NATIONAL UNIVERSITY OF ENGINEERING

COLLEGE OF ECONOMICS AND STATISTICAL ENGINEERING AND SOCIAL SCIENCES

STATISTICAL ENGINEERING PROGRAM

EF012 PROJECT WORKSHOP

Design of a Statistical Processing System for Analyzing the Evolution of COVID-19 in Peru by Estimating the Reproducibility Index and Mortality Indicators.

by:

Lesly Shirley RODRIGUEZ GONZALES

Karen BARROZO RAZO

Lesly MAZA GÔMEZ

Jennifer ASPARRIN RAMOS

Professor

Carlos RISCO FRANCO

LIMA – PERU

2020

Abstract

Currently, the emergence and rapid spread of COVID-19 has had serious effects in most countries, including Peru. Therefore, the Peruvian government has seen the need to take preventive measures in the face of this re- cent pandemic. Despite the preventive measures that Peru has taken, we continue to be one of the most affected countries in Latin America due to the high number of infections. Mandatory social isolation, bonds granted by the government, mandatory immobilization, among others; it has not given the expected results compared to countries that have taken similar measures. Peru continues to fight a tough battle against COVID-19 and the different social and economic problems that forces the government to take short-term measures. That is why the need to analyze the data of the newly infected not as static data but as real-time information. In addition, these decisions at this stage must be made in a differentiated way, because as mentioned, the country is not only fighting COVID-19 but also the economic problems which were a consequence of the quarantine; It is necessary to analyze the regions in a differentiated way to take new measures that counteract economic and social problems, as well as avoid new cases.

The following research will compare the evolution of COVID 19 in all the departments of Peru without including Lima and Callao, taking into account their population and demographic characteristics. To achieve this objective, the real-time reproducibility index of COVID-19 in Peru was estimated with a Bayesian approach, and then the departments of Peru were classified according to this index. In addition to the positive cases analyzed, the cases of deaths from Covid for the different regions of Peru were compared using demographic characteristics such as sex and age as a comparison attribute. The analysis is done with the data provided by the government, through MINSA. The following were specifically used: positive cases and deaths from Covid-19 from the different regions of Peru, excluding Metropolitan Lima and Callao.

The results first showed inconsistencies with the data provided by the MINSA, these inconsistencies are mainly due to the fact that the number of tests performed is not significant for the population, one of the reasons is that Peru has a percentage of asymptomatic patients of 40 %. However, information provided by the government is needed to have an overview and thus control the progress of the pandemic. Rt estimates are of great importance for making decisions in the event of a pandemic, therefore these estimates found for the different regions present upper and lower limits with low ranges in most of them, which means that the model does not have much uncertainty. In addition, it could be observed according to the daily estimates of the Rt that the measures taken by the government in some cases were correct and in another late and as of August 31 the regions that are still on alert are Moquegua, Cusco, Apurimac, Arequipa.

General Index

Chapter 1

Background of the Problem

1.1. Problem Statement

Despite the preventive measures that Peru took in the face of the spread of COVID 19, we continue to be one of the countries hardest hit in Latin America due to the high number of infections. Because at first there was not a large number of tests, which did not allow to indicate the real number of infections in the country, triggering their rapid spread.

Mandatory social isolation, bonds granted by the government, mandatory immobilization, among others; it has not yielded the expected results compared to countries that have taken similar measures. Peru´ continues to fight this tough battle against COVID-19 and the different social and economic problems that forces the government to take short-term measures. That is why the need to analyze the data of the newly infected not as static data but as information in real time. In addition, these decisions at this stage must be made in a differentiated way, because as mentioned, the country is not only fighting COVID-19 but also the economic problems which were a consequence of the quarantine; It is necessary to analyze the regions in a differentiated way to take new measures that counteract economic and social problems, as well as avoid new cases. That is why the need for conclusions based on the analysis of those infected per day in an adequate way, this will help to anticipate future outbreaks as much as possible, since more appropriate confinement measures are taken, taking take into account the evolution of the pandemic in the different regions of the country.

1.2. Problem formulation

How will the contagion rate allow the government to make the right decisions to stop the advance of the pandemic in each of the departments of Peru?

1.3. The systematization of the Problem

- Which departments currently need more attention from the government?
- How will the evolution curve of COVID 19 behave in the coming months in Peru´?
- What is the relationship between the population characteristics, sex and age, for the evolution of infections in the departments of Peru?
- What is the relationship between the population characteristics, sex and age of the deceased in the different departments of Peru´?
- Has there really been a significant change in the infections of the most affected departments with the measures taken by the government?
- What even stricter measures should be considered by the government, analyzing population characteristics such as age and sex of individuals in the different departments of Peru with the highest rate of infections?

1.4. Research Objectives

1.4.1. General Objective

To study the evolution of the reproducibility index (Rt) of COVID 19 in all departments of Peru without including Lima and Callao, taking into account their population and demographic characteristics.

2.4.2. Specific Objectives

1. Classify the departments in the different stages of the evolution of the pandemic according to the reproducibility index and population characteristics.
2. Estimate the actual number of deaths from covid, based on the deaths reported to SINADEF.
3. Compare the relationship between population characteristics, such as sex and age, and those infected and those who died from COVID 19.
4. Compare the measures applied by the government in relation to the evolution of COVID 19 in the departments of Peru.

1.5. Research Justification

1.5.1. Practical motivations

Given the health emergency that the country is experiencing, due to the spread of COVID 19, it is considered important to observe the evolution of the pandemic differently by department using the Covid-19 reproducibility index. Being able to compare Rt at the department level and / or observe how Rt changes in a place can help us measure how effective regional policies are in curbing the spread of the virus.

1.5.2. Methodological motivations

In this research work, it is proposed to carry out a descriptive-explanatory study, since it seeks to describe the evolution of the contagion rate curves of the departments of Peru, as well as to make use of the contrast of Hypothesis to determine the relationship of the measures implemented by the government and the evolution of the contagion curve.

1.5.3. Theoretical motivations

To estimate Rt is a function of how many new cases appear each day. The relationship between the number of cases yesterday and the number of cases today provides a clue as to how Rt would behave. However, you cannot rely too much on any day to try to guess Rt, since daily case counts, they present some irregularities (delayed tests, inconsistency in the number of tests). However, using Bayes' Theorem, we can take new information obtained from each day's count of cases to adjust our expectations of what Rt is, approaching the real value as more daily data becomes available.

Chapter 2

Conceptual Theoretical Framework

2.1. Background

The current situation of the country and the world in general, has led to an analysis of the development of various statistical techniques in relation to covid19, this fact allows us to have a large amount of information available. Some of the studies carried out on a potential threat of emerging infectious diseases and that will be important for the objective of our research are:

- Lu´ıs M. A. Bettencourt / Ruy M. Ribeiro - 2008

  "Bayesian real-time estimation of the epidemic potential of emerging infectious diseases"

  The study was based on H5N1 avian influenza in humans, with more than 50 countries affected worldwide, which emphasizes the need for new approaches specifically aimed at detecting and monitoring the disease. evolution of emerging infectious diseases.

  The aim of the paper is to output standard SIR class models in a form that directly relates to time series data for emerging infectious diseases. The model is based on terms of surveillance observables and immediately suggests a simple graphical estimation procedure for the effective reproductive number R (mean number of cases generated by an infectious individual) of standard epidemics. For emerging infectious diseases, which generally show large relative fluctuations in the number of cases over time, we developed a Bayesian scheme for the real-time estimation of the probability distribution of the effective reproductive number and showed how to use such inferences to formulate tests of significance in future epidemiological observations.

This methodology has been applied to case data from World Health Organization reports to set limits on the transmissibility of H5N1 influenza in humans and establish a statistical basis for monitoring its evolution in real time.

- Anne Cori, * Neil M. Ferguson, Christophe Fraser y Simon Cauchemez

  "A new framework and software for estimating spawning numbers that vary over time during epidemics."

  The main objective of the study was to develop a generic and robust tool to estimate the number of variable reproduction in time, implemented with ready-to-use software and without problems, some of the problems mentioned are:

  - When the data aggregation time step is small (e.g., daily data), the estimates of R can vary considerably in short periods of time, producing a substantial negative autocorrelation.
  - Different studies have developed methods to achieve more uniform estimates, but the results may be sensitive to the selected time step or the smoothing parameters.

  For this reason, this article seeks to make novel and statistically sound analytical estimates of R and incorporates uncertainty in the distribution of the series interval (the time between the onset of symptoms in a primary case and the onset of symptoms in secondary cases).

  This tool should help epidemiologists to quantify the temporal changes in the intensity of transmission of future epidemics through the use of surveillance data (it implies the collection of data on a health problem, its analysis and later the use of the same in the prevention of diseases and in the improvement of the health conditions of the population.)

- Qun Li, M.Med., Xuhua Guan, Ph.D., Peng Wu, Ph.D., Xiaoye Wang, MPH, Lei Zhou, M.Med., Yeqing Tong, Ph.D., Ruiqi Ren, M.Med. Kathy SM Leung, Ph.D., Eric HY Lau, Ph.D., Jessica Y. Wong, Ph.D., Xuesen Xing, Ph.D., Nijuan Xiang, M.Med., et al.
  "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia."
  In this investigation, extensive individual case reports were collected in China and key epidemiological parameters, including the incubation period (4.2 days), were estimated.

Then, 2 mathematical modeling approaches were designed to infer the dynamics of the outbreak in Wuhan using high-resolution data from domestic travel and infections. The planned goal is to design 2 mathematical modeling approaches to infer the dynamics of the outbreak in Wuhan. through the use of high-resolution data on domestic travel and infections; whose results show that the doubling time early in the epidemic in Wuhan was 2.3 to 3.3 days. Assuming a serial interval of 6 to 9 days, a median R0 value of 5.7 was calculated (95% CI 3.8–8.9). In the first model, the first arrival model, the probability of the arrival times of the first known cases in provinces outside of Hubei was calculated based on the exponentially growing population of infected people in Wuhan before late January. This calculation involved the use of domestic travel data to calculate the probability that an infected person will travel from Wuhan to a given province based on the unknown actual number of infected people in Wuhan and the probability that they traveled. The arrival times of the first infected people in different provinces will reflect the growth rate of the epidemic in Wuhan.

In the second model, the case count model, we counted the detection of additional people who were infected in Wuhan and received their diagnoses in other provinces, and those people were explicitly modeled using a SEIR h ' Deterministic-stochastic hybrid (susceptible-exposed-infectious-recovered) model. This model was then fitted to the new daily case count data reported outside of Hubei province during the period prior to substantial transmission outside the province.

2.1.1. Theoretical Bases

Time series

A time series defined as X_t is a sequence of observations of a variable X taken at various instants of time t. These observations come from a distribution that can be different at each instant of time. They are classified as:

- A series is stationary if the mean and variability remain constant over time.
  - Average: $E(X_t) = E(X_{t+m}), \ \forall m \in t$
  - Variance: $Var(X_t) = Var(X_{t+m}), \ \forall m \in t$

- A series is non-stationary if the mean and / or variability change over time. However, if said series is differentiated one or more times, the resulting series will be stationary.

Components of the series. The time series is the sum of several components:

$$X_t = T_t + S_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

- Trend (X_t): long-term series smooth behavior or movement.
- Seasonality (S_t): oscillation movements within the year.
- Random (ε_t): random variations (does not have a common pattern) around the above components, usually it should be white noise; that is, it is distributed with zero mean and valence σ_ε^2. If, in addition, the distribution that follows is normal, then it is Gaussian white noise.

The random component of a time series does not help us to reveal important characteristics and the components (trend and seasonality), there are two techniques that eliminate this noise are smoothing and filtering.

- Smoothing can be used to fill in missing values and make forecasts. The smoothing methods can be: moving averages, simple exponential, Holt's exponential and Holt Winter's exponential.
- Filtering can be used to remove unwanted noise. Filtering methods can be: linear (mean filter and Gaussian filter) and non-linear (ordered statistical filters)

Smoothing Methods For our case, we use moving average smoothing and it is defined as:

1. **Moving Average,** It is a calculation used to analyze a data set in point mode to create series of averages. Thus moving averages are a list of numbers in which each one is the average of a subset of the original data. Suppose we have a time series, X_t where t = 1, 2, ..., T, and let p a positive integer. The moving average of order p is defined as the following series:
   If p is odd:

   $$M_{t+(p+1)/2}^{p} = \frac{1}{p}(X_{t+1} + \ldots + X_{t+p})$$

   If p is odd, the moving averages are calculated with the previous expression and then the definitive ones are obtained:

$$M_t^p = \frac{M_{t-1/2}^p + M_{t+1/2}^p}{2}$$

Filtering Methods For our case, we will use the Gaussian filter with a moving window and it is defined as:

1.Gaussian filter is a linear filtering technique which is applied to suppress Gaussian noise in a time series. To apply the Gaussian filter, it is first contaminated with Gaussian noise of zero mean and variance σ_ε^2, then the mean square error between the contaminated series and the original is calculated to know its value before decontamination. An optimal sigma value σ_ε is sought, which will be different in each series and for the different variances of the noise with which they are contaminated, where it produces the maximum noise reduction with the minimum mean square error for the filtered series. Finally, the mean square error between the original series and the filtered one is calculated, which should give a lower value than that of the contaminated series.

Differentiation of the series to eliminate the trend consists in assuming that it evolves slowly over time, so that at instant t it must be close to instant t1. In this way, if we subtract the previous value from each value in the series, the resulting series will be approximately trend-free. This operation is called series differentiation and consists of moving from the original X_t series to the Y_t series by:

$$Y_t = X_t - X_{t-1}$$

Thus, the differentiated series turns out to be stationary. Time Series Models. The Box and Jenkins models are known by the generic name ARIMA, which is derived from its three components AR (Autoregressive), I (Integrated) and MA (Moving Averages).

The ARIMA model allows describing a value as a linear function of previous data and errors due to chance, in addition, it can include a seasonal component. That is, it must contain all the necessary elements to describe the phenomenon.

The Box and Jenkins methodology is summarized in four phases:

- **The first phase** consists of identifying the possible ARIMA model that the series follows, which requires:

    Decide which transformations to apply to convert the observed series into a stationary series.

    Determine an ARMA model for the stationary series, that is, the p and q orders of its autoregressive structure and moving average.

- **The second phase:** Provisionally selected a model for the stationary series, we proceed to the second estimation stage, where the AR and MA parameters of the model are estimated by maximum likelihood and their standard errors and model residuals are obtained.
- **The third phase** is the diagnosis, where it is verified that the residues do not have a dependency structure and follow a white noise process. If the residuals show structure, the model is modified to incorporate it and the previous stages are repeated until a suitable model is obtained.

- **The fourth phase** is prediction, once a suitable model has been obtained, predictions are made with it.

Definition of basic terms

Disease reproduction number (R_t) In epidemiology, R is the reproduction number of an infectious disease. It tells you, on average, how many people will be infected by 1 infected person before he recovers (or dies).

R = 3 it means that an infected person, on average, will infect 3 more (the disease grows).

R = 5 means that an infected person, on average, will infect 5 more (the disease grows)

R = 1 means that an infected person, on average, will infect 1 more (the disease remains)

R < 1 means that an infected person, on average, may not infect anyone (the disease dies)

The demographer Alfred Lotka proposed the reproduction number during the 1920s, as a measure of the reproduction rate of a given population. In the 1950s, epidemiologist George McDonald suggested using it to describe the transmission potential of malaria. He suggested that if R_0 is less than 1, then the disease will disappear from the population because on average an infected person will infect less than one susceptible person. On the other hand, if R_0 is greater than 1, the disease will spread.

Basic reproduction number (R_0): It consists of the expected number of new infectious subjects that a previously infected subject would cause on average during its period of transmissibility in a population. completely susceptible. The R0 does not include cases caused in turn by secondary cases. It is fairly constant and is considered to be unchanged by vaccination. It depends on three parameters:

- Number of contacts per unit of time (τ)
- Contact transmission probability (c)
- Duration of the period of transferability (d)

$$R_o = \tau.c.d$$

SIR Model

SIR models are standard epidemiological susceptible-infected-recovered (SIR) models were developed by Kermack and McKendrick in 1927 and have been applied in various epidemic scenarios. These models estimate the theoretical number of people susceptible to becoming ill (susceptible), the number of sick (infected) and the number of people who can no longer transmit the disease (removed), in a population. on over time. The basic assumptions of the SIR models are:

1. The population is homogeneous and of a fixed size.
2. At any given time, each individual can only belong to one of the following groups: infected, susceptible, or removed.
3. The interaction between individuals is random.
4. There is no external intervention that changes the contact rate of the population. It was assumed that the population per state (N) is constant and that the number of susceptible individuals S_t, infected I_t and removed R_t are time-dependent variables.

In this way, the susceptible, infected and removed fractions were defined using the following formulas, respectively:

$$s = \frac{S}{N}, \quad i = \frac{I}{N}, \quad r = \frac{R}{N}$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta si \quad .....(1)$$

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta si - \gamma i \quad .....(2)$$

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \gamma i \quad .....(3)$$

where β is the transmission rate, γ is the recovery rate (or the inverse of the infectious period) and N is the total size of the population, so that N = S + I + R. In the ordinary model of equation 1 it is assumed that there are no births or deaths. At the beginning of the outbreak or epidemic (t = 0) we assume that the population is made up entirely of susceptible individuals and a single infectious individual.

This model is appropriate for emerging infectious diseases, which describe the probabilistic progression of case numbers due to the concurrent effects of human transmission and multiple introductions from a reservoir. The model is constructed in terms of surveillance observables and immediately suggests a simple graphical estimation procedure for the effective reproductive number R (mean number of cases generated by an infectious individual) of epidemics. standard. For emerging infectious diseases, which typically show large relative fluctuations in the number of cases over time, a Bayesian scheme is developed for the real-time estimation of the probability distribution of the number of effective reproduction and shows how such inferences are used to formulate tests of significance in future epidemiological observations.

The derivation of the model is beyond the scope of this document, but it is enough for our purposes to summarize it in the following equation that is obtained at the end:

$$k_t = k_{t-1} exp\{\gamma(R_t - 1))\}$$

where kt represents the number of new cases observed in t and γ is the reciprocal of the series interval (the time interval between infection and subsequent transmission).

Poisson Distribution

For the estimation of the real-time reproducibility index (R_t), the process of arrival of infected will be modeled, for this we will use Poisson distribution, it is a discrete probability distribution that from a frequency of occurrence mean (λ) expresses the probability of a certain number of events (k) occurring during a certain period of time.

$$P(k/\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Bayesian Approach

It is an alternative approach to conventional statistical analysis of data, it is based on Bayes' theorem: Given two events A and B such that P (B) ≠ 0, then

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B/A)P(A)}{P(B)}$$

The objective of Bayesian inference is the reallocation of credibility or probability through the different possibilities of one or more parameters, the distribution of credibility initially reflects the prior knowledge about those possibilities, then as new information or data is obtained, the initial credibility is reassigned. The possibilities that are more aware of the data gain in credibility, while those that do not decrease in credibility, the Bayesian analysis is represented as:

$$P(\theta/\ Data) = \frac{P\ (\theta) P\ (Data\ /\theta)}{P\,(Data)}$$

- P (θ/ Data): posterior distribution
- P (θ): a priori distribution
- P (Data/θ) likelihood function
- P (Data) normalization constant

This can be expressed as:

$$P\ (\theta/Data) \alpha P\ (\theta) P\ (Data/\theta)$$

To understand Bayesian inference we must understand each of its components:

- **Likelihood Function**: The plausibility relates the unobservable parameters to the observed data. It is a function of the parameters of a statistical model that infers its value from a set of observations and is defined as: P (Data/θ) = L(θ/Data)
- **A Priori Distribution**: The a priori distribution plays an important role in Bayesian inference since it represents our knowledge about the parameter before collecting the data, the more knowledge as a priori we will have better results for the posterior distribution.
- **Posterior Distribution**: The posterior distribution is the updated knowledge that we have of the data, it is what we learned from the data and it represents our result.
- **Normalization Constant:** It is the marginal distribution that can be represented for continuous data $P\ (Data) = \int P\ (Data/\theta)P\ (\theta)d\theta$ and for discrete data $P\ (Data)\ = \sum P\ (Data/\theta)P\ (\theta)$

Bayesian analysis of a model AR(1)

Be a process Y_t; t = 1, 2, 3, .. of type AR (1). Namely $Y_t = \theta Y_{t+1} + \varepsilon_t$ con θϵ (0,1); ε1, ε2, ... r. v. independent and identically distributed, exponentials of parameter λ where λ> 0 and θ unknown parameters.

Therefore, each observation is a linear function of the previous one plus a noise", not negative, of the exponential type. Let us denote the model by ARE(1).

Let n observations $y_1, y_2, y_3, ..., y_n$ of the model described, where y0 is a known initial value. Since the errors are independent and with equal distribution, the joint density of the observations will be:

$$f\ (y_n/y_0, \lambda, \theta) = \lambda^n exp\ [-\lambda(S_n - \theta R_n)] \tag{2.1}$$

where

$$y_n = (y_1, y_2, y_3, ..., y_n)\ ,\ S_n = \sum_{1 \leq t \leq n} y_t\ \ R_n = y_0 + S_{n-1\ t} \in \mathbb{R}^+ (t = 1, 2, ..., n)$$

We wish to make inferences about the parameter $\Phi = (\lambda, \theta)$. From this point of view, (2.1) is the likelihood function of Φ based on the observations. It is easy to check that the maximum likelihood estimator.

for Φ is:

$$\widehat{\theta} = \theta^* \quad y \quad \widehat{\lambda} = \frac{n}{S_n - \widehat{\theta} R_n} = \left[\overline{y}_n - \widehat{\theta}(\frac{y_0}{n} + \overline{y}_{n-1})\right]^{-1}$$

where $\bar{y}_n$ is the sample mean and θ^* is defined below.

Suppose that λ and θ are independent a priori and that our initial opinions are expressed by a non-informative one for the autoregressive parameter θ, and by a Gamma distribution, G (λ / α, β), for the parameter exponential. Namely:

$$f(\Phi) \propto \lambda^{\alpha-1} e^{-\lambda\beta} \qquad (\lambda, \alpha, \beta > 0) \tag{2.2}$$

Therefore, since the likelihood is

$$l(\Phi/y_n, y_0) = \lambda^n \prod_{1 \le t \le n} exp\left[-\lambda(y_t - \theta y_{t-1})\right]$$

it will have $y_t - \theta y_{t-1} \geq 0 \forall t$, and therefore

$$\theta^* = \min_{1 \le t \le n} \left\{ \frac{y_t}{y_{t-1}}, 1 \right\}$$

With this, the posterior for Φ verifies:

$$f(\Phi/y_n, y_0) \propto \lambda^{n+\alpha-1} exp\left[-\lambda(\beta_n |- \theta R_n)\right] I_{(0,\infty)}(\lambda) I_{(0,\theta^*)}(\theta)$$

Cluster Analysis

One of the most basic techniques in data analysis consists of dividing the data into a set of groups so that those within the same group have similar characteristics. There are several grouping algorithms, differing from each other in the definition of what constitutes a group and what it does to find it.

Metrics for grouping methods

The basic tool for grouping methods are measurements that measure the similarity or dissimilarity between the objects to be grouped. To measure the closeness between objects, other objects or some statistical measure of central tendency (mean, median, mode, etc.) is taken as a reference. The decision of the measure to use in any application is often a matter of a very careful choice, which results not only in the importance of the type of variable (quantitative or qualitative) but also in the prior knowledge of the expert on the domain of study. There are two types of measures or proximity indices: those that measure the closeness between objects or similarity or those that measure the distance between objects or dissimilarity. Both indices are complementary, which means that the higher the similarity index between two objects, the lower their dissimilarity index.

Similarity Measures

The metrics to measure similarity are known as distance and it is used as a synonym for metric. A larger value of distance indicates that the objects are further apart. Generally, before applying these distance measures, a process of standardization or typification of the data is carried out previously, in order to eliminate the influence of the units of measures of the variables in the analysis. Let $x_i = x_{i1}, x_{i2}, x_{i3}, ..., x_{ip}$ and $x_j = x_{j1}, x_{j2}, x_{j3}, \; , x_{jp}$ two vectors of observations of pdimensional variables, then some measures of similarity are the following:

Euclidean distance: It is the best known and most used. Measure the geometric distance between the vectors. Is defined as:

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^{p} (x_{il} - x_{jl})^2}$$

Manhattan distance: It is the sum in absolute value of the distances between the components of the vectors. Is defined as:

$$d(x_i, x_j) = max \left| (x_{ik} - x_{jk})^2 \right|$$

Mahalanobis Distance: It is the sum in absolute value of the distances between the components of the vectors weighting by the covariance matrix. is defined as:

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^{p} (\bar{x}_i - \bar{x}_j)' S^{-1} (\bar{x}_i - \bar{x}_j)}$$

K-means grouping

K-means clustering (Forgy, 1965; MacQueen, 1967) is a type of unsupervised learning used when you have unclassified data. The objective of this algorithm is to find groups in the data, with the number of groups represented by the variable k. This algorithm works iteratively, assigning each data to one of the k groups or clusters, based on its characteristics. K-means is one of the simplest and most widely used algorithms that solve the clustering problem.

The main idea begins by defining k centroids, one for each cluster. These centroids are initially selected randomly. The next step is to select each of the input data and assign it to the closest centroid. When there is no longer any unassociated data, the initial grouping is complete. At this point, k new centroids must be recalculated according to the mean of each group resulting from the previous step. Then a new stage of assigning the data closest to these new centroids must begin. As a result of this iterative process, the centroids move little by little, until a moment comes when they stop moving and reach their final position.

Kruskal-Wallis test

The Kruskal-Wallis test (by William Kruskal and W. Allen Wallis) is a non-parametric method to test whether a group of data comes from the same population. Intuitively, it is identical to ANOVA with the data replaced by categories. It is an extension of the Mann-Whitney U test for 3 or more groups.

Since it is a non-parametric test, the Kruskal-Wallis test does not assume normality in the data, as opposed to the traditional ANOVA. It does assume, under the null hypothesis, that the data come from the same distribution. A common way in which this assumption is violated is with heteroscedastic data.

The statistic is given by:

$$K = (N-1)\frac{\sum_{i=1}^{g} n_i(\overline{r}_i - \overline{r})^2}{\sum_{g}^{n_i}\sum_{j=1}^{n_i}(\overline{r}_{ij} - \overline{r})^2}$$

Where:

- n_i is the number of observations in the group i.
- j=1 ij
- r_{ij} is the rank (among all observations) of observation j in group i.
- N is the total number of observations between all groups.
- $$\overline{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$$
- $\overline{r} = \frac{(N+1)}{2}$ is the average of $\overline{r}_{ij}$

A correction for missing values can be made by dividing K by $1-\frac{\sum_{i=1}^{G}(t_i^3-t_i)}{N^3-N}$ of groups of different repeated ranges, and t_i is the number of repeated observations in each group within group i that has repeated observations for a given value. This correction makes K change very little unless there are a large number of repeated observations.

Finally, the p-value (p-value) is approximated by $Pr(\chi^2_{g-1} \geq K)$. If any n_i is small (> 5) the distribution of K may be different from the chi-square.

2.2. Hypothesis formulation and operation

2.2.1. Formulation

General Hypothesis

H_0: The measures implemented by the government regarding social isolation did not produce significant changes in the evolution of COVID 19 in the departments of Peru.

H_1: The measures implemented by the government regarding social isolation have decreased the evolution of COVID 19 in the departments of Peru.

Specific hypotheses

For cases of contagion:

H_0: The measures implemented by the government regarding social isolation did not produce significant changes in the R_t reproducibility index of COVID 19 in the departments of Peru.

H_1: The measures implemented by the government regarding social isolation have decreased the R_t reproducibility index of COVID 19 in the departments of Peru.

For the deceased:

H_0: The measures implemented by the government regarding social isolation did not produce significant changes in the number of deaths from COVID 19 in the departments of Peru.

H_1: The measures implemented by the government regarding social isolation have decreased the number of deaths from COVID 19 in the departments of Peru.

2.2.2. Operationalization

- The measures implemented by the government encompass the established periods of social isolation and the compulsory social immobilization (distinction of sex and age) of the state of emergency in Peru.
- The reproducibility index Rt is our base indicator to measure the contagion rate in the different departments of Peru 'during the state of emergency.
- The total number of deaths from March 6 to August 31.

Chapter 3

General Aspects of the Investigation

3.1. About the Investigation

3.1.1. Type of study

Descriptive:

The contagion curves will be described by using time series with daily data on the number of contagions as part of the estimation of the Reproducibility Index, for each of the departments of Peru, not including Lima, during the period of March 15, 2020 (beginning of the state of emergency) to June 14, 2020 (selected cut-off date). To evaluate the measures established by the government and their impact generated in search of the control of the COVID 19 pandemic, previously research has been carried out on the modeling of the reproducibility index for the coronavirus pandemic, whose title is Peru - What factors are behind of the differential impact of COVID-19 in Peru? carried out by BBVA Research, whose file will serve as a guide for this study.

Explanatory:

It is known that currently according to a source from the MINSA, the contagion curve has not decreased as expected at the time of declaring the state of emergency in Peru, there are many causes that could explain this, according to the report whose title is Peru - What factors behind the differential impact of COVID-19 in Peru? carried out by BBVA Research, mentions the following causes:

1. Greater frequency of use of traditional markets for the supply of food in households.
2. Agglomerations in these traditional markets, one of the sources of infection, enhanced by the reduction in the number of hours of operation due to the curfew.
3. Low banking and crowds in bank offices for the collection of state transfers. Crowding is also enhanced by the curfew.
4. Greater labor informality and the incentive to break confinement to obtain income.
5. Greater overcrowding in homes, which makes it difficult to isolate positive cases.

However; It is currently known that the reality of Lima is not the same in all departments of Peru, therefore, the need to explain, through the modeling of the reproducibility index R_t, how the contagion curve has evolved in each department and what causes They have led to a decrease or increase in infections, allowing an evaluation to be generated as to which departments would already be able to restart their activities, continuing with protective measures to prevent the outbreak of this pandemic.

3.1.2. Research Method

Analysis and synthesis:

The study will be carried out individually for the contagion curve of each department of Peru except Lima, with this it will be possible to make conclusions regarding the expected division of the departments potentially available to restart their activities and departments that still present a growth curve very big.

3.1.3. Information gathering sources and techniques.

Secondary sources:

For the database, use will be made of a compilation of the accumulated number of infections in each department from the beginning of compulsory social isolation until June 14, 2020, information provided by the Ministry of Health.

3.1.4. Information processing

Statistical analysis techniques:

The main input for estimating the reproduction number in real time is the daily time series of new detected cases, provided in the Peruvian case by the Ministry of Health (MINSA).

3.2. Innovative nature of the Proposal

The effects of the pandemic due to the spread of COVID-19 are shaking socio-economic structures globally. The impacts are estimated in astronomical figures of loss of life and jobs. Our country, like others, is deploying sanitary and economic measures to halt the advance of the epidemic with great uncertainty even about the duration of the health emergency period and the real consequences of the pandemic on the population. From their side, the world of science, innovation and the business sector are deploying efforts to support the emergency and the actions of governments at breakneck speed. We are seeing in real time scientists collaborating worldwide, in a race against time, to develop a vaccine and antiviral medications, the industry modifying its production processes in order to contribute to the large-scale manufacturing of health supplies, and to innovators organizing to provide solutions to the most urgent problems. Although the period of confinement imposed on the population allows us to gain some time to reduce infections and thus lighten the burden on the health system, this period should be used to prepare a proposal on how the population could gradually return to activity. For us, the contagion rate (R_t) is taken as an important factor within the proposals that would be proposed as actions in the event of a possible re-growth; however, this would not guarantee that the objectives of slowing down the spread of the virus, taken by the state, will decrease, since it would be taken in a general way with respect to Peru. The proposal for calculating the contagion rate, in this project, has as an innovative purpose to carry out a projection of Rt values at the departmental level since this could generate greater control over those regions most affected by COVID-19 and monitor the evolution of infections intensively in order to get out of this situation and return to what would be a “new normal”.

3.3. Impact that the research will have on Public Policy.

The serious COVID-19 coronavirus pandemic has dominated public policy discussion in the country. The Peruvian government has taken emergency measures to stop the spread of the virus, such as suspending national travel, closing borders and imposing mandatory quarantines. Given the lack of immunity in the population to this new epidemic, the capacity of the health sector will be overwhelmed if infection rates are not stopped. The predictions of the epidemiological models change almost daily, but according to the policies that have been applied so far, they have not had the results expected by the government. In fact, the strict confinement that the Peruvian government has been applying is helping to some extent to combat the spread of COVID-19.

However, currently it is facing the economic and social pressures that have arisen due to strict confinements, but as is known, the evolution of COVID-19 in Peru is not advancing in the same magnitude in all regions of the country. That is why the application of focused confinement is an option to curb the contagion rate in Peru, it would also help economic reactivation in places where the contagion rate is lower; But for the state to establish policies of focused confinement, it is necessary to know how the contagion rate is advancing and will advance in each region of the country in real time.

3.4. Limitations of the Research Proposal

Although the real-time reproducibility index (Rt) is a useful measure to know the level of control of the pandemic, it has flaws because the data source to find this metric depends on the nu Number of new cases detected that is provided by the Ministry of Health (MINSA). These data suffer from irregularities for different reasons: tests can give false negatives, delay in test results, poor accessibility to health centers to perform the test on time, asymptomatic cases and those infected in serious condition that death comes to them before taking the test. However, the more data we obtain and the more coherent they are, the more certain we can be of the estimate and the decisions made regarding the result. In addition, there is an interval, the HDI (interval of higher density) that confirms that 95% of the true values are within this interval, this helps us to guarantee that the data does not deceive us, and it is not concluded something that could be a serious mistake.

Chapter 4

Methodology

4.1. Information sources and data quality

The data used for this investigation are those provided by the government, through MINSA. As of May 22, 2020, the MINSA open data is released on the topics of: Positive Cases due to Covid-19, and Deaths due to Covid, this Database is the result of data processing, which originates from the collection of information on COVID-19 by the National Institute of Health (INS) and the National Center for Epidemiology, Prevention and Control of Diseases (CDC).

1. **Confirmed cases**: According to the latest report from the National Center for Epidemiology, Prevention and Control of Diseases of the MINSA, a confirmed case is considered to be a case that, being a suspect, has a positive laboratory test for COVID-19, be it a reaction test Polymerase chain reverse transcriptase in respiratory samples RT-PCR and / or a rapid test for detection of IgM, IgG or IgG / IgM. For the suspect case, people with acute respiratory infection, who present a cough or sore throat and at least one or more of the following signs / symptoms are considered: General malaise, Fever, Headache, Difficulty breathing, Nasal congestion.
2. **Cases of deceased**: The cases of deaths that are registered by COVID-19 are only those in hospitals and those that are under surveillance according to the MINSA. The SINADEF is the one that registers the deceased with evidence, without evidence, with suspicion, at home, in hospital, etc.
3. **Number of tests performed**: The tests carried out by the government are focused, that is, in risk areas and vulnerable people. The samples are especially applied to newcomers to the country, returnees, people in prisons, shelters. In addition, to the PNP, Armed Forces,

Institutions, Municipalities, supply centers, transportation stations, cargo carriers, Health personnel. And in the first week of July, Operation Tayta began to be applied, which seeks to serve the most vulnerable populations in their homes, such as those over 60 years of age, and people with chronic diseases that aggravate their health condition if they become infected. However, these data published by the MINSA do not include the people sampled by PR the private IPRESSs that screen company workers in the framework of the economic reactivation, because the objective of this screening does not allow to identify new cases in suspicious persons Prepared by the National Center for Disease Epidemiology, Prevention and Control - MINSA.

In this context, it can be noted that the figures currently circulating are not scientifically comparable because they are not compiled in a systematic and representative way. The testing strategy is mainly case-based, which means that people with specific symptoms or a certain severity of the disease are especially tested. Now, in fact, the samples have been increasing throughout this stage that Peru is experiencing due to COVID-19, but the main problem is that the samples do not reach all the people who are infected by Covid-19 for two main reasons; there is a large percentage of asymptomatic people and screening is prioritized in regions with the highest number of infections at the time. This generates that other regions increase the number of infected and deceased silently.

4.1.1. Inconsistencies in the data of the deceased

For this investigation, the main source of data is the MINSA, regarding the report of deaths that it presents, inconsistencies have been found for the figure reported as of August 31, although it is true that the MINSA collects the information on the deaths that were previously mapped As confirmed infections, all those who did not undergo an autopsy or were studied before death are not part of the death count, so analyzing the database of the National System of Deaths (SINADEF), which shows the number of deaths reported daily for different causes, an exponential growth is observed for the months once the pandemic arrives in PERU.

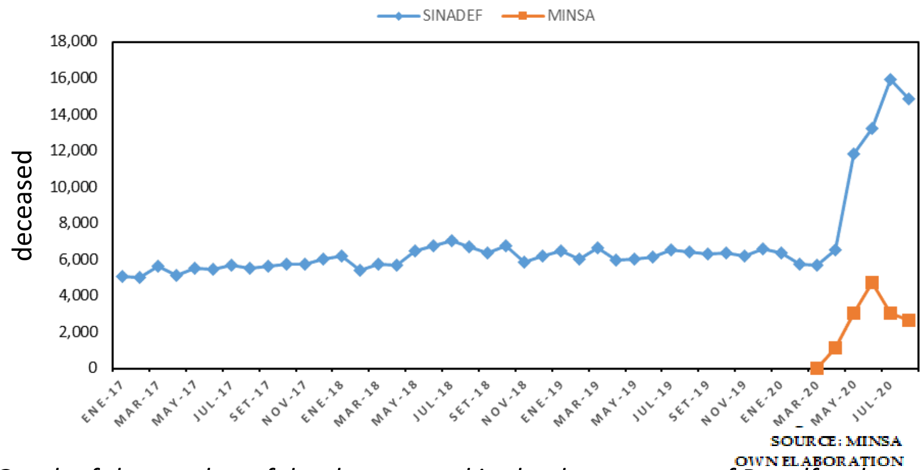


Figure 4.1. *Graph of the number of deaths reported in the departments of Peru 'for the period 2017 - 2020*

4.2. Situation of COVID-19 in Peru

4.2.1. Analysis of Contagions by COVID-19 in Peru

The index case (first case) of COVID-19 was a Peruvian with a history of travel to Spain, France and the Czech Republic, from which 8 confirmed cases of 1st generation (direct contacts), 2 confirmed cases of 2nd generation were detected (contacts of a confirmed case of the first generation) and 2 confirmed cases of the 3rd generation (contacts of a confirmed case of the second generation), one of the latter having died, according to the report published in the Epidemiological Analysis of the Current Situation of COVID- 19 in Peru, based on information from the Epidemiological Surveillance and Field Investigation of MINSA.

From March 6 to 15: the first case was detected in Lima (03/06/2020), followed by Arequipa, Huánuco, Ica, Cusco, Callao, La Libertad, Lambayeque, Ancash and Piura. Between March 16 and 30: new cases were reported in Loreto, Madre de Dios, San Martin, Junín, Tumbes, Cajamarca, Pasco, Ayacucho, Tacna and Huancavelica.

Despite having been one of the countries that earlier implemented measures to increase social distancing and restrict the mobility of its citizens, Peru is currently one of the countries hardest hit by COVID -19 in Latin America.

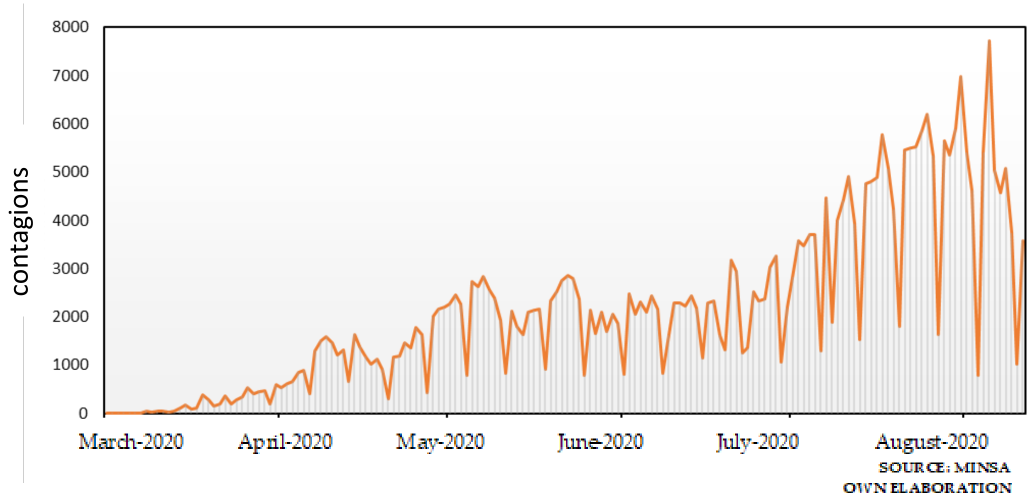
Regarding the measures, on March 16, the government decreed a state of health emergency, closed the country's borders and ordered that people leave only for essential actions such as buying food and medicine. Two days later, the government tightened the measures of social isolation by decreeing curfews at different times in all cities.

In the fifteenth week of compulsory social isolation, the government announced the extension of the state of emergency until July 31. This measure lifts the mandatory social isolation throughout the national territory and provides for a focused quarantine approach. According to Supreme Decree No. 116-2020-PCM, published on June 26 of this year, the regions that remained in compulsory isolation were Arequipa, Ica, Junín, Huánuco, San Martin, Madre de Dios and Ancash. However, social immobilization, or curfew, was maintained throughout the country.

After that, while Peru works to stop the spread of the virus in the country, the national executive decided to extend the term of the state of emergency and the duration of the targeted quarantine until the end of the month, specifically, to August 31.

There are five regions of the country in which the quarantine is maintained in all its provinces, which are Arequipa, Ica, Junín, Huánuco and San Martin. As for the rest of the territories, there are only some of the provinces that maintain the quarantine with the same strict measures that the Government of the nation dictated. Precisely, in these territories in which the quarantine is maintained, the curfew remains in force from 20:00 PM to 04:00 AM.

Making a count of the daily contagion cases that is carried out to August 31 in the departments of Peru without including Lima and Callao, the following graph is shown:



The evolution of the pandemic is a key factor in the extension of social isolation measures, which have a strong impact on the level of economic activity. Therefore, having a clear vision of the evolution of the disease is imperative to be able to estimate the performance of the economy this year. However, tracking the number of new confirmed cases daily is not enough. For example, in the last week these have decreased significantly, from 5 087 (28-Aug) to 3 587 (31-Aug). However, the number of deceased has a very random behavior, and it does not allow estimating a trend or season to forecast its values, to say that it was reduced from 64 (28-Aug) to 48 (31-Aug) does not show a significant reduction. This leads us to conclude that the lower number of new infections confirmed in recent days could be reflecting, to a large extent, a lower number of reported deaths.

The factors that would have driven the cases of COVID-19 in Peru.

Low social distancing and some sociodemographic characteristics have driven a higher number of cases and deaths in Peru, compared to other Latin American countries. These two factors have manifested themselves through five dimensions, according to the report published by BBVA Research:

1. The greater frequency of use of traditional markets for the supply of food in households.
2. Agglomerations in these traditional markets, one of the sources of infection, enhanced by the curfew.
3. Low banking and crowds in bank offices for the collection of state transfers, also enhanced by the curfew.
4. Greater labor informality and the incentive to break confinement.
5. Greater overcrowding in homes.

All this in the context of a health system whose public pillar was quite weak before the pandemic and which is currently under pressure from both the public and private spheres.

1. **The highest frequency of visits to traditional markets**

   One of the difficulties that households in Peru have and that explains the need to go out to markets frequently is the lack of food preservation infrastructure. Peru stands out as one of the Latin American countries with the lowest penetration of refrigerators in homes (50%), far behind other countries in the region. In the case of low-income households, the estimate is that only one in five households (21.9%) has this appliance at home. This, together with the lower price of fresh foods compared to processed ones, explains a high frequency of visits to traditional markets.

2. **Agglomerations in traditional markets turned them into a source of contagion.**

   Tests carried out in mid-May (mostly serological) to traders in the main traditional markets, wholesalers and retailers, have shown a high prevalence of positive cases (between 50 and 85% of traders). The case of the Lima wholesale fruit market, where 86% of its vendors tested positive, is particularly illustrative of what has been happening in the markets. In addition, curfews, which in most cities began at 6:00 p.m. (in some at 4:00 p.m.), limited opening hours (including closing all Sunday because the curfew applies all day) and contributed to the crowds on the days when you can shop.

3. **Low banking and crowds in bank branches.**

   Along with restrictions on people's mobility, the government approved a universal bond of around USD 220 for 6.8 million households (83% of all households), which included those with the lowest incomes. However, Peru is one of the Latin American countries with the least banking use: 57% of adults in Peru do not have a bank account, a percentage that increases to 73% of those adults who are among the 40% with the lowest income. Consequently, many of the beneficiaries had to withdraw government aid in person at banking institutions, with reduced opening hours due to the curfew. This has also generated crowds in bank offices.

4. **Greater labor informality and the incentive to break confinement.**

   About 71% of the economically active population in Peru is employed in the informal sector, one of the highest percentages among the largest economies in Latin America. Most of these activities (self-employed or employed) generate income on a day-to-day basis, so their operating logistics are also daily. Although we do not have statistics for the Peruvian case, in the case of Colombia (a country that also shows high levels of informality), 42% of the informal workers who work in the branches most affected by COVID-19 need to mobilize to be able to work. Therefore, the very structure of the labor market and economic activities have generated contagion opportunities, which have been enhanced through public transport.

5. **Greater overcrowding in homes.**

   Finally, an additional differential element in the Peruvian case is the overcrowding of the homes to which the infected return in markets, bank offices or public transport.

Positive cases of COVID-19, according to sex and age group

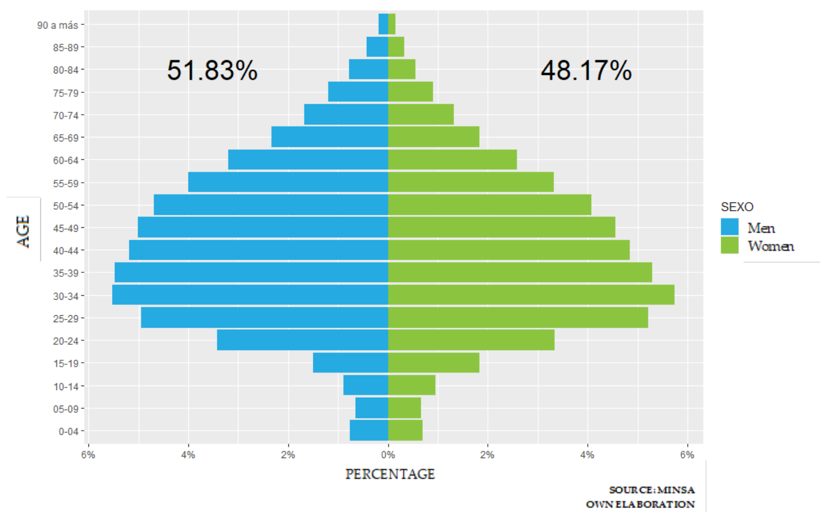


Figure 4.3. *Positive cases of COVID-19, according to sex and age group*

The population pyramid of the total infected in Peru is observed (not counting Metropolitan Lima and Callao), as of August 31. First you can see a difference between female and male sex; However, this difference is not very marked since the highest percentage is for the male group, which accumulates 51.83% of confirmed cases versus 48.17% for the female group. On the other hand, the age groups where the greatest number of infected people accumulate are in the range of 30 to 50 years of age for both sexes.

Age group	% 31 - may	% 30 - jun	% 31 - jul	% 31 - aug	Nº Contagions
Boy (0 – 11 years)	2.5 %	2.7 %	2.9 %	3.5 %	12,215
Teen (12 – 17 years)	1.9 %	2.1 %	2.5 %	2.9 %	10,302
Young (18 – 29 years)	18.8 %	18.8 %	18.5 %	18.6 %	65,015
Adult (30 – 59 years)	60.8 %	59.1 %	58.4 %	57.6 %	201,671
Elderly (60 to more years)	16.0 %	17.3 %	17.7 %	17.4 %	61,051

The adult and older age groups represent more than 75% of all COVID-19 cases, followed by the youth group. Children and adolescents are the groups that contribute the lowest% of cases. In the group of children, the percentage variation increased by 20% in the last month (it ranged from 2.9% to 3.5%), however, it did not reach values at the beginning of the pandemic, which was 3.2%. explained by the lower proportion of susceptible adults and older adults already affected in greater incidence so far as the pandemic goes.

The graph shows a linear time series of those infected daily by COVID-19, categorized by age ranges. The significant differences between the different age ranges are seen for the number of infected daily. During the last 15 days of March, the number of daily infected increases in a similar way. However, since April the slopes of each series have been differentiating. Three marked behaviors can be observed that will be described.

- On the one hand, the series of infected for adults (30-59 years) has a drastic change in slope from the first days of April, this growth is multiplicative until the end of June. Since June the series has an almost constant behavior. However, as of August, the series of adults showed significant growth, with peaks that reached up to 4,000 thousand infected per day.

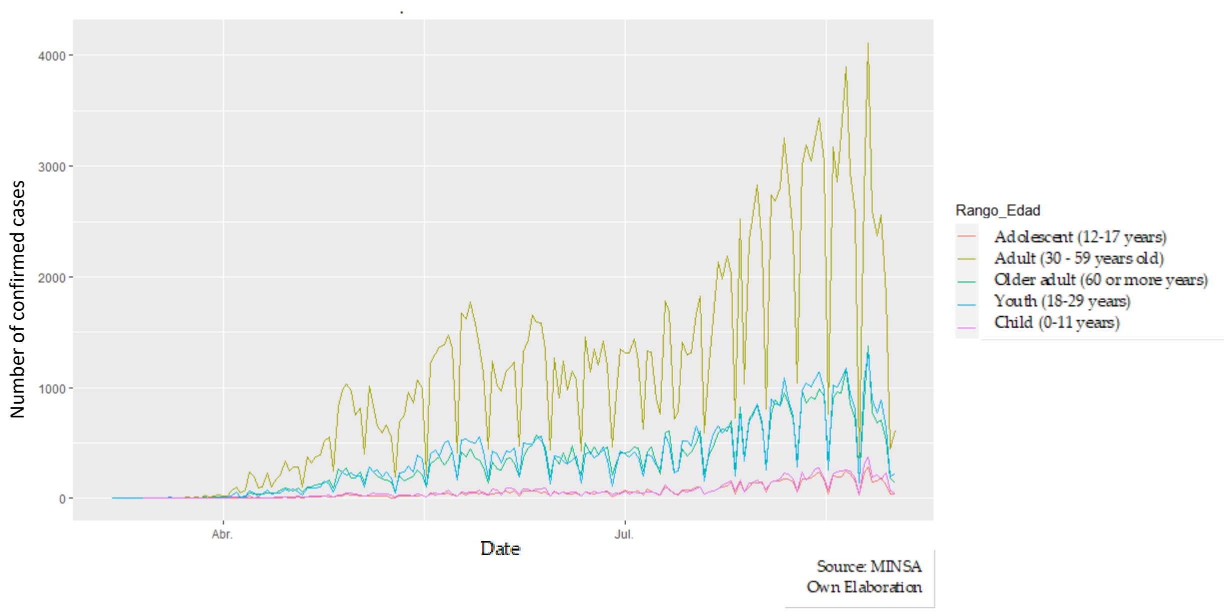


Figure 4.4. *Graph of daily cases by age ranges*

- On the other hand, there are the age ranges for young people and older adults, which have almost the same behavior. From the second week of April its slope began to increase but with a small slope compared to the series of adults. Daily cases reached more than a thousand cases per day. This first difference between the age series could be due to the restrictions applied by the government for these age groups.
- And finally, there are the age ranges of children and adolescents, both graphs have a very similar and constant behavior throughout the months that are being considered in the study. They have a slight increase from the month of August, but it is not significant compared to the other series. Furthermore, the daily cases for these age ranges do not exceed 300 daily cases.

The following graph shows a linear time series of those infected daily by COVID-19 categorized by sex. In the first weeks of the beginning of the pandemic, a marked difference between the series of both sexes cannot be visualized. However, the men's time series is slightly above the women's series for the months of May and June. In the month of August this has changed, since the series of women has slightly exceeded the series of men and in a few weeks, they coincided in numbers.

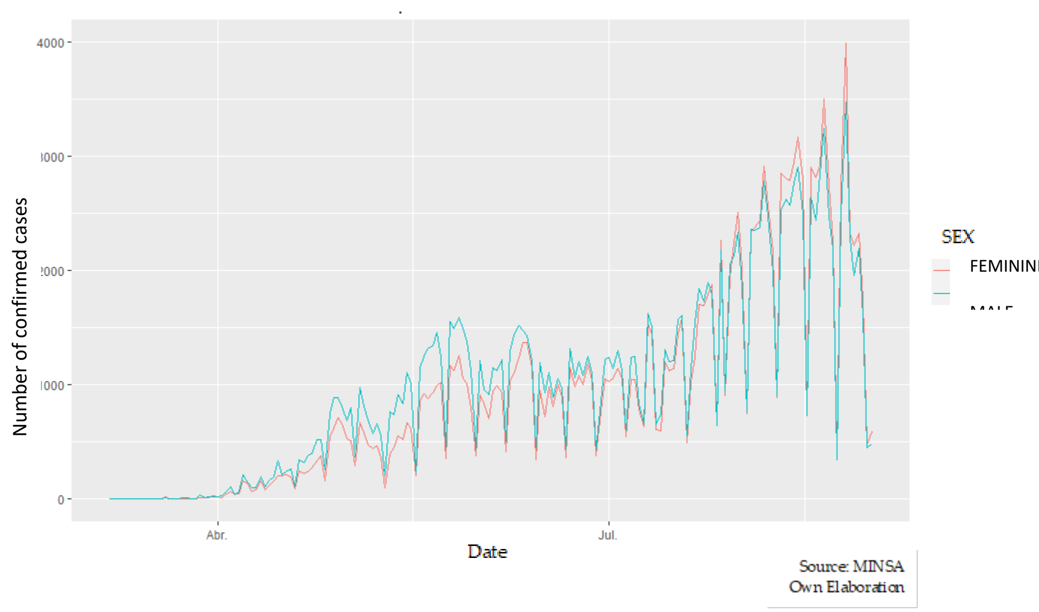


Figure 4.5. *Graph of daily infections differentiated by sex.*

The bar graph shows the number of new infections per day by department, those in red represent the departments with the highest risk of contagion such as Piura, Lambayeque, La Libertad, Loreto, Ancash, Ica and Arequipa, however only 3 of these remain in mandatory isolation until the end of July. The following departments are those that have regular risk Ucayali, San Martin, Junín, Huánuco, Madre de Dios to Ayacucho, those with the highest number of new infections per day are in isolation with the exception of Ucayali which is more prone to being a high zone Finally, the departments with the lowest risk of contagion are those in blue, note that the groups from highest to lowest risk have a certain characteristic: Higher risk: (60%) the northern zone Regular irrigation: (28%) jungle zone Less Risk: (12%) Sierra area.

Comparison of the impact of COVID-19 between the departments of Peru

To compare impact data between departments it is necessary to consider the number of population and not make a false step of using raw numbers. This process is called "normalization" of values. For example, for every 100,000 inhabitants.

The graph represents four heat maps for the number of infected at the end of the months of May, June, July and August for each departments of Peru, for this the accumulated data of confirmed cases (grouped by department) was used, but instead Using the raw cases, the cases per 100,000 inhabitants were used. Leaving aside Lima since the analysis of this research is based on the evolution of COVID-19 in the rest of Peru; The results obtained show how the state of the pandemic has changed over time.

- The first analysis of the accumulated infections as of May 31, you can see the marked difference between the departments of the jungle and coast regions, which present greater infections. On the other hand, there are the departments of the mountains where it can be seen that the pandemic is not so advanced, especially the departments of the southern zone.
- By the end of July it can be observed that the jungle regions such as Loreto, Ucayali and Madre de Dios are considerably increasing the number of infected. Like the departments of the north coast.
- By the end of July, the June scenario is still maintained, but it intensified in the departments of Ucayali and Madre de Dios. In addition, cases began to increase in the southern zone, especially in the departments of Ica, Arequipa and Moquegua.
- By the end of August, a scenario is very different from the rest of the other months. Ucayali and Moquegua have a large number of infected per thousand inhabitants.

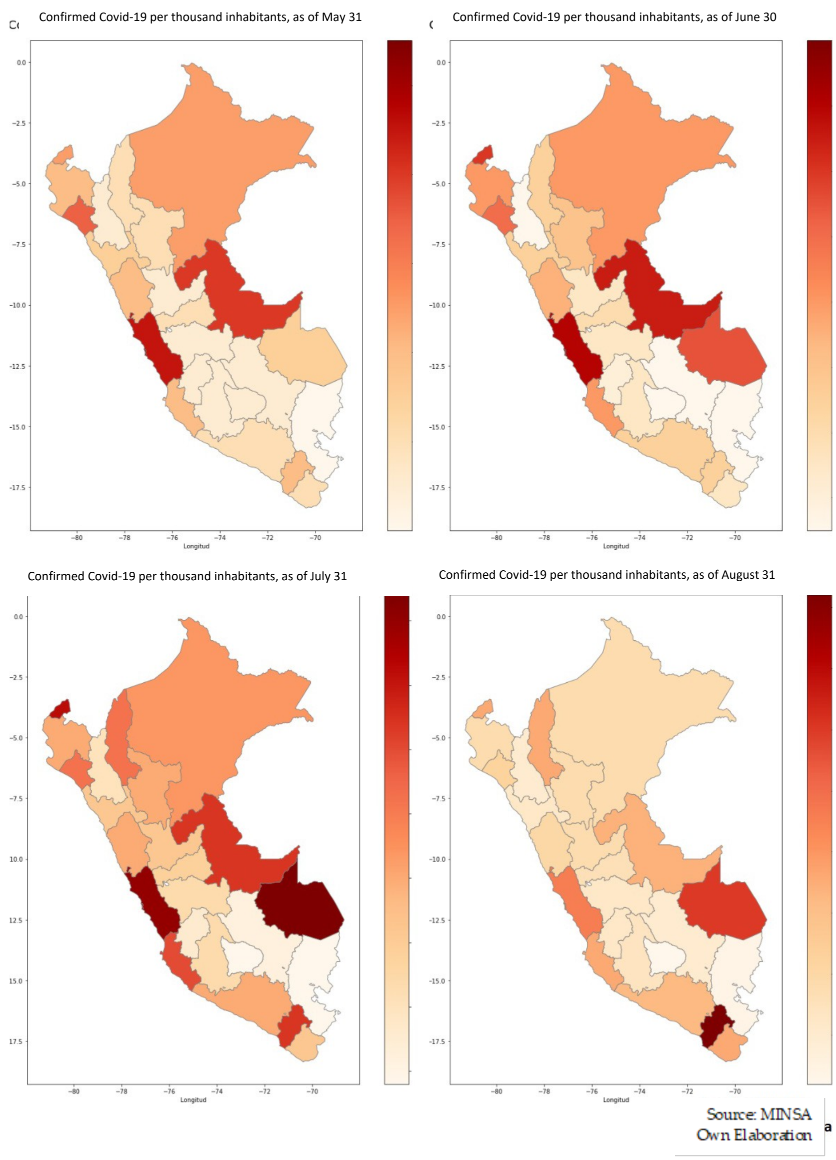


Figure 4.6. *Heat map of accumulated cases per thousand inhabitants, for the months of May, June, July and August.*

Cluster analysis of contagion cases for the departments of Peru

As the pandemic in Peru is known, it progresses differently in each department; therefore, it is important to recognize which departments are more advanced and which are not. The figure shows 3 groups of departments which are part of a different stage with respect to the advance of COVID-19.

The first group contains the departments of Ica, Loreto, Ancash, San Martin, Arequipa and La Libertad; These regions are the ones that have reached the highest number of infected with respect to the other regions. In the second group are regions such as Madre de Dios, Ucayali, Tumbes, Moquegua and Amazonas, which are in the middle part of the pandemic, but are on the way to reaching the highest number of infections. Finally, there are the regions of the mountains such as Huánuco, Cusco, Apurimac, Ayacucho, Cajamarca, Puno, Huancavelica and Pasco, which have minimal amounts of infected compared to the maximum amounts reached by other departments; Although these regions still have figures that are not large, these regions cannot be neglected.

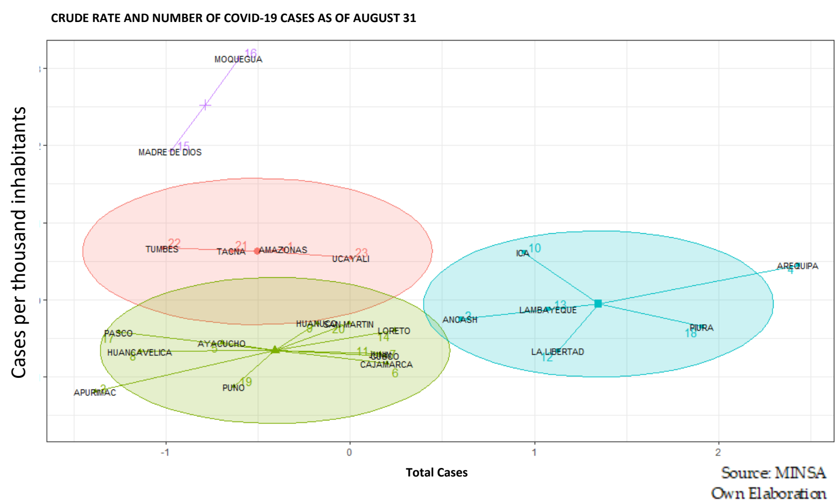


Figure 4.7. *Cluster of contagion cases for the departments of Peru*

4.2.2. Analysis of deaths from COVID-19 in Peru

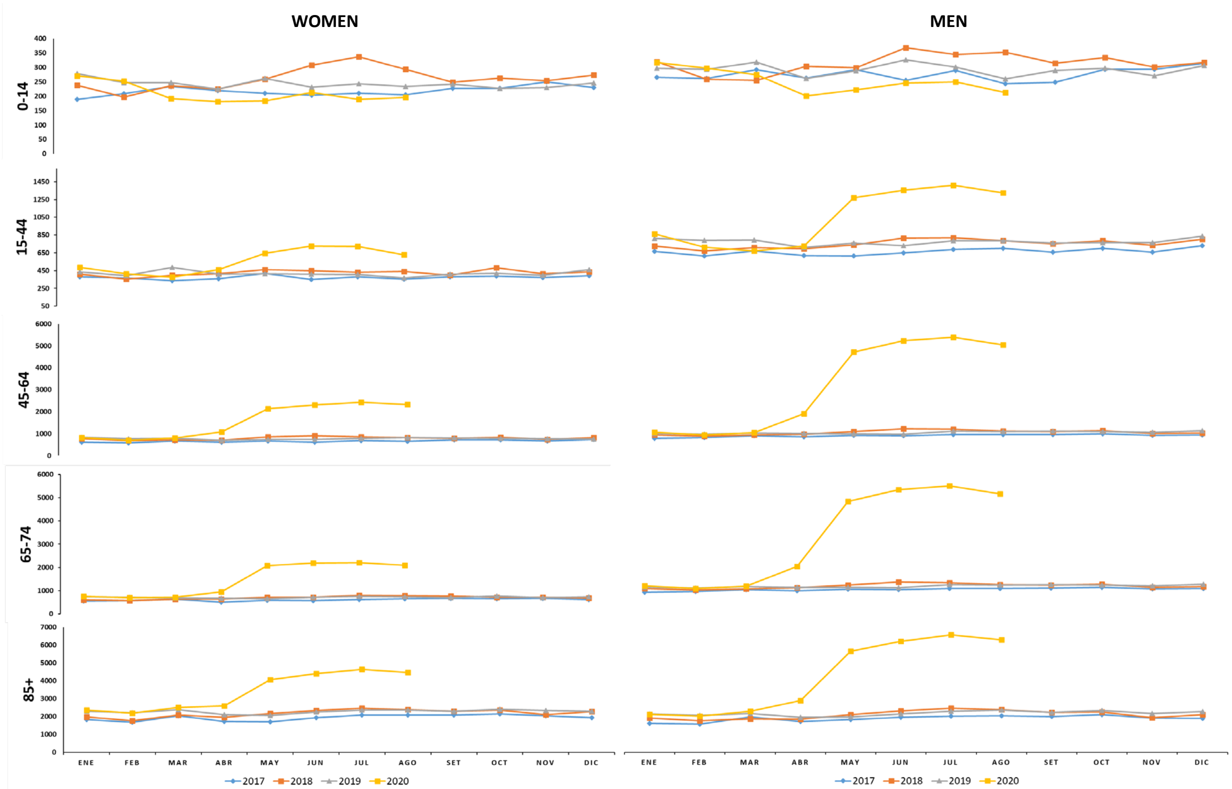


Figure 4.8. *Graph of the deceased by age ranges.*

Reported Deaths Women

In the following table we can see the deceased women reported in the SINADEF, by age groups for the year 2020 in the period from May to August in the departments of Peru without including Lima and Callao.

Age group	% 31 - may	% 30 - jun	% 31 - jul	% 31 - aug	No Deceased
Boy (0-14 years)	2.00%	2.20%	1.90%	2.00%	196
Teen - Young (15-44 years)	7.10%	7.40%	7.10%	6.40%	627
Adult (45-64 years)	23.40%	23.40%	23.80%	24.00%	2,333
Elderly (65-74 years)	22.90%	22.20%	21.70%	21.60%	2,101
Old man (85 to more years)	44.60%	44.80%	45.50%	45.90%	4,466

Table 4.1. *Women reported as deceased in the SINADEF*

Here we can see the marked difference between age groups in terms of deaths reported in the departments, where there has been a greater number of people over 85 years of age who have died, maintaining a percentage value of 44% - 45 % constantly, which is equivalent to 4,466 reported elderly women, followed by adults and older adults who together add up to a percentage similar to the elderly, a constant value of 45% is also established. To learn a little about what happened to the women, we will detail a summary of the bulletin published by the ombudsman.

In May, the compulsory social isolation measures were extended until June 30, after Supreme Decree N094 was issued, in the face of a panorama that the Government has called "new social coexistence", with the reopening of some sectors of production. And new schedules for total immobilization. In this context, the Ombudsman's Office once again warns that violence against girls, adolescents and adult women persists as a latent problem, although it continues to be made invisible by restrictive measures that hinder the They denounce and reduce the attention span of the members of the National Specialized Justice System.

A Despite the fact that these institutions have been implementing Legislative Decree 1470, which represents an important advance to reinforce the protection of victims of gender-based violence, there are still a large number of cases of adult women and minors reported as missing, in addition to femicides, completed and attempted.

Adult women reported missing In May, 76 adult women were reported missing. That is, 18 more than last month (an increase of 31%). This change in the figure occurs in a period of state of emergency similar to the previous one, although marked by new schedules for total immobilization and with the progressive adaptation to the D.L. 1470 of the complaints service. For example, the norm establishes priority attention to any form of violence against women regardless of risk and electronic channels, such as emails and WhatsApp lines to present complaints and dictate protection measures.

Against this background, it is worrying that, in most regions, the number of cases registered in their territory has increased. Lima continues to be the region with the most complaints and, this month, I presented an increase from 33 to 39 reports (18% more). Lambayeque follows, where last month no woman was reported as missing, but this month there are 7 cases. Finally, Arequipa and Junín present the same number as in April (4 each).

Likewise, in Piura and La Libertad, the number of cases tripled and in Huánuco it doubled. Additionally, regions such as Ancash, Ica, Moquegua, Ucayali and Madre de Dios, which did not register complaints in April, now appear with one each.

Girls and adolescents reported as missing In May, 207 minors were reported missing, of which 158 were girls and adolescents (76%). This number of women reflects an increase with respect to April (112) of more than 40%, which means that not only is there a greater impact due to the combination of gender and age, but that, in general, the problem has disappeared accentuating during the current health emergency and worrying about not knowing what happened to them.

It should be noted that the increase in cases has occurred in almost all regions, except Lima. However, the latter continues to be the department where the most complaints were filed (56), followed by Arequipa (15), Lambayeque (11), Callao (10) and Piura (8). These last two, in April, were also among the ones that reported the most cases. As for the increase in complaints about the disappearance of underage women, Ayacucho is concerned, where it went from 0 to 6 cases, and Arequipa (15), which increased its reports five times compared to April. For its part, Lambayeque almost quadrupled its figures, and Cajamarca and Madre de Dios went from 0 to 3 registered complaints.

Femicides and attempts During May, the Ombudsman's Office registered 9 femicides, one less than in April. Despite this slight reduction, it is worrying that five regions where there were no cases in April, now the occurrence of more than half of the total in May is reported: Tumbes with 2, and Cajamarca, Puno, Huancavelica and Piura with 1, respectively.

Restrictive measures have confined women to their homes, forcing them to live with their aggressor, which places them in a situation of vulnerability in the face of forms of violence that can lead to femicide. As there are no adequate care channels and the service is not declared essential in this pandemic, women are limited in their ability to report, which makes the real magnitude of the problem invisible. If in ordinary situations, violence against women practiced within a home presents high hidden figures, in this context, these are even higher.

The report of the Crime Observatory detailed that the largest number of victims occurred in the interior of the country, with three in the southern Andean region of Ayacucho, and another two, in each case, in the southern Arequipa, the northern Tumbes and the jungle. central country. Two cases also occurred in the Lima port of Callao and two in the South Lima area. One case per region was reported in the central Huánuco, the northern La Libertad, the Amazon Loreto, and in the southern Puno and Tacna. Thirteen victims were mothers and eight were between 25 and 34 years of age.

In 16 of the cases they were "intimate femicides", as researchers call those caused by aggressors linked to the victims, such as the spouse or ex-spouse, the husband or lover, the sentimental partner, as well as an acquaintance or a neighbor.

Deaths Reported Men

In the following table we can see the deceased men reported in the SINADEF, by age groups for the year 2020 in the period from May to August in the departments of Peru without including Lima and Callao.

Age group	% 31 - may	% 30 - jun	% 31 - jul	% 31 - aug	No Deceased
Boy (0-14 years)	1.40%	1.50%	1.40%	1.30%	214
Teen - Young (15-44 years) Adult (45-64 years)	8.20%	8.10%	8.00%	7.90%	1,345
Elderly (65-74 years)	29.80%	30.60%	30.00%	29.60%	5,029
Old man (85 to more years)	24.70%	23.50%	24.20%	24.40%	4,153
	35.90%	36.30%	36.50%	36.90%	6,274

Table 4.2. *Men reported as deceased in the SINADEF*

In the case of men, it is also evident that the elderly older than 85% have presented the highest percentage value for the table of reported deaths, followed by older adults and then adults; Unlike women, it can be observed that there is a lower percentage in terms of the elderly and the elderly, but this percentage value increases for the adult range, which has a constant value very close to 30%. In addition, the marked difference in the number of male deaths compared to women can be seen.

In order to know a little more in detail about what happened to men during the pandemic, the report published by ANDINA (Peruvian News Agency) will be detailed, who mention the following:

Covid-19 affects more men than women worldwide. According to reports from the World Health Organization (WHO), 60% of the deceased are men and 90% of them are over 60 years of age.

In Peru, 71.7% of deaths correspond to men and 28.3% to women, according to the Covid-19 Situation Room, a website through which the Ministry of Health (MINSA) details the daily evolution of the disease. As the head of the National Institute of Health (INS), Cesar Cabezas, explained to the Andean Agency, men have a genetic predisposition to contract certain diseases such as covid-19.

"Men are biologically different from women and therefore their immune response is also different. For that reason they suffer from severe forms of covid-19".

Diseases such as hepatitis B, liver cancer or cirrhosis tend to occur more in men and not so much in women, because they are protected by estrogens (female hormones). However, Cabezas warned, the explanation can also be multifactorial. Men are not culturally prepared to avoid putting their families at risk and also many suffer from pre-existing diseases such as hypertension, diabetes or heart conditions, in any case, the immune system between men and women is different, because, biologically, they are different, and in the case of women "being a mother makes a difference, since it makes her a guarantor of the human species".

"If women were weaker or died more frequently, the human species would not survive, its existence would be at risk"

Heinner Guio, physician and principal investigator of the INS, detailed that on the cell surface of the lungs, kidneys and liver there are receptors (proteins) that facilitate the entry of the spicules or spines (species of hooks) of the coronavirus covid-19. All the information of the human being is found in its 23 pairs of chromosomes. Pair number 23 is the one that contains the sex chromosomes X or Y: XY in the case of men and XX in the case of women, one of the receptors for this virus to enter the cell surface is called ACE-2 and It is produced by a gene of the same name that is found on the X chromosome.

"This being the case, it could be deduced that women (being XX) would have more ACE-2 and, therefore, a higher risk of infection or death from COVID-19 coronavirus. But in genetics there is imprinting, a phenomenon whereby a gene eliminates its copy. And that would happen with women, that is, the gene kills its copy; but in men (being XY) the virus can continue to live and harm the body".

If the ACE-2 receptor (which is like a lock and the spines of the coronavirus covid-19 the keys) is present in large quantities in the lungs of men, then it will trigger an exacerbated reaction of the immune system. The same happens in older adults.

However, people can suffer from genetic variation and the ACE-2 receptor can vary as well. This change would prevent the coronavirus spicules from "hooking" on it and, as a result, from being kept alive by a natural immune response.

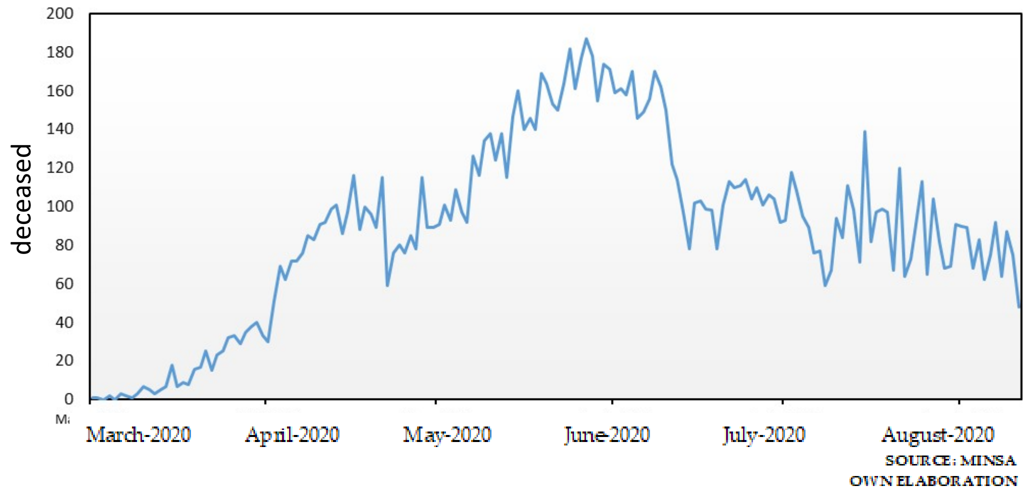


Figure 4.9. *Graph of deceased cases reported daily by COVID-19 in Peru*

In graph 4.9, you can see the total number of deaths from COVID 19 reported by MINSA, in the period March to August 2020 for all departments of Peru, not including Lima and Callao.

In this series it can be observed that for the months of May to June the highest number of deaths occurred; According to the newspaper Gestion, it mentions that the National Information System of Deaths (SINADEF) reported that last May 23,351 deaths were registered compared to 8,937 deaths in the same month of 2019. That is, 14 thousand more deaths were reported than last year.

According to the information on deaths in May 2019, 53.7% were male and 46.3% female, respectively. Meanwhile, 220 were classified for violent death. Compared to the same month of 2020, the figures show 63.1% (men) and 36.9% (women) and only 54 were classified as violent death.

Regarding the increase in deaths, Eduardo Gotuzzo, infectious disease physician and former director of the Alexander Von Humboldt Institute of Tropical Medicine of the Universidad Peruana Cayetano Heredia (UPCH) pointed out that there is a number of deaths that are not detailed and that it could be due to under-registration deaths amid the pandemic. However, he added that the deaths of patients with diseases such as cancer, diabetes, heart attack, stroke, hypertension, tuberculosis and AIDS, who did not receive treatment because the health system is now focused on care, can also be taken into account. of patients with coronavirus.

"Throughout the world there is a sub-registry that estimates a double or third of the cases. In my opinion, many of those people who have died at home could be that they had a heart attack and could not go to the hospital, that they had a cardiovascular complication or an accident because

that can happen on any day. Today people do not go to hospitals for fear of contagion or because there is no space and that is a bit what the figure says"

"In Peru, there is obviously an under-registration, and the probable number of deaths will always have to be evaluated based on suspecting that there is a significant number of patients who are registered by death, but are not registered by COVID-19"

For his part, Jose Incio, a specialist in comparative politics and research methodologies, considers that there is a very important difference in deaths with respect to the figures for 2019 and 2020. However, he pointed out that this situation does not occur in all regions.

"There are regions such as Loreto, Ucayali, Piura or Ancash that have this gap, but there are regions where the number of deaths in 2020 is very similar to the number in May 2019 or is even the same"

It is necessary to search, analyze, investigate and work better with the data to have a better answer. As mentioned above, it can be closely related to a kind of under-registration of deaths from COVID-19 and this warrants a descriptive analysis of certain ratios that allow a much clearer picture of what is happening with the deceased; For example, if the death data by domicile is analyzed, it can be much higher. So, they are dying at home and it is not being classified as such. In addition, the cases of people who need medical attention to treat pre-existing diseases and because the system is focused on the pandemic cannot be treated for fear of being infected, causing them to have a greater probability of dying.

Comparison of departments according to deaths from COVID 19

As for contagions, in order to graph the total number of deaths as of August 31, the data "normalization" was previously carried out, in order to accurately consider the precision of the death rate per thousand inhabitants for each one. of the departments. The figure represents a heat map of the number of deaths accumulated until August 31 by departments, for this the accumulated data of deceased cases (grouped by department) was used, but instead of using the raw cases, the cases per 1,000 inhabitants, the results obtained show that departments such as Ucayali, Madre de Dios, Lambayeque and Tumbes have the greatest impact due to the COVID-19 pandemic. Departments such as Lambayeque, La Libertad, Ancash, Ica are at a high point of death while Loreto, Ucayali, Madre de Dios, Arequipa are in the midst of deaths, therefore measures must be taken for these regions so that the impact can be controlled in a timely manner.

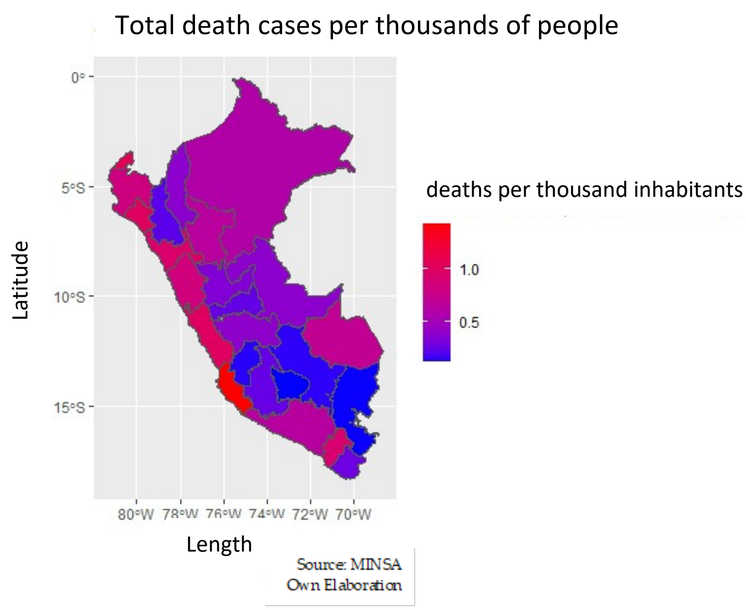


Figure 4.10. *Heat map of accumulated deaths per thousand inhabitants as of August 15, 2020*

Cluster analysis of deceased cases for the departments of Peru

As the pandemic in Peru is known, it progresses differently in each department; therefore, it is important to recognize which departments have the highest death rate and which are controlled. The figure shows 3 groups of departments which are part of a different stage with respect to the advance of COVID-19. The first group contains the departments of Lambayeque, La Libertad, Ancash, Ica and Piura; These regions are the ones that have reached the highest number of deaths with respect to the other regions. In the second group are regions such as Loreto, Arequipa, Ucayali, San Martin and Junín which are in the middle part of the pandemic, but are on the way to reaching the highest number of deaths. Finally, there are the regions of the mountains such as Moquegua, Amazonas, Huánuco, Cajamarca, Cusco, among others, which have low numbers of deaths compared to the maximum numbers reached by other departments. That is why the vitality of the present study, whose purpose is to be able to diagnose the groups with the highest death rate.

4.2.3. Analysis of hospital data

In the country, the professionals who are on the front lines against the coronavirus, face daily making decisions to prioritize the admission of patients to the Intensive Care Unit (ICU), that is why a graph of the availability of beds was made ICU during the pandemic, Figure 21 shows the evolution of ICU bed availability, where the number of available beds is from 100 to 200 per day, remaining constant over time and the beds in use increase day by day, Total number of ICU beds increasing every day means that new ICU beds are being purchased daily throughout the country.

UCI beds in the Essalud and Private sectors

It can be seen in figure 23 that the number of ICU beds for each month has been increasing for both ESSALUD and the private sector, however Essalud has had a much greater increase each month in contrast to the private sector.

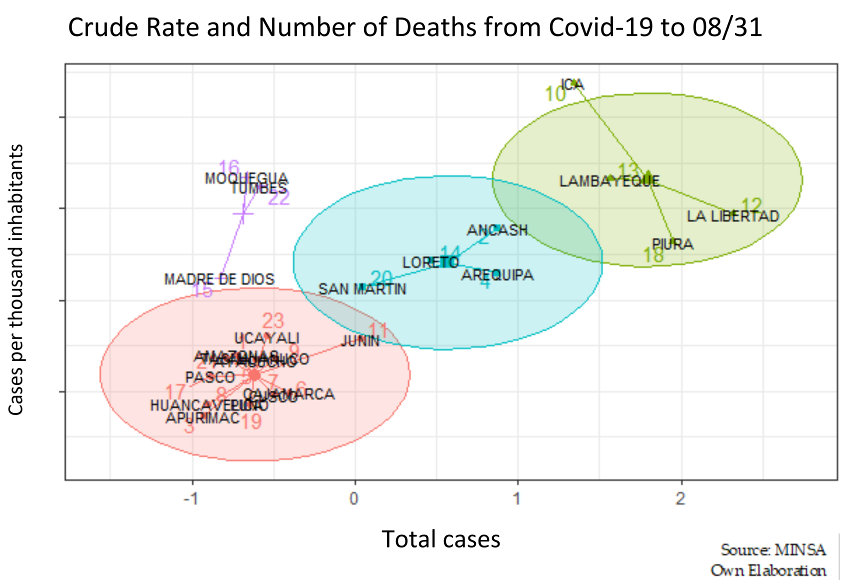


Figure 4.11. *Cluster of death cases for the departments of Peru as of August 31.*

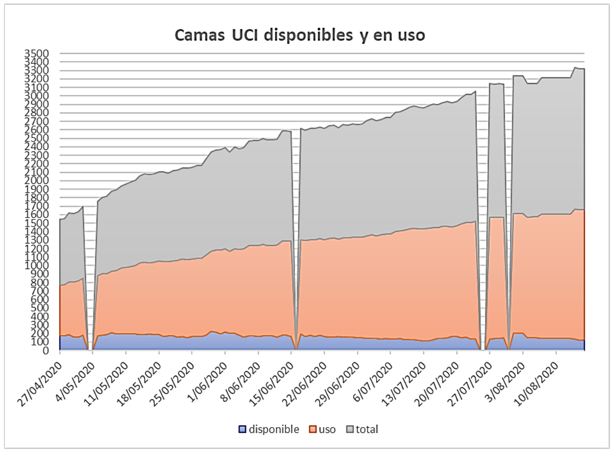


Figure 4.12. *Evolution of the number of ICU beds available and in use*

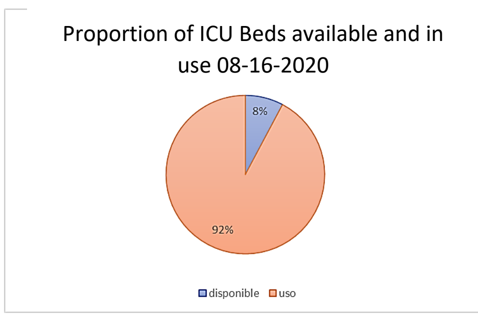


Figure 4.13. *Pie chart of the number of ICU beds available and in use as of August 16*

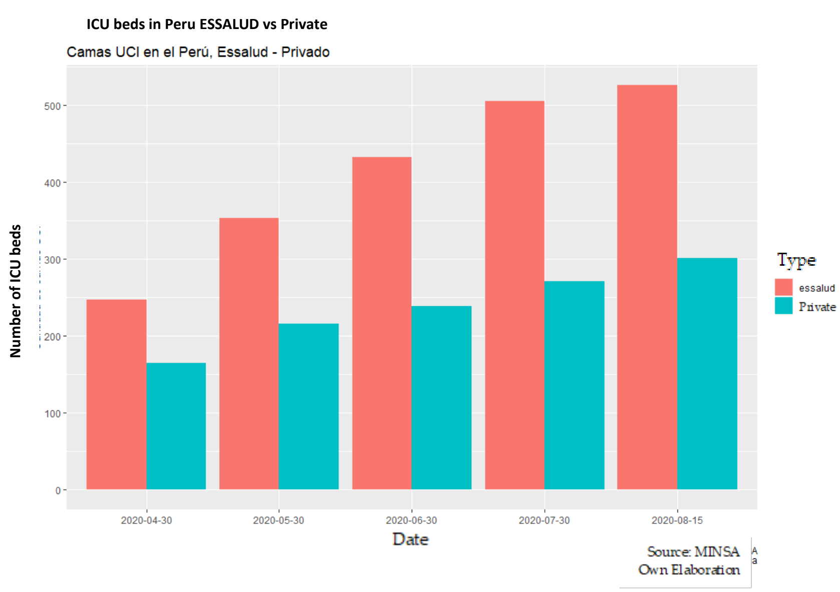


Figure 4.14. Number of ICU beds in ESSALUD and the private sector for each end of the month.

4.3. Estimation of the effective reproduction number (R_t)

The R_t, or effective reproduction number, is the average number of infections caused by each infected person at a given time, taking into account changes in the behavior of the population (quarantine, use of masks, teleworking, etc.). An R_t of 3-4 will infect virtually the entire population, while an R_t of 1.5 can still reach 60% of the population. Only if the Rt is less than 1, the epidemic will decrease in size until it is eliminated.

For the estimation of the R_t, the data provided by the MINSA for positive cases were used, which are published daily, in this case the data used for the estimation was from March 10 to August 31 for all departments of Peru. not including Metropolitan Lima and Callao. All the data processing and estimation procedure were carried out in Python.

4.3.1. Data Treatment

The first step was to group the data by date and region and then make a filter excluding the Callao and Metropolitan Lima region. This new database is a series of the number of new daily cases by region. Observing this new series of daily cases, the analysis should be started when there is a constant number of cases each day, because if there are zero cases in a day, this upsets the model (due to division by zero) and makes it start from new. Therefore, the last day zero of new cases was found and the construction of the model began the following day.

In addition, the notification of cases is very erratic depending on pending tests, inconsistency in the number of tests carried out on different days of the week, etc. A first step in working with the data will be to normalize them based on the number of tests taken each day to compensate for the variability in the latter and apply a moving average of seven days to deal with weekly seasonal effects and mitigate the impact of changes as much as possible. in the sampling criteria (focus on markets or public transport stations, for example). In order to remove the random component that "dirties" the data, after compensating for tests and averaging over the last 7 days, we applied a centered moving window Gaussian filter, with the window size equal to 7 days (one calendar week).

This and the following steps were applied to a selected department, making the results easier to see and understand. Arequipa department has been selected.

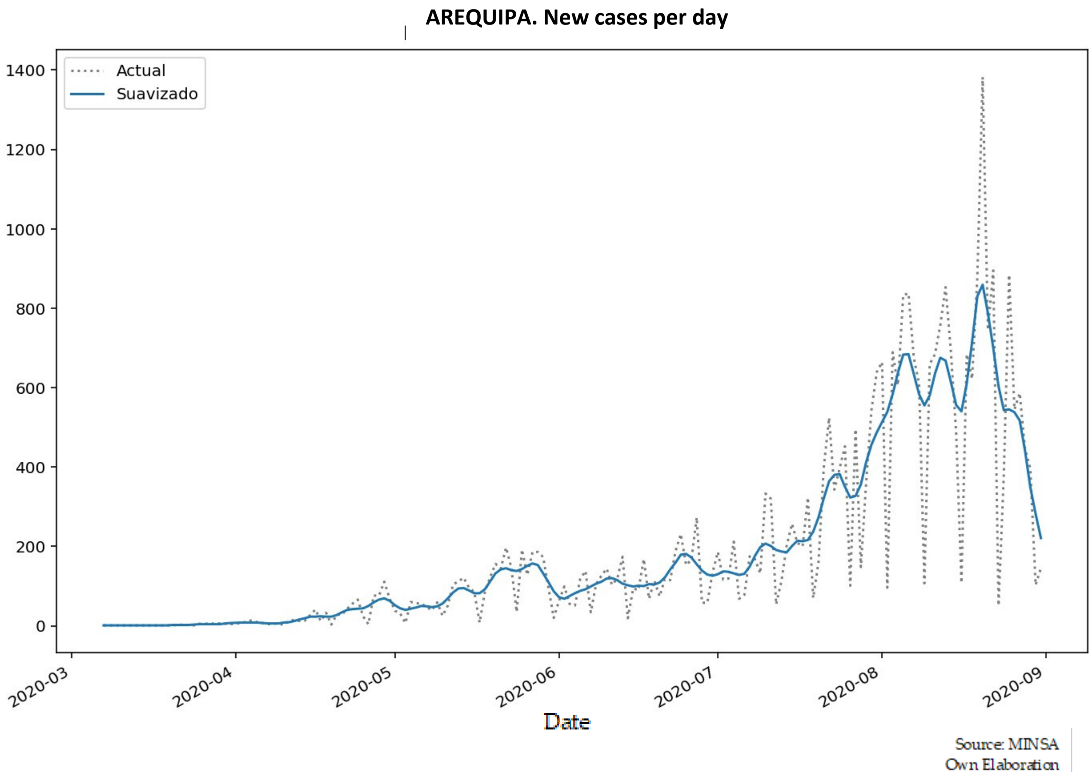


Figure 4.12. *Graph of the daily series of cases confirmed by COVID-19 in Arequipa and its smoothing.*

4.3.2. Posterior probability function P (k_t|R_t)

The original approach simply selects the later data from yesterday as the previous data from today. While intuitive, doing so does not allow our belief that the value of R_t has probably changed since yesterday. To allow for this change, a Gaussian noise was applied to the above distribution with some standard deviation σ. The higher σ, the more noise and the more we expect the value of R_t to deviate every day. Interestingly, applying noise over noise iteratively means that there will be a natural decay of the distant last. This approach has a similar windowed effect but is more robust and doesn't arbitrarily forget later ones after a certain time like my previous approach. Specifically, the windowing system computed a fixed R_t at each time t that explained the w days of surrounding cases, while the new approach computes a series of R_t values that explain all the cases, assuming that Rt fluctuates approximately σ each day.

However, there is still an arbitrary choice, Adam Lerer pointed out that we can use the maximum probability process to inform our choice. Is that how it works:

The maximum probability says that we would like to choose a σ that maximizes the probability of seeing our data k: P (k σ). Since σ is a fixed value, was left out of the notation, so I tried to maximize P (k) in all options σ.

Given the $P(k) = P(k_0, k_1, \ldots, k_t) = P(k_0)P(k_1)\ldots P(k_t)$) you need to define P (kt). It turns out that this is the denominator of Bayes's rule:

$$P(R_t|k_t) = \frac{P(k_t|R_t)P(R_t)}{P(k_t)}$$

To calculate it, we note that the numerator is actually just the joint distribution of k and R:

$$P(k_t, R_t) = P(k_t|R_t)P(R_t)$$

The distribution over R_t can be marginalized to obtain $P(k_t)$:

$$P(k_t) = \sum_{R_t} P(k_t|R_t)P(R_t)$$

Then, if we add the distribution of the numerator over all the values of R_t, we obtain P (kt). And since that is being calculated anyway while the later is being calculated, it will be followed separately.

Since we are looking for the value of σ that maximizes P (k) in general, we actually want to maximize:

$$P(R_t|k_t) \propto \exp\Big(\sum_{t=T-m}^{T} \log(\mathcal{L}(k_t|R_t))\Big)$$

where t is all times and i is each department

Since you are multiplying many small probabilities, it may be easier (and less error prone) to take the log of the values and add them together. remember that log ab = log a + log b. And since the logarithms increase monotonically, maximizing the log sum of the probabilities is the same as maximizing the product of the non-logarithmic probabilities for any choice of σ.

Function to calculate the posterior to calculate the posterior we follow these steps:

1. λ - is calculated the expected arrival rate for the Poisson process for each day.
2. Then the probability distribution of each day is calculated over all possible values of Rt
3. Then the matrix of the process as a function of the value of σ that was discussed previously.
4. The initial price is calculated because the first day does not have a previous day from which to take the later one.
5. Based on the information from the epidemiological center, a Gamma was chosen with a mean of 7.
6. It repeats from day 1 to the end, doing the following:
    - Calculate the previous one by applying the Gaussian to the previous one from yesterday.
    - Apply Bayes' rule by multiplying this a priori and the probability that we calculated in step.
    - Divide by the probability of the data (also Bayes rule)

Then, in Figure 4.13 you can see all the days of the posterior distribution graphed simultaneously. The later ones start without much confidence (wider) and become progressively more certain (narrower) about the true value of Rt

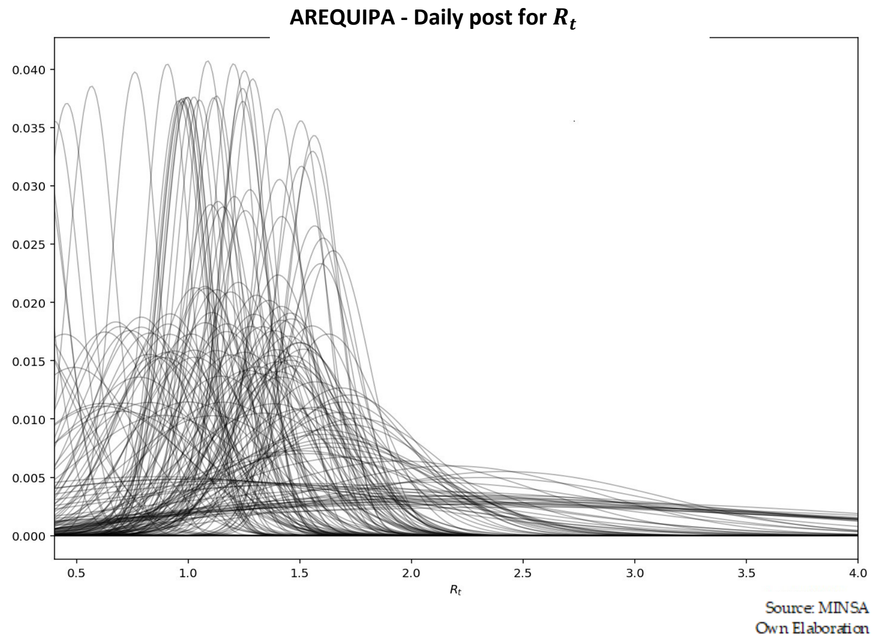


Figure 4.13. *Graph of the calculation of the day after to evaluate the value of R_t.*

4.3.3. Estimate R_t

The final step is to estimate the R_t. values and the higher density ranges that surround them. We need to simulate random values for R_t. using the posterior probabilities to calculate the highest density intervals.

Since our results include uncertainty, we would like to be able to see the most probable value of R_t along with its highest density interval.

Date	Most probable value	Low I.	High I.
7/03/2020	2	0.5	5.54
11/03/2020	1.96	0.5	5.09
19/03/2020	1.92	0.51	4.8
20/03/2020	2.14	0.59	5.02
...	...	...	...
29/08/2020	0.46	0.27	0.65
30/08/2020	0.54	0.36	0.76
31/08/2020	0.59	0.4	0.81

Finally, both the most probable values for R_t and the maximum density intervals over time can be plotted. This is the most useful representation since it shows how our beliefs change every day.

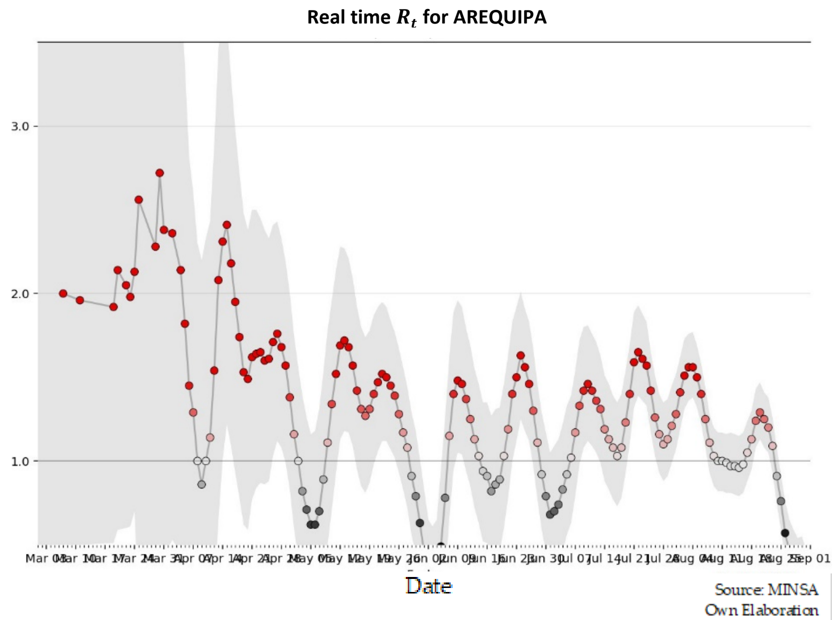


Figure 4.14. *Graph of the real estimate of Rt for Arequipa.*

The graph in Figure 4.14 shows two important things. First, the "most probable value" of R_t for each day, represented by the dots (the redder a dot, the higher and more dangerous the R_t value). Second, the range of values that R_t could actually be is represented by the gray bands (if there is a 'most probable value', then there are also other less probable but possible values).

4.3.4. Choose the optimal σ

The Python model will be made interactively for all departments, therefore a σ must be chosen for each of them interactively.

In previous lines the choice of an optimal σ was described, but it was assumed assuming a value for the department worked. Now you can evaluate each state with any sigma, you have the tools to choose the optimal σ which is to maximize the probability of the data P (k). Since we do not want to overfit in any state, we choose the sigma that maximizes P (k) in each state. To do this, all the record probabilities by state are summed for each sigma value and then the maximum is chosen.

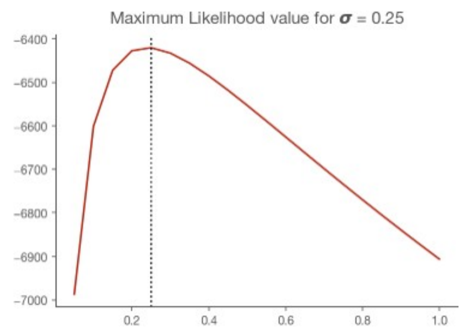


Figure 4.15. *Graph of the optimal σ value.*

Chapter 5

Results

5.1. Regional Analysis

5.1.1. Basic Reproductive Number R_0 and Effective Reproductive Number Rt

Currently we are working with Rt as an epidemiological indicator under a Bayesian approach. The data with which we work suffer from irregularities for different reasons (delayed tests, inconsistency in the number of tests that are carried out on the different days of the week), so to work them a normalization has been carried out according to the number of tests and a seven-day moving average has been applied to deal with weekly seasonal effects. The measures established by the Peruvian government are detailed below and, after that, the results obtained for the first objective set will be shown.

National State of Emergency Extended The government of Peru extended the current state of emergency for an additional 30 days starting on Tuesday, September 1. Children under the age of 14 will no longer be allowed outside for any length of time, and adults 65 and older must stay indoors except when absolutely necessary. Social gatherings of any kind are not allowed, even at home with the family. The curfew from 10:00 pm to 4:00 am from Monday to Saturday and the all-day quarantine on Sunday remain in force in most of Peru, with the exception of the following regions of Peru where there is a quarantine and a mandatory daily curfew from 8:00 pm to 4:00 am.

Quarantine areas (as of August 28, 2020): Amazonas (Bagua, Chachapoyas, Condorcanqui and Utcubamba); Ancash (Santa, Casma and Huarmey); Apurimac (Abancay); Arequipa (Cailloma, Camana, Castilla, Islay); Ayacucho (Huamanga, Huanta, Lucanas and Parinacochas); Cajamarca (Cajamarca and Jaen); Cusco (the entire region); Huancavelica (Huancavelica, Angaraes and Tayacaja); Huanuco (Huanuco, Leoncio Prado, Puerto Inca and Humalies); Ica (Ica, Pisco, Nasca and Palpa); Junin (Huancayo, Satipo and Chanchamayo); La Libertad (Trujillo, Viru, Sanchez Carrion, Pacasmayo, Chepen and Ascope); Lima (Barranca, Cañete, Huaura and Huaral); Mother of God (Tambopata); Moquegua (the entire region); Pasco (Pasco and Oxapampa); Puno (the entire region); and Tacna (the entire region).

National Health Measures extended for 90 days: On August 28, Peruvian President Martin Vizcarra signed a supreme decree that extended Peru's State of Sanitary Emergency for 90 days as of Tuesday, September 8. Social distancing and the use of masks will be necessary for the foreseeable future. The Ministry of Health of Peru (MINSA), the National Institute of Health and ESSALUD the Social Health Security continue to develop and implement an action plan for the surveillance, containment and attention of new cases of COVID-19 in Peru. The government of Peru will continue to purchase the goods and services necessary to combat the spread of COVID-19.

Classification of departments based on the research carried out: A Cluster analysis was carried out to be able to identify the departments with similar behavior according to the data provided by the MINSA, the grouping was carried out based on the predicted values of Rt, taking the mean value and the maximum reached in the confidence interval of each department as of August 31.

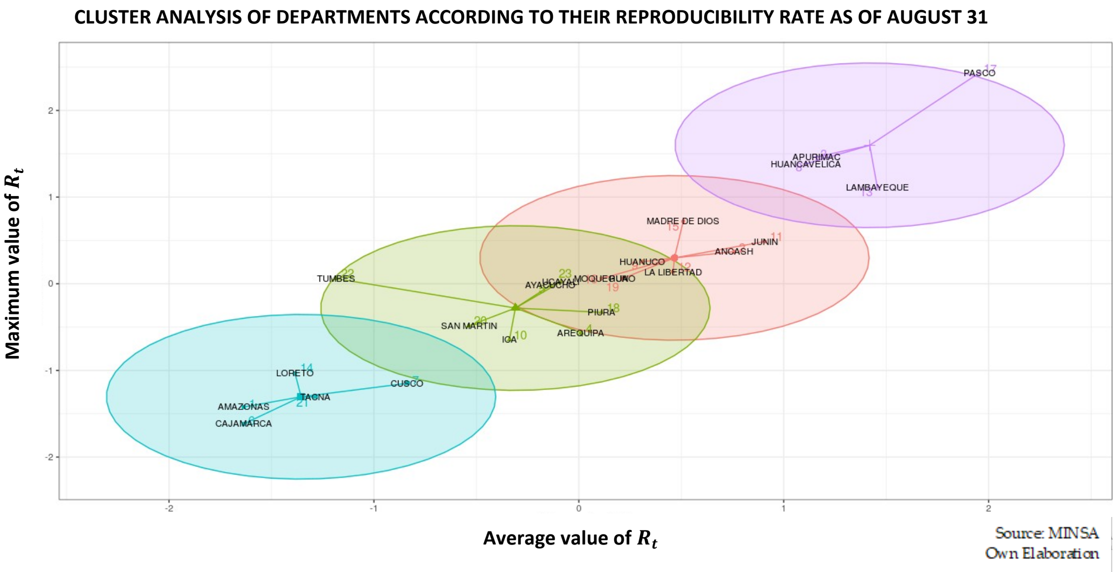


Figure 5.1. Cluster analysis of the departments based on their last estimated R_t as of August 31.

According to figure 5.1, 4 groups are observed, so we proceeded to identify the departments that belong to each group in the following table.

Group 1	Group 2	Group 3	Group 4
AMAZONAS	TUMBES	MOQUEGUA	APURIMAC
LORETO	SAN MARTIN	PUNO	HUANCAVELICA
TACNA	ICA	LA LIBERTAD	LAMBAYEQUE
CUSCO	AREQUIPA	HUANUCO	PASCO
CAJAMARCA	PIURA	MADRES DE DIOS	
	AYACUCHO	ANCASH	
	UCAYALI	JUNIN	

Table 5.1. *Classification of departments using Cluster analysis.*

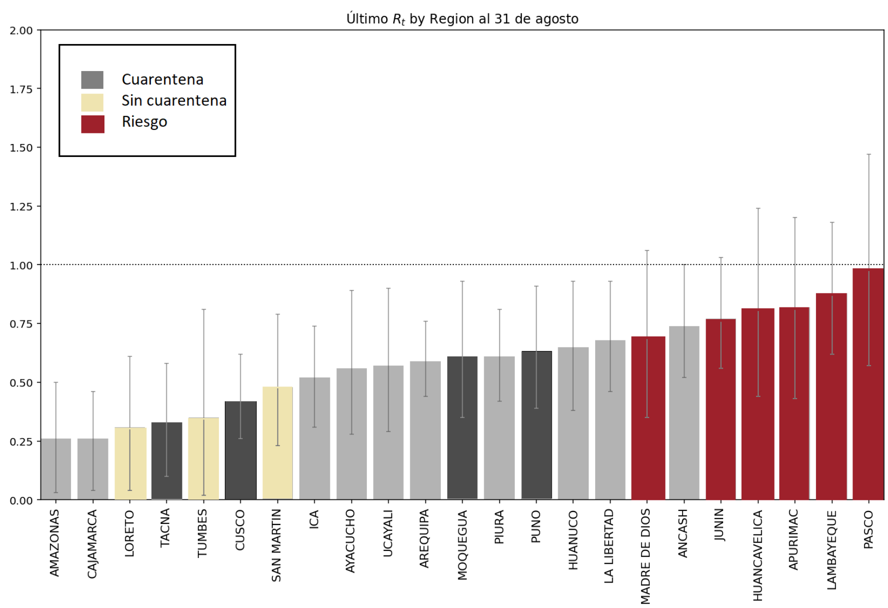
In order to determine intragroup similarity and dissimilarity between groups, we applied a non-parametric hypothesis test as follows:

- H_0 : The medians of the groups are all the same.
- H_a : At least one of the group medians is different from the others.

Accepting the hypothesis will allow us to affirm that the groups have the same median, which would imply a similar behavior within each group under the assumption of independence between groups. For this we will use the Kruskal Wallis test, with a value of 0.001, the χ_c^2 value calculated from the test was 15.48, so it would be concluded that at a significance level of 0.1%, it is not rejected that the medians of the groups are equal.

Based on the grouping, group 4 is defined as the group that as of August 31 presents a maximum value of R_t=1 so it would be recommendations for the departments of Apurimac, Huancavelica, Lambayeque and Pasco, which have not yet the measures are released since the departments are most prone to contracting the greatest number of infections.

Below is the last calculated reproducibility rate (R_t) estimated in real time for the departments of Peru until August 31st.



Fuente: MINSA
Elaboración propia

Figure 5.2. *R_t in real time estimated for all departments as of August 31.*

From what can be classified Junin, Huancavelica, Ayacucho, Lambayeque and Pasco, as the departments most prone to the outbreak of the pandemic, it would be necessary not to neglect the measures applied in those departments, such as curfew, mandatory social immobilization, among other. In order to dissipate contact between individuals and convert the currently calculated rate to zero.

For the other departments such as Madre de Dios, La Libertad, Huánuco, Puno, etc.; It is also necessary to continue with the measures, but it could be somewhat less strict than the provisions previously established.

Through this graph, you can see the difference between the measures implemented by the government to have a total quarantine in departments such as Moquegua, Tacna, Cusco and Puno, unlike the results obtained; It would be important to consider that there may be these inconsistencies, due to the limitation of accessing higher indicators in the population.

In order to reveal the details of the classification, we proceed to explain the diagnosis made to each department of Peru without including Lima and Callao, where the consistency of the investigation will be known a little more thoroughly.

Amazonas Region

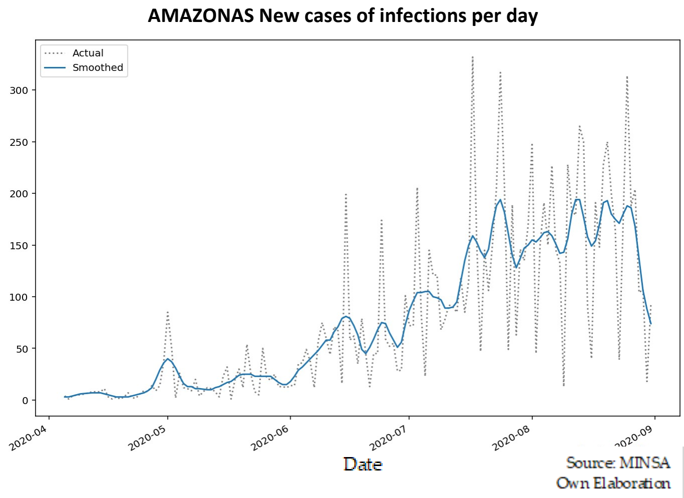


Figure 5.3. *Contagions confirmed in Amazon as of August 31.*

The spread of the coronavirus throughout the Peruvian territory was confirmed with different cases in some of the most remote communities of the Peruvian Amazon, located on the border with Brazil, an area without migration control.

By March, the indigenous people of the Peruvian Amazon declared themselves on alert before the expansion in the country of the coronavirus after the first cases were registered in the cities of the jungle part of Peru, which occupies 61% of the national territory.

The Central Ashaninka del Rıo Ene (CARE), which represents 18 native communities, announced on March 19 that it was declaring itself on maximum alert that it would not allow any foreigner to enter its space, under threat of being detained by its self-defense committees and expelled or handed over to the National Police or the armed forces.

There are at least six cases in Alto Purus of indigenous Sharanahuas who came to this area from the Brazilian side of the border. A similar situation occurs somewhat further north in the Alto Tamaya-Saweto native community of the Asheninka ethnic group, where there were also six cases in late June.

For this reason, in the first weeks of July there were several mobilizations of indigenous people in cities of the Amazon to protest the lack of health care for native populations since the emergency began, because until now the number of infected indigenous people and deaths from COVID-19 nationwide.

As of August, the COVID-19 situation room in Amazonas reported 21,199 autochthonous cases and 711 foreign cases, which would be justified mainly because the departments do not have control over the registration of samples to avoid duplicate cases.

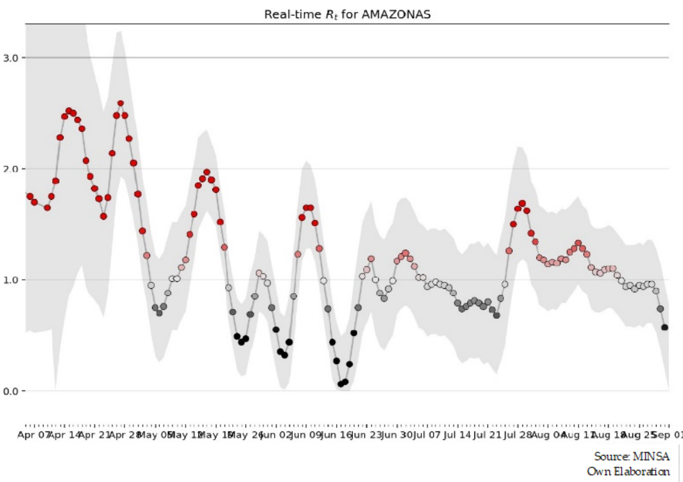


***Figure 5.4**. R_t real estimated for the department of Amazonas as of August 31.*

Estimating the R_t in real time, it is observed that in the first months after the pandemic began, the infection rate ranged from 2 to 3, that is, 1 person could infect 2 to 3 people, after which greater measures were implemented in terms of to social immobilization and quarantine, in addition the Amazon was in charge of closing the borders with other countries to avoid contagion and loss of the tribes of the department, therefore it can be observed that in the intermediate part in the period from May to July the value of the reproducibility rate fell enormously, below 1, after that it is known that the Peruvian government allowed domestic flights, in addition the economic crisis merits the struggle to support a family, allowing to generate income, that is why the rate is I raised it to close to 2, but the measures regarding focused quarantine were restored, which has allowed greater control for the department of Amazonas..

Ancash Region

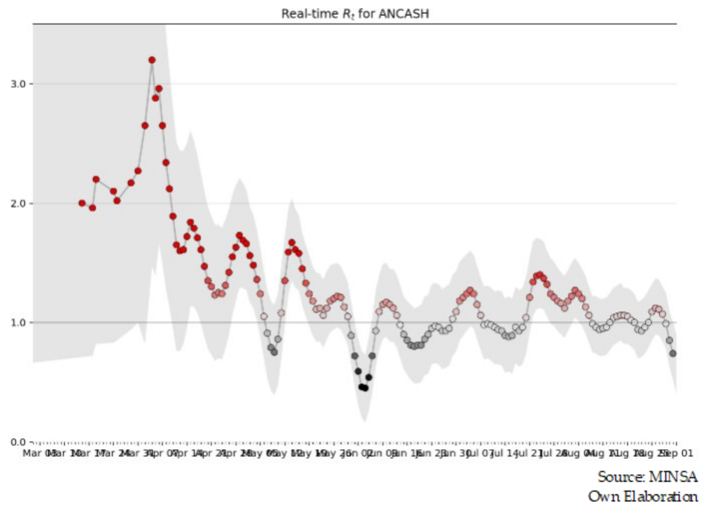


Figure 5.5. *R_t real estimated for the department of Ancash as of August 31.*

Evolution of R_t in the department of Ancash for the department of Ancash, a growing Rt is seen during the month of March, this may be due to the fact that initially no preventive measures were taken, on March 15 the quarantine began at the national level, consequently, a decline significant Rt from April 1, this behavior may be due to the fact that an infected person stops carrying the virus after 15 days and by not carrying it, they stop infecting other people. In the graph it is observed that the decline is continuous until the beginning of June, where the Rt tends to oscillate to the value of 1, this means that a person with COVID-19 can infect only one, suggests that the measures that this taking the government are giving results to control the pandemic, below are some of these measures of the Regional Government of Ancash (GRA)

- In a joint effort with the Central Government, the modern facilities of the modular hospital at the Rosas Pampa stadium in Huaraz were handed over, where there are currently 120 beds for hospitalization, five beds for intensive care, oxygenation, and triage area. and stations for nursing work.
- In coordination with the Ministry of Health, 73 beds and medical equipment were installed in the new field hospital implemented in the "Ramos Guardia" in Huaraz. The facilities have 26 beds available, where mild and moderate COVID-19 patients will be referred. In addition, multiparameter monitors, oxygen balloons and air conditioning system are available.
- Through the South Pacific Network, 8,000 doses of ivermectin were delivered to health networks and hospitals in the region. These samples were prepared at the Ancash Regional

Laboratory, located in the city of Casma and their distribution is free in all health establishments in Ancash.

- The purchase of ambulances was made for the rapid care of patients with covid 19.

Evolution of confirmed infections in the department of Ancash. In the figure, three behaviors of the series of very marked infections are observed:

- The first one that occurs from the beginning of the pandemic until the end of May where there is continuous growth, this occurs in general for most departments since it is the effect of the pandemic due to the lack of preventive measures at the beginning of the first registered infected that occurred **ANCASH New cases per day**

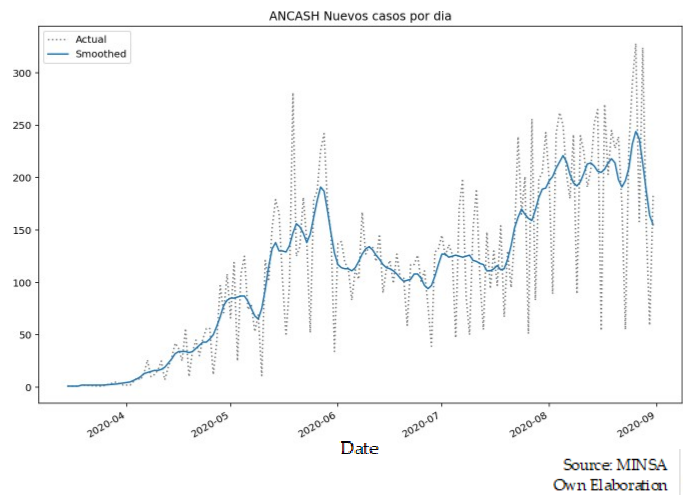


Figure 5.6. *Contagions confirmed in Ancash as of August 31.*

- The second is when the contagion curve begins to decline slightly until it maintains approximately 125 infected per day, this is a consequence of the quarantine that occurred from the fortnight of March to June 26.
- The third behavior is the growth of infections from the end of July, this fact occurred because the quarantine was lifted at the end of June and with a gap of 15 days the infections occurred, the growth is not exponential due to the measures that I take over the regional government of Ancash, making the numbers of infections not soar.

In general, the actions that the Regional government of Ancash is taking is working, the largest investment was made to improve hospitals and implementation with medical teams and ambulances, deployment of rapid response teams (campaigns) and personal protection materials for the first line of battle against COVID-19, it also cared about the training of its health personnel and new telephone services for the attention of suspected cases. These measures could be taken as a model in other regions so that the pandemic can be controlled more effectively.

Apurimac Region

The behavior of the COVID-19 pandemic in Apurimac has had a stable behavior and with very low numbers since the quarantine began in Peru, on average the values of daily cases ranged from 15 to 25 confirmed cases. However, a drastic change can be observed in the curve of daily cases from the fortnight of July, where the slope of the curve undergoes a significant change. To explain this behavior in the curve we will put ourselves in context, Apurimac is a department located in the southern highlands, by applying mandatory immobilization throughout the country, this region was isolated, its first case was reported just in April. But on July 15, interprovincial trips were opened, so it could be said that it was a reason for the increase in daily infections.

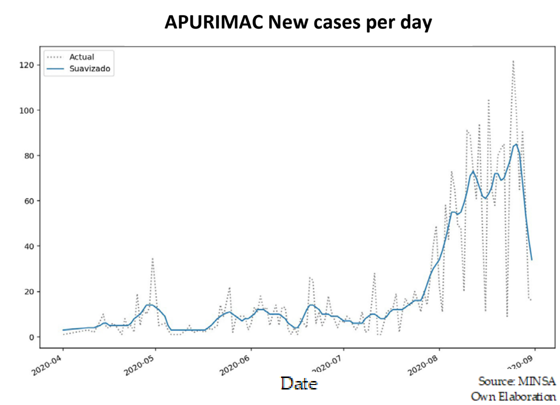


Figure 5.7. *Contagions confirmed in Apurimac as of August 31.*

Now, one way to explain this behavior a bit is by analyzing the real-time estimate of the reproducibility number in Apurimac, the estimates in the first months have a very changing behavior with values below 1 in one day and the next day above 1, in addition, its range is very large, therefore the uncertainty of the model is also high, this due to the small amount of data in those months, which shows that Apurimac in general lines was not as an alert region. However, from the first week of July the Rt was above 1 if it remained so until the end of August, which means that a person could infect one more and so on. The actions taken by the government in this case were late, since it was only in August that Apurimac entered the list of regions with targeted quarantine. And it continues in a focused quarantine, in this case only the province of Abancay, until September 30.

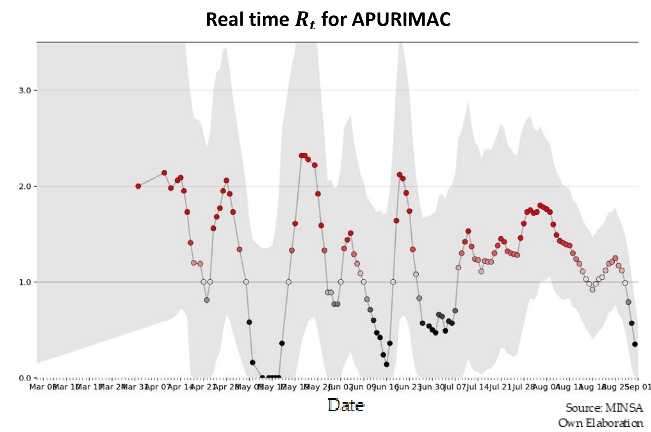


Figure 5.8. *Evolution of R_t in the department of Apurimac*

Arequipa Region

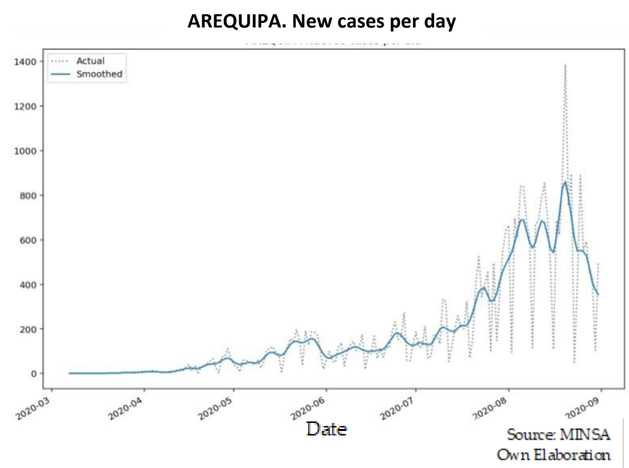


Figure 5.9. *Evolution of R_t in the department of Arequipa*

Evolution of confirmed infections in the department of Arequipa According to the Regional Health Management, and as can be seen in Figure 5.8, the number of cases in Arequipa, the second most populated city in Peru, has been increasing rapidly as of the first week of June. At the end of July and with more than 43 thousand cases and 853 deaths, this region of southern Peru had become another important epicenter of the pandemic in the country, mainly due to its weak health infrastructure, lack of medical personnel, its informal economy, non-compliance with the rules by the population and poor regional management. However, the highest peaks of infections occurred during the month of August; This is why, in the face of such figures, the national government took control of the health system due to the slow local response, beginning by extending the total quarantine in some provinces such as Arequipa, Ica, Junín. Huánuco and San Martin until August 31.

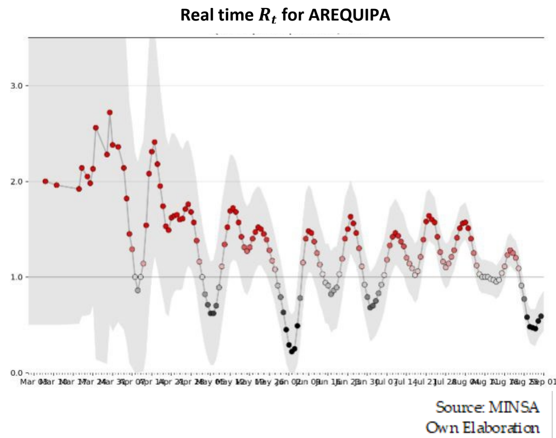


Figure 5.10. *R_t Evolution of the R_t in the department of Arequipa*

Evolution of the R_t in the department of Arequipa. The first case of the coronavirus disease pandemic in Arequipa, started on March 7, 2020. The first case was a 26-year-old man who came from the United Kingdom. At that time, the government still did not take the respective measures to mitigate the spread of the disease, that is why; that the Rt took values above 1, which indicates that a person with the virus can infect on average two other people, which indicates that the disease grows.

However, this reproducibility index has a slow decay until the beginning of July, where the Rt fluctuates around 1 and is not yet controlled between the first week of July and the beginning of August approximately; This is due to the fact that as the months passed and the population was left without economic resources, the residents of this southern Peruvian city began to take to the streets due to the need for work, which in turn has generated contact with other people. "Many did it out of need for work," said Oswaldo Zeballos, dean of the Arequipa College of Economists.

Ayacucho Region

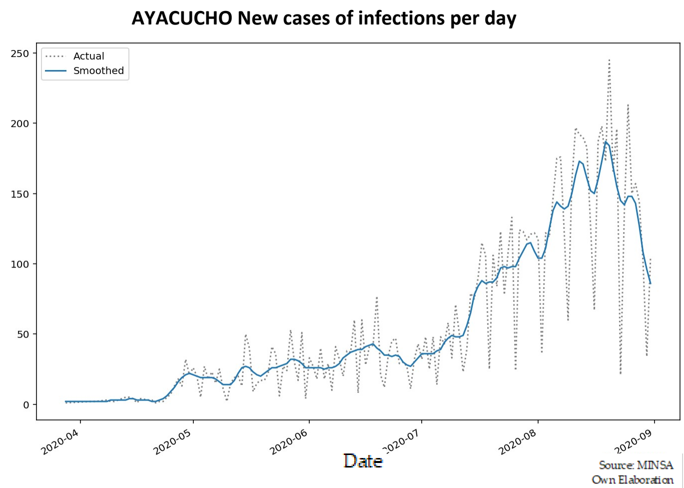


Figure 5.11. *Contagions confirmed in Ayacucho as of August 31*

For the department of Ayacucho, it can be seen that the increase in infections occurred from July to August, in the first months a slight number of infected people is observed and this could be due to the measures implemented by the regional government, to We will see them in detail.

- For the month of March, the regional government of Ayacucho began with awareness campaigns, as was experienced at the national level, giving specifications on the use of masks, hand washing with soap and water, social distancing and #YoMeQuedoEnHasa.
- A week later, the entry points to Ayacucho were closed, with the help of the PNP, the Peruvian army, self-defense committees and neighborhood councils.
- For the month of April, contributing to the inclusive campaign, they promoted awareness videos in Quechua, which allowed them to reach more residents.
- For the month of May, the number of tests carried out was intensified to 4734, which allowed to identify the people who contracted the virus and help with their speedy recovery. be cared for in the nearest health establishments, so medicines were supplied to the health posts, in order to help the patients, recover quickly.
- Another measure applied was the authorization of the operation of Hospital Antigo for the prevention of the spread of COVID-19 even with a lack of instruments for the care of its patients.

These are the main measures that allowed greater control in the COVID 19 contagion curve, for Ayacucho, however for the month of August, which is where its notable increase is seen, new measures were available by the Peruvian government, such as the new restriction on Sundays, the immobilization hours, and the awareness campaigns were intensified, in order to control the pandemic.

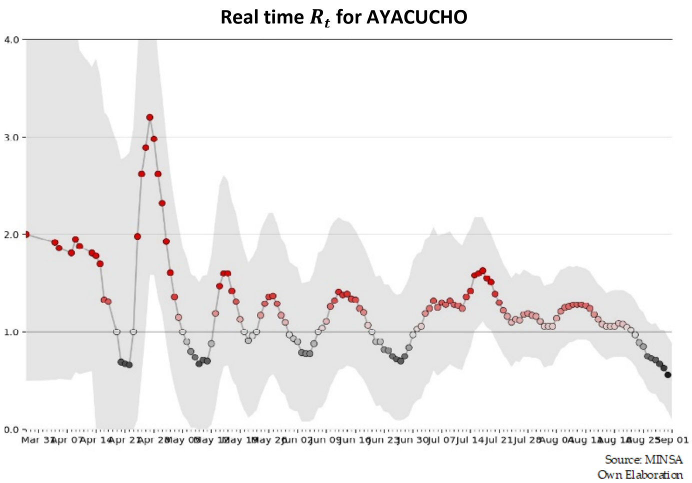


Figure 5.12. *R_t in real time estimated for the department of Ayacucho as of August 31.*

In this graph it can be seen that in the first month, the reproducibility rate Rt, reached a value of 2, which implied that 1 person could infect two others, so it was a considerable contagion rate at the beginning of the pandemic, the regional government was able to control, carrying out care and awareness campaigns, in addition to certain of the entry points to the city of Ayacucho, therefore the value of the rate dropped to a value less than 1; However, for the month of April, the contagion rate increased exponentially again, reaching a value greater than 3, that is, an Ayacucho could infect 3 or more other people, for this reason the supply of medicines to health posts for the prompt care of people with symptoms of COVID 19 as well as pre-existing diseases, in addition the Peruvian government decreed the differentiated exit both by sex and by age; All this added to the fact that the daily contagion rate could be controlled, reaching August 31, with a value less than 1.

Cajamarca Region

In the evolution of infections, it is observed that it is low in the first months, this is due to the low population density of the department of Cajamarca and the intervention of the peasant rounds with the creation of strict fences in small populations in addition to sanctions to the inhabitants They did not apply the norms established through physical exercises, for the month of August and September a growth is observed with a peak for the first of August this is due to the fact that the government allowed the return of the Cajamarca citizens who lived in Lima and by the lack of hospitals in many of the towns. Among the measures taken by the royal government of Cajamarca in the face of this problem were:

- Creation of community shelters so that newcomers will be quarantined for 15 days before entering any locality.
- Implementation of geo-covid to monitor the patient in order to take relevant actions such as rapid tests and health campaigns.
- Alliance with the private company and MINSA to manage 420 clinical beds, type I, II, III y IV.

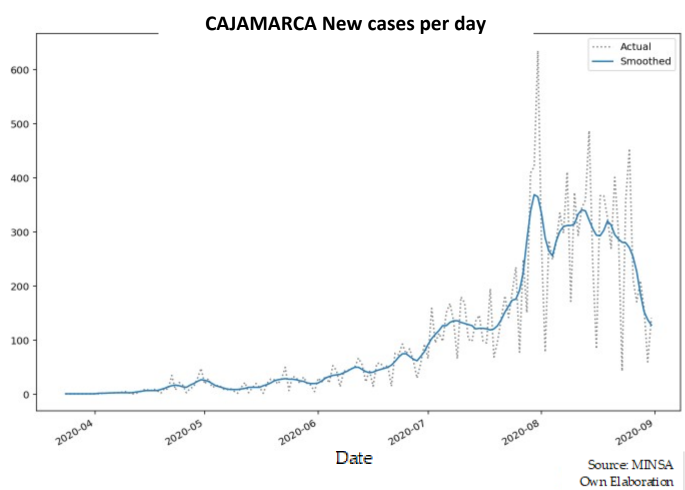


Figure 5.13. Contagions confirmed in Cajamarca as of August 31.

The R_t for the department of Cajamarca presents a high value for the month of May and this continues until the beginning of August where the reproducibility index begins to drop, this indicates that the measures that the government is taking are being effective, giving an R_t for the last date close to zero.

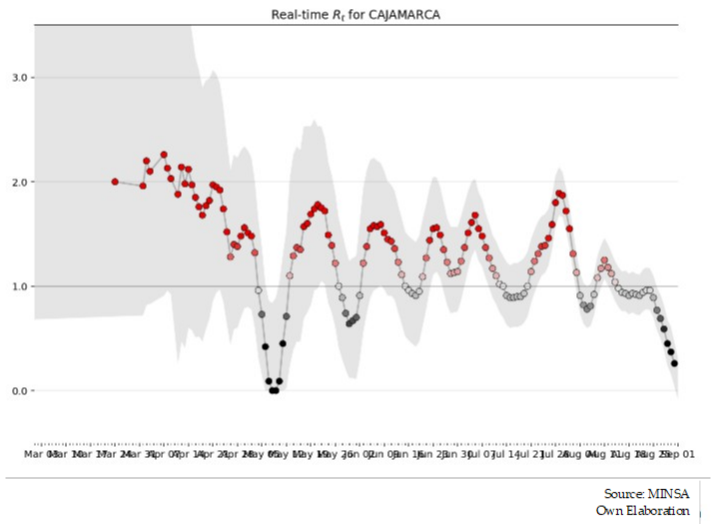


Figure 5.14. *R_t real estimated for the department of Cajamarca as of August 31.*

Cusco Region

The Cusco region, as can be seen in Figure 5.12, has had a very stable behavior and with very low daily confirmed cases compared to other regions. This behavior was maintained only until the third week of the month of July, as already mentioned, Cusco is a city that was completely isolated from cities like Lima, which was the focus of infections. However, when flights and interprovincial transport reopened in Peru as part of the economic reactivation, the Cusco region began to increase its confirmed cases of damage exponentially.

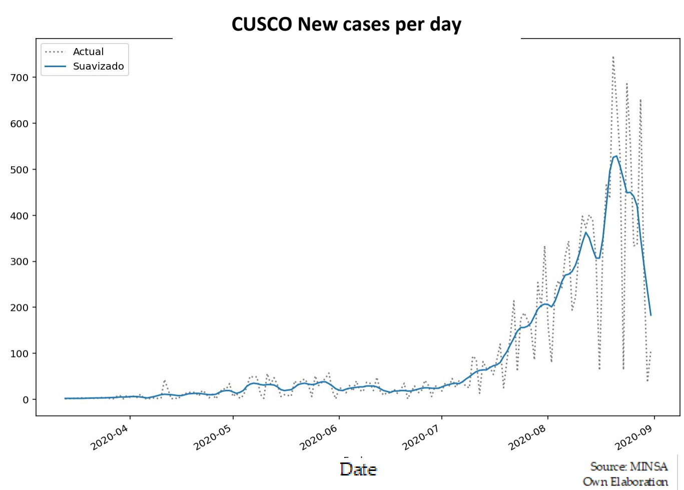


Figure 5.15. *Daily confirmed case series in Cusco as of August 31.*

In Figure 5.13, it can be observed that the Rt is very changeable at the beginning of the quarantine in Peru; However, from the third week of June the R_t remained above 1, that is, a person could rate 1 and that person to another and so on. But, until the second week of July the range of the confidence limits were very large, that is, there was a lot of uncertainty in the model, and from the third week of July that range was decreasing and the R_t was still above 1, that is, Cusco was already on alert. The measures taken for Cusco were late, however the regional government has constantly campaigned, which has generated that the exponential growth is not so drastic.

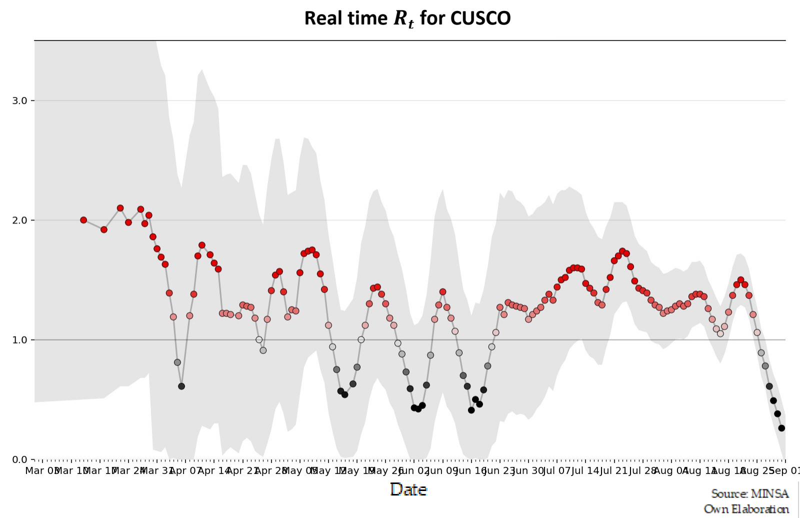


Figure 5.16. Evolution of Rt in the department of Cusco until August 31

Huancavelica Region

Evolution of confirmed infections in the department of Huancavelica As can be seen in Figure 5.16, the number of COVID-19 cases in the department of Huancavelica has been remarkably controlled until mid-July, which is where daily infections in the region begin to increase, one cause of this event is that on July 1, the Peruvian state begins the "Third Phase" that contemplates the opening of some 60 commercial activities; And although the number of infections has increased, this has not been exponentially as it has happened in other regions. However, we can also notice there is a slight increase in the number of infections in the first days of May, this is explained by the disorganized and uncoordinated mass transfers made by many people wanting to return to their places of origin as a result of evictions by of its tenants.

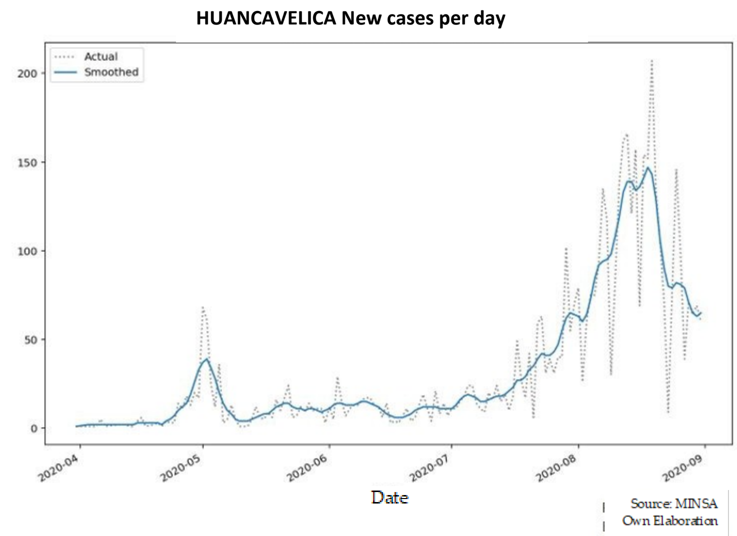


Figure 5.17. *Evolution of Rt in the department of Huancavelica until August 31*

Evolution of R_t in the department of Huancavelica One way to measure how the departments evolved with respect to the number of infections and thus be able to take measures such as targeted quarantine, is through the evolution of the R_t Reproducibility Index in real time. In the case of the department of Huancavelica, it is observed that this, at the beginning, is highly variable until the second week of July. As of July 15, we note that this Rt remains above 1 until the third week of August approximately, this due to the economic reactivation in more sectors of Peru. However, at the end of August this R_t is below 1, which means that an infected person can no longer infect another.

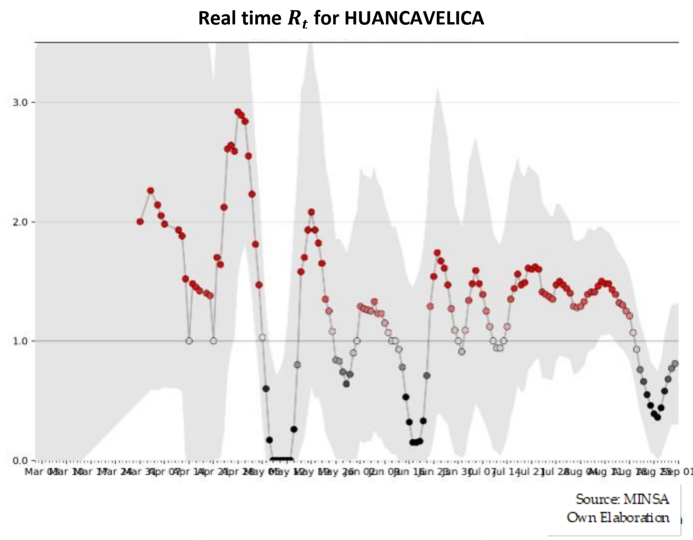


Figure 5.18. *Evolution of R_t in the department of Huancavelica until August 31*

Huánuco Region

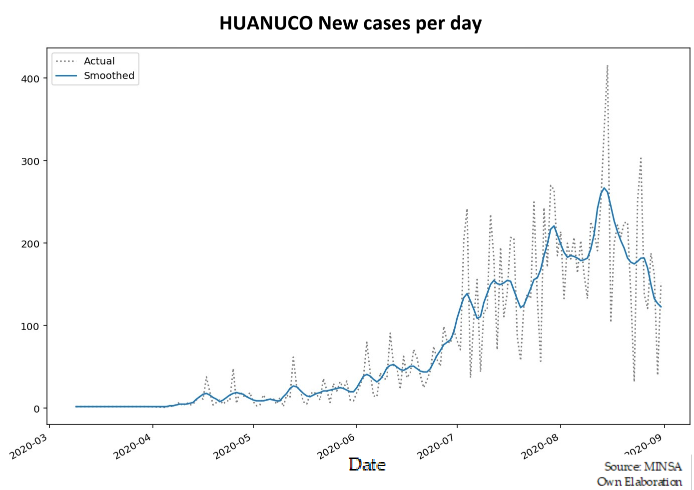


Figure 5.19. *Contagions confirmed in Huánuco as of August 31.*

For the month of July - August, it has been the months with the greatest increase in infections for the department of Huánuco, to understand a little about the measures implemented by the Regional Government, we will detail it below.

One of the first measures in March was implemented by ESSALUD, at Hospital II Huánuco, an external triage area and a differentiated topic for the care of suspected cases of coronavirus. The hours of attention were 7 in the morning to 12 at night. The same was done at the Hospital de Tingo Maria, Centro Medico Ambo and CAP III Metropolitano.

Likewise, the assistance and administrative personnel were provided with the necessary implements, in accordance with the technical standard for clinical care and management of COVID-19 cases. In addition, there are medicines, supplies and medical materials for the care of patients with symptoms of the virus.

As a preventive measure, visiting hours were restricted from 4 pm to 6 pm in the Hospitalization Service and the visit was limited to one family member per patient. The same measure has been given in the Emergency and ICU services (Intensive Surveillance Unit).

At the same time, liquid soap and paper towels were introduced for proper hand washing in all the toilet services of the 13 Huánuco healthcare centers.

It should be added that, according to the established protocol, all patients with fever, respiratory distress, and other discomforts are immediately evaluated by the doctors on duty, who proceed to perform the swab test. The sample was sent to the National Institute of Health (INS) in order to rule out the presence of the virus that causes the disease.

In the first days of June, the COVID 19 Surveillance and Control Plan in Huánuco was approved, which sought to propose surveillance measures for the restart of work in the department of Huánuco.

For the month of July, MINSA sent 45 oxygen balloons, necessary to save the lives of people affected by the new coronavirus, delivered to the Leoncio Prado Health Network and distributed at the Covid-19 Temporary Hospitalization Center, which was installed in Tingo Maria and houses 40 beds.

500 N95 masks were allocated for the exclusive protection of health personnel who provide care at this COVID-19 care center.

Due to the high demand for oxygen, the MINSA has also been managing an air bridge to alleviate any risk of shortages in the region and save the lives of patients with the new coronavirus.

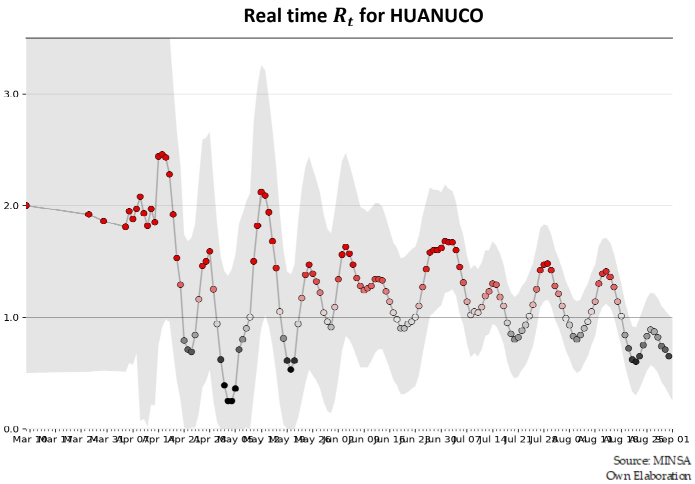


Figure 5.20. *R_t in real time estimated for the department of Huánuco as of August 31.*

In the case of the reproducibility rate R_t, it is observed that in the first months of March and April, it was where the rate became greater than 2, which implied that 1 person could infect more than two people, so If the measures had not been taken to be able to stop the expansion of COVID-19, the situation would have worsened in Huánuco, after that, after the implementation of measures for greater control by health, as well as the surveillance plan , the value of R_t has been oscillated between 1 and 2, thanks to the differentiated measures by sex and age, it was possible to decrease the value of the contagion rate in the following months, reaching a value lower than 1 for the month of August.

Ica Region

The evolution of infections for the department of Ica has a slight growth for the months of April to June, this is because the government aimed to gain time to equip its hospitals, taking measures such as mobilizing the most serious patients to Lima, application of COVID19 treatment and proper isolation both to people who tested positive and to people who presented symptoms of the disease, in order to lower their viral load and prevent them from worsening their health status and reaching emergency. This measure helped to ensure that the number of infected is not excessive, also the health system did not collapse and people did not die on the streets as in other departments, however, as of the fortnight of July, the cases continued to increase, so the government I take the quarantine measure so far.

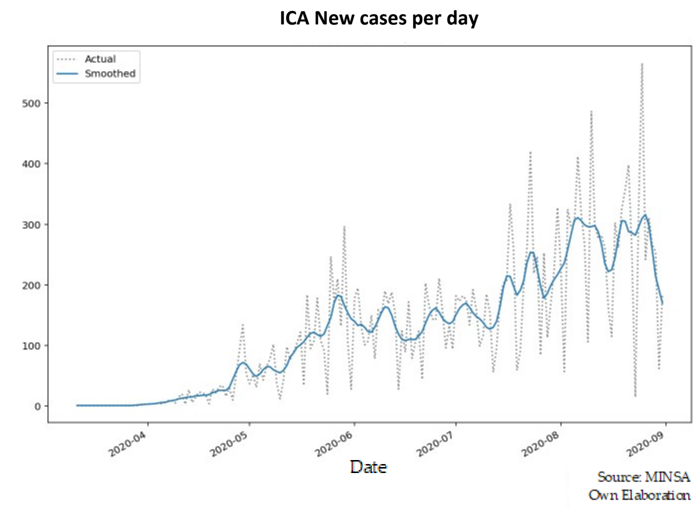


Figure 5.21. *Contagions confirmed in Ica as of August 31.*

The R_t for the department of Ica, unlike others, was reduced at the beginning of May much earlier, it is observed in the graph that the value of the reproducibility index oscillates in 1.5 and has the same behavior until the end of August, this is It is because the government kept its population in quarantine and continued with the measures it initially took.

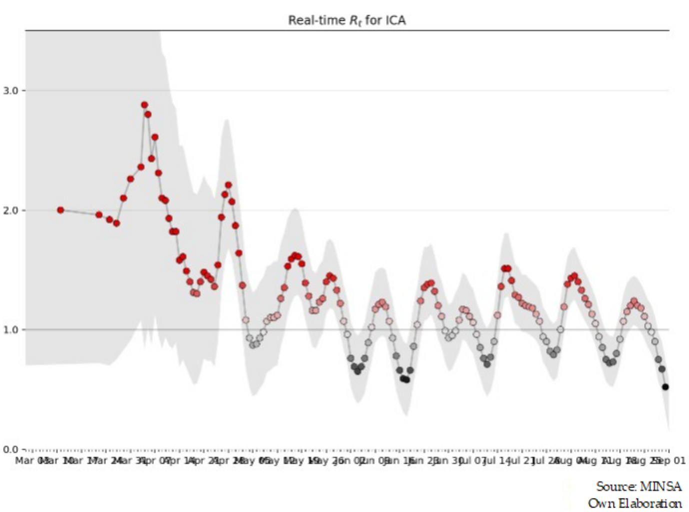


Figure 5.22. *R_t real estimated for the department of Ica as of August 31.*

La Libertad Region

Evolution of confirmed infections in the department of La Libertad Figure 5.22 shows coronavirus infections since March 15, the date on which the first case was reported in the department of La Libertad. A month after this fact, it is observed that the infections have been increasing rapidly, this product of the complete disinterest of the population of La Libertad for complying with the measures and the lack of police and army personnel. This high number of infections was maintained during the months of June, July and August.

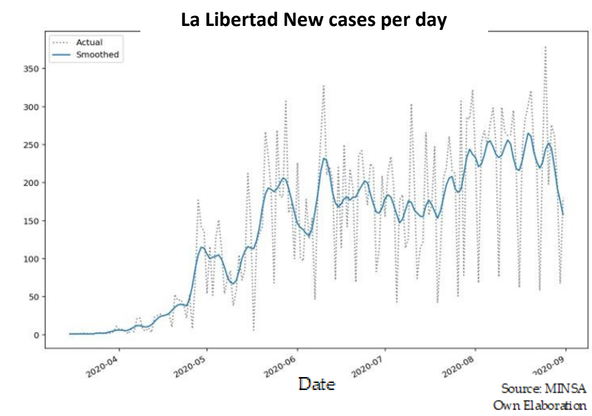


***Figura** 5.23.* *Evolution of Rt in the department of La Libertad until August 31*

Evolution of R_t in the department of La Libertad Next, in Figure 5.23, the Reproducibility Index R_t is observed throughout the pandemic in the La Libertad region, in its beginnings, as well as in other departments, it is highly variable, that is, it can reach values above or above below 1, however; for the months of July and August, this R_t greater than or equal to 1, which means that the disease remains in that region but does not disappear. Such results led the state to continue taking measures on these departments as well as others in the n Date ern part of Peru.

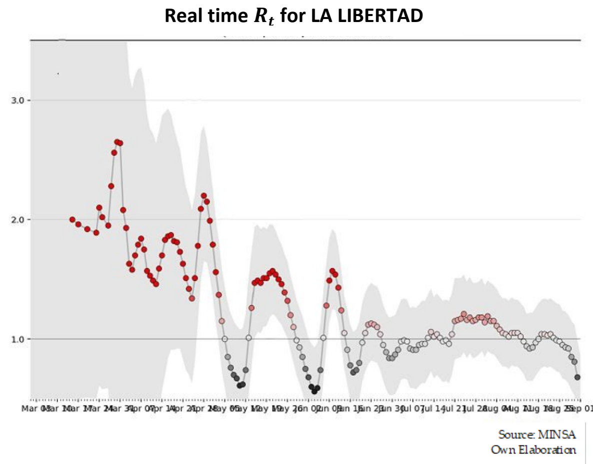


Figure 5.24. Evolution of R_t in the department of La Libertad until August 31

Lambayeque Region

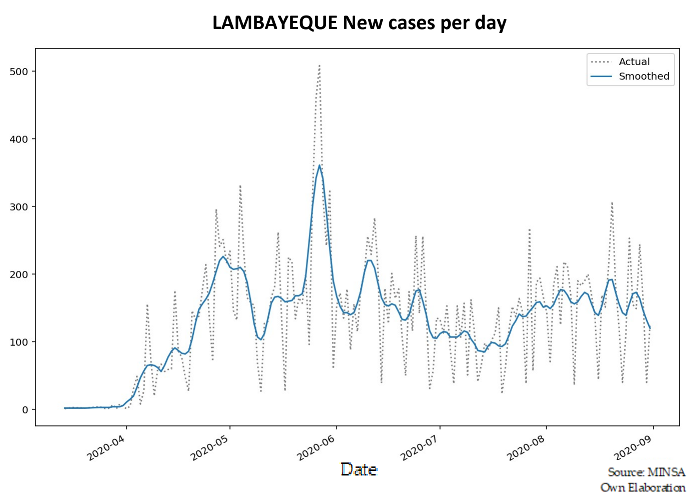


Figure 5.25. *Contagions confirmed in Lambayeque as of August 31.*

In Lambayeque it can be seen that the number of infections has almost remained constant around 200 daily cases registered during the months of March - August, since the arrival of the pandemic in Peru, according to the report published by the newspaper El Comercio, detailed below. In the first days of March, before quarantine, a young man born in the town of Calupe, in Tuman, and residing in Spain for several years, took a swab test and tested positive for COVID-19. The news quickly reached Lambayeque, where his family lives. At that time, more than 2,000 Spaniards had fallen ill and 55 had died. In Peru there were only 17 infected, the situation still seemed manageable. The northern region saw the coronavirus as an alien, remote, almost exotic evil. A few weeks later, everything was different, everything got worse.

The people of Lambayeque were so carefree, that they made mistakes, one after another. At that time, a 57-year-old from Chiclayo, recently arrived from Spain, was at the Luis Heysen hospital and they were going to do the screening test, but the nurses left him alone and he fled. If he had the virus, he spread it. The errors were many and very serious. Days later, when the government had already suspended school classes, the authorities discovered that at least one private school in Chiclayo was still giving classes, exposing students and teachers.

According to the report sent (MINSA), in Lambayeque for the month of April there were 642 confirmed cases of coronavirus (out of a total of 14,120) and 75 deaths (in all of Peru there were 348). That is, one in five victims of COVID-19 is from this northern region. There, the fatality rate was 11.68%, the highest in the country. It is the third region with the most cases and its situation is truly tragic.

The provinces of Chiclayo and Ferreñafe have the highest number of contagion cases, as well as victims. **"The population is aware of the crisis, but is not aware of it,"** says the provincial mayor of Chiclayo, Marcos Gasco, by telephone.

Regarding the implemented measures, it is detailed that:

Regarding health, for the month of March preventive control was available at the regional level, three telephone lines of Call Center were enabled so that people who presented symptoms of acute respiratory infection, fever, cough, sore throat can call so that health personnel come to assist them in their own homes, and avoid overcrowding in hospitals.

In the same way, for wholesale and retail traders, avoid hoarding and speculation of products in general and, especially basic necessities, to the detriment of the population of Lambayeque and not take undue advantage of the circumstances arising from the measures dictated by the Government. National in the fight against COVID-19 that affects our country.

For the month of May, the regional governor, Anselmo Lozano Centurion, formalized a request from the Covid 19 Command to the President of the Republic, which sustained the request for immobilization from May 8 to 12, 2020, 24 hours a day, based on in an epidemiological analysis carried out by the Health authorities.

The extreme measure suggested was based on the failure to comply with the isolation and social distancing measures, observing, on the contrary, a high traffic of people on the streets without the correct use of masks, an exponential increase in Covid positive people and patients who arrive at hospitals in a state of extreme severity, putting the health systems care capacity at risk.

In June, the Lambayeque Regional Health Management began vaccination against influenza and pneumococcus in adults and vulnerable populations with risk factors such as diabetes, high blood pressure, and cardiovascular diseases, among others.

For the month of August, through decree No. 000017-2020-GR.LAMB / GR, the Regional Government of Lambayeque, ordered that the Regional Health Management implement a series of measures in the 182 Health Centers prioritizing diagnosis and treatment To stop the advance of Covid 19. In this way, through these establishments, Covid-19 medical kits will be distributed free of charge to people with a positive diagnosis of Coronavirus (Covid19), consisting of essential drugs such as: Ivermectin, Azithromycin and Paracetamol o Ibuprofen, previous prescription issued by the corresponding Health Center.

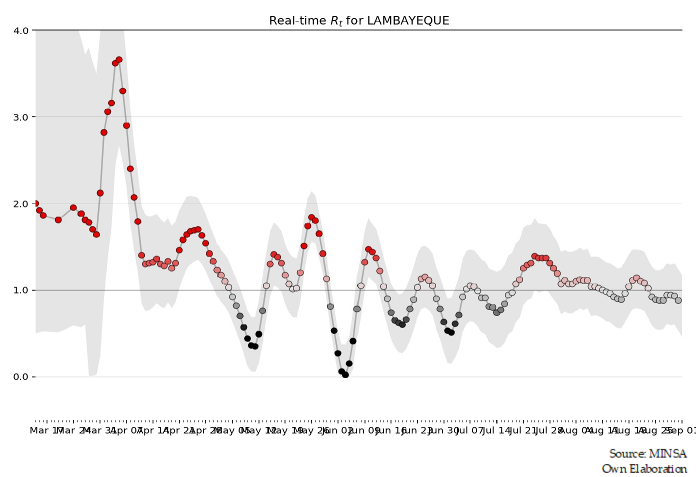


Figure 5.26. R_t in real time estimated for the department of Lambayeque as of August 31.

Regarding the value of the R_t Reproducibility rate, it can be observed that in the month of March it reached a value very close to 4, which implies that 1 person could infect another 4, therefore the exponential increase in infections was encouraged. , contributing to the large number of deaths, which placed Lambayeque as one of the most affected departments for the month of March and

April, after which the regional government promoted measures in order to prudently care for people with symptoms of COVID 19, and patients with pre-existing diseases, in order to reduce the lethality of the pandemic in the department, for this reason it has been possible to stabilize the contagion rate for this department, so much so that as of August it has a value of less than 1 with a 95% confidence interval, which would indicate that it could still be a value greater than 1, therefore the measurements for this department should not be neglected.

Loreto Region

For the department of Loreto, the growth of infections was exponential for the first months, this was due to the fact that the population did not comply with the isolation, the military and police who were in charge of taking people who were out of town to jail. Their houses, when exposed, ended up being infected, the Ministry of Health (MINSA) showed its concern with a donation of a 15-ton shipment, consisting of supplies and biosafety materials, as well as reagents for taking samples and medicines, having so many tests contagion data skyrocketed in early May. Among the measures that the Regional Government of Loreto took for the month of June were:

- Approval of the health project aimed at 89,177 citizens of indigenous communities, with a total investment of S / 29 million. Likewise, the expansion of hospital capacity with 300 hospital beds and 25 ICU beds.
- The mobilization of 102 health professionals for the cities of Yurimaguas, Requena and Caballococha.
- Distribution of 4450 molecular tests and 62 775 rapid tests for diagnosis; the distribution of 999 568 biosafety kits for health personnel.
- Acquisition of 900 oxygen cylinders for the care of patients with moderate and severe disease

Let us remember that the department of Loreto is large with a population of 883 510 inhabitants, so despite all these measures the growth of contagion oscillates at 150 daily for the month of August, however, it no longer has a steep growth. big. The evolution of the Rt for Loreto is very high for the first months of March, April and May this is due to the reasons explained above, and this is reduced from the month of June where the Rt values oscillate to 1 where it indicates that there are good control of the pandemic.

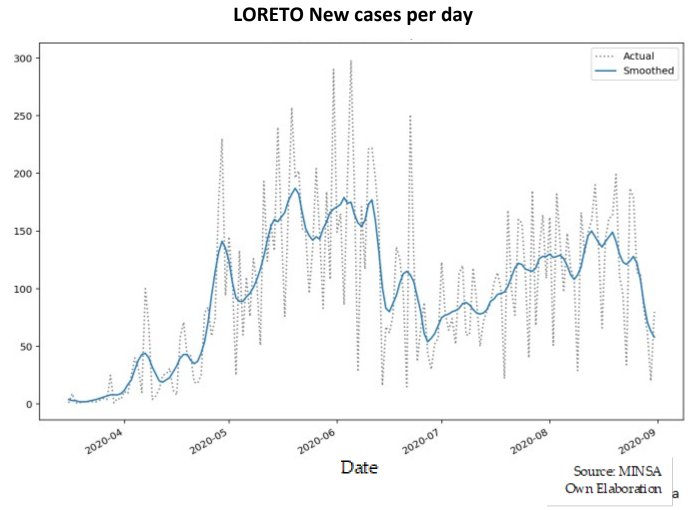


Figure 5.27. Contagions confirmed in Loreto as of August 31.

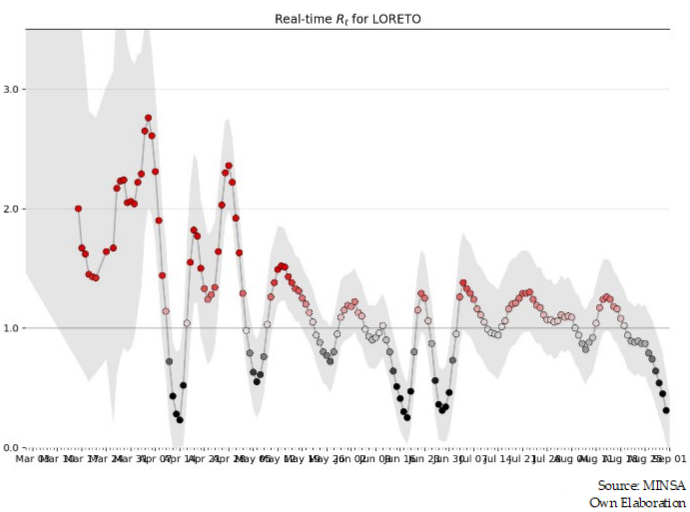


***Figure 5.28**. R_t real estimated for the department of Loreto as of August 31.*

Moquegua Region

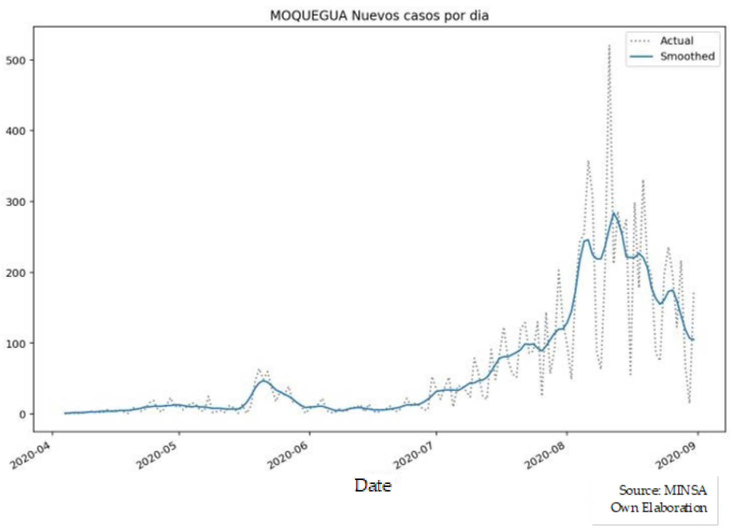


Figure 5.29. *Evolution of R_t in the department of Moquegua until August 31*

Evolution of confirmed infections in the department of Moquegua the number of daily cases for Covid-19 in the department of Moquegua has not been significant in the first months since the first case was reported, the population took the necessary measures as well as focused on having the greatest number of tests in order to have more accurate information about the situation, becoming the first region in the country in number of Covid-19 tests per 10,000 inhabitants. However, in the third week of May, it indicates that the massive transfers that the population began to carry out caused a slight increase in the number of infections, which later; it was controlled due to the quick response of the government.

Evolution of R_t in the department of La Moquegua R_t

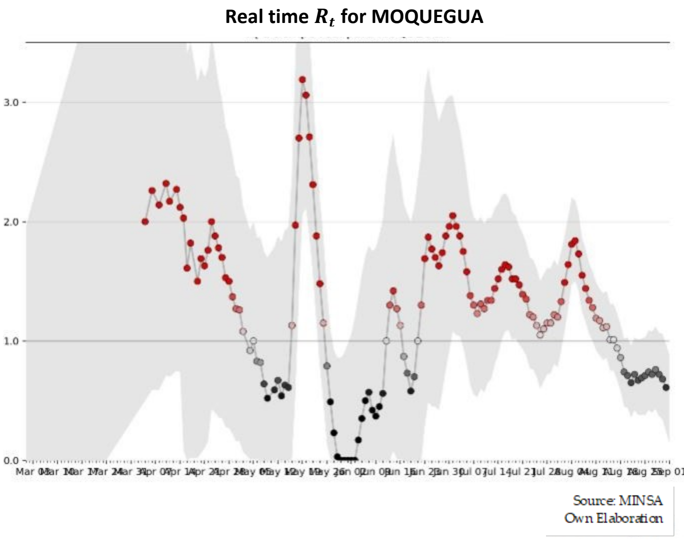


Figure 5.30. *Evolution of the Rt in the department of Moquegua until August 31*

As can be seen in figure 5.29, Moquegua presents a high R_t in the last two weeks of May, and as mentioned above, this as a consequence of the massive mobilization of the population to want to return to their places of origin. Already from mid-June to August it is observed that this R_t remains above 1 and even reaching 2, which indicates that a person infected with the Covid-19 virus can infect 1 or 2 people making the epidemic go increasing, before this, the Social Health Security (EsSalud) sent a group of 30 health professionals to Moquegua for the detection and timely treatment of covid-19 patients in addition to a shipment with 6,000 rapid tests and 6,000 covid-kits. 19. This caused the Rt to be below 1 for the last week of August.

Pasco Region

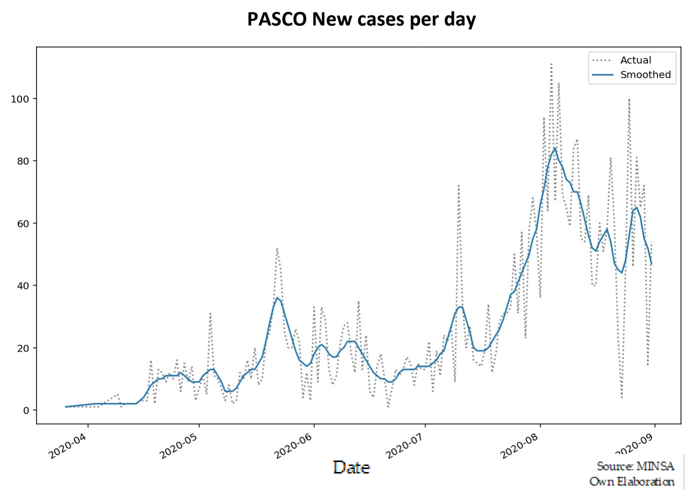


Figure 5.31. *Contagions confirmed in Pasco as of August 31.*

Pasco is one of the regions that has better assimilated the COVID 19 pandemic. According to the report of the Regional Health Directorate (DIRESA) of the region for the month of June, to date there are 1,332 infected people and 18 deaths, the latter figure has not risen since June 21; that is, it has been almost 8 days without counting another death due to the virus. Why after three months after the virus appeared, Pasco maintains low numbers of infected and dead?

According to the director of the DIRESA de Pasco, Alcedo Jorges, this is due to the measures they have applied: "We have strengthened primary health care, we have a more active epidemiological surveillance, that is, we are looking for cases, there is coordination with private entities for the purchase of rapid and molecular tests and we carry out communication campaigns to inform the population about the measures adopted and the risks of leaving home if you are positive", he pointed out.

Jorges maintains that in Pasco they are betting on a continuous epidemiological monitoring that helps to detect and quickly isolate positive cases and their families. This measure can be carried out because, both from the Regional Government and from private entities, tests are being purchased to achieve a larger-scale screening. He also points out that monitoring has been strengthened: "The follow-up teams do daily clinical monitoring of these patients to see their status. They do this on a daily basis when it is by phone, and for those who are patients at risk, it is done in person every three days", he pointed out.

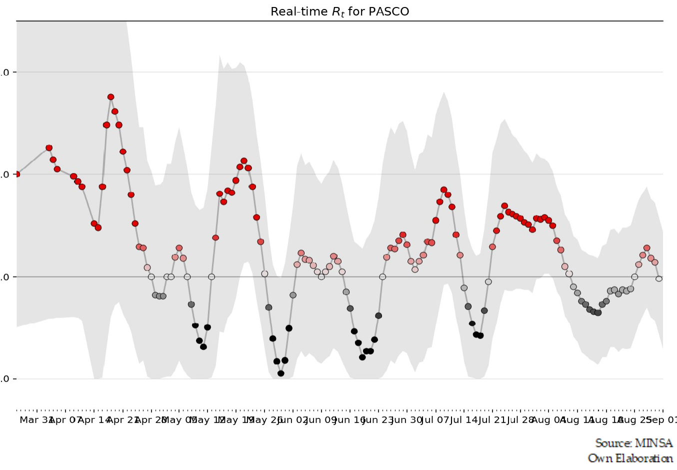


***Figure 5.32**. R_t in real time estimated for the department of Pasco as of August 31.*

Regarding the value of R_t for Pasco, it is observed that the first months it reached a value close to 3, but that later it tended to fall, which would allow the value of Rt to be lower for the month of August. a 1, however, it is necessary to consider that the 95% confidence interval used for the estimation would mention that we could still possibly be above a higher rate 1.

Piura Region

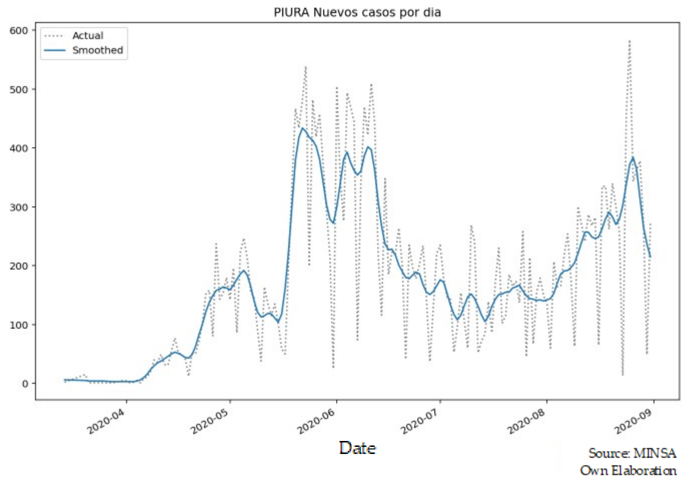


***Figure 5.33**. Contagions confirmed in Piura as of August 31.*

Evolution of R_t in the department of Piura The figure shows that Piura had a continuous growth until the beginning of May, but this has been triggered since the fortnight of May when the hospitals began to collapse, there were no longer beds, the patients were treated in the corridors of the hospitals, the despair of the population to be cared for made it continue to grow, due to these problems the Regional government of Piura decided to take measures that helped control the pandemic, among the measures and actions they took:

- Installation of isotanks, which is a cryogenic storage tank for oxygen that stores 60 thousand cubic meters of liquid oxygen, distributed by medical gas networks for each of the 320 beds of the Santa Rosa and Cayetano Heredia Hospitals in Piura.
- Creation of the “DOCTOR AT YOUR HOME” program for May 29, a Social Support Program (PAS) that consists of bringing medical care to the homes of vulnerable populations with coronavirus symptoms.
- Tele-orientation and telemonitoring, which is the communication through technologies (social networks, video calls) carried out by the nursing staff of health establishments to guide the relatives of the Covid-19, this was carried out as of July 1.
- Creation of new temporary hospitals and ambulances for several provinces on July 6.

These measures taken by the regional government of Piura made the R_t values that were in values of more than 3 fall below 1 for the month of June, it means that the pandemic was more controlled, this is observed in the graph of the evolution of the R_t. Let us remember that for July 1, a focused quarantine is applied, this is not for Piura that is why by July 15 the contagion rate begins to grow exceeding the value of 1.

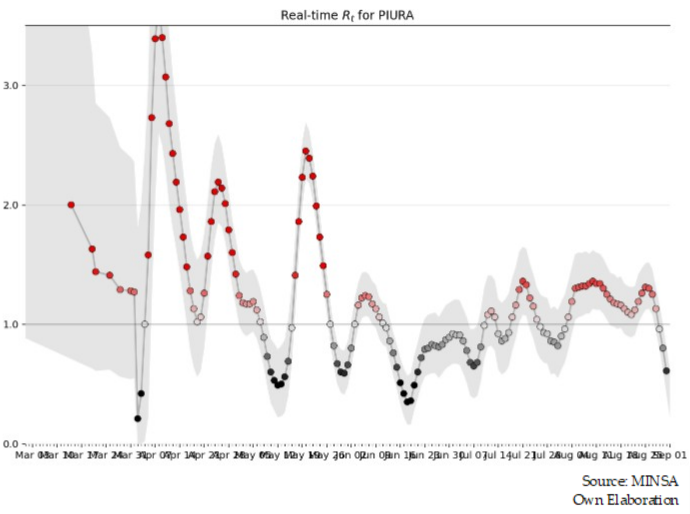


Figure 5.34. *R_t real estimated for the department of Piura as of August 31.*

San Martin Region

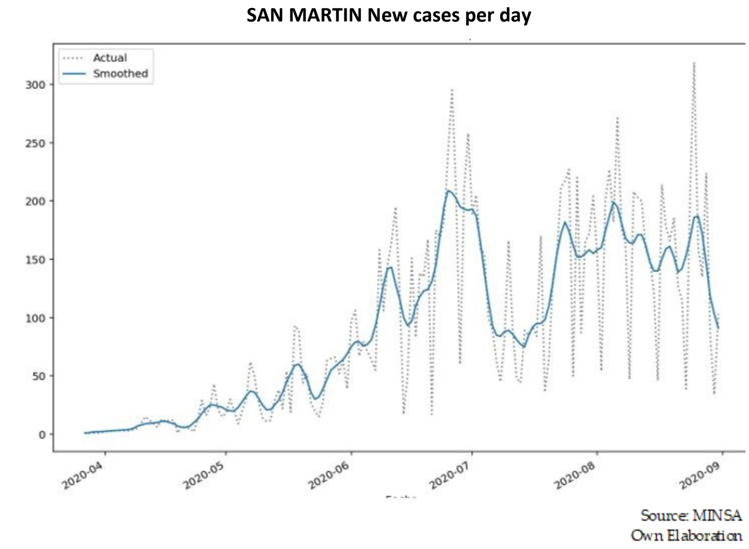


Figure 5.35. *Evolution of R_t in the department of San Martin until August 31*

Evolution of confirmed infections in the department of San Martin. The department of San Martin is also one of the most affected in the jungle region. Figure 5.34 shows us how it is that since mid-April the number of infections has been increasing continuously, this began with the transfer of a large number of people to San Martin, which made it difficult for the authorities of both the State and the regional governor to organize. since the lack of rapid tests by the Executive, did not allow transferring several groups to their hometowns. All this, added to the lack of beds and oxygen in the hospitals, made its growth exponential. In the month of July, and thanks to the two medical brigades made up of 37 health professionals, including doctors, nurses, technologists and pharmacists, they allowed to reinforce the attention to coronavirus cases, causing it to decrease, however by the end of July these contagions rose again, this could be the result of the lack of care of the population at the beginning of the economic reactivation.

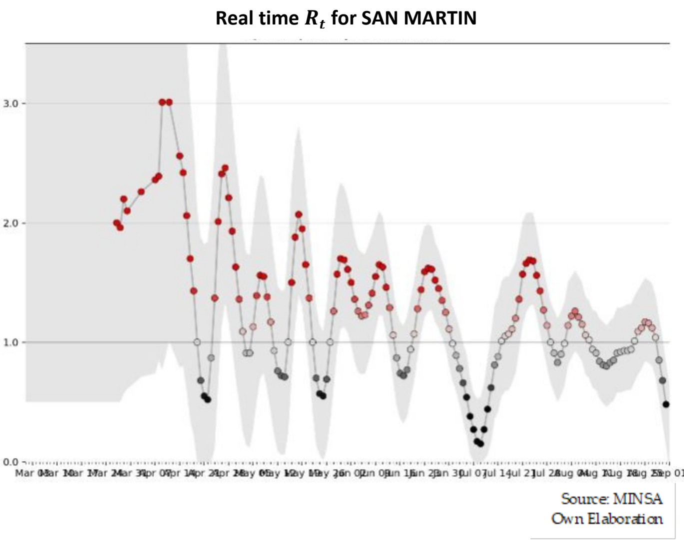


Figure 5.36. Evolution of R_t in the department of San Martin until August 31

Evolution of R_t in the department of San Martín. In San Martin, the reproducibility index has varied notably since the first case was reported, a 46-year-old Native American who had had contact with tourists from the Netherlands. Figure 5.35 shows us that this Rt in the month of June has been above one, for the following month this index has been below 1 due to the arrival of medical brigades as support. By the end of July and August, the Rt has been oscillating over one until finally being less than 1 for the last week of August, which would indicate that an infected person, on average, may not infect anyone; which is favorable for the region.

Tacna Region

The evolution of infections in Tacna is slow during the months of April, May, June and July, this is due to the measures taken by the government such as the training of the assistance and administrative personnel of the ESSALUD Assistance Network, also on Wednesdays The markets are closed for a deep cleaning and disinfection by the same merchants, and the arrival of 25 health professionals mobilized to the areas most impacted by the pandemic Due to its high number of infections for the month of August, Tacna is in quarantine.

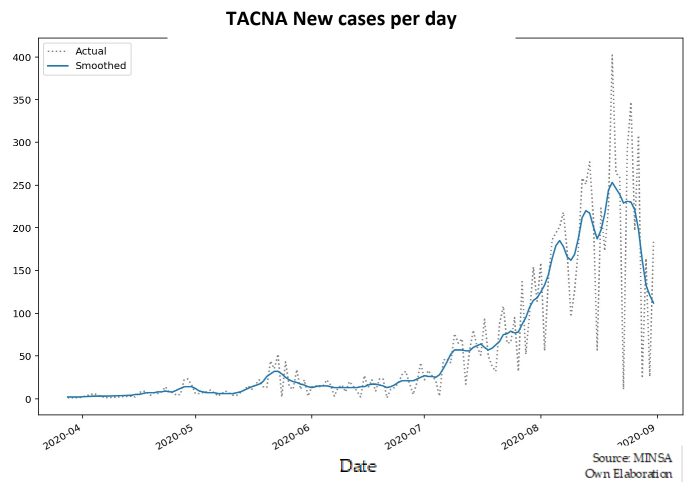


Figure 5.37. *Contagions confirmed in Tacna as of August 31.*

Figure 5.37 shows us how the reproducibility index has evolved in the department of Tacna. This index shows the variability at the beginning of the pandemic, however; In the months of July and August this index was above 1, this due to overflows in hospitals and the lack of medical equipment.

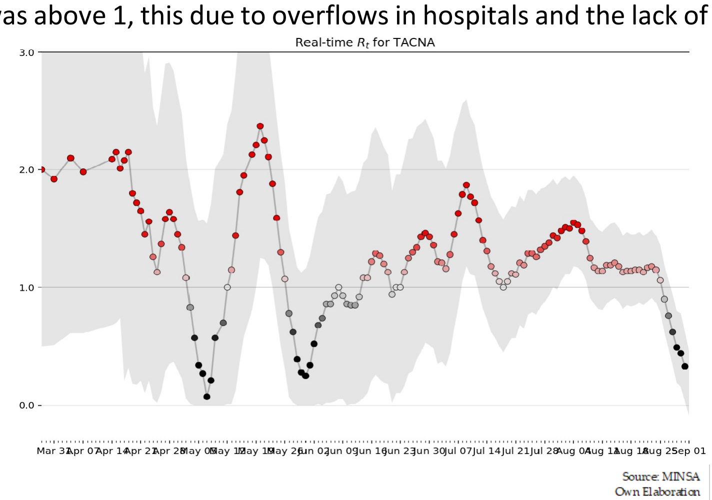


Figure 5.38. R_t in real time estimated for the department of Tacna as of August 31.

Tumbes Region

The evolution of infections in Tumbes increases constantly until the beginning of July where the highest peak of 140 infections is had, despite the quarantine extended to this date, infections continued to increase, this was due to the little responsibility of the population to comply isolation, crowds in all types of establishments without due distance. On June 4, the regional government of Tumbes did not restart the activities of the productive and commercial conglomerates as in the other departments because the number of infected did not reduce, the government, seeing that the population of Tumbes did not comply with the isolation, knew that they would come Higher daily infections, which is why rapid tests and drug kits were massively distributed in campaigns, this led to a reduction in infections, maintaining 40 per day. On July 1, the quarantine for the Tumbes population ended, because the pandemic was already under control.

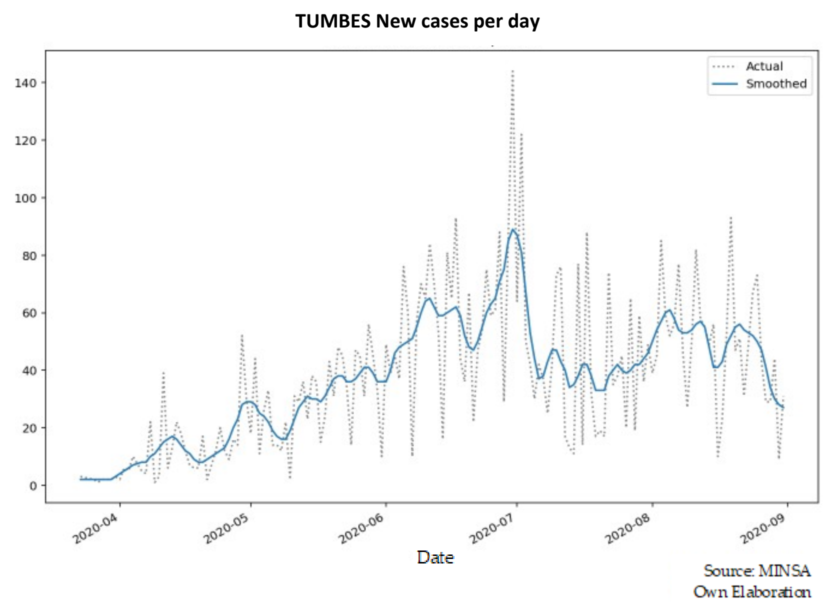


Figure 5.39. *Contagions confirmed in Tumbes as of August 31.*

In the graph of the reproducibility index, it is observed that, like the evolution of infections, the R_t is high, the shaded part indicates that it can reach values of 3, that is, a person with COVID19 can infect up to 3 people. In the last month a low R_t is observed, this means that measures are being taken that are controlling the pandemic.

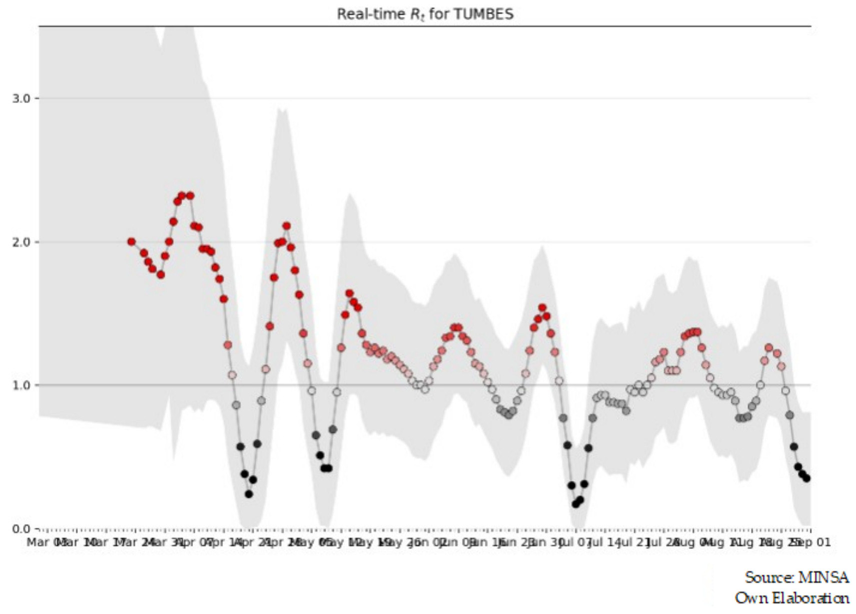


Figure 5.40. *R_t real estimated for the department of Tumbes as of August 31.*

Ucayali Region

The Ucayali region has been one of the regions with a not very stable behavior over time, that is, it has had a strong wave of infections at the end of May, this due to the fact that there was too much misinformation in the Jungle in addition to the collapse of health centers. Due to this high behavior in the contagion curve, Ucayali was one of the first departments to have a focused forty-year-old since the curve did not have a significant decrease.

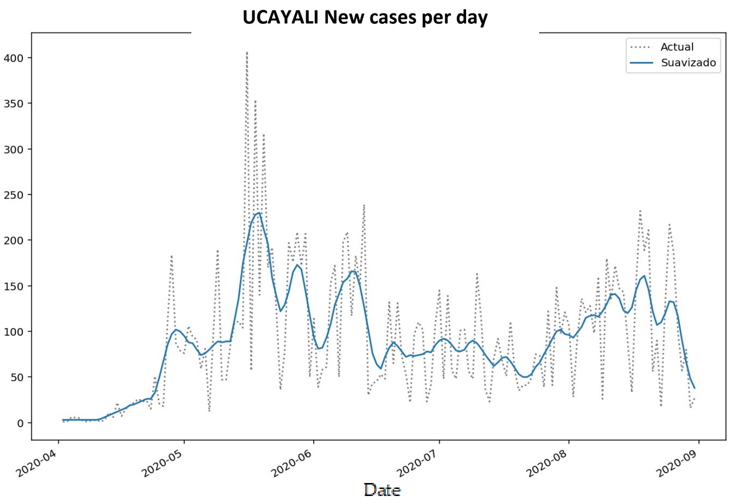


Figure 5.41. *R_t real estimated for the department of Tumbes as of August 31.*

Given the measures implemented by the central and regional government to stop the advance of COVID-19, there was a decrease in the number of reproducibility for the month of June, however, as seen in Figure 5.41 for that month, the Rt is below of 1 but its maximum density interval is above 1, so the measurements should not have been stopped or relied entirely on Ucayali. This carelessness on the part of the authorities is reflected in the abrupt rise of the R_t for the month of August where this and its interval are above 1, that is, one person could infect another and so on.

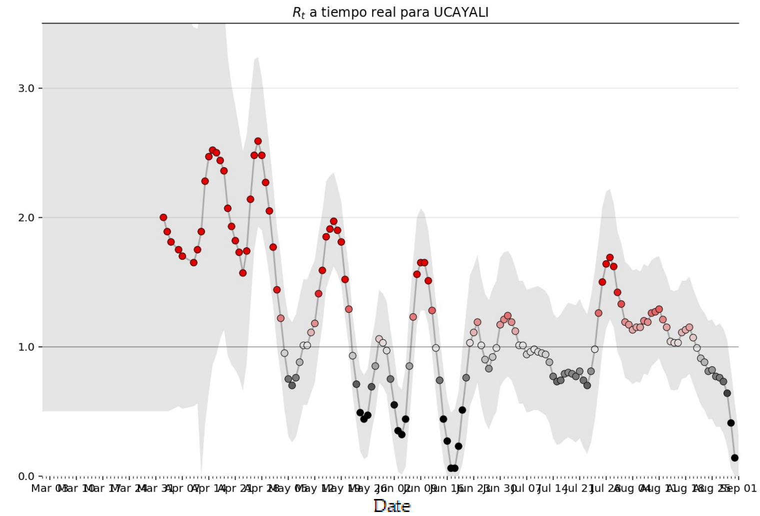


Figure 5.42. *R_t real estimated for the department of Tumbes as of August 31.*

5.2. Analysis of the Deceased

As mentioned above, it is considered that there is not a good count of the real number of deaths from COVID 19, so it was sought to approximate the real number of deaths from COVID 19 by using time series with the ARIMA modeling technique with the data of deaths reported to SINADEF in the period from January 2017 to March 2020, from there to issue forecasts for the period April - July 2020, in order to calculate an approximate number of deaths from COVID 19 based on the difference reported.

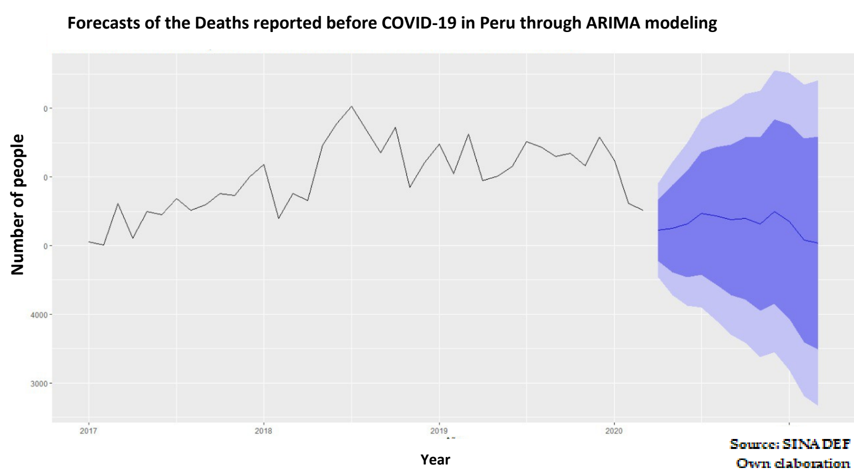


***Figure 21**. Estimated Time Series of death cases for the departments of Peru reported to SINADEF in the period 2017 – 2020*

Figure 21 shows the forecasts made in the ARIMA model with a confidence interval of 90% and 80%. After that, we made the comparison of the series of deaths reported to SINADEF and the estimated series. It can be seen that the estimated series has a lot of similarity to the real series, so we can infer based on

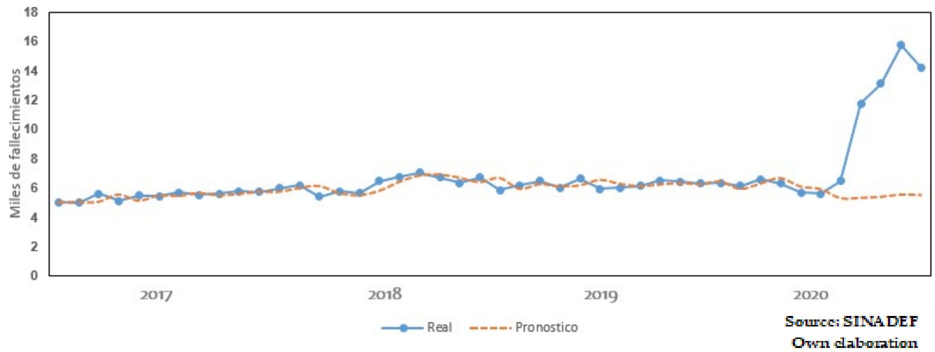


Figure 21. *Estimated Time Series vs Real Series of death cases for the departments of Peru reported to SINADEF in the period 2017 – 2020*

The forecasts are that if the pandemic had not reached Peru, the number of deaths reported monthly would follow the trend of the estimated series, but as a result of the pandemic, the death curve grew exponentially.

Calculating the differences between the real and the predicted we have 23,858 deaths in the period April - August 2020.

Month	**Deceased (in thousands)**
April	1.17
May	6.41
June	7.72
July	10.16
August	8.67

Table 5.2. *Deaths predicted from SINADEF data, by ARIMA modeling*

Conclusions

1. Regarding the exploratory analysis carried out, it was observed that as of August 31, the behavior of the daily contagion curve had very different slope changes, for the months of July and August it had the highest peaks. In addition, it is observed that these daily cases have a weekly seasonal behavior, that is, in the week there are different amounts of daily cases, the explanation is for the days in which the tests are taken and the sampling criteria changes, since there are days in which massive operations were carried out and this is what produces the variability in the days. On the other hand, comparing the number of total infections by sex, it can be observed that the difference is around 3% which is not significant and in the case of age groups, the group called adults (30-50 years) presents the highest number of accumulated infections with respect to the other age groups.

2. In the case of the deceased, it can be observed that the daily cases of the deceased reached their highest peaks between the month of May and the first week of June and for the subsequent dates a constant and slightly decreasing series has been seen. As another point to rescue, the significant difference between the deaths of COVID-19 by sex can be observed, since the study showed that as of August 31, the difference between the two was 40%, being the male sex the one that reported the highest cases. In addition, the age groups that had the highest number of deaths were both men and women who had ages, which ranged from 50 to 80 years of age.

3. With the study, it was observed that different stages of the COVID-19 pandemic have been lived and are being experienced in Peru, these states are experienced differently in each department.

The difference is in the case of daily infections and the number of infected per thousand inhabitants. It was noted that departments such as Loreto, Ucayali, Madre de Dios, Piura and Lambayeque present very similar daily contagion curves and which had high daily contagions from May to the end of July. And on the other hand, there are the departments whose daily contagion curves until the second week of July did not exceed 150 daily infections, departments such as Cusco, Apurimac, Puno, Moquegua and Ayacucho just the third week of July began to drastically increase the cases of daily infections, mainly due to the reactivation of domestic flights and interprovincial transport.

4. 4 groups were defined according to the Rt, the first group has been designated as high risk that the population continues to be infected since they have the maximum value of their credibility intervals above 1, that is, a person can infect 1 person on average, these departments are Apurímac, Pasco, Lambayeque and Huancavelica. The second group made up of the departments of Madre de Dios, Junín, Ancash, Huánuco, La Libertad, Puno and Moquegua are called departments at risk, and their maximum values of their credibility intervals are slightly above 1 or close to 1 On the other hand are the group of departments that is in a moderate state of the pandemic, in this group are the departments of Tumbes, Ucayali, Ayacucho, Piura, Arequipa, Ica, San Martin. Finally, there are the departments where the pandemic is controlled according to its Rt, this group is made up of Loreto, Tacna, Amazonas, Cajamarca and Cusco.

5. It must be taken into account that the groups mentioned in point number 4 have been classified based on the Bayesian estimate of R_t, carried out in the present investigation, however, differences are observed with respect to the measures implemented by the Peruvian state to throughout the pandemic, since by August 31 the departments in total quarantine have been Puno, Tacna, Moquegua, Cusco and focused quarantines for the other departments, however, it is suggested that the total quarantine measures be implemented in the departments of Apurímac, Pasco, Lambayeque and Huancavelica because its R_t estimated in real time as of August 31 is greater than 1.

6. In the analysis of the deceased, a significant inconsistency has been noted in the cases reported by the MINSA and the deaths registered by the SINADEF. When making the projections using the ARIMA time series modeling technique, it was noted that there is an excess of deaths compared

to previous years, this excess can be explained by COVID-19 or other factors, however comparing the data from the MINSA can be seen that the amount they register is much less than it should be in reality, it is around 4 times the official figures. This is explained by many reasons, such as the percentage of asymptomatic that exist is 40%, many die at home or already in the hospital, but since they were not registered before with a test that confirmed their positivity to COVID-19, they are simply not considered. , in addition to taking into account people with pre-existing diseases, who due to the priority that COVID 19 presented in many health centers, especially the public, have not been able to continue their treatments, such as cancer, chronic diseases, etc. It should also be taken into account that being confined within their homes, the cases of femicides have increased considerably; Although all these cases have not necessarily been diagnosed as COVID 19, they could be considered as a population directly affected by COVID 19.

Bibliography

1. Hugo PEREA & Juan RUIZ HUGO VEGA (2020), “**Peru - What factors are behind the differential impact of COVID-19 in Peru? "**

2. Bettencourt & Ribeiro (2008), **“Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases”**

3. Vaidyanathan (2020), **“Estimating COVID-19 Rt in Real-Time”**

4. Perone (2020), **“COVID-19 Time varying reproduction numbers estimation for Brazil”**

5. A´vila Ayala, R. M. (2016), **"Inference of compartmental epidemiological models in social networks." [Master's Thesis, Center for Research in Mathematics, A.C.]**

6. Ridenhour B, Kowalik JM, Shay DK. (2014), **“Unraveling R0: Considerations for Public Health Ap- plications. AmJ Public Health.”**

7. Ridenhour, Kowalik y Sahay. (2007), **“The basic reproductive number and its considerations in public health. "**

8. Ombudsman (2020), **"What happened to them during compulsory isolation?”**

9. J. Aronson, J. Brassey, K. Mahtani, on behalf of the Oxford COVID-19 Evidence Service Team. **“When will it be over?”: an introduction to viral reproduction numbers, R_0 and R_e. (2020)** https://www.cebm.net/covid- 19/when-will-it-be-over-an-introduction-to-viral-reproduction-numbers-r0-and-re

10. Peruvian Technical Standard ISO/ TR 10017 – 2004

11. NTP-ISO3534-1: 2015 Statistics. Vocabulary and symbols

Appendix

Phases of an influenza pandemic

According to the WHO, there are 6 phases in an influenza pandemic:

Phase 1: A virus circulating in animals has not been reported to cause human infections.

Phase 2: An animal influenza virus circulating among animals, domestic or wild, is known to cause human infection and is therefore considered a potential pandemic threat.

Phase 3: A reassociated animal or animal-human influenza virus has caused sporadic cases or small clusters of illness in people, but without sufficient person-to-person transmission to sustain sustained outbreaks at the community level. In some cases, limited person-to-person transmission can occur, for example, when there is close contact between an infected person and an unprotected health worker. However, limited transmission in such restricted circumstances does not indicate that the virus has acquired the level of human-to-human transmissibility necessary to cause a pandemic.

Phase 4: It is characterized by the proven person-to-person transmission of a reassociated animal or animal-human influenza virus capable of causing “community-level outbreaks”. The ability to cause sustained outbreaks of the disease in a community is a sign of a significantly increased risk of a pandemic. Any country that suspects or has verified an event of this nature should urgently consult with the WHO so that the situation can be jointly assessed and so that the affected country can decide whether to initiate an operation of rapid containment of the pandemic. Phase 4 indicates a significant increase in the risk of a pandemic, without necessarily being unavoidable.

Phase 5: It is characterized by person-to-person spread of the virus in at least two countries in a WHO region. Although most countries will not be affected at this stage, the declaration of phase 5 is a strong signal that a pandemic is imminent and that there is little time to complete the organization, communication and implementation of planned mitigation measures.

Phase 6: The pandemic phase is characterized by community-scale outbreaks in at least one other country in a different WHO region, in addition to the criteria defined in phase 5. The declaration of this phase indicates that a global pandemic is underway.

In the case of Peru, the epidemiological phases due to the new coronavirus, according to the Peruvian government were:

Phase 1 (The first case arrives): In this phase, the disease enters the country with patient zero, a carrier of the virus who arrives from abroad and who traveled to a country in a risk zone. The first patient is isolated and an investigation is started to identify her activity and the people with whom they may have been in contact. In this context, protocols are defined and the hospitals to attend to the cases were determined.

Phase 2 (Targeted transmission or containment): More imported cases are registered by people who arrive from abroad and infect people who were in contact with the zero patient. It begins to identify those people who bring this infection imported from abroad to achieve adequate containment; In this scenario, an attempt is made to implement measures for citizens to prevent the virus from spreading at the household, community level, at the level of collective spaces with basic hygiene measures and by avoiding concentrations.

Phase 3 (Community transmission or community contagion): The first people with the disease have spread to a close group and these to others. In this phase, the positive cases increase and there is greater expansion, patients who are not related to imported patients begin to be identified. One of the measures taken to stop the uncontrolled advance of COVID-19 at this stage is the state of emergency and the mandatory social immobilization (quarantine) that was decreed throughout the country.

Phase 4 (Sustained transmission): In this phase there is an exponential growth of infected cases in the population, infections that affect many localities. Health authorities must guarantee citizens the timely management of the emergency and guarantee resources.

INTERNATIONAL STANDARDS

- NTP-ISO3534-1: 2015 Statistics. Vocabulary and symbols.

  Part 1. General statistical terms and terms used in the calculation of probabilities. 2nd Edition

  POISSON DISTRIBUTION:

$$P(X = x) = \frac{\lambda x}{x!} e^{-\lambda}$$

  where:

  x = 0,1, . . . , with parameter λ > 0

- GUIDANCE ON STATISTICAL TECHNIQUES FOR THE STANDARD NTP-ISO 9001:2001

Time Series Analysis

1. What is it?

Time series analysis is a family of methods for studying a collection of observations made sequentially in time. Time series analysis is used here to refer to analytical techniques in applications such as:

- searching for 'lag' patterns by statistically analyzing how each observation correlates with the immediately preceding observation, and repeating this for each successive separation period;
- the search for cyclical or seasonal patterns, to understand how causal factors in the past could repeatedly influence the future;
- the use of statistical tools to predict future observations or to understand which are the factors that have contributed the most to the variations in a time series.

Even though the techniques used in time series analysis may include simple "trend graphs", in this Peruvian Technical Standard such elementary graphs are listed among the simple graphical methods cited in Descriptive Statistics"

2. What is it used for?

Time series analysis is used to describe patterns in time series data, to identify outliers (that is, extreme values whose validity should be investigated), either to help understand the patterns or to make adjustments, and to detect turning points in a trend. Another use for it is to explain patterns in one time series with those of another time series, with all the goals inherent in regression analysis.

Time series analysis is used to predict future time series values, usually with some upper and lower limits, known as the prediction interval. It is widely used in the control area and is frequently applied in automated processes. In that case, a probability model is fitted to the time series history, future values are predicted, and then specific process parameters are adjusted to keep the process on track, with as little variation as possible.

International definition of health

According to the Center for Evidence-Based Medicine (CBEM), Nuffield University of Oxford's Department of Primary Care Health Sciences in the article **"When will it end?": Introduction to viral reproduction numbers, R_0 and R_e**

The basic reproductive number R_0 is the number of cases expected to occur on average in a homogeneous population as a result of infection by a single individual, when the population is susceptible at the beginning of an epidemic, before generalized immunity begins to develop. and before any immunization attempt has been made. So, if one person develops the infection and passes it on to two others, the R_0 is 2.

The zero in R_0 means that it is estimated when there is zero immunity in the population, although not everyone will necessarily be susceptible to infection, although that is the usual assumption. In an epidemic with a completely new virus, the earlier measurements are made, the calculated value is likely to be closer to the true value of R_0, assuming high-quality data. For this reason, it is better to speak of the transmissibility of the virus at the moment in which the effective reproduction number is measured, using a different symbol, R_e.

The effective reproduction number R_t is the number of people in a population that can be infected by an individual at any specific time. It changes as the population becomes more and more immunized, either by individual immunity after infection or by vaccination, and also as people die.

R_t is affected by the number of people with the infection and the number of susceptible with whom infected people are in contact. People's behavior (for example, social distancing) can also affect R_t.

The number of susceptible decreases as people die or become immunized from exposure. The sooner people recover or die, the lower the value of R_t will be at any given time.

Bayesian analysis of an AR(1) model with exponential errors

According to the Bayesian Analysis of an AR (1) Model with Exponential Errors *, developed by the Department of Mathematics (Statistics and Operations Research), Faculty of Sciences of the University of Malaga, is an article recovered by the Institute of Spanish Statistics (1990)

An AR (1) model for positive Time Series with non-normal errors is constructed and analyzed from a Bayesian point of view. To do this, a model with exponential errors and stationarity given by means of a positive autoregressive dependence is introduced. Given an informative a priori and a non-informative one for the learning object parameters (autoregressive and exponential), the corresponding a posteriori and predictive distributions are obtained, which are non-standard.

The posterior function for Φ:

$$f(\Phi/y_n, y_0) \propto \lambda^{n+\alpha-1} exp\left[-\lambda(\beta_n - \theta R_n)\right] I_{(0,\infty)}(\lambda) I_{(0,\theta^*)}(\theta)$$

where, $\Phi = (\lambda, \vartheta)$

Research Limitations

Limitations of the Methodology

Non-representative sample

Different governments have reported numbers of infected since the beginning of the crisis. However, it soon became clear that these figures were clearly skewed to the downside, for different reasons:

1. There are asymptomatic patients: they do not have symptoms, but are or have been infected. If there are no symptoms, no tests are done.
2. The inability to do sufficient diagnostic tests. While states like Germany have been able to carry out a large number of tests, other countries like Spain or Italy have not been able to carry out as many tests as they would have liked. This has led to recommending patients with symptoms to confine themselves to their home in case of suspicion, without knowing whether they have been infected.

The main evidence that the official number of infected is incorrect is the disproportion between infected and deaths. Taking the example of Spain, on April 20, 200,210 infected and 20,852 deaths were reported, 10.4%. It is known that Covid-19 does not have such a high mortality rate, which can only mean that the number of infected is much higher than that reported.

There have been several people who, using statistical models, have tried to estimate the real number of infected and alert us to the situation that was coming our way.

As early as March 10, authors such as Tomas Pueyo warned rulers around the world about the seriousness of the situation. Simply reviewing (1) the existing data, (2) the experience of the countries that were being most affected at that time (China, South Korea, Italy and Iran) and (3) knowing the exponential growth that infections tend to have. Viruses, Pueyo painted us the discouraging panorama that awaited us, especially if the governments did not take isolation measures.

Bort, in his article, made an estimate based on the official number of infected and another based on the number of deaths. The first estimate had the major drawback of the unreliability of the official data, which weakened the analysis. On the contrary, the estimate based on the number of deaths was based on a much more reliable data. Unfortunately, later we have discovered that not even the death data is reliable, since for many weeks the deaths derived from Covid-19 have not been properly computed, especially in nursing homes.

If we want to reliably estimate the real number of people infected with covid-19, we need a representative sample of people. Yes, a sample, the same as we use to do an electoral poll or a brand recall test.

A sample, if it has been suitably designed, makes it possible to accurately estimate the real incidence of a disease in the population, regardless of factors such as symptoms or saturation of the hospital system. The principle is very basic: we select people at random and do a test to detect the disease. People are selected regardless of whether they have symptoms or not, in all regions, of all ages, and the result is extrapolated to the entire population.

Subregisters of Deceased

In Peru 'the official records of deaths from Covid-19 only include cases confirmed by "molecular and rapid tests." However, a report by IDL-Reporteros compared the statistics provided to the WHO with that of the National Information System of Deaths (SINADEF) on those who died due to death.

From March to June 15, the MINSA reported 6,860 deaths, but the SINADEF, also a government entity, counted at least 24,262 deaths from the pandemic. Of these, 16,213 were confirmed fatalities by Covid-19 and 8,049 were classified as “suspicious”, “probable” or “not confirmed”. As in other countries, many died before being tested or admitted to hospitals.

The radical contrast in deaths is due to the fact that SINADEF "records cases that doctors certified as having died from the virus, either due to their symptoms and / or the result of rapid or molecular tests." of death are registered by medical teams there are variations such as "pneumonia due to Covid-19"; "Covid-19 infection"; "SARS COV 2 coronavirus disease"; "Atypical pneumonia" and others.

In this way, SINADEF records warn that so far in June an average of 494 people have died from the virus every day. According to IDL Reporteros, this information should be contrasted with the general database, since there are "typing" errors in the public database with online certificates, where "about 30% of death certificates are on paper".

The Minister of Health, Victor Zamora, admitted in April before the Health, Inspection and Monitoring Commission of the Covid-19 actions of Congress that “in all the countries of the world there are different degrees of under-registration. There is no desire to hide this information, but there may be weaknesses in terms of the quality of the record”.

Tool Limitations

- Peruvian Technical Standard ISO/ TR 10017 - 2004
  Limitations of the use of time series analysis:

The same limitations and precautions cited for regression analysis are also valid for time series analysis. When modeling a process to understand causes and effects, great skill is needed to select the most appropriate model and to use diagnostic tools to improve it.

The inclusion or omission of a single observation or a small set of observations in the analysis can have a significant influence on the model. Thus influencing observations should be understood and distinguished from outliers in the data.

The different techniques for estimating time series can have different degrees of success, depending on the patterns of the time series and the number of periods for which you want to forecast, in relation to the number of periods for which you have data. time series data. The selection of a model should take into account the objective of the analysis, the nature of the data, the relative cost, and the analytical and predictive properties of the different models.