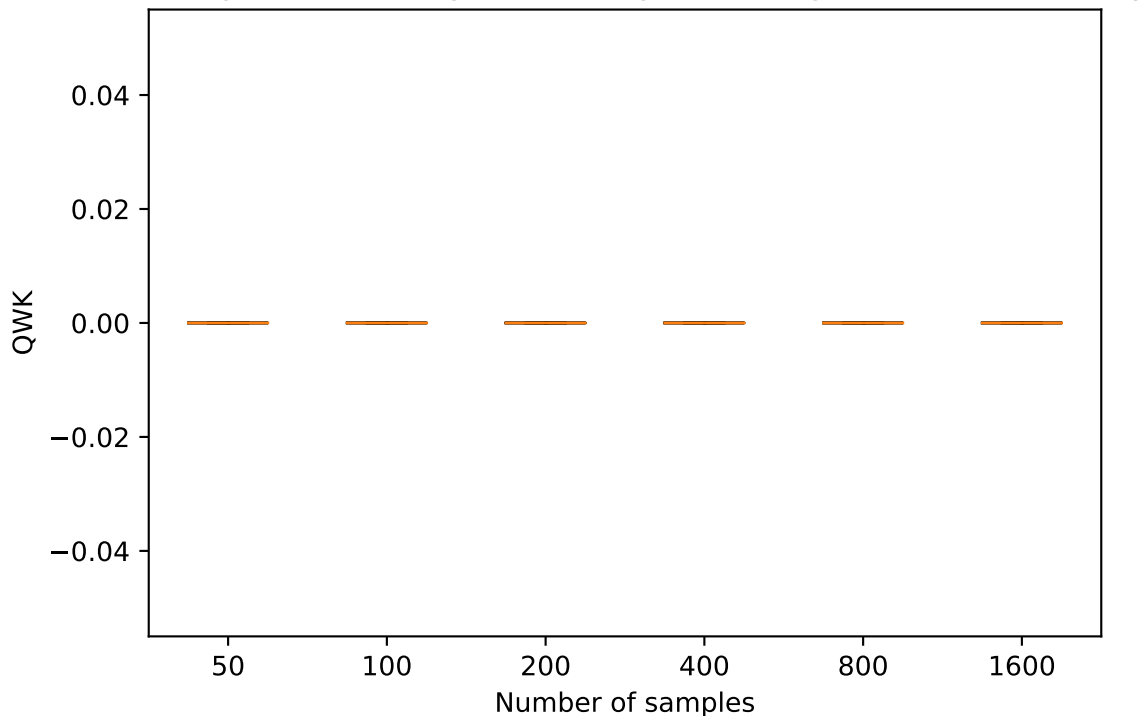


BERT Model trained with samples created by human experts rating answers created by gpt4 used for testing



Each box represents exactly 6 kappa values