RPA\_Process\_Suitability\_Checklist All Teams.xlsx

Spreadsheet

Is there a statistically significant difference between the human score, and the GPT scores in the spreadsheet?

ChatGPT

The t-statistic for the comparison between human scores and GPT scores is approximately 2.40, and the p-value is about 0.018. Since the p-value is less than 0.05, we can conclude that there is a statistically significant difference between the human scores and the GPT scores in your dataset. ​

​



tabulate the scores by team and create a table

ChatGPT

Here is the table showing the average Human and GPT scores, tabulated by each team:

markdown

Team Human Score GPT Score

1 1.5 1.5

2 1.8 1.5

3 2.0 1.6

4 2.0 1.8

5 2.1 1.6

This table represents the mean scores for both Human and GPT assessments across different teams. ​

​



are the scores independent of team?

ChatGPT

The ANOVA test resulted in an F-statistic of approximately 1.51 and a p-value of about 0.216. Since the p-value is greater than 0.05, we can conclude that there is no statistically significant difference in human scores across different teams. This suggests that the scores are independent of the team. ​

​



redo the t-test and the ANOVCA test after removing the scores from Team 1

ChatGPT

After removing the scores from Team 1:

1. **T-Test (Human vs GPT Scores):**
   * T-statistic: Approximately 2.80
   * P-value: About 0.006

The p-value is less than 0.05, indicating a statistically significant difference between human scores and GPT scores without considering Team 1.

1. **ANOVA Test (Human Scores Across Teams):**
   * F-statistic: Approximately 0.39
   * P-value: About 0.759

The p-value is greater than 0.05, suggesting that there is no statistically significant difference in human scores across the remaining teams, even after removing Team 1. This implies that the human scores are still independent of the team. ​

1. ​



do the humas score lower or higher than the machine?

ChatGPT

The t-test conducted earlier, comparing human scores and GPT scores, resulted in a significant p-value, indicating a difference between the two sets of scores. To determine whether human scores are generally higher or lower than the GPT scores, we can look at the means of each.

Let's calculate the average scores for both human and GPT ratings, excluding Team 1, to see which is higher.

After excluding Team 1, the average human score is approximately 1.98, while the average GPT score is around 1.63. This indicates that, on average, the human scores are higher than the GPT scores in your dataset. ​

​



Did you take into account to paired nature of the data?

ChatGPT

The "paired" nature of the data, where each human score corresponds to a GPT score for the same characteristic, should indeed be considered in the statistical analysis. Initially, I conducted a two-sample t-test, which assumes independent samples. However, for paired data like this, a paired sample t-test is more appropriate, as it accounts for the fact that each pair of scores (human and GPT) is related to the same characteristic and team.

Let's perform a paired sample t-test on the data, excluding Team 1, to accurately compare the human scores with the GPT scores.

Using a paired sample t-test, which accounts for the paired nature of the data, the t-statistic is approximately 3.01, and the p-value is about 0.0046. This p-value is still less than 0.05, indicating a statistically significant difference between the human scores and the GPT scores, even when considering the paired nature of the data. This reinforces the conclusion that, on average, human scores differ significantly from GPT scores in your dataset, with human scores being higher after excluding Team 1. ​

​