Average Diversity (%) - Average Across All Models (Poem)

Statistical test results:
Best VS: VS-Standard
vs Direct: *** (p=0.000)
vs CoT: *** (p=0.000)
vs Sequence: ns (p=0.217)
vs Multi-turn: ** (p=0.009)