Teaser to Power analysis by data simulation

T. B.

Simulate data for a two-sample comparison, starting with a fixed sample size and fixed difference between the samples.

```
set.seed(123) # random seed to make code reproducible
samplesize <- 40
sampdiff <- 0.5
samp1 <- rnorm(n = samplesize, mean = 0, sd = 1) #random numbers following normal
#distribution of mean 0 and standard deviation 1
samp2 <- rnorm(n = samplesize, mean = sampdiff, sd = 1)
```

T-test comparing the two samples. To stay very basic, we will just look at the p-value from the test:

```
testres <- t.test(samp1, samp2)
testres$p.value
```

```
## [1] 0.0341346
```

That's "significant", we can distinguish the difference between the samples from noise.

To compute the power, we need many repeats of this process:

```
nbsimul <- 1000
set.seed(123) # random seed to make code reproducible
samplesize <- 40
sampdiff <- 0.5
storepvalues <- vector(length = nbsimul)
for (i in 1:nbsimul)
{
  samp1 <- rnorm(n = samplesize, mean = 0, sd = 1)
  samp2 <- rnorm(n = samplesize, mean = sampdiff, sd = 1)
  testres <- t.test(samp1, samp2)
  storepvalues[i] <- testres$p.value # store the pvalue from this data set test
}
```

Statistical power is the proportion of significant results, that is p-values below our significance threshold (typically 5%), which we can compute as:

```
mean(storepvalues<0.05)
```

```
## [1] 0.598
```

So the power is estimated to be 59.8% for a sample size of 40 per sample, a difference of 0.5 between the samples, and a defaut t-test.