UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Design and Analysis of Experiments

07 - Paired Design

Version 2.11

Felipe Campelo
http://www.cpdee.ufmg.br/~fcampelo

Graduate Program in Electrical Engineering

Belo Horizonte
March 2015

"*I am driven by two main philosophies: know more today about the world than I knew yesterday and lessen the suffering of others. You'd be surprised how far that gets you.*"

Neil deGrasse Tyson
1958 -
American astrophysicist and author.

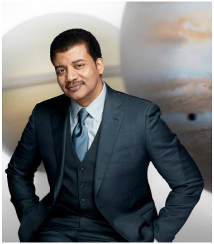

Image: `http://goo.gl/hzzFUJ`

Comparison of two means

Dependent populations

Suppose the following situation: a young researcher develops an optimization algorithm (A) for a given family of problems, and wants to compare its convergence speed against a method that represents the state-of-the-art (B).

The researcher implements both methods and wants to determine whether the proposed one has a better average performance for problems of that particular kind.

The measurements are made under homogeneous conditions (same computer, same operational conditions, etc.) and the time is measured in a way that is not sensitive to other processes running in the system.

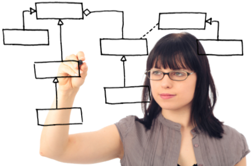

Image: `http://goo.gl/xwqig9`

This problem has some important questions worth considering:

- What is the actual question of interest?
- What is the *population* for which that question is relevant?
- What are the independent observations for that population?
- What is the relevant sample size for the experiment?

Comparison of two means

Paired design

The variability due to the different test problems is a strong source of spurious variation that can and must be controlled;

An elegant solution to eliminate the influence of this nuisance parameter is the *pairing* of the measurements by problem:

- Observations are considered in pairs (A, B) for each problem;
- Hypothesis testing is done on the sample of *differences*;

Comparison of two means
Paired design

Let y_{Aj} and y_{Bj} denote paired observations of average time for methods A and B, for each problem instance j. The *paired differences* of the observations are simply $d_j = y_{Aj} - y_{Bj}$.

If we model our observations as an additive process:

$$y_{ij} = \underbrace{\mu + \tau_i}_{\mu_i} + \beta_j + \varepsilon_{ij}$$

where μ is the grand mean, τ_i is the effect of the i-th algorithm on the mean, β_j is the effect of the j-th problem, and ε_{ij} is the model residual, then:

$$\begin{aligned} d_j &= \cancelto{0}{\left(\mu + \beta_j - \mu - \beta_j\right)} + \tau_A - \tau_B + \varepsilon_{Aj} - \varepsilon_{Bj} \\ &= \mu_D + \varepsilon_j \end{aligned}$$

The hypotheses of interest can now be defined in terms of μ_D, e.g.:

$$\begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D \neq 0 \end{cases}$$

which can now be treated as a test of hypotheses for a single sample: the population of interest is the differences in average times until convergence for the problems under investigation.

Comparison of two means

Paired design

Some other important questions worth considering:

- In this example the minimally interesting effect size δ^* must be expressed in terms of *average time gains across problems* (not within individual instances);

- The most important sample size to consider in this situation refers to the *number of problem instances*, and not necessarily to the number of within-problems repeated measures;

- The number of repetitions within each problem will have an impact on the uncertainty associated to each observation (that is, to each value of mean time to convergence for each algorithm on each problem), and should be selected with some care[b].

[b] Alternatively, we can set it as arbitrarily large, particularly in cases where the cost of repetitions is small. As much as I hate to admit it, the lazy heuristic of setting it as ≥ 30 should be enough in most algorithmic studies. A more methodologically sound approach to setting this is under development, and will be included in future versions of these lecture notes.

Bibliography

Required reading

1.
2. J.P. Simmons, L.D. Nelson, and U. Simonsohn, *False-Positive Psychology : Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant*, Psychological Science 22(11):1359-1366, 2011 - `http://goo.gl/9e0cdw`

Recommended reading

1. L. Lehe and V. Powell, *Simpson's Paradox* - `http://vudlab.com/simpsons/`

About this material

Conditions of use and referencing

Please reference this work as:

Felipe Campelo (2015), *Lecture Notes on Design and Analysis of Experiments*.
Online: https://github.com/fcampelo/Design-and-Analysis-of-Experiments
Version 2.11, Chapter 6; Creative Commons BY-NC-SA 4.0.

```
@Misc{Campelo2015-01,
  title={Lecture Notes on Design and Analysis of Experiments},
  author={Felipe Campelo},
  howPublished={\url{https://github.com/fcampelo/Design-and-Analysis-of-Experiments}},
  year={2015},
  note={Version 2.11, Chapter 6; Creative Commons BY-NC-SA 4.0.},
}
```