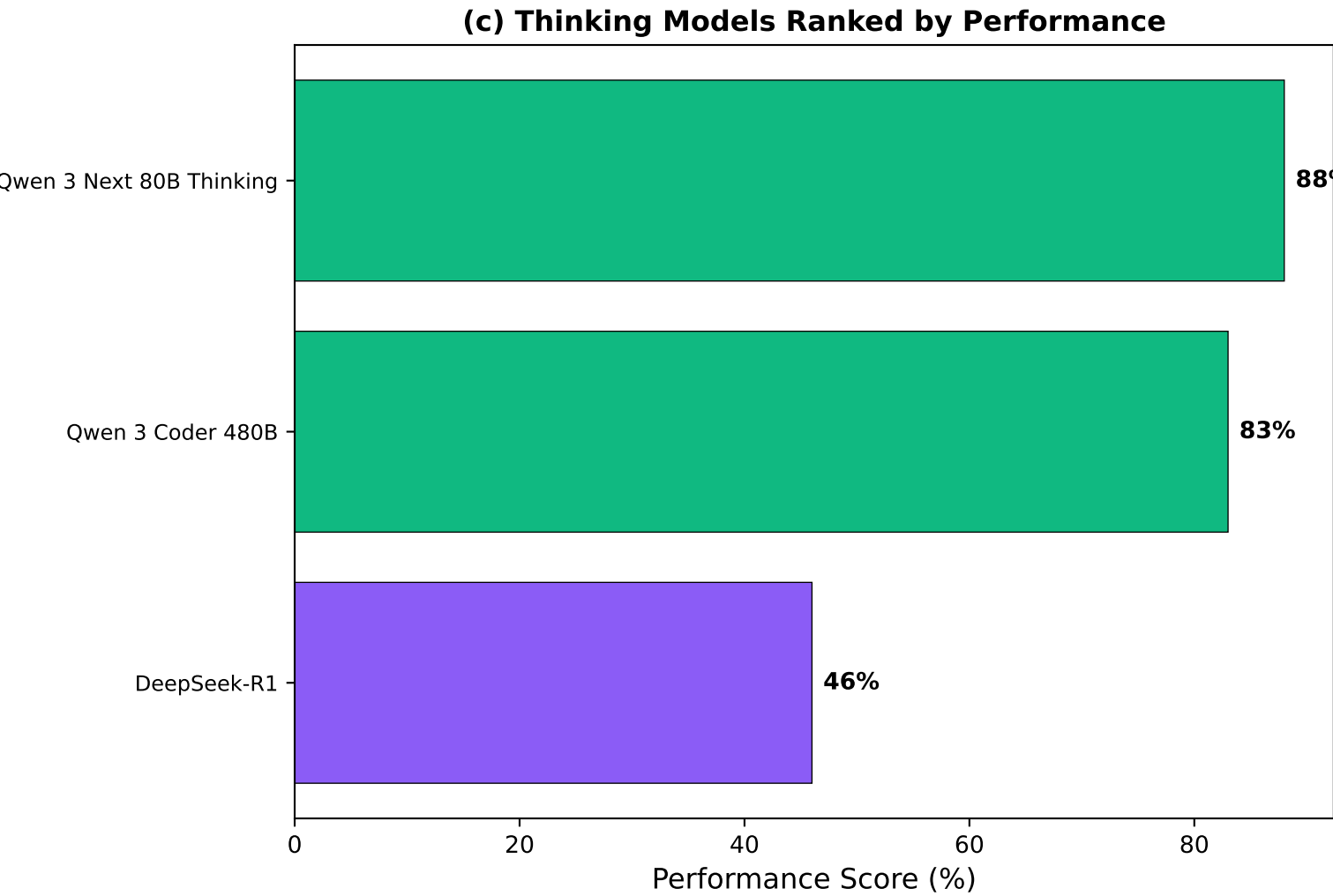
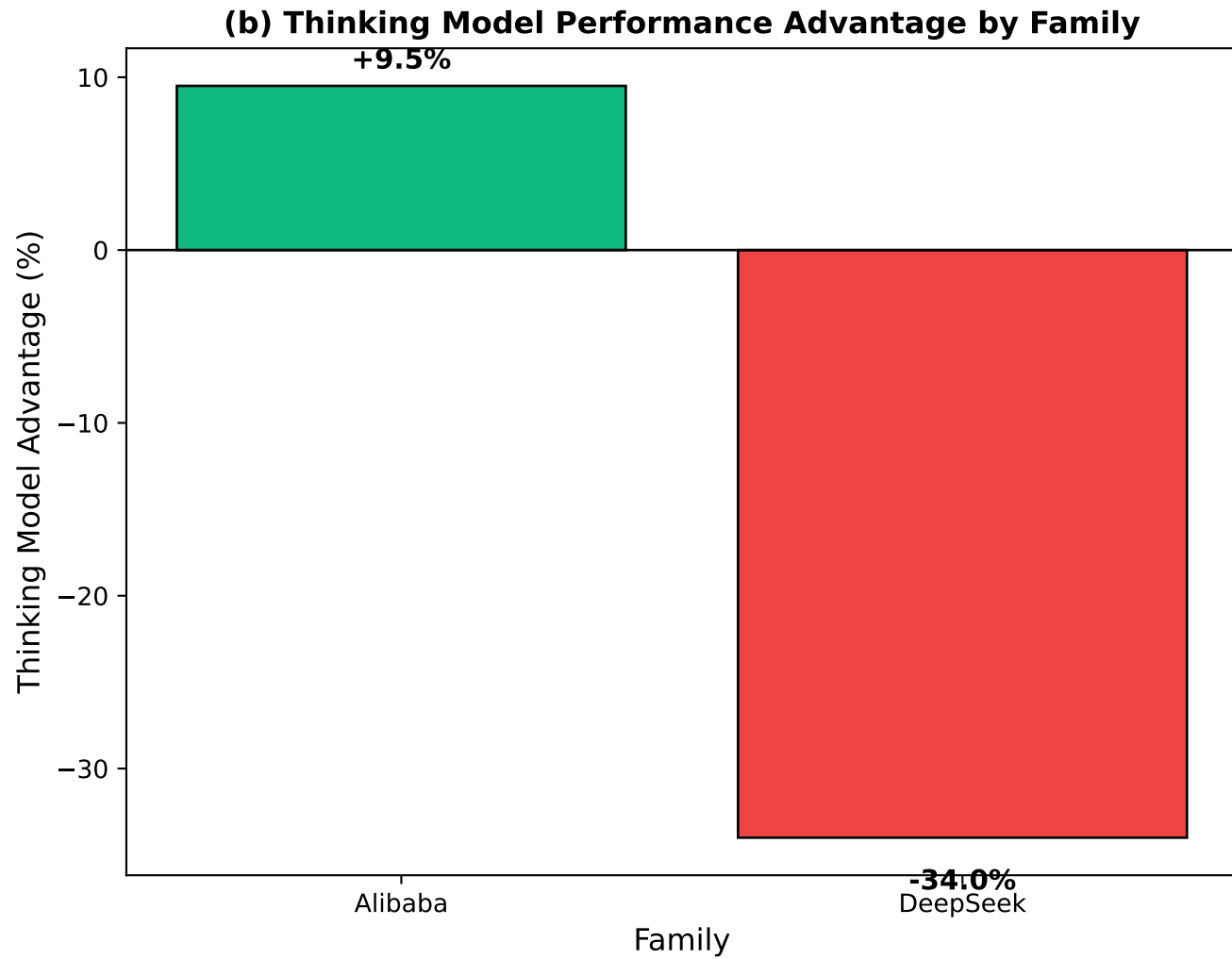
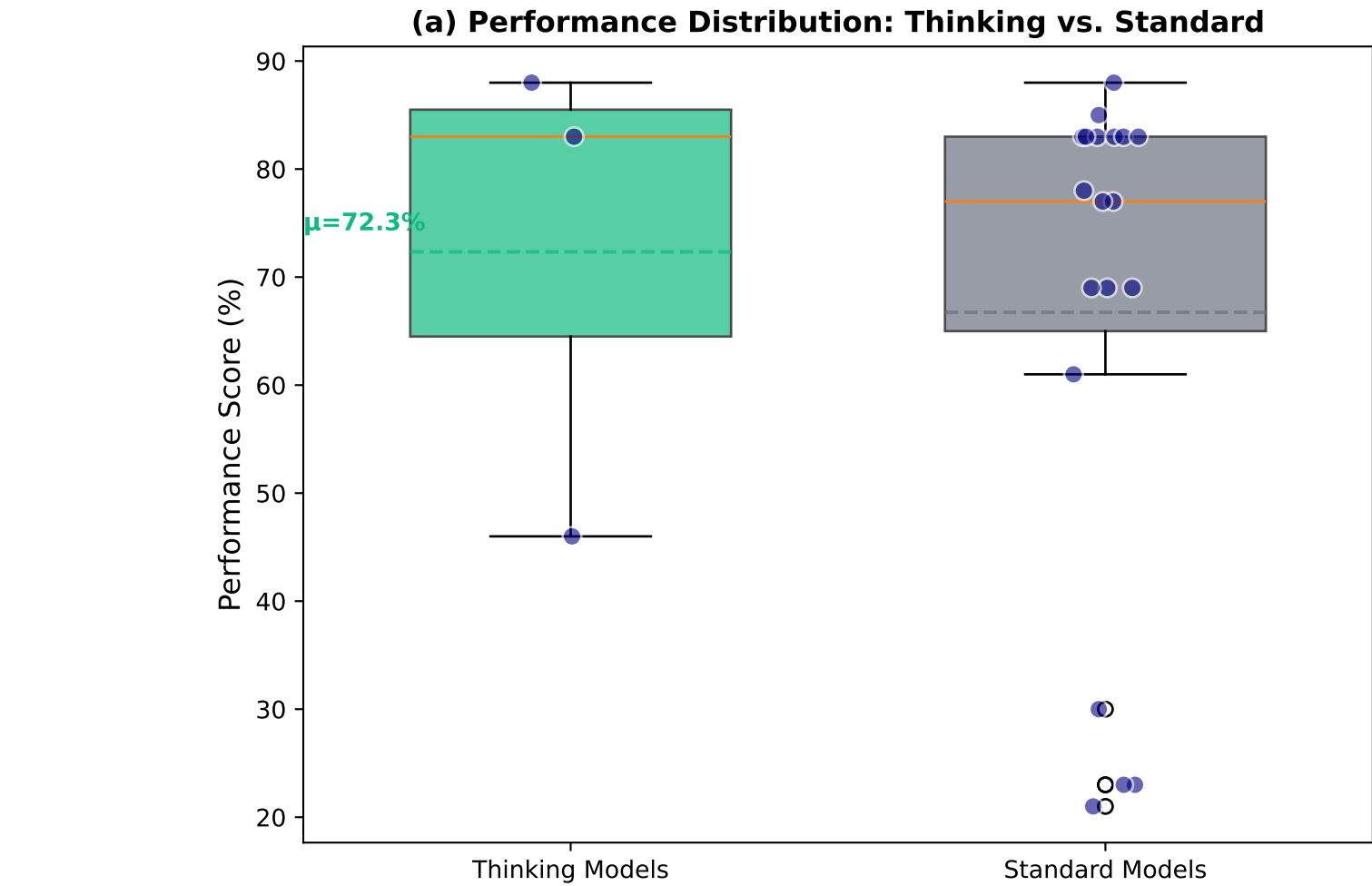


Figure 13: Thinking vs. Non-Thinking Model Comparative Analysis



(d) Statistical Significance Testing

Statistical Test	Result
Sample Size (Thinking)	3
Sample Size (Standard)	19
Mean Difference	+5.60%
t-statistic	0.383
t-test p-value	0.7059
Mann-Whitney U	36
M-W p-value	0.2325
Cohen's d	0.238

Interpretation: Not Significant difference ($p=0.706$), Small effect size ($d=0.24$)