Why temperature cannot corelate with acute COVID-19 spreading and death toll. Data from Spain

Rafael Franco[1,2] and *Miguel Galo Fernández*[3]

1. Department of Biochemistry and Molecular Biomedicine. University of Barcelona. Barcelona. Spain.

2. CiberNed. Centro de Investigación en Red, Enfermadades Neurodegenerativas. National Health Institute Carlos III, Madrid. Spain.

3. Statistical advisor (www.galo.es). miguel@galo.es. Murcia. Spain.

Correspondence
Rafael Franco
Department of Biochemistry and Molecular Biomedicine.
Diagonal 643.
Facultat de Biologia. Prevosti Building
rfranco123@gmail.com; rfranco@ub.edu
Tel +34 934012108

Index of contents

1. Abstract

Appearance of outbreaks due to some viruses is seasonal; in temperate (non-tropical) areas of the World, viruses are usually silent in summer while they reappear in winter. Based on this fact it is believed that SARS-CoV-2 will disappear in summer. Recent attempts to putatively prove this hypothesis has led to promising data. In this paper we present data showing that the kinetics of current acute COVID-19 pandemics does not correlate with temperature. Using absolute number of cases and death rate and average temperature in April in either Spanish provinces/regions or in Catalonian comarcas (comarca being similar to Counties in the United Sates), we find that the data are independent. What our data tells is that temperature is not a key factor in current evolution of this particular pandemic. Although there cannot be any scientific evidence supporting the hypothesis of virus disappearance in summer, there are still hopes that this might happen. Time will tell.

Keywords Temperature, SARS-CoV-2, COVID-19, incidence rate, time-series regression, Barcelona, Spain.

2. Introduction

COVID-19 pandemics is affecting world population in a way that was not known in recent history. The serious condition having a death toll of at least 1 % (Onder et al., 2020) and affecting the economy of population worldwide (McKibbin and Fernando, 2020) is managed mainly by confinement to avoid person-to-person infection, i.e. to reduce infectivity rate.

It is known that the first year of appearance of a new virus the affectation is higher due to the lack of previous exposure to the pathogen. When the pathogen reappears, the immune system of previously-exposed individuals impedes infection. Then, already exposed individuals rarely become infected nor transmit the infection in the next season (see (Hayward et al., 2015) and references therein). This natural mechanism of defense against new pathogens is reinforced by the development of vaccines or of specific drugs, both of which are only available upon testing for safety and efficacy, i.e. some years after the appearance of the new virus. In this scenario the main objective is to wait for infection decay and be ready for reappearance. For many pathogens outbreaks are seasonal. The main example is provided by the influenza virus; in temperate (not tropical) areas of the World flu appears in winter and disappears in summer (Roussel et al., 2016). We recommend the review of (Cannell et al., 2008) to understand flu epidemiology.

To the best of our knowledge there is not any real explanation of why the influenza virus becomes undetectable in the months with higher temperature. Probably there are multiple factors that impact, among them a lower latency time. In fact RNA viruses such as the influenza and SARS-type, have a limited life when outside an infected body; accordingly higher temperatures inactivate quickly the virus outside the infected individual thus decreasing the infective power. Also, the ultraviolet rays of the Sun, which is appears more hours in summer than in winter, kill pathogens. It likely that human bodies in summer have more mechanism of defense against infection. The mechanism are unclear but one that has been put forward is a higher activity of epithelial cell of the lung to trap/remove particles. Taken together it is tempting to speculate that COVID-19 will disappear in summer. This will be known in the future; in the meantime we can try to make correlations between temperature and COVID-19 infection. A recent report states that the temperature affect the infectivity rate (Tobías and Molina, 2020). Our paper aimed at further contrasting this possibility and at correlating cases and temperature in a Country with very high COVID-19 incidence, Spain. As of today May 12, the death rate per million population is 572, right after San Marino, Belgium and Andorra. If we take only Madrid and Catalonia regions the death toll per million population is the highest in the world and comparable to that of San Marino (¿1,200) (https://www.worldometers.info/coronavirus/; date: May 12, 2020). Our results indicate that cases/deaths in this first COVID-19 infection in Spain do not correlate with temperature.

3. Methods

3.1. Data collection

All data are taken from official National and/or regional sources, mainly from the INE (Spanish National Institute for Statistics) and METEOCAT (Catalonian service for weather data). No data of temperature is available for 2020, therefore we took data from the most recent year for which temperatures are available. Data are available in the accompanying Excel. We assumed that differences on average temperatures in April in 2020 versus those in previous years are not biased in any specific Spanish location. We have confirmed by comparing average temperate in some few years backwards, that despite global warming, average temperatures have not significantly changed in Spanish regions in short periods of time. Figures were prepared using Excel and its linear regression utility. When indicated, data provided as supplementary material of the (Tobías and Molina, 2020) paper were used.

Statistical analysis The IBM SPSS® Statistics package was used to calculate the Pearson and other statistical correlation coefficients and to calculate p in contrast of hypothesis on correlation between cases/deaths versus temperature. Significance was considered when p¡0.05.

3.2. Results

Apparent positive correlation between cases and April average temperature in Catalonia.

As indicated in Methods we took the number of cases and of COVID-19-related death from official sources and the temperatures publicly available; in Catalonia there is not any information for 2019 or 2020 but for 2018. All data used in statistical analysis are found in Supplementary material (Excel file). Within Catalonia data were taken by comarcas that are roughly equivalent to what counties are in the US. Figure 1 shows the point cloud chart of deaths per 10,000 population and temperature variation (respect to average) and the line resulting from linear regression fit. The temperature taken for a given comarca is the average temperature in April. Despite the apparent positive correlation that is suggested by the positive slope of the linear plot, statistical analysis shows that the null hypothesis holds, i.e. the two compared sets of data are independent.

Taking into account a significance level of $\alpha = 5\,\%$ the test of Shapiro-Wilk for n=39 shows that the temperature variation is normally distributed and that the population per comarca does not follow a normal distribution. This result already shows that it is unlikely that the two sets of data correlate. Spearman test for the 39 pair of death rate/temperature variation data, provides a correlation coefficient (r) of 0.107 and, indeed the two sets of data are independent because the p value is 0.581. As a control the same type of analysis was performed taking into account the death rate and the population. The correlation coefficient (r) of -0.297 and a p=0.06, suggest independence that would turn into dependence if instead of $\alpha = 5\,\%$ a $10\,\%$ value is assumed. Whatever the case, this result was not expected as we were expecting that the death rate would be proportional to the population. The reason for independence may be due to the fact that deaths are not necessarily counted in the comarca where it happens but in the hospital where the patient dies. A similar comparison using more homogeneous data, i.e. taking paired values of population in a Spanish province and deaths occurring in each of the same province, there is a dependence of number of deaths and population (see below). Similar results are obtained with cases as deaths are roughly proportional to cases. It is however important that cases/deaths are counted using the same criteria. This requirement was not met in April but homogeneous criteria across Catalonia and Spain was ruling when the data was collected by May 1, 2020.

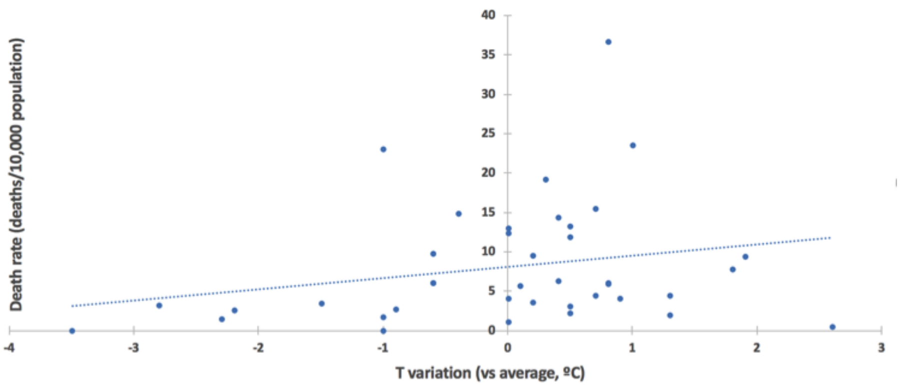

Figura 1: Plot of deaths versus April temperature variation in Catalonian “comarcas” (see Supplementary material for raw data). Death are presented per 10,000 population and X axis indicate temperature variation for each “comarca” in April respect to April average temperature in Catalonia (in year 2018). Data as of May 1, 2020.

The apparent positive correlation contrasts with visual inspection of the heat map of deaths in Catalonian “comarcas”. Temperatures in April in Catalonia are colder going North (up in Figure 2) and West (left in Figure 2). Areas hardest hit by the epidemic are not necessarily those in North West areas. By comparing relatively important cities with similar latitude, Lleida with is the capital of a province and Barcelona, another capital of province, it is true that cases are most abundant in the warmer city (Barcelona). However, this trend is completely opposite when we one goes South from Barcelona, where warmer cities do present much less cases than in Barcelona or than -the even colder- Girona. Again it is unlikely that COVID-19 incidence in Catalonia in April correlated with temperature.

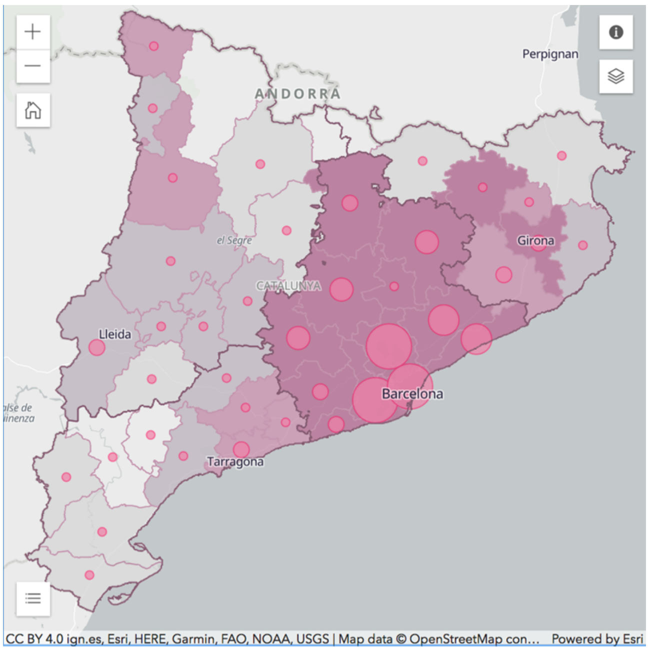

Figura 2: Heat map of death rates in Catalonia. Relatively low death rates (normalized by million population) are in grey color scale and relatively high death rates are in pink/magenta color scale; the darker the color the higher the death rate. The diameter of the pink circle is roughly proportional to the normalized death rate. Taken on May 12, 2020 from National.cat on-line newspaper (https://www.elnacional.cat/es/salud/coronavirus-mapa-contagios-catalunya-municipios _ 485916 _ 102.html). A similar image but taken on May1rst if available in (Franco, 2020)

4. Apparent negative correlation between cases and April average temperature in Spain

We next wanted to address whether the apparently positive trend in Catalonia was also found in the while Country, Spain. We took the number of cases and of COVID-19-related death from official sources and the temperatures publicly available; in Spain there is not any information for 2020 but for 2019. All data used in statistical analysis are found in Supplementary material (Excel file). Within Spain data is taken by provinces except when death toll is not available per provinces but for various closely-located provinces constituting a region. Then we used either data from provinces except for Catalunya, Galicia, Comunidad Valenciana and Pais Vasco . It should be reminded that Catalonia has data for every comarca but not for provinces; therefore we used average data for the whole region (and similarly for Galicia, Comunidad Valenciana and Pais Vasco). Data from these regions were also included in the analysis. Figure 3 shows the point cloud chart of deaths per 10,000 population and temperature variation (respect to average) and the line resulting from linear regression fit. The temperature taken for a given province/region is the average temperature in April. Despite the apparent negative correlation suggested by the negative slope of the linear plot, statistical analysis shows that the null hypothesis holds, i.e. the two compared sets of data are independent.

Taking into account a significance level of $\alpha = 5\,\%$ the test of Shapiro-Wilk for n=41 shows that the temperature variation is normally distributed and that the population per provinces/regions does not follow a normal distribution. This result already shows that it is unlikely that the two sets of data correlate. Spearman test for the 41 pair of death rate/temperature variation data, provides a correlation coefficient (r) of -0.16 and, indeed the two sets of data are independent because the p value is 0.158. As a control the same type of analysis was performed taking into account the death rate and the population. The correlation coefficient (r) of -0.16 and a p=0.001 shows dependence, i.e. the higher the population the higher the death toll. Similar results are obtained with cases as deaths are roughly proportional to cases. A note of caution is needed as before along April 2020 the criteria to count COVID-19 cases and number of deaths changed. At the date of this study (May 1rst) the criteria to count cases and deaths was the same all around Spain (by Government rules)

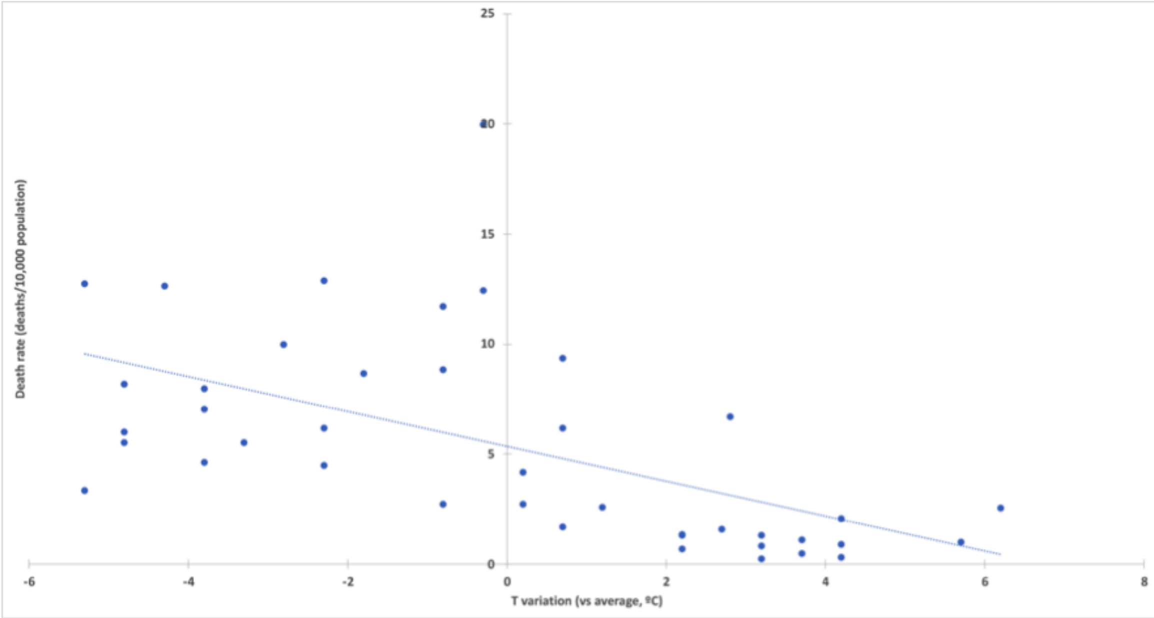

Figura 3: Plot of deaths versus April temperature variation in Spain provinces/regions (see Supplementary material for raw data). Death are presented per 10,000 population and X axis indicate temperature variation for each province/region in April respect to April average temperature in Spain (in year 2019). Data from regions were considered when official death rates were not available for individual provinces in given region. Data as of May 1, 2020.

5. Discussion

Data on this paper shows that there is not any correlation between the spread of SARS-CoV-2 averaged on the month when COVID-19 peaked in Spain and the average temperature in different regions of the Country this specific month (April). Just by eye looking at the heat map in Figure 2, such correlation is unlikely occurring. In fact, just taken into account four main cities in Catalonia, i.e. those that are capital of the four provinces, it is evident that cases are fewer in warmer Tarragona than in Barcelona but also in colder Girona and much colder Lleida (Franco, 2020). Furthermore, the number of cases and normalized number of deaths is lesser in Lleida, which is colder than Girona. Going to data in overall Spain, there is a seemingly negative correlation between death and average temperature that might lead to accept that the higher the temperature the lesser the death toll. But in Catalonia whose capital is Barcelona in which the report by (Tobías and Molina, 2020) seems to prove such negative correlation, our careful analysis shows the opposite, i.e. the higher the temperature the worse the COVID-19-related numbers. In fact all those data can only be reconciled by the fact that there is not any correlation. Linear regression assumes that data can be properly fitted to a straight line but this is not the case as clearly deduced from the position of the "experimental" points in Figures 1 and 3. These data points cannot be represented by a straight line.

Google Scholar search for "temperature and COVID-19" provides 9690 hits as of May 12, 2020. The first hit is a recent report stating that high temperature and high humidity reduces COVID-19 transmission. The paper seemingly demonstrates that 1°C of temperature increase decreases by 0.0225 what authors call "daily effective reproductive number" (R), being the decrease of 0.0158 when the humidity increases by 1 % (Wang et al., 2020). Although there are flaws in the statistics and the approach misses that viral transmission is not immediate, these numbers are modest when compared with those of (Tobías and Molina, 2020) who seemingly demonstrate that for every °C of temperature increase there is (in Catalonia) a 7.5 % reduction in the COVID-19 "incidence" rate.

Wishes, which are legitimate, do not fit when fueling predictions disguised of rigorous research. Not surprisingly virtually all the reports go in the same direction. The fact that other SARS viruses slow down progression depending on the temperature and humidity (Chan et al., 2011) is irrelevant in the case of SARS-COV-2. We do not know and cannot know now. Common behavior of previous viruses cannot serve to predict the behavior of a new virus. But even today when we know that flu disappears in summer, there is controversy on the real correlation between flu spread and temperature, humidity and wind speed (Peci et al., 2019). There are two types of misunderstanding underlying data handling in the (Wang et al., 2020), in the (Tobías and Molina, 2020) and in virtually all the recent papers "demonstrating" correlation between COVID-19 cases and temperature. Actually they are linked by the fact of considering immediate (daily) changes in temperature (or in humidity). The now famous epidemic curve has a similar inverted U shape in all countries. Neither the rise nor the decay has anything to do with temperature; in fact, the first highest slope in the rising of COVID-19 cases in April was in warmer Italy, and few days later Italy was surpassed by more warmer Spain. In the most acute phase on pandemics Spain had the highest slope despite being the warmer country compared with Countries whose peak of COVID-cases were preceding Spain. In summary, virus spreading in a Country with high average temperature in April was higher than in colder countries. This goes against the conclusions of the (Tobías and Molina, 2020) paper. A second reason is also related with immediate data collection as any viral infection takes days to be detected. Virus transmission depends on the viral exposure that, for COVID-19 in turn depends on the confinement measures and how they are enforced. Moreover, virus-derived symptoms do not immediately appear but take few days. Last but not least it is required that the cases are real and, much unfortunately, there has been a shortage of tests in Spain, thus impeding to know the real number of infected people and missing asymptomatic cases.

According to the US centers for disease control (CDC), nonpharmacological interventions in a viral pandemics such as any of those produced by influenza viruses aim at: "1) delay the exponential growth in incident cases and shift the epidemic curve to the right in order to "buy time" for production and distribution of a well-matched pandemic strain vaccine, 2) decrease the epidemic peak, and 3) reduce the total number of incident cases, thus reducing community morbidity and mortality" (Centers for Disease Control and Prevention, 2007). These are sensible rules to implement while waiting that SARS-CoV-2 will, in a temperate Country like Spain, disappear in summer. Disappearance or not will be empirically known. Even if it disappears for some months preparation for next-season outbreaks also requires interventions such an epidemiological study to detect and make a profile for asymptomatic carriers and to early detect the beginning of next COVID-19 outbreak (Franco, 2020).

6. References

Cannell, J.J., Zasloff, M., Garland, C.F., Scragg, R., Giovannucci, E., 2008. On the epidemiology of influenza. Virol. J. doi:10.1186/1743-422X-5-29

Centers for Disease Control and Prevention, 2007. Interim Pre-pandemic Planning Guidance: Community Strategy for Pandemic Influenza Mitigation in the United States — Interim Pre-Pandemic Planning Guidance: Community Strategy for Pandemic Influenza Mitigation in the United States — Early , Targeted , La, CDC stacks.

Chan, K., Malik Peiris, J., Lam, S., Poon, L., Yuen, K., Seto, W., 2011. The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus. Adv. Virol. Article ID, 1–7. doi:10.1155/2011/734690

Franco, R., 2020. COVID-19 Management in Spain. Errors to Avoid and the Need of A Post-Confinement Longitudinal Epidemiological Study. doi:10.20944/PREPRINTS202005.0137.V1

Hayward, A.C., Wang, L., Goonetilleke, N., Fragaszy, E.B., Bermingham, A., Copas, A., Dukes, O., Millett, E.R.C., Nazareth, I., Nguyen-Van-Tam, J.S., Watson, J.M., Zambon, M., Johnson, A.M., McMichael, A.J., 2015. Natural T cell-mediated protection against seasonal and pandemic influenza: Results of the flu watch cohort study. Am. J. Respir. Crit. Care Med. 191, 1422–1431. doi:10.1164/rccm.201411-1988OC

McKibbin, W.J., Fernando, R., 2020. The Global Macroeconomic Impacts of COVID-19: Seven Scenarios. SSRN Electron. J. doi:10.2139/ssrn.3547729

Onder, G., Rezza, G., Brusaferro, S., 2020. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA - J. Am. Med. Assoc. doi:10.1001/jama.2020.4683

Peci, A., Winter, A.L., Li, Y., Gnaneshan, S., Liu, J., Mubareka, S., Gubbay, J.B., 2019. Effects of absolute humidity, relative humidity, temperature, and wind speed on influenza activity in Toronto, Ontario, Canada. Appl. Environ. Microbiol. 85. doi:10.1128/AEM.02426-18
Roussel, M., Pontier, D., Cohen, J.-M., Lina, B., Fouchet, D., 2016. Quantifying the role of weather on seasonal influenza. BMC Public Health 16, 441. doi:10.1186/s12889-016-3114-x

Tobías, A., Molina, T., 2020. Is temperature reducing the transmission of COVID-19? Environ. Res. 186, In the Press. doi:10.1016/j.envres.2020.109553

Wang, J., Tang, K., Feng, K., Lv, W., 2020. High Temperature and High Humidity Reduce the Transmission of COVID-19. SSRN Electron. J. doi:10.2139/ssrn.3551767

Cuadro 1: **Tabla en la que se basa el Estudio estadístico en España**

Provincia/Autonomía	Defunciones	TCD	Población	TMA	Δt
Galicia	557	2.06	2,699,499	17.5	4.2
Asturias	279	2.73	1,022,800	12.5	-0.8
Cantabria	193	3.32	581,078	8.0	-5.3
País Vasco	1,321	5.98	2,207,776	8.5	-4.8
León	366	7.96	460,001	9.5	-3.8
Zamora	77	4.46	172,539	11.0	-2.3
Salamanca	329	9.97	330,119	10.5	-2.8
Valladolid	321	6.18	519,546	11.0	-2.3
Palencia	74	4.60	160,980	9.5	-3.8
Ávila	129	8.18	157,640	8.5	-4.8
Segovia	195	12.73	153,129	8.0	-5.3
Soria	112	12.64	88,636	9.0	-4.3
Burgos	197	5.52	356,958	8.5	-4.8
Zaragoza	594	6.16	964,693	14.0	0.7
Huesca	92	4.17	220,461	13.5	0.2
Teruel	74	5.52	134,137	10.0	-3.3
Navarra	460	7.03	654,214	9.5	-3.8
Cataluña	5,137	6.69	7,675,217	16.1	2.8
Madrid	8,292	12.44	6,663,394	13.0	-0.3
Toledo	614	8.84	694,844	12.5	-0.8
Ciudad Real	990	19.97	495,761	13.0	-0.3
Albacete	454	11.70	388,167	12.5	-0.8
Cuenca	253	12.89	196,329	11.0	-2.3
Guadalajara	223	8.65	257,762	11.5	-1.8
Cáceres	368	9.34	394,151	14.0	0.7
Badajoz	90	1.34	673,559	15.5	2.2
Comunidad Valenciana	1,266	2.53	5,003,769	19.5	6.2
Murcia	132	0.88	1,493,898	17.5	4.2
Islas Baleares	193	1.68	1,149,460	14.0	0.7
Huelva	43	0.82	521,870	16.5	3.2
Sevilla	255	1.31	1,942,389	16.5	3.2
Córdoba	100	1.28	782,979	15.5	2.2
Jáen	162	2.56	633,564	14.5	1.2
Almería	49	0.68	716,820	15.5	2.2
Granada	248	2.71	914,678	13.5	0.2
Málaga	261	1.57	1,661,785	16.0	2.7
Cádiz	135	1.09	1,240,155	17.0	3.7
Ceuta	4	0.47	84,777	17.0	3.7
Melilla	2	0.23	86,487	16.5	3.2
Tenerife	102	0.99	1,032,983	19.0	5.7
Las Palmas de Gran Canaria	34	0.30	1,120,406	17.5	4.2

Cuadro 2: TCD: Tasa cruda de defunciones por cada 10.000 habitantes, TMA: temperatura media en el mes de abril, Δt : desviación de la temperatura respecto a la media

7. Estudio estadístico en provincias de España

7.1. La normalidad de las poblaciones

Hacemos el siguiente contraste de hipótesis:

$H_0 \equiv$ Las muestras X proceden de una población normal.

$H_1 \equiv$ Las muestras X **NO** proceden de una población normal.

X puede ser una de las siguientes muestras: defunciones, tasa cruda de defunciones por cada 10.000 habitantes, población, temperatura media en abril, desviación de la temperatura respecto a la media.

Efectuamos un test de normalidad con el programa IBM SPSS Statistics versión 24.0.0.0 obteniendo los siguientes resultados:

Pruebas de normalidad

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Defunciones	,375	41	,000	,377	41	,000
Tasa cruda def. 10000h	,156	41	,014	,892	41	,001
Población	,281	41	,000	,599	41	,000
TMA	,115	41	,196	,952	41	,085
Deltat	,115	41	,196	,952	41	,085

a. Corrección de significación de Lilliefors

Para un nivel de significación $\alpha = 5\,\%$ el p-valor para defunciones, tasa cruda de defunciones por cada 10.000 habitantes y población es menor que 0.05 (el p-valor viene dado por sig.) por lo que el test es significativo, rechazamos la hipótesis nula con lo que estas tres muestras **NO** proceden de una población normal. Miramos el p-valor para el test de Shapiro-Wilk ya que el tamaño de cada muestra es de 41 individuos (menor de 50). Para la temperatura media en el mes de abril y la desviación de la temperatura respecto a la media el p-valor es 0,085 > 0,05, lo que nos dice que para estas dos muestras el test no es significativo y por tanto SÍ proceden de una población normal.

Si aplicamos el Teorema Central del Límite, al ser el tamaño de la muestra de 41 individuos (mayor de 30), podemos considerar que las muestras de defunciones, tasa cruda de defunciones por cada 10.000 habitantes y población tienen unas medias que proceden de una normal. Por el número de datos en cada muestra podemos considerar que proceden de una normal estas muestras. Lo mismo se puede decir para el resto de las muestras.

7.2. Correlación entre fallecimientos y temperatura

Hacemos el siguiente contraste de hipótesis:

$H_0 \equiv$ Las muestras de temperatura y de fallecimientos son independientes.

$H_1 \equiv$ Las muestras de temperatura y de fallecimientos están relacionadas.

Efectuamos un test de correlación con el programa IBM SPSS Statistics versión 24.0.0.0. El test de correlación que empleamos es el de de Pearson, indicado cuando las poblaciones son normales y paramétricas. Obtenemos los siguientes resultados:

Correlaciones

		Defunciones	TMA
Defunciones	Correlación de Pearson	1	,049
	Sig. (bilateral)		,761
	N	41	41
TMA	Correlación de Pearson	,049	1
	Sig. (bilateral)	,761	
	N	41	41

El p-valor es 0,761 $>$ 0,05, con un nivel de confianza del 95 % no podemos rechazar la hipótesis nula, **la temperatura y los fallecimientos son independientes**.

7.3. Correlación entre fallecimientos y desviación media de la temperatura de abril

Hacemos el siguiente contraste de hipótesis:

$H_0 \equiv$ Las muestras de desviación de la temperatura respecto a la media y de fallecimientos son independientes.

$H_1 \equiv$ Las muestras de desviación de la temperatura respecto a la media y de fallecimientos están relacionadas.

Efectuamos un test de correlación con el programa IBM SPSS Statistics versión 24.0.0.0. El test de correlación que empleamos es el de de Pearson, indicado cuando las poblaciones son normales y paramétricas. Obtenemos los siguientes resultados:

Correlaciones

		Defunciones	Deltat
Defunciones	Correlación de Pearson	1	,049
	Sig. (bilateral)		,761
	N	41	41
Deltat	Correlación de Pearson	,049	1
	Sig. (bilateral)	,761	
	N	41	41

El p-valor es 0,761 $>$ 0,05, **la desviación de la temperatura respecto a la media y los fallecimientos son independientes**.

7.4. Correlación entre el número de habitantes y las defunciones

Planteamos el siguiente contraste de hipótesis:

$H_0 \equiv$ Las muestras de población y defunciones son independientes.

$H_1 \equiv$ Las muestras de población y defunciones están relacionadas.

Efectuamos un test de correlación con el programa IBM SPSS Statistics versión 24.0.0.0. El test de correlación que empleamos es el de de Pearson, indicado cuando las poblaciones son normales y paramétricas. Obtenemos los siguientes resultados:

Correlaciones

		Defunciones	Población
Defunciones	Correlación de Pearson	1	,848**
	Sig. (bilateral)		,000
	N	41	41
Población	Correlación de Pearson	,848**	1
	Sig. (bilateral)	,000	
	N	41	41

**. La correlación es significativa en el nivel 0,01 (bilateral).

Como el p-valor es 0,000 $<$ 0,05 se rechaza la hipótesis nula, por tanto hay una relación entre el nº de habitantes de la provincia/comunidad autónoma y el número de defunciones. Al ser el coeficiente de correlación $r = 0{,}848 > 0$ la relación es directa, es decir, a mayor población de la provincia/comunidad autónoma, mayor número de fallecimientos.

Cuadro 3: Tabla en la que se basa el estudio estadístico en las comarcas catalanas

comarca	defunciones	TD	habitantes	temperatura	Δt
Val d'Aran	6	5.94	10,093	20.30	0.81
Alta Ribagorça	1	2.63	3,802	17.30	-2.19
Pallars Jussà	17	13.00	13,080	19.50	0.01
Pallars Sobirà	0	0.00	6,932	18.50	-0.99
Noguera	17	4.38	38,770	20.20	0.71
Segrià	119	5.67	209,818	19.60	0.11
Pla d'Urgell	15	4.09	36,693	20.40	0.91
Garrigues	5	2.65	18,833	18.60	-0.89
Urgell	15	4.09	36,693	19.50	0.01
Solsonès	2	1.48	13,469	17.20	-2.29
Segarra	8	3.47	23,052	18.00	-1.49
Conca de Barberà	12	5.99	20,042	18.90	-0.59
Priorat	1	1.08	9,245	19.50	0.01
Ribera d'Ebre	1	0.46	21,865	22.10	2.61
Terra Alta	2	1.74	11,490	18.50	-0.99
Baix Ebre	24	3.09	77,596	20.00	0.51
Montsià	15	2.22	67,436	20.00	0.51
Baix Camp	69	3.61	190,973	19.70	0.21
Tarragonès	161	6.27	256,730	19.90	0.41
Alt Camp	27	6.10	44,296	20.30	0.81
Conca de Barberà	12	5.99	20,042	18.90	-0.59
Garraf	144	9.54	150,887	19.70	0.21
Alt Penedès	144	13.28	108,411	20.00	0.51
Anoia	443	36.69	120,738	20.30	0.81
Vallès Occidenta	1144	12.36	925,237	19.50	0.01
Baix Llobregat	1273	15.41	825,963	20.20	0.71
Barcelonès	4373	19.19	2,278,437	19.80	0.31
Maresme	444	9.81	452,690	18.90	-0.59
Vallès Oriental	484	11.82	409,638	20.00	0.51
Bages	422	23.59	178,885	20.50	1.01
Osona	239	14.86	160,821	19.10	-0.39
Berguedà	91	23.07	39,446	18.50	-0.99
Selva	133	7.75	171,617	21.30	1.81
Cerdanya	0	0.00	18,192	16.00	-3.49
Ripollès	8	3.19	25,087	16.70	-2.79
Garrotxa	83	14.41	57,590	19.90	0.41
Gironès	182	9.39	193,908	21.40	1.91
Baix Empordà	59	4.39	134,359	20.80	1.31
Alt Empordà	28	1.98	141,339	20.80	1.31

Cuadro 4: TD: tasa cruda defunciones por cada 10.000 habitantes. Δt : desviación de la temperatura respecto a la media.

8. Estudio estadístico en Comarcas Catalanas

8.1. Normalidad de las muestras de defunciones, temperatura y población

Hacemos el siguiente contraste de hipótesis:

$H_0 \equiv$ La muestra X proceden de una población normal.

$H_1 \equiv$ Las muestra X **NO** proceden de una población normal.

X puede ser la muestra de fallecimientos, temperatura, desviación de la temperatura respecto a la media o población.

Efectuamos un test de normalidad con el programa IBM SPSS Statistics versión 24.0.0.0 obteniendo los siguientes resultados:

Pruebas de normalidad

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
defunciones	,360	39	,000	,373	39	,000
temperatura	,168	39	,007	,950	39	,083
delta_t	,168	39	,007	,950	39	,083
habitantes	,329	39	,000	,472	39	,000

a. Corrección de significación de Lilliefors

Figura 4: Prueba de normalidad

Como nuestra muestra tiene un tamaño de 39 individuos, nos fijamos en el test de normalidad de Shapiro-Wilk, el de Kolmogorov-Sirnov está indicado para un tamaño de la muestra superior a 50.

a) Para las defunciones y la población el p-valor (sig.) es de $0{,}000 < 0{,}05$, el test es significativo, se rechaza la hipótesis nula. Con un grado de significación $\alpha = 5\,\%$ **no podemos afirmar que las muestras de defunciones y población procedan de una población normal**.

b) Para la temperatura y desviación de la temperatura respecto a la media, el p-valor (sig.) es de $0{,}083 > 0{,}05$, el test NO es significativo, se acepta la hipótesis nula. Con un grado de significación $\alpha = 5\,\%$ **podemos afirmar que las muestras de temperatura y desviación de la temperatura respecto a la media procede de una población normal**.

Si aplicamos el Teorema Central del Límite, al ser el tamaño de la muestra de 41 individuos (mayor de 30), podemos considerar que las muestras que aparecen en la tabla del comienzo, tienen unas medias que proceden de una normal.

8.2. Correlación entre fallecimientos y temperatura

Hacemos el siguiente contraste de hipótesis:

$H_0 \equiv$ Las muestras de temperatura y de fallecimientos son independientes.

$H_1 \equiv$ Las muestras de temperatura y de fallecimientos están relacionadas.

Efectuamos un test de correlación con el programa IBM SPSS Statistics versión 24.0.0.0. El test de correlación que empleamos es el de de Pearson, indicado cuando las poblaciones son normales y paramétricas. Obtenemos los siguientes resultados:

Correlaciones

		temperatura	defunciones
temperatura	Correlación de Pearson	1	,107
	Sig. (bilateral)		,518
	N	39	39
defunciones	Correlación de Pearson	,107	1
	Sig. (bilateral)	,518	
	N	39	39

Figura 5: Test de independencia

Como el p-valor es 0,518 > 0,05, el test no es significativo y no se puede rechazar la hipótesis nula. Por tanto con un nivel de confianza del 95 % podemos afirmar que las poblaciones temperatura y defunciones son independientes una de otra.

8.3. Correlación entre fallecimientos y desviación de la temperatura respecto a la media

Hacemos el siguiente contraste de hipótesis:

$H_0 \equiv$ La muestra de fallecimientos y desviación de la temperatura respecto a la media son independientes.

$H_1 \equiv$ La muestra de fallecimientos y desviación de la temperatura respecto a la media están relacionadas.

Efectuamos un test de correlación con el programa IBM SPSS Statistics versión 24.0.0.0. El test de correlación que empleamos es el de de Pearson, indicado cuando las poblaciones son normales y paramétricas. Obtenemos los siguientes resultados:

Correlaciones

		defunciones	delta_t
defunciones	Correlación de Pearson	1	,107
	Sig. (bilateral)		,518
	N	39	39
delta_t	Correlación de Pearson	,107	1
	Sig. (bilateral)	,518	
	N	39	39

Figura 6: Prueba de normalidad para la población

Como el p-valor es 0,518 > 0,05, el test no es significativo y no se puede rechazar la hipótesis nula. Por tanto con un nivel de confianza del 95 % podemos afirmar que las poblaciones desviación de la temperatura respecto a la media y defunciones son independientes una de otra.

8.4. Correlación entre fallecimientos y población

Hacemos el siguiente contraste de hipótesis:

$H_0 \equiv$ Las muestras de población y de fallecimientos son independientes.

$H_1 \equiv$ Las muestras de población y de fallecimientos están relacionadas.

Efectuamos un test de correlación con el programa IBM SPSS Statistics versión 24.0.0.0. El test de correlación que empleamos es el de de Pearson, indicado cuando las poblaciones son normales y paramétricas. Obtenemos los siguientes resultados:

Correlaciones

		defunciones	habitantes
defunciones	Correlación de Pearson	1	,980**
	Sig. (bilateral)		,000
	N	39	39
habitantes	Correlación de Pearson	,980**	1
	Sig. (bilateral)	,000	
	N	39	39

**. La correlación es significativa en el nivel 0,01 (bilateral).

Figura 7: Test de independencia

Como el p-valor es 0,000 < 0,05 el test es significativo, se rechaza la hipótesis nula . Con un nivel de significación del 5 % afirmamos que el número de fallecimientos en cada comarca depende de su población. Además esta dependencia es directa, al ser el coeficiente de correlación positivo (su valor es 0.980), al aumentar la población aumenta también el número de muertos.

8.5. Correlación entre tasa cruda de fallecimientos y población

Hacemos el siguiente contraste de hipótesis:

$H_0 \equiv$ Las muestras de población y tasa cruda fallecimientos son independientes.

$H_1 \equiv$ Las muestras de población y tasa cruda de fallecimientos están relacionadas.

Efectuamos un test de correlación con el programa IBM SPSS Statistics versión 24.0.0.0. El test de correlación que empleamos es el de de Pearson, indicado cuando las poblaciones son normales y paramétricas. Obtenemos los siguientes resultados:

Correlaciones

		defunciones	tasastandar10000
defunciones	Correlación de Pearson	1	,395*
	Sig. (bilateral)		,013
	N	39	39
tasastandar10000	Correlación de Pearson	,395*	1
	Sig. (bilateral)	,013	
	N	39	39

*. La correlación es significativa en el nivel 0,05 (bilateral).

Figura 8: Test de independencia

Como el p-valor es $0{,}013 < 0{,}05$ el test es significativo, se rechaza la hipótesis nula . Con un nivel de significación del 5 % afirmamos que la tasa cruda de fallecimientos en cada comarca y su número de habitantes están relacionados. Como el coeficiente de correlación es $r = 0{,}395$, a mayor población corresponde una mayor tasa cruda de fallecimientos.