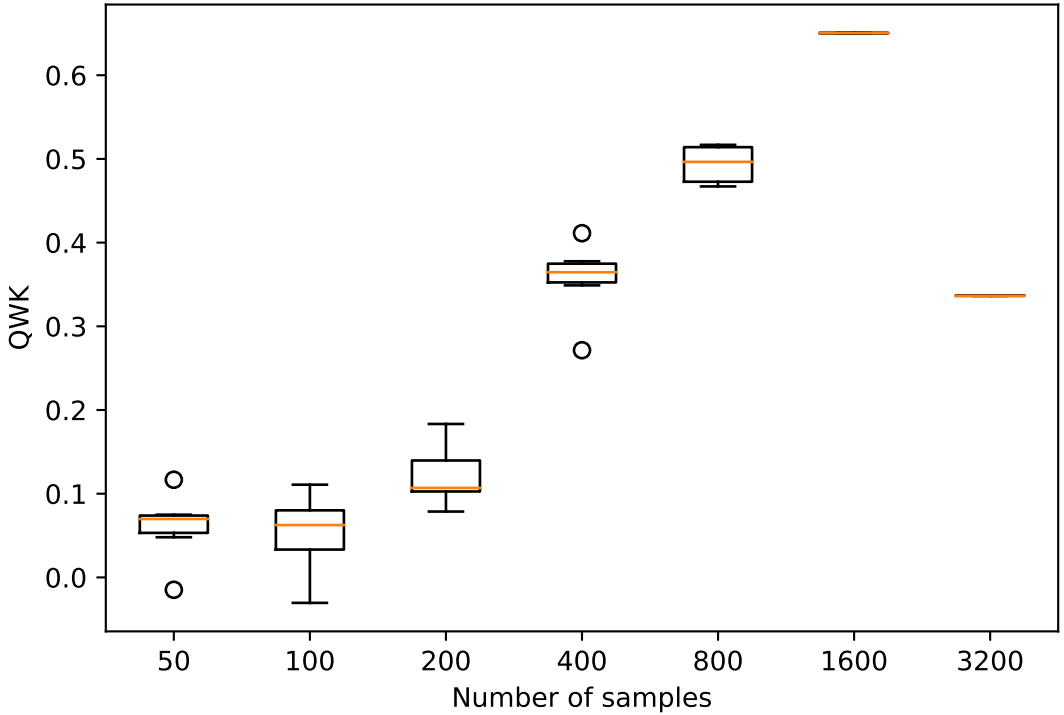
BERT Model trained with samples created by text-davinci-003 rating answers created by human experts used for testing



Each box represents exactly 6 kappa values