Statistical determination of COVID-19 mortality in age groups in the Ecuadorian Highlands

Juan Santillán-Lima¹*, Fernando Molina-Granja², Patricio Santillán-Lima³, Daniela Caichug-Rivera⁴, Raúl Lozada-Yánez⁵ and Washington Luna-Encalada⁶

²,³ Universidad Nacional de Chimborazo. Facultad de Ingeniería. Ecuador.
⁵,⁶Escuela Superior Politécnica de Chimborazo, Facultad de Informática y Electrónica, Riobamba, Ecuador

Abstract

INTRODUCTION: A problem plaguing humanity is the countless deaths from COVID-19. Conducting a statistical study based on deaths data during the pandemic will allow us to identify the age groups most vulnerable to death. This information can be used by public and private health organizations to take preventive measures in the most vulnerable population.

OBJECTIVES: To determine the age groups most vulnerable to COVID-19 in the Ecuadorian highlands during the periods 2020-2021.

METHODS: The "Bootstrap method to calculate the confidence interval for proportions" was applied to the age groups of those who died from COVID in the Ecuadorian highlands in 2020 and 2021. The method is implemented in R Studio.

RESULTS: The adjusted confidence intervals suggest that there is a 95% confidence that the age group most prone to death from COVID-19 is the Elderly with a point estimate of 72.70% in 2020 and 76.61% in 2021. The main subgroup of 80 to 90 years followed by 65 to 70 years.

CONCLUSION: Adjusted confidence intervals indicate that there is a 95% confidence that the age subgroup most prone to death from COVID-19 is 80 to 90 years with a point estimate of 16.13% in 2020 and 16.85% in 2021. And 65 to 70 years with a point estimate of 12.98% in 2020 and 12.57% in 2021.

Keywords: Bootstrap by proportions, Confidence Intervals, COVID-19, Death from COVID-19, Elderly.

Received on 19 February 2023, accepted on 2 October 2023, published on 12 October 2023

Copyright © 2023 Santillán-Lima et al., licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.9.4133

1. Introduction

On 30 December 2019, China announced the existence of a new disease known as COVID-19.(1)Due to its high contagion rate, the World Health Organization on 11 March 11, 2020. (2)and entered the equator on 29 February of the same year.(3).

Corresponding author. Email: juankasl@outlook.com

Covid-19 spread very quickly through Ecuador, causing 23,793 deaths in 2020, due to causes of COVID-19 or suspected COVID-19, followed by ischemic heart disease with 15,639 deaths and 7900 people due to diabetes mellitus. While in the year 2021, the three main causes of death were: 16,610 cases due to confirmed COVID-19, 13,002 due to ischemic heart disease, and 5,564 due to diabetes mellitus. (4).
We must also take into account that COVID-19, despite the existence of a vaccine, cannot be eliminated, and we must learn to coexist with said virus. In recent weeks, the coronavirus has become a problem that few talk about. However, the virus still has an important capacity to become an issue again, especially for older adults or people with comorbidities. An example of this was what happened in November and December, when infections increased significantly and long lines were again seen at the vaccination centers.

Carrying out a statistical study, based on the data on deaths in the pandemic, will allow us to know the age groups with the greatest vulnerability to death from COVID, this information can be used by doctors and public and private health organizations in Ecuador to take preventive measures in the most vulnerable population.

Although the COVID pandemic has ended, the alert for this disease has not ended in this sense, in June the WHO issued an epidemiological alert(6) which includes SARS-CoV-2 in which it recommends that the Member States strengthen and integrate surveillance on SARS-CoV-2 and adopted the necessary measures for the prevention and control of serious cases. This indicates that we must continue to be vigilant against this disease.

In article(7) The non-interventionist prospective observational study was used as a technique. In(8) the relationship between COVID-19 mortality and altitude of residence of the deceased is studied. Using the ecological observational study and descriptive statistics, analyzing 201 cantons of Ecuador using the INEC Statistical Registry of General Deaths 2020. This, applying a multiple linear regression model, found that there is a statistically significant association between COVID-19 mortality and altitude in cantons located at more than 2,000 meters above sea level.

Statistics has been applied to the analysis of various areas such as education (9–12), administration (13,14) and health (15,16). Within the applications of the bootstrap method we find its use to find confidence intervals, for example, in education.(17), psychology(18), health(19,20) among others.

In the statistical area, Garca, J. & Fuentes, H. carry out an analysis of COVID-19 and the probability of Death(21). Applying to the analysis of mental health and COVID, the structural equation model has been used for the mediator analysis. The indirect effects were analyzed using the bootstrap method. (22) However, it differs from the bootstrap that is proposed in our research. An interesting example is the study of mortality in patients over 65 years of age hospitalized for COVID-19, using bivariate analysis and multivariate logistic regression to determine the risk factors associated with hospital mortality. (23).

One of the most important statistics tasks is to determine, as precisely as possible, the probability distribution of certain random variables of interest. All this leads to the following research objective to determine the age groups with the highest vulnerability to death from COVID-19 in the mountains of Ecuador in the period 2020-2021. Therefore, the databases of the Statistical Registry of General Deaths will be taken into account. The year 2020 and 2021 of the National Institute of Statistics and Census of Ecuador. A statistical algorithm based on the bootstrap method will be used to calculate the confidence interval for the proportions. (24,25) which will be applied to determine the age groups of those who died of COVID-19 and other causes in Ecuador in the periods 2020 and 2021.

2. METHODOLOGY

The focus of this research is quantitative since a statistical analysis will be carried out based on the data and, based on these results, the age groups most vulnerable to dying from COVID will be interpreted. The type of research will be descriptive, since the age groups most vulnerable to dying from COVID-19 will be described based on data on deaths in the Ecuadorian highlands in 2020 and 2021.

The population is the databases of people who died in the years 2020 and 2021 and the sample in our case is the database limited to people who died in the Ecuadorian highlands. These databases have information on the deceased, containing 45 variables of which the age at death, sex, ethnicity, place of occurrence of the deceased and basic cause of death stand out. They can be downloaded at(4)

2.1. Procederes

- Download of the INEC ECUADOR database, analysis of the database and its variables, analysis of the variables to be taken into account, application of refinement techniques of the obtained database, and construction of a new database with the data necessary for the study.
- Statistical comparison of deaths from COVID and other causes in the periods 2020 and 2021 in the Ecuadorian highlands. The main statistics of deaths from COVID and other causes in the periods 2020 and 2021 in Ecuador will be analyzed, grouping them by year and sex.
- Apply the bootstrap method to calculate the confidence interval for the proportions of the age groups of those who died from COVID-19 and other causes in the Ecuadorian highlands in the periods 2020 and 2021. The method is implemented in R Studio.

2.2. Investigation Techniques and Instruments

INEC physical forms

For data collection, there is the "Statistical Report on General Deaths" (physical and electronic), which is designed by the INEC together with the MSP and is the instrument for collecting information.

The forms are distributed to each one of the Zones of the Ministry of Public Health, to the Hospitals and Clinics of the public and private sectors, Forensic Services, as well as to the offices belonging to the General Directorate of Civil Registry, Identification and Documentation. from the country(26).
2.3. Bootstrap algorithm to calculate the confidence interval for proportions

The Bootstrap method is a statistical procedure that is used to approximate the sampling distribution (usually from a statistic) \((27)\). To do this, we proceed by resampling, that is, obtaining samples using some random procedure that uses the original sample. Its main advantage is that it does not require hypotheses about the mechanism that generates the data. The interval estimation does not provide a point estimate with a number but gives us an interval at which a certain confidence is deposited that it contains the parameter. Thus, instead of saying that the graduation average of a university degree "should be close to 6.50", an interval will be constructed, which will say, for example, "there is a 95% confidence that is deposited that it contains the parameter. Thus, instead of saying that the graduating average of a university degree "should be close to 6.50", an interval will be constructed, saying that the graduation average of a university degree "should be close to 6.50", an interval will be constructed, saying that the graduation average of a university degree "should be close to 6.50", an interval will be constructed, saying that the graduation average of a university degree "should be close to 6.50", an interval will be constructed, saying that the graduation average of a university degree "should be close to 6.50", an interval will be constructed, saying that the graduation average of a university degree "should be close to 6.50", an interval will be constructed, saying that the graduation average of a university degree "should be close to 6.50", an interval will be constructed, saying that

Based on the general aspects of this method, the confidence interval for the proportions of the nominal or ordinal scales of a qualitative variable is calculated using the steps in the following algorithm. \((24,25)\):

1. Given the sample:
   \[ \bar{X} = (X_1, X_2, \ldots, X_n) \]  
   Where is a quantitative variable, has a scale value \(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k\) \(\bar{X}_n\).

2. For each \(i = 1, 2, \ldots, n\) throw
   \[ U_i \sim U(0,1) \text{ and do } X_i^* = X_i \left[ u_{[1]} + 1 \right] \]  

3. Get
   \[ \hat{p}_j = \frac{1}{n} \sum_{i=1}^{n} Y_i^* \]  
   \[ Y_i^* = \left\{ \begin{array}{ll} 1 & \text{if } X_i = \varepsilon_j \\ 0 & \text{if } X_i \neq \varepsilon_j \end{array} \right. \]

4. Calculate the bootstrap statistic:
   \[ j = 1, 2, \ldots, kR_j^* = \left( \hat{p}_j - p_j \right) \left( \frac{1 - p_j}{n} \right) \]  
   \[ \sqrt{ \beta \left( 1 - p_j \right) / n} \]  

5. Repeat steps 2, 3, and 4 \(B\) times to obtain bootstrap replicas
   \[ \ldots, j = 1, 2, \ldots, kR_j^{*(1)} R_j^{*(B)} \]  

6. Order the values of the bootstrap statistic in increasing order:
   \[ j = 1, 2, \ldots, k ; b = 1, 2, \ldots, B \]
   \[ BR_j^{* (b)} \]

7. Calculate the lower and upper critical points of the significance level \(\alpha\):
   \[ \text{pcinf} = \left[ R_j^{* (b)} \right]_{t = \alpha / 2} \]
   \[ \text{pcsup} = \left[ R_j^{* (b)} \right]_{t = 1 - (1 - \alpha / 2)} \]

8. Calculate the lower and upper limits of the proportions of the of the confidence interval of nominal or ordinal scale of a qualitative variable, with the significance level \(\alpha\):
   \[ \text{lim.inf}j = \text{pcinf} + \left( \frac{p_j (1 - p_j)}{n} \right) \]
   \[ \text{lim.sup}j = \text{pcsup} + \left( \frac{p_j (1 - p_j)}{n} \right) \]

Where is the scale sample proportion of the \(j\)-th of the qualitative variable, \([x]\) an integer part of \(x\), is the uniform distribution on the interval \((0, 1)\). Therefore, the confidence interval corresponding to the proportion of the \(j\)-th scale is given by \(\hat{p}_j \bar{X}_n(0,1)\)

\[ \text{IC}_j = \left( \text{lim.inf}j, \text{lim.sup}j \right) \text{ for } j = 1, 2, \ldots, k \]

3. RESULTS AND DISCUSSION

Two databases of the Statistical Registry of General Deaths were used, which correspond to the vital events of deaths that occurred and/or were registered in the national territory, in the years 2020 and 2021. Within these databases, 4 variables are used: sex, age, cause103, and prov_fail. Of which the first two are numerical, and the others are categorical, cause103 is broken down from the Condensed List of 103 Groups - Tenth Revision - ICD - 10 which has classified diseases. In this list, as of 2020, category 104 was added, which corresponds to 104 confirmed and suspected COVID-19.

Within the age variable, there are data 999 which represents the lack of information on age, for which we proceeded to eliminate these data, since the percentage within the bases is 0.018\%, and 0.046\% for 2020 and 2021 respectively, which does not alter the database. When analyzing age, there are no atypical data since in the three databases there are deaths at a very advanced age. Within the age variable, there is data 999 which represents the lack of information on age, for which we proceeded to eliminate these data, since the percentage within the bases is 0.018\%, and 0.046\% for 2020 and 2021 respectively, which does not alter the database. When analyzing age, there are no atypical data since in the three databases there are deaths at a very advanced age.

Therefore, 2 new databases were built with the variables sex, age, cause103, and prov_fail Taking into consideration only the deceased in the provinces of the Ecuadorian highlands.

3.1. Statistical determination of COVID-19 mortality in age groups in the Ecuadorian Highlands
When analyzing the main statistics (mean, standard deviation, median, range, asymmetry, and kurtosis) of COVID deaths and from other causes in the periods 2020 and 2021 in Ecuador, grouping them by year and sex, it was found that the main statistics of The ages of death from COVID-19 and other causes in Ecuador remained relatively stable; no significant changes were observed in age patterns between men and women for most of the causes analyzed.

It is important to note that these results are based solely on the calculated descriptive statistics and do not take into account other factors such as mortality rates, geographic distribution, or changes in the population.

The evolution of the main causes of death in the highlands of Ecuador since 2020 and 2021 shows that COVID-19 had a significant impact, representing 23.62% (10,522) and 22.60% (10,307) of all deaths, respectively. This increase in deaths nationwide from 75,228 (2019) to numbers greater than 100 thousand deaths(4).

There was a higher number of deaths of men compared to women in the years 2020 and 2021. In general, there is a significant impact of COVID-19 on deaths in Ecuador. Furthermore, cardiovascular diseases appear to be the main cause of death in the country.

### 3.2. Age groups of death by COVID

The age groups of people who died from COVID-19 are determined, which are distributed from childhood to the elderly; the number of people that make up these age groups can be seen in Table 1.

Once the age groups were determined, the Bootstrap Algorithm was applied to calculate the confidence interval for the proportions of people who died from COVID-19 in both 2020 and 2021, with 1000 re-samples, obtaining the following results (Table 2).

#### Table 1. Distribution of age groups.

<table>
<thead>
<tr>
<th>Age groups</th>
<th>2020 Male</th>
<th>2020 Female</th>
<th>2020 Total</th>
<th>2021 Male</th>
<th>2021 Female</th>
<th>2021 Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Childhood</td>
<td>39</td>
<td>25</td>
<td>64</td>
<td>23</td>
<td>14</td>
<td>37</td>
</tr>
<tr>
<td>Youths</td>
<td>56</td>
<td>31</td>
<td>87</td>
<td>33</td>
<td>24</td>
<td>57</td>
</tr>
<tr>
<td>Adults</td>
<td>1919</td>
<td>783</td>
<td>2702</td>
<td>1988</td>
<td>949</td>
<td>2937</td>
</tr>
<tr>
<td>Older adults</td>
<td>4989</td>
<td>2680</td>
<td>7669</td>
<td>4360</td>
<td>2916</td>
<td>7276</td>
</tr>
<tr>
<td>Total</td>
<td>7003</td>
<td>3519</td>
<td>10,522</td>
<td>6404</td>
<td>3903</td>
<td>10,037</td>
</tr>
</tbody>
</table>

| Note: Children from 0 to 14 years, Youth from 15 to 27 years, Adults from 28 to 59 years, Adults over 60 years |

#### Table 2 Confidence intervals of the age groups of people who died of COVID

<table>
<thead>
<tr>
<th>Age group</th>
<th>2020 Lower limit</th>
<th>2020 Upper limit</th>
<th>2020 Point Estimate</th>
<th>2021 Lower limit</th>
<th>2021 Upper limit</th>
<th>2021 Point Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Childhood</td>
<td>0.45%</td>
<td>0.62%</td>
<td>24.95%</td>
<td>0.23%</td>
<td>0.39%</td>
<td>27.62%</td>
</tr>
<tr>
<td>Youths</td>
<td>0.74%</td>
<td>0.97%</td>
<td>26.46%</td>
<td>0.46%</td>
<td>0.68%</td>
<td>29.30%</td>
</tr>
<tr>
<td>Adults</td>
<td>0.60%</td>
<td>0.80%</td>
<td>25.71%</td>
<td>0.35%</td>
<td>0.54%</td>
<td>28.46%</td>
</tr>
<tr>
<td>Older Adults</td>
<td>0.54%</td>
<td>0.80%</td>
<td>25.71%</td>
<td>0.74%</td>
<td>0.97%</td>
<td>26.46%</td>
</tr>
</tbody>
</table>

| Note: 95% confidence that the interval. CI (α = 0.05). Childhood 0 to 14 years, Youth from 15 to 27 years, Adults 28 to 59 years, Older Adults over 60 years and older |

Adjusted confidence intervals suggest that there is 95% confidence that the age group most likely to die from COVID-19 is the Older Adults group. This is because the confidence interval for the proportion of deaths in this group CI (α = 0.05) = [0.7201; 0.7339] % with a point estimate of 72.70% in 2020 and CI (α = 0.05) = [0.6977; 0.7144] % with a point estimate of 70.61% in 2021 are wider than the intervals of the other groups ("Children", "Youth", "Adults") of both years. Following our results, the age ranges most affected by COVID-19 are people over 60 years of age. (29) and 65 years(twenty-one).

The "Childhood" age group is the least likely to die from COVID-19 since the adjusted confidence intervals suggest that there is a 95% confidence that 0.60% will be concentrated in 2020 and 0.35% in 2021 of deaths.

Therefore, an age subgroup was also determined within the group of Older Adults (see Table 5), with intervals of 5 years up to 80 years, then 10 years and 90 years onwards. By applying the Bootstrap Algorithm to calculate the confidence interval for the proportions of people who died from COVID-19 in both 2020 and 2021, with 1000 re-samples, obtaining the following results (Table 3).
Table 3  Confidence intervals of the age subgroups of older adults who died of COVID

<table>
<thead>
<tr>
<th>Age group</th>
<th>0-60</th>
<th>60-65</th>
<th>65-70</th>
<th>70-75</th>
<th>75-80</th>
<th>80-90</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2020</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower limit</td>
<td>26.24%</td>
<td>11.06%</td>
<td>12.35%</td>
<td>11.65%</td>
<td>11.82%</td>
<td>17.84%</td>
<td>4.57%</td>
</tr>
<tr>
<td>Upper limit</td>
<td>27.88%</td>
<td>12.23%</td>
<td>13.61%</td>
<td>12.91%</td>
<td>13.10%</td>
<td>19.26%</td>
<td>5.36%</td>
</tr>
<tr>
<td>Point Estimate</td>
<td>27.06%</td>
<td>11.64%</td>
<td>12.98%</td>
<td>12.28%</td>
<td>12.46%</td>
<td>18.55%</td>
<td>4.96%</td>
</tr>
<tr>
<td><strong>2021</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower limit</td>
<td>28.51%</td>
<td>11.40%</td>
<td>11.93%</td>
<td>11.62%</td>
<td>10.54%</td>
<td>16.13%</td>
<td>5.09%</td>
</tr>
<tr>
<td>Upper limit</td>
<td>30.32%</td>
<td>12.66%</td>
<td>13.21%</td>
<td>12.87%</td>
<td>11.89%</td>
<td>17.57%</td>
<td>6.01%</td>
</tr>
<tr>
<td>Point Estimate</td>
<td>29.41%</td>
<td>12.03%</td>
<td>12.57%</td>
<td>12.24%</td>
<td>11.22%</td>
<td>16.85%</td>
<td>5.55%</td>
</tr>
</tbody>
</table>

Note: 95% confidence that the interval. CI (α = 0.05)
In graphs 1 and 2, we can notice that the confidence intervals of both years overlap in each age group. This suggests that there is no statistically significant difference in the proportions of deaths from COVID-19 between the two years for the age groups represented.

It is important to note that the confidence intervals indicate the uncertainty associated with the estimates, and in this case, there is no clear separation between the two-year confidence intervals. This means that we cannot say with certainty that there are significant differences in the proportions of deaths between the two years for these specific age groups.

The “Bootstrap to calculate the confidence interval for proportions” method, is a robust method since the calculations obtained agree with investigations of vulnerability to death from COVID-19, which have applied various methods(19).

**CONCLUSIONS**

Adjusted confidence intervals suggest that there is 95% confidence that the age group most likely to die from COVID-19 is the "AMayor" (Older Adults) group. This is because the confidence interval for the proportion of deaths in this group CI ($\alpha = 0.05$) = [0.7201; 0.7339] % with a point estimate of 72.70% in 2020 and CI ($\alpha = 0.05$) = [0.6977; 0.7144] % with a point estimate of 70.61% in 2021 are wider than the intervals of the other groups.

Within the age subgroups of "Older Adults", the adjusted confidence intervals indicate that there is a 95% confidence that the age subgroup most likely to die from COVID-19 is the 80 to 90 years with a CI ($\alpha = 0.05$) = [0.1784423; 0.1926024] % with a point estimate of 16.13% in 2020 and CI ($\alpha = 0.05$) = [0.1612858; 0.1757478] % with a point estimate of 16.85% in 2021. In second place is the subgroup of 65 to 70 years with a CI ($\alpha = 0.05$) = [0.1234954; 0.1361284] % with a point estimate of 12.98% in 2020 and CI ($\alpha = 0.05$) = [0.1193482; 0.1320973] % with a point estimate of 12.57% in 2021.

**Acknowledgements.**

Secretaría de Educación Superior, Ciencia, Tecnología e Innovación del Ecuador: CZ03-000392-2918;

**References**


