

The Relationship Between the Number of New COVID-19 Cases and the Number of Patients on Ventilators*

Simple Linear Model Project

Siling Guo

September 23, 2025

This study explores the relationship between the number of new COVID-19 cases and the number of patients on ventilators. A simple linear regression model was established based on the data provided by the County of Santa Clara Public Health Department between 2020-03-27 and 2021-05-23. The model shows a significant positive correlation between the two variables; there was an expected increase of 0.97 ventilator patients for each additional COVID-19 case ($\beta_1=0.97$, $p<0.001$). The number of new COVID-19 cases is a significant predictor of ventilator demand (R Squared=0.62). This model can help predict the potential demand for medical devices during a pandemic.

1 Introduction

The 2020-2023 COVID-19 pandemic magnified the imbalance between the limited amount of medical resources and the surging number of new incoming medical patients worldwide. The pandemic significantly disrupted the health service system (Filip et al. 2022), especially in resource-limited countries. Even though the COVID-19 pandemic has ended, understanding the correlation between new COVID-19 cases and the number of patients requiring ventilators is useful for analyzing the potential demand for medical devices during a pandemic. This project aims to analyze the impact of the number of COVID-19 cases on ventilator usage using the linear regression model based on the data set published by the County of Santa Clara Public Health Department from 2020-03-27 to 2021-05-23 (“COVID-19 Hospitalizations by Date | County of Santa Clara” n.d.). First, we will discuss the research background and proposal followed by the usage and processing of the data. Next, we will explain the setup and

*Project repository available at: <https://github.com/silingguo/MATH261A-project-template>.

validation of the model and its limitations before showing the results with relevant tables and graphs along with our conclusion. Through analysis, we found a significant positive correlation between the increase in the number of COVID-19 cases and the growth in the number of patients requiring ventilators.

2 Data

The observational unit in this study is the daily hospital data from 2023/3/27-2021/5/23, provided by the County of Santa Clara Public Health Department. The sample size is 423 days. The two key variables are the number of new COVID-19 cases and the number of patients on ventilators, which are continuous variables measured in people. The data is generated by the County of Santa Clara Public Health Department and posted on the County of Santa Clara Open Data Portal. There are no missing values in the core variables. Besides, there are a small number of outliers that were retained as they likely reflect peak pandemic conditions. Overall, the data are clean, so the only processing performed was converting date strings into a datetime format. According to the data summary (Table 1), the daily average number of new COVID-19 cases was 28.4 with a standard deviation of 28 and a maximum of 124. This data has significant fluctuations. The daily average number of patients using ventilators is 179.7 with a standard deviation of 34.5. This data has smaller fluctuations. Overall, the trend in the number of patients using ventilators is generally consistent with the trend in the number of new COVID-19 cases (Figure 1). Furthermore, due to the unique course of the COVID-19 disease, not all confirmed patients require ventilators immediately, resulting in a delayed peak in demand for medical equipment. We can observe this time delay in the data. This trend allows us to predict demand for medical equipment.

Table 1: Data Summary of Covid-19 New Cases and Number of patients on ventilators

Data Summary of Covid-19 New Cases and Number of patients on ventilators							
Mean (covid_new)	Mean (vents_pts)	SD (covid_new)	SD (vents_pts)	Max (covid_new)	Max (vents_pts)	Min (covid_new)	Min (vents_pts)
28.38958	179.7419	28.0269	34.50896	124	283	0	123

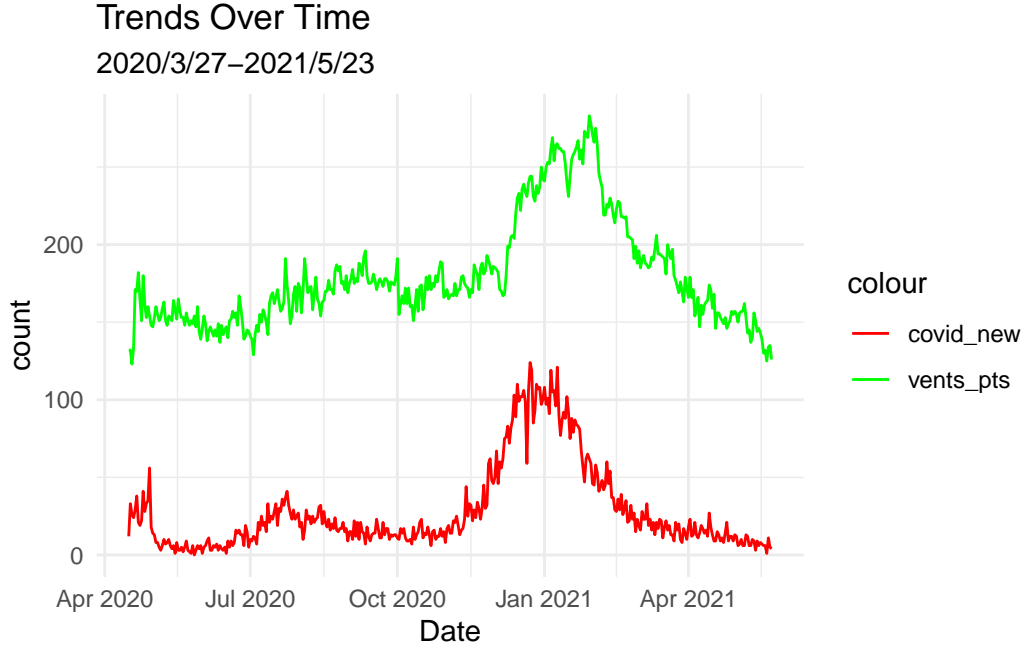


Figure 1: Trends Over Time

3 Methods

The number of new COVID-19 cases and the number of patients on ventilators are continuous variables. According to the trend chart of the data, it suggests a potential linear correlation between two variables. Therefore, we apply a simple linear regression model. $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ Here, y_i covid_new: The number of new COVID-19 cases on a given day X_i (Vents_pts): The number of patients on ventilators on a given day β_0 (intercept):The average number of patients on ventilators when there are no new COVID-19 cases. β_1 (slope): The expected change in the number of patients using ventilators for each additional new COVID-19 case. In this model, we use the ordinary least squares (OLS) method to estimate the parameters, and apply the `lm()` function in R to fit the model. The OLS method assumes that the error term ε_i follows a normal distribution. The QQ-Plot (Figure 2) indicates that the majority of the residuals closely align with the reference line. A few data points fall below the expected quantiles on the left, while a few data points fall above the expected quantiles. This suggests that there might be outliers. In general, the error terms are approximately normally distributed.

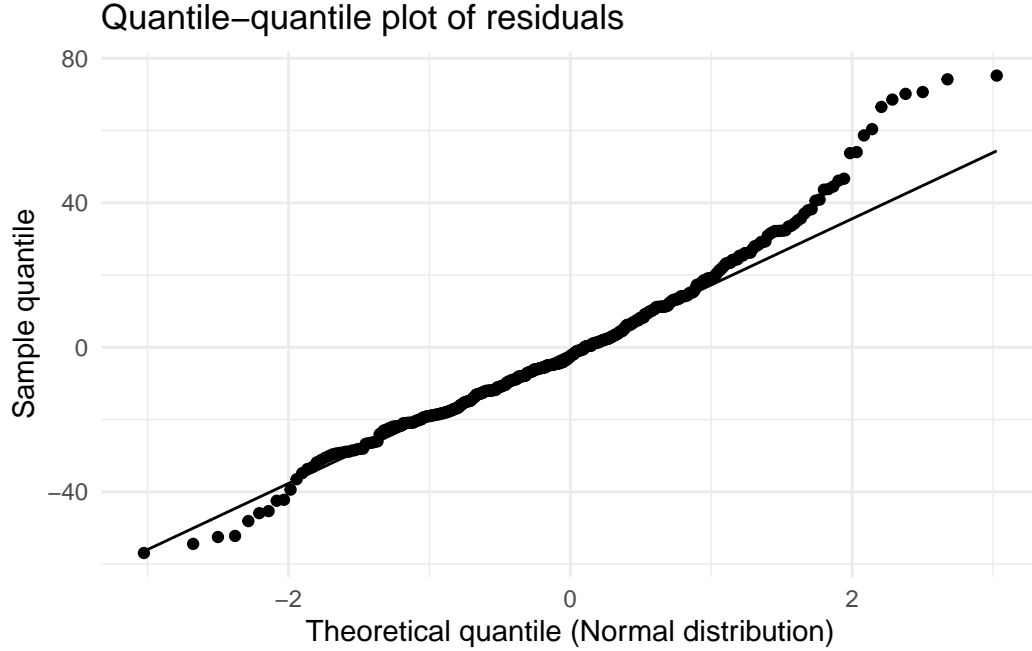


Figure 2: Quantile-quantile plot of residuals

4 Results

Figure 3 indicates that the number of patients on ventilators becomes increasingly dispersed as the number of new COVID-19 cases rises, and the model demonstrates a moderate fit. From the data summary (Table 2), we find $F(1, 401) = 655.6$, $p\text{-value} < .001$, and estimate of $\text{covid_new} = 0.97$, indicating that the linear regression model is highly significant, and the predictor variable (x) can explain the variation in the dependent variable (y). Multiple R-squared: 0.6205 and Adjusted R-squared: 0.6195 indicate a model fit of approximately 60%. This result is likely due to the failure to account for the lag effect in ventilator usage, the proportion of severe COVID-19 cases, and seasonal factors, which can all affect ventilator use. This study's results show that the number of newly confirmed COVID-19 cases is a key factor influencing the demand in ventilator usage. Specifically, for every new COVID-19 patient during this time period, there is an expected increase of 0.97 ventilator patients. This finding provides insights into the relationship between epidemics the medical equipment demand, contributing to better planning for the stockpiling, deploying and resource planning of medical equipment during epidemics. We can improve the model's fit by adding in additional unaccounted factors.

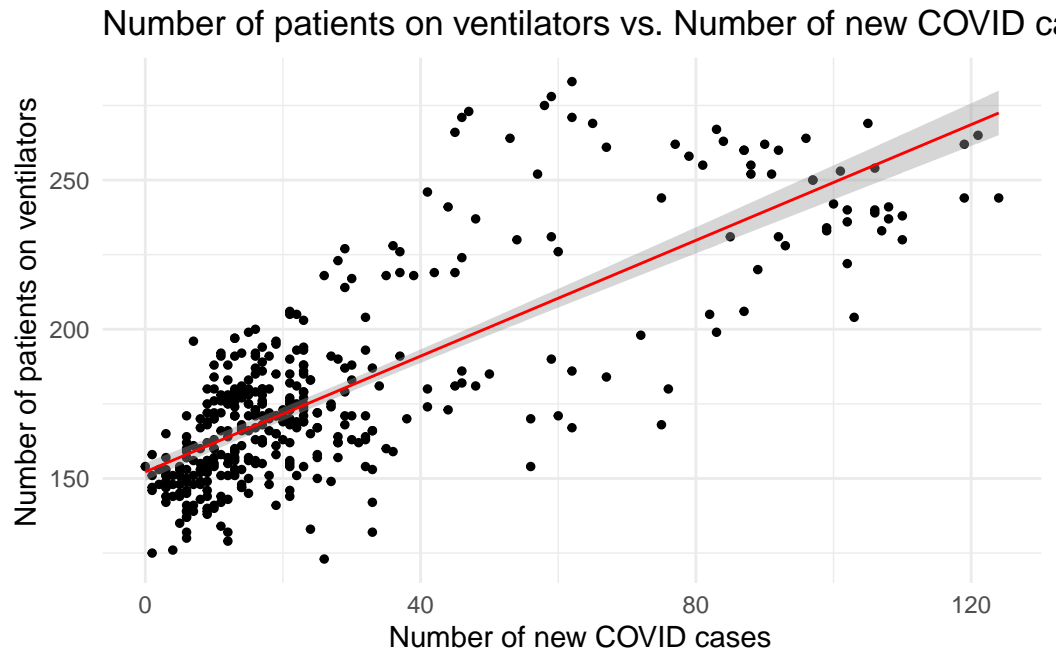


Figure 3: Number of patients on ventilators vs. Number of new COVID cases

Table 2: Summary of Simple Linear Regression Model

Summary of Simple Linear Regression Model				
Variable	Estimate	Standard error	T-value	P-value
(Intercept)	152.2070210	1.51018188	100.78721	<0.001
covid_new	0.9698952	0.03787873	25.60527	<0.001
Residual Standard Error: 21.3 R-squared: 0.62 Adjusted R-squared: 0.62 F-statistic: 655.63				

References

- “COVID-19 Hospitalizations by Date | County of Santa Clara.” n.d. Accessed September 23, 2025. https://data.sccgov.org/COVID-19/COVID-19-hospitalizations-by-date/5xkz-6esm/about_data.
- Filip, Roxana, Roxana Gheorghita Puscaselu, Liliana Anchidin-Norocel, Mihai Dimian, and Wesley K. Savage. 2022. “Global Challenges to Public Health Care Systems During the COVID-19 Pandemic: A Review of Pandemic Measures and Problems.” *Journal of Personalized Medicine* 12 (8): 1295. <https://doi.org/10.3390/jpm12081295>.