A web-based personal assistant for designing statistically sound experiments

Kathy Ruggiero and David Banks

Department of Statistics

9 July 2015
All too often…
Inevitably...

- Wrong data is collected
  - Cannot answer the questions for which they were intended

  OR

- Data is inefficiently collected
  - Comparisons between treatment means made with lower precision than may otherwise have been possible
The solution

“Use statistically designed experiments to ensure that the “right answers” are found with minimum effort, subjects and other resources.”

• Three fundamental principles (Fisher, 1926):
  – Replication, to enable separation of signal and noise
  – Randomisation, to eliminate bias and induce independence
  – Blocking, to control for systematic nuisance sources of variation
The problem

• Many software packages with experimental design capabilities but…
  – User must know *a priori* the “layout” of the design, e.g.
    ▪ Completely randomised design
    ▪ Randomized complete block design
    ▪ Balanced incomplete block design
    ▪ Row-column design
    ▪ Split-plot design
    ▪ Etc., etc., etc.!

• Large body of literature, largely inaccessible to novices
Our idea

“An expert system with the embedded requisite knowledge which interactively guides users through the thinking needed to develop an efficient statistically designed experiment.”

Conception\textsuperscript{C2D}: Concept 2 Design

Start

Objectives

Measurable responses

Experimental

Title of Experiment

Enter title

Your name (optional)
Conception\textsuperscript{C2D}: Concept 2 Design

Start \rightarrow Objectives

Measureable responses \rightarrow Experimental factors

Experimental material \rightarrow Research questions

Nuisance variables \rightarrow Generalisability

PDF Summary \rightarrow Finish

Design of Experiment

Kathy Ruggiero k.ruggiero@auckland.ac.nz
Example: Simulating a high $\text{CO}_2$ world

Experiment: Simulate a high $\text{CO}_2$ world

Identify/quantify metabolite, protein and gene transcript abundances

**Goal:** Elucidate molecular mechanisms involved in calcification
How were samples obtained?

Male:Female pair

Divide into 3 cultures

Until 4-arm stage

Control = 380 ppm (current level)
Mid = 540 ppm
High = 1000 ppm
How were samples obtained?

The sea urchin, *Evechinus chloroticus* (Kina) x 8 Male:Female pairs

Express and mix gametes from single male:female pair

Fertilized eggs

We will focus on proteomics and assume one sample is taken from each tank.
Conception C2D: Concept 2 Design

Title of Experiment

Molecular effects of high CO2 on calcification of sea urchin larvae

Your name (optional)

Kathy Ruggiero

Your institution's name (optional)

University of Auckland

Kathy Ruggiero k.ruggiero@auckland.ac.nz
The most important question?

Why am I collecting this data?

How will each variable you measure enable you to answer your research question(s)?

Kathy Ruggiero k.ruggiero@auckland.ac.nz
Why this experiment?

Why is this information being collected?
To direct you towards writing focused statements about the investigative questions you want your experiment to answer.

Objectives of your experiment

Specify each objective of your experiment in the boxes below. New boxes can be added by clicking the Add objective button.

1. Elucidate molecular mechanisms in sea urchin larvae

Add objective
What will you measure?

Why is this information being collected?
To direct you towards thinking about the variable(s) you will measure on the subjects of your experiment.

Measurable Responses

Enter the names of the variables that you will measure in your experiment and, if appropriate, their corresponding units of measurement. Enter N/A for numerical variables with unit of measurement unknown or not available.

1. Metabolite intensity  
2. Protein abundance  
3. Transcript abundance

Kathy Ruggiero k.ruggiero@auckland.ac.nz
Factors you will deliberately vary?

Why is this information being collected?
To direct you towards thinking about the experimental factors you will study in your experiment for the purpose of learning how they may help to explain observed changes in your measurable responses.

Conception\textsuperscript{C2D}: Concept 2 Design

Experimental Factors

How many \textit{experimental factors} will you investigate in your experiment?

1

An experimental factor refers to a variable (e.g. disease status) whose values (e.g. healthy and diabetic) will be varied in the experiment with the goal of understanding how changes in the factor's values explain differences in the responses in the measured variables.
Factors you will deliberately vary?

Enter the name and *levels* of each experimental factor you will investigate in your experiment. If any control(s) are required to assess the effects of a particular experimental factor(s) on the response variables, include the control(s) in the level name.

**Factor 1**

Name: CO2 level

**Levels**

- 380 ppm
- 540 ppm
- 1000 ppm

The levels of an experimental factor (e.g. disease status) are the discrete possible values (e.g. healthy, diabetic) it takes in an experiment.
Factors you will deliberately vary?

Each row of the following table represents one of the 6 experimental treatments, formed by combining one level from each factor you are electing to study in this experiment.

<table>
<thead>
<tr>
<th>Treatment #</th>
<th>CO2</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>380 ppm</td>
<td>21 degC</td>
</tr>
<tr>
<td>2</td>
<td>380 ppm</td>
<td>26 degC</td>
</tr>
<tr>
<td>3</td>
<td>540 ppm</td>
<td>21 degC</td>
</tr>
<tr>
<td>4</td>
<td>540 ppm</td>
<td>26 degC</td>
</tr>
<tr>
<td>5</td>
<td>1000 ppm</td>
<td>21 degC</td>
</tr>
<tr>
<td>6</td>
<td>1000 ppm</td>
<td>26 degC</td>
</tr>
</tbody>
</table>

Showing 1 to 6 of 6 entries
The subjects of your experiment?

Why is this information being collected?
To direct you towards thinking about:
1. Who or what are the subjects of your experiment?
2. How will you apply the experimental treatments to them?
3. How will they be managed for making measurements on them?

To which of the following will you apply the experimental factors in your study?

- Humans
- Animals
- Plants
- Cell cultures
- Soil cores

Other - please specify: sea urchin larvae

You have selected sea urchin larvae as the subjects of your study. How will each treatment in your study be applied to these subjects?
The subjects of your experiment?

You have selected sea urchin larvae as the subjects of your study. How will each treatment in your study be applied to these subjects?

- Each treatment will be independently applied to individual subjects
- Each treatment will be independently applied to groups of subjects

For each variable in your list of measurable responses, what will you make measurements on?

- Each individual subject, i.e. one measurement per subject per measurable response variable.
- The group of subjects, i.e. one measurement per group of subjects per measurable response variable.

The experimental unit in your experiment is a group of subjects.
The observational unit in your experiment is a group of subjects.
Other experimental variables?

Why is this information being collected?

To direct you towards thinking about:

1. The source of each portion of experimental material. I.e. is each derived from:
   a. an independent source?
   b. a common source?
   c. a combination of both independent and common sources?

2. How the experimental material will be:
   a. processed and/or
   b. arranged in time and/or space
   in order to make measurements on them.

3. How 1 and 2 may contribute to systematic differences in the observed values on your measurable variables.
Other experimental variables?

The *experimental unit* in your experiment is a group of subjects.

The *observational unit* in your experiment is a group of subjects.

Group of subjects = culture or tank
3 per parent-pair

Male:Female parent-pair

Common source: *A male:female pair*

Kathy Ruggiero k.ruggiero@auckland.ac.nz
Other experimental variables?

How will the experimental material will be:
  a. processed and/or
  b. arranged in time and/or space
in order to make measurements on them?

8 samples can be simultaneously analysed in a single mass
spec run (ignoring isobaric tags used to label samples)

Can each run accommodate a sample from each
male:female pair?

Yes
Other experimental variables?

How will the experimental material will be:

a. processed and/or

b. arranged in time and/or space

in order to make measurements on them?
Other experimental variables?

Cells with same fill pattern = Samples analysed in a run
Cells with same fill colour = Samples from common source

Every cell colour (M:F pair) occurs with every pattern (run)

Row-Column design
**Other experimental variables?**

Allocation of CO2 levels to row-column arrangement

- Full complement of CO2 levels in columns
- CO2 levels balanced within runs
- Biological variation orthogonal to run-to-run variation

<table>
<thead>
<tr>
<th>Male:Female Pair</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run 1</td>
<td>380</td>
<td>1000</td>
<td>540</td>
<td>380</td>
<td>1000</td>
<td>540</td>
<td>380</td>
<td>540</td>
</tr>
<tr>
<td>Run 2</td>
<td>540</td>
<td>380</td>
<td>1000</td>
<td>540</td>
<td>380</td>
<td>1000</td>
<td>540</td>
<td>1000</td>
</tr>
<tr>
<td>Run 3</td>
<td>1000</td>
<td>540</td>
<td>380</td>
<td>1000</td>
<td>540</td>
<td>380</td>
<td>1000</td>
<td>380</td>
</tr>
</tbody>
</table>
Our design versus theirs?

Design from Conception\textsuperscript{C2D}

- 3 runs
- 12 residual d.f.
- 98.4% efficiency
- CO2 levels estimated independently of tags

Their design

- 4 runs (extra $4500)
- 14 residual d.f.
- 100% efficiency
- CO2 levels partially confounded with tags
Output

• Summary of input (PDF)

Experiment summary

Experiment title: Molecular effects of high CO2 on sea urchin larvae

Experiment Title: Kathy Ruggiero

Experiment Title: University of Auckland

Objectives

Enter the scientific objectives of your experiment.

1. Elucidate molecular mechanisms involved in changes in calcification of sea urchin larvae due to high CO2

Measurable responses

Kathy Ruggiero k.ruggiero@auckland.ac.nz
Output

- Statistical design (CSV)

<table>
<thead>
<tr>
<th>Run</th>
<th>Pair</th>
<th>CO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>380</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>540</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1000</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>380</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>540</td>
</tr>
</tbody>
</table>
What’s next

• Beta test it
  – Is it accessible and easy-to-use?
    ▪ Want to try it out?
    ▪ Email me: k.ruggiero@auckland.ac.nz

• Make it “smart”
  – An adaptive expert system which learns from its interactions with end-users

• Make it available on smart devices
Thank you!