

Manual Statistical Application - Proefcentrum Hoogstraten

Dieter Baets

V1 - October 2020

Introduction

Proefcentrum Hoogstraten vzw. (Belgium) is a research company specialized in greenhouse field trials and practical research for three crops: strawberry, sweet pepper and tomato. In order to ensure the quality of the research conclusions and repeatability a standardised statistical analysis tool was developed as a R-Shiny web application. With this Shiny application researchers of Proefcentrum Hoogstraten are able to perform frequently used analysis on their data.

This manual helps researchers to use the application as intended. For further questions or inquiries please contact me directly [dieterbaets\(at\)gmail\(dot\)com](mailto:dieterbaets(at)gmail(dot)com) or [dieter\(dot\)baets\(at\)proefcentrum\(dot\)be](mailto:dieter(dot)baets(at)proefcentrum(dot)be).

Install requirements

Before running the Shiny application on your own machine please ensure following prerequisites:

- at least R-version 4.0.0 (Arbor Day) or higher
- at least RStudio 1.2.5033 (Orange Blossom) or higher

First install the correct R-version on your own machine before installing R-Studio. This ensures that all R-related documents are recognised by RStudio as an IDE. After installing the software download the code on the github repo (github.com/dbaets/PCH_StatisticalApplicaton). Unzip the code in a folder of your own choosing, best to keep it on your own c:// harddrive for example under documents. Look for install folder and open the *install.Rproj* file. In this R project open the *R_installpackages.R* file and run the script by selecting the entire code and clicking on *Run* in the right top corner of the scripts pane. Wait until all necessary applications are installed.

Running Shiny application

After installing all the required packages you can now run the shiny application. Open the *statisticalapp* folder and open the *PCH_statisticalApplication.Rproj* file. By doing so you are ensuring that all necessary scripts for the application are located in the right place so the app can run smoothly. Open the *app.R* file and click on **Run App** in the right upper corner of the scripting pane. If all goes well you are now officially running the *PCH - Statistical application* and you can perform the necessary statistical analysis.

Performing statistical analysis using statistical application

Preparing data

In order to perform any analysis using the *PCH - statistical application* you need to put the data in to the correct format and on the correct named sheet otherwise the shiny app cannot load the data and will throw an error. Please download the correct template beforehand.

Loading data

First select a input file using the **Select file** button and select an input type using the radiobuttons on the *Input* pane before loading the file using the **Load file** button.

The screenshot shows a web application interface for loading data. It is divided into two main sections. The left section, titled 'Templates', contains two buttons: 'Botrytis' and 'General template'. Below this is a 'Select file' section with a 'Select file' button and a text input field containing '[1] ""'. A red arrow points to the 'Select file' button with the text '1. Select file to load data'. The right section, titled 'Load data', contains a 'Load file' button. A red arrow points to this button with the text '3. Load data'. Below the 'Load file' button is a 'Select input type' section with a list of radio buttons: 'Botrytis', 'Uitval', 'Drukplekgevoeligheid', 'Productie', 'Sortering', 'Aantal bloemtakken', 'Gewaslengte', and 'Algemene proef'. A red box surrounds this list, and a red arrow points to it with the text '2. choose input type of data'.

Figure 1: Input pane to load data for analysis

Running analysis

When the data is loaded correctly into the application you are automatically taken to the next pane called *statistics*. On this pane you only have to press the green button *Run Statistics*. If everything goes well you are redirected to the *output* pane where you can download the results using the green *Output Results* button. The output excel is redirected to the same folder where you loaded the input data from.

If the app jumps automatically to the next pane the analysis went without errors.

Interpreting output

output file

following tabs are generated in the outputfile:

- **inputdata:** the inputdata used to perform the analysis
- **teststatistic:** summary of the test statistic used to differentiate between different objects. The app chooses automatically between parametric and non parametric tests according to the prerequisites for each tests.
- **sigletters:** if the p-value of the test statistic is <0.05 you can interpret the results using a post-hoc test. For each object a letter or p-value is placed in a table with the summary statistics. If the letters are different from each other the different objects are different from each other.
- **summary:** table with a summary of the data. This summary is ideal as a starting point to make graphs.

reporting results

When writing a report please include following text under the title *statistical analysis*:

All results were statistically confirmed by the statistical software R (version 4.0.0, 24/04/2020) (R Core Team, 2019), with RStudio (RStudio Team, 2016) as an integrated development environment (IDE). In order to analyse the results following packages were used: tidyverse (Wickham, 2017), multcomp (Hothorn et al., 2008) lsmeans (Lenth, 2016), rcompanion (Mangiafico, 2019) and car (Fox and Weisberg, 2019). The statistical tests have the goal to evaluate the difference between the different objects. When the data and residues were normally distributed an analysis of variance (ANOVA) was used. In the case both assumptions were not fulfilled a Kruskal-Wallis test was used as a non-parametric test. With each result it is clearly mentioned which test is used for the analysis. When a significant difference occurred a Tukey post-hoc test between the different objects was performed. A significant result is shown by the use of different letters.

with following references:

Fox, J. en Weisberg, S., 2019. An {R} companion to applied regression, Third Edition. Thousand Oaks.

Hothorn, T., Bretz, F. en Westfall, P., 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50(3)

Lenth, R., 2016. Least-squares means: the R package lsmeans. *Journal of statistical software*, 69(1)

Mangiafico, S., 2020. Rcompanion: functions to support extension education program evaluation. Package version: 2.3.0 R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: www.R-project.org

RStudio Team, 2016. RStudio: Integrated Development for R. RStudio Inc., USA, Boston. URL: www.rstudio.com Wickham, H., 2017. Tidyverse: easily install and load the ‘Tidyverse’. Package version: 1.2.1

The results of an anova can be best reported in the following format:

$$F(5, 18) = 1.288, p = 0.312$$

- **5:** DF object, degrees of freedom at the level of the objects
- **18:** DF residuals, degrees of freedom of the residuals
- **1.288:** F-value in *TestStatistic* table
- **0.312:** p-value in *TestStatistic* table

The results of an Kruskal-Wallis should be reported in following format:

$$\chi^2(3) = 5.889, p = 0.117$$

- **3:** DF, degrees of freedom
- **5.889:** F-value in *TestStatistic* table
- **0117:** p-value in *TestStatistic* table

Types of inputdata

Botrytis

excelsheetname: *statistiek*

productie

excelsheetname: *statistiek_productie*

	A	B	C	D	E	F	G	H	I	J	K
1	datum	objectnr	objectnaam	herhaling	aantal_stuks	aantasting_geen	aantasting_25	aantasting_50	aantasting_75	aantasting_100	
2	datum1	1	Murano	1	52	49	2	0	0	1	
3	datum1	1	Murano	2	52	50	1	0	0	1	
4	datum1	1	Murano	3	52	47	2	1	1	1	
5	datum1	1	Murano	4	52	48	2	2	0	0	
6	datum1	2	Verity	1	49	45	4	0	0	0	
7	datum1	2	Verity	2	51	45	3	2	1	0	
8	datum1	2	Verity	3	39	37	2	0	0	0	
9	datum1	2	Verity	4	52	50	1	0	1	0	
10	datum1	3	M. Champio	1	34	31	1	1	1	0	
11	datum1	3	M. Champio	2	23	21	0	0	1	1	
12	datum1	3	M. Champio	3	10	8	2	0	0	0	
13	datum1	3	M. Champio	4	35	34	0	0	1	0	
14	datum1	4	08-06-10	1	52	48	1	0	3	0	
15	datum1	4	08-06-10	2	52	49	2	0	0	1	
16	datum1	4	08-06-10	3	52	51	1	0	0	0	
17	datum1	4	08-06-10	4	52	50	1	0	1	0	
18	datum1	5	CIV 621	1	52	47	1	2	2	0	
19	datum1	5	CIV 621	2	52	51	1	0	0	0	
20	datum1	5	CIV 621	3	52	46	3	1	1	1	
21	datum1	5	CIV 621	4	52	49	1	1	1	0	
22	datum2	1	Murano	1	52	43	8	0	1	0	
23	datum3	1	Murano	2	52	42	7	2	1	0	
24	datum4	1	Murano	3	52	47	4	0	1	0	
25	datum5	1	Murano	4	52	44	4	2	2	0	
26	datum6	2	Verity	1	52	45	3	2	2	0	
27	datum7	2	Verity	2	52	48	3	1	0	0	
28	datum8	2	Verity	3	52	48	2	1	1	0	
29	datum9	2	Verity	4	52	47	4	1	0	0	
30	datum10	3	M. Champio	1	52	51	1	0	0	0	
31	datum11	3	M. Champio	2	52	46	4	1	1	0	
32	datum12	3	M. Champio	3	52	48	0	2	2	0	
33	datum13	3	M. Champio	4	52	44	5	0	2	0	

Figure 2: botrytis

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	objectnr	objectnaam	herhaling	kg_pl	kg_m2									
2	1	Murano	1	1.871867	9.359333									
3	1	Murano	2	1.679633	8.398167									
4	1	Murano	3	1.913467	9.567333									
5	1	Murano	4	1.5222	7.611									
6	2	Verity	1	1.3855	6.9275									
7	2	Verity	2	1.524367	7.621833									
8	2	Verity	3	1.316667	6.583333									
9	2	Verity	4	1.276033	6.380167									
10	3	M. Champ	1	2.135067	10.67533									
11	3	M. Champ	2	2.165767	10.82883									
12	3	M. Champ	3	2.313533	11.56767									
13	3	M. Champ	4	2.0618	10.309									
14	4	Bravura	1	2.001133	10.00567									
15	4	Bravura	2	1.854367	9.271833									
16	4	Bravura	3	1.897033	9.485167									
17	4	Bravura	4	1.8579	9.2895									
18	5	Cantus	1	1.757233	8.786167									
19	5	Cantus	2	1.564833	7.824167									
20	5	Cantus	3	1.5042	7.521									
21	5	Cantus	4	1.488367	7.441833									
22														
23														
24														
25														
26														
27														
28														
29														
30														

middenoogstdatum oogstverloop per sortering uitval Rendement **statistiek_productie** statistiek_sortering +

Figure 3: productie

sortering

exelsheetname: *statistiek_sortering*

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	objectnr	objectnaai	herhaling	groot2A	grootA	klein	misvormd	rot						
2	1	Murano	1	5.101786	0.455357	4.282143	0.155893	0.032679						
3	1	Murano	2	4.444828	0.406897	3.660345	0.122241	0.053448						
4	1	Murano	3	5.005085	0.40339	4.145763	0.120678	0.054576						
5	1	Murano	4	4.203571	0.176786	3.553571	0.146786	0.073929						
6	2	Verity	1	4.381667	0.575	1.853333	0.106667	0.010833						
7	2	Verity	2	4.96	0.455	2.135	0.043167	0.028667						
8	2	Verity	3	3.905	0.543333	2.05	0.0775	0.0075						
9	2	Verity	4	4.136667	0.555	1.615	0.0605	0.013						
10	3	M. Champ	1	4.381667	0.771667	5.44	0.062	0.02						
11	3	M. Champ	2	3.875	0.786667	6.046667	0.100833	0.019667						
12	3	M. Champ	3	4.055	1.015	6.349667	0.102833	0.045167						
13	3	M. Champ	4	3.968421	0.696491	6.068421	0.03614	0.082105						
14	4	Bravura	1	5.96	0.96	2.91	0.154167	0.0215						
15	4	Bravura	2	5.248333	0.88	2.871667	0.1605	0.111333						
16	4	Bravura	3	5.601667	0.908333	2.841667	0.112833	0.020667						
17	4	Bravura	4	5.119167	1.13	2.941667	0.088833	0.009833						
18	5	Cantus	1	5.155	0.815	2.525	0.251667	0.0395						
19	5	Cantus	2	4.421667	0.798333	2.32	0.2355	0.048667						
20	5	Cantus	3	4.186667	0.765	2.323333	0.209667	0.036333						
21	5	Cantus	4	4.15	0.658621	2.567241	0.276379	0.046207						
22														
23														
24														
25														
26														
27														
28														
29														
30														
	...	middenoogstdatum		oogstverloop per sortering		uitval	Rendement	statistiek_productie					statistiek_sortering	

Figure 4: sortering

algemene proef

exelsheetname: *statistiek*

	A	B	C	D	E	F	G	H
1	objectnr	objectnaam	herhaling	meting				
2	1	a	1	10				
3	1	a	2	2				
4	1	a	3	3				
5	1	a	4	4				
6	2	b	1	4.1				
7	2	b	2	8				
8	2	b	3	2.1				
9	2	b	4	3.1				
10								
11								
12								
13								
14								
15								

Figure 5: sortering