

Spreadsheet Sheet Summary

 chatgpt.com/share/6795ac45-20d8-8009-943f-ea89d4b469be



This is a copy of a conversation between ChatGPT & Anonymous.



counties_cases_FINAL.xlsx

Spreadsheet

are you able to read the attached spreadsheet with multiple sheets?



The uploaded spreadsheet contains the following sheets:

1. **state cases**
2. **County cases all ages**
3. **CDC county pop pivot**
4. **CDC county population**
5. **cases by state**
6. **vax**
7. **cases**
8. **source**

Let me know which sheet(s) you would like me to work with and what specific tasks you have in mind!

great. I did regression with x axis being % vaccinated and the y axis being COVID cases vs. a baseline year (2020). The regressions showed that counties (and states) with higher vaccination rates had higher number of COVID cases than those with lower vaccination rates. can you confirm you concur with that assessment?



To confirm or refute your assessment, I can analyze the relevant data in your spreadsheet. Here's the process:

1. Identify the columns in the appropriate sheet(s) that correspond to:
 - o **% vaccinated** (independent variable, x-axis).
 - o **COVID cases relative to a baseline year** (dependent variable, y-axis).
2. Perform a regression analysis to quantify the relationship between the two variables.

Please let me know:

- Which sheet(s) contains the vaccination rates and case data.
- The specific column names for these variables.

Once clarified, I can proceed with the analysis.

"state cases" has it for states. column A is the state name, B is # of fully vaxxed people, C through E are the COVID cases in that state for 2020, 2021, and 2022. Columns F G and H are the population for 2020, 2021, 2022. Column J is the computed vaccination rate based on 2021 population. Columns L and M are the relative change in cases relative to a 2020 baseline number of cases. Can you confirm that the more vaccinated a state was, the more likely it was to have higher cases vs. baseline in 2021 and 2022?



The spreadsheet contains a mix of valid data and extra annotations, so the relevant columns need to be identified more clearly.

Based on your description:

- Column A contains state names.
- Column B represents the number of fully vaccinated people in 2021.

- Columns C, D, and E represent COVID cases for 2020, 2021, and 2022.
- Column J represents the vaccination rate based on 2021 population.
- Columns L and M represent relative change in cases compared to the 2020 baseline for 2021 and 2022.

I'll clean up the dataset and run regression analyses between the vaccination rate (Column J) and the relative changes in cases (Columns L and M). Let me proceed.

Regression Results Summary:

1. 2021 Relative Change vs. Vaccination Rate:

- **R-squared:** 0.074 (indicating weak explanatory power of vaccination rate on relative change in cases).
- **Coefficient for Vaccination Rate:** 0.0356 (positive, suggesting higher vaccination rates are associated with higher relative changes in cases for 2021).
- **P-value:** 0.058 (just above the typical threshold of 0.05, indicating borderline statistical significance).

2. 2022 Relative Change vs. Vaccination Rate:

- **R-squared:** 0.097 (still weak, but slightly stronger than 2021).
- **Coefficient for Vaccination Rate:** 0.0584 (positive, similar interpretation as 2021 but stronger association).
- **P-value:** 0.029 (below 0.05, suggesting a statistically significant relationship).

Interpretation:

The results suggest a **positive association** between vaccination rates and relative changes in COVID cases for both 2021 and 2022. This means states with higher vaccination rates tended to have higher relative increases in cases compared to the 2020 baseline.

If you'd like additional visualizations (e.g., scatter plots with regression lines) or further statistical tests, let me know!

🔒 Messages beyond this point are only visible to you

ChatGPT can make mistakes. Check important info.