Experiment Design for Computer Sciences (0AL0400)

Topic 09 - Review

Claus Aranha
caranha@cs.tsukuba.ac.jp

University of Tsukuba, Department of Computer Sciences

Version 2021.1 (Updated June 10, 2021)

Lecture Outline

Let's review the process of experiment design that was studied in this course

- Hints for Designing your Experiment
- Hints for Statistical Analysis

Motivating Example

In my research, I need to take pictures of plants, and then process them using a segmentation network.

The camera creates very high resolution pictures in RAW format. However, my network reads the picture in a lower resolution format.

I have several choices for the transformation of the picture:

- What algorithm I should use to scale down the picture (linear scaling, cubic scaling, etc);
- What final resolution to save the pictures as.

I have to choose the scaling algorithm and the final resolution. These choices will effect the time to convert the pictures, and the amount of disk space needed. The choice may also have an effect on the performance of the algorithm.

What should I do?

Point 1: Choose an experiment that you care about

If you don't care about the question/answer of your experiment, it will be very hard to do a good experiment.

- You won't be able to imagine what results are important or irrelevant;
- You will not be able to think about the factors that can influence the experiment.
- You will not be able to decide which factors to fix or vary; and the parameters of variation.

Point 2: Think about the output variable

The Output variable is the result of your experiment. It is the value that you are interested in observing.

When you think about the output variable, consider the characteristics of this variable.

- Is the variable discrete or continuous?
- What is the range of values for this variable? What are the typical values?
- What is the typical variation for this variable?
- What is the size of difference that is important for you and for your experimental question?

The answer to some of the questions above may come from your experience. The answer to some of the questions above can be "I don't know". That is okay, we make experiments and analyse data to find the answers to these questions.

If you have multiple output variables, you might also want to think if you want to make separate experiments.

Point 3: The experimental Factors

The experimental factors are the conditions that could change the result of the experiment.

When you think about each experimental factor, you have to consider their characteristics as well.

- Are these factors discrete or continuous?
- What is the range of possible values? What are the typical values?
- What is the relationship between the factor and the output variable?
- What are the values of the factor that interest you?

Point 3: The experimental Factors (2)

Based on these characteristics, you should separate the factors as follows:

- Factors that you will study in the experiment. These factors you will set to specific values, because you are interested in the influence of these factors in the outcome of the output variable.
- Factors that you want to control in the experiment. These factors have important influence in the output variable, but you do not want to study their influence. So you will have to fix their value during the experimental design, or limit their effect using techniques such as pairing (studied in class) or blocking (not studied in class)
- Other factors, that you cannot control, but you can estimate their influence in the output variable.

Point 4: Expected results

It is helpful to try to predict what is the expected result of the experiment before you execute the experiment. When you learn the result, you can more clearly define if the result confirmed your expectations, or revealed something surprising.

By considering the expected result, you can discover if you have missing knowledge about your experiment that you need to study, and if there are parts of the experiment design that are missing.

Point 5: Pre-experimental choices

Before you begin your experiment, you should make several choices:

- Decide the Statistical Test (Depends on type of output variable, number of factors and factor levels, characteristics of the problem);
  - One sample: z-test or t-test;
  - two samples: t-test, paired or not paired;
  - multiple samples: ANOVA;
  - output variable is not a normal random variable: non-parametric tests;
- Define Hypothesis and critical values;
- Decide the Test Parameters (confidence, power, minimal difference of interest), depend on the experiment characteristics;
- Calculate sample size;

Point 6: Collecting Data and Descriptive Statistics

During data collection, make sure to pay attention to any unusual things that happen during the experiment.

Describe your initial results using descriptive statistics, such as mean and standard deviation of the output variable, boxplots, confidence interval, etc.

A good visualization here can make the next step much easier.

Point 7: Statistical test analysis and conclusions

Perform your statistical test, calculate the test statistic and the p-value.

Make any necessary calculations to verify that your output variable follows the necessary assumptions of the statistical test.

When reporting the result, do not forget to include not only the p-value, but also the descriptive statistics.

Compare the final result with your initial prediction. What did you learn?

About these Slides

These slides were made by Claus Aranha, 2021. You are welcome to copy, re-use and modify this material.

These slides are a modification of "Design and Analysis of Experiments (2018)" by Felipe Campelo, used with permission.

Individual images in some slides might have been made by other authors. Please see the following references for those cases.

Image Credits I