Exercises 5 solutions

1. The L_1 estimator has

   (a) objective function $\rho_{L_1}(u) = |u|$, where u stands for the standardised residuals *or the unstandardised ones* as the definition of the L_1 estimator in the lecture notes is based on the unstandardised residuals. See (c) for why it doesn't matter.

   (b) $\psi_{L_1}(u) = -1$ if $u < 0$, $\psi_{L_1}(u) = 1$ if $u > 0$, **and** $\psi_{L_1}(0)$ not defined.

   (c) If e has a double exponential distribution with mean 0 then $Y = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + e$ has a double exponential distribution with mean $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$. Hence, the likelihood for an independent sample $Y_1, \ldots, Y_N$ from Double Exp with mean $\mu_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$ is given by

   $$L(\boldsymbol{\beta}; y_1, \ldots, y_N) = \prod_{i=1}^{N} \frac{\lambda}{2} \exp\{-\lambda|y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}|\} = \frac{\lambda^N}{2^N} \exp\left\{-\lambda \sum_{i=1}^{N} |y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}|\right\}$$

   which is maximal when $\sum_{i=1}^{N} |y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}|$ is minimal (for any given λ).

2. Weight functions:

   (a) the L_1 weights are

   $$w_{L_1}(u) = \begin{cases} -\frac{1}{u}, & u < 0, \\ \frac{1}{u}, & u > 0, \end{cases}$$

   $w_{L_1}(0)$ not defined (it becomes clear why the L_1 estimator is numerically difficult).

   (b) the Bisquare weights are

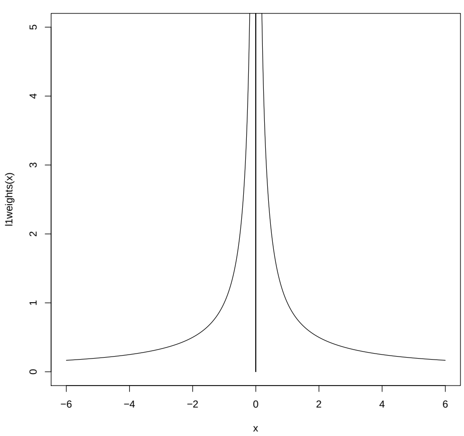
   $$w_B(u) = \begin{cases} \frac{1}{c^2}\left\{1 - \left(\frac{u}{c}\right)^2\right\}^2, & |u| < c \\ 0, & \text{otherwise,} \end{cases}$$

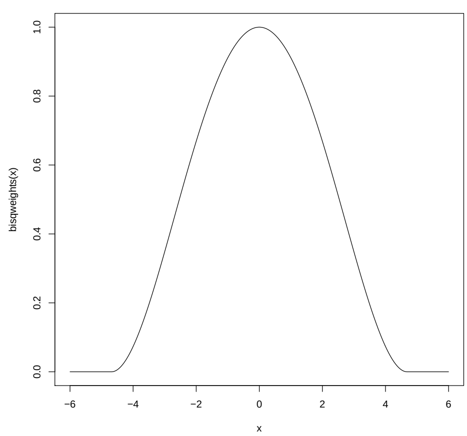
   (hence the name **Bi**square).

   (c) (for your information) the least squares weights are

   $$w_{LS}(u) = 1.$$

Implications: Least Squares treats all observations the same (with same weight); Bisquare gives full weight only to zero residuals, decreasingly less the further away they are and zero weight if they are outside $[-c, c]$ — of the estimators considered here it is the only one that can give zero weight to extreme observations. The L_1 estimator gives inverse proportional weight to residuals the further away from zero they are.





3. Robust regression analyses of the Turnip with Model 2 (including an x_2^2–term).

   (a) The estimators for the regression parameters are more or less similar, with the same signs. However, the absolute values are larger for the robust estimators (LMS even larger as MM), and the p-values for the (robustified MM) t-tests are smaller for x1 and x3. Weighting down outliers, the robust estimators estimate the main bulk of the data more precisely, so that the estimated residual standard errors are much smaller. Obviously, the influence of variables x1 and x3 is clearer weighting down the outliers.
   These results together mean that there are no bad leverage effects (cf. leverage plot for LS Model 2 – none stick out too much)[1].

   (b) Outliers in least squares (model 2): no.s 10, 15, 19 (residual plot and Cook's distance).
   The MM-estimator gives only observation 19 zero weight (one could interpret this by saying that only this observation really looks clearly incompatible with the normal model assumption), but 8, 10, 15 and 20 are weighted down considerably as well.
   So there is not too much discrepancy, but the MM detects more critical observations than least squares (cf. "masking effect").

   (c) The residual plots are not much different but it is noticable that for the robust estimators, especially LMS, there are 'core' observations that have residuals very close to zero and others that are allowed to be much further away — if one takes into account how these estimators work the reason is pretty clear. The MM-estimator is a compromise between LMS and LS in the sense that it has fewer very small residuals than the LMS, but the larger ones (not only those of the suspected outliers) are generally better.
   Also it is noticable that there is a small range of x1-values (roughly between 1.8 and 2) where there is much more variability and many more observations — it might be of interest to further investgate this (e.g. ask subject matter experts for possible explanations).

[1] The MM-estimator did not produce a plot of robust Mahalanobis distances for these data because the robust covariance matrix estimator was singular; due to the structure of x2, more than half of the points lie on a lower dimensional hyperplane of x-space.