



(RESEARCH ARTICLE)



Statistical analysis: An approach by applying nonparametric tests to a Brazilian commercial egg production line

Carvalho Cleginaldo Pereira de *

University of the State of São Paulo – Faculty of Engineering of Guaratinguetá, Industrial Engineering Department, Avenida Dr. Ariberto Pereira da Cunha 333, Pedregulho, Guaratinguetá, São Paulo, Brazil, Zip Code:12.516-410.

Publication history: Received on 31 July 2020; revised on 10 August 2020; accepted on 11 August 2020

Article DOI: <https://doi.org/10.30574/gjeta.2020.4.2.0049>

Abstract

The present study refers to the statistical analyzes applied to the production of commercial eggs carried out in a Poultry Company in the interior of the State of São Paulo, Brazil. Currently, the concern with consumer demands is growing in the food sector, which makes companies seek to increase the level of quality of their products. In this study, the application of statistical analysis was applied to the data obtained from the software that integrates a crack detection accessory, the Crack Detector of 12 egg production lines for 6 months. It was concluded that, when performing the nonparametric tests, Spearman and Kruskal tests, the data are correlated and there is a difference between the analyzed cracks among the quality control equipments. As result of this research work a calibration plan for the measurement equipments was suggested, in order to increase the reliability in the quality control inspection process, reducing the amount of non-conform products delivered to the customers market.

Keywords: ANOVA; Nonparametric tests; Egg production; Crack Detector; Spearman tests; Kruskal tests

1. Introduction

Statistical analyzes are extremely important for the validation of the results obtained in studies, and one of the central objectives of science is to identify and justify the best explanation for a given phenomenon [1], so it becomes an important tool for companies to assess the impacts of certain variables on the process of manufacturing their product or providing the final service.

Consumers in the food industry are influenced by the added value of the product. In Brazil, a large part of egg production is sold on the domestic market, however the sector is adapting, more and more, to increase exports [AVISITE. Brazilian Poultry Union [UBA], https://avisite.com.br/legislacao/anexos/protocolo_de_boas_praticas_de_producao_de_ovos.pdf, Last accessed on: 22/05/2019]. Brazilians consume 192 eggs per capita and the export of this same product caused the sector to generate US \$ 7, 490 million in 2017 [AHORA DO OVO. Brazilian egg consumption is 192 eggs per capita increaisng 1% in Brazil, <http://www.ahoradoovo.com.br/no-mundo-do-ovo/noticias/?id=1320%7Cbrasileiro-ja-consome-192-ovos-per-capita-consumo-cresce-1-no-brasil>, Last accessed on: 20/06/2019, BRAZILIAN ANIMAL PROTEIN ASSOCIATION [ABPA], Annual Report 2018, www.abpa-br.com.br Last accessed on: 15/05/2019]. This leads to a greater concern of the poultry industry with the quality of its products and, applying statistical studies in order to adapt to meet the demands of consumers, is essential to ensure competitiveness [2].

In poultry farms, there is still a lack of technologies to automate and improve quality processes in egg production. Quality control at these companies is still essentially based on human inspection, which occurs at low rates (12 eggs per person) and is subject to human error [3].

* Corresponding author

E-mail: cleginaldopcarvalho@hotmail.com

Thus, in July 1973, Crack Detector was invented by George N. Bliss, which is an accessory that can be integrated into production lines and allows you to recognize cracks by the vibration that differs from a cracked egg to an egg in marketing conditions, and is currently widely used in poultry companies [4]. Crack Detector, in addition to preventing cracked eggs from reaching consumers, provides data with which it is possible to work by carrying out statistical analysis in order to increase the level of product quality.

This study aims to improve the knowledge about statistical analysis and its importance as a tool to increase the production quality index in poultry companies. In addition, it also aims to verify whether Crack Detectors read the same or different and whether they need repairs to deliver better products to consumers.

2. Theoretical Background

2.1. Statistics Analysis

Statistical methods are efficient direct approaches that offer objectivity and accuracy, which leads to giving more importance to facts than to abstract concepts [5]. Statistics provides tools that formalize and standardize procedures to obtain certain conclusions [6].

In studies involving statistical analysis, the methods and tests that best suit the analysis must be chosen to achieve the objective of the work.

2.1.1. Hypothesis Test

In order to reach an objective decision, it is necessary to carry out the hypothesis test, establishing hypotheses that can be rejected, reviewed or accepted.

The hypothesis test examines two hypotheses, one null, which is the hypothesis to be tested and an alternative. And, according to the sample data, the test determines whether to accept or reject the null hypothesis [7].

2.1.2. Nonparametric Tests

Data that is “free of distribution” is called non-parametric data and does not assume that the data analyzed was extracted from a normally distributed population. And, for these data, non-parametric tests are used that are identified as “ranking tests” that can be used with scores, which are not numerically accurate, but which are indicated by “rankings” in ranking.

These tests are commonly used with small samples, which makes it easier to use them in a pilot study or in studies that are not possible to have large samples [6].

2.1.3. Spearman's Ordinal Correlation Test

The Spearman Ordinal Correlation test has as main requirement that the paired data of the variables under study originate from a simple random sample of the population, therefore, the Spearman coefficient (r_s) is used to test the association of the variables in the population. Thus, using the r_s value, it is tested whether the variables are independent or whether there is an association between them [8].

2.1.4. Kruskal –Wallis Test

The Kruskal-Wallis test, developed in 1952 by William H. Kruskal and Wilson A. Wallis, is used to determine whether three or more independent groups are from different populations or not, as well as “The one-way analysis of variance” (ANOVA). However, ANOVA is used for parametric data, with populations normally distributed, and Kruskal-Wallis for nonparametric data, selected at random.

In addition, the Kruskal-Wallis test is also used to determine whether three or more independent groups are equal or differ in some variable of interest regarding the ordinal level of data, the interval level or the relationship levels of the available data [9].

3. Material and methods

The present work is a case study and was carried out by statistically analyzing the data from the 12-line Crack Detector software, over a period of 6 months. The average of each Crack Detector for each month was also compared, thus making a month-to-month analysis.

A coleta Data collection was carried out in a poultry company located in the interior of the state of São Paulo, Brazil.

Initially, an outlier test was performed to remove values outside the usual data distribution. Soon after, the Graphical Summary test was performed analyzing the P-Value of each production line.

As this is a study with a small sample size, non-parametric tests were used. To verify whether the production lines differed from each other, the Kruskal-Wallis Test was performed and to verify if there was a correlation, the Spearman Ordinal Correlation Test was performed. All analyzes were made from the graphics and information obtained in the Minitab statistical software version 2017.

4. Results and discussion

The equipment under study, Crack Detector, has an internal defect detection module, based on the vibrations emitted when the egg passes through the device, which allows checking whether an egg is broken or not. This detection module is important, as it replaces the human work of visual inspection of defects, allowing the automation of this task, producing agility, efficiency in the egg screening stage and standardization in the classification [10].

We have a company with a production plant with 12 production lines, at the end of each line, a Crack Detector identifies if there are defects in the eggs. Each equipment works in the same way.

The work carried out by Crack Detector identifies four common types of defects that occur in chicken eggs: dirt, blood stains, cracks and leaking of yolk. In this work, we analyzed the data collected from the 12 production lines during 6 months to identify these errors, as shown in Figures 1 and 2.



Figure 1 Defect – dirty shells



Figure 2 Defect– shell with cracks

The results obtained from the automatic visual inspection performed by Crack Detector can be seen in the series of data commented in the sequency.

The first test to be done is the outlier test and its data are contained in Figure 3.

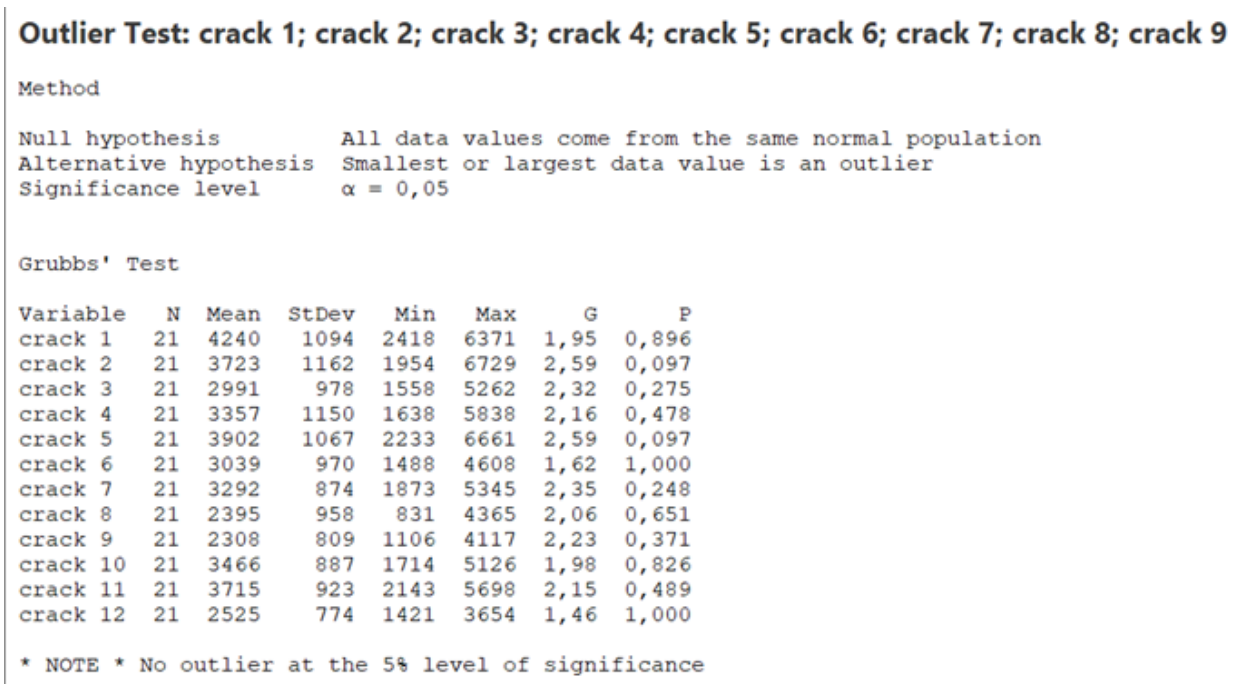


Figure 3 Outlier Test

crack 1	crack 2	crack 3	crack 4	crack 5	crack 6	crack 7	crack 8	crack 9	crack 10	crack 11	crack 12
4768	4827	4209	4670	5074	4289	4361	3634	3656	4057	4248	3174
4912	4490	3219	3434	3872	3311	3195	2805	2831	4066	4165	1460
3599	3160	2319	2518	3169	2387	2726	2726	1899	1714	2588	1460
3587	3128	2270	2387	2999	2370	2955	2340	2294	3374	3061	2165
4252	3636	2879	3041	3664	2740	2892	2052	1945	2787	2825	1881
6371	6729	5262	5477	4608	5345	4365	4117	5126	5698	3622	
3420	2593	2330	2460	2978	2175	2252	1788	1815	2323	2639	1792
2418	1954	1558	1638	2233	1586	1873	1776	1335	2199	2143	1421
3500	2809	2209	3680	3299	2454	2723	3267	2779	3268	3656	2391
5805	5026	4165	4445	4382	4639	3538	3301	4852	5336	3591	
5358	4570	3823	3967	4136	3828	4311	3110	2852	4270	4279	2746
4093	4065	3179	4022	4142	3932	3871	3397	3026	3928	3902	2394
4365	3835	3284	3318	4156	3297	3411	2433	2271	3245	3431	2106
2903	2402	2075	2069	2567	2363	2458	1484	1749	2516	2553	1759
3229	3265	2476	2746	3944	2975	3132	2602	2548	3692	3788	3387
5606	4831	3897	3764	4638	3754	3923	2203	2250	3935	4580	3287
3144	2580	1635	2010	2758	2147	2454	1490	1648	2724	2991	1991
3480	2820	2348	2125	3604	1488	2496	831	1106	2840	3545	2480
4064	3280	2476	5838	4477	4016	3057	968	1293	3782	4009	2845
4173	3105	2892	2774	3847	1845	3335	992	1340	3873	4169	3654
5995	5073	4303	4123	3796	3876	3714	2496	2416	4215	4401	3427

Figure 4 Variance of the Outliers

If the outlier test found any data to be an outlier, that data will be replaced by the average of the other data obtained.

The data from the Graphical Summary Test for the month of September are contained in Figure 5.

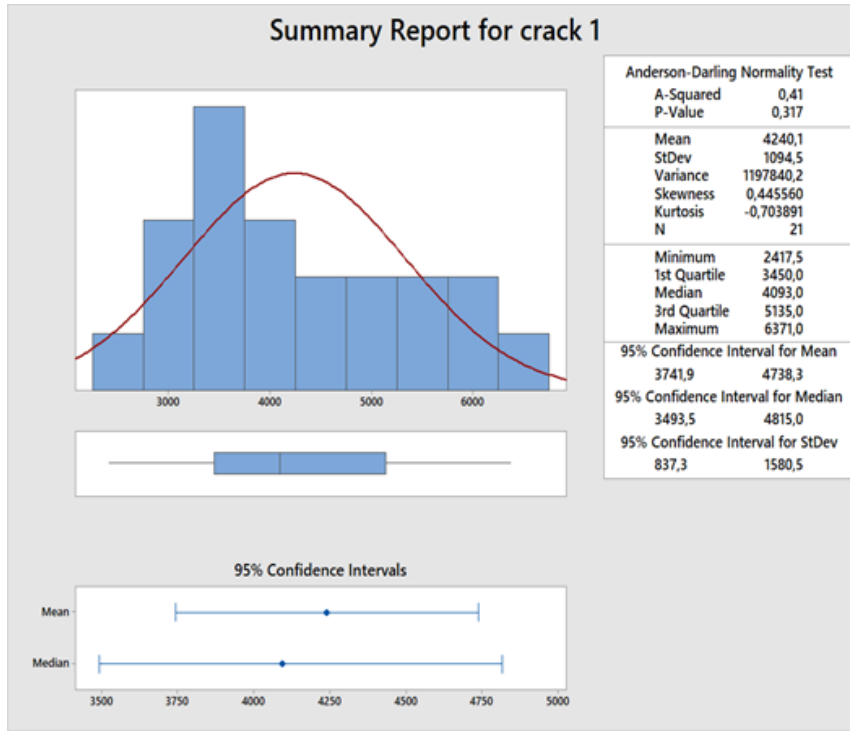


Figure 5 Graphic Summary

In Figure 5 it can be seen that the Summary Report provides several important data for several statistical analyzes, however, in the case in question, the P-Value (P-Value) was analyzed, which, despite having given a value greater than 0.05 which indicates that the data follow a normal distribution, due to the low amount of data analyzed (21 for each Crack Detector face), it is necessary to assume that this indication is a false positive, since the error β , for this amount of data, is very high, so there is a need to treat the data as nonparametric data.

The data for the Kruskal-Wallis test for the month of September are contained in Figure 6.

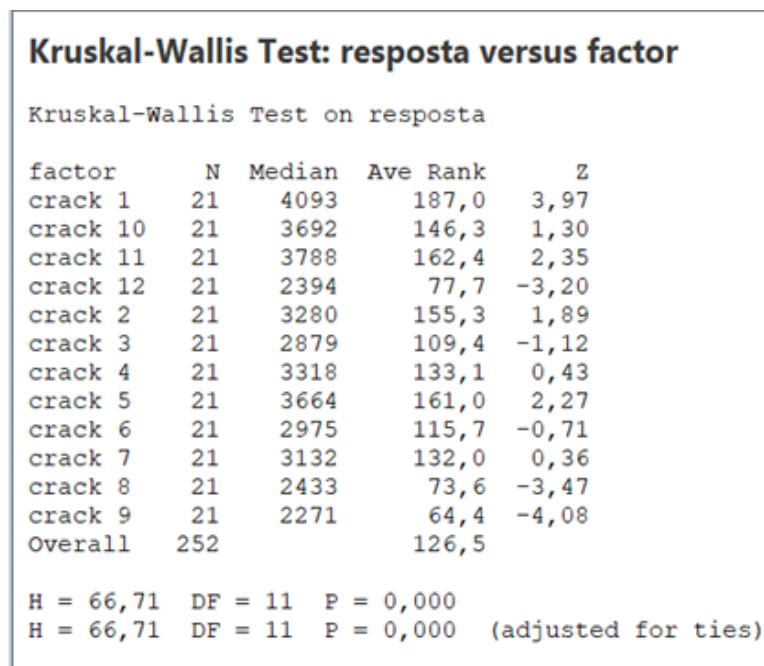


Figure 6 Kruskal - Wallis Test

When assuming that the data follow a non-parametric distribution, the Kruskal-Wallis Test is performed in order to replace the “The one-way analysis of variance” (ANOVA). The hypotheses tested in the Kruskal-Wallis test, at a level of 5% significance, are:

H0: the populations from which the samples came are identical;

H1: the populations from which the samples came are not identical.

The P value provided in Figure 6 indicates that, at the 5% significance level, there is sufficient statistical evidence to reject the hypothesis (H0) that there are no significant differences between the Crack Detectors installed in the 12 different egg production lines.

The data from the Spearman Correlation Test for the month of September are contained in Figures 7 and 8.

Spearman Rho: crack 1; crack 2; crack 3; crack 4; crack 5; crack 6; crack 7; crack 8; ...

	crack 1	crack 2	crack 3	crack 4	crack 5	crack 6	crack 7	crack 8
crack 2	0,945 0,000							
crack 3	0,935 0,000	0,952 0,000						
crack 4	0,800 0,000	0,839 0,000	0,810 0,000					
crack 5	0,601 0,004	0,679 0,001	0,709 0,000	0,720 0,000				
crack 6	0,770 0,000	0,881 0,000	0,797 0,000	0,930 0,000	0,653 0,001			
crack 7	0,883 0,000	0,923 0,000	0,924 0,000	0,825 0,000	0,720 0,000	0,845 0,000		
crack 8	0,564 0,008	0,652 0,001	0,563 0,008	0,592 0,005	0,300 0,186	0,709 0,000	0,669 0,001	
crack 9	0,627 0,002	0,718 0,000	0,652 0,001	0,631 0,002	0,381 0,089	0,756 0,000	0,751 0,000	0,958 0,000
crack 10	0,835 0,000	0,856 0,000	0,853 0,000	0,806 0,000	0,603 0,004	0,788 0,000	0,901 0,000	0,587 0,005
crack 11	0,842 0,000	0,847 0,000	0,865 0,000	0,817 0,000	0,640 0,002	0,747 0,000	0,890 0,000	0,506 0,019
crack 12	0,567	0,592	0,669	0,626	0,507	0,510	0,718	0,247

Figure 7 Spearman Correlation Test

	crack 9	crack 10	crack 11
crack 10	0,713 0,000		
crack 11	0,605 0,004	0,958 0,000	
crack 12	0,353 0,116	0,739 0,000	0,835 0,000

Cell Contents: Spearman rho
P-Value

Figure 8 Spearman Correlation for all data

After the result obtained by the Kruskal-Wallis test, the Spearman Correlation Test was performed with the aim of analyzing the interaction between Crack Detectors. The calculation of the Spearman coefficient was very effective, since it is very sensitive to outliers in continuous variables - which confirmed, once again, the thesis that there are differences between the samples and proves the Kruskal test month-to-month analysis.

The same tests and analyzes were used when comparing the average of the months, also obtaining results similar to those obtained in the analysis of Crack Detectors within each month.

The Outlier Test and its data are contained in Figure 9.

Outlier Test: Crack 1; Crack 2; Crack 3; Crack 4; Crack 5; Crack 6; Crack 7; Crack 8; Crack 9

Method

Null hypothesis All data values come from the same normal population
 Alternative hypothesis Smallest or largest data value is an outlier
 Significance level $\alpha = 0,05$

Grubbs' Test

Variable	N	Mean	StDev	Min	Max	G	P
Crack 1	6	3881	447	3337	4315	1,22	1,000
Crack 2	6	3397	437	2737	3789	1,51	0,555
Crack 3	6	2710	352	2313	3243	1,51	0,550
Crack 4	6	2901	436	2220	3357	1,56	0,453
Crack 5	6	2638,8	202,1	2447,5	3014,7	1,86	0,069
Crack 6	6	2779	316	2267	3171	1,62	0,354
Crack 7	6	2945	325	2446	3292	1,54	0,506
Crack 8	6	2115	291	1570	2395	1,87	0,059
Crack 9	6	1968	459	1374	2475	1,29	1,000
Crack 10	6	2882	621	2112	3515	1,24	1,000
Crack 11	6	3189	675	2446	4052	1,28	1,000
Crack 12	6	2514,4	230,9	2229,1	2853,5	1,47	0,642

* NOTE * No outlier at the 5% level of significance

Figure 9 Outlier Test month by month

Figure 10 shows the averages of each Crack in all the months analyzed.

Average value of the outlier column (column 5): 2638,82.

Crack 1	Crack 2	Crack 3	Crack 4	Crack 5	Crack 6	Crack 7	Crack 8	Crack 9	Crack 10	Crack 11	Crack 12
4257,36	3722,83	2990,86	3357,38	2638,82	3039,24	3291,52	2395,02	2308,05	3466,00	3714,55	2525,36
4182,00	3733,17	3243,45	3212,21	3014,69	3171,19	3245,07	2173,64	2158,43	3514,52	3575,62	2490,52
4314,95	3788,71	2698,71	2960,67	2620,36	2788,31	3093,21	2199,02	2055,93	3317,43	4052,10	2679,19
3842,38	3387,48	2601,48	3117,71	2633,88	2665,07	2849,98	2053,83	1373,91	2519,05	2773,88	2853,50
3349,71	3014,81	2313,00	2219,67	2447,48	2742,29	2747,29	1570,14	1435,10	2111,81	2446,10	2229,10
3336,69	2737,41	2413,86	2541,02	2477,71	2266,81	2445,52	2295,71	2474,95	2361,83	2570,52	2308,64

Figure 10 Variance of the Outliers month-to-month

The Graphical Summary test data for the month-to-month analysis is contained in Figure 11.

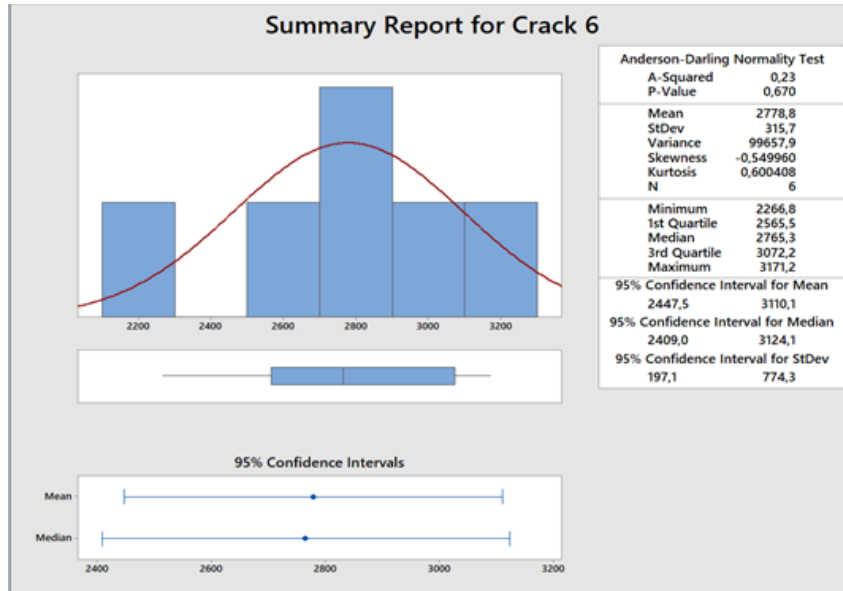


Figure 11 Graphic Summary month-to-month

In the same way as analyzed for the month of September (Figure 5), Figure 11 provides the Summary Report among the observed months, analysis month by month. As well as performed for a single month, the P-Value (P-Value) of the sample, which, despite having also given a value greater than 0.05 indicating that the data follows a normal distribution, it was necessary to assume that this indication is a false positive due to the β error, with the need to treat the data as nonparametric data.

The data from the Kruskal-Wallis test for the month-to-month analysis are contained in Figure 12.

Kruskal-Wallis Test: resposta versus factor

Kruskal-Wallis Test on resposta

factor	N	Median	Ave Rank	Z
crack 1	6	4012	66,0	3,61
crack 10	6	2918	38,0	0,18
crack 11	6	3175	47,7	1,37
crack 12	6	2508	26,0	-1,28
crack 2	6	3555	57,0	2,51
crack 3	6	2650	33,2	-0,41
crack 4	6	3039	40,5	0,49
crack 5	6	2627	31,3	-0,63
crack 6	6	2765	37,3	0,10
crack 7	6	2972	43,2	0,81
crack 8	6	2186	9,2	-3,34
crack 9	6	2107	8,7	-3,40
Overall	72		36,5	

H = 43,13 DF = 11 P = 0,000

Figure 12 Kruskal - Wallis Test month-to-month

Analogously to the test performed in Figure 6, the Kruskal-Wallis Test was performed in order to replace the “The one-way analysis of variance” (ANOVA) for the month-to-month analysis. The hypotheses tested in the Kruskal-Wallis test month by month are the same as those tested for a single month, namely:

H0: the populations from which the samples came are identical;

H1: the populations from which the samples came are not identical.

The P value provided in Figure 12 provides the same result as found previously, that, at the 5% level of significance, there is sufficient statistical evidence to reject the hypothesis (H0) that there are no significant differences between the Crack Detectors installed in the different 12 egg production lines.

The data from the Spearman Correlation Test of the month-to-month analysis are contained in Figures 13 and 14.

Spearman Rho: Crack 1; Crack 2; Crack 3; Crack 4; Crack 5; Crack 6; Crack 7; Crack 8; ...

	Crack 1	Crack 2	Crack 3	Crack 4	Crack 5	Crack 6	Crack 7	Crack 8
Crack 2	0,943 0,005							
Crack 3	0,714 0,111	0,771 0,072						
Crack 4	0,600 0,208	0,543 0,266	0,886 0,019					
Crack 5	0,543 0,266	0,600 0,208	0,943 0,005	0,943 0,005				
Crack 6	0,714 0,111	0,771 0,072	0,829 0,042	0,657 0,156	0,714 0,111			
Crack 7	0,829 0,042	0,771 0,072	0,886 0,019	0,886 0,019	0,829 0,042	0,886 0,019		
Crack 8	0,314 0,544	0,143 0,787	0,429 0,397	0,486 0,329	0,314 0,544	0,143 0,787	0,371 0,468	
Crack 9	-0,086 0,872	-0,143 0,787	0,257 0,623	0,200 0,704	0,143 0,787	0,086 0,872	0,086 0,872	0,829 0,042
Crack 10	0,714 0,111	0,771 0,072	1,000 *	0,886 0,019	0,943 0,005	0,829 0,042	0,886 0,019	0,429 0,397
Crack 11	0,943 0,005	0,886 0,019	0,771 0,072	0,657 0,156	0,600 0,208	0,600 0,208	0,771 0,072	0,543 0,266

Figure 13 Spearman Correlation Test month-to-month

Crack 12	0,600 0,208	0,543 0,266	0,429 0,397	0,543 0,266	0,486 0,329	0,086 0,872	0,429 0,397	0,143 0,787
Crack 10	Crack 9 0,257 0,623	Crack 10 0,771 0,072	Crack 11 0,657 0,156					
Crack 11	0,143 0,787	0,771 0,072						
Crack 12	-0,371 0,468	0,429 0,397	0,657 0,156					

Cell Contents: Spearman rho
P-Value

Figure 14 Spearman Correlation Test month by month for the whole data

In the month-to-month analysis, the objective is to assess whether, among the analyzed months, any of them stands out, positively or negatively, thus leading to a rejection of the H0 hypothesis of the Kruskal-Wallis test indicating that,

at the 5% significance level, there is statistical evidence to affirm that there are significant differences between the 6 months analyzed.

5. Conclusion

With this study, it was initially sought to improve knowledge about Statistical Analysis applied to the quality of commercial egg production, showing the importance of applying statistical tests in order to verify and ensure the quality of the final product.

Analyzing the data that the Crack Detector Software provided in the six months, it was possible to conclude, through statistical tests, that the equipment and the months analyzed differ from each other, showing that a limitation of the study for data validation is based on calibration, which should have a standard for all equipment.

The present research work, although only using Crack Detector as equipment, presents as a practical implication that statistical tests, such as those presented, are important tools in support of quality assurance.

It is recommended for future work, the application of different statistical tools for the same data, followed by a critical comparison of the results, as well as the application of the same tests in other models of production systems, in order to validate the results of this work.

Compliance with ethical standards

Acknowledgments

The author wish to acknowledge the University of the State of São Paulo, Faculty of Engineering of Guaratinguetá, for its assistance during the research period.

Disclosure of conflict of interest

The author of the article claims that he has no conflict of interest.

References

- [1] Bausman W and Halina M. (2018). Not null enough: pseudo-null hypotheses in community ecology and comparative psychology. *Biology & Philosophy*, Springer Nature B.V, 33-30.
- [2] Detofol DF, Rauta J and Winck CA. (2018). Logística aplicada na produção de ovos comerciais. *Revista Visão: Gestão Organizacional*, Caçador – Santa Catarina, Brazil, 7(1), 52-69.
- [3] Machado DS. (2009). Sistema de Inspeção Visual Automática Aplicado ao Controle de Qualidade de Ovos em Linhas de Produção. Dissertação de Mestrado, Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, Brazil.
- [4] Diamond International Corporation. (1973). Crack Detector, 3, 744,399.
- [5] Sena TTO, et al. (2017). Controle estatístico do processo no setor de embalagem: um estudo de caso. *Revista da Jornada de Pós-Graduação e Pesquisa, Congrega - Urcamp*, Brazil, 1(1), 15-25.
- [6] Siegel S and Castellan JN. (2006). *Estatística não-paramétrica para ciências do comportamento*, segunda edição. São Paulo: ARTMED.
- [7] Costa Neto PLO. (1977). *Estatística*. Editora Edgard Blucher.
- [8] Marques AB, et al. (2018). A relação entre competências para inovar e competitividade na indústria de eletromédicos no Brasil. *Revista Gestão & Produção*, São Carlos, Brazil, 25(3), 545-556.
- [9] Santos M, et al. (2016). Análise da demanda dos itens de alto valor agregado: aplicação do teste de Kruskal - Wallis em uma multinacional do setor de óleo e gás. *Revista Produção Industrial & Serviços*, Universidade Estadual de Maringá – UEM Campus Sede - Paraná – Brasil, 3(1), 01-14.
- [10] Montgomery DC. (2009). *Introduction to Statistical Quality Control*, Six edition. New York: John Wiley and Sons.

- [11] Vanegas LH and Paula GA. (2016). Log-symmetric distributions: statistical properties and parameter estimation. Brazilian Journal of Probability and Statistics, São Paulo, Brazil, 30(2), 196-220.

How to cite this article

Carvalho CP. (2020). Statistical analysis: An approach by applying nonparametric tests to a brazilian commercial egg production line. Global Journal of Engineering and Technology Advances, 4(2), 11-21.
