Teresa Ventaja

Higher Diploma in Science in Data Analytics

Covid-19 insights, anova & financial analysis and Regression

Business Data Analysis

A circuit board

Description automatically generated

INDEX

[1. Personal Insights COVID-19 Coronavirus Pandemic 2](#_Toc39612716)

[2. Are apartment rents significantly different in the four regional Irish cities? 5](#_Toc39612717)

[3. Central Statistics Office Data and Building Activity 8](#_Toc39612718)

[4. Central Statistics Office Data continued 10](#_Toc39612719)

[BIBLIOGRAPHY 12](#_Toc39612720)

# 1. Personal Insights COVID-19 Coronavirus Pandemic

Little did we know what the future was going to bring us back in January 2020. In Europe, we heard in the news about a new disease creating difficulties in certain areas of China. However, we were not watchful about the possibility of a worldwide extension. Now we criticize some governments for not being mindful about the statistical probabilities of Covid-19 expansion. Entropy is a real, and up to certain extend, we are surrounded by chaos, but in 21st Century, science is advanced enough to have a lot to say. In this case, we had numbers to reflect on, statistical figures emerged since the very beginning of the pandemic, and we have enough computing capacity to measure what is happening.

Descriptive statistics is often underrated, delegated to a second place in favor of predictions. Time is money, and so data is. Big data is widely used to calculate the future, mostly in the economic world. However, one of the lessons that we can get from this pandemic is that an accurate read of the present can help making appropriate decisions. Another key point we learned the hard way is about the capacity of intensive health care systems:

A close up of a map

Description automatically generated[[1]](#footnote-1)

The most exponential was the curve of critical cases, the least capable we are of saving lives. At the time we were seeing this graph in the news, we started to be bombarded by statistical figures in the medias. This enormous amount of numbers is not always easy to interpret, especially because they are collected with different criteria. Among the most popular statistics, we have new infections and new deaths, usually shown on linear and logarithmic scales:

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

Since the growth of the epidemic is exponential, the logarithmic scale line shows more stability on showing historical series[[2]](#footnote-2)[[3]](#footnote-3). Data collection methodology is something else to take into consideration when comparing country data and time series, but the effective reproduction rate (Rt)[[4]](#footnote-4) has been considered the key metric to predict the transmission of the illness. However, it is common that even within the same country the methodology changes at a certain point[[5]](#footnote-5), but generally speaking, the more testing is done, the more reliable is the data[[6]](#footnote-6). On the top of that, analyzing all this data bring its own complexities[[7]](#footnote-7).

To illustrate an example, we can ask ourselves how do we know the actual number of deaths. Most countries do not have the capacity to make an autopsy to every death body and are not even counting deaths resulted from epidemic spots such us retirement homes[[8]](#footnote-8)[[9]](#footnote-9). Yes, we can use death per million ratios to compare countries[[10]](#footnote-10), but the differences we find are more than significant. On the other hand, we praise South Korea for being an example of successful measurement and containment of the epidemic[[11]](#footnote-11), but that scenario seems more complicated to be implemented in Western countries due to its ethical consequences. However, this topic is open to discussion[[12]](#footnote-12), and that discussion is not a new one, but the classic freedom vs. security paradox.

In terms of predictions, an article in Science[[13]](#footnote-13) foresees that resurgence on contagion could happen as late as 2024. A similar study from CIDRAP[[14]](#footnote-14) agrees that at least this pandemic will have sporadic spikes during the next 2 years. Even if it does not, we will need to stay vigilant about critical care capacity. Another relevant report for understanding different scenarios on the spread of the pandemic is the famous infectious disease report of the Imperial College of London[[15]](#footnote-15)[[16]](#footnote-16).

Economic figures do not look good either. Goldman Sachs report[[17]](#footnote-17) anticipated a drop on GDP in the US, which will undoubtedly expand to the EU, and it even seems unavoidable to speak about a global recession[[18]](#footnote-18)[[19]](#footnote-19). At an early stage, best optimistic forecasts talked about a V-shaped recovery of the economy, but that does not seem realistic anymore[[20]](#footnote-20). Historically, the Spanish flu has shown a W-shaped upturn, more probable with premature withdrawal of public-health or economic-stimulus measures[[21]](#footnote-21).

A screenshot of a cell phone

Description automatically generated[[22]](#footnote-22)

Finally, understanding the transmission of the virus is essential for prevention. A modeling on aerodynamics of saliva droplets for walking and running[[23]](#footnote-23) indicates that the initially though social distance was not enough. 2 meters while walking is more convenient than 1. The scarcity of face masks has led to inadequate recommendations from the authorities, but scientifically[[24]](#footnote-24) and historically[[25]](#footnote-25), it is no doubt that they have an essential role in transmission prevention. Most importantly, at least since 1840’s we know that hand washing is key for preventing any infectious diseases[[26]](#footnote-26), and we are discovering more and more about how long the virus survives on surfaces[[27]](#footnote-27). Nevertheless, we focus our hopes about ending this unprecedented catastrophe on developing a vaccine, which is estimated to be in the market no earlier than the beginning of 2021[[28]](#footnote-28). Fortunately, we are now seeing major advancements on the discovery of effective antibodies which may offer potential for prevention and treatment of COVID-19[[29]](#footnote-29).

# 2. Are apartment rents significantly different in the four regional Irish cities?

Details on the calculations can be found as follows:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A picture containing white

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. *Specify the null and alternate hypotheses to determine if there is a significant difference between the mean rents in the three cities.*

H0 (null hypotheses): µ Cork = µ Galway = µ Waterford

H1 (alternative hypotheses): at least 2 cities means differ

1. *Using the two-bed apartment quarterly rent data for the three selected cities, calculate the test statistic to test your hypotheses and report your result.*

Clearly F > C (79.16 > 3.10) and as such we reject H0 in favour of H1 and infer that at least one pairing of groups are different to each other with 95% confidence.

1. *Report your overall outcome, clearly explain your reason for choosing this particular statistical test, your reasoning behind the process you followed, any decisions you made and any limitations on the outcome.*

I used the ANOVA test of hypotheses[[30]](#footnote-30), because we have 3 independent variables (3 independent samples). Having used independent t-tests for pairs of samples would have led to an explosion on type 1 error, but this ANOVA test is ideal in this case because it is the statistical technique for comparing means for multiple independent populations (usually 3 or more). The design is completely randomised, so I performed the calculations on excel, and after I used the Data Analysis tool-pack. Fortunately, the results matched, so I can infer that the calculations were correct.

I encounter 2 limitations, the first was on how to create a formula to interpret the null hypotheses. I did not know how to represent it with mathematical symbols, so I wrote manually “at least 2 city means differ”. Otherwise the formula would have been too big. Another limitation was finding the critical values on the ANOVA table. Some tables were not big enough to cover every single critical value, and they start to jump from 20 to 30, for example, not showing 21, 22, etc. I found this table on the internet[[31]](#footnote-31), which helped me to find the proper value.

# 3. Central Statistics Office Data and Building Activity

I calculated the Simple Moving Average, Weighed Moving Average and Exponential Moving Average. Following the instructions of the exercise I had to use Q1 and Q3 starting on 2015[[32]](#footnote-32). One of the limitations I found was not having enough values to follow the instructions “use six periods for averaging calculations”. I simply did not have enough values to do so, so I decided to base 2 calculations on less than 6 periods. Every decision I took is explained in Comments in Excel, to makes it easy to understand what I did and why.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

My results indicate that the Exponential moving average differs from the Simple Moving Average, decreasing by 12.16 basic points. However, the Weighed Moving Average and the Exponential Moving Average only differ by 0.2 basic points, so they are very similar. Averages are trend-following indicators[[33]](#footnote-33). The most relevant for this analysis in my opinion is the weighed moving average, because it gives more relevance to recent data, which is useful to describe the behaviour of the price. The exponential moving average is a good predictor even when there are sudden changes on prices, whereas the simple moving average is a good descriptor of a long-term time series.

# 4. Central Statistics Office Data continued

To build this graph, I had to trim data and get rid of Q2 and Q4 to make it match with the data I built on Question 3.

A screenshot of a cell phone

Description automatically generated

I created the following scatter plot in Excel:

We can see that a single increase in a basic point in the Production in Building and Construction Index will result in a 58% increase on the Residential Property Price Index. On the other hand, R2 value indicates that the model explains 82% of the variability of the Residential Property Price index around its mean.

# BIBLIOGRAPHY

* Abdelmalek, M., Francescani, C. and Folmer, K. (2020): “How accurate is the US coronavirus death count? Some experts say it's off by 'tens of thousands'”, abcNEWS. Available at: <https://abcnews.go.com/Health/accurate-us-coronavirus-death-count-experts-off-tens/story?id=70385359> [Accessed 05/05/2020]

Health Emergencies Preparedness and Response (2020): “Advice on the use of masks in the context of COVID-19”, *World Health Organization*. Available at: <https://www.who.int/publications-detail/advice-on-the-use-of-masks-in-the-community-during-home-care-and-in-healthcare-settings-in-the-context-of-the-novel-coronavirus-(2019-ncov)-outbreak> [Accessed 05/05/2020]

* Akpan, N (2020): “How to measure your nation’s response to coronavirus”, *National Geographic.* Available at: <https://www.nationalgeographic.com/science/2020/05/coronavirus-how-to-measure-your-nation-response-cvd/> [Accessed 05/05/2020]

Blocken, B., Malizia, F., Van Druenen, T., Marchal, T. (2020): “Towards aerodynamically equivalent COVID-19 1.5 m social distancing for walking and running”, *Eindhoven University of Technology.* Available at: <http://www.urbanphysics.net/COVID19_Aero_Paper.pdf> [Accessed 05/05/2020]

* Cohn, S. (2020): “Face masks: what the Spanish flu can teach us about making them compulsory”, *The Conversation*. Available at: <https://theconversation.com/face-masks-what-the-spanish-flu-can-teach-us-about-making-them-compulsory-137648> [Accessed 05/05/2020]

Crook, C. (2020): “About That V-Shaped Recovery ...”, *Bloomberg*. Available at: <https://www.bloomberg.com/opinion/articles/2020-04-28/v-shaped-economic-recovery-from-coronavirus-is-not-happening> [Accessed 05/05/2020]

Ferguson, M., Dorigatti, I., Bhatt, S. et al (2020): “Report 20 - Using mobility to estimate the transmission intensity of COVID-19 in Italy: A subnational analysis with future scenarios”, *Imperial College of London*. Available at: <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-20-italy/> [Accessed 05/05/2020]

* Fleming, S. (2020): “South Korea's Foreign Minister explains how the country contained COVID-19”, *World Economic Forum.* Available at: <https://www.weforum.org/agenda/2020/03/south-korea-covid-19-containment-testing/> [Accessed 05/05/2020]

Frankel, J. (2020): “Avoiding A W-Shaped Recession”, *THE ASEAN POST.* Available at: <https://theaseanpost.com/article/avoiding-w-shaped-recession> [Accessed 05/05/2020]

Harilal, K. N. (2020): “World Economy and Nation States post COVID-19”, *Economic & Political WEEKLY*. Available at: <https://www.epw.in/journal/2020/18/commentary/world-economy-and-nation-states-post-covid-19.html> [Accessed 05/05/2020]

* Hasell, J., Ortiz-Ospina, E., Mathieu, E., Ritchie, H., Beltekian, D., Macdonald, B. and Roser, M. (2020): *“To understand the global pandemic, we need global testing – the Our World in Data COVID-19 Testing dataset”,* Our World in Data. Available at: <https://ourworldindata.org/covid-testing> [Accessed 05/05/2020]

Hatzius, J. et al. (2020): “US Daily: A Sudden Stop for the US Economy”, *Goldman Sachs*. Available at:

<https://www.goldmansachs.com/insights/pages/gs-research/us-daily-20-mar-2020/report.pdf> [Accessed 05/05/2020]

* Hirsch, C. and Martuscelli, C. (2020): “The challenge of counting COVID-19 deaths”, *POLITICO.* Available at: <https://www.politico.eu/article/coronavirus-the-challenge-of-counting-covid-19-deaths/> [Accessed 05/05/2020]

Kissler, S. M., Tedijanto, C., Goldstein, E., H. Grad, Y. and Lipsitch, M. (2020): “Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period”, *Science.* Available at: <https://science.sciencemag.org/content/early/2020/04/24/science.abb5793> [Accessed 05/05/2020]

* Knight, W. (2020): “The Value and Ethics of Using Phone Data to Monitor Covid-19”, *WIRED*. Available at: <https://www.wired.com/story/value-ethics-using-phone-data-monitor-covid-19/> [Accessed 05/05/2020]
* 2018: “Linear Graphs VS Logarithmic Graphs”, *Diaman R&D.* Available at: <https://www.diaman.eu/single-post/2018/01/13/Linear-Graphs-VS-Logarithmic-Graphs> [Accessed 05/05/2020]

Moore, K. A., Lipsitch, M., Barry, J. M., Osterholm, M. T. (2020): “COVID-19: The CIDRAP Viewpoint”, *University of Minnesota*. Available at: <https://www.cidrap.umn.edu/sites/default/files/public/downloads/cidrap-covid19-viewpoint-part1_0.pdf> [Accessed 05/05/2020]

2020: “Production in Building and Construction Index”, *Central Statistics Office*. Available at: <https://www.cso.ie/en/statistics/construction/productioninbuildingandconstructionindex/> [Accessed 02/05/2020]

2020: “Residential Tenancies Board (RTB) statistics hosted by the CSO”, *Central Statistics Office.* Available at: <https://statbank.cso.ie/px/pxeirestat/Statire/SelectVarVal/Define.asp?maintable=RIQ02&PLanguage=0> [Accessed 02/05/2020]

* Roser, M., Ritchie, H., Ortiz-Ospina, E. and Hasell, J. (2020): “Total confirmed COVID-19 deaths per million people, May 5, 2020”, *Our World in Data*. Available at: <https://ourworldindata.org/grapher/total-covid-deaths-per-million> [Accessed 05/05/2020]

2020: “SARS-CoV-2 Transmission”, *worldometer*. Available at: <https://www.worldometers.info/coronavirus/transmission/> [Accessed 05/05/2020]

* Shields, C. (2020): “Change in Covid-19 counting method sees surge in case and death data”, *CatalanNews by ACN.* Available at: <https://www.catalannews.com/society-science/item/change-in-covid-19-counting-method-sees-surge-in-case-and-death-data> [Accessed 05/05/2020]
* Singhal, S., Finn, P., Kumar, P., Craven, M. and Smit, S. (2020): “Critical care capacity: The number to watch during the battle of COVID-19”, *McKinsey & Company*. Available at <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/critical-care-capacity-the-number-to-watch-during-the-battle-of-covid-19> [Accessed 05/05/2020]
* 2005: “Table of critical values for the F distribution (for use with ANOVA)”, *University of Sussex.* Available at: <http://users.sussex.ac.uk/~grahamh/RM1web/F-ratio%20table%202005.pdf> [Accessed 02/05/2020]
* 2007: “Technical Analysis Averages”, *Stocata*. Available at: <http://stocata.org/ta_en/averages.html> [Accessed 02/05/2020]

Thanh Le, T., Andreadakis, Z., Kumar, A., Gómez Román, R., Tollefsen S., Saville, M. and Mayhew, S. (2020): “The COVID-19 vaccine development landscape”, *Nature*. Available at: <https://www.nature.com/articles/d41573-020-00073-5> [Accessed 05/05/2020]

Viana, I. (2020): “¿Por qué los médicos se negaron a lavarse las manos hasta casi el siglo XX?”, *ABC Historia*. Available at: <https://www.abc.es/historia/abci-coronavirus-medicos-negaron-lavarse-manos-hasta-casi-siglo-202004061902_noticia.html?ref=https%3A%2F%2Fwww.google.com%2F> [Accessed 05/05/2020]

Wang, C., Li, W., Drabek, D., Okba, N. M. A., Van Haperen, R., Osterhaus, A. D. M. E., Van Kuppeveld, F. J. M., Haagmans, B. L., Grosveld, F. and Bosch, B. J. (2020): “A human monoclonal antibody blocking SARS-CoV-2 infection”, *Nature.* Available at: <https://www.nature.com/articles/s41467-020-16256-y> [Accessed 05/05/2020]

Winck, B. (2020): “The IMF says its forecast for the COVID-19 recession might now be too optimistic”, *World Economic Forum*. Available at: <https://www.weforum.org/agenda/2020/04/imf-economy-coronavirus-covid-19-recession/> [Accessed 05/05/2020]

* Wolkewitz, M. and Puljak, L. (2020): “Methodological challenges of analysing COVID-19 data during the pandemic”, *BMC*. Available at: <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-020-00972-6> [Accessed 05/05/2020]

1. Singhal, S., Finn, P., Kumar, P., Craven, M. and Smit, S. (2020) [↑](#footnote-ref-1)
2. 2018: “Linear Graphs VS Logarithmic Graphs” [↑](#footnote-ref-2)
3. This is useful to compare exponential growths over time, whereas a linear scale would tend to the infinite (although the world population is not infinite, the graphical representation would initially appear so). [↑](#footnote-ref-3)
4. Akpan, N (2020) [↑](#footnote-ref-4)
5. Shields, C. (2020) [↑](#footnote-ref-5)
6. Hasell, J., Ortiz-Ospina, E., Mathieu, E., Ritchie, H., Beltekian, D., Macdonald, B. and Roser, M. (2020) [↑](#footnote-ref-6)
7. Wolkewitz, M. and Puljak, L. (2020) [↑](#footnote-ref-7)
8. Hirsch, C. and Martuscelli, C. (2020) [↑](#footnote-ref-8)
9. Abdelmalek, M., Francescani, C. and Folmer, K. (2020) [↑](#footnote-ref-9)
10. Roser, M., Ritchie, H., Ortiz-Ospina, E. and Hasell, J. (2020) [↑](#footnote-ref-10)
11. Fleming, S. (2020) [↑](#footnote-ref-11)
12. Knight, W. (2020) [↑](#footnote-ref-12)
13. Kissler, S. M., Tedijanto, C., Goldstein, E., H. Grad, Y. and Lipsitch, M. (2020) [↑](#footnote-ref-13)
14. Moore, K. A., Lipsitch, M., Barry, J. M., Osterholm, M. T. (2020) [↑](#footnote-ref-14)
15. Ferguson, M., Dorigatti, I., Bhatt, S. et al (2020) [↑](#footnote-ref-15)
16. Social distancing measures, community surveillance, contact tracing and the early isolation of infections are key for reducing the risk of resurgence in transmission. [↑](#footnote-ref-16)
17. Hatzius, J. et al. (2020) [↑](#footnote-ref-17)
18. Winck, B. (2020) [↑](#footnote-ref-18)
19. Harilal, K. N. (2020) [↑](#footnote-ref-19)
20. Crook, C. (2020) [↑](#footnote-ref-20)
21. Frankel, J. (2020) [↑](#footnote-ref-21)
22. Hatzius, J. et al. (2020) [↑](#footnote-ref-22)
23. Blocken, B., Malizia, F., Van Druenen, T., Marchal, T. (2020) [↑](#footnote-ref-23)
24. Health Emergencies Preparedness and Response (2020) [↑](#footnote-ref-24)
25. Cohn, S. (2020) [↑](#footnote-ref-25)
26. Viana, I. (2020) [↑](#footnote-ref-26)
27. 2020: “SARS-CoV-2 Transmission” [↑](#footnote-ref-27)
28. Thanh Le, T., Andreadakis, Z., Kumar, A., Gómez Román, R., Tollefsen S., Saville, M. and Mayhew, S. (2020) [↑](#footnote-ref-28)
29. Wang, C., Li, W., Drabek, D., Okba, N. M. A., Van Haperen, R., Osterhaus, A. D. M. E., Van Kuppeveld, F. J. M., Haagmans, B. L., Grosveld, F. and Bosch, B. J. (2020) [↑](#footnote-ref-29)
30. Data source: <https://statbank.cso.ie/px/pxeirestat/Statire/SelectVarVal/Define.asp?maintable=RIQ02&PLanguage=0> [↑](#footnote-ref-30)
31. Table source: http://users.sussex.ac.uk/~grahamh/RM1web/F-ratio%20table%202005.pdf [↑](#footnote-ref-31)
32. Data Source: https://www.cso.ie/en/statistics/construction/productioninbuildingandconstructionindex/ [↑](#footnote-ref-32)
33. 2007: “Technical Analysis Averages” [↑](#footnote-ref-33)