Example of the Application of an DOE to Coronavarius Data Analysis

Jorge Luis Romeu, Ph.D.
Romeu@cortland.edu
https://web.cortland.edu/romeu/
Syracuse, April 30, 2020

Introduction

This **example illustrates** how **Design of Experiments/DOE** can be used in the analysis and research of Covid-19 Data, to identify and assess the effects of significant factors in Virus Containment. **The Data has been made-up; its objective is to illustrate the Power of DOE in Covid-19 problems.**

This type of analysis can be implemented at County or Regional levels, with incoming data, to verify that the containment measures in use are working as they should and if so, to quantify their results.

DOEs may be implemented using a pre-established statistical design (e.g. full, fractional factorials, etc.) Some professional statistical guidance is necessary. Otherwise, DOEs can be implemented in a similar manner as EVOP (Evolutionary Operations) taking advantage of successive implementation of various containment measures. Results should be interpreted with care, using professional statistical help.

In the present example the DOE analysis has been implemented using an Excel Spread Sheet, to avoid the use of expensive statistical SW. A tutorial on the calculations of the DOE matrix, using a Spread Sheet as done below, can be found in: https://web.cortland.edu/matresearch/FFDOEOverview2007.pdf

The DOE Example

Assume we collect weekly infection rate data from several counties or regions, that implement different levels of containment measures. The three measures (here-on Factors) analyzed are:

(A) Suggested Social Distancing, when persons are in congested places, with two options (levels): in Operation (denoted as -1) and None (denoted +1);

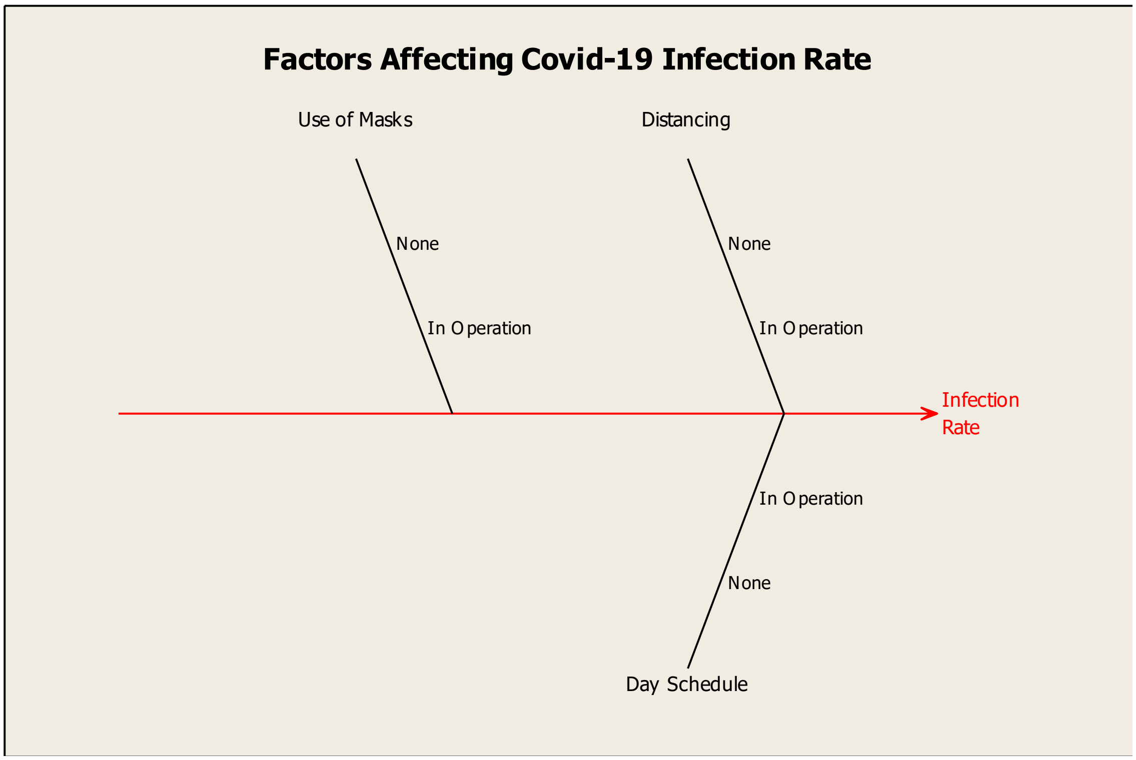
(B), Suggested Use of Face Masks when persons are in congested places, with two options (levels): in Operation (denoted as -1) and None (denoted +1);

(C), Suggested Use of Birthday Schedule (e.g. those born on odd years should go out on M/W/F, and those born on even years, should go on T/Th/S), with two levels: in Operation (-1) and None (+1);

We use three replications (measurements) of three weekly Infection Rates in regions where these three above-defined factors (A, B, C) have been used in specified combinations (-1, -1, -1), in different periods of time (weeks). For example, in the matrix below, Run 1 (-1, -1, -1) means that the data comes from a county/region where, that week, Social Distancing, Use of Masks and Day Schedule were all in Operation

Let the <u>Response</u> of interest be the <u>*Effect of* Factors Social Distancing, Face Masks and Day Schedule *on* the *Infection Rate*</u>. We will measure this effect by <u>*comparing the measured rate with*</u> *a* ***standard/desired rate (say* 5%.)** *below which Community Spread can be effectively contained***.** Responses are computed, for example: if *actual infection* rate was *7%*, the *value recorded* in our analysis would be *7 - 5 = 2%*

Below we show an *Ishikawa* (Cause and Effect) Chart for the above problem:



The graph expresses how the three factors, at their two stated levels, can effect infection rate.

<u>Design of Experiments Results</u>

We show below the DOE <u>Analysis Table.</u> There are eight runs (eight possible combinations of the three Factors A, B, C, at their -1 and +1 levels. Each run is a line, where under the respective columns A, B, C are recorded the levels at which these applied when the data was collected. There are three *replications* denoted Y1, Y2, Y3, corresponding to the three different weeks in which said *factors* were operating at said *levels*. Replications are then averaged and their variance is calculated.

The *Effects,* for the different column values (*Factors*), are obtained by algebraically adding the eight row Averages, according to the +/- signs below each Effect column.

For example, for Factor A (Social Distancing), we would add: -1.5 + 1.7 … -4.1 + 4.9 = 0.23. This value is the increment over desired 5%; and its 95% Confidence Interval (-.3, 0.76) is the region where such +/- increment lies, 95% of the times (remember, the data is only a *sample*).

Table of DOE Analysis Results:

Design of Experiments **Covid-19** IV/2020

Factorial Experiments 2^3 (Three Replics/Treatment) **Run Results**

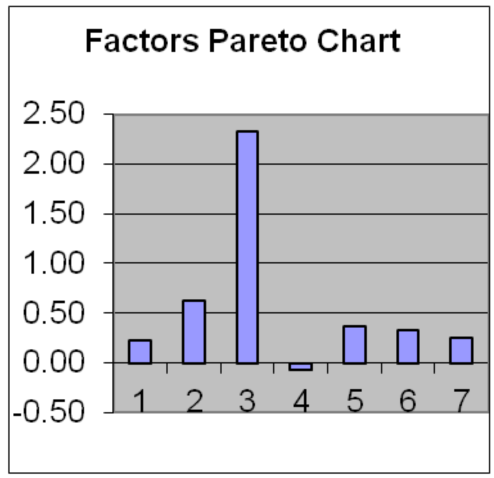
Run	A	B	C	AB	AC	BC	ABC	Y1	Y2	Y3	Avg.	Var.
1	-1	-1	-1	1	1	1	-1	1.50	1.42	1.72	1.547	0.024
2	1	-1	-1	-1	-1	1	1	1.56	1.73	1.87	1.719	0.024
3	-1	1	-1	-1	1	-1	1	1.71	2.75	2.02	2.161	0.286
4	1	1	-1	1	-1	-1	-1	1.98	1.64	1.50	1.704	0.062
5	-1	-1	1	1	-1	-1	1	2.52	4.12	3.61	3.417	0.673
6	1	-1	1	-1	1	-1	-1	4.08	3.86	3.57	3.835	0.064
7	-1	1	1	-1	-1	1	-1	4.19	3.49	4.90	4.193	0.492
8	1	1	1	1	1	1	1	5.71	5.02	4.19	4.976	0.577
TotSum								23.24	24.04	23.38	23.55	2.20
SumY+	12.23	13.03	16.42	11.64	12.52	12.44	12.27					
SumY-	11.32	10.52	7.13	11.91	11.03	11.12	11.28					
AvgY+	3.06	3.26	4.11	2.91	3.13	3.11	3.07					
AvgY-	2.83	2.63	1.78	2.98	2.76	2.78	2.82					
Effect	**0.23**	**0.63**	**2.32**	**-0.07**	**0.37**	**0.33**	**0.25**					

Factors Analyzed:

Factor A:	Social Distance
Factor B:	Use Face Mask
Factor C:	Day Scheduling

Factors	**SocialDistance**	**Use Face Masks**	**DayScheduling**
Low Level	Implemented	Implemented	Implemented
HighLevel	None	None	None

Response: **InfectionRate**



Var. of Model	0.28	**StdDv**	0.52
Var. of Effect	0.05	**StdDv**	0.21
Student T (0.025;DF) =		2.473	
C.I. Half Width =		**0.530**	

Significant Factors & 95% CI Limits:

Factor	**A**	**B**	**C**	**AB**	**AC**	**BC**	**ABC**
Signific.	**No**	**Yes**	**Yes**	**No**	**No**	**No**	**No**
LwrLimit	-0.30	0.10	1.79	-0.60	-0.16	-0.20	-0.28
UprLimit	0.76	1.16	2.85	0.46	0.90	0.86	0.78

Data Analysis Interpretation:

Two of the three Main Effects are statistically significant (i.e. they increase the infection rate over 5%). *Effect A, Social Distancing, is not statistically significant*: its Effect is 0.23; its 95% Confidence Interval (CI) is -0.3 to 0.76, and covers zero. Social Distancing helps maintain the infection rate at the desired rate of 5%. *Effect B, Use of Face Masks, is mildly significant*. Its 95% CI shows its use may allow an infection rate increase 0.1 to 1.16% over the desired 5%. *Effect C, Day Scheduling is statistically significant*. Its 95% CI shows Social Distancing allows an infection rate increase 1.79 to 2.85% over 5%. Thus, *Social Distancing is the most helpful tool to help keep infection rates at the desired 5% levels*. *More replications will help* determine if Wearing Masks will actually become Not Significant, as currently it is barely significant.

THE ABOVE RESULTS ARE JUST FOR ILLUSTRATION, AND DO NOT REPRESENT REAL DATA ANALYSES.