Experiment Design for Computer Sciences (0AL0400)

Topic 08 - Statistical Recipes

Claus Aranha
caranha@cs.tsukuba.ac.jp

University of Tsukuba, Department of Computer Sciences

Version 2023.1 (Updated June 21, 2023)

Outline

In this material we study a few extra topics related to statistical testing that came up in previous lectures.

Second Order Interaction Effects

What are Interaction Effects?

When analyzing data from an experiment, a common question of interest is **how do the factors affect the response variable?**.

For example, we want to know how do the values of the parameters F and CR affect the performance of the *Differential Evolution* algorithm.

Main Effects are how the factors individually affect the response variable:[1]

- Convergence time decreases as we increase the value of CR;
- Convergence time changes in a parabole with F, and is minimal when $F \approx 0.5$;

Interaction Effects are how the factors affect the response variable in combination;

- When $CR \approx 0.5$, minimal convergence time happens for $F \approx 0.6$
- When $CR \approx 0.8$, minimal convergence time happens for $F \approx 1.1$

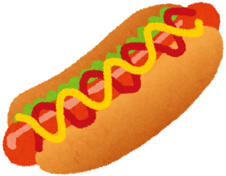
[1] hypothetical statements

Simple Example – Chocolate survey

Imagine that we are conducting a survey to discover what is the best food condiment.

We perform a survey, where we give the person a random food with a random condiment, and ask them their satisfaction.

- Independent Factor: **Food:** Hotdog, Ice Cream
- Independent Factor: **Condiment:** Chocolate Sauce, Mustard
- Response Variable : Enjoyment



Chocolate survey data

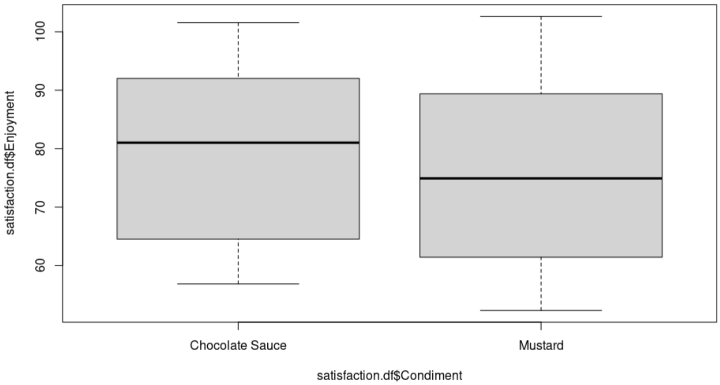
Here is what the data looks like (you can download data and code from manaba)

```
% cat Condiments.csv
"Enjoyment","Food","Condiment"
81.9269574232529,"Hot Dog","Mustard"
84.9397743293292,"Hot Dog","Mustard"
90.2864793280143 8,"Hot Dog","Mustard"
89.5618015166550 2,"Hot Dog","Mustard"
97.6768259188006 6,"Hot Dog","Mustard"
83.6171299693461 8,"Hot Dog","Mustard"
89.2108604651020 6,"Hot Dog","Mustard"
90.766866722188 3,"Hot Dog","Mustard"
102.620440307723 65,"Hot Dog","Mustard"
85.7439003642255 1,"Hot Dog","Mustard"
96.592358894779 2,"Hot Dog","Mustard"
```

Visualizing the results with a boxplot

At first glance, it seems that the condiment does not make a big difference?

```
> satisfaction.df <- read.csv("Condiments.csv")
> boxplot(satisfaction.df$Enjoyment ~ satisfaction.df$Condiment)
```



Means across factors

We suspect that something is wrong here, so we plot the means for each combination of food and condiment.[2]

```
> aggregate(Enjoyment ~ Food + Condiment, satisfaction.df, mean)
       Food         Condiment Enjoyment
1   Hot Dog Chocolate Sauce  65.31661
2 Ice Cream Chocolate Sauce  93.04810
3   Hot Dog           Mustard  89.60569
4 Ice Cream           Mustard  61.30891
```

It appears that the Enjoyment depends on both the food and the condiment!
Unfortunately, ice cream with mustard does not seem to be a very popular choice...

[2]It is always good to be curious about your data and explore it in depth.

Measuring Interaction Effects

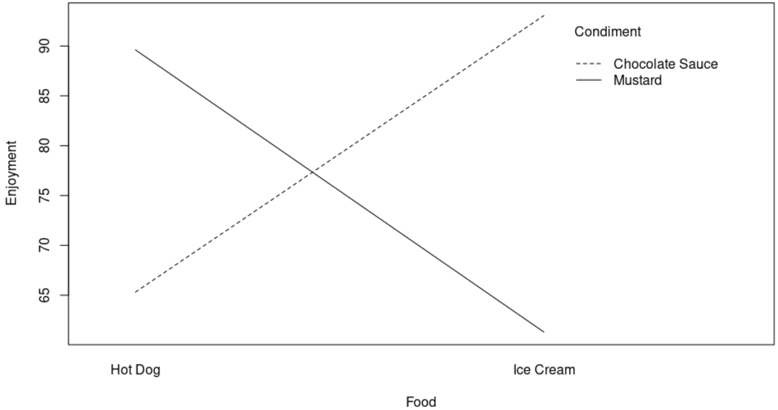
We can get a statistical measure of our interaction effect by using the ANOVA analysis on a linear model that includes the effect.

```
> Food.lm <- lm(formula = Enjoyment ~ Food + Condiment + Food*Condiment,
+                data = satisfaction.df)
> # ANOVA (effects of each factor on the final result)
> summary(aov(Food.lm))
               Df Sum Sq Mean Sq F value  Pr(>F)
Food            1      2       2   0.064 0.80136
Condiment       1    278     278  11.071 0.00135 **
Food:Condiment  1  15696   15696 626.153 < 2e-16 ***
Residuals      76   1905      25
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F (and P) values on the Food:Condiment row indicate the strong interaction effect.

Visualization of Interaction Effects

```
interaction.plot(satisfaction.df$Food, satisfaction.df$Condiment,
                 satisfaction.df$Enjoyment,
                 xlab = "Food", ylab = "Enjoyment", trace.label = "Condiment")
```



Another Example: Smoking, Abestos and Cancer

The data below was presented in Hilt et al. (1986), about the risk of lung cancer depending on whether a person smokes, and whether they are exposed to abestos.

Table 1 Risk of lung cancer by smoking and asbestos status

	No asbestos	Asbestos
Non-smoker	0.0011	0.0067
Smoker	0.0095	0.0450

We can observe that the risk of Lung Cancer increases much more when both Abestos and Smoking are present. More than we would expect from either factor alone.

Interaction effects can be both negative and positive.

Interaction Effects on DE Parameters

Let's see a more complex example, that is closer to computer science.

Differential Evolution (DE) is a popular meta-heuristic algorithm that can be used to solve blackbox optimization problems.

The it is known that the performance of DE depends on the problem being solved, as well as its two control parameters: F and Cr. Therefore, the user is encouraged to **fine tune** the algorithm prior to application.

In this example, we run a simplified experiment to find optimal parameter values for DE for a sample problem.

DE Description

Differential Evolution is an algorithm that rely on two parameters:

- CR: Ranges from 0 to 1, recommended value is 0.5;
- F: Ranges from 0 to 2, recommended value is 0.8;

It is recommended that the algorithm is fine-tuned for specific problems. In this case, we will use check the performance of the algorithm in a benchmark function under different parameter values.

Experimental Design

Scientific inquiry: Understand how the convergence rate of DE changes as we set different values of the parametes CR and F. In specific, try to detect if there are any interactions between these parameters.

- **Response Variable**: Number of iterations until a target solution quality is reached.
- **Independent Factor 1**: CR, levels: 0.1, 0.3, 0.5, 0.7, 0.9
- **Independent Factor 2**: F, levels: 0.2, 0.4, 0.8, 1.0, 1.2
- **Noise Factor**: Random seed (probabilistic algorithm)

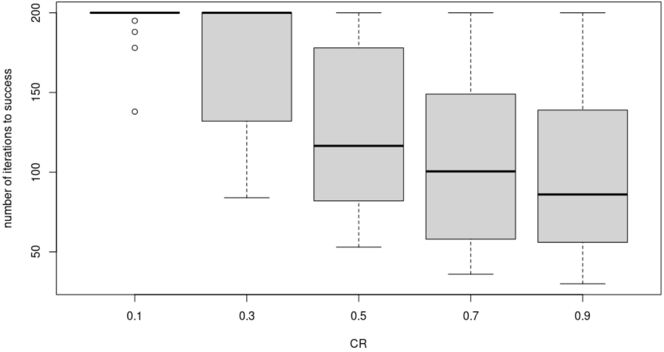
Since this is an exploratory experiment, we do not set specific values for α, β, etc. These values should be set when specific null and alternate hypothesis are considered.

Initial observation of the parameters – CR

First, let's observe how changing the parameters affect the response variable:

Notes:

- This is a box plot of the response value, grouped by CR value.
- It includes different values of F in the results, so interaction effects might be a concern.
- Still, we see a generally downwards trend for higher CR values **For this problem.**

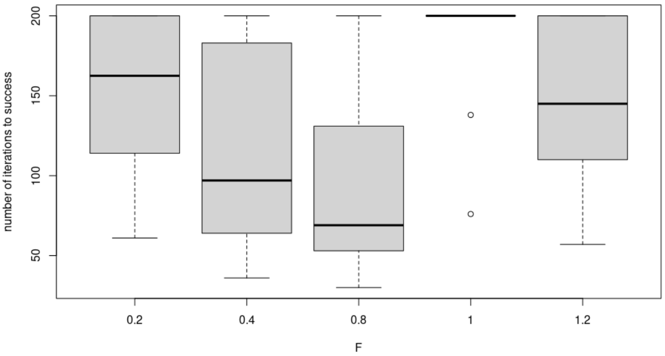


Initial observation of the parameters – F

Same boxplot observation, focusing on the parameter F:

Notes:

- The parameter F shows a clearly non-linear relationship with the output variable.
- The anomalous behavior for $F = 1$ should be investigated more carefully (but we won't do it here).



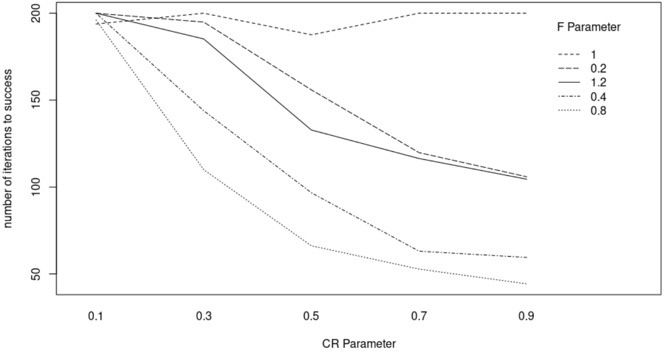
Measuring Interaction Effect with a linear regression

```
> DE.lm <- lm(formula = R ~ F + CR + F*CR,              <-- Linear Reg Formula
              data = tab.results)                           Includes F*CR, the
> summary(DE.lm)                                            iteraction factor.
(...)
Residuals:
    Min      1Q  Median      3Q     Max
-90.107 -36.083   7.101  31.577 103.793                 <-- Residuals don't look
(...)                                                       too skewed.
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   212.31      12.94  16.409  < 2e-16 ***
F             -13.15      15.97  -0.823  0.41124        <-- no "linear" effect!
CR           -177.01      22.52  -7.859 1.21e-13 ***
F:CR           74.56      27.81   2.681  0.00783 **     <-- Detects a significant
(...)                                                       effect from changing
                                                            both factors together
```

Interaction Graph – CR by F

A visual investigation of interaction effects might be more informative.

```
interaction.plot(tab.results$CR, tab.results$F, tab.results$R)
```

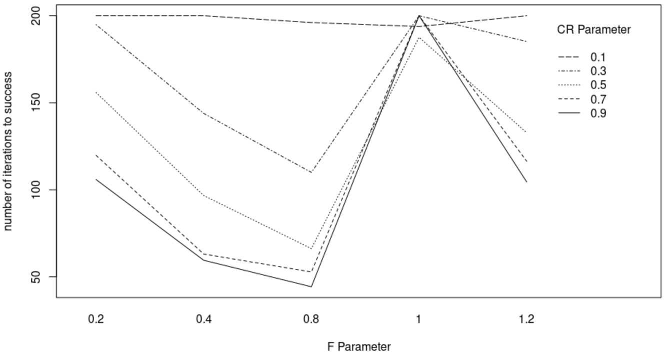


- $F = 0.2$ and $F = 1.2$ show low convergence compared to $F = 0.4$ and $F = 0.8$. But the shape of the curve seems roughly similar.
- Again we see the anomalous behavior of $F = 1$

Interaction Graph – F by CR

A visual investigation of interaction effects might be more informative.

```
interaction.plot(tab.results$F, tab.results$CR, tab.results$R)
```



- Here we see that the curves show roughly the same shape, and convergence speed increasing as CR increases;
- Here the anomalous behavior of $F = 1$ is more pronounced.

Early Conclusions

- The linear regression model indicates a strong effect for a CR*F factor.
- However, this effect is not so clear in a visual inspection.
  - This may be because the relationship between F and the output variable is not linear.
- The recommendation **from these results** would be to try $F = 0.8$, and CR as high as possible.
  - Maybe worth investigation $CR > 0.9$;
- Maybe worth it investigating $F = 1.0$ for bugs in the code, since a scaling factor should not have such an anomalous behavior.
- A good experiment many times raises as many questions as it answers!

Caveats of this case study

- **Non-linear effects**: F shows a clear non-linear relationship with the output variable.
  - Note how the difference cause the linear model to return non-reliable information!
  - Good opportunity to investigate non-linear models.
- Second order interaction is already though. Third order interaction and above become exponentialy harder to analyze;
- Eventually, it becomes more practical to use sequential models like IRACE and SMAC;

Suggested Extra Reading

- Bartz-Bielstein: "Experimental Methods for the Analysis of Optimization Algorithms", In depth discussion about parametrization and analysis of algorithms like DE;
- Paul Teetor: "R Cookbook", Quick recipes for a variety of data visualizations in R;
- Statistics by Jim: "Interaction Effects".
  A short, intuitive description of interaction effects.
  `https://statisticsbyjim.com/regression/interaction-effects/`
- T.J. VanderWeele, M. J. Know: "A Tutorial on Interaction.
  A more in-depth discussion of interaction in several models, but focused on medicine use cases. `https://www.hsph.harvard.edu/wp-content/uploads/sites/603/2018/04/InteractionTutorial_EM.pdf`

Statistical Model for Equality Testing

General Idea

About these Slides

These slides were made by Claus Aranha, 2022. 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

[Page 5] Hotdog and icecream images from irasutoya
[Page 11] Lung Cancer table from VanderWeele and Know: A Tutorial on Interaction