**COVID-19 Mortality Analysis**

This project is my attempt to explore real COVID-19 case data and answer a simple but important question: **were certain groups of people more likely to die from the virus?**

The dataset has just over a thousand patient records. Each row represents a case, with information like age, gender, country, when symptoms started, whether they were hospitalized, and ultimately if they survived or died. Before diving into any analysis, I had to clean things up. The “death” column, for example, was messy sometimes it showed dates, sometimes it was just a number. To make this easier to work with, I created a new variable called death dummy where **1 = the patient died** and **0 = they survived**.

Once that was done, I started looking at patterns.

**What I Found**

* **Mortality rate**  
  Out of the 1,085 cases, around **6% of patients died**. It’s a sobering figure, even in this limited dataset.
* **Age matters — a lot**  
  When I compared the average age of people who survived versus those who didn’t, the difference was huge:
  + Survivors: about **48 years old** on average
  + Non-survivors: about **69 years old** on average  
    To make sure this wasn’t just a random fluke, I ran a statistical test (a t-test). The result gave me a p-value so small it was basically zero, which means the difference is very unlikely to be by chance. **Conclusion: older patients were significantly more likely to die.**
* **Gender plays a role too**  
  Then I looked at men and women separately.
  + About **8.5% of men** in the dataset died
  + Only **3.7% of women** died  
    Again, I ran a t-test. This time the p-value came out to about 0.002, which is still well below the usual 0.05 cutoff. That tells us the gender difference is statistically significant too. **Conclusion: men had a higher risk of dying compared to women.**

**Why the t-tests matter**

The averages and percentages already point in a clear direction, but the t-tests are what make the claims solid. They check whether the differences we see (older vs younger, men vs women) are just noise or if they really mean something. Both tests came back strongly significant, which gives us confidence that age and gender truly affected outcomes in this dataset.

**Takeaway**

From this analysis, it’s clear that **age and gender both played a major role in COVID-19 mortality**. Older patients were far more vulnerable, and men were more likely to die than women. These patterns in the data line up with what health organizations and studies later confirmed on a global scale.

All of the work was done in **R**, using the Hmisc package for quick descriptive stats and base R functions for hypothesis testing.

Bottom of Form