Skincells Data

Introduction

The Skincells Data is consisted of 118 observations collected in 4 different days.

The central question for this analysis is to consider any evidence of a differential effect of time exposed to radiation on number of cells in a culture. But, since observations were made on different days, with different conditions and amount of time samples, we want to take a look at how different are those days data collection. Table 2 below shows us that day 1 has more samples for 0.0 and 0.5 times than the other days and less samples for 3.0 and 3.5

Table 1: Quantity of Samples per time, per day:

day	time	n
1	0.0	5
1	0.5	4
1	1.0	4
1	1.5	3
1	2.0	3
1	2.5	5
1	3.0	3
1	3.5	2

day	time	n
2	0.0	3
2	0.5	4
2	1.0	4
2	1.5	5
2	2.0	4
2	2.5	2
2	3.0	4
2	3.5	3

day	time	n
4	0.0	4
4	0.5	3
4	1.0	4
4	1.5	5
4	2.0	3
4	2.5	3
4	3.0	5
4	3.5	4

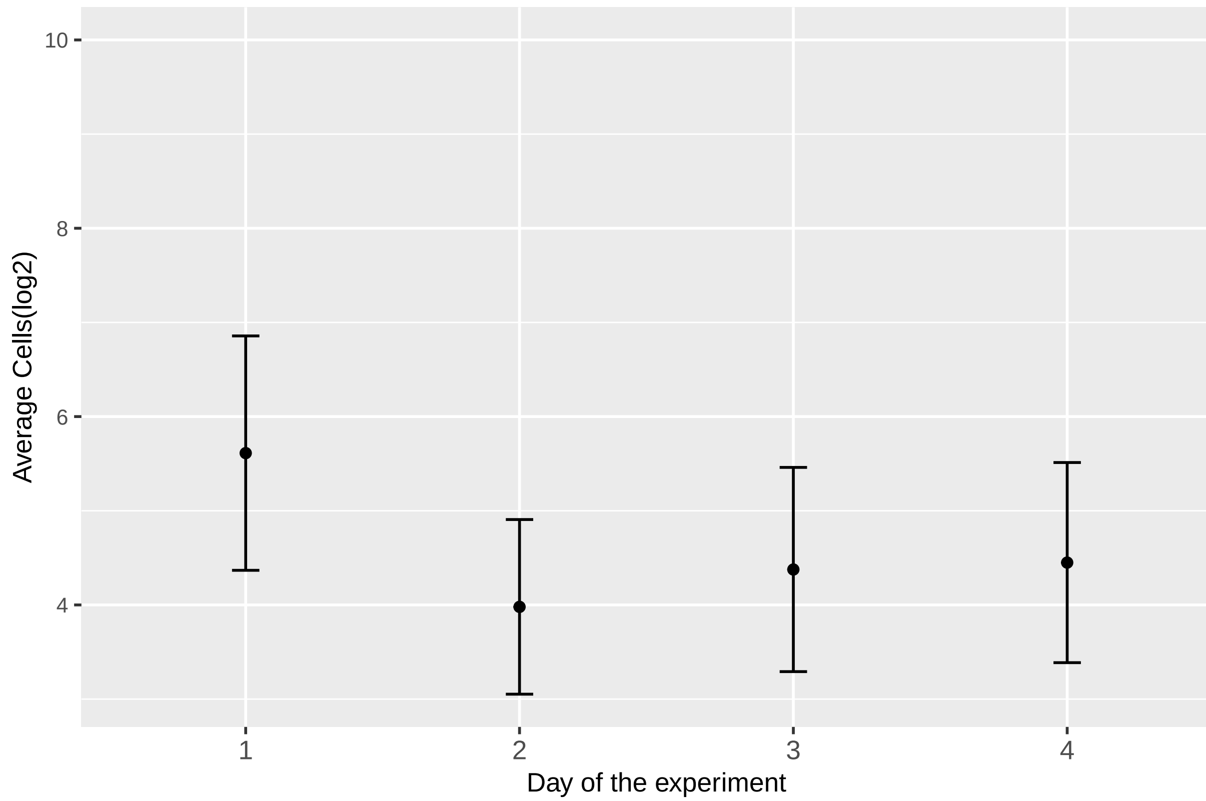
day	time	n
3	0.0	5
3	0.5	3
3	1.0	3
3	1.5	4
3	2.0	2
3	2.5	4
3	3.0	5
3	3.5	3

We also would like to know if there is a difference in the number of samples collected on each day, we noticed that only day 4 has 2 samples (31 total) more than the other days (29 samples).

Finally, we want to take a look into how many cells were alive on average each day to see if there is a difference.

We can see this difference on Plot 1. Day 1 had an average of cells higher than the other days and day 2 slightly lower. We needed to make more tests though to access if it's statistically relevat

Plot 1: Average number of cells by time exposed to radiation and its 95% CIs



Model 1

A linear regression model was fitted to these data with time as a continuous predictor and day as a categorical predictor. To allow maximum flexibility in the initial model, an interaction of time and day was included. Here we noticed straight away that time exposed to radiation seems to have a direct influence on the decrease of the number of cells with a p-value of 3.17e-05 and decreasing the number by -1.7451 for each extra minute exposed to radiation.

The model is:

$$y_i = \beta_0+\beta_1(time_i)+\beta_2(\delta_{ib})+\beta_3(\delta_{ic})+\beta_4(\delta_{id})+\beta_5(\delta_{ib}\text{Ö}time_i)+\beta_6(\delta_{ic}\text{Ö}time_i)+\beta_7(\delta_{id}\text{Ö}time_i)+\varepsilon_i$$

where y_i is the number of cells on experiment i, $\delta_{ib} = 1$ if the experiment i was done on day 2 or zero it it was done other day (and similarly for δ_{ic} and d).

This model allows for different straight lines (differing in both intercept and slope) but, with a test of the null hypothesis H0 : $\beta_5 = \beta_6 = \beta_7 = 0$ (see R code) yielding a p-value of 0.72 we failed to reject the null hypothesis and, therefore, we can assume that a common slopes model is adequate for this data and we no longer need to analyse the current model.

Model 2

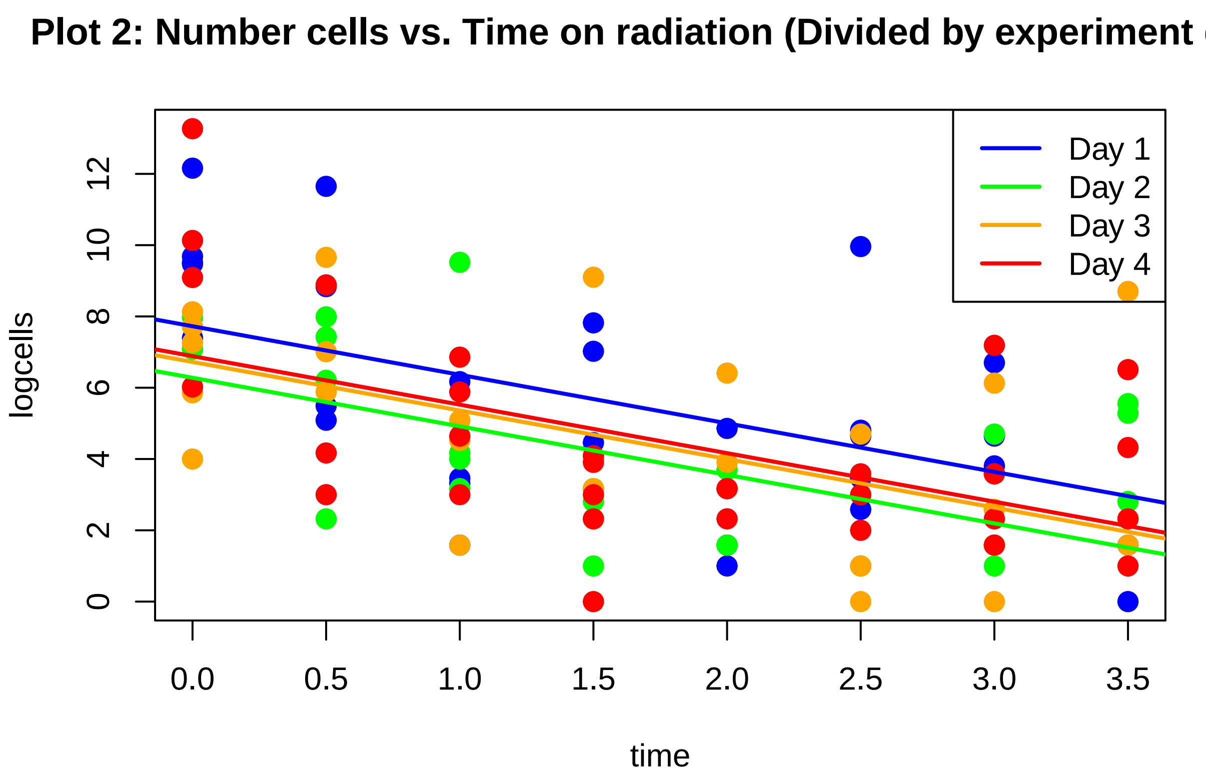
The following common slope model was then fitted to the data:

$$y_i = \beta_0 + \beta_1(time_i) + \beta_2(\delta_{ib}) + \beta_3(\delta_{ic}) + \beta_4(\delta_{id}) + \varepsilon_i$$

This leads to the following regression line equations for each day

$$day1 = \beta_0 + \beta_1 \text{Ö} time$$
$$day2 = (\beta_0 + \beta_2) + \beta_1 \text{Ö} time$$
$$day3 = (\beta_0 + \beta_3) + \beta_1 \text{Ö} time$$
$$day4 = (\beta_0 + \beta_4) + \beta_1 \text{Ö} time$$

This model allows for a common slope for the relationship between radiation time and number of cells for each day, but different intercepts. The effect of time was statistically significant yielding a p-value of 2.016e-10. The slope parameter for time was -1.36, suggesting that there was an average decrease in cells of -1.36 for each minute extra exposed to radiation. The effect of day still unclear since there is a difference between day 1 and 2 but not between other days. The current plotted looks like this Plot 2:



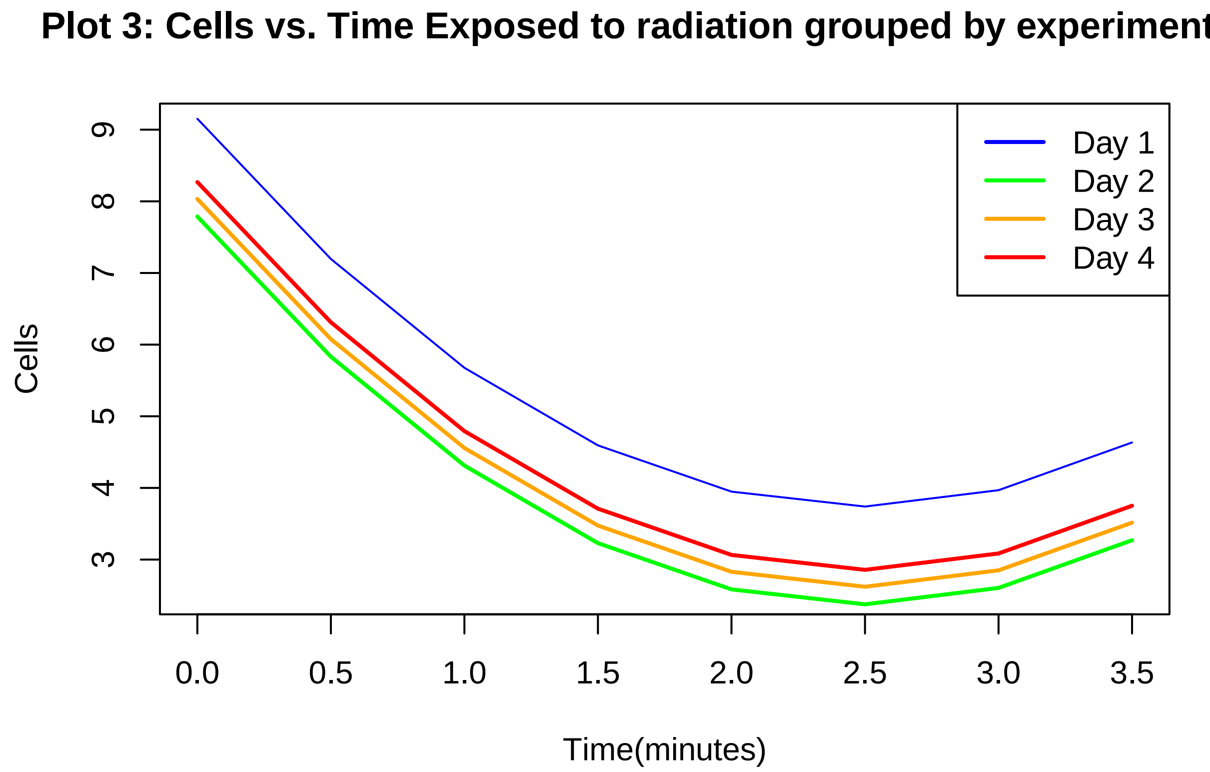
Model 3

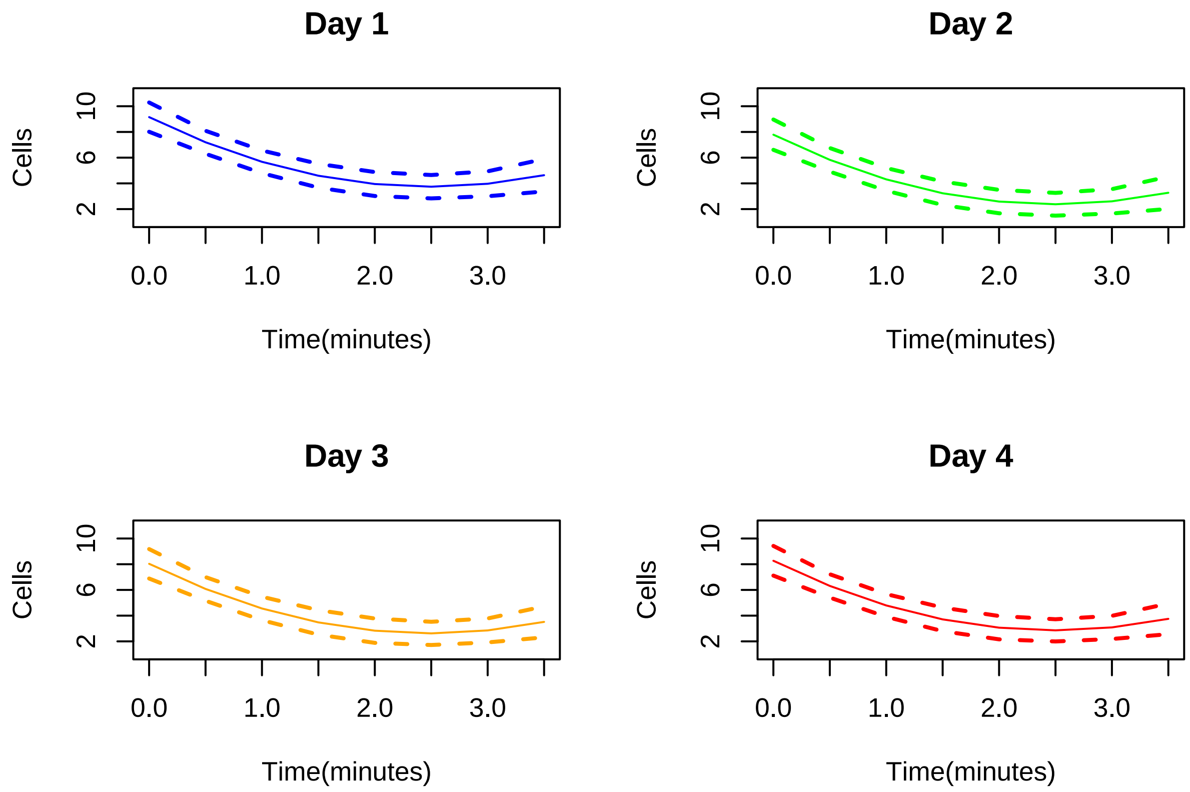
We could improve our model by adding a quadratic effect to time. Our model is the following now:

$$y_i = \beta_0 + \beta_1(time_i) + \beta_2(\delta_{ib}) + \beta_3(\delta_{ic}) + \beta_4(\delta_{id}) + \beta_5(time_i^2)\varepsilon_i$$

In this model we can notice that not only day 2 difference to day 1 is statistically relevant (p-value=0.02) as the day 3 difference to day 1 is now borderline (p-value=0.0565) but, the most interesting finding here is that after 2.5 minutes of exposure to radiation, cells seem to start increasing in number again as we can see on Plot 3 and Plot 4.

```
## [1] "blue"   "green"  "orange" "red"
```




Day 1
Day 2
Day 3
Day 4
Cells
Time(minutes)
0.0
1.0
2.0
3.0
2
6
10