

General description of the Project

Imagine you are a primary researcher who spends a great deal of time in meetings reviewing and discussing quantitative research. You likely hear from stakeholders that they want to compare significance across two distinctive populations. In your reporting you have represented your sig testing based on 95% significance. However often times you make a correction within your stats package and there are borderline cases that are called out. In most tools in the space, you select your significance level and then put in the sample size and percentage.

This application will take the sample sizes and percentages of comparable populations to determine if the difference is statistically significant and if so at what confidence interval. I also envision an output of a table that provides the level of different it would take to be statistically significant at confidence intervals between 80% and 99%. and loop through the calculations to come up with the level at which the numbers are significantly different. This arms researchers with an on-the-fly tool to help conversations.

After which the stakeholder asks to have the specific data in a slide. I will work to put in use python-pptx to export the information from the application to a power point slide that can be edited and altered as needed. This will give the research the ability to leave the meeting with an updated set of slides that they can promptly send to meeting participants as a follow up.

I plan to use Tkinter for the graphical interface of the application, where as you have 4 total input boxes (Sample size A, Percentage A, Sample Size B, Percentage B). This will also have a compute action button that when clicked cycles through significance calculations. The graphical interface would also return a statement if significant or not and at what level. This will likely require the scipy.stats package to accomplish the calculations.

Task Vignettes

Task Vignette 1: Evaluating Customer Satisfaction Differences Across Two Regions

Scenario:

James the researcher is analyzing customer satisfaction data from two regions. Region A had a survey response of 1,400 people, with 72% indicating satisfaction. Region B

had a survey response of 1,600 people, with 75% satisfaction. James wants to know if the satisfaction rates differ significantly across regions and at what confidence level as a result of a question from Tina who is the Global head of service.

Goal:

Determine whether there is a statistically significant difference in satisfaction rates between Region A and Region B, and identify the level of confidence at which the difference is significant.

Steps:

1. Input "1400" in the "Sample Size A" field.
2. Input "72" in the "Percentage A" field.
3. Input "1600" in the "Sample Size B" field.
4. Input "75" in the "Percentage B" field.
5. Press the "Compute" button to calculate the results.
6. View the significance output to see at what confidence level the difference is statistically significant.

Task Vignette 2: Understanding the Table Output for Confidence Intervals

Scenario:

John is preparing for a meeting with the head of product Sampath to discuss the results of a large-scale product satisfaction survey. Two product lines, A and B, were rated by different groups. Product A had 2,500 respondents with 68% rating it favorably, while Product B had 2,800 respondents with 72% providing favorable feedback. John wants to be ready for any questions on the survey by having a detailed breakdown of the confidence intervals at which the difference in satisfaction becomes significant.

Goal:

Generate a table that shows the confidence intervals (from 80% to 99%) and the minimum percentage difference required for the results to be statistically significant at each level.

Steps:

1. Enter "2500" in the "Sample Size A" field.
2. Enter "68" in the "Percentage A" field.
3. Enter "2800" in the "Sample Size B" field.
4. Enter "72" in the "Percentage B" field.
5. Click the "Compute" button to run the significance calculations.
6. Review the table output, which displays the required percentage difference for significance at confidence intervals from 80% to 99%.
 - For each row of the table:
 - The first column lists the confidence level (80%, 85%, 90%, 95%, 99%).

- The second column shows whether the difference in percentages between Product A and Product B is statistically significant at that level.
 - The third column shows the minimum percentage difference required for the results to be significant at that level.
7. Use the table to explain to Sampath at what confidence levels the satisfaction difference between Product A and Product B is significant, and how much the percentages would need to differ to reach significance at higher confidence levels (e.g., 99%).

Example Table Output:

Confidence Level	Statistically Significant	Required % Difference
80%	Yes	1.5%
85%	Yes	1.8%
90%	Yes	2.0%
95%	Yes	2.5%
99%	No	3.0%

Explanation to Stakeholders:

- "At an 80% confidence level, the 4% difference between Product A and Product B (68% vs 72%) is significant, but to reach 99% confidence, the difference would need to be at least 3%."

Technical flow

1. Input Capture: User enters sample sizes and percentages for two populations.
2. Z-Score Calculation: Backend computes the z-score using the pooled proportion formula.
3. P-Value Calculation: Backend computes the p-value based on the z-score and checks against predefined confidence levels.
4. Display Results: GUI displays whether the result is significant and at which confidence level.
5. Table Output: Optional table output shows significance at various confidence levels and required percentage differences.
6. Error Handling & Edge Cases: Validate that the user has entered valid numeric inputs in the fields. If any of the input fields are empty, non-numeric, or negative: Display an error message in the GUI asking for proper input

Somewhat unknown: Unsure of the complete GUI workflow and will need to think through all of the