

p-value of Dunn test for β (order 1)

| | GPT-3.5 | GPT-4 | Human | LaMA-13B | LaMA-30B | LaMA-65B | LLaMA-7B |
|----------|---------|-------|-------|----------|----------|----------|----------|
| GPT-3.5 | 1.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| GPT-4 | 0.20 | 1.00 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 |
| Human | 0.00 | 0.29 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| LaMA-13B | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.49 | 1.00 |
| LaMA-30B | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 0.41 |
| LaMA-65B | 0.00 | 0.00 | 0.00 | 0.49 | 1.00 | 1.00 | 0.00 |
| LLaMA-7B | 0.00 | 0.00 | 0.00 | 1.00 | 0.41 | 0.00 | 1.00 |
| | PT-3.5 | GPT-4 | human | A-13B | A-30B | A-65B | MA-7B |