

# Rouge Score - Post-hoc Nemenyi Test p-values

**Claude 3.7 Sonnet** 1.0000 0.0000 0.0227 0.9751 0.9996

**Llama 3.1 8B** 0.0000 1.0000 0.0703 0.0000 0.0000

**Llama 3.3 70B** 0.0227 0.0703 1.0000 0.1152 0.0120

**Ministral 8B** 0.9751 0.0000 0.1152 1.0000 0.9307

**Qwen 2.5 72B** 0.9996 0.0000 0.0120 0.9307 1.0000

**Claude 3.7 Sonnet**  
**Llama 3.1 8B**  
**Llama 3.3 70B**  
**Ministral 8B**  
**Qwen 2.5 72B**

