

Applied Robust Statistics

David J. Olive

Southern Illinois University
Department of Mathematics
Mailcode 4408
Carbondale, IL 62901-4408
dolive@math.siu.edu

January 16, 2005

Contents

Preface	v
1 Introduction	1
1.1 Outlier....s	4
1.2 Applications	5
1.3 Complements	18
1.4 Problems	19
2 The Location Model	25
2.1 Four Essential Location Estimators	25
2.2 A Note on Notation	29
2.3 The Population Median and MAD	30
2.4 Robust Confidence Intervals	37
2.5 Large Sample CIs and Tests	40
2.6 Some Two Stage Trimmed Means	43
2.7 Asymptotics for Two Stage Trimmed Means	48
2.8 L, R, and M Estimators	53
2.9 Asymptotic Theory for the MAD	55
2.10 Summary	59
2.11 Complements	62
2.12 Problems	64
3 Some Useful Distributions	72
3.1 The Binomial Distribution	73
3.2 The Burr Distribution	74
3.3 The Cauchy Distribution	74
3.4 The Chi Distribution	75
3.5 The Chi-square Distribution	75

3.6	The Double Exponential Distribution	77
3.7	The Exponential Distribution	78
3.8	The Two Parameter Exponential Distribution	79
3.9	The Extreme Value Distribution	80
3.10	The Gamma Distribution	81
3.11	The Half Normal Distribution	83
3.12	The Logistic Distribution	84
3.13	The Lognormal Distribution	84
3.14	The Normal Distribution	85
3.15	The Pareto Distribution	87
3.16	The Poisson Distribution	88
3.17	The Power Distribution	88
3.18	The Rayleigh Distribution	89
3.19	The Student's t Distribution	89
3.20	The Truncated Extreme Value Distribution	90
3.21	The Uniform Distribution	91
3.22	The Weibull Distribution	91
3.23	Complements	92
3.24	Problems	93
4	Truncated Distributions	100
4.1	The Truncated Exponential Distribution	103
4.2	The Truncated Double Exponential Distribution	105
4.3	The Truncated Normal Distribution	105
4.4	The Truncated Cauchy Distribution	108
4.5	Asymptotic Variances for Trimmed Means	109
4.6	Simulation	113
4.7	Complements	121
4.8	Problems	123
5	Multiple Linear Regression	126
5.1	A Graphical Method for Response Transformations	128
5.2	Assessing Variable Selection	138
5.3	A Review of MLR	153
5.4	Complements	166
5.5	Problems	167

6	Regression Diagnostics	189
6.1	Numerical Diagnostics	189
6.2	Graphical Diagnostics	192
6.3	Outlier Detection	196
6.4	A Simple Plot for Model Assessment	199
6.5	Complements	206
6.6	Problems	207
7	Robust and Resistant Regression	215
7.1	High Breakdown Estimators	215
7.2	Two Stage Estimators	217
7.3	Estimators with Adaptive Coverage	219
7.4	Theoretical Properties	220
7.5	Computation and Simulations	227
7.6	Resistant Estimators	232
7.7	Complements	235
7.8	Problems	237
8	Robust Regression Algorithms	239
8.1	Inconsistency of Resampling Algorithms	242
8.2	Theory for Concentration Algorithms	247
8.3	Elemental Sets Fit All Planes	256
8.4	Complements	266
8.5	Problems	268
9	Resistance and Equivariance	270
9.1	Resistance of Algorithm Estimators	270
9.2	Advice for the Practitioner	274
9.3	Desirable Properties of a Regression Estimator	275
9.4	The Breakdown of Breakdown	278
9.5	Complements	286
9.6	Problems	286
10	Multivariate Models	287
10.1	The Multivariate Normal Distribution	288
10.2	Elliptically Contoured Distributions	292
10.3	Sample Mahalanobis Distances	295
10.4	Affine Equivariance	297

10.5	Breakdown	298
10.6	Algorithms for the MCD Estimator	299
10.7	Theory for CMCD Estimators	301
10.8	Complements	313
10.9	Problems	315
11	CMCD Applications	320
11.1	DD Plots	320
11.2	Robust Prediction Regions	328
11.3	Resistant Regression	331
11.4	Robustifying Robust Estimators	335
11.5	Complements	336
11.6	Problems.	337
12	1D Regression	340
12.1	Estimating the Sufficient Predictor	343
12.2	Visualizing 1D Regression	349
12.3	Predictor Transformations	361
12.4	Variable Selection	362
12.5	Complements.	372
12.6	Problems	375
13	Generalized Linear Models	382
13.1	Introduction	382
13.2	Multiple Linear Regression	384
13.3	Logistic Regression	388
13.4	Loglinear Regression	397
13.5	Inference	404
13.6	Variable Selection	411
13.7	Complements	416
13.8	Problems	416
14	Stuff for Students	435
14.1	Tips for Doing Research	435
14.2	R/Splus and Arc	438
14.3	Projects	446
14.4	Hints for Selected Problems	455