Robust Statistics : Statistics 4360/5360

Instructor: Chris Field, Chase 108, 494-3339, field@mathstat.dal.ca
Time/location: MWF 9:30-10:30, LSC 212
Office Hours: Monday 2:30-3:30, Wednesday 10:30-11:30, Thursday 11:30-12:30
Marking scheme:
30% based on Assignments
20% based on Article review/Project
20% based on midterm
30% based on Final Exam

Outline

1. Overview of robustness
   Chapter 1, Hampel et al
   Concepts of Robustness article
   Ozone depletion example

2. Basic tools
   Scale and location/scale
   Welsh 5.1-5.4
   Influence function article, JASA(1974) V69, p383-393

3. Linear models : M and GM-estimates
   Welsh 5.5-5.6

4. Least median of squares, regression depth
   Rousseeuw and Leroy Chapter 3
   Regression depth article, JASA(1999) V94, p388-433

5. Covariance estimation, multivariate methods
   Outlier detection
   Rousseeuw and Leroy Chapter 7

6. Robust Testing
   Handbook V15 Chapters 3 and 4

7. Generalized linear models
   Cantoni and Ronchetti article, JASA (2001) V96, p1022-1030

8. Mixed Models
   Handbook V15 Chapter 13

9. Model Selection
   Model selection paper

10. Robustness against Dependence
    Dependence in regression paper

Final Exam Statistics 5360, April 12, 2004

Time allowed : 10:00am to 10:00pm

1. The data set urine (description attached) gives physical characteristics of urines (6 predictors) with and without crystals. Carry out a classical and robust analysis to determine the best model for predicting presence/absence of crystals. Comment on any differences in the results.

2. The following data gives the number of lettuce plants which grow in a factorial experiment with three levels of nitrogen and three levels of phosphate. The data is given as follows:

	n_0	n_1	n_2
p_0	449	413	326
p_1	409	358	291
p_2	341	278	312

   (a) Carry out a least squares analysis, an analysis with lmRob and an analysis using least median of squares to determine the effects of the nitrogen and the phosphate. Compare the results and determine what causes the differences among the analyses. Note that you need to create indicator variables to do this as a regression.

   (b) For both the least squares and the lmRob, construct confidence intervals for the differences $p_2 - p_0$ and $p_2 - p_1$.

   (c) Carry out a robust Wald test for the phosphate effects. If we are testing that the last q βs are 0 ($H_0 : \beta_{(2)} = 0$), then the test statistic has the form $\hat{\beta}_{(2)}^T \hat{V}_{(2,2)}^{-1} \hat{\beta}_{(2)}$ where $\hat{\beta}_{(2)}$ is the estimate of $\beta_{(2)}$ under the full model and $\hat{V}_{(2,2)}$ is the q by q submatrix of the covariance matrix of the estimates under the full model. Note that the covariance matrix is the object cov returned by lmRob.

3. The attached data gives body weight and brain weight for a selection of mammals. For this question use the 11 observations from observation 6 to observation 16.

   (a) Find the regression depth of the least squares fit obtained by regressing log(brain wt) on log(body wt).

(b) Find the regression line with maximum depth.

4. Assume we are interested in estimating the parameters α and β from a Gamma distribution but are concerned that the model may not be exactly true.

   (a) Determine whether the maximum likelihood estimates for the parameters are robust.

   (b) If they are not, propose a robust procedure outlining the steps of an algorithm you would need to compute it.

   (c) Indicate how you would assess the performance of the mle relative to the robust estimate at the model Gamma distribution and a contaminated Gamma, $(1-\epsilon)\Gamma(\alpha, \beta) + \epsilon\Gamma(\alpha, 10\beta)$. Write out any integrals that you would need to evaluate.