after assessment in one school/group or a concurrent assessment in two schools/group.

**We need to specify a measure of variability here. For proportions this is not necessary since it is assumed to be a function of the proportion.**

**Question 1** How many students do you need to assess in order to detect a difference of 40m with a power of 80% and a significance level of 5%? Use the function `pwr.t.test` for the first approach. Assume a standard deviation of 62m.

```
delta <- 40
mysd <- 62
mysig <- 0.05
mypower <- 0.8
scalcl <- pwr.t.test(sig.level = mysig,
  alternative = "two.sided",
  type = "paired",
  power = mypower,
  d = delta/(mysd*sqrt(2)) )
print(scalcl)
```

##

```
##      Paired t test power calculation
##
##          n = 39.67736
##          d = 0.4561979
##      sig.level = 0.05
##          power = 0.8
##      alternative = two.sided
##
## NOTE: n is number of *pairs*
```

*Hint:* The function uses the so-called Cohen's d, which is the effect divided by the standard deviation. Since here we look at differences in distance for the same person we need to multiply the standard deviation of a single distance, i.e. 62, by the square root of 2.

**Question 2** Now we look at the second approach. Assessing on group at the beginning of a semester and then again the end is considered suboptimal because many students are sedantary during the summer breaks. Hence, we need a control group from a different school/group to finish the assessment within one semester. Consider that you have 60 students in your first group. How many students do you need to assess in the control group in order to detect a difference of 50m with a power of 80% and a significance level of 5%? Use the function `pwr.t2n.test` for the first approach. Assume a standard deviation of 60m.

```
delta <- 25
mysd <- 62
mysig <- 0.05
mypower <- 0.8
nschool1 <- 60
scalc2<-pwr.t2n.test(n1= nschool1,
                     sig.level = mysig,
                     alternative = "two.sided",
                     power = mypower,
                     d = delta/mysd)
print(scalc2)
```

```
##
##      t test power calculation
##
##          n1 = 60
##          n2 = 255.0041
##          d = 0.4032258
##      sig.level = 0.05
##          power = 0.8
##      alternative = two.sided
```

*Hint:* Here, each assessment is only done once on every person, hence we do not need a multiplier for the standard deviation.

You can see that the total sample size in the second approach is much larger, compare 40 to 316, unequal group sizes have this effect. Sometimes it is nevertheless necessary to consider group sizes of 1:2 or, as here 1:3.

**Question 3** Now consider that you run this test in the whole school district over two semesters. You will assess 12 year old 2500 boys after a first “normal” semester and then again after the “intervention” semester. How big is the smallest effect that you can detect at 80% power?

```
ndistri <- 2500
scal3 <- pwr.t.test(n = ndistri,
  sig.level = mysig,
  alternative = "two.sided",
  power = mypower)
scal3$d*(mysd*sqrt(2))
```

```
## [1] 6.94992
```

What do you think about the size of the effect?