A QUICK INTRO TO CONTROLLED EXPERIMENTS

Source: Class of Michel Beaudouin-Lafon on experimental approach

CONTROLLED LAB EXPERIMENTS

"Observe" a phenomenon in a lab setting following a controlled experiment protocol

Form a hypothesis
Isolate the object of the study and its environment
Control its causes and effects
Observe (measure) user responses
Assess if the correlation between causes and effects is consistent with the experimental hypothesis

\+ Full control allows reliable interpretation
\- Isolation of the de object of the study (External validity ↓)

EXAMPLES

1 - Target pointing performance

Fitts Hypothesis : pointing time is proportional to the index of difficulty of the target
Experiment : "abstract" pointing task
we vary the size and distance of the target
we measure the movement time

2 - Comparison of linear and circular menus

Hypothesis : circular menus are faster
Experiment : "abstract" selection task
we vary the number of menu items
we measure the time to select an item

FORMULATE A HYPOTHESIS

Problem :

we cannot prove that a hypothesis is true
we can only prove statistically that a hypothesis is false !

Solution :

Let H1 be the experimental hypothesis
We define H0, the null hypothesis, as the opposite of H1
We try to prove that H0 is false
If we succeed, we did not prove H1, but we proved that H1 is not contradicted

EXAMPLES

Target pointing performance

H1: movement time TM is proportional to the target's ID
H0: movement time has no connection to the ID
We measure TM for several ID values
We measure the correlation between TM and ID
If it is significant (greater than a threshold), H0 is invalidated

Comparison of linear (LM) and circular menus (CM)

H1: selection time ST is faster for CM than for LM
H0: there is no significant difference between ST(CM) and ST(LM)
We measure the ST for different sizes of LM and CM
We measure the difference of selection time
If it is significant (greater than a threshold), H0 is invalidated

WHAT CAN WE SAY IF **H0** IS INVALIDATED ?

Not much more ...

In particular

WE CANNOT SAY THAT **H1** IS VALIDATED !

It failed to show that **H1** is false,
H1 remains valid until proven otherwise

It simply says that the experiment reinforces or supports **H1**

WHAT CAN WE SAY IF **H0** IS VALIDATED ?

NOTHING, ***nothing*** and **nothing** !

In particular,

WE CANNOT SAY THAT **H0** IS TRUE !

WE CANNOT SAY THAT **H1** IS INVALIDATED !

We can simply say that the experiment is inconclusive

CONTROLLING CAUSES AND EFFECTS (1/2)

The *independent variables* (or *factors*) are those we have to vary or control

The combinations of the variables define the conditions

In our experiment :

Type of menu : linear, circular
Size of menu : 3, 6, 9, 12, 15 items
Subject : expert, novice, intermediate
=> 2 x 5 x 3 = 30 conditions

CONTROLLING CAUSES AND EFFECTS (2/2)

The *dependent variables* (or *measures*) are those that we measure

To make a reliable statistical analysis, we have to get measures for each condition

Typically : in the order of 30

In our experiment :

Time to select an item
Number of errors

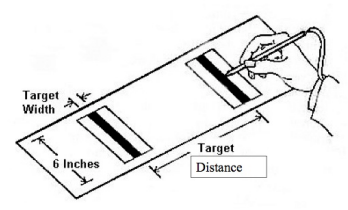
ISOLATE THE OBJECT OF THE STUDY

The trickiest part of the experiment design !

Simplify the task as much as possible to eliminate bias and external factors ... without distorting the study objective

Example : Fitts point task

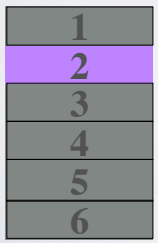
we test one dimension of movement
target = vertical band
reciprocal pointing : back and forth between two targets

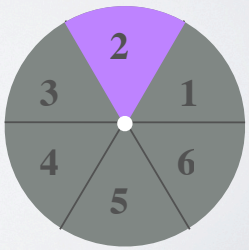


ISOLATE THE OBJECT OF THE STUDY

In our experiment (removing bias) :

Same labels for menu items
Same menu position (center of the screen)
Highlight the item to select instead of having to find it





PLAN THE EXPERIMENT : ASSIGNING SUBJECTS

Goal : To control the variability of participants

Random assignment to conditions

Recognize subject classes and introduce an independent variable if necessary (Example : level of experience of participants)

Counterbalance the presentation order to control fatigue or learning effects

Use enough participants (number determined by the type of statistical test, the degree of confidence desired, and the variability of data)

ANALYZING THE RESULTS

Statistical treatment of collected data

Ensure that the chosen statistical test is valid for the data (Anova, χ^2...)

Distribution of population	(normal)
Type of data	(ordinal)
Sampling procedure	(random)
Sample size	(n=30)

Determine the degree of confidence of results

"The hypothesis that the type of menu (linear, circular) makes no difference is rejected with a confidence of 0.95"