Estimating excess of deaths during COVID-19 pandemics - A step-by-step tutorial and guidance

Working paper 01/20 - Global Platform for the Rapid Generation and Transfer of Knowledge| on COVID-19 and older adults in low and middle-income countries (GP-Older-COVID)

Lucas Sempé* Peter Lloyd-Sherlock†

15 July 2020

1 Introduction

This tutorial provides a step-by-step guide for the generation of robust excess mortality estimates. Excess mortality estimates the number of additional deaths occurring over a given time period under specific conditions (in this case, the presence of COVID-19), compared to the number of deaths we might reasonably expect for the same period, based on historical data. This captures both deaths directly attributed to COVID-19 and those resulting less directly from the pandemic (such as conditions that go untreated due to reduced access to health services). Together, these provide a complete view of the mortality impact of the pandemic over a specific period of time (Leon et al. 2020).

As part of the tutorial, we provide an example from Peru, one of the countries that has been most affected by the pandemic to date. Additionally, we pay particular attention to patterns of excess deaths by age groups, given the strong association between old age and risk of COVID-19 mortality. As of 13 July 2020, official data report 12,054 deaths directly caused by COVID-19 since the pandemic arrived in Peru on March 6th. Of these, 8,354 were reported to have been people aged 60 or more. Our case study will provide a check on the robustness of these official data, as well as an estimate of excess deaths in Peru not reported as caused by COVID-19.

2 Data analysis framework

Data on mortality are usually collected and reported by government agencies such as the civil registries, statistical or health agencies and data collection responsibilities may be divided between them. This reflects the complexity of registering deaths, where medical, legal, administrative, technological cultural factors are intertwined. As a result, there is a significant amount of under-registration of deaths and misidentification of its causes. Also death registration may not occur in a timely way, which further impedes analysis.

Figure 1 sets out a data framework required for estimating excess mortality in the context of COVID-19 pandemic. This shows five interlinked layers of information that have are required. Analysis should start with the outer layer, which refers to the total population, and then proceed inwards, following the arrow.

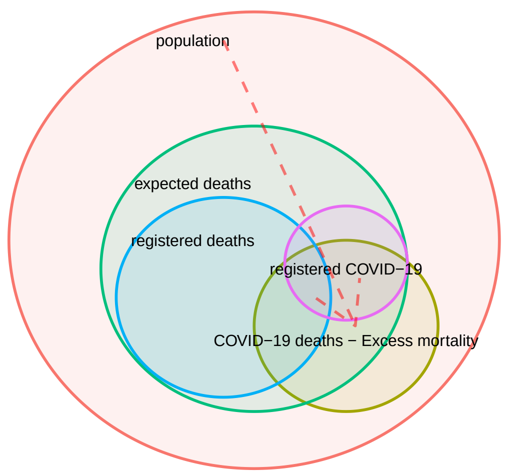
The first layer refers to disaggregated population estimates derived from population projections and/or census data. Obtaining good population estimates may seem a very simple process, but it is not without challenges, as will be shown below.

*University of East Anglia & Univesidad Católica San Pablo - email:l.sempe@uea.ac.uk
†University of East Anglia

The second layer refers to numbers of deaths. Due to incomplete death registration in most low and middle-income countries, mortality data are usually derived from expected deaths based on demographic projections and are measured in terms of death counts and mortality ratios. Mortality estimations are usually inconsistent between different sources reliability and availability of data (Adair and Lopez 2018) and because population projections vary significantly.

The third layer refers to deaths officially registered by different agencies such as hospitals, health professionals or other civil and community authorities. The fourth layer refers to the pandemic itself and provides death whose cause is officially registered as COVID-19.

The fifth layer refers any deaths that were either directly caused by COVID-19 or were more indirectly caused by the pandemic. These include both those deaths that are reported and those not included in official counts, to provide an estimate of excess mortality during the pandemic.



Source: Author's own

Figure 1: Data framework analysis - Excess of mortality

As seen in Figure 1, there are different potential sources of mismatch between the registration and the identification of deaths. These are summarised in Table 1. Both the potential false negatives shown in Table 1 are resolved by the excess of mortality analysis, since it captures deaths indirectly related to the pandemic. It is, however, possible that a number of deaths during the period of interest are due to causes that do not usually occur with the same frequency in previous years (such as an outbreak of an infectious disease like dengue or a natural disaster). These causes may then be wrongly attributed to COVID-19 and the pandemic. This shows the importance of supplementing statistical data with contextual epidemiological information from the field (perhaps obtained from local health professionals). If substantial numbers of deaths are potentially due to causes unrelated to the pandemic and unrelated to previous years, it will be necessary to develop an adjusted model of analysis (a method not dealt with in this tutorial).

Table 1: Possibles sources of bias in estimation of excess of deaths by COVID-19

	Death by COVID-19	Death by other causes
Registered as COVID-19	OK	False positive
Not registered as COVID-19	False negative	False positive
Expected but not registered	False negative	-

3 Empirical strategy

Our estimation of excess mortality utilises time-series analysis applying univariate models to both forecast missing values prior to 2020 and to forecast values for 2020. Time series models assume a linear relationship between certain variables of interest y, such as the mortality rate, expected mortality or registered mortality, both over a specific period of time t and in relation to previous values of the same variable, y_{t-n}. Two different models are used in forecasting population and mortality (Hyndman and Athanasopoulos 2019; Li and Hyndman 2020), namely, non-seasonal Autoregressive integrated moving average models (ARIMA) and a Random walk model with drift (RWD).

ARIMA models combine three elements: autoregressions (AR), differencing (I), and moving averages (MA). The AR refers to linear regressions of the dependent variable against itself on previous periods of time. The Ipart refers to the process of transforming data into a stationary time series, where the mean, variance, and autocorrelation become constant over time. This is needed to eliminate seasonality and longer-run data trends. Finally, the MA indicates the addition of past forecast errors to the model. Combining these three features provides the best possible model fit. The formal notation for the model is given by:

$$y'_t = c + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \cdots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{1}$$

Where y is the predicted value at time t, c refers to the average of changes between consecutive observations, ϕ represent autoregressive parameters, θ represent the moving average parameters and ϵ_t is the model error term, known as white noise. The variance of the error term ϵ_t enables us to compute confidence intervals for the forecasted values.

On other hand, the RWD is a simple model where the forecasted values increase or decrease over time depending on changes derived from consecutive observations. The amount of change over time (called the drift) is set by the average change seen in the historical time series data. The notation is:

$$y_t = c + y_{t-1} + \epsilon_t \tag{2}$$

In the case that c is positive, then the average change is an increase in the value of y_t, while the opposite occurs when $c < 0$.

Non-seasonal ARIMA models are generally classified as ARIMA(p, d, q), where parameter p refers to number of time lags of the autoregressive model, d is the number of times the data have been differentiated, and q is the number of lagged values for the error term that are added or subtracted to y. The RWD is a special case of ARIMA, equivalent to a ARIMA$(0, 1, 0)$ where $\phi_1 = 1$ and $c \neq 0$ in equation (1).

Once the model parameters from (1) and (1) are estimated, we forecast mean values up to the year 2020. As we are dealing with uncertainty, we estimate a range of values which are likely to include the parameter value with a specific level of confidence (in this case, 95%).

The estimated parameter 95% CI interval is given by $\hat{y}_{t+h|T} \pm 1.96\sqrt{(\hat{\sigma}_h)}$ where $\hat{y}_{t+h|T}$ are the predicted values on future periods h and $\hat{\sigma}$ is the standard deviation of the residuals. In case of the first forecast, the 95% CI is $\hat{y}_{t+1|T} \pm 1.96\hat{\sigma}_h$.

Finally, to deal with the propagation of uncertainty when we combine estimations and their confidence intervals, we adjust the residual standard deviations σ such as in:

$$\sigma(\hat{\beta}' * \hat{\beta}'') = \sqrt{(\frac{\sigma_{\hat{\beta}'}}{\hat{\beta}'})^2 + (\frac{\sigma_{\hat{\beta}''}}{\hat{\beta}''})^2} \tag{3}$$

Where β' and β'' represent any pair of forecasted parameters. Finally, in order to check the robustness of the ex-post forecasting, when available, we compare forecasted values with values in official statistics to establish the accuracy of the predictions.

We use the R package `fable`, which allows for automated searches through the model space to identify the best ARIMA model from a set of different p, d and q parameters, which lowest information criteria, in this case, the Akaike information criterion corrected for small sample sizes. In all cases, model parameters and fit are found in the Appendix.

4 Application: Peru

Peru has three natural regions: a narrow dry coast in the west, the Andean region and a large rainforest region to the east, which is the less densely populated. Administratively, it is divided into 25 regions (see Figure 2).

Source: https://github.com/juaneladio/peru−geojson

Figure 2: Map of Peru: administrative subdivisions

The application of the methodological approach is illustrated in the case of measuring COVID-19 deaths over months of March to June 2020 in Peru. Peru is an interesting case of study for different reasons. First, anonymised individual level data for mortality and COVID administrative data is on the public domain. Second, there has been a recent change in the death notification system, which allows for comparing spatio-temporal trends on deaths registration ("Peru Implementation Working Group. Peru: An exceptional example of CRVS system advancement. CRVS country reports." 2018). Finally, official statistics and projections at sub-level and disaggregated by age groups are not update or available. This feature, common across many countries, allows us to apply our method towards addressing missing data points.

Data come from the National Statistics Agency (INEI)and the Ministry of Health (MoH). Table 2 shows missing data, which we will therefore need to estimate to be able compute excess mortality. National level data are available and up-to-date, and can therefore serve as a benchmark for our projections. However, disaggregated local data are not available. Mortality projections were recently adjusted at the national level, and are not consistent with official data for the regional level. Consequently, we also apply a second source of mortality data, the Global Burden of Disease study (GBD), which for Peru includes deaths by sex, range of age and causes from 2008 to 2017 (Dicker et al. 2018).

Table 2: Data availability by categories and time

Data	Level	Prior to 2016	2017	2018	2019	2020
Population estimations	National	yes	yes	yes	yes	yes
Population estimations	Regional	not available	not available	not available	not available	yes
Expected mortality	National	yes	yes	yes	yes	yes
Expected mortality	Regional	not available	not available	not available	not available	missing
Registered mortality	National	not available	yes	yes	yes	yes
Registered mortality	Regional	not available	yes	yes	yes	yes
COVID mortality	National	-	-	-	-	yes
COVID mortality	Regional	-	-	-	-	yes
Deaths - Global Burden disease	National	yes	yes	yes	yes	yes

4.1 First layer: Population

4.1.1 Step 1: Finding and knowing the data

Data on population are usually derived from a combination of periodic census data and projections based on fertility, mortality, migration, life expectancy rates, among other things.

In 2020 INEI estimates the population of Peru was around 32.8 million, with a growth rate of around 1% year. Since older people account for the majority of COVID-19 deaths falls, it is important to understand the evolution of population age structure over time. Figure 3 shows how the population aged 65 and over has significantly increased over the last two decades, with yearly growth rates of over 2.5%. By contrast, younger cohorts show a declining growth rate, and an absolute decline from 2000.

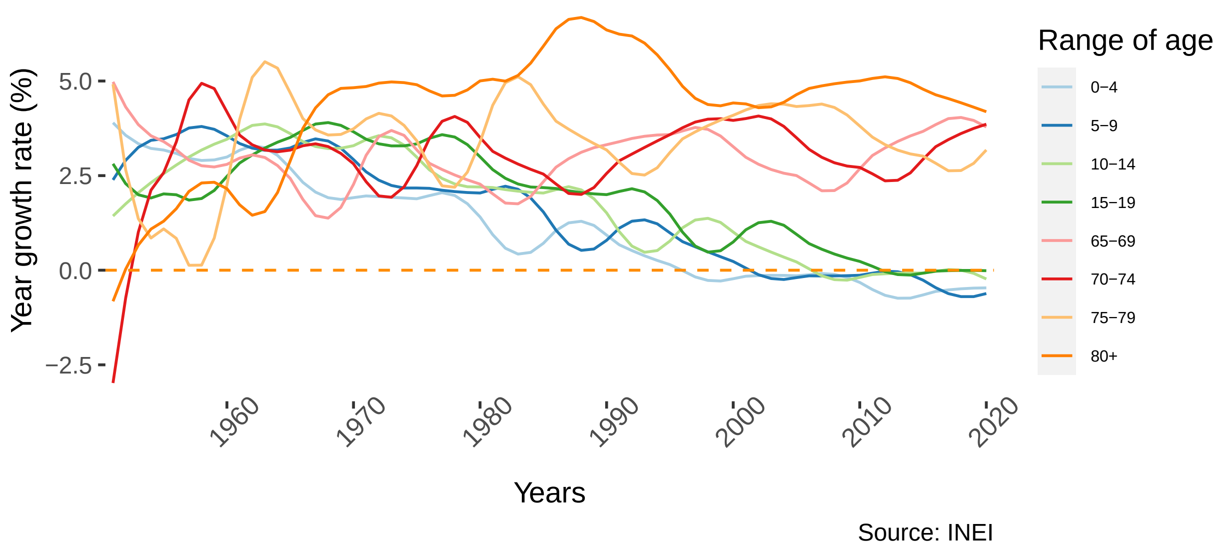


Figure 3: Yearly growth rate by selected range of age - Peru

4.1.2 Step 2: Analysis and forecasting

In Peru there are no official population estimates disaggregated by regions or age. Therefore, we use the two models ARIMA and RWD, to estimate and forecast population by regions and age group up to year

2020, $\widehat{\text{pop}}$. We compare aggregated forecast values against official population projections by INEI. Model parameters and residual analysis can be found in Section 6.

Figure 4 shows small differences between the aggregated forecasted values and INEI projections for 2020 across the different age groups. The mean difference is -1.2% and -0.7% for the RWD and ARIMA models, respectively. The interquartile range also shows lower values, 5.95% and 4.0%, which suggests predicted values are close to official estimates in all cases. The analysis of outliers shows an overestimate of 82,496 people aged 30-39 years in Lima, which represents a difference of 5% between predicted and official values. To understand the magnitude and potential bias of this gap, we compute the number of deaths represented by that population. As the mortality rate of Peru is 5.8 deaths per thousand, this represents 14.2 overestimated deaths. The plot shows that the RWD model tends to underestimate official figures: this will be of use in the analysis, since official estimates have been adjusted downwards in recent times.

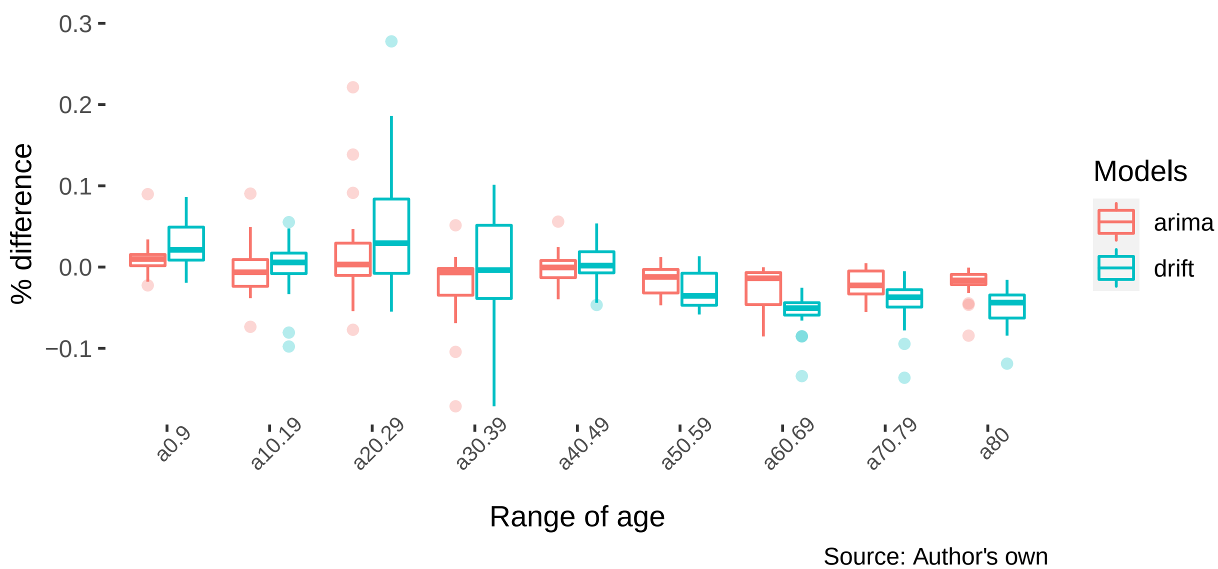


Figure 4: Difference between population estimates for 2020 by forecasting model (ARIMA and RWD drift) by age group

4.2 Second layer: Expected deaths

4.2.1 Step 1: Finding and knowing the data

Expected mortality gives an overall estimate of deaths over a certain period. It is usually measured by the mortality rate (MR), which shows the number of deaths for a fixed population over a year per 1000 people.

Figure 5 shows the evolution of MR over the last two decades in Peru and selected South American countries. In 2018 the expected mortality rate in Peru was 5.83, one of the lowest rates in the region. Peru's MR has remained stable since 2000, unlike Bolivia (which saw a significant decrease), or Chile (an increase).

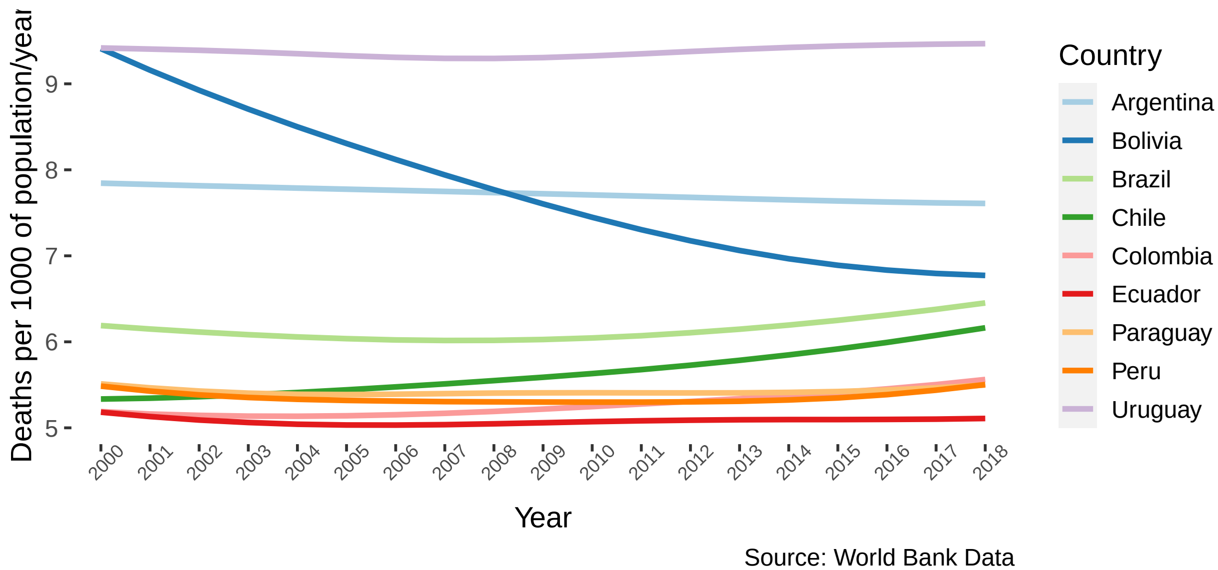


Figure 5: Mortality rate per 1000 population in selected South American countries from 2000 to 2018

While Peru's national MR trend appears to have been stable over the last two decades, a different picture is observed over a longer time period across Departments. Figure 6 shows significant falls many regions, such as Amazonas. In urban centres such as Lima and Callao rates are lower and more stable.

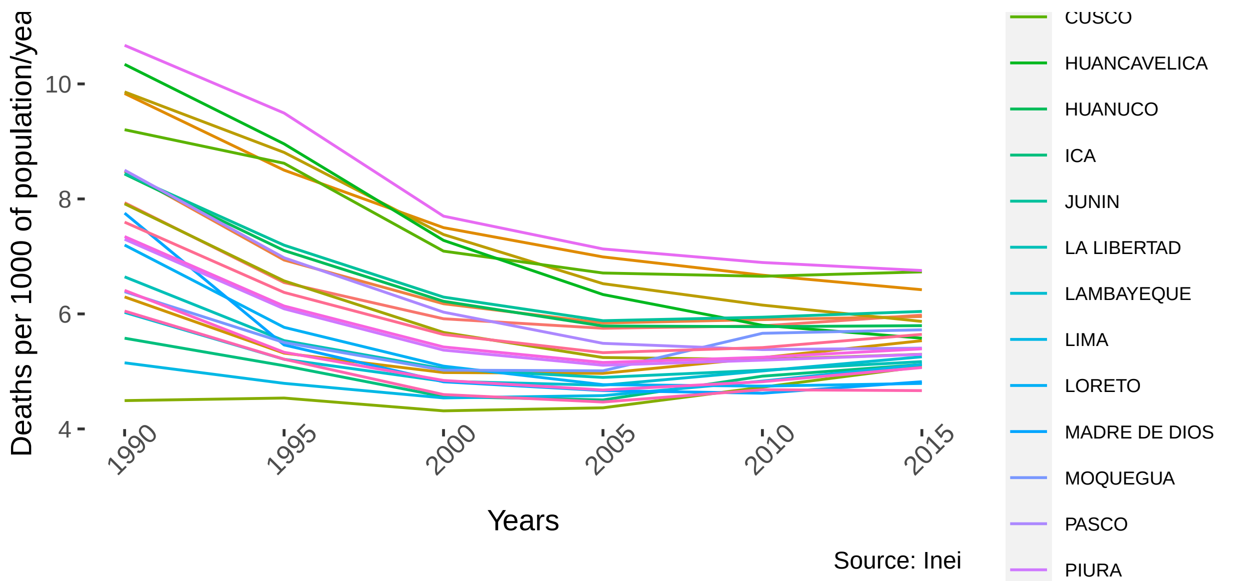


Figure 6: Mortality rate (crude estimates) per 1000 population by region in Peru from 1990 to 2015

4.2.2 Step 2: Analysis and forecasting

To estimate the expected mortality we need two set of values: namely, the mortality rate for the period of interest, MR and the population, such as in $MR = \frac{\text{Deaths}}{\text{population}}$ over a certain period of time.

Data on expected mortality in Peru are more limited than data on population. The INEI only presents projections at the regional level for every 5 years, without disaggregation into age groups. Due to those limitations, we estimate expected mortality by region and age group by using the same analytical tools used above. We compare the aggregation of predicted values with INEI projections for 2020. Model parameters and residual analysis can be found in Section 6.

Figure 7 compares aggregate regional ($\widehat{\mathrm{MR}}$) estimates and INEI's projected mortality rates for 2020, whereby RWD slightly underpredicts rates while ARIMA shows a less clear pattern and higher gaps between predicted and official figures, especially in regions with higher mortality rates.

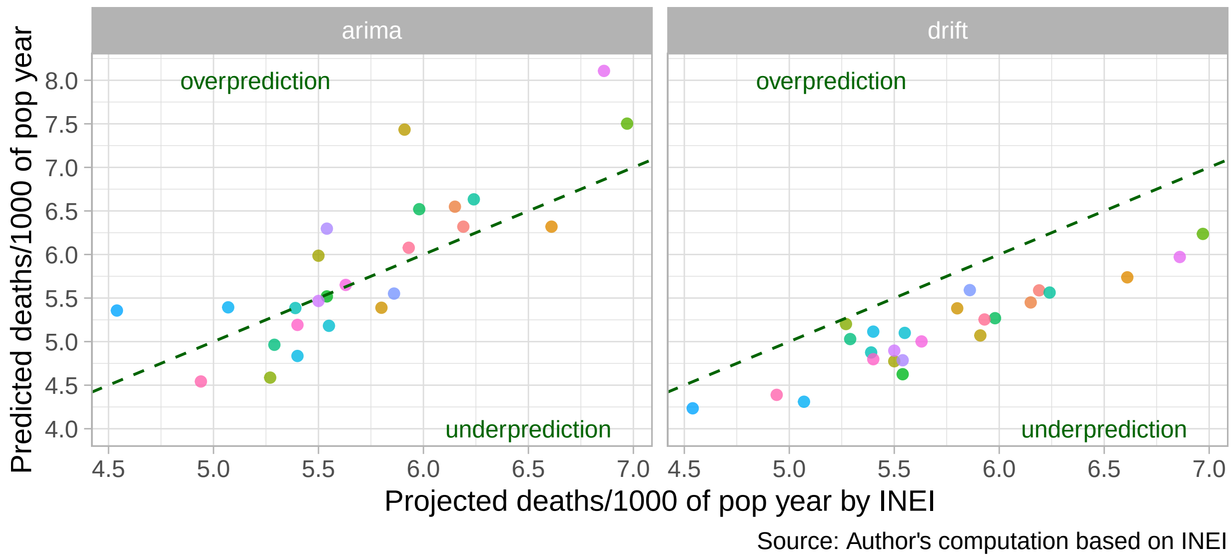


Figure 7: Predicted and projected crude mortality rates by Department - model ARIMA and RWD drift

We then forecast average expected deaths 95% confidence intervals based on previous estimates for population and mortality ratios. We adjust for uncertainty propagation caused by computing values with their respective errors, by computing standard errors following rules of time series data as $\sqrt{\sigma_h}$ based on equation (3), as follows:

$$\widehat{\mathrm{Deaths}}_{\mathrm{Exp}_{\mathrm{INEI}}} = \widehat{\mathrm{MR}} * \widehat{\mathrm{pop}} \pm 1.96 \sqrt{\widehat{\mathrm{MR}} * \widehat{\mathrm{pop}} \sqrt{(\frac{\hat{\sigma}_{\mathrm{MR}}}{\widehat{\mathrm{MR}}})^2 + (\frac{\hat{\sigma}_{\mathrm{pop}}}{\widehat{\mathrm{pop}}})^2}} \tag{4}$$

Table 3 shows the results of four sets of expected mortality estimates taken from the combination of population and mortality rate forecasts. The average of estimated values falls between 169,557 and 184,680, with a 95% CI ranging from 138,937 to 225,361 under RWD models to forecast $\widehat{\mathrm{MR}}$ and $\widehat{\mathrm{pop}}$. As previously mentioned, the selection of mortality ratio models strongly influences the range of estimated values, whereas the population models do not lead to large variations.

We find the RWD range of forecasted values coincide with recent INEI projections for expected mortality, which estimate an average of 172,000 deaths across the period 2015 - 2020. However, earlier more detailed estimates also by INEI projected 191,411 deaths for the same year, which is closer to the ARIMA models forecasting. This suggests the need to estimate excess mortality using both sets of models.

A second source of expected mortality data is the GBD, which provides expected mortality estimates by age group from 1990 to 2017. In 2017, the GBD estimated 141,759 deaths (95% uncertainty interval 123,632 to 161,881). This suggests the existence of different scenarios, which are more conservative compared to the previous model.

In this case, the forecasting exercise is divided into two steps. First, we disaggregate the number of deaths at the regional level. For that, we use INEI projections, which are highly correlated to registered deaths

Table 3: Expected mortality based on population and mortality ratio models

Año	model.mort	model.pop	exp.mortality	lower.CI.mortality	upper.CI.mortality
2020	arima	arima	183856.7	180296.6	187416.8
2020	arima	drift	184679.7	181110.4	188248.9
2020	drift	arima	168914.9	165803.4	172026.3
2020	drift	drift	169629.2	166508.7	172749.7

from SINADEF (r(23)=.993,p<.001). Secondly, we forecast estimated mortality, $\widehat{\text{Deaths}}_{\text{Exp}_{\text{GBD}}}$, by regions and age group up to 2020 based on the expected values provided by the GBD. The computation of a 95% CI interval for in 2020 is given as follows:

$$\widehat{\text{Deaths}}_{\text{Exp}_{\text{GBD}}} = \text{Deat}\widehat{\text{hs}}_{\text{Exp}_{\text{t+h|T}}} \pm 1.96\sqrt{\hat{\sigma}_{\text{Deaths}_{\text{Exp}_{\text{t+h|T}}}}} \tag{5}$$

Where $\text{Deat}\widehat{\text{hs}}_{\text{Exp}_{\text{t+h|T}}}$ are the predicted values on future periods h and $\hat{\sigma}$ is the standard deviation of the residuals.

$\widehat{\text{Deaths}}_{\text{Exp}_{\text{GBD}}}$ for 2020 falls in a range from 133,504 to 161,805 and 134,133 to 156,715 under the RWD and ARIMA models respectively. Table 4 shows how the 95% CI increases significantly for 2020, which implies a greater range of uncertainty.

Table 4: Forecast of expected mortality based on GBD

.model	year	mean	low	up
arima	2018	141498.2	132923.9	150072.5
arima	2019	143475.0	133312.8	153637.2
arima	2020	145423.9	134132.9	156714.9
drift	2018	143723.8	136333.8	151113.8
drift	2019	145689.1	134672.7	156705.4
drift	2020	147654.3	133503.6	161805.0

4.3 Third layer: Registered deaths

4.3.1 Step 1: Finding and knowing the data

There are two sources of registered death data: INEI's death registry and the MOH's SINADEF, which has been running since 2017 ("Peru Implementation Working Group. Peru: An exceptional example of CRVS system advancement. CRVS country reports." 2018). SINADEF permits us to assess gaps between expected and registered mortality, and whether these gaps vary over time. However, coverage of SINADEF is not complete, reaching 74.0% of total deaths in 2018. This may be due to challenges of implementing this new system, especially in more remote areas that lack IT equipment and connectivity. The unreliability of SINADEF data at the regional level can be seen in Figure 8, which shows almost no deaths recorded by SINADEF in Lambayeque in 2019. SINADEF data for Lima appear less inconsistent, and appear to indicate the increased application of the SINADEF system over time. However, these data do not permit forecasting of registered deaths, since many exogenous factors could influence future figures, and the lack of clear patterns, which characterise stationary time series.

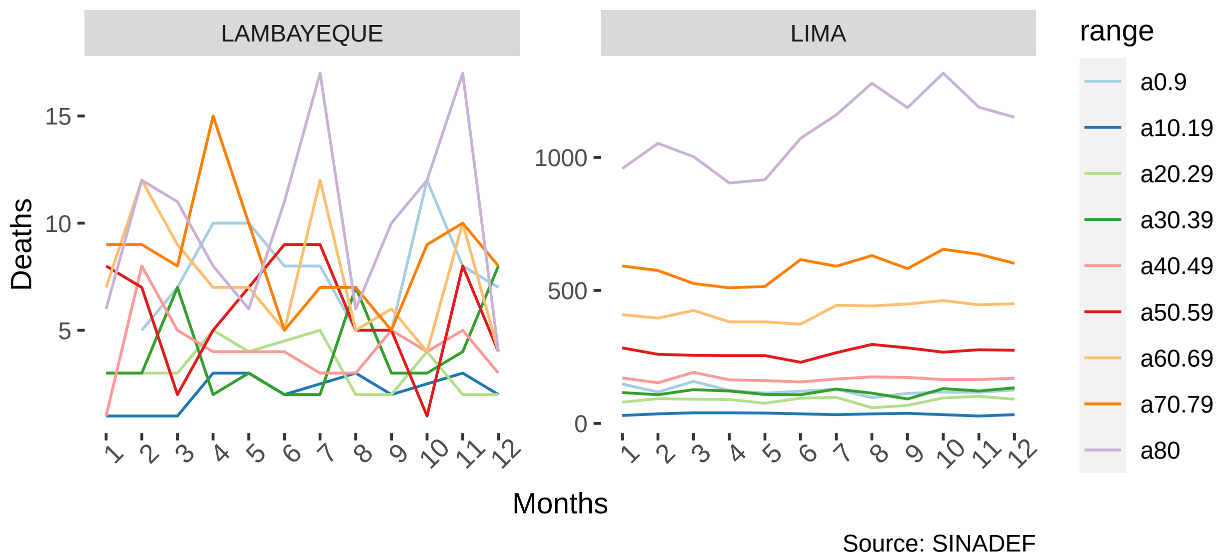


Figure 8: Monthly deaths by age group in Departments of Lambayeque and Lima in 2019

4.3.2 Step 2: Quantification of the completeness of registered deaths

One major gap in mortality estimations is between the number of expected deaths and registered deaths. Quantifying the completeness of a death registration system shows whether coverage is suboptimal. Also, the use of incomplete death registration data to compute excess deaths may under-estimate the true number of excess deaths, which should be adjusted by the registration completeness rate. In this step, we estimate the completeness of registration by regions and age.

For INEI data, we compute the completeness of death registration, $\widehat{\text{Comp.reg}}_{\text{INEI}}$ as a ratio of expected $\text{Deaths}_{\text{Exp}_{\text{INEI}}}$ and registered deaths, $\text{Deaths}_{\text{Reg}}$, as follows:

$$\widehat{\text{Comp.reg}}_{\text{INEI}} = \frac{Deaths_{Reg}}{\widehat{\text{Deaths}}_{\text{Exp}_{\text{INEI}}}} \tag{6}$$

As we are computing this rate combining two previous estimations, we compute a 95% CI addressing the propagation uncertainty based on (3), as follows:

$$\widehat{\text{comp.reg}}_{\text{INEI}} = \frac{\text{Deaths}_{\text{Reg}}}{\widehat{\text{Deaths}}_{\text{Exp}_{\text{INEI}}}} \pm 1.96\sqrt{\widehat{\text{MR}} * \widehat{\text{pop}}\sqrt{(\frac{\hat{\sigma}_{\text{MR}}}{\widehat{\text{MR}}})^2 + (\frac{\hat{\sigma}_{\text{pop}}}{\widehat{\text{pop}}})^2}} \tag{7}$$

Similarly, the completeness rate based on GBD estimations, $\widehat{\text{Comp.reg}}_{\text{GBD}}$ is a ratio between registered deaths, $\text{Deaths}_{\text{Reg}}$ and $\widehat{\text{Deaths}}_{\text{Exp}_{\text{GBD}}}$, as follows:

$$\widehat{\text{Comp.reg}}_{\text{GBD}} = \frac{\text{Deaths}_{\text{Reg}}}{\widehat{\text{Deaths}}_{\text{Exp}_{\text{GBD}}}} \tag{8}$$

Unlike the previous case, estimating the 95% CI does not require adjusting for propagation of uncertainty. Therefore, we compute the 95% CI based on the residuals of the time series, $\widehat{\text{Deaths}}_{\text{GBD}}$, which refers to the forecast values at year 2019, such as in:

$$\widehat{\text{Comp.reg}}_{\text{GBD}} = \frac{\text{Deaths}_{\text{Reg}}}{\widehat{\text{Deaths}}_{\text{Exp}_{\text{GBD}}}} \pm 1.96\sqrt{\hat{\sigma}_{\text{Deaths}_{\text{GBD}}}} \tag{9}$$

Figure 9 shows completeness rates and 95% CIs based on the INEI and the GBD drift models for 2019, by region for people aged 60-69. The plot is divided into four levels. Regions in the red area have less than 50% registration; those in the yellow area 50% to 85%; and those in the green area 85% to 115%. An additional gray area for over 115% shows 6 GBD estimations, which suggests GDB expected mortality may be underestimating deaths in comparison to INEI.

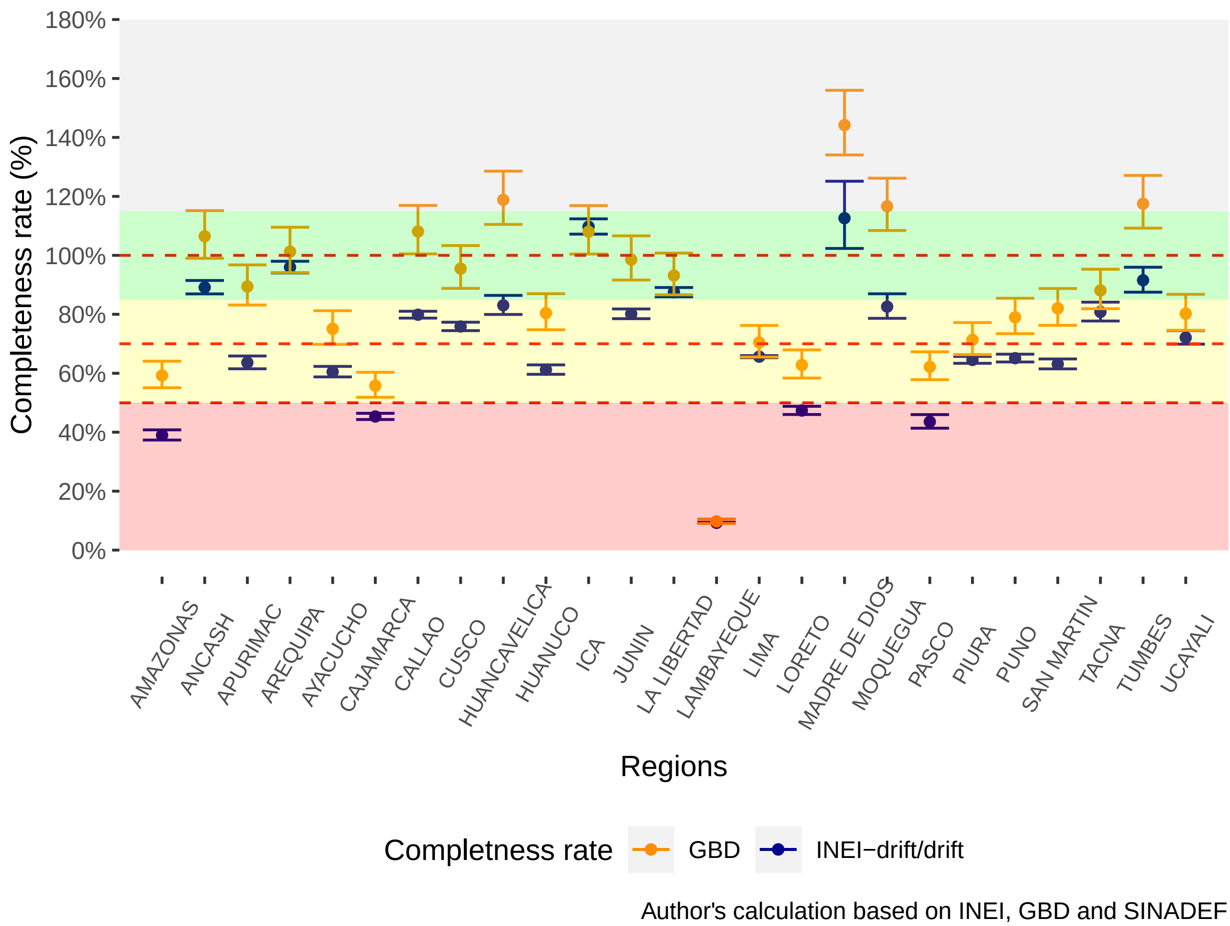


Figure 9: Estimates of completeness rate of registered deaths from SINADEF in the age group 60-69 years by Department in 2019

4.4 Fourth layer: Deaths registered as COVID-19

4.4.1 Step 1: Finding and knowing the data

By July 13 there had been 12,504 deaths registered as caused by COVID-19 according to the MoH. Figure 10 shows three periods of daily reported deaths. An exponential increase between March and April, a significant decrease during May and a sudden increase during June, with figures stabilizing at around 170 to 200 reported deaths per day. It is beyond the scope of this analysis to address the causes of the apparent decline in May, which may have been caused by different factors such as reporting patterns, the effects of quarantine and curfews or the under-identification of COVID-19 deaths. This erratic pattern shows the importance of computing excess mortality as a more robust approach.

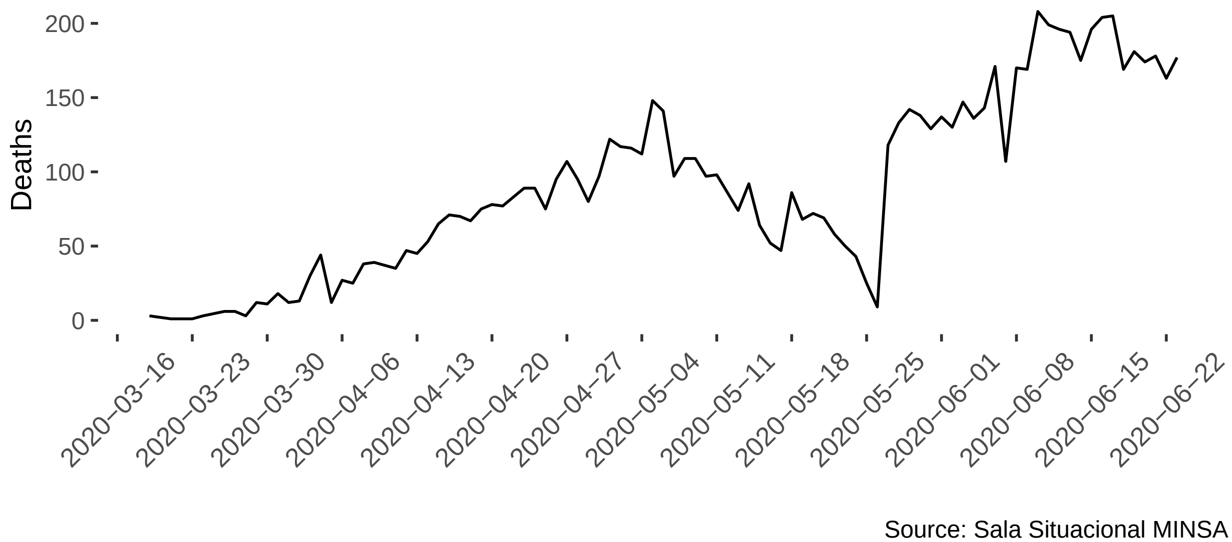


Figure 10: Number of registered daily deaths due to COVID-19 in Peru from March 16 to June 22, 2020

4.4.2 Step 2: Analysis

The death count shows great variability between regions. Table 5 shows that the highest numbers are in the most populated part of the country, Lima, with 5,611 deaths, followed by other coastal regions such as Piura and Lambayeque, with 845 and 799 deaths, respectively. Conversely, 6 regions report less than 30 registered deaths caused by COVID-19. Since the populations of different regions vary, it is more useful to compare the death rate per million people. Figure 11 shows how regions of Ica, Callao, Lambayeque, Lima, Madre de Dios and Tumbes report more 500 deaths per million, while an important number of less populated Andean regions shows rates 10 times fewer than the previous.

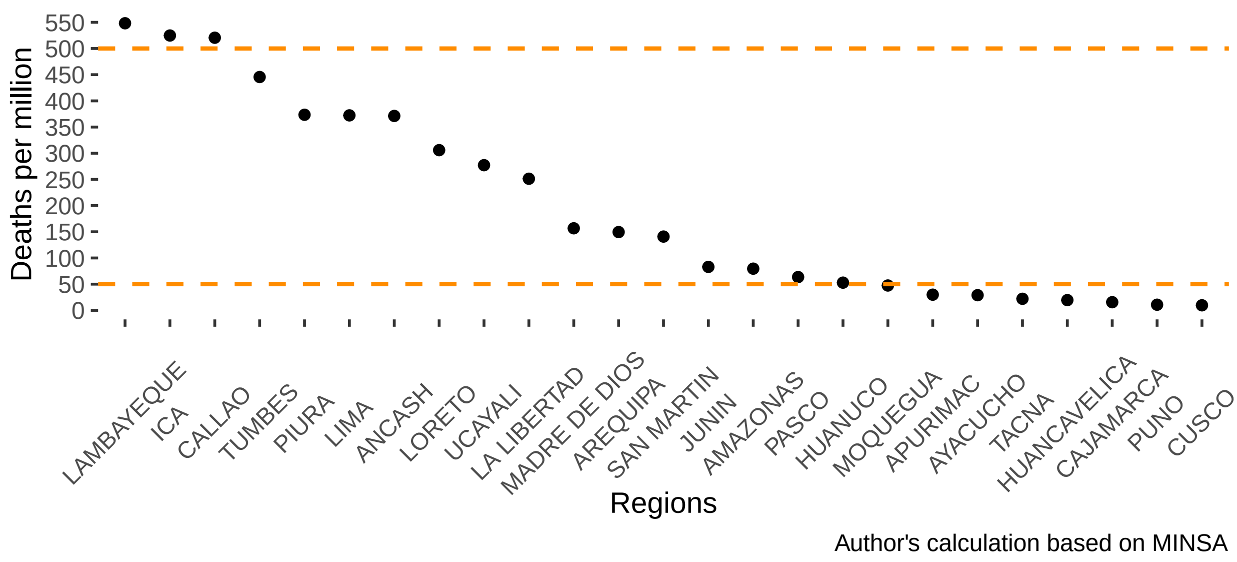


Figure 11: Deaths COVID-19 per million - 23 June 2020

Table 6 shows that over 69% of deaths reported as caused by COVID-19 were among people aged over 60. This age pattern is consistent across all regions regardless of population size or geographical location.

Table 5: Number of deaths reported by region as caused by COVID-19 until July 3rd

DEPARTAMENTO	deaths.COVID19
AMAZONAS	34
ANCASH	437
APURIMAC	14
AREQUIPA	203
AYACUCHO	21
CAJAMARCA	24
CALLAO	563
CUSCO	13
HUANCAVELICA	10
HUANUCO	47
ICA	433
JUNIN	116
LA LIBERTAD	496
LAMBAYEQUE	718
LIMA	3950
LORETO	332
MADRE DE DIOS	24
MOQUEGUA	9
PASCO	20
PIURA	715
PUNO	16
SAN MARTIN	126
TACNA	8
TUMBES	112
UCAYALI	145

Table 6: Number of deaths reported by age as caused by COVID-19 until July 3rd

range	deaths.COVID19	%total
a0.9	25	0.2911717
a10.19	17	0.1979967
a20.29	73	0.8502213
a30.39	250	2.9117167
a40.49	700	8.1528069
a50.59	1606	18.7048684
a60.69	2428	28.2785931
a70.79	2097	24.4234801
a80	1390	16.1891451

4.5 Fifht layer: Excess mortality

4.5.1 Estimating excess mortality

Excess mortality is a subset of the total count of deaths over a certain period of time. Figure 12 shows each component of our analytical framework.

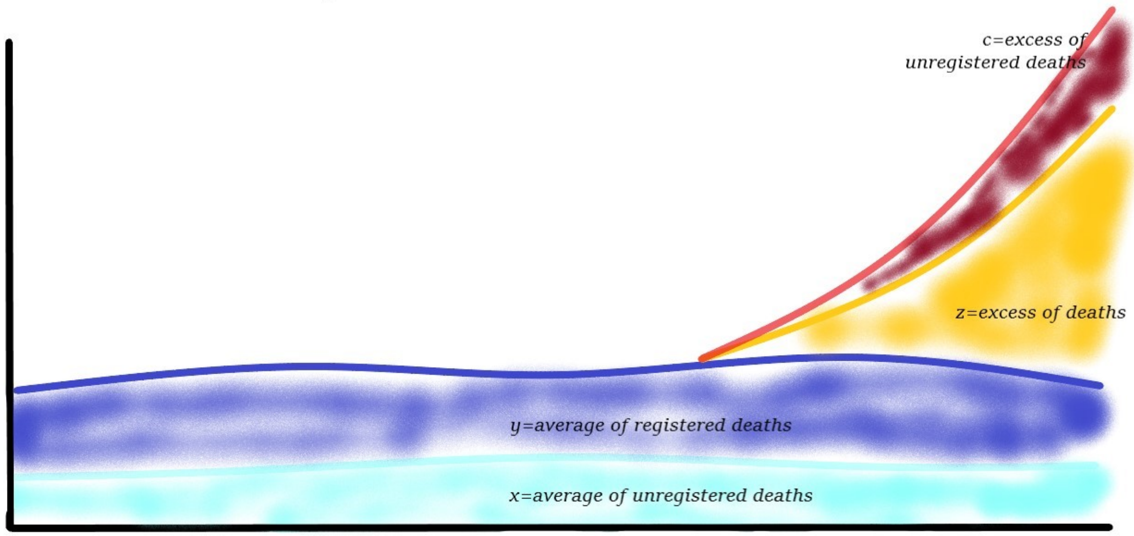


Figure 12: Understanding excess of mortality

We compute excess deaths from an unobserved counterfactual scenario, which projects average weekly deaths from previous years to the present. This requires assessing the potential existence of differences between 2020 and previous years. In some cases there is visual evidence of a pattern change. In other cases, we need to implement more sophisticated tools to detect possible differences over time.

To do this, we fit a Bayesian state-space time series model (Scott and Varian 2014) using data prior to the pre-intervention and developing a conterfactual time series synthetic control (Abadie, Diamond, and Hainmueller 2010). The inferential part of the method relies on the use of Markov chain Monte Carlo (MCMC) to simulate the counterfactual posterior distribution. In this case, we compare the expected value of the observed data y_{t} in time points $\mathrm{t} = \mathrm{t}_0 + 1 + ..., \mathrm{t}$ to the average of samples from the forecasted posterior distribution $\hat{y}_t$ of each draw τ. More details about the method and the implementation packages `CausalImpact` and `BSTS` are found in (Brodersen et al. 2015; Scott and Varian 2014). Synthetic control methods have been widely used in the field (Bouttell et al. 2018), while this specific method has been already used to evaluate health interventions (Bruhn et al. 2017; Kurz et al. 2019).

In our case, we are looking to evaluate changes in the number of deaths registered since the outbreak of COVID-19. We state as our null hypothesis, H_0 the absence of a significant change in the number of registered deaths during 2020 at the regional level, as:

$$\begin{cases} H_0 : \phi_t^\tau = y_t - \hat{y}_t^\tau = 0 \\ H_a : \phi_t^\tau = y_t - \hat{y}_t^\tau \neq 0. \end{cases} \tag{10}$$

Now we turn to the computation of excess mortality for areas where there is robust evidence of excess deaths. Our approach decomposes the estimation of total excess of deaths $\mathrm{Excess.d_T}$ into two dimensions: one related to the registered death count and other related to the average and the additional non-registered deaths. The decomposition is:

$$\widehat{\mathrm{Excess}}.\mathrm{d_T} = \widehat{\mathrm{Excess}}.\mathrm{d_{Reg}} + \widehat{\mu}_{\mathrm{Deaths_{Not.reg}}} \tag{11}$$

The computation of the first term of Equation (11), the excess of registered mortality, $\widehat{\mathrm{Excess}}.\mathrm{d_{Reg}}$, is based only on data provided by SINADEF. We compare the weekly average of deaths from 2018 and 2019 to the number of weekly deaths in 2020. We compare values from mid-March, when the first COVID-19 death case was registered, until the current time. This can be formally notated as:

$$\widehat{\text{Excess.d}}_{\text{Reg}} = \frac{1}{n}\sum_{week=12}^{n} \text{Reg.deaths}_{2020} - 1/n \sum_{week=12}^{n} \text{Reg.deaths}_{2017-2019} \tag{12}$$

The second term consists in the estimation of the recent historic average of not registered deaths in the country, $\widehat{\mu}_{\text{Deaths}_{\text{Not.reg}}}$. T This relies on the assumption that under-registration in recent years is a reasonable counterfactual to estimate missing registration in 2020. We explore visual patterns and rates to evaluate the robustness of the assumption. We follow a deterministic approach and assume under-registered COVID-19 deaths are also captured in this term. The term is computed as the product of the registration completeness ratio, $\widehat{\text{Comp.reg}}$ and the average of registered deaths on 2018-2019, $\widehat{\text{Comp.reg}}$, as follows:

$$\widehat{\mu}_{\text{Deaths}_{\text{Not.reg}}} = (1 - \widehat{\text{Comp.reg}}) * \widehat{\text{Excess.d}}_{\text{Reg}} \tag{13}$$

We now address three cases of atypical data behaviour.

First, some regions of Peru may not be significantly affected by COVID-19. This case corresponds to the non rejection of H_0 based on the hypothesis tests of (10). A second case occurs when there is no solid evidence suggesting under-registration of deaths over time in some regions or age groups, which is given by $\widehat{\text{Excess.d}}_{\text{Not.reg}} \leq 0$. This occurs usually in areas or sub-categories with very small populations. A third possibility is the scenario where less deaths occur due to a positive but unintended effect of quarantine or curfews (such as fewer road accidents or violent crimes). In this case, certain groups are not affected by an excess of deaths, and then $\widehat{\text{Excess.d}}_{\text{Reg}} \leq 0$. In these cases, we compute excess deaths without adding the terms from Equation (13) referring to unregistered figures (as they would provide add negative values to the sum, biasing the results). To address uncertainty, we estimate the total excess deaths only computing a 95% CI based on the standard error of the parameter, where n represents the number of observations at the regional level.

In the other cases, we compute the overall magnitude of mortality adding both terms. Equation (14) summarises the estimation of $\widehat{\text{Excess.d}}_{\text{T}}$ based on INEI data, where the 95% CI are provided by prior estimations of lower and upper bounds from Equations (4) and (7), conditionally to (10) and $\widehat{\text{Excess.d}}_{\text{Reg}} \leq 0 \vee \widehat{\text{Excess.d}}_{\text{Reg}} \geq 0$. as follows:

$$\begin{cases} \widehat{\text{Excess.d}}_{\text{T}} = \widehat{\text{Excess.d}}_{\text{Reg}} \pm 1.96\sqrt{(\hat{\sigma}_{\frac{\text{Excess.d}_{\text{Reg}}}{\text{n}-1}})}, & \text{if } \phi(\tau) = 0 \vee \widehat{\text{Excess.d}}_{\text{Reg}} \leq 0, \\ \widehat{\text{Excess.d}}_{\text{T}_{\max}} = \widehat{\text{Excess.d}}_{\text{Reg}} + \Big(\widehat{\mu}_{\text{Deaths}_{\text{Not.reg}}} + 1.96\sqrt{\widehat{\text{MR}} * \widehat{\text{pop}}\sqrt{(\frac{\hat{\sigma}_{\text{MR}}}{\widehat{\text{MR}}})^2 + (\frac{\hat{\sigma}_{\text{pop}}}{\widehat{\text{pop}}})^2}}\Big) & \text{and} \\ \widehat{\text{Excess.d}}_{\text{T}_{\min}} = \widehat{\text{Excess.d}}_{\text{Reg}} + \Big(\widehat{\mu}_{\text{Deaths}_{\text{Not.reg}}} - 1.96\sqrt{\widehat{\text{MR}} * \widehat{\text{pop}}\sqrt{(\frac{\hat{\sigma}_{\text{MR}}}{\widehat{\text{MR}}})^2 + (\frac{\hat{\sigma}_{\text{pop}}}{\widehat{\text{pop}}})^2}}\Big), & \\ \text{if } \phi(\tau) \neq 0 \wedge \widehat{\text{Excess.d}}_{\text{Reg}} \geq 0. & \end{cases} \tag{14}$$

In case of GBD, total excess of mortality $\widehat{\text{Excess.d}}_{\text{T}}$ and the 95% CI are estimated from Equations (5) and (11), conditionally to (10) as follows:

$$
\begin{cases}
\widehat{\text{Excess.d}}_{\text{T}} = \widehat{\text{Excess.d}}_{\text{Reg}} \pm 1.96\sqrt{\hat{\sigma}_{\text{Excess.d}_{\text{Reg}}}}, & \text{if } \phi(\tau) = 0 \vee \widehat{\text{Excess.d}}_{\text{Reg}} \leq 0, \\
\widehat{\text{Excess.d}}_{\text{T}} = \widehat{\text{Excess.d}}_{\text{Reg}} + \widehat{\mu}_{\text{Deaths}_{\text{Not.reg}}} + \widehat{\text{Excess.d}}_{\text{Not.reg}} & \text{and} \\
\widehat{\text{Excess.d}}_{\text{T}_{\text{max}}} = \widehat{\text{Excess.d}}_{\text{Reg}} + \left(\widehat{\mu}_{\text{Deaths}_{\text{Not.reg}}} + 1.96\sqrt{\hat{\sigma}_{\text{Excess.d}_{\text{Reg}}}}\right) & \text{and} \\
\widehat{\text{Excess.d}}_{\text{T}_{\text{min}}} = \widehat{\text{Excess.d}}_{\text{Reg}} + \left(\widehat{\mu}_{\text{Deaths}_{\text{Not.reg}}} - 1.96\sqrt{\hat{\sigma}_{\text{Excess.d}_{\text{Reg}}}}\right), & \text{if } \phi(\tau) \neq 0 \wedge \widehat{\text{Excess.d}}_{\text{Reg}} \geq 0.
\end{cases}
\tag{15}
$$

Our analysis makes two assumptions. First, it analysis assumes that comparison between years is not invalidated by specific time-bound mortality events such as additional disease outbreaks or wars. A second assumption is that registration patterns remain stable and any changes over time are due to death rates. Both assumptions can be assessed by visualising the data and analysing changing trends. More complex analysis such as change-point detection, computing growth rates or advanced time series analysis could also be valuable, although the absence of stationary data cannot be easily solved without in-depth knowledge of real causes. We use SINADEF data for 2018 to 2020 but not for 2017. First, as a new system, the magnitude of data collected is not stable in comparison to other years. In case of Lambayeque, after visually and analytically assessing a significant change in trend in 2020, we only use data from 2019 and 2020, as the inclusion of the prior year on the average distorts results for some subgroups, resulting in negative excess mortality.

4.5.2 Step 2: Results

We plot the time series of SINADEF by region to visually assess if our assumptions hold. Figure 13 shows almost all regions present a stable trend over time, with negative peaks at the end of each year (around weeks 53, 106, 154) due to interrupted reporting during national holidays. A few regions such as Lima, Cusco and Arequipa show a slight increase from 2017, due to their large populations compared to other regions, which is expected to create a clearer learning curve in terms of adherence to the new system. By 2020, most regions show significant increases in deaths. The vertical orange line shows the week for the first confirmed COVID-19 death. Four regions provide no evidence of COVID-19 deaths: Ayacucho, Cusco, Puno and Tacna. Lambayeque shows a very erratic trend through 2019, which suddenly changes at the beginning of the pandemic in mid-March.

4.5.2.1 Weekly deaths by region registered in SINADEF To assess whether there has been a change of patterns, we perform a Bayesian Times series univariate analysis on weekly averages from week 2 (avoiding the seasonal under-reporting of week 1) to week 28 of 2020. We set week 16 as the cut-off point, which is three weeks later than the first registered death by COVID-19 in the country. This relies on observed data that the spread of COVID-19 across regions was uneven. We draw MCMC 5,000 samples to increase the accuracy of inference and set a Bayesian one-sided tail-area probability p <. 05 to reject H_0. Table 7 shows only Puno present p.>.05 and it will be excluded from the estimates of excess of mortality (as they may provide positive values that bias the final estimates). This case will receive a different treatment, as stated in Equations (14) and (15).

With reference to age, Figure 14 similar patterns across all groups, following the population-wide results of stable trends with negative yearly peaks. The effect of the pandemics is clear across older cohorts, while the opposite applies for younger ages. In some groups, such as people aged 20-39, there is a delayed negative effect of the pandemic. The last plot of the series reflects cases without an age identification and thus represent white noise (NA). They have a very small y-scale and they do not affect the overall trends.

We decompose regional data into different components: seasonal patterns, trend and residual components using locally estimated scatterplot smoothing (LOESS). This enables us to identify specific seasonal effects (Robert B Cleveland et al. 1990). We set a periodic window that assumes no major changes in patterns over the last three years and apply robust smoothing to deal with possible data outliers. Figure 15 shows

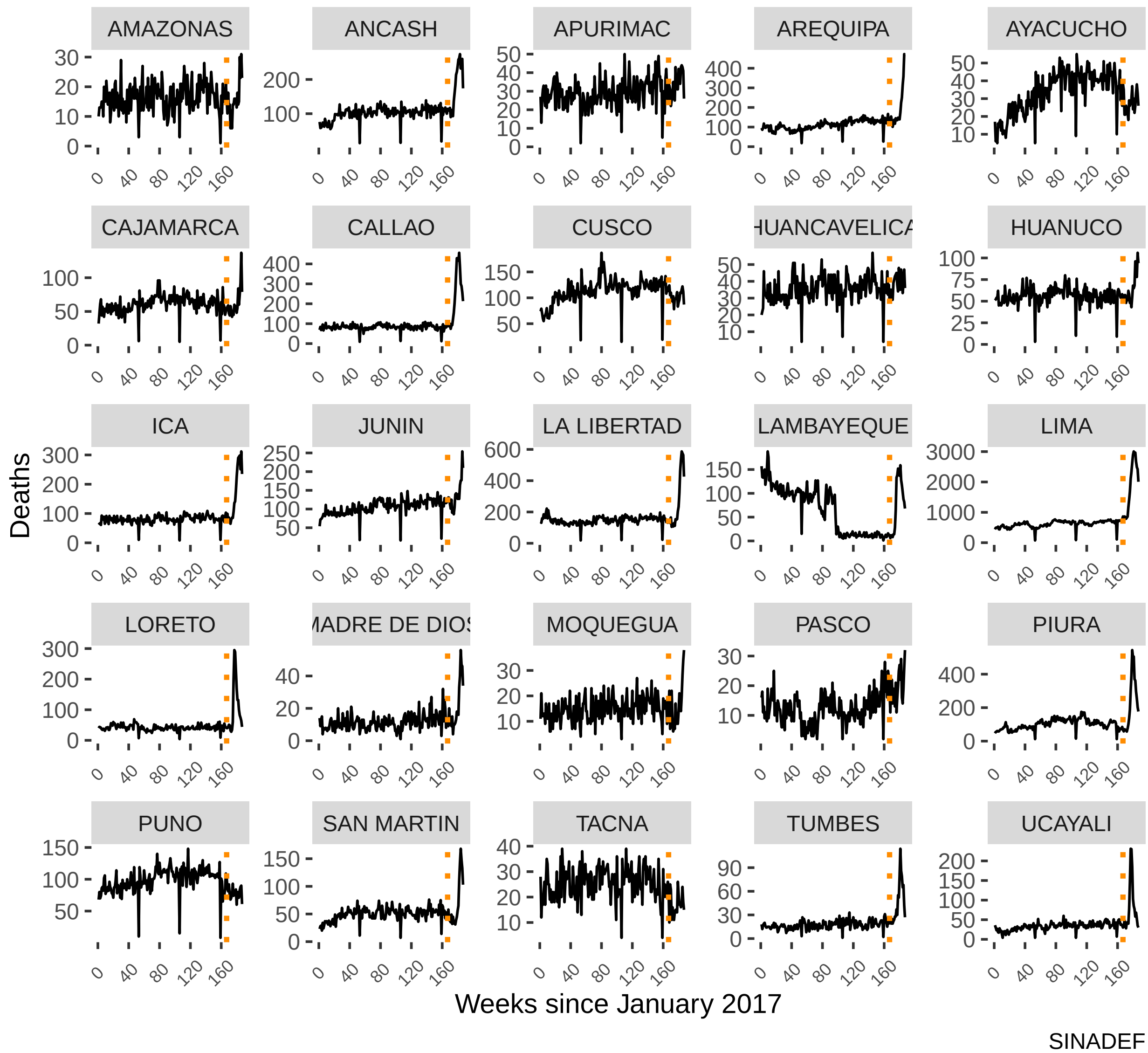


Figure 13: Weekly deaths by region registered in SINADEF

Table 7: Average of predicted deaths under Bayesian Time Series Analysis counterfactual estimation and significance tests

Departamento	Pred	Pred.lower	Pred.upper	Pred.sd	p
AMAZONAS	7.1015316	2.260177	11.951297	2.465145	0.0002008
ANCASH	99.8941089	88.870161	111.121816	5.704213	0.0002002
APURIMAC	28.4900685	20.384792	36.804274	4.195315	0.0232932
AREQUIPA	112.6800115	90.211598	135.371305	11.578070	0.0002000
AYACUCHO	10.1242816	0.547666	19.934549	4.963521	0.0002000
CAJAMARCA	41.1034224	32.533619	49.754380	4.371634	0.0002000
CALLAO	103.9850479	90.541118	118.015579	7.105448	0.0002000
CUSCO	70.8261649	58.890741	83.409814	6.317878	0.0002008
HUANCAVELICA	34.3933787	26.140355	42.664821	4.233648	0.0438000
HUANUCO	46.2989879	38.514061	54.144275	3.956012	0.0002006
ICA	68.9795616	56.825611	81.024259	6.190525	0.0002000
JUNIN	105.0365438	92.082474	118.070088	6.626338	0.0002002
LA LIBERTAD	86.9884047	68.792908	105.796931	9.429550	0.0002009
LAMBAYEQUE	17.9117845	10.644342	25.249844	3.718587	0.0002002
LIMA	1026.5074242	941.287186	1112.218121	43.675649	0.0002002
LORETO	33.0271116	24.030397	42.455007	4.683722	0.0002000
MADRE DE DIOS	0.9805459	-4.673200	6.671235	2.892826	0.0002009
MOQUEGUA	7.0372713	2.451331	11.703004	2.362294	0.0002000
PASCO	10.9370637	5.343668	16.654026	2.879414	0.0008024
PIURA	66.9194108	54.110030	79.823199	6.495689	0.0002006
PUNO	75.2992563	60.155543	90.907603	7.807562	0.3608826
SAN MARTIN	27.3370412	18.008682	36.874402	4.866695	0.0002009
TACNA	8.1267233	2.514384	13.707266	2.883408	0.0006008
TUMBES	17.6133287	12.915063	22.217275	2.371243	0.0002006
UCAYALI	31.1195139	22.562714	40.063186	4.470999	0.0002002

the results of the decomposition exercise for one region, Cusco. It shows both a strong influence of long-term trends and the absence of weekly seasonality. The same pattern is found in all other regions. Model parameters are found in the Appendix.

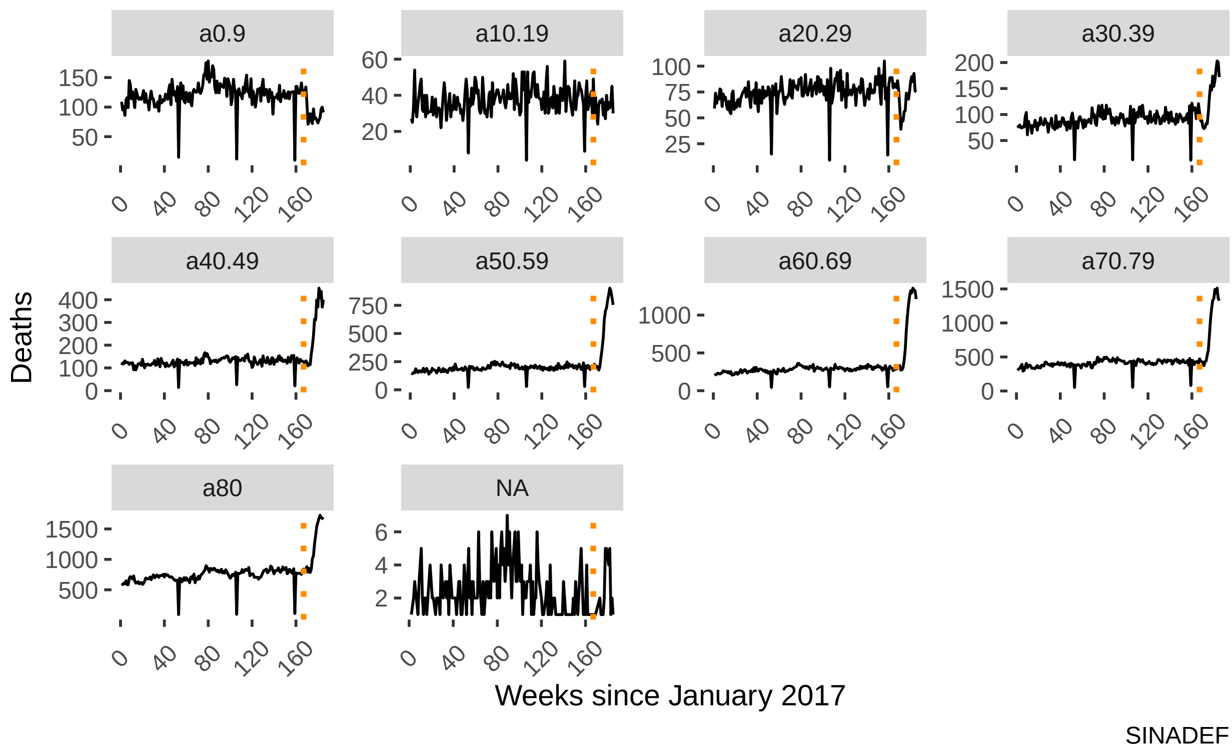


Figure 14: Weekly deaths by range of age registered in SINADEF

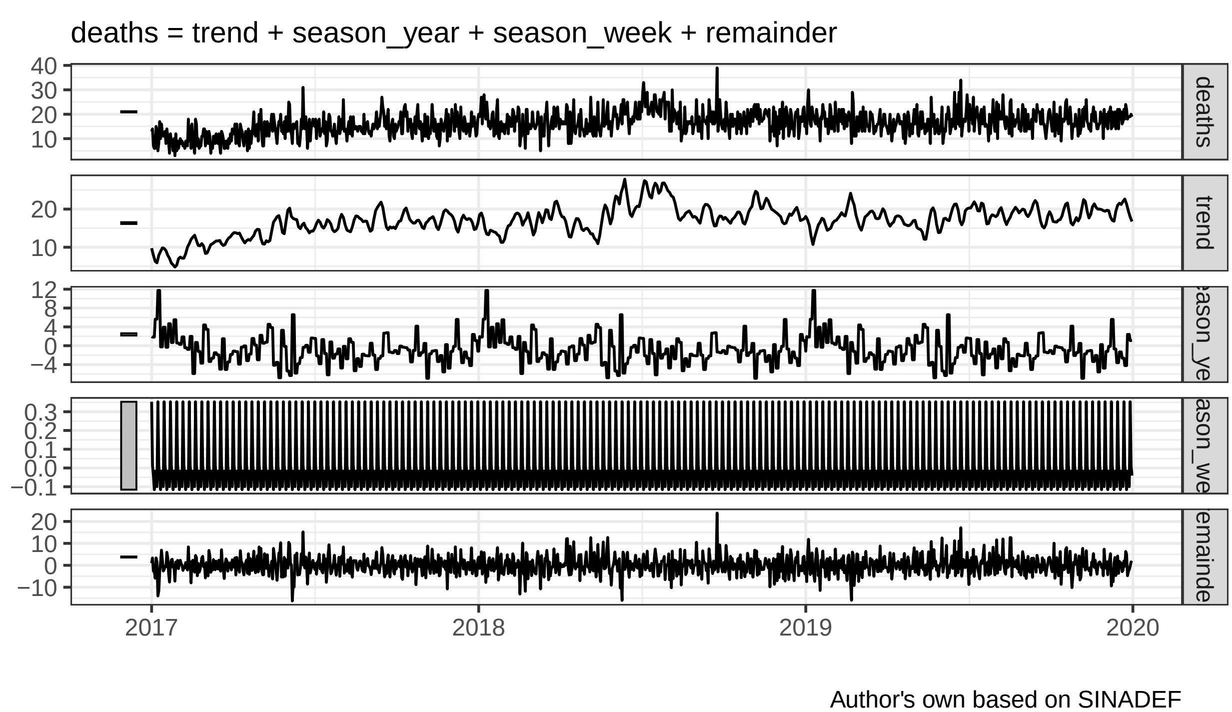


Figure 15: STL decomposition - CUSCO

The estimations of excess of registered deaths based on SINADEF raises to 41,700 deaths by week 28 (ending 12.7.20). Table 8 presents excess deaths disaggregated by regions. Lima has the highest excess with 24,397 deaths over the previous average, followed by Callao and Piura (3,083 and 2,612). Some regions (Amazonas, Ayacucho, Cajamarca, Cusco, Puno and Tacna, among others) have negative excess mortality. This will affect the final computation strategy, as mentioned above. The region of Puno reports the largest positive scenario with 468 fewer deaths than the average over previous years, followed by Cusco (338).

Table 8: Excess of deaths based on SINADEF - weeks 12 to 28 in 2020 by regions

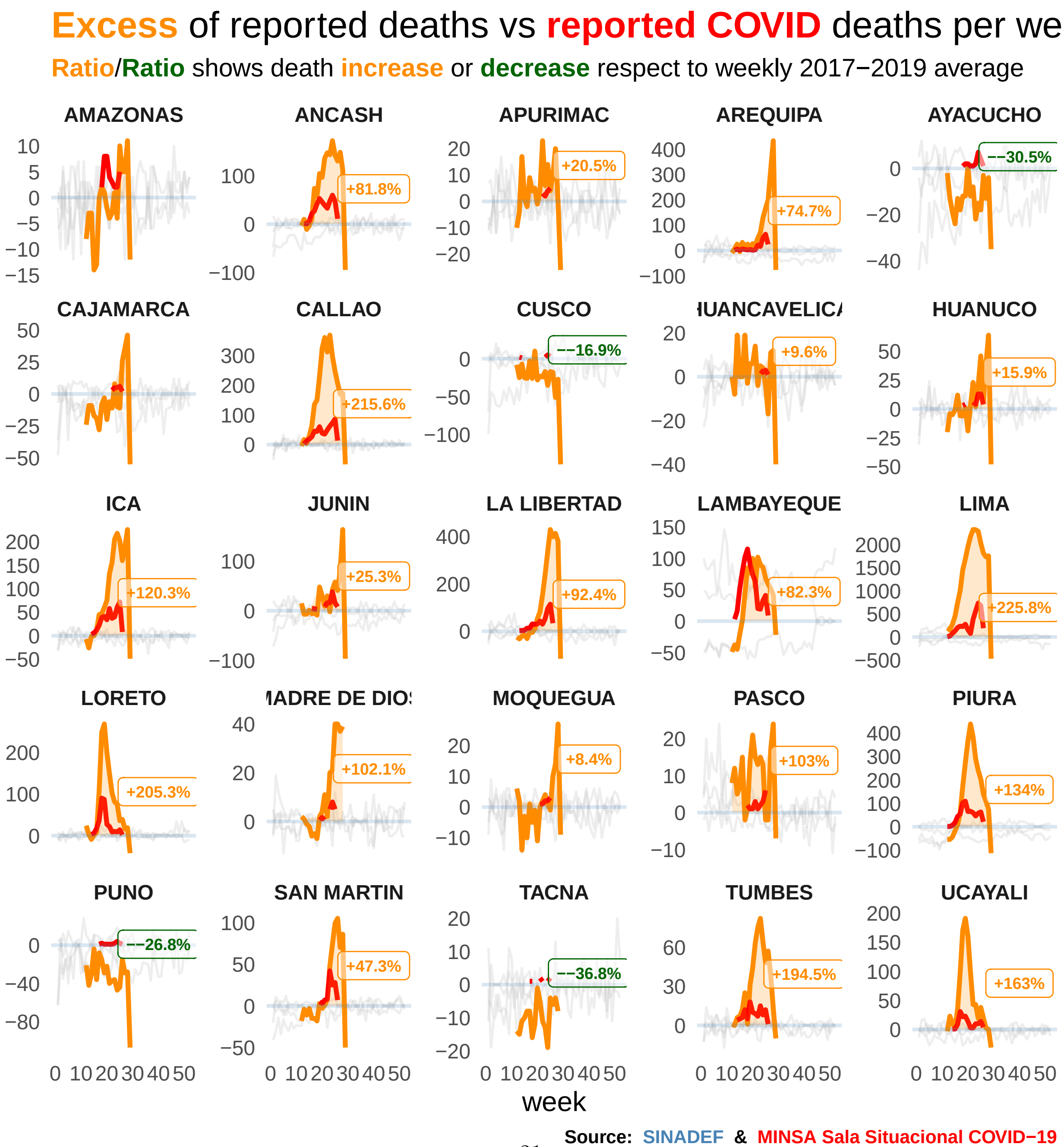
Departamento	s
AMAZONAS	-12.0
ANCASH	1484.0
APURIMAC	105.5
AREQUIPA	1549.0
AYACUCHO	-181.0
CAJAMARCA	-52.0
CALLAO	3083.0
CUSCO	-338.0
HUANCAVELICA	49.0
HUANUCO	182.5
ICA	1707.0
JUNIN	483.5
LA LIBERTAD	2371.5
LAMBAYEQUE	1305.0
LIMA	24397.0
LORETO	1353.0
MADRE DE DIOS	165.0
MOQUEGUA	40.5
PASCO	118.5
PIURA	2612.5
PUNO	-468.0
SAN MARTIN	416.0
TACNA	-133.5
TUMBES	536.0
UCAYALI	925.5

Table 9 shows excess mortality for age groups. Deaths among people aged up to 30 are below average. By contrast, excess mortality for people aged 60 or over surpasses 10,000.

Combined analysis of SINADEF and reported COVID-19 deaths shows an important gap for all regions with excess mortality. Figure **??** shows that Lima, Callao, Piura, Ucayali, Loreto, and Tumbes have over double the number of deaths compared to previous years. In Lambayeque COVID-19 deaths make up 50.6% of total mortality, while in Lima and Piura they make up 15.8% and 18.6%. This may mean that reporting of deaths not caused by COVID-19 is relatively low in Lambayeque or that Lima and Piura are reporting a lower share of deaths caused by COVID-19. Since it is unsafe to assume stable reporting for the region of Lambayeque, we compare 2020 weeks against values in 2017 in the final analysis. This gives a better estimate for the expected number of registered deaths.

Table 9: Excess of deaths based on SINADEF - weeks 12 to 28 in 2020 by regions

range	s
a0.9	-690.5
a10.19	-24.0
a20.29	-47.5
a30.39	760.0
a40.49	2913.0
a50.59	6902.0
a60.69	11328.0
a70.79	10719.0
a80	9839.5



We now present our estimate of total excess deaths, which include the non-registered terms in the equation. The estimates consider different scenarios, using INEI and GBD mortality and population projections, and fitting data using the ARIMA and RWD models.

Table 10 summarises aggregated estimates, with mean values of around 54,600 deaths for all the regressions. For example, the combination of ARIMA models gives an expected value of 54,645 95% CI (45,534 to 63,739). This clearly suggests a significant downward bias of identification of COVID-19 deaths in official sources(reporting 12,504 deaths). Estimated non-registered notifications add 13,000 to excess registered mortality (equivalent to 23.7% of non-registered deaths). This level of underestimate is slightly higher than for the GBD estimates (Dicker et al. 2018), which generate around 19.6% additional deaths. Comparing death registrations from INEI and SINADEF in 2016 and 2017 shows SINADEF registrations still lag behind previous years.

Table 10: Estimated of excess of deaths based on INEI - weeks 12 to 28

model.pop	model.mort	Total excess	Lower.CI95%	Upper.CI95%	Deaths SINADEF	Covid deaths
arima	arima	54644.80	45533.42	63738.39	78624	8586
arima	drift	54866.65	45727.64	63984.48	78624	8586
drift	arima	54712.03	45603.95	63802.61	78624	8586
drift	drift	54941.02	45805.67	64055.56	78624	8586

Table 11 shows that 76% of total excess deaths are among people aged 60 or over, with an expected value of 42,000 additional deaths across the models. For people 80 or over reported COVID-19 deaths account for the lowest share of total estimated excess deaths (15.9%), possible due to higher rates of misreporting cause of death for this age group.

Table 11: Estimated of excess of deaths based on INEI - weeks 12 to 28: age groups

range	Excess	Lower.CI	Upper.CI	SINADEF	Covid	Covid:Excess
a0.9	-683.84937	-4212.0421	2844.285	1446	25	-0.0365578
a10.19	-12.87355	-3000.6521	2974.784	583	17	-1.3205369
a20.29	-19.99634	-1345.9894	1305.855	1204	73	-3.6506677
a30.39	1003.55196	824.2866	1181.999	2421	250	0.2491152
a40.49	3802.84215	3648.9058	3954.323	5205	700	0.1840729
a50.59	9058.84391	8874.8965	9238.502	10386	1606	0.1772853
a60.69	14889.54887	14686.6907	15087.655	16409	2428	0.1630674
a70.79	13982.86757	13792.9600	14169.433	18075	2097	0.1499692
a80	12623.86828	12264.3681	12981.550	22895	1390	0.1101089

Table 12 presents estimations of excess of deaths by region. Lima shows 29,309 95% CI (25760, 32857) deaths followed by Callao, Piura, Lambayeque and La Libertad, averaging 2,500 to 3,100 deaths. Six regions yield negative values: Amazonas, Ayacucho, Cajamarca, Cusco, Puno and Tacna. However, in cases like Amazonas, Ayacucho and Tacna the upper bound of 95% are close or over 0.

Table 12 shows estimated excess deaths by region. For Lima there were 31,933 95% CI (28,077, 35,786) excess deaths, followed by Callao, Piura, Lambayeque and La Libertad, averaging 2,500 to 3,100 deaths. Four regions yield negative expected values: Ayacucho, Cusco, Puno and Tacna. In cases like Amazonas and Cajamarca the lower levels of the 95% CI are lower than 0.

Table 13 shows estimates based on GBD data, which, at around 46,000, are slightly lower than those based on INEI data. Confidence intervals are significantly larger than for the models based on INEI data, signifying a 19% under-registration rate. Estimations show same patterns than INEI models across age groups, with 75% of excess deaths occurring among people aged 60 and Lima, Piura and Callao reporting the largest shares of deaths.

Table 12: Estimated of excess of deaths based on INEI - weeks 12 to 28: regions

Departamento	Total excess	Lower.CI95%	Upper.CI95%	Deaths SINADEF	Covid deaths
AMAZONAS	5.679648	-50.47077	61.79354	279	34
ANCASH	1868.131371	1595.76289	2139.16577	3277	437
APURIMAC	155.583989	97.76301	213.35216	587	14
AREQUIPA	1613.198315	1315.77771	1909.63284	3553	203
AYACUCHO	-181.000000	-375.83894	13.83894	472	21
CAJAMARCA	8.264441	-136.97967	153.45506	1097	24
CALLAO	3352.695019	2577.55135	4125.93972	4515	563
CUSCO	-334.137683	-508.85412	-159.43866	1723	13
HUANCAVELICA	71.546043	44.87256	97.93794	687	10
HUANUCO	287.840845	224.49392	351.04312	1136	47
ICA	1501.554288	1198.92541	1801.78014	3104	433
JUNIN	654.940015	509.83833	799.69506	2403	116
LA LIBERTAD	2863.842721	2226.64696	3499.76514	4927	496
LAMBAYEQUE	2594.687652	1986.99107	3202.28350	2856	718
LIMA	31932.512306	28077.89874	35786.12689	35182	3950
LORETO	2209.635656	2082.70799	2335.58950	2032	332
MADRE DE DIOS	181.159976	124.63632	235.08053	392	24
MOQUEGUA	47.450977	40.95871	53.71032	284	9
PASCO	200.045641	179.24623	220.63487	343	20
PIURA	3740.486752	3090.87961	4389.44053	4524	715
PUNO	-468.000000	-801.95217	-134.04783	1321	16
SAN MARTIN	604.674970	530.66790	678.47160	1268	126
TACNA	-133.001306	-212.58624	-53.45341	282	8
TUMBES	594.991855	528.17644	659.39315	859	112
UCAYALI	1272.019979	1186.31090	1357.19540	1521	145

Table 13: Estimated of excess of deaths based on GBD - weeks 12 to 28

.model	Total excess	Lower.CI95%	Upper.CI95%	Deaths SINAEF	Covid deaths
arima	51188.24	40057.27	62019.83	78624	8586
drift	51713.74	40697.12	62446.05	78624	8586

Figure 16 shows estimates by region and age group. The general trend is for higher rates of excess deaths among older people, although this is not the case for some regions Ayacucho, Cajamarca, Moquegua and Tacna. The reasons for this counter-intuitive result are not explored in this paper.

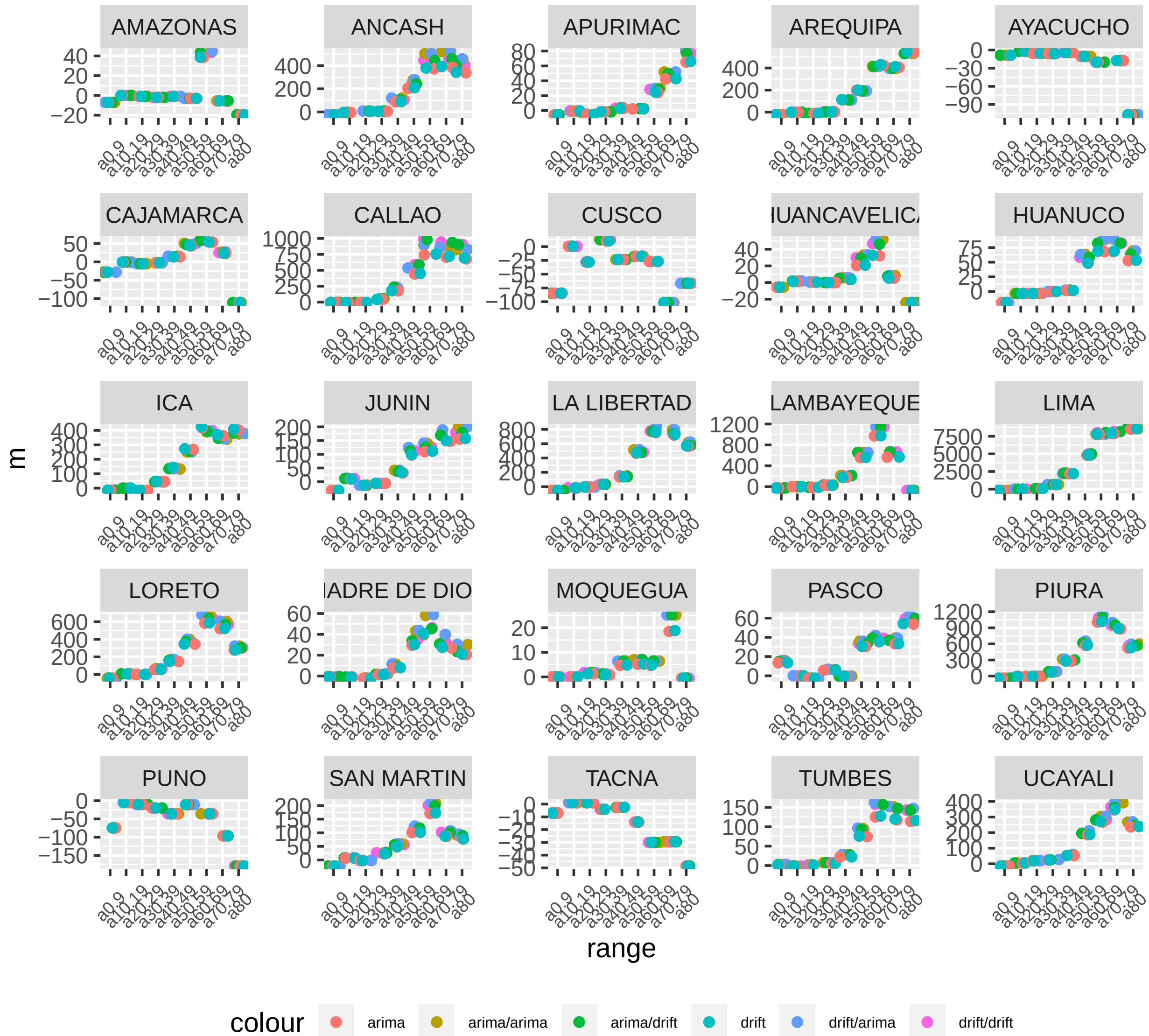


Figure 16: Total excess of deaths by regions - different models

Table 14 shows projections based on current excess deaths, assuming all other things remain equal. This shows an increase to 195,000 expected deaths, which is 21.7 % higher than INEI estimates for 2020.

Table 14: Projected number of deaths Estimated of excess of deaths based on GBD - weeks 12 to 28

model.pop	model.mort	Ratio Excess	Projected 2020
arima	arima	0.7631009	195025.3
arima	drift	0.7600154	195817.1
drift	arima	0.7621632	195265.3
drift	drift	0.7589866	196082.5

Our comparisons across results need to take into account differences in population sizes and age structures. To address this problem, we compute standardised adjusted mortality rates across regions using direct standardisation (Curtin and Klein 1995). We use population estimates by region and age group, as well as total estimated deaths. Table 15 shows adjusted mortality ratios by region until week 28 of 2020. Ucayali, Tumbes, Madre de Dios, Loreto and Callao have rates of over 4, while Puno, Amazonas, Tacna and Ayacucho show the lowest adjusted rates -less than 1.

Table 15: Age standardised mortality rates by regions 2020 - up to week 28

Departamento	std.d
AMAZONAS	0.9099395
ANCASH	3.3827087
APURIMAC	2.0884530
AREQUIPA	2.3462059
AYACUCHO	0.8513508
CAJAMARCA	1.1206732
CALLAO	4.0117474
CUSCO	1.3960288
HUANCAVELICA	2.4557700
HUANUCO	2.1443842
ICA	3.1431579
JUNIN	2.4831039
LA LIBERTAD	3.0124320
LAMBAYEQUE	3.4998962
LIMA	3.9156776
LORETO	4.2204854
MADRE DE DIOS	4.3573846
MOQUEGUA	1.4774693
PASCO	2.3053489
PIURA	3.6591048
PUNO	0.9561755
SAN MARTIN	2.7730377
TACNA	0.8711758
TUMBES	4.7830820
UCAYALI	4.8830833

5 Conclusion

This tutorial and case study provides a framework and tools to estimate the burden of death caused by the COVID-19 pandemic. This tutorial sets out an analytical path for progressively gathering and analyzing information to generate robust estimates of excess mortality. This approach can be applied in all countries, although the complexity of analysis will vary according to the availability and accuracy of data. There is strong evidence that excess COVID-19 mortality has disproportionately affected people at older ages (Dowd et al. 2020). Consequently, we include population age structure in our analysis.

The case of Peru shows that significant under-registration of deaths caused by COVID-19. Official figures correspond to only 20-22% of total excess mortality and registered deaths. Different models and databases show that Lima, Piura, Callao are the worst affected regions and people 60 and over the most affected age group. Adjusted analysis shows that excess mortality is higher in less populated regions such as Tumbes and the amazon states of Ucayali, Loreto and Madre de Dios. Some regions, especially in the Andes, do not appear to have been significantly affected by the pandemic by the end of June, in terms of mortality.

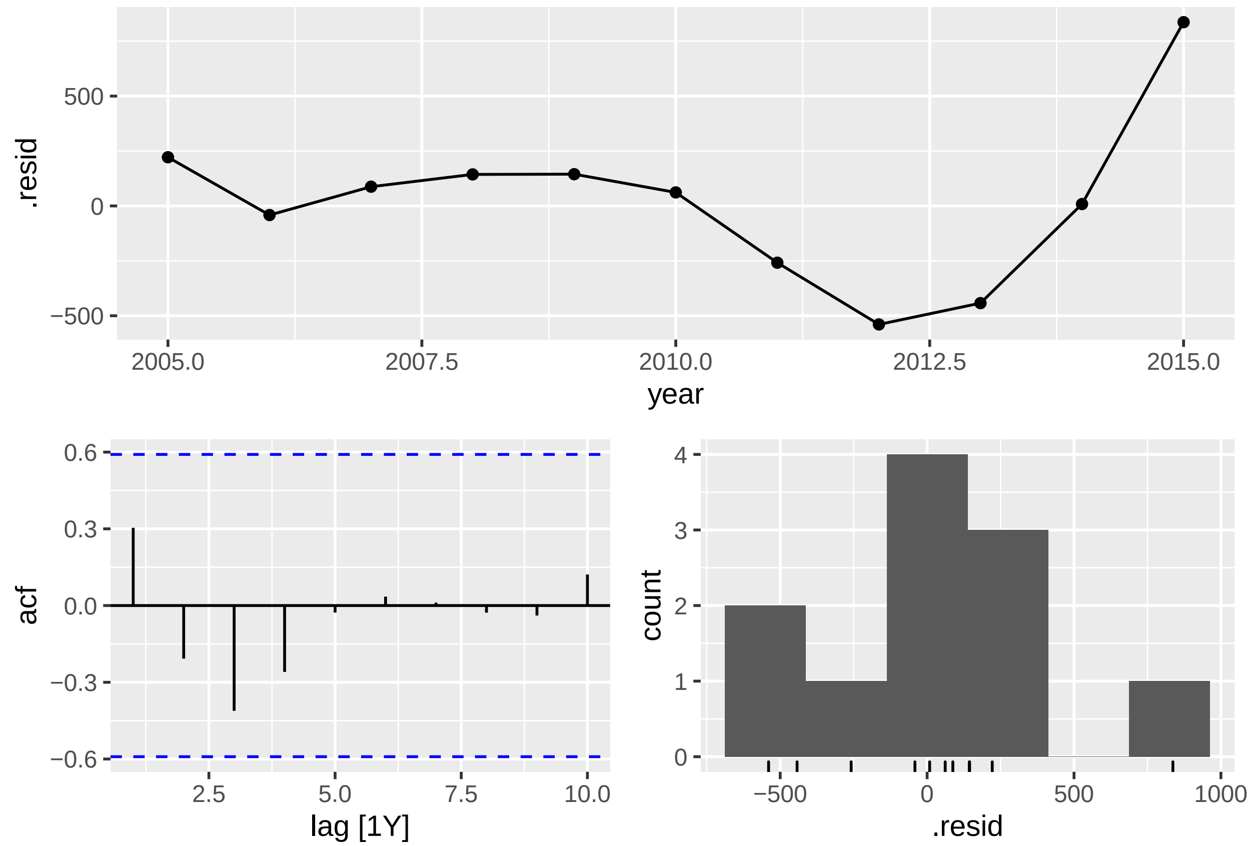
There are important limitations in our method. Estimates are based on provisional data, which are incomplete. Our estimates of registration completeness assume no variation across age groups, which may not be the case. Our time series analysis produces higher CIs when forecasting over extended periods, reducing the precision of estimates. Lags in reporting deaths may reduce the validity of our cutoffs. Finally, we present a conservative scenario, allowing for the existence of negative as well as positive excess deaths.

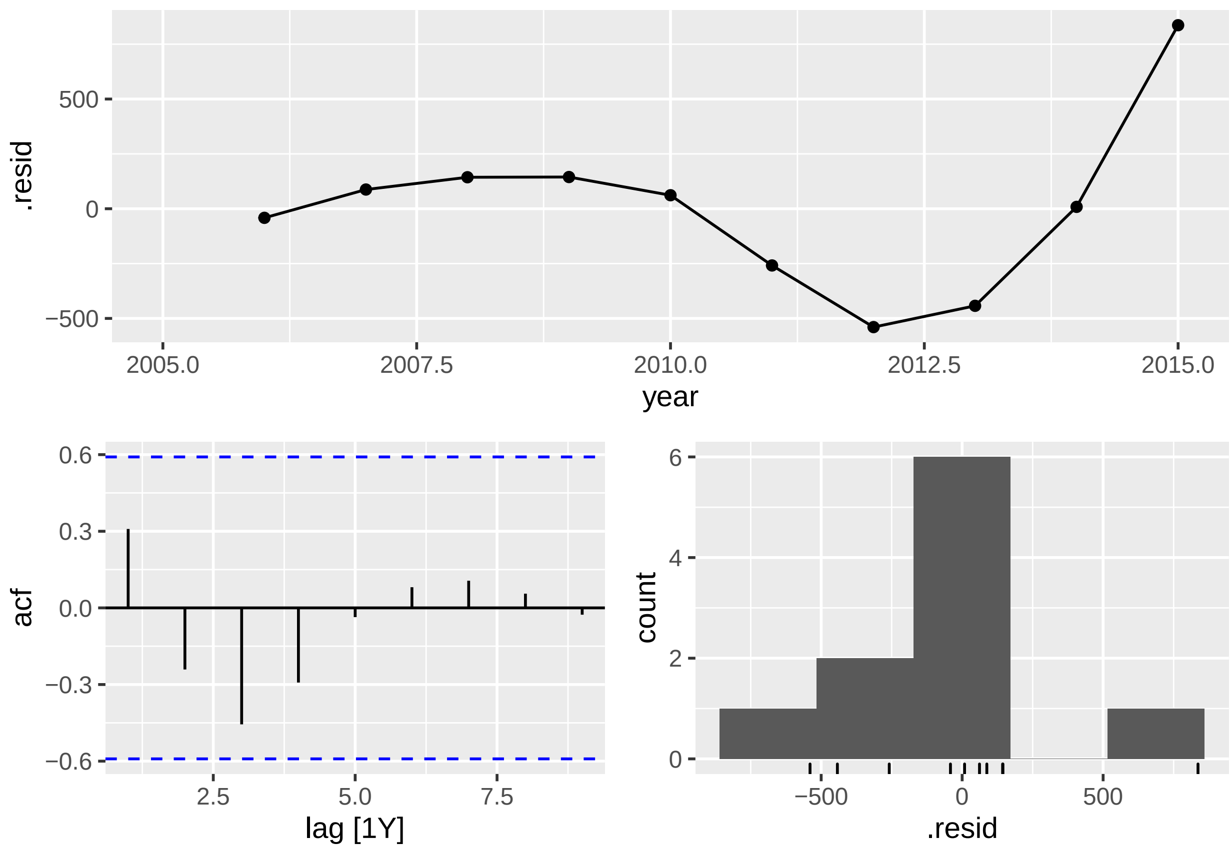
6 Appendix

6.0.1 Model fit and residuals analysis: population INEI

```
## # A tibble: 450 x 5
##    Departamento range  .model lb_stat lb_pvalue
##    <chr>        <chr>  <chr>    <dbl>     <dbl>
##  1 AMAZONAS     a0.9   arima    3.62    0.460
##  2 AMAZONAS     a0.9   drift   10.9     0.0272
##  3 AMAZONAS     a10.19 arima    0.774   0.942
##  4 AMAZONAS     a10.19 drift   14.1     0.00694
##  5 AMAZONAS     a20.29 arima    1.85    0.764
##  6 AMAZONAS     a20.29 drift   13.6     0.00860
##  7 AMAZONAS     a30.39 arima    4.64    0.327
##  8 AMAZONAS     a30.39 drift    8.73    0.0681
##  9 AMAZONAS     a40.49 arima    3.18    0.528
## 10 AMAZONAS     a40.49 drift    7.70    0.103
## 11 AMAZONAS     a50.59 arima    9.87    0.0426
## 12 AMAZONAS     a50.59 drift   10.1     0.0389
## 13 AMAZONAS     a60.69 arima    6.58    0.160
## 14 AMAZONAS     a60.69 drift   11.9     0.0183
## 15 AMAZONAS     a70.79 arima    2.26    0.689
## 16 AMAZONAS     a70.79 drift   13.5     0.00926
## 17 AMAZONAS     a80    arima    1.56    0.815
## 18 AMAZONAS     a80    drift   12.1     0.0168
## 19 ANCASH       a0.9   arima    0.507   0.973
## 20 ANCASH       a0.9   drift    2.12    0.714
## 21 ANCASH       a10.19 arima    5.01    0.286
## 22 ANCASH       a10.19 drift    6.84    0.144
## 23 ANCASH       a20.29 arima   11.6     0.0206
## 24 ANCASH       a20.29 drift    8.79    0.0667
## 25 ANCASH       a30.39 arima    3.78    0.436
## 26 ANCASH       a30.39 drift    2.97    0.563
## 27 ANCASH       a40.49 arima    5.10    0.277
## 28 ANCASH       a40.49 drift    8.75    0.0675
## 29 ANCASH       a50.59 arima   10.8     0.0284
## 30 ANCASH       a50.59 drift   11.1     0.0253
## 31 ANCASH       a60.69 arima    2.28    0.685
## 32 ANCASH       a60.69 drift   11.4     0.0220
## 33 ANCASH       a70.79 arima   11.8     0.0191
## 34 ANCASH       a70.79 drift   13.4     0.00938
## 35 ANCASH       a80    arima    4.63    0.327
## 36 ANCASH       a80    drift   11.7     0.0197
## 37 APURIMAC     a0.9   arima    8.82    0.0658
## 38 APURIMAC     a0.9   drift   11.1     0.0259
## 39 APURIMAC     a10.19 arima   16.0     0.00305
## 40 APURIMAC     a10.19 drift   14.0     0.00723
## 41 APURIMAC     a20.29 arima    2.17    0.704
## 42 APURIMAC     a20.29 drift   13.0     0.0114
## 43 APURIMAC     a30.39 arima    4.44    0.349
## 44 APURIMAC     a30.39 drift    4.16    0.385
```

```
## 45 APURIMAC     a40.49 arima    4.96    0.291
## 46 APURIMAC     a40.49 drift   13.4     0.00966
## 47 APURIMAC     a50.59 arima    2.97    0.562
## 48 APURIMAC     a50.59 drift    2.21    0.696
## 49 APURIMAC     a60.69 arima    4.29    0.368
## 50 APURIMAC     a60.69 drift    8.28    0.0818
## # ... with 400 more rows
```

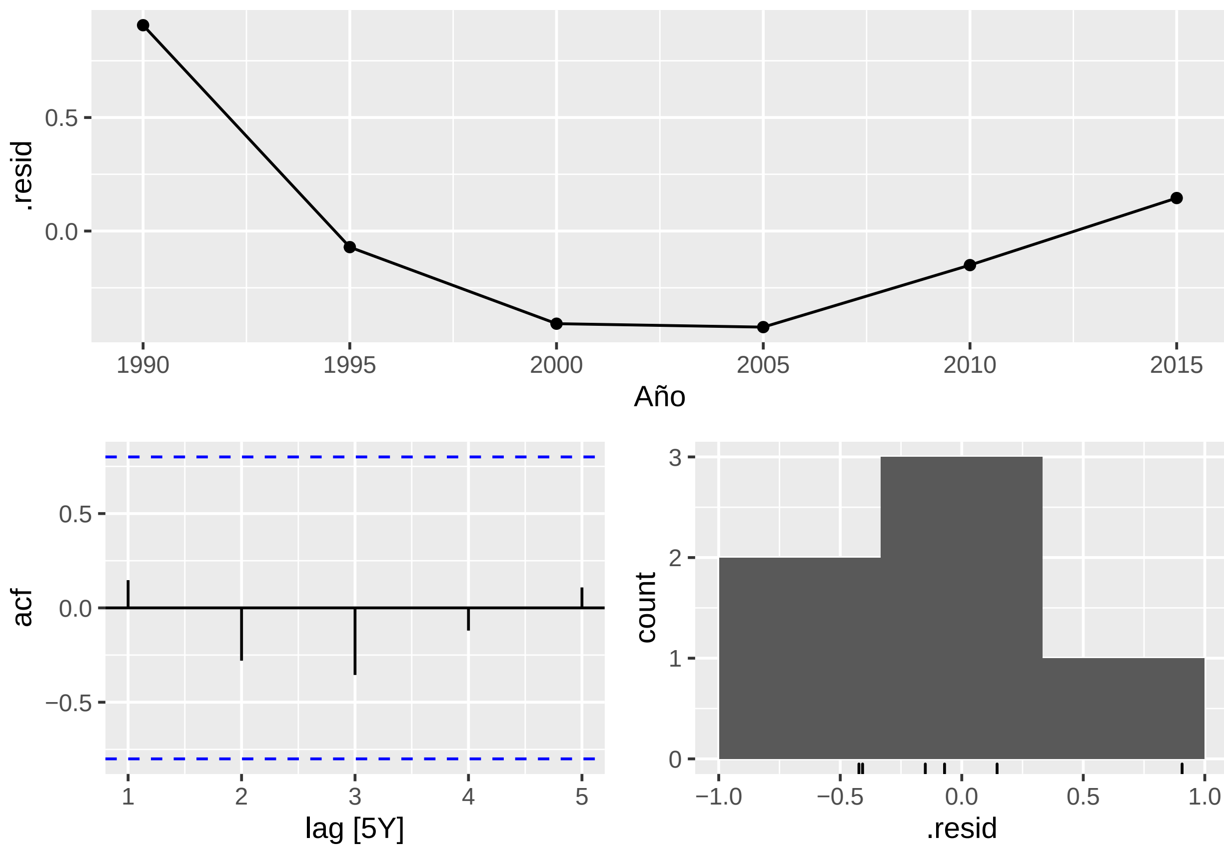




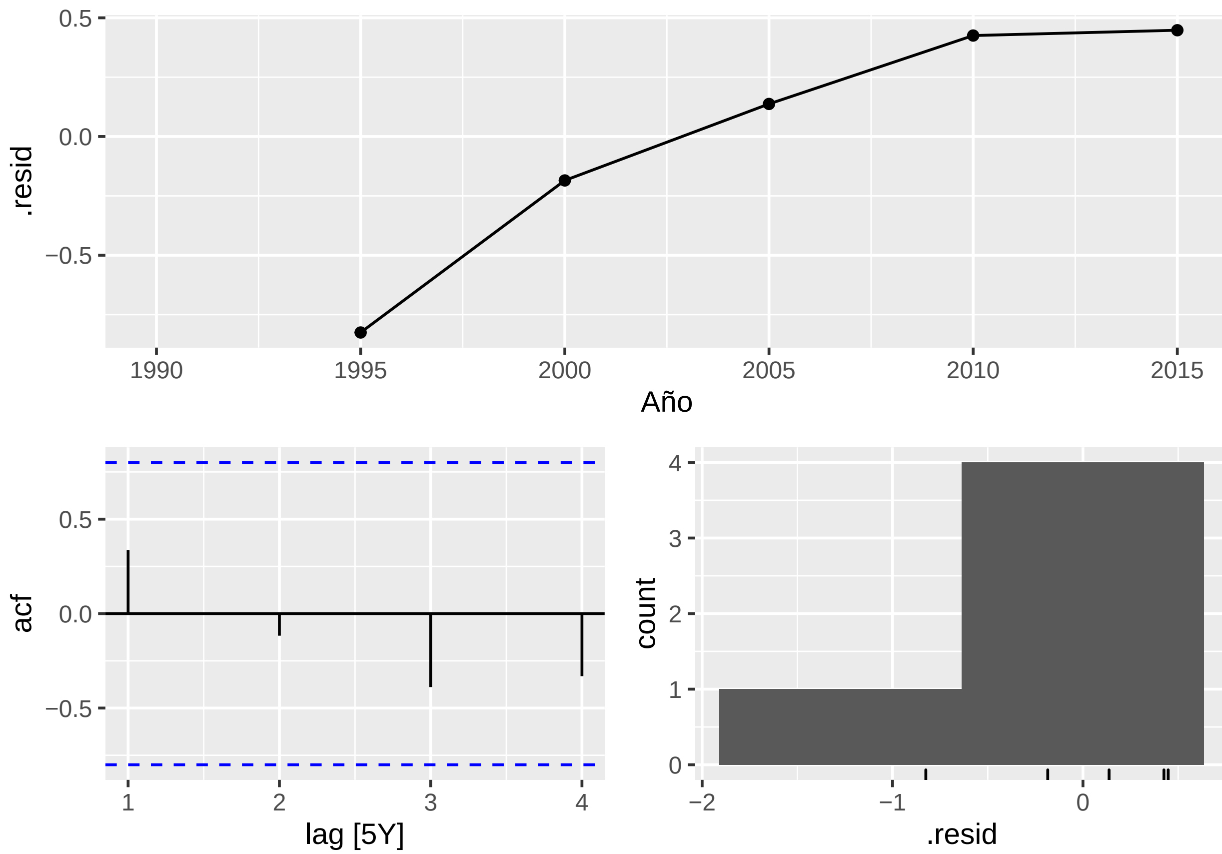
6.0.2 Model fit and residuals analysis: mortality

```
## # A tibble: 50 x 4
##    Departamento  .model lb_stat lb_pvalue
##    <chr>         <chr>    <dbl>     <dbl>
##  1 AMAZONAS      arima     3.56    0.468
##  2 AMAZONAS      drift     7.68    0.104
##  3 ANCASH        arima     4.05    0.400
##  4 ANCASH        drift     7.75    0.101
##  5 APURIMAC      arima     3.43    0.488
##  6 APURIMAC      drift     8.76    0.0673
##  7 AREQUIPA      arima     3.52    0.475
##  8 AREQUIPA      drift     7.64    0.106
##  9 AYACUCHO      arima     8.11    0.0875
## 10 AYACUCHO      drift     7.40    0.116
## 11 CAJAMARCA     arima     5.48    0.241
## 12 CAJAMARCA     drift     9.53    0.0491
## 13 CALLAO        arima     3.43    0.489
## 14 CALLAO        drift     6.08    0.193
## 15 CUSCO         arima     7.33    0.120
## 16 CUSCO         drift     3.69    0.450
## 17 HUANCAVELICA  arima     1.28    0.864
## 18 HUANCAVELICA  drift     8.25    0.0829
## 19 HUANUCO       arima     5.29    0.259
## 20 HUANUCO       drift     9.22    0.0558
## 21 ICA           arima     6.41    0.170
```

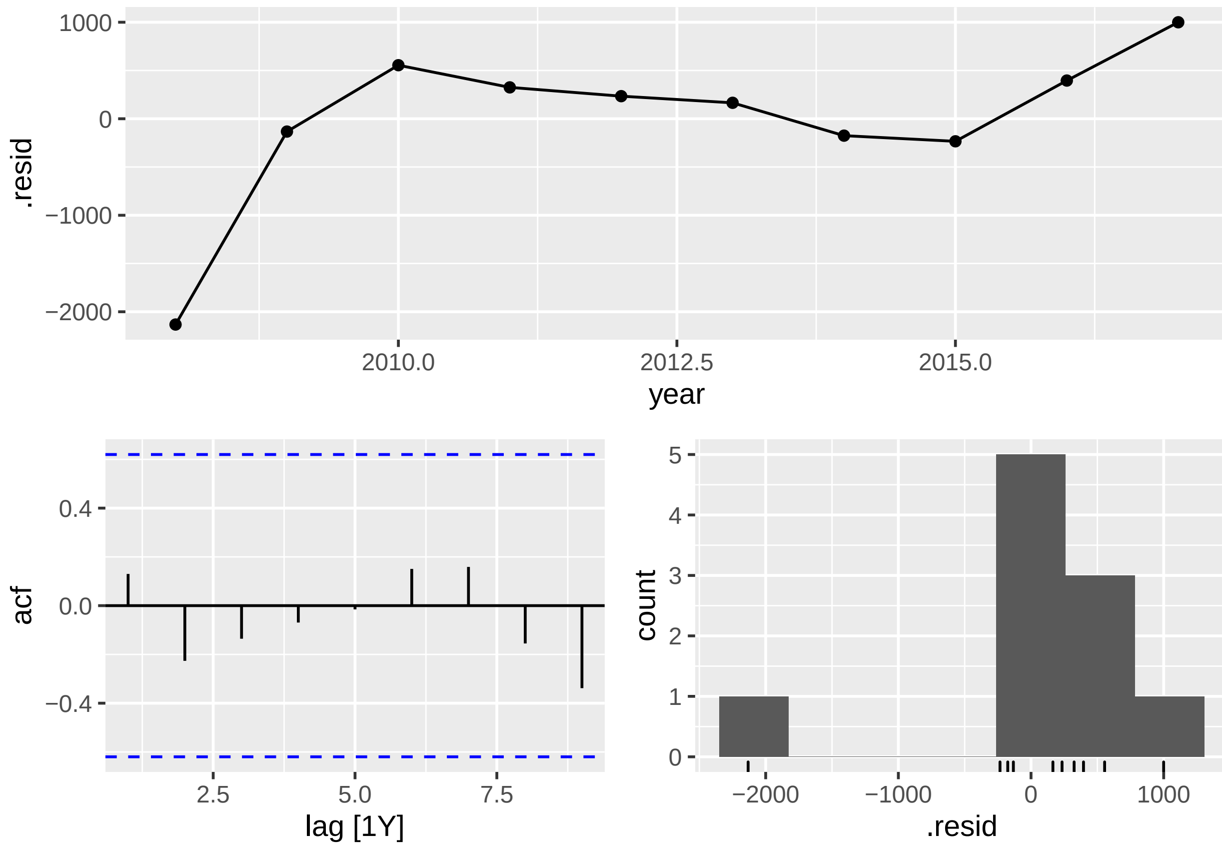
```
## 22 ICA           drift     8.05    0.0899
## 23 JUNIN         arima     5.33    0.255
## 24 JUNIN         drift     9.62    0.0472
## 25 LA LIBERTAD   arima     3.36    0.499
## 26 LA LIBERTAD   drift     7.71    0.103
## 27 LAMBAYEQUE    arima     3.90    0.419
## 28 LAMBAYEQUE    drift     8.62    0.0712
## 29 LIMA          arima     6.91    0.140
## 30 LIMA          drift     9.76    0.0447
## 31 LORETO        arima     4.27    0.371
## 32 LORETO        drift     7.60    0.108
## 33 MADRE DE DIOS arima     2.55    0.636
## 34 MADRE DE DIOS drift     5.35    0.253
## 35 MOQUEGUA      arima     5.42    0.247
## 36 MOQUEGUA      drift     6.31    0.177
## 37 PASCO         arima     5.65    0.227
## 38 PASCO         drift     9.21    0.0561
## 39 PIURA         arima     4.79    0.310
## 40 PIURA         drift     8.76    0.0673
## 41 PUNO          arima     7.09    0.131
## 42 PUNO          drift     5.88    0.208
## 43 SAN MARTIN    arima     4.39    0.356
## 44 SAN MARTIN    drift     8.99    0.0614
## 45 TACNA         arima     3.75    0.441
## 46 TACNA         drift     8.50    0.0749
## 47 TUMBES        arima     5.21    0.267
## 48 TUMBES        drift     7.96    0.0932
## 49 UCAYALI       arima     3.44    0.488
## 50 UCAYALI       drift     9.45    0.0508
```

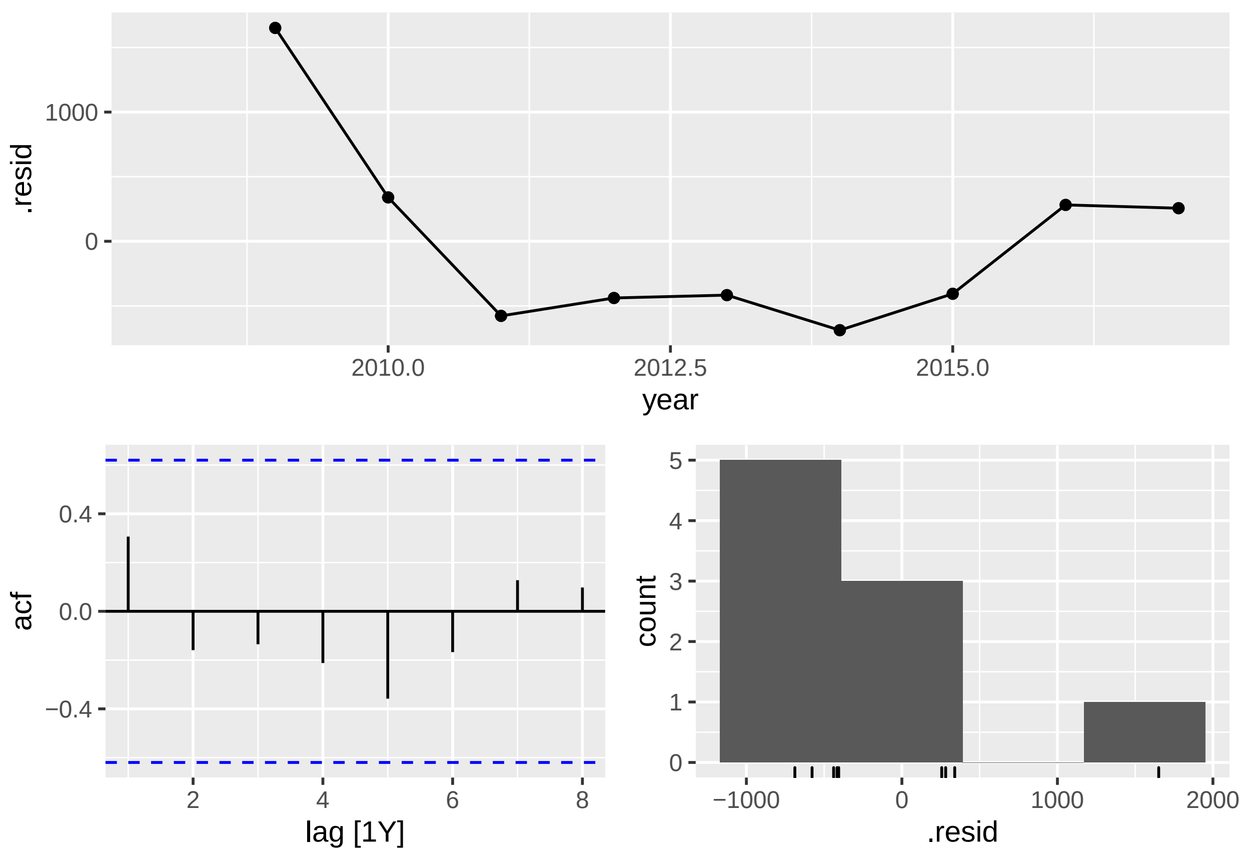


.resid
0.5
0.0
1990
1995
2000
2005
2010
2015
Año
acf
0.5
0.0
−0.5
1
2
3
4
5
lag [5Y]
count
3
2
1
0
−1.0
−0.5
0.0
0.5
1.0
.resid


.resid
0.5
0.0
−0.5
1990
1995
2000
2005
2010
2015
Año
acf
0.5
0.0
−0.5
1
2
3
4
lag [5Y]
count
4
3
2
1
0
−2
−1
0
.resid

6.0.3 Model fit expected mortality GBD





range	.model	lb_stat	lb_pvalue
a0.9	arima	1.2330083	0.9416717
a0.9	drift	0.9116141	0.9693751
a10.19	arima	7.6403049	0.1772051
a10.19	drift	3.4794053	0.6265061
a20.29	arima	4.1651456	0.5258925
a20.29	drift	1.9638133	0.8541292
a30.39	arima	4.4000599	0.4933654
a30.39	drift	2.9813480	0.7028616
a40.49	arima	7.5793945	0.1809903
a40.49	drift	2.3383179	0.8006230
a50.59	arima	1.9039871	0.8622644
a50.59	drift	3.6734880	0.5973142
a60.69	arima	0.8508486	0.9736773
a60.69	drift	5.5542843	0.3520301
a70.79	arima	1.4095876	0.9232616
a70.79	drift	5.8866397	0.3174061
a80	arima	9.2488066	0.0995422
a80	drift	9.0136099	0.1085225

6.0.4 Model fit STL decomposition SINADEF

6.0.5 Excess total INEI regions

References

Abadie, Alberto, Alexis Diamond, and And Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program." *Journal of the American Statistical Association* 105 (490): 493–505. https://doi.org/10.1198/jasa.2009.ap08746.

Adair, Tim, and Alan D. Lopez. 2018. "Estimating the completeness of death registration: An empirical method." *PLoS ONE* 13 (5): e0197047.

Bouttell, Janet, Peter Craig, James Lewsey, Mark Robinson, and Frank Popham. 2018. "Synthetic control methodology as a tool for evaluating population-level health interventions." *Journal of Epidemiology and Community Health* 72 (8): 673–78. https://doi.org/10.1136/jech-2017-210106.

Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. 2015. "Inferring causal impact using bayesian structural time-series models." *Annals of Applied Statistics* 9 (1): 247–74. https://doi.org/10.1214/14-AOAS788.

Bruhn, Christian A. W., Stephen Hetterich, Cynthia Schuck-Paim, Esra Kürüm, Robert J. Taylor, Roger Lustig, Eugene D. Shapiro, Joshua L. Warren, Lone Simonsen, and Daniel M. Weinberger. 2017. "Estimating the population-level impact of vaccines using synthetic controls." *Proceedings of the National Academy of Sciences of the United States of America* 114 (7): 1524–9. https://doi.org/10.1073/pnas.1612833114.

Curtin, Lester R, and Richard J Klein. 1995. *Direct standardization (age-adjusted death rates)*. 6. US Department of Health; Human Services, Public Health Service, Centers ....

Dicker, Daniel, Grant Nguyen, Degu Abate, Kalkidan Hassen Abate, Solomon M. Abay, Cristiana Abbafati, Nooshin Abbasi, et al. 2018. "Global, regional, and national age-sex-specific mortality and life expectancy, 1950-2017: A systematic analysis for the Global Burden of Disease Study 2017." *The Lancet* 392 (10159): 1684–1735. https://doi.org/10.1016/S0140-6736(18)31891-9.

Dowd, Jennifer Beam, Liliana Andriano, David M. Brazel, Valentina Rotondi, Per Block, Xuejie Ding, Yan Liu, and Melinda C. Mills. 2020. "Demographic science aids in understanding the spread and fatality rates of COVID-19." *Proceedings of the National Academy of Sciences of the United States of America* 117 (18): 9696–8. https://doi.org/10.1073/pnas.2004911117.

Hyndman, R, and G Athanasopoulos. 2019. *Forecasting: principles and practice*. 3rd ed. Melborne, Australia.

Kurz, Christoph F., Martin Rehm, Rolf Holle, Christina Teuner, Michael Laxy, and Larissa Schwarzkopf. 2019. "The effect of bariatric surgery on health care costs: A synthetic control approach using Bayesian structural time series." *Health Economics (United Kingdom)* 28 (11): 1293–1307. https://doi.org/10.1002/hec.3941.

Leon, David A, Vladimir M Shkolnikov, Liam Smeeth, Per Magnus, Markéta Pechholdová, and Christopher I Jarvis. 2020. "COVID-19: a need for real-time monitoring of weekly excess deaths." *The Lancet* 395 (10234): e81.

Li, Han, and Rob Hyndman. 2020. "Assessing Longevity Inequality in the U.S.: What Can Be Said About the Future?" *SSRN Electronic Journal*, 1–18. https://doi.org/10.2139/ssrn.3550683.

"Peru Implementation Working Group. Peru: An exceptional example of CRVS system advancement. CRVS country reports." 2018. March. Melbourne, Australia: The University of Melbourne; Civil Registration; Vital Statistics Improvement, Bloomberg Philanthropies Data for Health Initiative.

Robert B Cleveland, William S. Cleveland, Jean E. McRae, and Irma Terpenning. 1990. "STL: A Seasonal-Trend decomposition Procedure Based on Loess." http://www.nniiem.ru/file/news/2016/stl-statistical-model.pdf.

Scott, Steven L., and Hal R. Varian. 2014. "Predicting the present with Bayesian structural time series." *International Journal of Mathematical Modelling and Numerical Optimisation* 5 (1-2): 4–23. https://doi.org/10.1504/IJMMNO.2014.059942.

Table 16: Model fit population INEI

Departamento	range	.model	term	estimate	std.error	statistic	p.value
AMAZONAS	a0.9	arima	ar1	2.757741e+00	0.2627990	1.049373e+01	0.0000005
AMAZONAS	a0.9	arima	ar2	-2.546586e+00	0.5386097	-4.728073e+00	0.0006213
AMAZONAS	a0.9	arima	ar3	7.869771e-01	0.2795229	2.815430e+00	0.0168045
AMAZONAS	a0.9	arima	ma1	6.571431e-01	0.4755400	1.381888e+00	0.1944245
AMAZONAS	a0.9	arima	constant	1.431607e+02	98.6818444	1.450730e+00	0.1747666
AMAZONAS	a10.19	arima	ar1	2.720592e+00	0.0978614	2.780046e+01	0.0000000
AMAZONAS	a10.19	arima	ar2	-2.621606e+00	0.1850259	-1.416886e+01	0.0000000
AMAZONAS	a10.19	arima	ar3	8.904879e-01	0.1032087	8.628033e+00	0.0000032
AMAZONAS	a10.19	arima	constant	8.473559e+02	34.2248245	2.475852e+01	0.0000000
AMAZONAS	a20.29	arima	ar1	2.610032e+00	0.2135079	1.222452e+01	0.0000001
AMAZONAS	a20.29	arima	ar2	-2.445307e+00	0.3892194	-6.282593e+00	0.0000598
AMAZONAS	a20.29	arima	ar3	8.043516e-01	0.2157773	3.727692e+00	0.0033368
AMAZONAS	a20.29	arima	ma1	6.691210e-01	0.3768175	1.775717e+00	0.1034150
AMAZONAS	a20.29	arima	constant	2.290744e+03	41.0057408	5.586399e+01	0.0000000
AMAZONAS	a30.39	arima	ar1	8.317636e-01	0.1512684	5.498596e+00	0.0002623
AMAZONAS	a30.39	arima	constant	1.457050e+02	27.0185076	5.392784e+00	0.0003045
AMAZONAS	a40.49	arima	ar1	7.939404e-01	0.1732111	4.583658e+00	0.0010050
AMAZONAS	a40.49	arima	constant	2.568133e+02	12.6877334	2.024107e+01	0.0000000
AMAZONAS	a50.59	arima	constant	8.735000e+02	38.1774954	2.287997e+01	0.0000000
AMAZONAS	a60.69	arima	ar1	1.875195e+00	0.0315923	5.935602e+01	0.0000000
AMAZONAS	a60.69	arima	ar2	-9.750777e-01	0.0289986	-3.362496e+01	0.0000000
AMAZONAS	a60.69	arima	ma1	3.717597e-01	0.2475209	1.501933e+00	0.1640153
AMAZONAS	a60.69	arima	constant	4.692915e+01	2.8774274	1.630941e+01	0.0000000
AMAZONAS	a70.79	arima	ar1	1.622911e+00	0.1073952	1.511159e+01	0.0000000
AMAZONAS	a70.79	arima	ar2	-9.184459e-01	0.0965296	-9.514653e+00	0.0000025
AMAZONAS	a70.79	arima	ma1	8.111050e-02	0.4141343	1.958556e-01	0.8486458
AMAZONAS	a70.79	arima	constant	7.240616e+01	2.6012649	2.783498e+01	0.0000000
AMAZONAS	a80	arima	ar1	1.862031e+00	0.0332592	5.598547e+01	0.0000000
AMAZONAS	a80	arima	ar2	-9.761466e-01	0.0257886	-3.785192e+01	0.0000000
AMAZONAS	a80	arima	constant	1.164520e+01	0.8047098	1.447130e+01	0.0000000
ANCASH	a0.9	arima	constant	-1.508800e+03	40.2320031	-3.750248e+01	0.0000000
ANCASH	a10.19	arima	ar1	1.827095e+00	0.1202178	1.519821e+01	0.0000000
ANCASH	a10.19	arima	ar2	-8.828519e-01	0.1263681	-6.986350e+00	0.0000378
ANCASH	a20.29	arima	ar1	1.846786e+00	0.0787111	2.346284e+01	0.0000000
ANCASH	a20.29	arima	ar2	-9.246935e-01	0.0802366	-1.152459e+01	0.0000004
ANCASH	a30.39	arima	constant	1.199100e+03	104.3135426	1.149515e+01	0.0000004
ANCASH	a40.49	arima	ar1	7.713430e-01	0.1799327	4.286842e+00	0.0015941
ANCASH	a40.49	arima	constant	5.768922e+02	34.2109630	1.686279e+01	0.0000000
ANCASH	a50.59	arima	constant	1.860000e+03	111.4704427	1.668604e+01	0.0000000
ANCASH	a60.69	arima	ar1	1.872541e+00	0.0289375	6.470977e+01	0.0000000
ANCASH	a60.69	arima	ar2	-9.625976e-01	NaN	NaN	NaN
ANCASH	a60.69	arima	ma1	7.075156e-01	NaN	NaN	NaN
ANCASH	a60.69	arima	constant	9.366286e+01	6.3929997	1.465085e+01	0.0000000
ANCASH	a70.79	arima	ar1	1.776352e+00	0.1637284	1.084938e+01	0.0000007
ANCASH	a70.79	arima	ar2	-7.811910e-01	0.1657812	-4.712180e+00	0.0008262
ANCASH	a80	arima	ar1	1.941881e+00	NaN	NaN	NaN
ANCASH	a80	arima	ar2	-9.450035e-01	NaN	NaN	NaN
ANCASH	a80	arima	ma1	8.640383e-01	0.2973579	2.905719e+00	0.0156793
APURIMAC	a0.9	arima	constant	1.036136e+05	217.5796192	4.762102e+02	0.0000000
APURIMAC	a10.19	arima	ar1	1.964042e+00	NaN	NaN	NaN
APURIMAC	a10.19	arima	ar2	-9.654890e-01	NaN	NaN	NaN
APURIMAC	a10.19	arima	constant	1.685135e+02	NaN	NaN	NaN
APURIMAC	a20.29	arima	ar1	2.562937e+00	0.1958369	1.308710e+01	0.0000000

Table 17: Model fit mortality rate INEI

Departamento	.model	term	estimate	std.error	statistic	p.value
AMAZONAS	arima	constant	6.3195490	0.3135255	20.1564106	0.0000010
ANCASH	arima	constant	6.5487943	0.3834326	17.0793876	0.0000026
APURIMAC	arima	ar1	1.8625168	0.1740367	10.7018617	0.0000393
APURIMAC	arima	ar2	-0.8755477	0.1696462	-5.1610213	0.0020926
APURIMAC	arima	ma1	1.5136338	NaN	NaN	NaN
APURIMAC	arima	ma2	0.5341339	NaN	NaN	NaN
APURIMAC	arima	constant	0.1782021	0.0233164	7.6427818	0.0002619
AREQUIPA	arima	constant	5.3890608	0.1838113	29.3184488	0.0000001
AYACUCHO	arima	constant	7.4332902	0.5938814	12.5164564	0.0000159
CAJAMARCA	arima	constant	5.9859947	0.4012241	14.9193299	0.0000057
CALLAO	arima	constant	4.5869860	0.1053998	43.5198853	0.0000000
CUSCO	arima	constant	7.5021782	0.4167626	18.0010813	0.0000019
HUANCAVELICA	arima	ar1	1.8158450	0.3630773	5.0012631	0.0024493
HUANCAVELICA	arima	ar2	-0.8282803	0.3399848	-2.4362271	0.0507311
HUANCAVELICA	arima	ma1	0.3949038	0.7006805	0.5636003	0.5934489
HUANCAVELICA	arima	ma2	0.0385093	1.0426176	0.0369352	0.9717348
HUANCAVELICA	arima	constant	0.1418951	0.1235756	1.1482460	0.2945772
HUANUCO	arima	constant	6.5200369	0.3978019	16.3901607	0.0000033
ICA	arima	constant	4.9636328	0.1486123	33.3998879	0.0000000
JUNIN	arima	constant	6.6329956	0.3761317	17.6347691	0.0000021
LA LIBERTAD	arima	constant	5.3857540	0.2439788	22.0746778	0.0000006
LAMBAYEQUE	arima	constant	5.1816889	0.1709420	30.3125638	0.0000001
LIMA	arima	constant	4.8354043	0.0964952	50.1103078	0.0000000
LORETO	arima	constant	5.3937096	0.3596821	14.9957699	0.0000055
MADRE DE DIOS	arima	constant	5.3569624	0.4516718	11.8602991	0.0000217
MOQUEGUA	arima	constant	5.5516229	0.1910622	29.0566285	0.0000001
PASCO	arima	constant	6.2966512	0.4617546	13.6363597	0.0000097
PIURA	arima	ar1	1.9113458	NaN	NaN	NaN
PIURA	arima	ar2	-0.9515705	0.0756670	-12.5757601	0.0000155
PIURA	arima	ma1	0.7402194	0.8544024	0.8663592	0.4195839
PIURA	arima	ma2	0.3294593	1.3773999	0.2391893	0.8189197
PIURA	arima	constant	0.4791433	NaN	NaN	NaN
PUNO	arima	constant	8.1073604	0.5989815	13.5352446	0.0000101
SAN MARTIN	arima	ar1	1.9332177	NaN	NaN	NaN
SAN MARTIN	arima	ar2	-0.9632981	0.0463478	-20.7841028	0.0000008
SAN MARTIN	arima	ma1	0.7439952	0.6293392	1.1821847	0.2818551
SAN MARTIN	arima	ma2	0.4678362	0.8977050	0.5211469	0.6209318
SAN MARTIN	arima	constant	0.4479199	NaN	NaN	NaN
TACNA	arima	constant	5.1918996	0.2381897	21.7973284	0.0000006
TUMBES	arima	ar1	1.2874156	0.2086273	6.1708863	0.0008320
TUMBES	arima	ar2	-0.6427666	0.2977517	-2.1587338	0.0742064
TUMBES	arima	ma1	0.8684073	0.7867929	1.1037305	0.3119979
TUMBES	arima	ma2	0.1885324	NaN	NaN	NaN
TUMBES	arima	constant	1.8388821	0.1722865	10.6733942	0.0000399
UCAYALI	arima	ar1	1.9573821	NaN	NaN	NaN
UCAYALI	arima	ar2	-0.9783820	0.0239543	-40.8437397	0.0000000
UCAYALI	arima	ma1	0.7146891	0.6019498	1.1872903	0.2799825
UCAYALI	arima	ma2	0.8451716	0.9572400	0.8829255	0.4112520
UCAYALI	arima	constant	0.4338673	NaN	NaN	NaN
AMAZONAS	drift	b	-0.3909672	0.2849865	-1.3718795	0.2420128
ANCASH	drift	b	-0.5051168	0.3001792	-1.6827176	0.1677193
APURIMAC	drift	b	-0.6826128	0.2088459	-3.2684992	0.0308327
AREQUIPA	drift	b	-0.1522529	0.2361637	-0.6446924	0.5542381

Table 18: Model fit GBD

range	.model	term	estimate	std.error	statistic	p.value
a0.9	arima	ar1	0.9477953	0.0701530	13.5103957	0.0000001
a0.9	arima	constant	778.8166638	91.4702754	8.5144235	0.0000068
a10.19	arima	ar1	0.9097290	0.1021388	8.9067903	0.0000045
a10.19	arima	constant	251.9582934	20.4485956	12.3215451	0.0000002
a20.29	arima	ar1	0.8214739	0.1534372	5.3538120	0.0003219
a20.29	arima	constant	885.3616626	36.2468421	24.4258978	0.0000000
a30.39	arima	ar1	0.7312402	0.1878718	3.8922294	0.0029987
a30.39	arima	constant	1609.9919617	47.3090727	34.0313574	0.0000000
a40.49	arima	constant	8255.4219129	78.4484308	105.2337419	0.0000000
a50.59	arima	constant	12059.7561682	123.7740046	97.4336752	0.0000000
a60.69	arima	constant	17226.9073975	207.6111235	82.9768035	0.0000000
a70.79	arima	constant	25586.8717660	251.0217344	101.9309018	0.0000000
a80	arima	constant	1751.2088564	307.0020471	5.7042253	0.0002927
a0.9	drift	b	-476.1246706	143.4437186	-3.3192438	0.0105519
a10.19	drift	b	-59.7774487	33.8786015	-1.7644603	0.1156584
a20.29	drift	b	-49.0140409	55.5060191	-0.8830401	0.4029689
a30.39	drift	b	-24.9097299	69.7976150	-0.3568851	0.7304083
a40.49	drift	b	13.3889831	87.2920217	0.1533815	0.8818958
a50.59	drift	b	157.9807104	136.7392233	1.1553430	0.2812887
a60.69	drift	b	304.4965253	159.4000113	1.9102666	0.0924932
a70.79	drift	b	347.9980711	245.1434960	1.4195689	0.1935033
a80	drift	b	1751.2088563	325.6227986	5.3780290	0.0006632

Table 19: Model fit STL: head

Departamento	.model	FECHA	deaths	trend	season_
AMAZONAS	STL(deaths ~ season(window = "periodic"), robust = TRUE)	2017-01-01	2	2.333251	0.090
AMAZONAS	STL(deaths ~ season(window = "periodic"), robust = TRUE)	2017-01-02	1	2.382619	-0.976
AMAZONAS	STL(deaths ~ season(window = "periodic"), robust = TRUE)	2017-01-03	3	2.431986	-0.976
AMAZONAS	STL(deaths ~ season(window = "periodic"), robust = TRUE)	2017-01-04	2	2.490166	-0.976
AMAZONAS	STL(deaths ~ season(window = "periodic"), robust = TRUE)	2017-01-05	1	2.548346	-1.013
AMAZONAS	STL(deaths ~ season(window = "periodic"), robust = TRUE)	2017-01-06	1	2.511014	-1.013

Table 20: Excess total INEI per region

Departamento	model.pop	model.mort	mean	lower.CI95%	upper.CI95%
AMAZONAS	arima	arima	5.6796482	-50.47077	61.79354
AMAZONAS	arima	drift	4.4599948	-51.73953	60.61940
AMAZONAS	drift	arima	5.6908098	-50.46363	61.80803
AMAZONAS	drift	drift	4.4726169	-51.73319	60.63724
ANCASH	arima	arima	1868.1313705	1595.76289	2139.16577
ANCASH	arima	drift	1640.5828061	1363.77183	1915.88247
ANCASH	drift	arima	1874.6188127	1602.48851	2145.43088
ANCASH	drift	drift	1648.3785159	1371.84917	1923.41487
APURIMAC	arima	arima	155.5839892	97.76301	213.35216
APURIMAC	arima	drift	148.7014933	90.39537	206.91853
APURIMAC	drift	arima	155.3491233	97.48360	213.15846
APURIMAC	drift	drift	148.4428264	90.09854	206.69461
AREQUIPA	arima	arima	1613.1983146	1315.77771	1909.63284
AREQUIPA	arima	drift	1611.2670125	1312.83875	1908.63580
AREQUIPA	drift	arima	1614.8366307	1317.46889	1911.22111
AREQUIPA	drift	drift	1612.9074111	1314.53275	1910.22554
AYACUCHO	arima	arima	-181.0000000	-375.83894	13.83894
AYACUCHO	arima	drift	-181.0000000	-375.83894	13.83894
AYACUCHO	drift	arima	-181.0000000	-375.83894	13.83894
AYACUCHO	drift	drift	-181.0000000	-375.83894	13.83894
CAJAMARCA	arima	arima	8.2644406	-136.97967	153.45506
CAJAMARCA	arima	drift	-0.5461133	-145.95626	144.80539
CAJAMARCA	drift	arima	8.6480780	-136.58002	153.82384
CAJAMARCA	drift	drift	-0.0651676	-145.45691	145.26916
CALLAO	arima	arima	3352.6950187	2577.55135	4125.93972
CALLAO	arima	drift	3687.1942937	2926.05210	4447.20177
CALLAO	drift	arima	3375.7399356	2601.17406	4148.43447
CALLAO	drift	drift	3707.5156497	2946.76976	4467.14203
CUSCO	arima	arima	-334.1376831	-508.85412	-159.43866
CUSCO	arima	drift	-335.4848256	-510.22303	-160.76385
CUSCO	drift	arima	-334.1212165	-508.83680	-159.42296
CUSCO	drift	drift	-335.4650170	-510.20242	-160.74479
HUANCAVELICA	arima	arima	71.5460427	44.87256	97.93794
HUANCAVELICA	arima	drift	60.7426702	33.66582	87.51119
HUANCAVELICA	drift	arima	72.8579011	46.29077	99.15729
HUANCAVELICA	drift	drift	62.3078172	35.35546	88.96759
HUANUCO	arima	arima	287.8408453	224.49392	351.04312
HUANUCO	arima	drift	263.3475875	199.64728	326.89830
HUANUCO	drift	arima	287.8811349	224.53086	351.08636
HUANUCO	drift	drift	263.3974432	199.68682	326.95744
ICA	arima	arima	1501.5542882	1198.92541	1801.78014
ICA	arima	drift	1526.9695056	1224.58888	1826.94204
ICA	drift	arima	1512.8416533	1210.60187	1812.70419
ICA	drift	drift	1538.1091092	1236.11668	1837.71939
JUNIN	arima	arima	654.9400151	509.83833	799.69506
JUNIN	arima	drift	585.2346392	439.45798	730.66576
JUNIN	drift	arima	658.6236111	513.64309	803.26483
JUNIN	drift	drift	589.6264393	443.98562	734.92936
LA LIBERTAD	arima	arima	2863.8427214	2226.64696	3499.76514
LA LIBERTAD	arima	drift	2659.2451971	2018.90158	3298.21592
LA LIBERTAD	drift	arima	2880.7236156	2243.98818	3516.20839
LA LIBERTAD	drift	drift	2677.8930452	2038.05495	3316.38294
LAMBAYEQUE	arima	arima	2594.6876523	1986.99107	3202.28350