Tool for One-way and Two-way CATANOVA and ORDANOVA for Analysis of Variation in a Cross-balanced Design and Power of a Test

Leuven - ENBIS 2024
16/09/2024, 14:30

Yariv N. Marmor, Tamar Gadrich
(BRAUDE - College of Engineering ,Karmiel)

Francesca Pennecchi
(Istituto Nazionale di Ricerca Metrologica - INRIM)

D. Brynn Hibbert
(School of Chemistry, UNSW Sydney)

Anastasia A. Semenova
(V.M. Gorbatov Federal Research Center for Food Systems)

Ilya Kuselman
(Independent Consultant on Metrology)

Outline

- Definitions of nominal and ordinal data.
- Analysis of variation of qualitative (nominal) data and semi-quantitative (ordinal) data.
- Examples.
- Explanation of the tool (simulation).
- Power of the test and associated risks of incorrect decisions.
- Continuation of the previous examples.
- Summary and Future Research

Definitions of Categorical Data

- Nominal Data
  - Categorical data with unordered exhaustive and disjoint labeled by *K* exclusive categories.
  - Nominal data have no magnitude: occurrences can be only **equal** or **unequal**, i.e., belong to the same or different categories.
  - Examples: Color of a spot test; Gender; Blood Type; Sequence of amino acids in a polypeptide; Type of weld imperfection.
- Ordinal Data (*K* exclusive sorted categories)
  - Categorical data that are ordered according to their inherent magnitude. Ordinal data can enter into empirical relations only; can be **equal** or **unequal**, **greater than** or **less than**.
  - Examples: Octane number of fuel; Fault severity according to experts.
  - Ordinal data are arranged according to ordinal scales (e.g., 1 to 5). Ordinal data should not be treated as numbers since the distance between 1 and 2 may differ from 2 and 3 or 3 and 4.

Analysis of Variation of Nominal and Ordinal Data – Layout (Gadrich and Marmor 2021)

We assume a cross-balances design of I levels of factor *X1*, J levels of factor *X2* and K categories:

- $p = (p_1, p_2, \dots, p_K)$ is the theoretical probability of response.
- $\sum p_k = 1;$
- In **Ordinal data** also $F_k = \sum_{q=1}^{k} p_{\text{q}}$; $F_K = 1$

Factor $X1$	1	$\dots$	Factor $X2$ j	$\dots$	J	Total counts
1	n_{11k}	$\dots$	n_{1jk}	$\dots$	n_{1Jk}	J
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
i	n_{i1k}	$\dots$	n_{ijk}	$\dots$	n_{iJk}	J
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
I	n_{I1k}	$\dots$	n_{Ijk}	$\dots$	n_{IJk}	J
Total counts	I	$\dots$	I	$\dots$	I	N

In general, a cross-balanced design may contain *(i, j)* cells with the same number *n > 1* of replicate responses and the total number of responses $N = nIJ$. This design allows to test the interaction between the factors.

Analysis of Variation of Ordinal Data – Variation Definition (Gadrich and Marmor 2021)

- Treating *N* responses as a statistical sample, and the no. of responses n_{ijk} as a random variable, then $\hat{p}_{ijk} = \frac{n_{ijk}}{N}$ (and $\hat{F}_{ijk} = \sum_{q=1}^{k} \hat{p}_{ijq}$) denote the sample relative (cumulative) frequency of responses belonging to (up to) the *k*-th category.
- The sample total cumulative relative frequency of all responses belonging to the k-th category is denoted by: $\hat{F}_{..k} = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{F}_{ijk} \quad k = 1, 2, \ldots, K$.
- $\hat{F}_{i.k} = \frac{1}{J} \sum_{j=1}^{J} \hat{F}_{ijk}$ $(\forall i, k)$ and $\hat{F}_{.jk} = \frac{1}{I} \sum_{i=1}^{I} \hat{F}_{ijk}$ $(\forall j, k)$ denote the sample cumulative relative frequency of responses up to the *k*-th category at level *i* of factor *X1* and at level *j* of factor *X2*, respectively.
- The observed (sample) total normalized (to [0,1] interval) variation of the response variable $\boldsymbol{Y}$, is estimated in the two-way ORDANOVA as (and its decompositions):

$$\hat{V}_{\mathrm{T}} = \frac{4}{(K-1)} \sum_{k=1}^{K-1} \hat{F}_{..k}\left(1 - \hat{F}_{..k}\right) = \hat{C}_B + \hat{V}_W = \hat{C}_{X1}^{B} + \hat{C}_{X2}^{B} + \hat{C}_{X1*X2}^{B} + \hat{V}_W$$

Analysis of Variation of Nominal and Ordinal Data – Summary (Gadrich and Marmor 2021)

- For ordinal data (Based on ORDANOVA):
  - $\hat{V}_{\mathrm{T}} = \frac{4}{(K-1)} \sum_{k=1}^{K-1} \hat{F}_{..k}\left(1 - \hat{F}_{..k}\right)$
  - $\hat{C}_{X1}^{\mathrm{B}} = \frac{4}{(K-1)} \sum_{k=1}^{K-1} \frac{1}{I} \sum_{i=1}^{I} \left(\hat{F}_{i.k} - \hat{F}_{..k}\right)^2$
  - $\hat{C}_{X2}^{\mathrm{B}} = \frac{4}{(K-1)} \sum_{k=1}^{K-1} \frac{1}{J} \sum_{j=1}^{J} \left(\hat{F}_{.jk} - \hat{F}_{..k}\right)^2$
  - $\hat{C}_{X1*X2}^{\mathrm{B}} = \frac{4}{(K-1)} \sum_{k=1}^{K-1} \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\hat{F}_{ijk} - \hat{F}_{i.k} - \hat{F}_{.jk} + \hat{F}_{..k}\right)^2$
- Similarly, for nominal data (Based on CATANOVA):
  - $\hat{V}_{\mathrm{T}} = \frac{K}{(K-1)} \sum_{k=1}^{K-1} \hat{p}_{..k}\left(1 - \hat{p}_{..k}\right)$
  - $\hat{C}_{X1}^{\mathrm{B}} = \frac{K}{K-1} \sum_{k=1}^{K} \frac{1}{I} \sum_{i=1}^{I} (\hat{p}_{i.k} - \hat{p}_{..k})^2$
  - $\hat{C}_{X2}^{\mathrm{B}} = \frac{K}{K-1} \sum_{k=1}^{K} \frac{1}{J} \sum_{j=1}^{J} \left(\hat{p}_{.jk} - \hat{p}_{..k}\right)^2$
  - $\hat{C}_{X1*X2}^{\mathrm{B}} = \frac{K}{K-1} \sum_{k=1}^{K} \frac{1}{J} \sum_{j=1}^{J} \left(\hat{p}_{ijk} - \hat{p}_{i.k} - \hat{p}_{.jk} + \hat{p}_{..k}\right)^2$

Analysis of Variation of Nominal and Ordinal Data (Gadrich and Marmor 2021)

- The null hypothesis H_0 of homogeneity of the responses states that the probability of classifying the responses as belonging to the *k*-th category does not depend on the levels of the first nor the second factor, i.e., $\hat{p}_{ijk} = \hat{p}_k \; \forall i, j, k$:

$$\frac{E\left[\hat{V}_T\right]}{df_T} = \frac{E\left[\hat{C}^B_{X1}\right]}{df_{X1}} = \frac{E\left[\hat{C}^B_{X2}\right]}{df_{X2}} = \frac{E\left[\hat{C}^B_{X1*X2}\right]}{df_{X1*X2}} = \frac{V_T}{N}$$

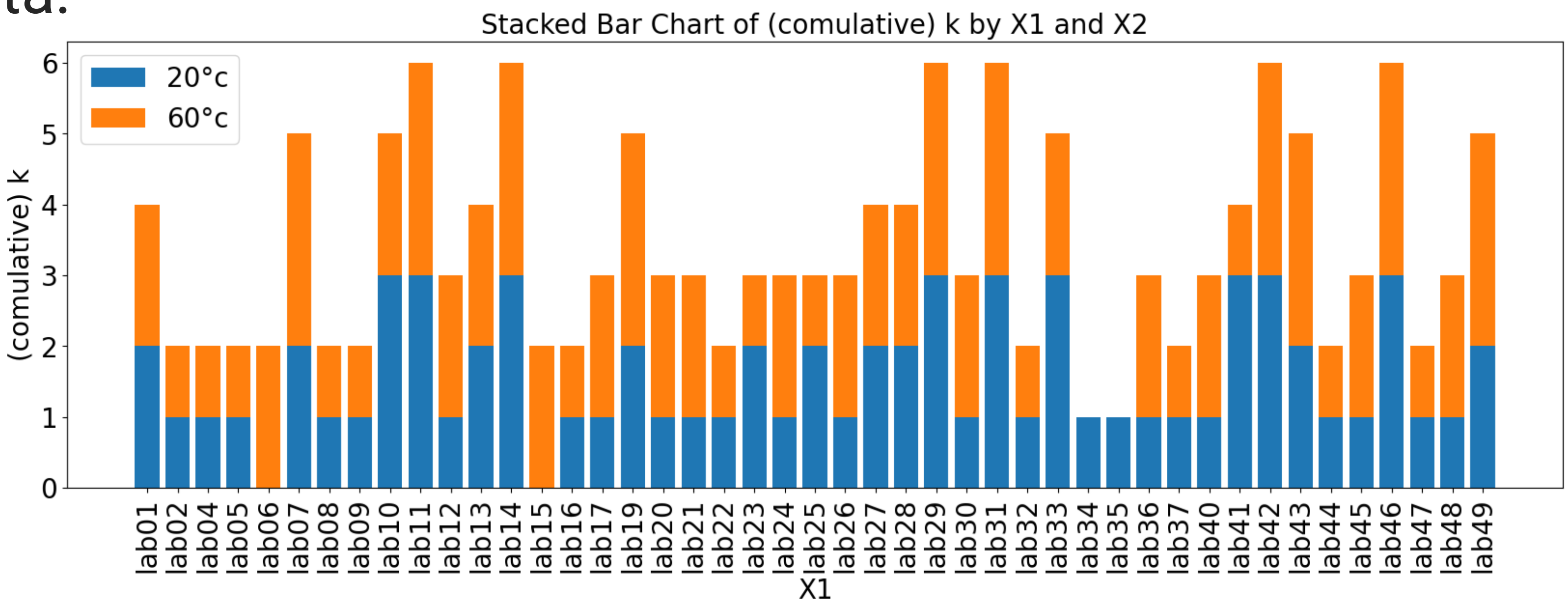
To test the statistical significance of both the factor effects the following significance indices (test statistics) have been defined:

$$\widehat{SI}_{X1} = \frac{\hat{C}^B_{X1}}{df_{X1}} \Big/ \frac{\hat{V}_T}{df_T};\; \widehat{SI}_{X2} = \frac{\hat{C}^B_{X2}}{df_{X2}} \Big/ \frac{\hat{V}_T}{df_T};\; \widehat{SI}_{X1*X2} = \frac{\hat{C}^B_{X1*X2}}{df_{X1*X2}} \Big/ \frac{\hat{V}_T}{df_T}$$

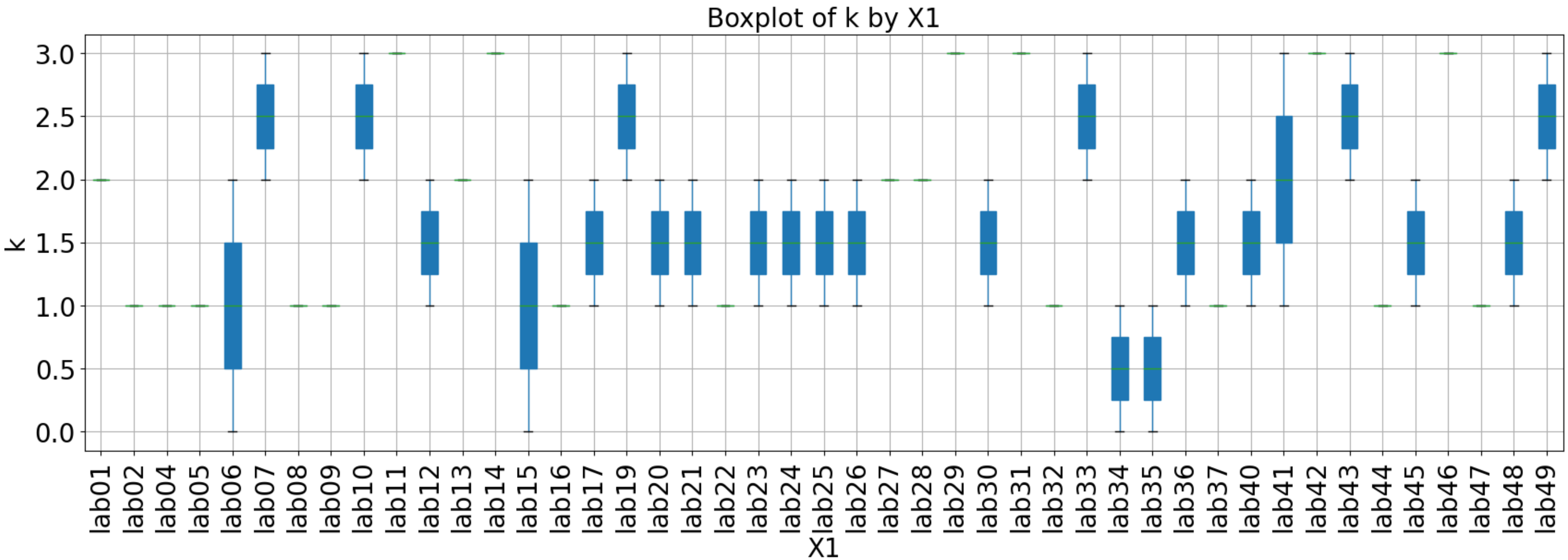
$$df_{\mathrm{T}} = N - 1;\, df_{X1} = I - 1;\, df_{X2} = J - 1;\, df_{X1*X2} = (I-1)(J-1)$$

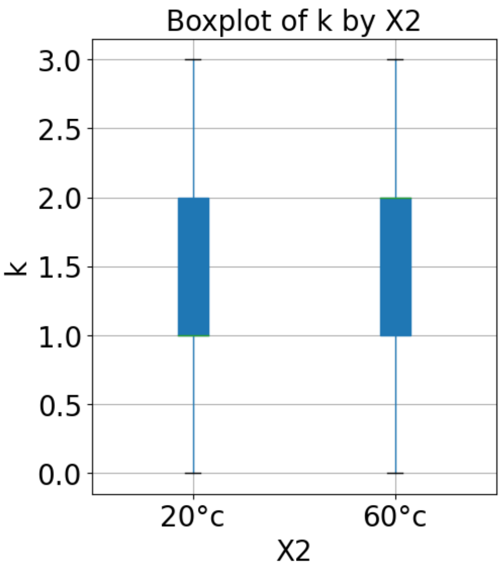
Example 1 (Water Chlorine)

- Problem outline (Gadrich et al., 2022):
  - Intensity of chlorine odor of bottled drinking water.
  - Water was kept at room temp ($20^{o}c$) or heated ($60^{o}c$) – *X2*.
  - 49 laboratories, 45 sent result for each temperature – *X1*.
  - The odor intensity according to the national standard: a) imperceptible odor, b) very weak, c) weak, d) noticeable, e) distinct, and f) very strong – *k* (0-5).
  - The data:



Example 1 (Water Chlorine) - Results





<u>Two-way Anova (treating *k* as a continuous number)</u>

	SSE	df	F	p-value
X1	48.400000	44.0	3.193548	0.000096
X2	1.344444	1.0	3.903226	0.054486
Residual	15.155556	44.0	NaN	NaN

As the data is ordinal, two-way Anova "works" well only when the order of magnitude is evenly distributed among options

Example 1 (Water Chlorine) – Simulation Tool

Model Type:	Ordinal
Number of Iterations:	10,000
α:	0.05
w:	0.3

I:	45
J:	2
K:	6
n:	1

Two-way ORDANOVA

P: 0.0444 ; 0.4222 ; 0.3222 ; 0.2111 ; 0.0000 ; 0.0000

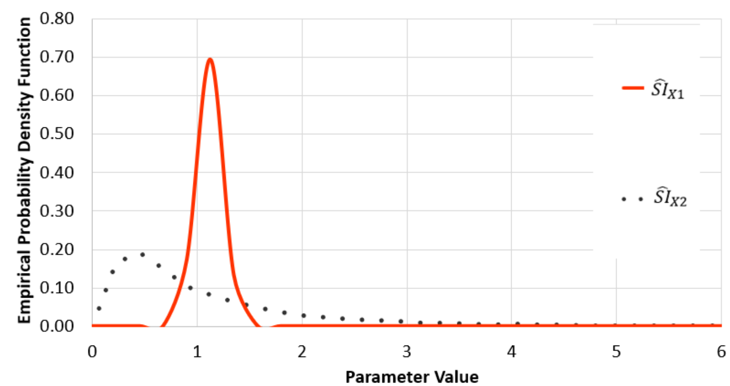
Get File

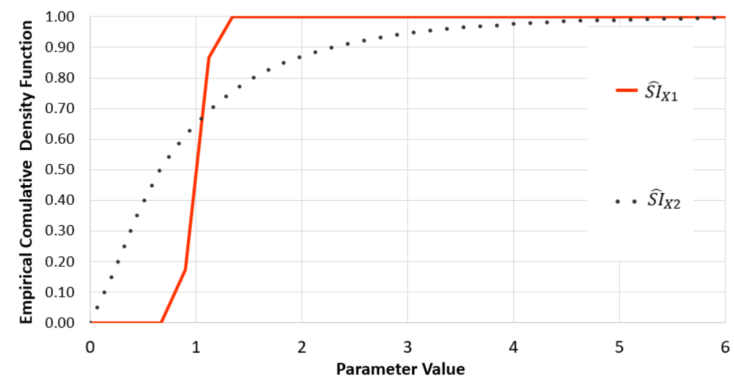
Run Model

File Name:
data_water_Chlorine.txt

Two-way ORDANOVA without replication

	Variation Component	df	Significance index	P-value	SI_{crit}	Power
X1	0.2463	44	1.3601	0.00060	1.1830	0.10350
X2	0.0100	1	2.4236	0.08981	3.1093	0.28830
Interaction	n/a	n/a	n/a	n/a	n/a	n/a
Within	0.11	44				
Total	0.3663	89				





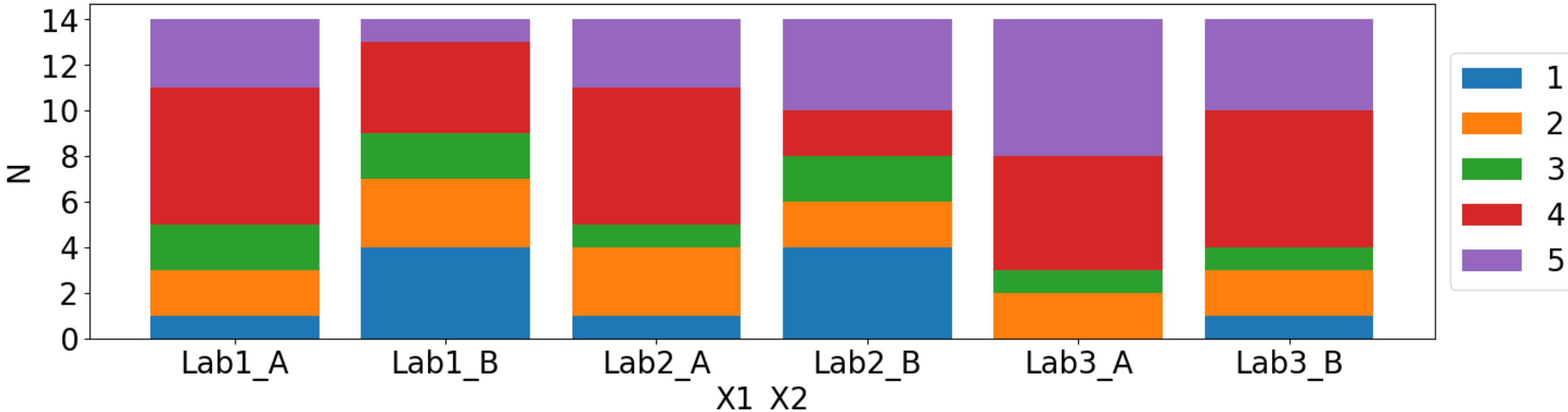
“similar” results as the two-way Anova

Example 2 (Weld Imperfection)

- Problem outline (Gadrich, Kuselman and Andrić, 2020):
  - 3 laboratories – *X1*.
  - Asked to classify weld imperfections according to ISO 6520-1 classes: 1) cracks, 2) cavities, 3) inclusions, 4) lack of fusion/penetration, and 5) geometrical shape errors – *k*.
  - Each imperfection was examined by an experienced technician (A) and a trained novice (B) – *X2*.
  - Each of the examinators were given 14 samples - *n*.
  - The data (number of examination results by class):

Class	Laboratory						Total
	*L*1		*L*2		*L*3		
	A	*B*	*A*	*B*	*A*	*B*	
1	1	4	1	4	0	1	11
2	2	3	3	2	2	2	14
3	2	2	1	2	1	1	9
4	6	4	6	2	5	6	29
5	3	1	3	4	6	4	21
Total	14	14	14	14	14	14	84

Example 2 (Weld Imperfection) - Results

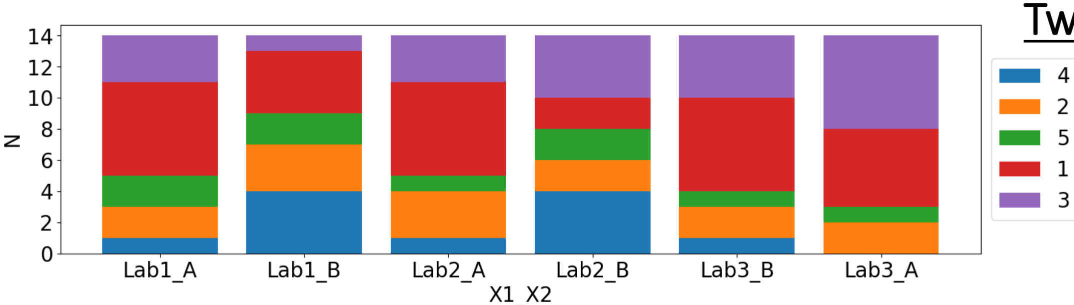


Two-way Anova (treating *k* as a continuous number)

	SSE	df	F	p-value
X1	9.809524	2.0	2.773692	0.068597
X2	7.440476	1.0	4.207664	0.043599
X1*X2	1.238095	2.0	0.350078	0.705735
Residual	137.928571	78.0	NaN	NaN

Because the data is nominal, there is no meaning for analyzing *k* as continuous number

k replacements: 5 with 3 and 1 with 4



Two-way Anova (*k* as a continuous number)

	SSE	df	F	p-value
X1	2.452381	2.0	0.663200	0.518081
X2	4.297619	1.0	2.324418	0.131402
X1*X2	4.023810	2.0	1.088162	0.341893
Residual	144.214286	78.0	NaN	NaN

Example 2 (Weld Imperfection) - Simulation Tool

Model Type:	Nominal
Number of Iterations:	10,000
α:	0.05
w:	0.3

I:	3
J:	2
K:	5
n:	14

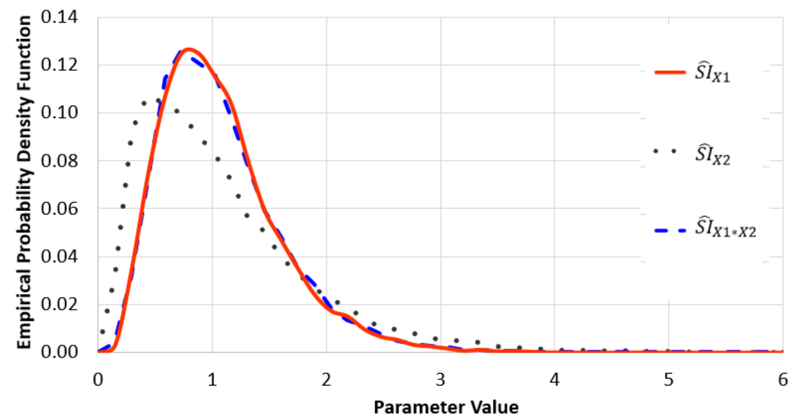
Two-way CATANOVA

P: 0.1310 ; 0.1667 ; 0.1071 ; 0.3452 ; 0.2500

Get File

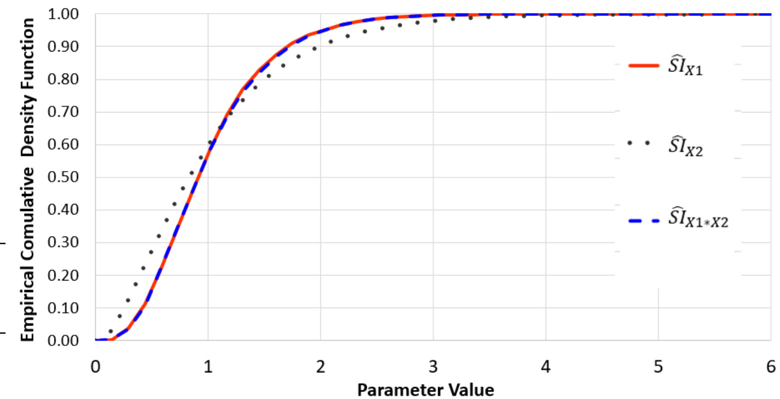
Run Model

File Name:
Data_welding.txt



Two-way CATANOVA with replication

	Variation Component	df	Significance index	P-value	SI_{crit}	Power
X1	0.0191	2	0.8337	0.34088	1.9384	0.43180
X2	0.0149	1	1.2969	0.74522	2.3719	0.50800
Interaction	0.0128	2	0.5558	0.24263	1.9384	0.42970
Within	0.9056	78				
Total	0.9524	83				



Power of the Test and Associated Risk of Incorrect Decision (Gadrich et al., 2024)

- The alternative hypotheses H_1 are that one or both the studied factors influence the probability vector $\boldsymbol{p}$:

$$\frac{E[\hat{C}^B_{X1}]}{df_{X1}} \geq \frac{V_T}{N} \text{ or } \frac{E\left[\hat{C}^B_{X2}\right]}{df_{X2}} \geq \frac{V_T}{N} \text{ or } \frac{\hat{C}^B_{X1*X2}}{df_{X1*X2}} \geq \frac{V_T}{N}$$

- The statistics $df_l\,\widehat{SI}_{Xl}$ (*l* = 1, 2) distributions, for nominal variables, are asymptotically approximated by the chi-square distributions $\chi^2_{df_l}$ with df_1=(*K*-1)(*I*-1) and df_2=(*K*-1)(*J*-1) respectively. They have the following parameters:

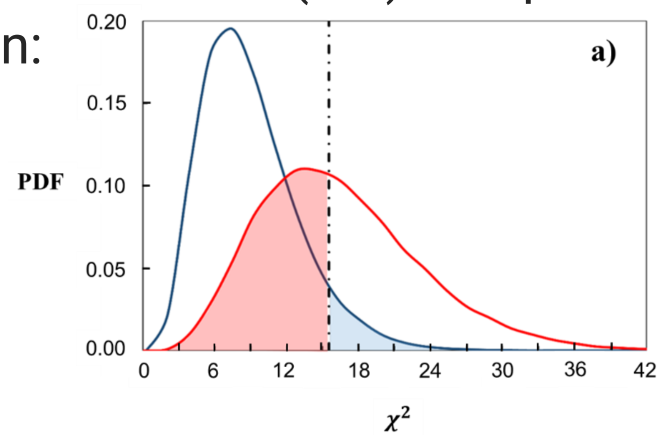
$$E\left[df_l\,\widehat{SI}_{Xl}\right] = df_l \; and \; VAR\left[df_l\,\widehat{SI}_{Xl}\right] = 2\,df_l$$

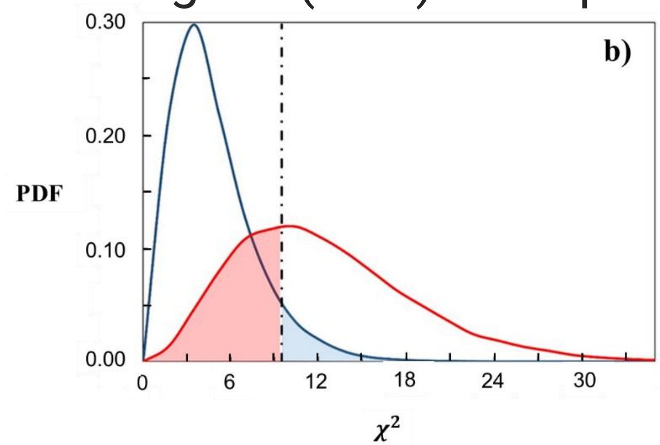
- The alternative hypothesis H_1 corresponds to the shifted distribution of the statistics $df_l\,\widehat{SI}_{Xl}$ which would be valid under the null hypothesis H_1. It is denoted further $df_l\,\widehat{SI}_{Xl,\lambda}$, where λ is the parameter of non-centrality (distribution shift). It has the following parameters:

$$E\left[df_l\,\widehat{SI}_{Xl,}\right] = df_l + \lambda \text{ and } VAR\left[df_l\,\widehat{SI}_{Xl,}\right] = 2\,df_l + 4\,\lambda$$

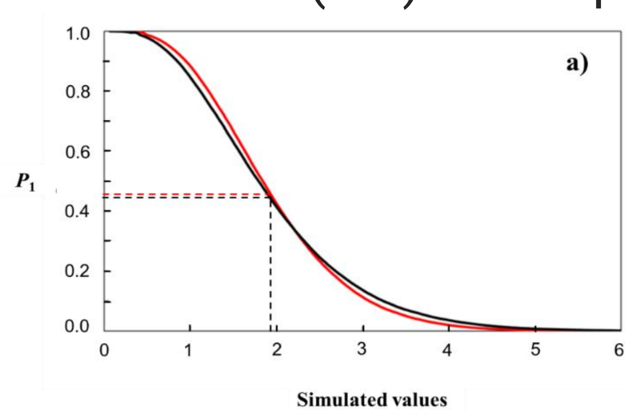
Power of the Test and Associated Risk of Incorrect Decision (Gadrich et al., 2024)

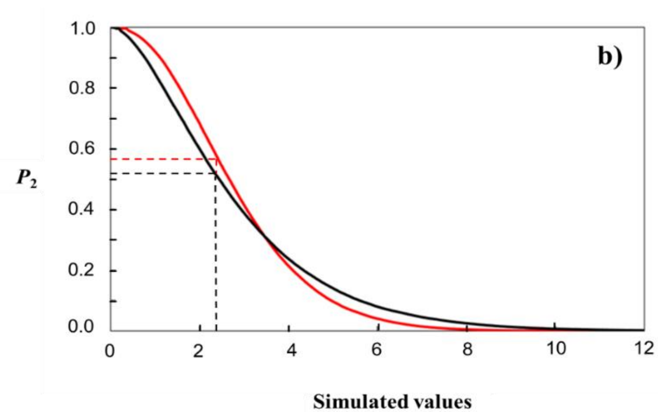
- Examples of modified (red) compares with original (blue) chi-square distribution:





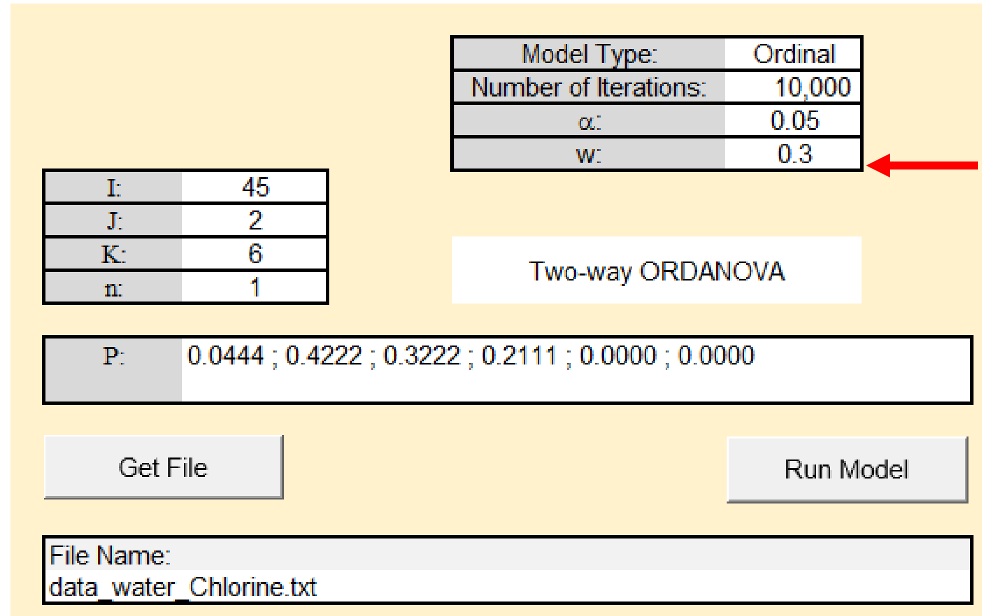
- Consider transformation on the $\widehat{SI}$ distribution such that $\widehat{SI}^{M}_{Xl}$=(1+λ/df_l)$\widehat{SI}_{Xl}$, and the power of the factor Xl is $P_l = 1 - CDF_{\widehat{SI}^{M}_{Xl}}(SI^{crit}_{Xl})$. The simulated values of modified distribution (red) is compared with the transformed *SI* distribution (black):





Power of the Test and Associated Risk of Incorrect Decision (Gadrich et al., 2024)

- Simulation tool was modified to account for the power calculation:
  - where w is the effect of the statistical sample size for the chi–square test: $w = 0.1$ is considered as a small effect, 0.3 – medium, and 0.5 – a large effect
  - In this example, As the sample size $N = IJ$ is equal for both factors $X1$ and $X2$, the same λ is applicable ($\lambda = w^2N$)



Two-way ORDANOVA without replication

	Variation Component	df	Significance index	P-value	SI_{crit}	Power
X1	0.2463	44	1.3601	0.00060	1.1830	0.10350
X2	0.0100	1	2.4236	0.08981	3.1093	0.28830
Interaction	n/a	n/a	n/a	n/a	n/a	n/a
Within	0.11	44				
Total	0.3663	89				

Reference

- Gadrich, Tamar, Yariv N. Marmor, Francesca Pennecchi, David Brynn Hibbert, Anastasia A. Semenova, and Ilya Kuselman. "Power of a test for assessing interlaboratory consensus of nominal and ordinal characteristics of a substance, material, or object." *Metrologia* 61(4):045004 (2024).
- Gadrich, Tamar, Ilya Kuselman, Francesca R. Pennecchi, D. Brynn Hibbert, Anastasia A. Semenova, Pui Sze Cheow, and Vladimir N. Naidenko. "Interlaboratory comparison of the intensity of drinking water odor and taste by two-way ordinal analysis of variation without replication." *Journal of Water and Health* 20, no. 6 (2022): 1005-1016.
- Gadrich, Tamar, and Yariv N. Marmor. "Two-way ORDANOVA: analyzing ordinal variation in a cross-balanced design." *Journal of Statistical Planning and Inference* 215 (2021): 330-343.
- Gadrich, Tamar, Ilya Kuselman, and Ivana Andrić. "Macroscopic examination of welds: interlaboratory comparison of nominal data." *SN Applied Sciences* 2, no. 12 (2020): 2168.

Summary and Future Research

There is a value for using ORDANOVA when the order of magnitude is unknown. Or when the power of test is needed.

CATANOVA is needed when the response is nominal.

The tool is easy to use (up to two parameter. Need to expand for more).

A guide for use is in preparation.

Thank You for your attention!

E-mail: *myariv@braude.ac.il*