DEPARTMENT OF STATISTICS
2011/12 - Term 1

# STAT 547N (ROBUSTNESS)

**Description**: Statistician often handle data of uneven quality, that is, data contaminated by a fraction outliers and other types of "bias generating" contamination. Classical MLE and least squares procedures commonly used to process these data are very sensitive to contamination in the data. The classical approach for dealing with contaminated data is "to clean the data first and then apply classical statistical procedures". This approach may not be feasible/practical when dealing with a large number of variables and in occasions it may not work at all (e.g. some outliers may not be detected because of masking and swamping).

Modern robust procedures address the question of data quality. When it is found that the model cannot fit all the data well, the robust procedure will automatically search for the largest fraction of data that can be well fit by the model. Robustness theory is also concerned with the issue of measuring the sensitivity of statistical procedures and the development tools to assess "sensitivity". An advantage of robust methods is that they produce reasonable fits in the presence of outliers and other departures from model's assumptions. On the other hand, to apply robust methods we need fast and reliable algorithms. The development of such algorithms is then an area of current research interest.

In this course, the students will learn about the main robust procedures recently implement in available R libraries. They will also learn some basic theoretical concepts of the modern robustness theory. Students will analyze real data sets using robust and non-robust methods and compare their findings.

**Instructor**: Ruben Zamar

**Prerequisite**: Stat grad student. Other students, please consult with the instructor.

**Evaluation**: will be based on 3 Assignments and a final project.

**Textbook**: There is no textbook for this course. A set of photocopied material from course notes and several papers will be made available to the students throughout the course.

**References**:

Maronna, Martin and Yohai (2006). Robust Statistics: Theory and Methods. Wiley Series in Probability and Statistics.

Rousseeuw and Leroy (1987). Robust regression and outlier detection. Wiley.
Hampel, Ronchetti, Rousseeuw and Stahel (1986). Robust statistics: The approach based on influence functions. Wiley
Huber (1981). Robust statistics. Wiley

## Some Topics Covered in This Course

- **Location and dispersion models**

- S– and M–estimates.
- Influence function and contamination sensitivity.
- Breakdown point.
- Maxbias function.
- Huber's and Hampel's optimality problems: Optimally robust estimates.

- **Regression Model**

- S-estimates.
- MM–estimates.
- Tau–estimates

- **Multivariate Location and Scatter**

- M-Estimates

- S-Estimates

- Coordinatewise and pairwise estimates

- **Local and global robustness**

- One–step GM–estimates.

- **Robust inference**

- Robust confidence intervals.
- Robust testing of hypothesis.

- **Diagnostic tools and outlier tests based on robust procedures**