Experimental Design II: replication and power exercise 2

Lieven Clement & Alexandre Segers

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

Contents

```
library(tidyverse)
```

1 Puromycin data

Data on the velocity of an enzymatic reaction were obtained by Treloar (1974).

The number of counts per minute of radioactive product from the reaction was measured as a function of substrate concentration in parts per million (ppm) and from these counts the initial rate (or velocity) of the reaction was calculated (counts/min/min). The experiment was conducted once with the enzyme treated with Puromycin, and once with the enzyme untreated.

Here, we will focus again on the enzyme treated data.

```
data(Puromycin)
Puromycin <- Puromycin %>%
  filter(state=="treated")
```

There was a linear association between the log10 substrate concentration and the reaction rate

```
Puromycin %>%
  ggplot(aes(x=conc %>% log10,y=rate)) +
  geom_point() +
  stat_smooth(method = "loess",col="red") +
  stat_smooth(method='lm',col="black") +
  ylab("Reaction Rate (counts/min)") +
  xlab("log10(Substrate concentration) (log10 ppm)")
```
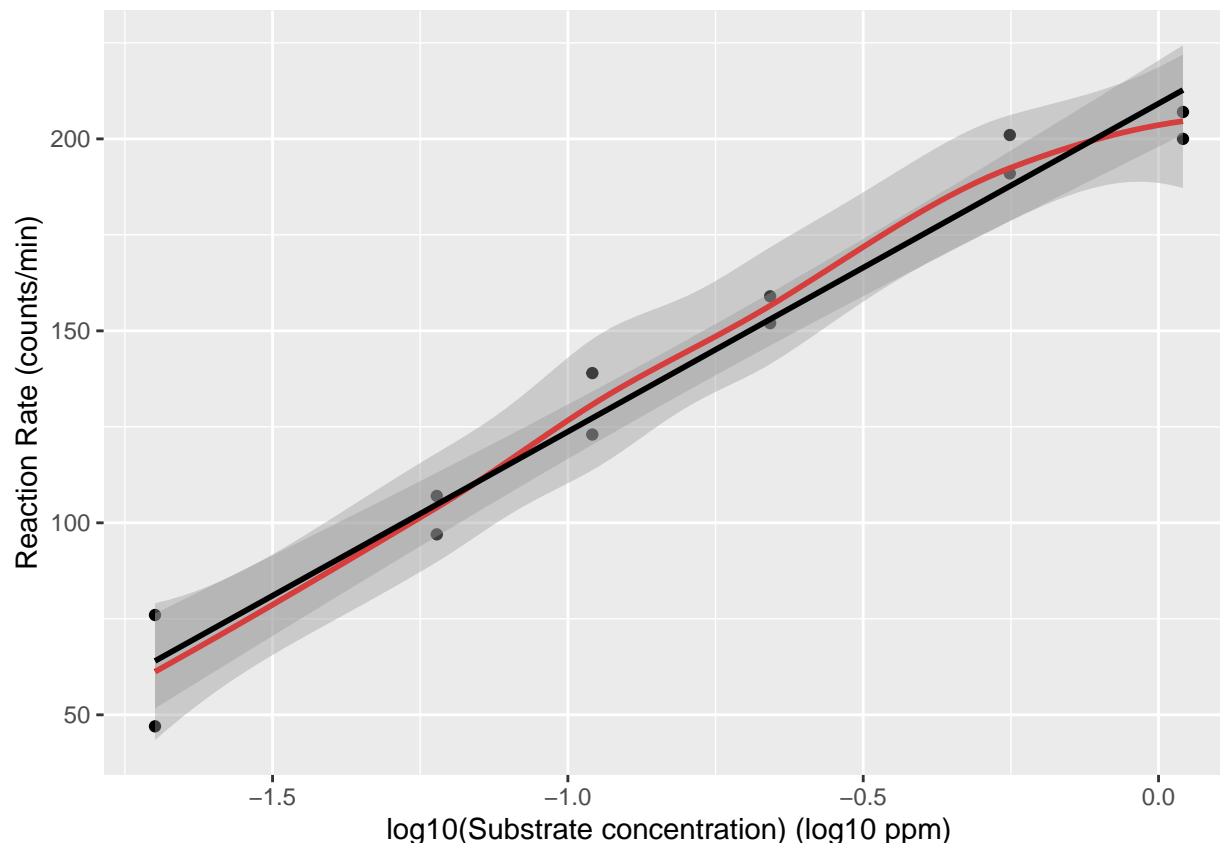
Note, that the researchers have chosen 6 different substrate concentrations and conducted an experiment where they assessed the initial reaction rate twice for every concentration.

1. Use the data to calculate the power to pick up an association that is as least as strong as the association you observed in the dataset when using an experiment with the same design.

2. Use the data to calculate the power to pick up an association where the reaction rate increases on average with 10 counts/min when the substrate concentration is 10 times higher ($\beta_1 = 10$).

3. Use the data to calculate the number of repeats you need for each concentration to pick up an association where the reaction rate increases on average with 10 counts/min when the substrate concentration is 10 times higher with a power of at least 90%. ($\beta_1 = 10$)

4. Suppose that you would setup an experiment with a design similar with the same concentrations as in the puromycin dataset and you have the following restriction: you need to use each concentration at least once and can setup at most 12 reactions, how would you choose your design points? Calculate the power for this design when the effect size is 10 counts/min per 10 times increase in the substrate concentration ($\beta_1 = 10$).

1.1 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))
}
```

1.2 Power to pick up the same effect size as we observed in the data set with the same design

```
mod1 <- lm(rate ~ conc %>% log10, Puromycin)
betas <- mod1$coefficients
nSim <- 10000
form <- ~ conc %>% log10
sd <- sigma(mod1)
contrast <- matrix(c(0,1),ncol=1)
rownames(contrast) <- names(mod1$coefficients)
alpha <- 0.05

power <- simFast(form, Puromycin, betas, sd, contrasts = contrast, alpha = alpha, nSim = nSim)
power
```

```
[1] 1
```

1.3 Power for $\beta_1 = 10$

```
mod1 <- lm(rate ~ conc %>% log10, Puromycin)
betas <- mod1$coefficients
betas[2] <- 10
nSim <- 10000
form <- ~ conc %>% log10
sd <- sigma(mod1)
contrast <- matrix(c(0,1),ncol=1)
rownames(contrast) <- names(mod1$coefficients)
alpha <- 0.05

power <- simFast(form, Puromycin, betas, sd, contrasts = contrast, alpha = alpha, nSim = nSim)
power
```

```
[1] 0.4167
```

The power to pick up a slope of $\beta_1 = 10$ for this experiment is only 41.7%.

1.4 Calculate the number of repeats needed per concentration to obtain a power of 90% to pick up an effect of $\beta_1 = 10$.

```
mod1 <- lm(rate ~ conc %>% log10, Puromycin)
concentrations <- Puromycin %>%
  pull(conc) %>%
  unique


betas <- mod1$coefficients
betas[2] <- 10
nSim <- 10000
form <- ~ conc %>% log10
sd <- sigma(mod1)
contrast <- matrix(c(0,1),ncol=1)
rownames(contrast) <- names(mod1$coefficients)
alpha <- 0.05
powers <- data.frame(n = 1:10, power=NA)
for (i in 1:10)
{
  simData <- data.frame(conc = rep(concentrations,each=i))
  powers[i,2] <- simFast(form, simData, betas, sd, contrasts = contrast, alpha = alpha, nSim = nSim)
}
powers %>%
  ggplot(aes(n,power)) +
  geom_line() +
  geom_hline(yintercept = .9, lty = 2)
```
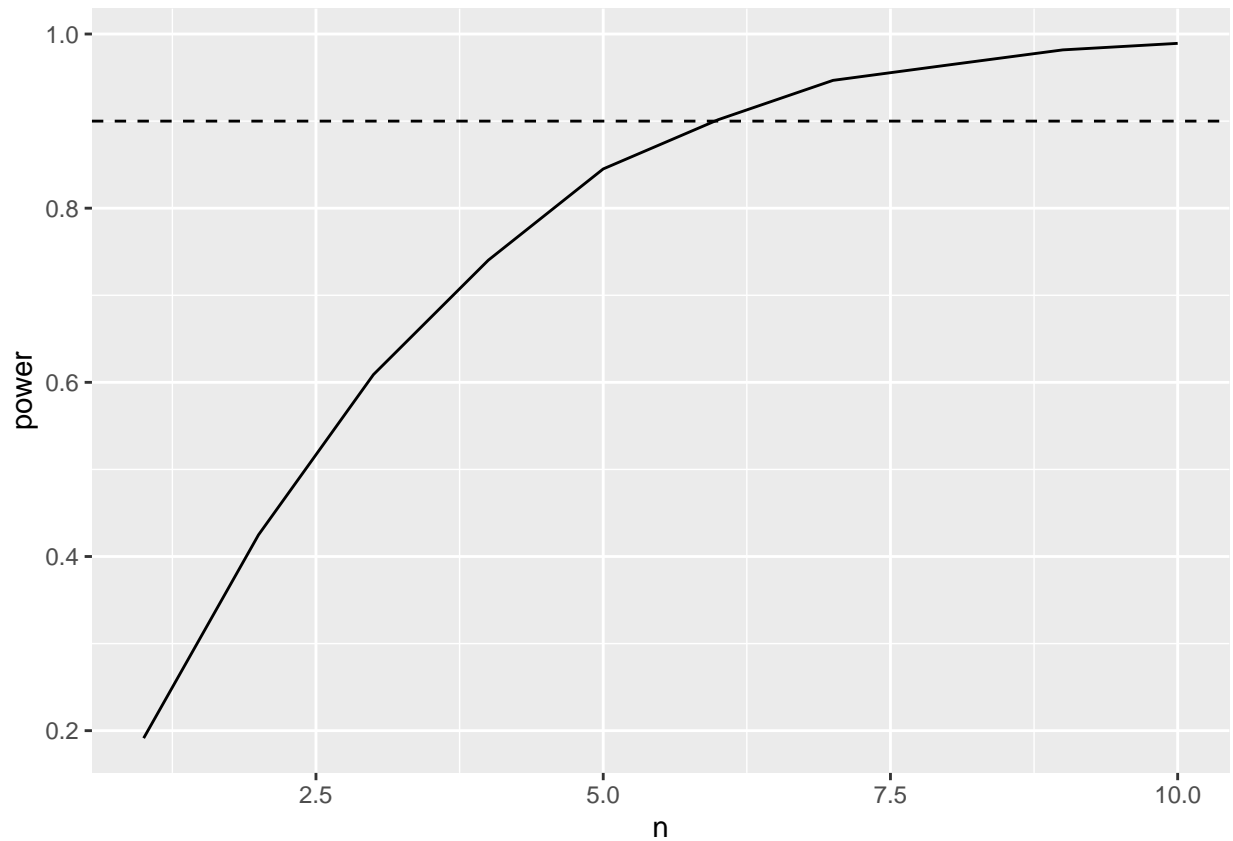
We need 6 repeats for each concentration to obtain a power above 90%.

2 Optimal design with 12 reactions

```r
concentrations <- Puromycin %>%
  pull(conc) %>%
  unique

betas <- mod1$coefficients
betas[2] <- 10
nSim <- 10000
form <- ~ conc %>% log10
sd <- sigma(mod1)
contrast <- matrix(c(0,1),ncol=1)
rownames(contrast) <- names(mod1$coefficients)
alpha <- 0.05

simData <- data.frame(conc = c(concentrations,rep(min(concentrations),3),rep(max(concentrations),3)))
simData
```

```
    conc
1   0.02
2   0.06
```

```
 3  0.11
 4  0.22
 5  0.56
 6  1.10
 7  0.02
 8  0.02
 9  0.02
10  1.10
11  1.10
12  1.10
```

```
powerOpt <- simFast(form, simData, betas, sd, contrasts = contrast, alpha = alpha, nSim = nSim)
powerOpt
```
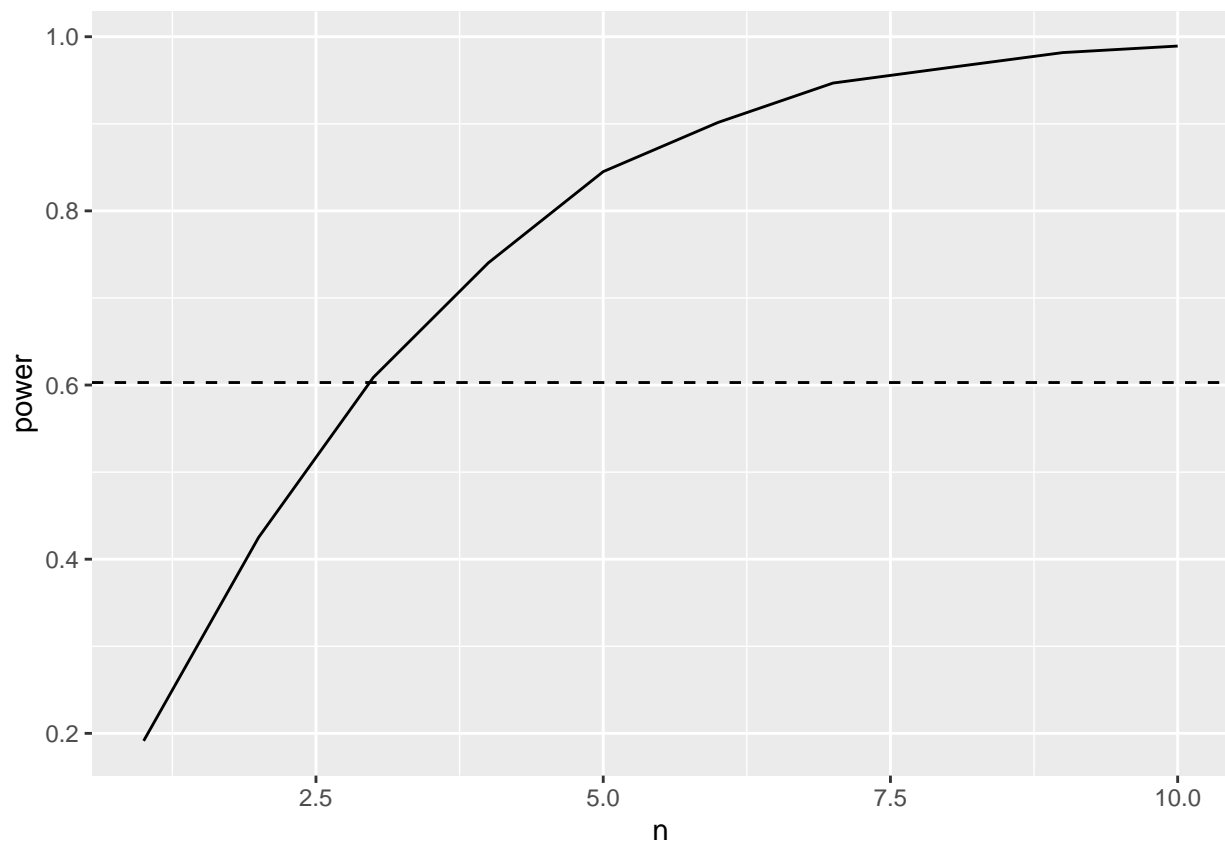
```
[1] 0.6029
```

Note that the power for a design where we repeat each concentration 1 time and the minimum and maximum concentration 4 times is considerably higher than that for the designs where we repeat all data points.

```
powers %>%
  ggplot(aes(n,power)) +
  geom_line() +
  geom_hline(yintercept = powerOpt, lty = 2)
```


Indeed, the power for our optimal design with 12 reactions is as high as the power for an experiment where you would repeat every concentration 3 times for which we need to conduct 18 reactions!