This file documents the significant differences in success rates between PushGP (a1), GPT-4o Data-only prompting (a2), and GPT-4o Text-only prompting (a3) on each PSB2 question based on Pairwsie Chi-Squared Test. See Section 6.1 for the discussion and interpretation of these results. 1, 2, 3 in row and column headers refer to a1, a2, and a3. Cells give the p value of differences in success between the treatment in the column header versus the treatment in the row header.

$Basement

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 1 | - |
| 3 | 1 | - |

P value adjustment method: holm

$`Bouncing Balls`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm

$Bowling

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 2.2e-07 | - |
| 3 | 2.2e-06 | 0.62 |

P value adjustment method: holm

$`Camel Case`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm

$`Coin Sums`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 0.00063 | - |
| 3 | 0.00063 | - |

P value adjustment method: holm

$`Cut Vector`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm

$`Dice Game`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm

$`Find Pair`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 1 | - |
| 3 | 1 | - |

P value adjustment method: holm

$`Fizz Buzz`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 2.8e-11 | - |
| 3 | 2.4e-16 | 0.012 |

P value adjustment method: holm

$`Fuel Cost`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 8e-14 | - |
| 3 | 8e-14 | - |

P value adjustment method: holm

$GCD

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | <2e-16 | - |
| 3 | <2e-16 | - |

P value adjustment method: holm

$`Indices of Substring`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm

$Leaders

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm

$Luhn

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm

$Mastermind

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 1 | - |
| 3 | - | 1 |

P value adjustment method: holm

$`Middle Character`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 1.6e-09 | - |
| 3 | 3.4e-14 | < 2e-16 |

P value adjustment method: holm

$`Paired Digits`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 1 | - |
| 3 | 3.9e-09 | 1.1e-08 |

P value adjustment method: holm

$`Shopping List`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm

$`Snow Day`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm

$`Solve Boolean`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 0.0038 | - |
| 3 | 8e-05 | 0.2282 |

P value adjustment method: holm

$`Spin Words`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm

$`Square Digits`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | <2e-16 | - |
| 3 | <2e-16 | 0.0019 |

P value adjustment method: holm

$`Substitution Cipher`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 0.00018 | - |
| 3 | 0.00018 | - |

P value adjustment method: holm

$Twitter

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | 0.023 | - |
| 3 | 0.023 | - |

P value adjustment method: holm

$`Vector Distance`

Pairwise comparisons using Pairwise comparison of proportions

data: c(a1, a2, a3) out of c(100, 100, 100)

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| 2 | - | - |
| 3 | - | - |

P value adjustment method: holm