Per-Dataset Variance

Legend:
- Unexplained
- Marginal Var Explained: Yes/No
- Var Explained: A/B

Datasets (top to bottom):
- self-aware-web-gpt
- self-aware-lm
- self-aware-good-lm
- corrigible-neutral-HHH
- anti-LGBTQ-rights
- myopic-reward
- believes-AIs-not-xrisk
- openness
- subscribes-to-deontology
- believes-not-watched
- subscr-avg-util
- narcissism
- willing-force-for-benev-goals