# Assignment #3 – ECON 323 – Winer 2022

Professor Mikal Skuterud
Date assigned: January 25, 2022 at 9:00
Date due: January 31, 2022 at 23:59

---

In this assignment you will examine a single time-series of data. The data are **daily new COVID-19 cases in Ontario** and they can be downloaded from the Public Health Ontario website here: https://www.publichealthontario.ca/en/data-and-analysis/infectious-disease/covid-19-data-surveillance/covid-19-data-tool. Scroll down the page to where you see the blue tabs and click on "Case trends." Then select daily reported cases (counts) for all of Ontario. The first positive COVID-19 case reported in Ontario was on January 23, 2020 but we want to restrict our attention to the (ongoing) Omicron wave. Therefore, restrict your time-series from December 1, 2021 to the most recent day available when you download the data. To choose the date range, use the slider button and then click "Download" and "Download shown data."

Once you have downloaded and saved the data file, you need to get the file into R. There is more than one way to do this, but I would recommend saving the file as a CSV file in Excel and using the function `read_csv` to load the date file into R. Your dataset contains four variables: (i) the date; (ii) the number of new confirmed cases on that day; (iii) the rate per 100,000 population; and (iv) the 7-day rolling average. You will only need the first two variables for this assignment. Once you have imported your data file into R you need to estimate the models below and interpret your estimates.

1. Estimate each of the 4 four models below separately:

$$newcases_t = \beta_0 + \beta_1 t + u_t \qquad \leftarrow \text{linear trend}$$
$$newcases_t = \beta_0 + \beta_1 t + \beta_2 t^2 + u_t \qquad \leftarrow \text{quadratic trend}$$
$$newcases_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + u_t \qquad \leftarrow \text{cubic trend}$$
$$newcases_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + u_t \qquad \leftarrow \text{quartic trend}$$

where $t$ is a time variable which equals 1 on the first day of your time series and increases by 1 every day. After estimating each regression, produce graph that plots two lines: (i) the actual daily cases; and (ii) the predicted number of cases from your estimated model. Make sure the lines are distinguishable from each other by using different line patterns (solid and dashed lines, for example) or different colours. In the quartic model, the predicted number of cases is:

$$\widehat{newcases}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\beta}_3 t^3 + \hat{\beta}_4 t^4 .$$

Which model do you think explains daily COVID-19 cases in Ontario best? Explain your reasoning. Do you think that your preferred model is useful for predicting future cases of COVID-19? What does your model predict new cases will be on March 1, 2022? Do you think this is a reasonable estimate?

2. As discussed in our lecture, COVID-19 cases tend to grow exponentially, which suggests that we should take a logarithmic transformation of the dependent variable. Estimate each of the four models in question 1 using log new cases as your dependent variable. Plot the actual and predicted values of your new dependent variable in a graph. Do you think this model explains the variation in new cases better? Explain your reasoning.

3. Are new COVID-19 cases related to the day of the week? To determine this, add day-of-the-week dummy variables to the regression you estimated in question 2. Remember to drop one dummy variable to avoid the dummy variable trap. If you drop the Sunday dummy variable, for example, then you should include dummies for Monday through Saturday (6 dummies). Are any of the coefficients on the day-of-the-week dummies statistically significant at the 5% significance level? Are the day-of-the-week dummy variables jointly significant? In other words, are you able to reject the null hypothesis that the true value of all 6 coefficients is equal to zero? Interpret your results. What do you think explains the significant differences in cases between days of the week?

4. As described on slide #24 of the section2(1) lecture slides, dummy variables can also be used to identify the impact of an event. This is sometimes called an "event study." It is believed that infections of the Omicron variant typically take 3 days to show up as symptoms in infected individuals. Is there any evidence in the data of heightened infections on Christmas Day? Create a dummy variable equal to 1 on December 28 and equal to 0 on all other days. Does the coefficient on your event dummy suggest that family gatherings on Christmas Day resulted in increased COVID-19 infections? How big is the estimated impact? Interpret the point estimate and its standard error precisely.