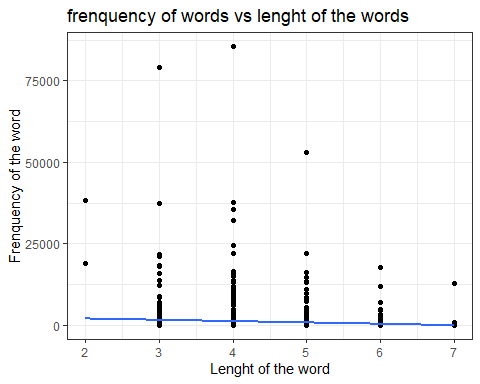
Homework #11

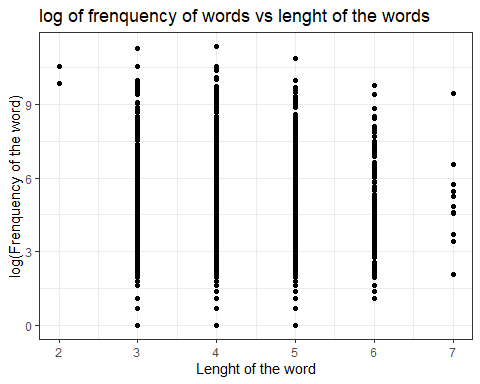
Henrique Magalhaes Rio

11/19/2019

# Question 1

## Part 1 (a)

 ## Part 2 (b)



## Part 3 (c)

=5.65479 =-0.146

## Part 4 (d)

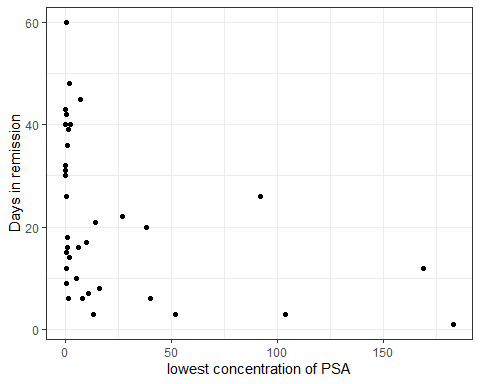
A difference of 1 percentage point in the lenght of the words is associated with a negative difference of 0.146 (95% CI: -0.236,-0.0562) in the estimated average of the logarithm of word frenquency.

## Part 4 (e)

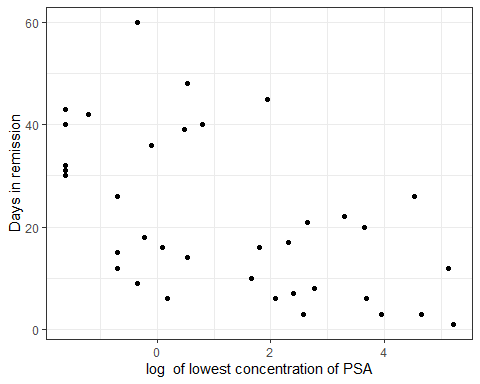
A difference of 1 percentage point in the lenght of the words is associated with a multiplicative difference of -0.146 (95% CI: -0.236,-0.0562) in the geometric mean of word frequency.

# Question 2

## Part 1 (a)



## Part 2 (b)



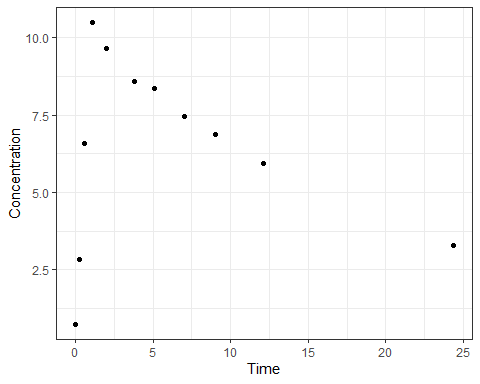
## Part 3 (c)

-3.93 is the difference in time spent in remession after treatment for a e-fold difference in the lowest concentration of PSA

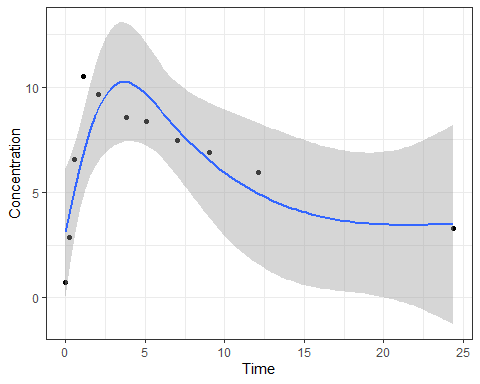
# Question 3

## Part 1 (a)

## Part 2 (b)



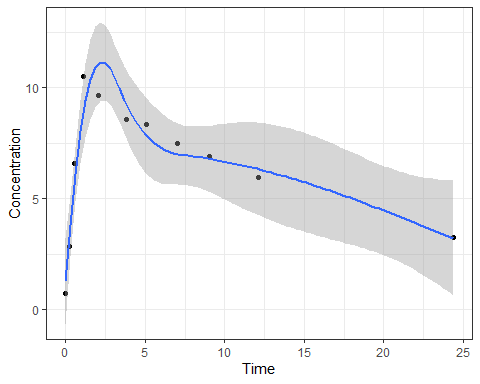
## Part 3 (c)



## Part 4 (d)

=3.10 =-3.29 =7.97 =-3.94

## Part 5 (e)



## Part 6 (f)

=1.28 =5.86 =0.821 =14.1 =-3.93

## Part 7 (g)

based on the the model with 4 splines (=0.927) is a lot better then the model with 3 splines (=0.6895)

# Appendix

# Insert packages you need here  
library(knitr)  
library(ggplot2)  
library(dplyr)  
library(tidyverse)  
library(broom)  
library(splines)  
  
  
words <-read.csv("english\_words.csv")  
  
  
ggplot(words,aes(y=WrittenFrequency,x=LengthInLetters)) +theme\_bw()+ geom\_point() +geom\_smooth(method=lm, se=FALSE)+ggtitle("frenquency of words vs lenght of the words")+xlab("Lenght of the word")+ylab("Frenquency of the word")  
wordslm <- lm(WrittenFrequency~LengthInLetters, data =words)  
  
  
  
ggplot(words,aes(y=log(WrittenFrequency),x=LengthInLetters)) +theme\_bw()+ geom\_point() +ggtitle("log of frenquency of words vs lenght of the words")+xlab("Lenght of the word")+ylab("log(Frenquency of the word)")  
  
  
  
  
wordslm2 <- lm(log(WrittenFrequency)~LengthInLetters, data =words)  
  
psa <-read.csv("psa\_complete.csv")  
  
  
ggplot(psa,aes(y=obstime,x=nadirpsa)) +theme\_bw()+ geom\_point() + ylab("Days in remission")+ xlab("lowest concentration of PSA")  
  
  
ggplot(psa,aes(y=obstime,x=log(nadirpsa))) +theme\_bw()+ geom\_point()+ ylab("Days in remission")+ xlab("log of lowest concentration of PSA")  
  
  
psalm <- lm(obstime~log(nadirpsa),data=psa)  
  
  
subject1 <- subset(Theoph, Subject==1)  
ggplot(subject1,aes(y=conc,x=Time)) +theme\_bw()+ geom\_point()+ylab("Concentration")  
ggplot(subject1,aes(y=conc,x=Time)) +theme\_bw()+ geom\_point() +ylab("Concentration")+ geom\_smooth(aes(x=Time, y=conc),  
data=subject1,  
method="lm",  
formula=y~ns(x, df=3))  
  
  
sublm <- lm(conc~ns(Time,3), data = subject1)  
ggplot(subject1,aes(y=conc,x=Time)) +theme\_bw()+ylab("Concentration") +geom\_point() + geom\_smooth(aes(x=Time, y=conc),  
data=subject1,  
method="lm",  
formula=y~ns(x, df=4))  
  
  
sublm2 <- lm(conc~ns(Time,4), data = subject1)