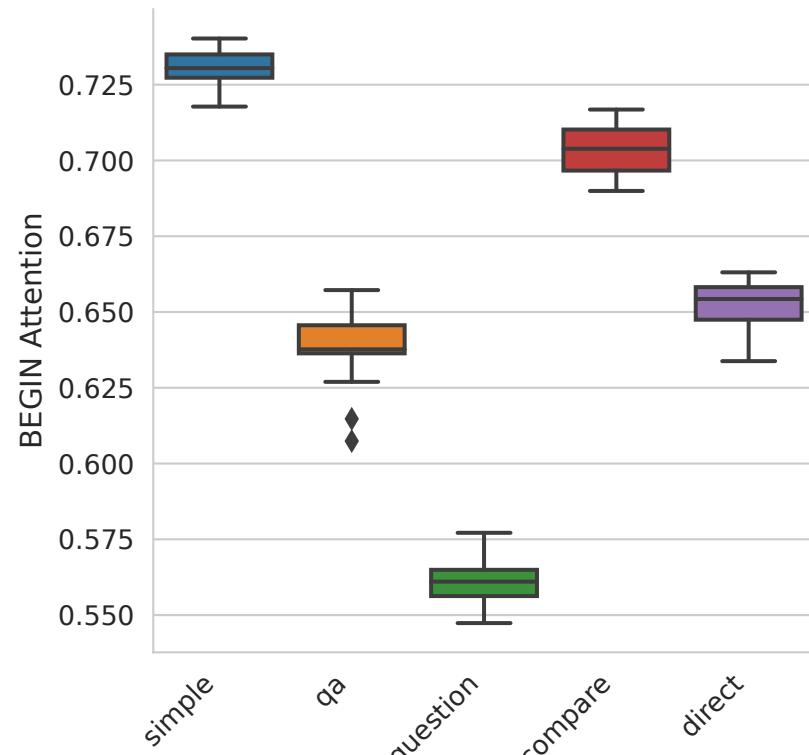
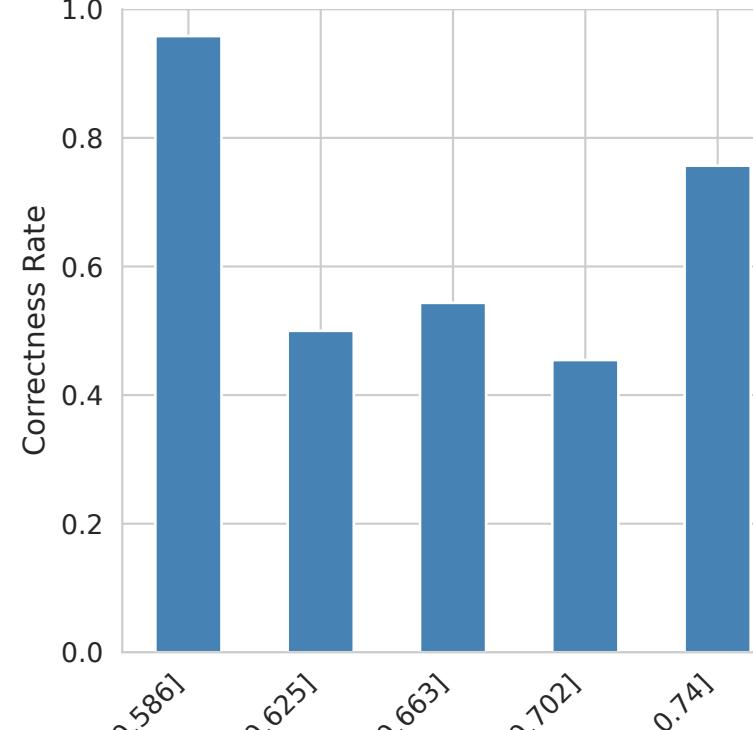


Attention Pattern Quantification: BEGIN Token Attention and Correctness

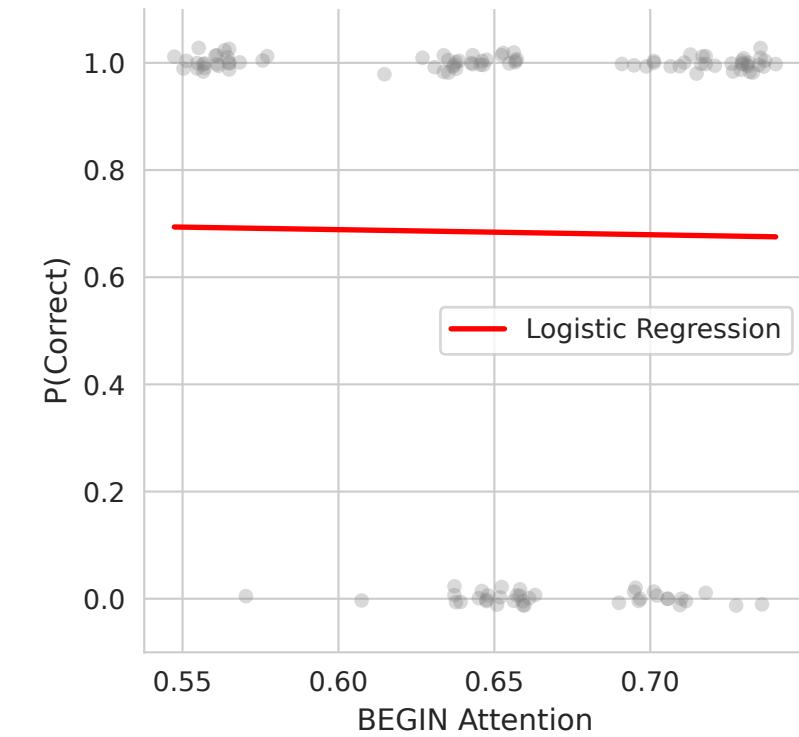
BEGIN Token Attention by Format



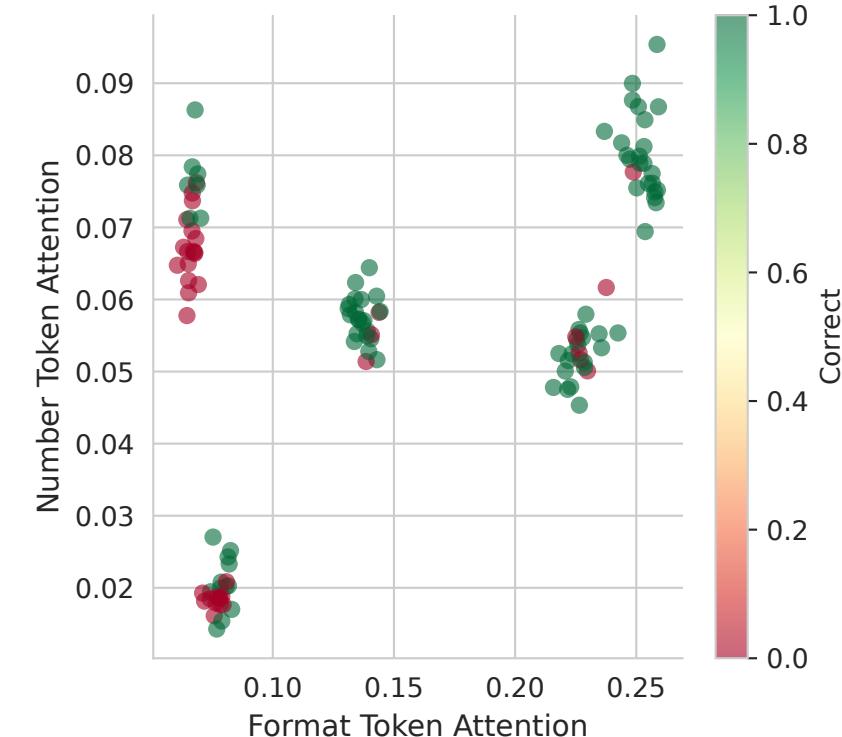
Correctness Rate by BEGIN Attention Level



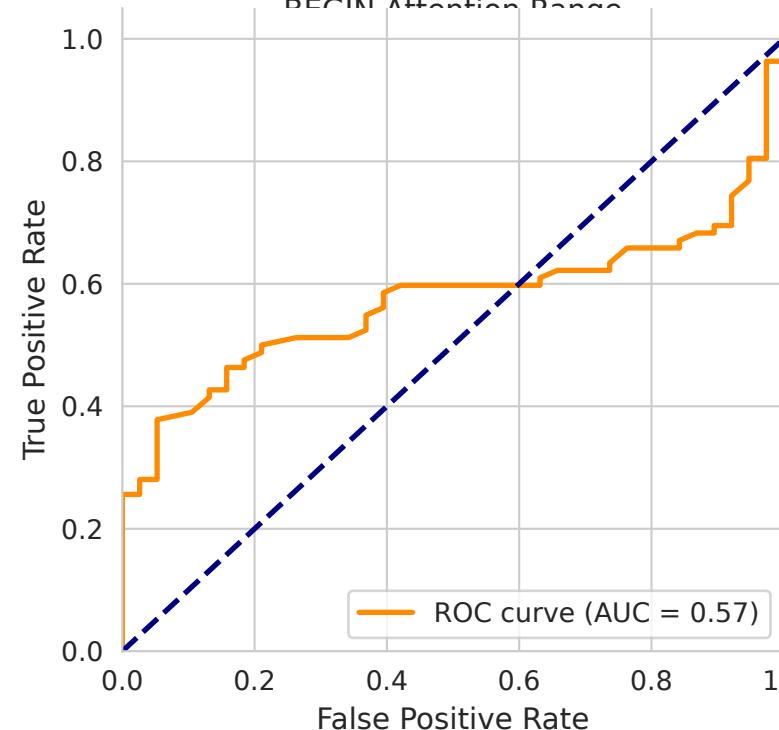
BEGIN Attention vs Correctness



Format vs Number Attention Trade-off



ROC Curve: BEGIN Attention Predicting Correctness



KEY FINDINGS:

- Correlation (BEGIN attention, correctness): $r = -0.146$ ($p < 0.1123$)
- Logistic Regression: Coefficient = -0.440 ROC AUC = 0.573
- Format Comparison: Simple format: 73.0% BEGIN Q&A format: 63.9% BEGIN Difference: 9.1%
- Interpretation: Higher BEGIN attention strongly predicts correct decimal comparison