

Experimental Design II: replication and power exercise 1

Lieven Clement & Alexandre Segers

statOmics, Ghent University (<https://statomics.github.io>)

Contents

1	Power	1
2	Rodents	2
2.1	Data exploratie	2
3	Analyse of the data	4
4	Power of the test to detect the same effect size as observed in our dataset with our experimental design?	5
4.1	Simulation function	5
4.2	Simulation	6
5	Power for a balanced design	6
6	Required sample size to obtain a power of 90 %?	7
7	Impact of effect size	9

```
library(tidyverse)
```

1 Power

The power of a test is defined as:

$$P(p < \alpha | H_1)$$

This is the probability to reject the nullhypothesis at the significance level α given that the alternative hypothesis is true.

The power depends on:

- the real effect size in the population $\mathbf{L}^T \beta$.
- the number of observations: SE and df.
- Choice of designpoints

- Choice of significance-level α .

We will evaluate the power using simulation.

2 Rodents

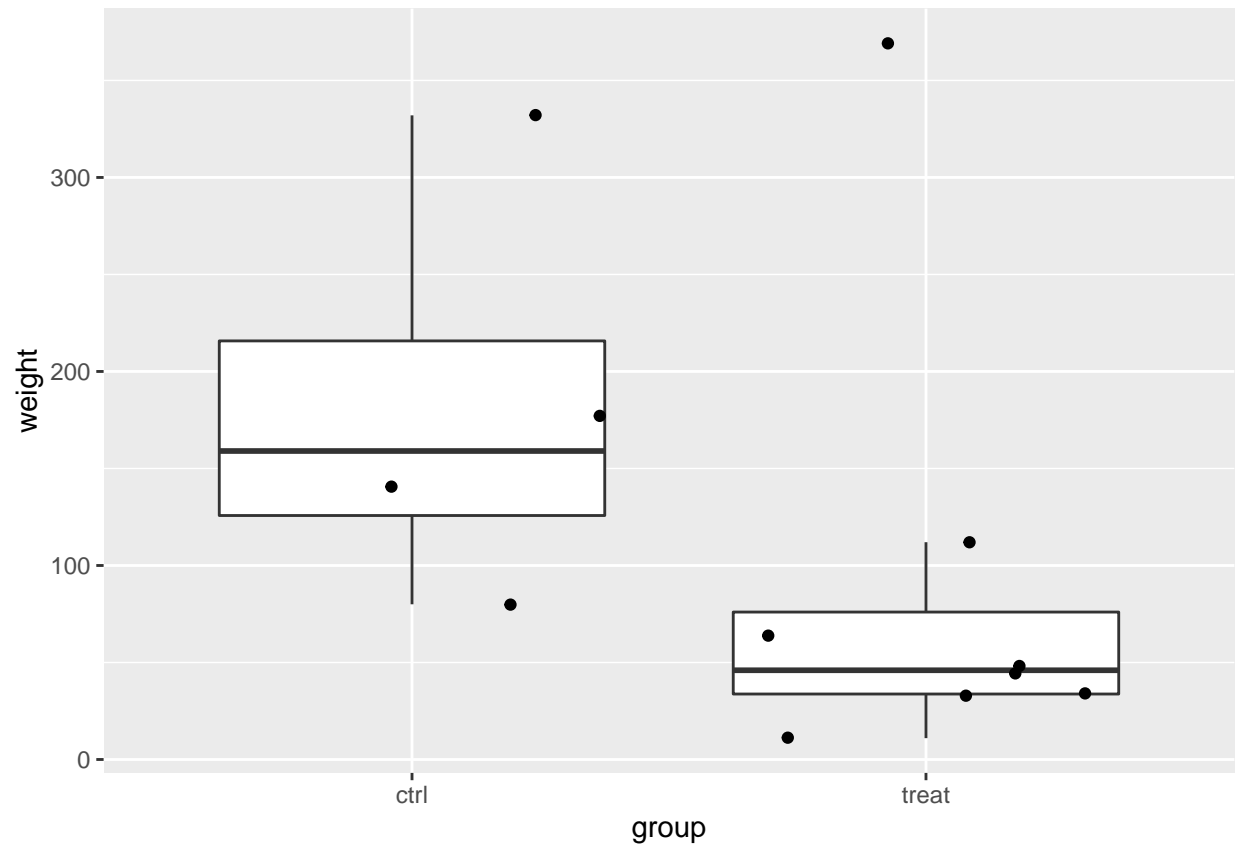
A biologist examined the effect of a fungal infection on the eating behavior of rodents. Infected apples were offered to a group of eight rodents, and sterile apples were offered to a group of 4 rodents. The amount of grams of apples consumed per kg body weight are given in the dataset below.

```
rodents <- data.frame(weight=c(11,33,48,34,112,369,64,44,177,80,141,332),group=as.factor(c(rep("treat",8),rep("ctrl",4))),
rodents
```

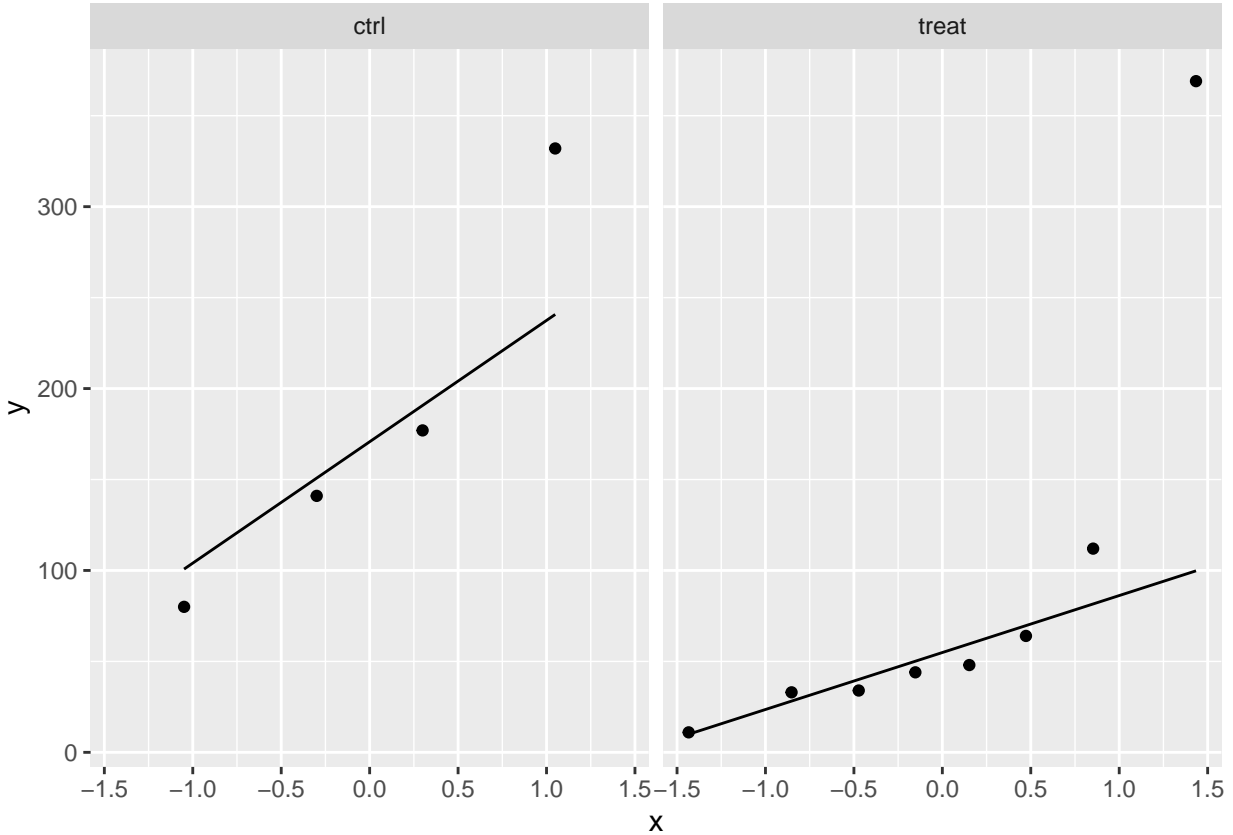
```
##   weight group
## 1     11 treat
## 2     33 treat
## 3     48 treat
## 4     34 treat
## 5    112 treat
## 6    369 treat
## 7     64 treat
## 8     44 treat
## 9    177  ctrl
## 10     80  ctrl
## 11    141  ctrl
## 12    332  ctrl
```

2.1 Data exploratie

```
rodents %>%
  ggplot(aes(x=group,y=weight)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter()
```



```
rodents %>%  
  ggplot(aes(sample = weight)) +  
    geom_qq() +  
    geom_qq_line() +  
    facet_wrap(~ group)
```



In the data exploration we do not have enough data to evaluate the assumptions.

Suppose that the assumptions are valid and that standard deviation in the population would be equal to the ones you observed in the experiment.

1. What is the power of the experiment if the effect size and standard deviation in the population would be equal to the ones you observed in the experiment
2. What would the power be if number of rodents would be balanced in both groups
3. How many observations would you need to pick up the treatment effect with a power of 80%?
4. How many observations would you need to pick up the treatment effect of 60 g/kg with a power of 80%?

3 Analyse of the data

We will model the data using a linear model with one dummy variable.

$$y_i = \beta_0 + \beta_1 x_{t,i} + \epsilon_i$$

with $x_{p,i} = 0$ if the rodent is subjected the control treatment with sterile apples and $x_{t,i} = 1$ if rodent receives the treatment with infected apples.

- Estimated effect size?

The average difference amount of apples eaten in g/kg body weight between infected and sterile apples is:

$$\hat{\beta}_1 = \bar{y}_t - \bar{y}_c$$

- H_0 : rodents eat consume on average the same amount of apples per kg body weight when they are fed with sterile or with infected apples.
- H_1 : the average amount of apples in g/kg body weight is different when rodents are fed with sterile then as when they are fed with infected apples.

```
lm1 <- lm(weight ~ group, rodents)
summary(lm1)

##
## Call:
## lm(formula = weight ~ group, data = rodents)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -102.500  -55.625  -41.438    1.531   279.625
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   182.50      57.03   3.200  0.00949 **
## grouptreat    -93.12      69.85  -1.333  0.21204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 114.1 on 10 degrees of freedom
## Multiple R-squared:  0.1509, Adjusted R-squared:  0.06601
## F-statistic: 1.777 on 1 and 10 DF,  p-value: 0.212
```

With the current study and under when we assume that the assumptions of the model hold, we conclude that the amount of apples that rodents on average consume does not differ significantly between the group that was fed with sterile apples and the group that was fed with infected apples.

4 Power of the test to detect the same effect size as observed in our dataset with our experimental design?

4.1 Simulation function

Function to simulate data similar to that of our experiment under our model assumptions.

```
simFast <- function(form, data, betas, sd, contrasts, alpha = .05, nSim = 10000)
{
  ySim <- rnorm(nrow(data)*nSim, sd=sd)
  dim(ySim) <- c(nrow(data), nSim)
  design <- model.matrix(form, data)
  ySim <- ySim + c(design %*% betas)
  ySim <- t(ySim)
```

```

### Fitting
fitAll <- limma::lmFit(ySim, design)

### Inference
varUnscaled <- c(t(contrasts)%*%fitAll$cov.coefficients)%*%contrasts)
contrasts <- fitAll$coefficients %*%contrasts
seContrasts <- varUnscaled^.5*fitAll$sigma
tstats <- contrasts/seContrasts
pvals <- pt(abs(tstats), fitAll$df.residual, lower.tail = FALSE)*2
return(mean(pvals < alpha))
}

```

4.2 Simulation

```

betas <- lm1$coefficients
nSim <- 10000
form <- ~ group
sd <- sigma(lm1)
contrast <- limma::makeContrasts("grouptreat", levels = names(lm1$coefficients))

## Warning in limma::makeContrasts("grouptreat", levels = names(lm1$coefficients)):
## Renaming (Intercept) to Intercept

contrast

##           Contrasts
## Levels      grouptreat
## Intercept           0
## grouptreat           1

#contrast <- matrix(c(0,1), ncol=1)
#rownames(contrast) <- names(mod1$coefficients)

alpha <- 0.05
power <- simFast(form, rodents, betas, sd, contrasts = contrast, alpha = alpha, nSim = nSim)
power

## [1] 0.2239

```

We observe that the experiment is severely underpowered. We only have a power of 22.4% to pick up the treatment effect.

5 Power for a balanced design

```

betas <- lm1$coefficients
nSim <- 10000
form <- ~ group
sd <- sigma(lm1)
contrast <- limma::makeContrasts("grouptreat", levels = names(lm1$coefficients))

```

```
## Warning in limma::makeContrasts("grouptreat", levels = names(lm1$coefficients)):
## Renaming (Intercept) to Intercept
```

```
n1 <- n2 <- nrow(rodents)/2
predictorData <- data.frame(group = rep(c("ctrl", "treat"), c(n1, n2))) %>% as.factor()
powerBalanced <- simFast(form, predictorData, betas, sd, contrasts = contrast, alpha = alpha, nSim = nS
powerBalanced
```

```
## [1] 0.2455
```

We observe that the power is larger for the balanced design. We could also have known this from formula of the standard error from the two-sample t-test.

$$SE = \hat{\sigma} \sqrt{1/n1 + 1/n2}$$

Indeed,

```
sqrt(1/sum(rodents$group=="treat") + 1/sum(rodents$group=="ctrl"))
```

```
## [1] 0.6123724
```

```
sqrt(1/n1 + 1/n1)
```

```
## [1] 0.5773503
```

So the SE is larger when the design is not balanced.

6 Required sample size to obtain a power of 90 %?

```
set.seed(1400)
betas <- lm1$coefficients
nSim <- 10000
form <- ~ group
sd <- sigma(lm1)
alpha <- 0.05
contrast <- limma::makeContrasts("grouptreat", levels = names(lm1$coefficients))
```

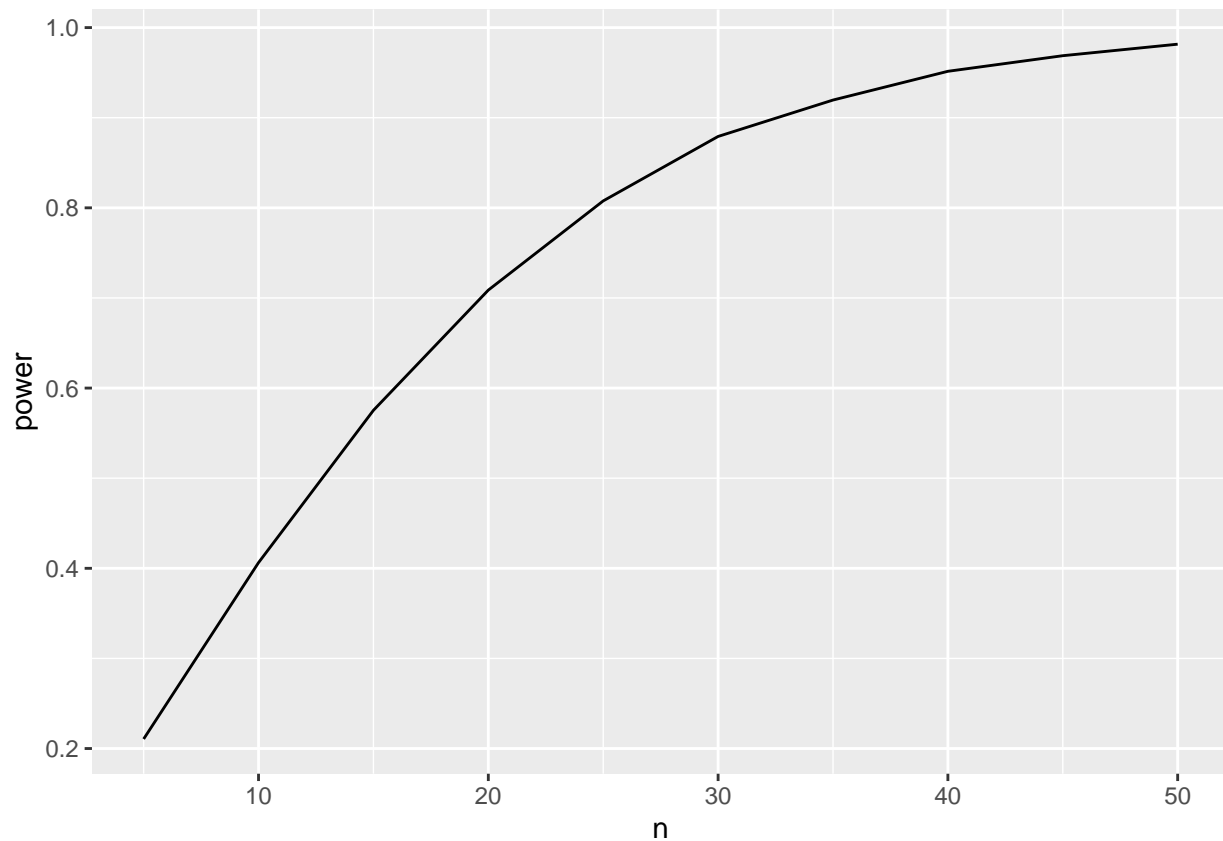
```
## Warning in limma::makeContrasts("grouptreat", levels = names(lm1$coefficients)):
## Renaming (Intercept) to Intercept
```

```
power <- data.frame(n=seq(5, 50, 5), power=NA)

for (i in 1:nrow(power))
{
  n1 <- n2 <- power$n[i]
  predictorData <- data.frame(group = rep(c("ctrl", "treat"), c(n1, n2))) %>% as.factor()
  power$power[i] <- simFast(form, predictorData, betas, sd, contrasts = contrast, alpha = alpha, nSim =
}
power
```

```
##      n power
## 1     5 0.2106
## 2    10 0.4063
## 3    15 0.5753
## 4    20 0.7087
## 5    25 0.8077
## 6    30 0.8792
## 7    35 0.9196
## 8    40 0.9515
## 9    45 0.9688
## 10   50 0.9816
```

```
power %>%
  ggplot(aes(x=n,y=power)) +
  geom_line()
```



Through simulations we show that we need about 32-33 observations to obtain a power of about 90%.

This is similar to what we would obtain with the close form formula that can be applied for a two group design

```
power.t.test(delta = lm1$coef[2], sd = sigma(lm1),power=.9)
```

```
##
##      Two-sample t test power calculation
##
```



```
##          n = 32.52035
##          delta = 93.125
##          sd = 114.067
##          sig.level = 0.05
##          power = 0.9
##          alternative = two.sided
##
## NOTE: n is number in each group
```

7 Impact of effect size

Suppose that we would like to pick up an effect size of $\beta_1 = 60g/kg$. how many samples would be required in each group to obtain a power of 90%? Note, that

- we do a two-sided test so the sign of the effect size is arbitrary.
- the intercept in the power analysis is also arbitrary so we could also set it at 0.

```
set.seed(1400)
betas <- c(0,60)
nSim <- 10000
form <- ~ group
sd <- sigma(lm1)
power2 <- data.frame(n=seq(5,100,5),power=NA)
alpha <- 0.05
contrast <- limma::makeContrasts("grouptreat",levels = names(lm1$coefficients))
```

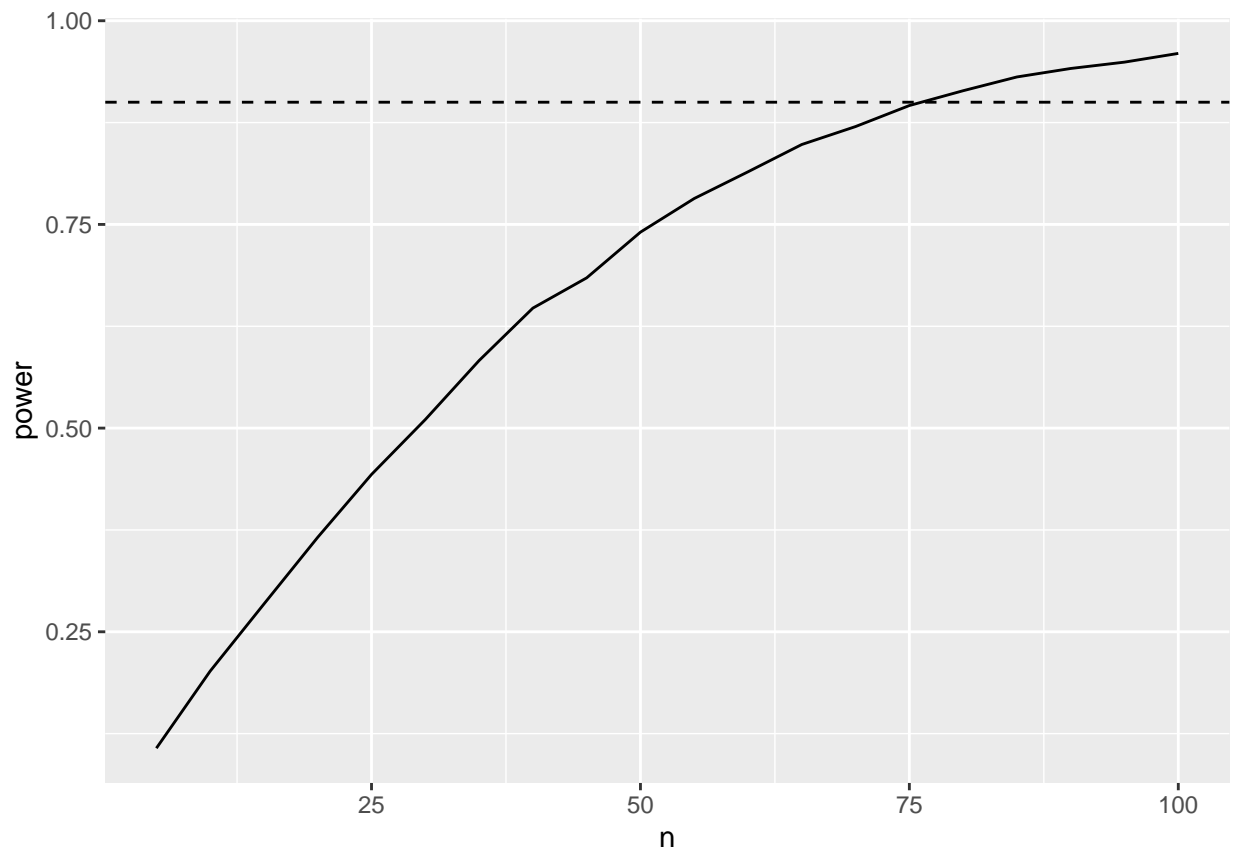
```
## Warning in limma::makeContrasts("grouptreat", levels = names(lm1$coefficients)):
## Renaming (Intercept) to Intercept
```

```
for (i in 1:nrow(power2))
{
  n1 <- n2 <- power2$n[i]
  predictorData <- data.frame(group = rep(c("ctrl","treat"),c(n1,n2)) %>% as.factor)
  power2$power[i] <- simFast(form, predictorData, betas, sd, contrasts = contrast, alpha = alpha, nSim = nSim)
}
power2
```

```
##      n  power
## 1    5 0.1071
## 2   10 0.2017
## 3   15 0.2841
## 4   20 0.3660
## 5   25 0.4431
## 6   30 0.5105
## 7   35 0.5829
## 8   40 0.6473
## 9   45 0.6844
## 10  50 0.7405
## 11  55 0.7818
## 12  60 0.8146
## 13  65 0.8481
```

```
## 14 70 0.8701
## 15 75 0.8959
## 16 80 0.9141
## 17 85 0.9310
## 18 90 0.9414
## 19 95 0.9492
## 20 100 0.9599
```

```
power2 %>%
  ggplot(aes(x=n,y=power)) +
  geom_line() +
  geom_hline(yintercept = .9, lty=2)
```



We observe that we need between 75-80 observations to obtain a power of 90%.

This is confirmed with the power functions for the two sample t-test.

```
b1 = - 60
power = .9
power.t.test(d = b1, sd = sigma(lm1), type='two.sample', power = power)
```

```
##
##      Two-sample t test power calculation
##
##              n = 76.926
##              delta = 60
```

```
##          sd = 114.067
##      sig.level = 0.05
##          power = 0.9
##      alternative = two.sided
##
## NOTE: n is number in each group
```

Note, that we would require a much larger sample size. This is because the desired effect size that we would like to pick up is small compared to the variability (standard deviation) in the population.