Classroom Experiments

James M. Curran

Dept. of Statistics, University of Auckland

2^{nd} July 2015

j.curran@auckland.ac.nz
http://www.stat.auckland.ac.nz/~curran

After being given the task of setting how much money they would like to be given for reading through a 25 year-old set of notes to a half-full room of hungover teenagers, over two-thirds said they would like as much as was legally possible.

– thedailymash.co.uk, "Oh, go on then, say universities" published 2011-04-06

STATS 201/208 at the University of Auckland

The name of our course is **Data Analysis**

- A first introduction to data analysis
- A second year course for undergraduates
- Almost all participants will have completed a first year Statistics course
- But probably not any mathematics, or any computer science
- Taught three times per year
- Ostensibly tailored for Science and Business
- Each semester contains single class of 250+ students

STATS 201/208 at the University of Auckland

Our emphasis is primarily on the linear model, with extensions to generalized linear models, and a separate section on descriptive time series analysis.

Most of the modelling in the course can be summed up by

$$Y_i = \mathbf{X}\beta + \boldsymbol{\epsilon}$$
$$g(E[Y_i]) = \mathbf{X}\beta$$
$$\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \sigma^2\mathbf{I}) \equiv \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)^\dagger$$

Aside: We spend *forever* trying to stop the students making a fetish out of normality and equality of variance assumptions.

† for Gaussian family only obviously

Broad overview

- Regression
- Analysis of grouped data
- Analysis of categorical data
- Time Series

Aims and Issues

Aims

- Practice not theory
- Real data sets in class
- New real data sets in assignments, tests, exams
- *Not achieved yet:* No small data sets

Issues

- Students get pre-cleaned, pre-vetted data sets
- Almost all students have no experience in collecting data
- Similarly they have not (yet) encountered any situation where they need to analyze data

Teaching experimental design

1. We teach ANOVA as a method for analyzing grouped data
2. Only up to two-way ANOVA with interaction
3. No two-way without replication
4. ANOVA table as a hypothesis testing and reality checking tool
5. ANOVA identity $\textit{Total SS} = \textit{Model SS} + \textit{Residual SS}$

What really students think of formal experimental design

Can you blame them?

We need to get students involved in experimentation

Hardly novel: Box, Cobb, Montgomery and others

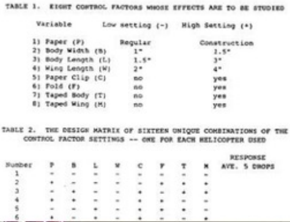


TABLE 1. EIGHT CONTROL FACTORS WHOSE EFFECTS ARE TO BE STUDIED

Variable	Low setting (-)	High Setting (+)
1) Paper (P)	Regular	Construction
2) Body Width (B)	1"	1.5"
3) Body Length (L)	1.5"	3"
4) Wing Length (W)	2"	4"
5) Paper Clip (C)	no	yes
6) Fold (F)	no	yes
7) Taped Body (T)	no	yes
8) Taped Wing (M)	no	yes

TABLE 2. THE DESIGN MATRIX OF SIXTEEN UNIQUE COMBINATIONS OF THE CONTROL FACTOR SETTINGS -- ONE FOR EACH HELICOPTER USED

Number	P	B	L	W	C	F	T	M	RESPONSE AVE. 5 DROPS
1	-	-	-	-	-	-	-	-	
2	+	-	-	-	-	+	+	+	
3	-	+	-	-	+	-	+	+	
4	+	+	-	-	+	+	-	-	
5	-	-	+	-	+	+	+	-	
6	+	-	+	-	+	-	-	+	

We need to get students involved in experimentation

We need an experiment that:

- Can be carried out by each student and they collect their **own** data
- Does not require any specialized equipment
- Takes a relatively short amount of time
- Not prone to failure
- Two factors, sufficient replication
- Easy to understand

Primary question: Which is faster?

Method A

```
methodA = function(listLength, sampleSize){
  l = vector(mode = "list", length = listLength)
  for(i in 1:listLength){
    l[[i]] = matrix(rnorm(sampleSize), nrow = 2)
  }
}
```

Method B

```
methodB = function(listLength, sampleSize){
  l = vector(mode = "list", length = listLength)
  lapply(l, function(x){x = matrix(rnorm(sampleSize),
                          nrow = 2)})
}
```

Other factors

- `listLength`, `sampleSize`
- **Learning point:** List length is confounded
- We held `sampleSize` constant (500)
- Allowed `listLength` to vary, $1 \times 10^4, 2 \times 10^4, \ldots, 10^5$
- 10 reps per treatment = 200 observations per student
- Random number seeds set to student ID number
- Execution time 10–15 minutes in computer lab

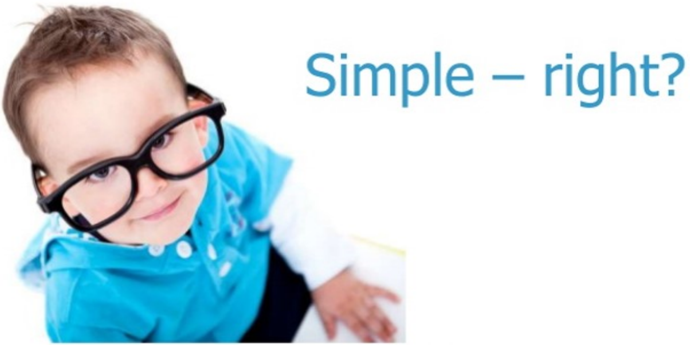
Model

$$time \sim method * listLength$$

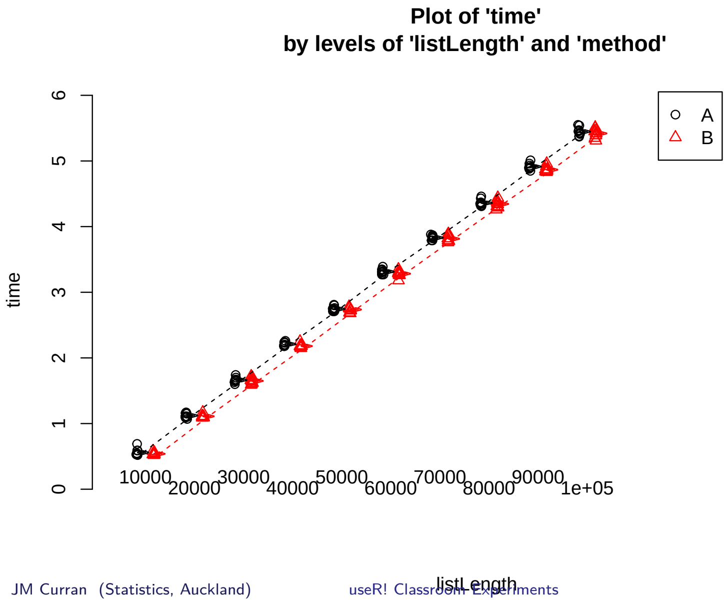
Instructions

- Students were provided with an R script that ran the experiment
- They only had to change the random number seed, and then run the script (copy and paste into console for most of them)
- They were instructed to **not do anything** while the script was running
- Script provided a progress bar, and at the end executed `file.choose()`

What could go wrong?



Interaction plot

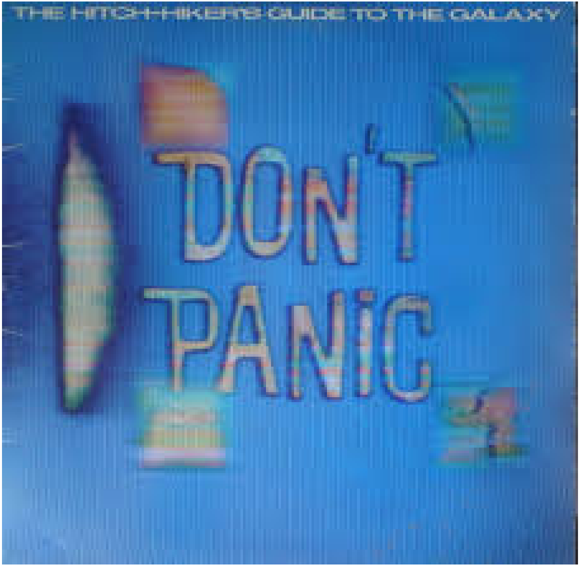


ANOVA table

```
Analysis of Variance Table

Response: time
                   Df Sum Sq Mean Sq    F value    Pr(>F)
method              1   0.03   0.025    14.1538 0.0002277 ***
listLength          9 483.44  53.716 30035.7964 < 2.2e-16 ***
method:listLength   9   0.01   0.001     0.3897 0.9389665
Residuals         180   0.32   0.002
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

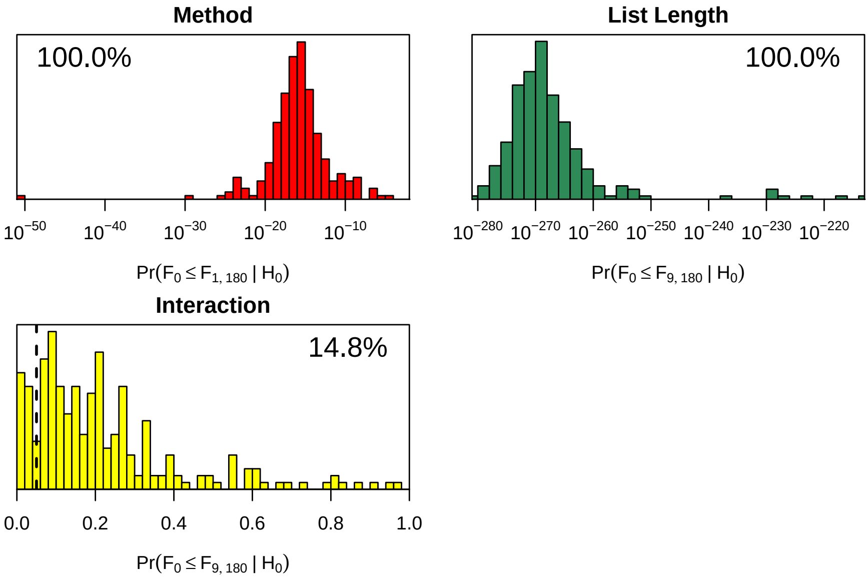
However rumours were coming back from lab tutors that students were finding non-significant method effects.


THE HITCH-HIKER'S GUIDE TO THE GALAXY
DON'T
PANIC

Marking time

- We asked one of our more skilled markers just to count up how many people (in the set of scripts that he had marked) found a significant `method` main effect.
- Out of 13 scripts (approx 5% of class), only 7
- Are my results a fluke?
- The students' results differ only by the random number stream. I have access to all the seeds, hence I can reproduce their analyses (under the naïve assumption they fitted the same model)
  - Rewrote script to accept ID number as a command line argument
  - Submitted 250 jobs to our cluster
  - collated 3 P-values from each analysis (two main effects + interaction)

Results


Method
100.0%
10^−50 10^−40 10^−30 10^−20 10^−10
Pr(F0 ≤ F1, 180 | H0)
List Length
100.0%
10^−280 10^−270 10^−260 10^−250 10^−240 10^−230 10^−220
Pr(F0 ≤ F9, 180 | H0)
Interaction
14.8%
0.0 0.2 0.4 0.6 0.8 1.0
Pr(F0 ≤ F9, 180 | H0)

Lessons learned

- It is possible to carry out large scale experimentation in class/assignments
- It is important to do this
- Do not trust people to follow instructions (obvious with the benefit of hindsight)
- This probably will not address lack of domain knowledge (although we might address this by discussing the experiment in advance in class)

Thanks

To consult a statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination: he may be able to say what the experiment died of. – R. A. Fisher