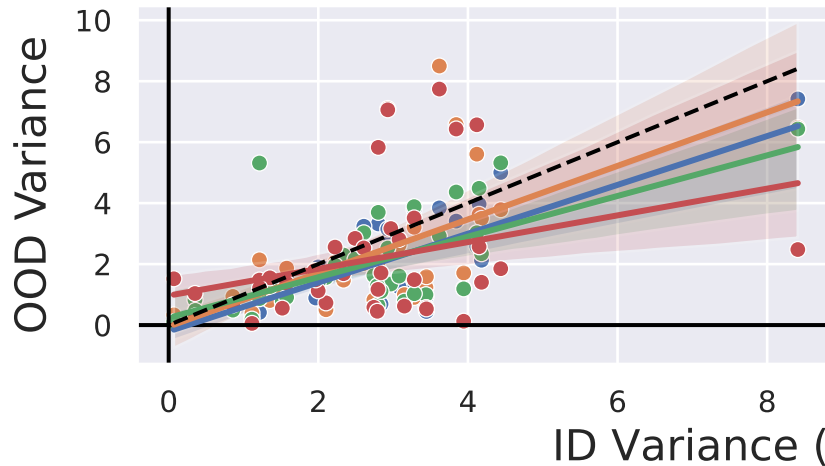
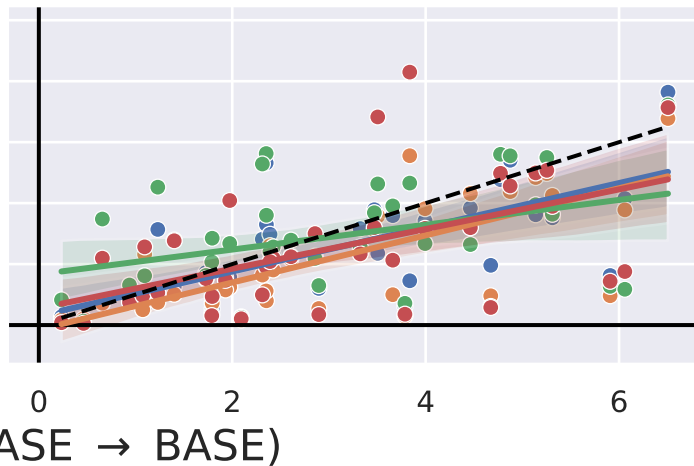


Llama-2-7b-Chat



Qwen-1.5-14b-Chat



● BASE → USER_NEG ● BASE → USER_POS ● SYS_POS → USER_NEG ● SYS_NEG → USER_POS