**Introduction to robust statistics**

Robust, paramteric and non-parametric.

Regression assumptions

1. Independent
2. Linear
3. Constant variance
4. X is normally distributed
5. Normal errors

Library(lattice)

xyplot(Y~X1|X2, data = data)#for categorical plotting

After fitting a model look at the residual plots to see if they look normal.

QQ-plot: normal distribution is the line. –Assumption 5.

Cook’s Distance: How influential a point is. See if all the points are within the red lines.

library(robustbase)

lmrob(Y~X)#for a robust linear regression.

abline(lmrob(Y~X))#to see the fitted robust line

Treat robust models as a different set of models in comparisons.

library(MNM)#for covariance matrix plotting

Parametric model: Finite space. Suppose we want to estimate 2 parameters N(u,sigma^2). U is in +- inf and sigma is in (0,+inf).

Non-parametric: Not necessarily finite. The distribution is not symmetric. More general.

Semi-parametric: Between parametric and non-parametric

Robust statistics are in parametric.

Testing different estimators: Unbiasedness = minimum variance

Luento 2

Run a classical method and then a robust model to see if they differ. If they don’t differ then use the classical and if they do then use robust.

Statistic to see if distribution has fat tails: kurtosis. A normal distribution has a kurtosis excess of 0 but a kurtosis of 3.

Median absolute deviance in R in mad(). Measure of spread that should be reported with median.

Lecture 3

MSE is variance + squared bias.

Lecture notes for L8

Location

m(x+c) = myy(x)+c

m(cx) = cmyy(x)

Scale

sigma(x+c)=sigma(x)

sigma(cx)=abs(c)\*sigma(x)

sigma is scale

M estimates are starting points and we build on from there.