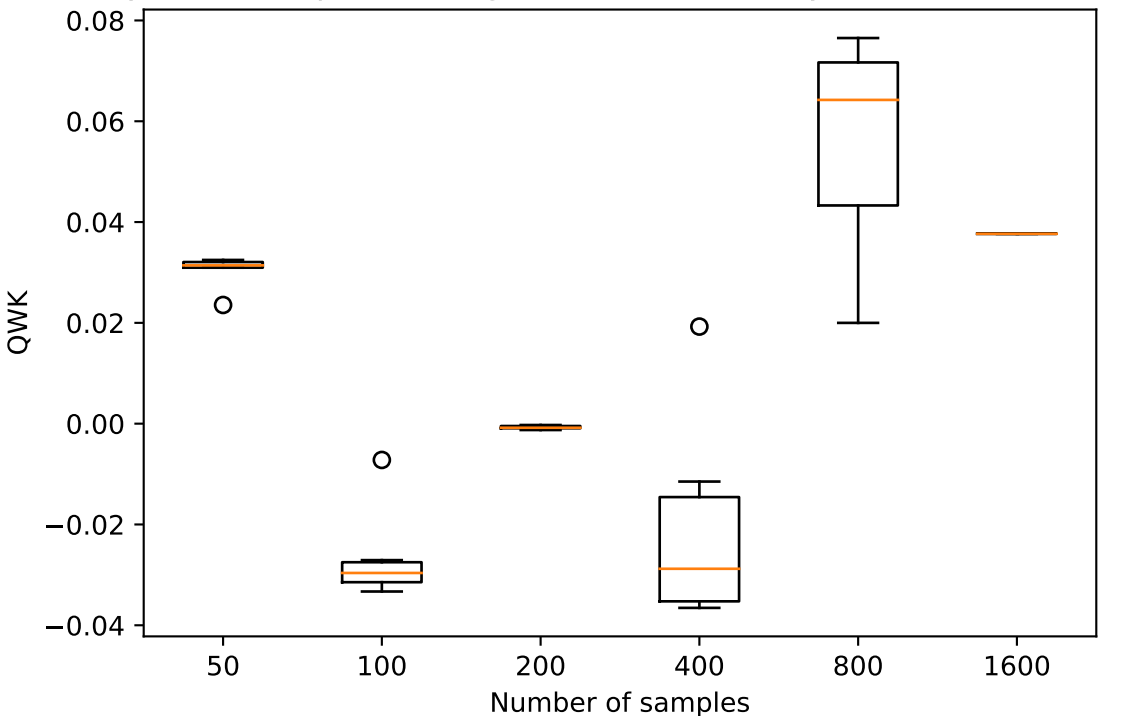


BERT Model trained with samples created by human experts rating answers created by text-davinci-003 annotated by gpt-3.5-turbo used for testing



Each box represents exactly 6 kappa values