

Skin Cells

Jerry Kiely

26 February 2017

Introduction

Researchers in the Radiation and Environmental Science Centre are conducting research into the effect of solar radiation on the mortality of human skin cells. Colonies of human skin cells are placed in medium and transferred to wells on experimental plates. These plates are then exposed to radiation in a solar simulator for various amounts of time from 0 (control) to 3.5 minutes. After exposure, the number of live cells in a sample are counted under a microscope and the total number of live cells in the colony is extrapolated from the sample result. The purpose of this experiment is to assess the effect of the varying times of exposure to radiation on cell death. Due to inherent variability in the response of cells, the researchers replicate their experiments a number of times. There is also concern that environmental conditions within the lab from day to day may have an impact on the results. Therefore, the experiment is repeated over a number of days. Initially the experiment is designed with complete balance; i.e. the same number of replications at each exposure time on each day. But due to experimental loss, complete balance was not reflected in the final data (e.g. some plates were contaminated and had to be discarded).

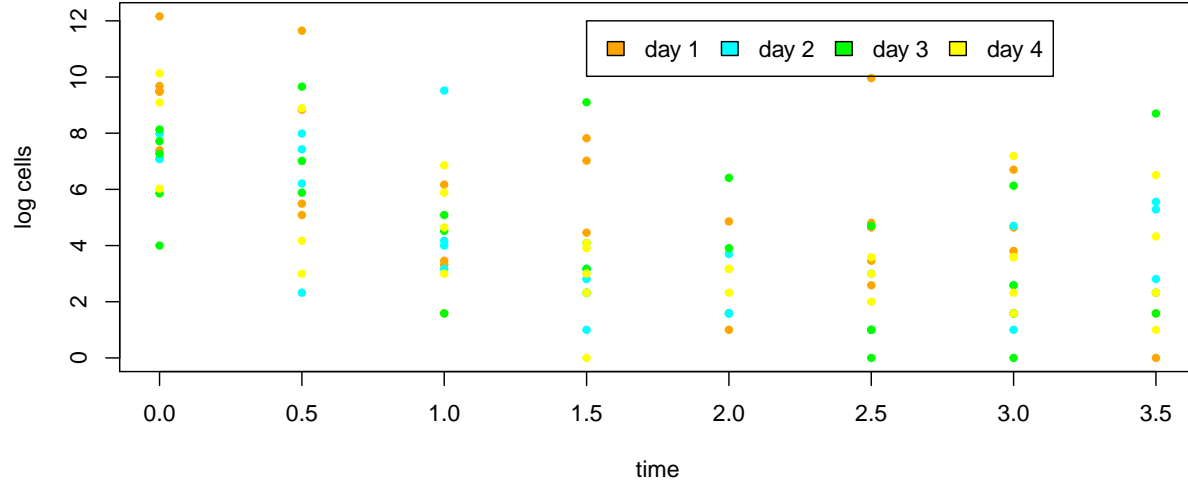
The dataset is `skincells.xls` and is available on Webcourses. The variables are:

Variable	Description
day	The day (number code) that observation was recorded.
time	The amount of radiation exposure in minutes the colony was exposed to in the solar simulator.
logcells	The logarithm (base 2) of the number of live cells in the colony extrapolated from the sample result under the microscope.

The central question is to establish whether the environmental conditions within the lab have an impact on the results. Below is a table of the means for each of the day, and the corresponding confidence intervals for each day:

Day	Mean	Lower 95%	Upper 95%
1	8.3201268	6.6060231	10.0342305
2	5.8858886	4.3942513	7.3775259
3	6.5115999	4.8604817	8.162718
4	6.8478244	5.1716733	8.5239755

the above data would suggest that there is not much difference between the means, with maybe the means on day 1 and day 2 having the greatest difference.



The above scatterplot of the data with days having different colours would seem to confirm this.

First Model

Initially, a linear model was fitted with time as a continuous predictor, and day as a categorical predictor, with an interaction allowed between day and time.

$$y_i = \beta_0 + \beta_1(time_i) + \beta_2(\delta_{i2}) + \beta_3(\delta_{i3}) + \beta_4(\delta_{i4}) + \beta_5(\delta_{i2} \times time_i) + \beta_6(\delta_{i3} \times time_i) + \beta_7(\delta_{i4} \times time_i) + \epsilon_i$$

The null hypothesis may be stated as follows:

$$\begin{aligned} H_0 : & \quad \beta_5 = \beta_6 = \beta_7 = 0 \\ H_a : & \quad \beta_5, \beta_6, \text{ and } \beta_7 \text{ are different to } 0 \end{aligned}$$

The F-test statistic was found to be 0.4456585 on 3 and 110 degrees of freedom, yielding a p-value of 0.7208564. Therefore we fail to reject the null hypothesis and assume no interaction between the predictors (i.e. that the common slopes model is adequate).

Second Model

Secondly, a linear model was fitted with time as a continuous predictor, and day as a categorical predictor, with no interaction between day and time.

$$y_i = \beta_0 + \beta_1(time_i) + \beta_2(\delta_{i2}) + \beta_3(\delta_{i3}) + \beta_4(\delta_{i4}) + \epsilon_i$$

The null hypothesis may be stated as follows:

$$H_0 : \quad \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_a : \quad \beta_2, \beta_3, \text{ and } \beta_4 \text{ are different to } 0$$

The F-test statistic was found to be 2.4559368 on 3 and 113 degrees of freedom, yielding a p-value of 0.0667058. Therefore, although close, we fail to reject the null hypothesis and assume no significant difference between the days.

Third Model

Thirdly, a linear model was fitted with time as a continuous predictor, and a quadratic effect of time.

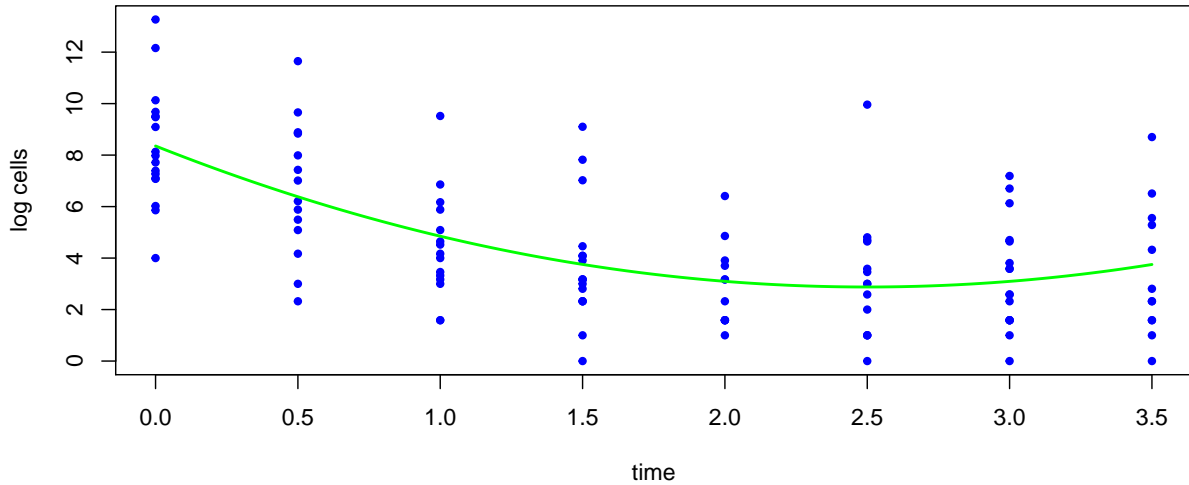
$$y_i = \beta_0 + \beta_1(time_i) + \beta_2(time_i^2) + \epsilon_i$$

The null hypothesis may be stated as follows:

$$H_0 : \quad \beta_2 = 0$$

$$H_a : \quad \beta_2 \text{ is different to } 0$$

The F-test statistic was found to be 23.2011782 on 1 and 115 degrees of freedom, yielding a very small p-value of 0.00000449. Therefore we reject the null hypothesis and assume a quadratic effect.



The final plot of the data would seem to validate the research.

Comparison between days

Although it is clear that day is not a predictor, there is grounds to exclude day 1 because the results on that day would seem to be significantly different from the results of the other days. Specifically, the differences between day2, day 3, and day 4 are not significant, but the difference between day 1 and day 2

may be statistically significant, with a p-value of 0.0248125. Therefore day 1 and day 2 may have significant differences, and perhaps day 1 should be excluded.

Below is a plot of the data excluding day 1, and with it's associated fitted line.

