

Analysis of cancer data from the Central Bureau of Statistics using Two-Way ANOVA

Ola Hasan Al-Lahaibat¹, Hadeel Salim Alkutubi², Noor Sabah Murad³, Nabaa Mohammed Alshamary⁴

¹Faculty of Pharmacy, Jabir Ibn Hayyan University for Medical and Pharmaceutical sciences, Najaf, Iraq,

^{2,3}College of Computer Science and Mathematics, University of Kufa, Najaf, Iraq,

⁴Faculty of science, University of Kufa, Najaf, Iraq,

¹ola.h.allahibi@jmu.edu.iq, ²hadeel.alkutubi@uokufa.edu.iq

^{3,2}noors.alfatlawy@uokufa.edu.iq, ⁴Nabaam.alshammary26@uokufa.edu.iq

Abstract: In this research paper, data on health indicators published in the environmental statistics for Iraq for the year 2020, by the Central Statistical Organization in Iraqi Ministry of Planning, were used. The health data analyzed includes the number of infections and the number of deaths from the cancer cases among males and females, in addition to analyzing data distributed among all Iraqi governorates for the years 2017 and 2018. Two-way analysis of variance (ANOVA) was used to demonstrate the degree of interaction between all independent variables and to test for significant differences between the number of cases and deaths for the period 2013-2018, as distributed across the various Iraqi governorates. Finally, the study demonstrated clear significant differences between the number of cancer cases among males and females for the years 2017 and 2018, distributed across the various Iraqi governorates. Furthermore, significant differences were found in the number of cancer cases between the years 2013-2018.

Keywords: Two-way analysis of variance, Statistical data from the Iraqi Ministry of Planning.

1. Introduction

Analyzing data statistically is involved in many fields in various sciences due to its importance in monitoring phenomena and determining methods for collecting, classifying, analyzing, and interpreting data to reach sound decisions in the field and to conduct reliable scientific research with its results. In this research, we used statistical data published by the Iraqi Ministry of Planning/ Central Statistical Organization for the year 2020/ The Private data analysis of case of cancerous tumors between males and females and there is also the number of cancerous cases distributed among the different governorates of Iraq for the years 2017 and 2018. We used the two-way analysis of variance method, which is considered one of the statistical methods commonly used in life applications because of its advantages, as it allows researchers to examine the effects of the interaction between two independent variables on a dependent variable. This is important because it helps determine whether the impact of an independent variable on the dependent variable is affected by the presence of another independent variable. There are many researchers who used the method of two-way variance and factorial analysis, and among them are H Zhang et al [9], studied the importance of the effects of multiple behavioral factors: target direction (left and right) and obstacle (presence or absence) and their interaction on the dependency between the group of neurons in the dorsal motor PMD region of the monkey during the reaching movement. They used ANOVA to investigate the importance of the effects of these two behavioral factors and their interaction on repetition. In 2013 [1], researcher Al-Maliki studied the effect of seaweed extract. The experiment was conducted in the Department of Horticulture and Biozyme Engineering on the growth and yield of two varieties of Lahana plants according to the design of a 4×2 factorial experiment under a randomized complete block design. The results showed a significant superiority of the Lucky Ball variety as well. Treatment with three times the addition of Biozyme extract, the highest values in the total weight of the Lucky Ball variety of the Lahana plant. In 2016 [4], researcher Karto Iskandar and others examined the performance of content creation speed in four units of the knowledge management system of BINVS University. The test was conducted to obtain the average value of speed performance in 3 branches of BINUS University. The speed performance was analyzed using a two-way analysis of variance and it was concluded that there are differences in the average value of speed performance for each unit, and there is no difference in the average value of speed performance for each university campus, and there is also no interaction between campuses and units in speed performance.

2. Materials and Method

In this section, the first item explains shows the two-way analysis of variance. As for the second item, we present the data used in this research and the statistical methods for analyzing it.

2.1. Two – Way Analysis of Variance (ANOVA)

Two-way ANOVA is the most commonly used analysis technique when an experiment has a quantitative outcome and two categorical explanatory variables that are defined such that each experimental unit (subject) can be exposed to any combination of one level of one explanatory variable and one level of the other explanatory variable. Since there are two distinct explanatory

variables, a change in one variable may have an additive model effect where changes in the other variable do not depend on its level or an interaction model effect where changes in the other variable do [2]. To obtain a two-way analysis of the variance table, it is necessary to calculate the sum of the squares of the errors for the parameters used in the experiment and also to calculate the sum of the squares of the experimental error. All of this is done by using the following equations (SST, SSR, SSC, and SSE), as follows: $SST = \sum_i^r \sum_j^c Y_{ij}^2 - rc\bar{Y}_{..}^2$, $SSR = c \sum_i^r \bar{Y}_{i.}^2 - rc\bar{Y}_{..}^2$, $SSC = r \sum_j^c \bar{Y}_{.j}^2 - rc\bar{Y}_{..}^2$, $SSE = SST - SSR - SSC$. Finally, complete the Two – way analysis of variance (ANOVA) table and find the calculated value F. The statistical conclusion of the experiment is made by comparing the calculated F value with its tabulated value. If the calculated value is greater than the tabulated value, then the null hypothesis is rejected and the alternative hypothesis is accepted. [1], [4], [5].

2.2. Research Data and Methods of Analysis

The data used in this research were obtained from the Department of Environmental Statistics 2020 / published by the Iraqi Ministry of Planning / Central Statistical Organization for the year 2020. In this section, the data used in the research will be presented through statistical tables, and then the analysis used in this research will be explained. [2], [3].

Table 1: The total number of cases of cancerous tumors by gender and governorate in 2017.

Governorates	Male	Female	Total
Ninawa	640	1071	1711
Kirkuk	436	661	1097
Diyala	440	647	1087
Al Anbar	388	517	905
Baghdad	3200	4538	7738
Babil	607	784	1391
Karbala	623	704	1327
Wasit	434	523	957
Salah Ad Din	292	419	711
Najaf	546	704	1250
Al Qadisiyyah	384	507	981
Al Muthana	251	331	582
Dhi Qar	663	772	1435
Maysan	247	296	543
Basra	951	1243	2194
Dahuk	417	414	831
As Sulaymaniyah	838	953	1791
Arbil	798	883	1681
Unknown	347	554	901
Total	12502	16521	29023

The data in the above table was analyzed using two-way analysis of variance. As for the rest of the tables subject to the two-way analysis of variance (Table 2, Table 3, Table 4, Table 5), we will follow the same steps above in part 2.1. The results of the analysis will be presented and discussed in the Results and Discussion section (Table 6, Table 7, Table 8, Table 9, Table 10).

Table 2: Total number of deaths from cancerous tumors by sex and governorate in 2017.

Governorates	Male	Female	Total
Ninawa	20	19	39
Kirkuk	79	63	142
Diyala	185	212	397
Al Anbar	106	59	165
Baghdad	1235	1300	2535
Babil	362	398	760
Karbala	254	241	495
Wasit	212	189	401
Salah Ad Din	32	21	53
Najaf	118	127	245
Al Qadisiyyah	81	92	173
Al Muthana	14	11	25
Dhi Qar	385	368	753
Maysan	51	40	91
Basra	397	385	782
Dahuk	0	1	1
As Sulaymaniyah	1	0	1
Arbil	---	---	---
Unknown	35	52	87
Total	3567	3578	7145

(---): Data not available.

Table 3: The total number of cases of cancerous tumors by gender and governorate in 2018.

Governorates	Male	Female	Total
Ninawa	997	1208	2205
Kirkuk	362	499	861
Diyala	579	851	1430
Al Anbar	469	642	1111
Baghdad	3008	4399	7407
Babil	613	846	1459
Karbala	509	682	1191

Wasit	406	521	927
Salah Ad Din	345	498	843
Najaf	623	814	1437
Al Qadisiyyah	449	546	995
Al Muthana	282	383	665
Dhi Qar	737	804	1541
Maysan	355	417	772
Basra	1022	1293	2315
Dahuk	561	555	1116
As Sulaymaniyah	1018	1245	2263
Arbil	908	1038	1946
Unknown	369	649	1018
Total	13612	17890	31502

Table 4: Total number of deaths from cancerous tumors by sex and governorate in 2018.

Governorates	Male	Female	Total
Ninawa	300	362	662
Kirkuk	75	85	160
Diyala	287	286	573
Al Anbar	145	156	301
Baghdad	1516	1444	2960
Babil	394	367	761
Karbala	274	256	530
Wasit	224	237	461
Salah Ad Din	120	130	250
Najaf	210	245	455
Al Qadisiyyah	114	123	237
Al Muthana	11	15	26
Dhi Qar	353	385	738
Maysan	171	174	345
Basra	434	403	837
Dahuk	183	249	432
As Sulaymaniyah	112	161	273

Arbil	61	89	150
Unknown	81	61	142
Total	5065	5228	10293

Table 5: The total number of cases of cancerous tumors by governorate for the years (2013-2018).

Governorates	Years					
	2013	2014	2015	2016	2017	2018
Ninawa	2015	1953	1623	636	1711	2205
Kirkuk	719	1023	877	878	1097	861
Diyala	980	995	963	1054	1087	1430
Al Anbar	1010	388	447	474	905	1111
Baghdad	5859	6927	7044	7308	7738	7407
Babil	1257	1352	1398	1311	1391	1459
Karbala	993	1089	1166	1328	1327	1191
Wasit	767	827	784	852	957	927
Salah Ad Din	658	547	593	540	711	843
Najaf	1094	1060	1126	1161	1250	1437
Al Qadisiyyah	722	766	721	933	891	995
Al Muthana	437	468	379	462	582	665
Dhi Qar	1193	1294	1216	1414	1435	1541
Maysan	590	623	605	662	543	772
Basra	1610	1965	1969	2132	2194	2315
Dahuk	388	758	766	768	931	1116
As Sulaymaniyah	1085	1311	1181	1051	1791	2263
Arbil	795	977	1116	1583	1681	1946
Other sites	1136	1275	1295	1009	901	1018
Total	23308	25598	25269	25556	29023	31502

Source: Iraqi Ministry of Health and Environment/Cancer Council

3. Results and Discussion

This section presents the results of the statistical analysis of the data used in this research, tabulated and arranged within the tables presented in the previous section. The results of the statistical analyzes were discussed (each table separately).

Table 6: Two-way analysis of variance table for the number of cancer cases in 2017.

Source of variation	df	SS	MS	F _{cal.}	F _{tab.}
Between row rates	18	22095468.67	1227526.04	29.291	2.94

(Between governorates)					
Between column rates (Between males and females)	1	424375.92	424375.92	10.126	8.10
Error	18	754327.58	41907.09		
Total	37	23274172.17			

The two-way analysis of variance table above shows that there is a significant difference in the total number of cases of cancerous tumors between males and females, and there is also a significant difference in the number of cancerous cases distributed among the different governorates of Iraq for the year 2017.

Table 7: Two-way analysis of variance table for the number of deaths from cancer in 2017.

Source of variation	df	SS	MS	F _{cal.}	F _{tab.}
Between row rates (Between governorates)	17	3033919.23	178465.84	581.15	2.94
Between column rates (Between males and females)	1	74.82	74.82	0.244	8.10
Error	17	5220.68	307.09		
Total	35	3039214.73			

The table above shows that there is a significant difference in the total number of deaths due to cancerous tumors between the different governorates of Iraq. While there is no clear significant difference in the number of deaths between males and females.

Table 8: Two-way analysis of variance table for the number of cancer cases in 2018.

Source of variation	df	SS	MS	F _{cal.}	F _{tab.}
Between row rates (Between governorates)	18	19825007	1101389.28	25.655	2.94
Between column rates (Between males and females)	1	481077.15	481077.15	11.206	8.10
Error	18	772755.85	42930.88		
Total	37	21078840			

The analysis of variance table above indicates that there is a significant difference in the number of cases of cancerous tumors in Iraq in 2018 between males and females, in addition to the presence of another significant difference in the number of infections between all governorates of Iraq.

Table 9: Two-way analysis of variance table for the number of deaths from cancer in 2018.

Source of variation	df	SS	MS	F _{cal.}	F _{tab.}
Between row rates (Between governorates)	18	3567589.34	19821.63	34.927	2.94
Between column rates (Between males and females)	1	699.35	699.35	1.2323	8.10

Error	18	10215.15	567.51
Total	37	3578503.84	

From the table above, we note that there is a significant difference in the number of deaths from cancerous tumors in Iraq in 2018 between the different governorates of Iraq. While there is no significant difference in the number of deaths in Iraq between males and females in 2018.

Table 10: Two-way analysis of variance table for the number of cancerous tumors for the years (2013-2018).

Source of variation	df	SS	MS	F _{cal.}	F _{tab.}
Between row rates (Between governorates)	18	218171671.87	12120648.44	205.25	2.37
Between column rates (Between the years)	5	2646236.87	529247.37	8.962	3.34
Error	90	5314774.13	59053.045		
Total	113	226132682.87			

The two-way analysis of variance table above shows that there are significant differences in the total number of cancerous tumors for the years (2013-2018) between males and females. There are also clear significant differences in the number of cancer cases between all governorates of Iraq.

4. Conclusion

The study revealed significant differences between the number of cancer cases among males and females in 2017 and 2018, in addition to significant differences in the number of cases distributed across the various Iraqi governorates. Regarding the number of deaths, it was concluded that there were no significant differences between male and female cases in 2017. However, a significant difference was observed in the number of deaths in 2018. Regarding the number of cancer cases for the period 2013-2018, it was concluded that there was a clear significant difference in the number of cases between the years under study and distributed across the various Iraqi governorates.

References

- [1] Al-Maliki, Abdul-Hussein Qasim. (2013). "The effect of Biozyme marine algae extract on the growth and yield of two varieties of Lahana plants grown in desert areas" . Basra Research Journal (Operations) .Volume 39-4, Pages (88-97).
- [2] Al-Sheikhli, Rawaa Nouri Hussein . (2017) . "The relationship between the regression variance analysis table and the full factorial experiments analysis table" A master's thesis in statistics submitted to the College of Administration and Economics - University of Karbala.
- [3] Al Shmary Nabaa, Alkutubi H. and Kadham S. 2025. Alayalysis of coronavirus data using a randomized complete block design and tukey's honest significance test. Mathematical Modelling of Engineering Problems. 12(2), pp: 395-402.
- [4] Alkutubi H, Murad N, Al Shmary N, Al Luhabet O. 2025. Private data analysis of cases of viral hepatitis in Iraq for the year 2019 using a factorial experiment and a randomized complete block design. World journal of pharmacy and pharmaceutical sciences. 14(2), pp: 1447-1456.
- [5] Al Kutubi H. 2024. An Analytical Study of the Number of Communicable Diseases and Health Institutions in Iraq using Two-Way Analysis of Variance, Friedman and Scheffe Tests. Communications on applied nonlinear analysis. 31(4), pp: 42-71.
- [6] Fujikoshi, Yasunori. (2019). "Two-way ANOVA models with unbalanced data" Discrete mathematics. Elsevier. 116(1). pp (315-334).
- [7] Kitsche, A., & Schaarschmidt, F. (2015). Analysis of statistical interactions in factorial experiments. Journal of Agronomy and Crop Science, 201(1), 69-79.
- [8] Mead Fadel Aliwi, Obaid Mahmoud Al-Zubaie. (2017). "Using the RSM response surface method for factorial design in improving processes and achieving optimization by applying it to the Al-Fattayah Cement Factory in the city of Derna – Libya" The Scientific Journal of Cihan. vol. (1) No (3).