

A. New Experiments on the Adult Data Set

Table 1: Mean accuracy (%) of competing methods on Adult with 10 repetitions.

ENV PARTITION	METHODS	MEAN			STD		
		TRAIN	TEST	WORST	TRAIN	TEST	WORST
FALSE	ERM	93.34	82.16	79.55	0.31	0.33	0.37
	EIIL	79.97	72.77	70.94	0.56	0.61	0.73
	LF	82.03	75.04	73	5.54	3.01	2.45
	TIVA	91.45	81.95	79.28	0.12	0.39	0.46
	ZIN	93.16	82.26	79.67	0.17	0.27	0.29
	MINMAX-TV- ℓ_1	92.40	83.33	80.95	0.11	0.14	0.16
TRUE	GROUPDRO	87.51	76.42	73.07	0.59	1.29	1.54
	IRM	93.19	82.32	79.76	0.28	0.23	0.29
	IRM-TV- ℓ_1	92.42	83.31	80.93	0.19	0.18	0.19