Robust regression

- Second-level with Robust regression
  - Robust regression techniques are a class of estimators that **are relatively insensitive to the presence of one or more outliers** in the data
    - If extreme values on predictors (high variance) can be very dangerous to estimate
  - Outliers in the data can create violations of normality and equality of variance assumptions, and they can have a disproportionately large impact on the statistical solution
  - Thus, **the robust regression** when the group effects are estimated is a good tool to overcome these issues

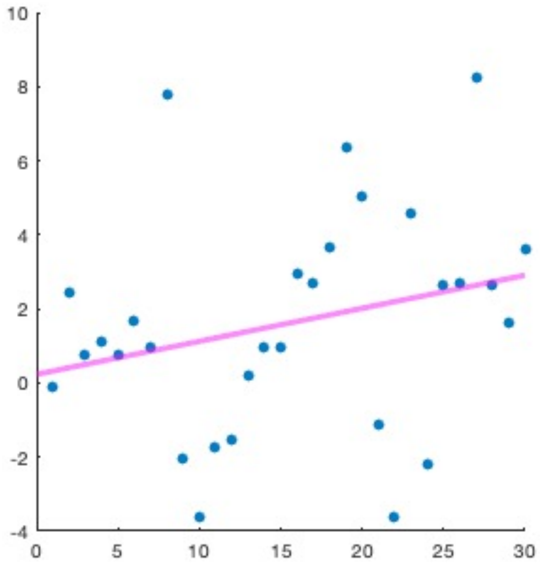
When outlying values are present in the data, violations of distributional assumptions can lead to reduced power and increased false positive rates.

https://www.youtube.com/watch?v=lFdQMW7mEGo

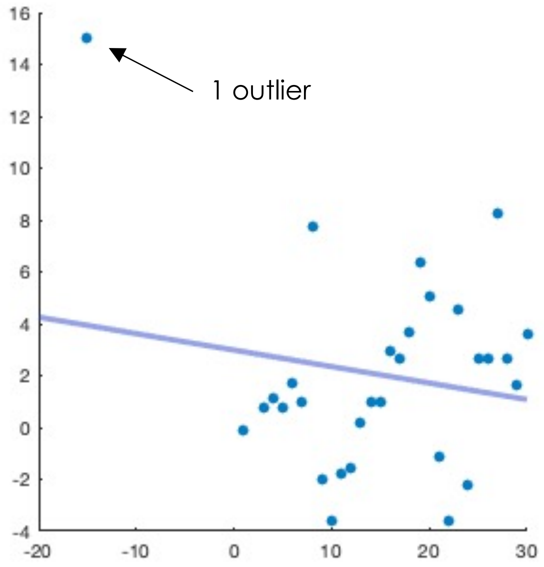
Wager, T. D., Keller, M. C., Lacey, S. C., & Jonides, J. (2005). Increased sensitivity in neuroimaging analyses using robust regression. Neuroimage, 26(1), 99-113.

Robust regression: examples

Data with linear relationship (n=30)



Same data, with one outlier




```
datax = [1:30];
datay = [0.1:0.1:3] + randn(1,30).*2.5;

scatter(datax,datay);
l1 = lsline;
l1.LineWidth =3;
%%
datax2 = datax;
datax2(2) = -15;
datay2 = datay;
datay2(2) = 15;
scatter(datax2,datay2);
l2 = lsline;
l2.LineWidth = 3;
%%
create_figure;
subplot(1,2,1);
scatter(datax,datay,'filled');
l1 = lsline;
l1.LineWidth =3;
l1.Color = [1 0 1 0.5];
subplot(1,2,2);
scatter(datax2,datay2,'filled');
l2 = lsline;
l2.LineWidth = 3;
l2.Color = [0.1 0.2 0.8 0.5];
```

MATLAB

https://www.youtube.com/watch?v=lFdQMW7mEGo

Robust regression and Neuroimaging

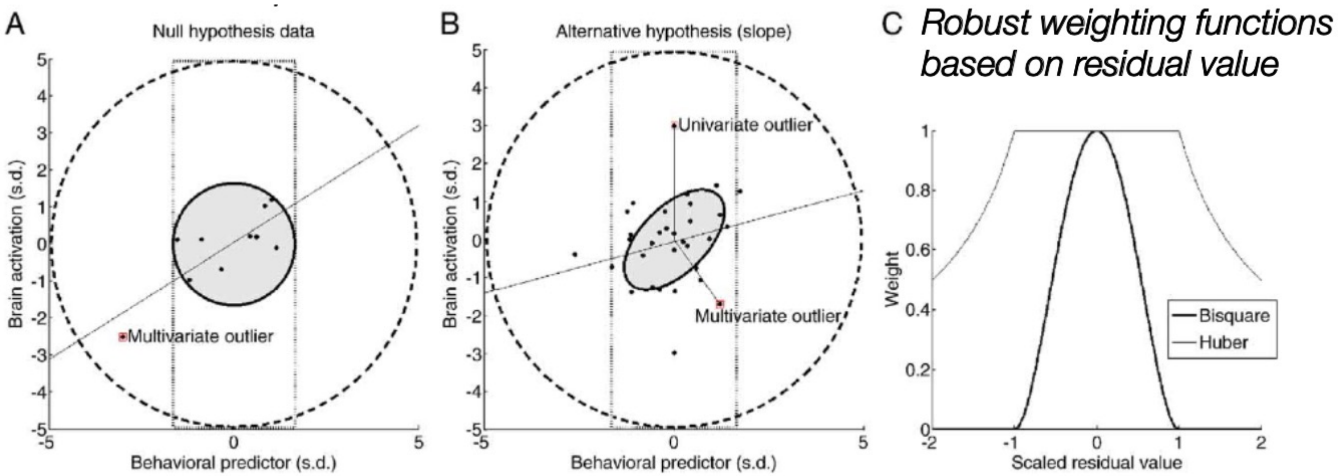
- Robust techniques are particularly useful when **a large number of regressions** are tested and assumptions cannot be evaluated for each individual regression, such as with neuro- imaging data.
- There are **three principal reasons** why robust regression techniques may be particularly important for analyzing neuroimaging data
  - First, there are good reasons to suspect that artifactual outliers are common in such data
  - Second, it is often unfeasible to check assumptions for each individual regression analysis due to the number of separate regression analyses performed (Luo and Nichols, 2003, provide a solution), and thus an efficient robust algorithm that dampens the effects of outliers would be advantageous
  - Finally, robust techniques may **increase statistical power** (decreasing the false negative rate) and may prevent false positives in the presence of outliers or skew in the data.

Wager, T. D., Keller, M. C., Lacey, S. C., & Jonides, J. (2005). Increased sensitivity in neuroimaging analyses using robust regression. Neuroimage, 26(1), 99-113.

Robust regression: Iterative generalized least squares

- Iterative algorithm (Iterative generalized least squares; IGLS):
  1) Weight based on inverse of leverages
  2) Fit weighted least squares model
  3) Scale and weight residuals
  4) Re-fit model
  5) Iterate steps 2-4 until convergence
  6) Adjust variances or degrees of freedom for P-values

Examples of data with outliers



inciples of fMRI (Lindquist and Wager)

Wager, T. D., Keller, M. C., Lacey, S. C., & Jonides, J. (2005). Increased sensitivity in neuroimaging analyses using robust regression. Neuroimage, 26(1), 99-113.

Robust regression: Iterative generalized least squares

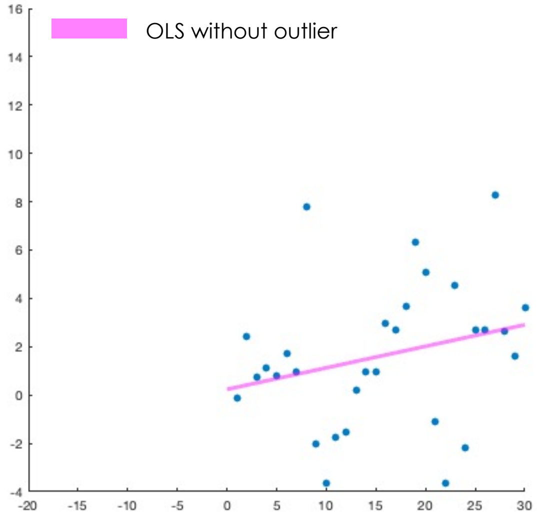
- Iterative algorithm (Iterative generalized least squares; IGLS):
  1) Weight based on inverse of leverages
  2) Fit weighted least squares model
  3) Scale and weight residuals
  4) Re-fit model
  5) Iterate steps 2-4 until convergence
  6) Adjust variances or degrees of freedom for P-values

intercept

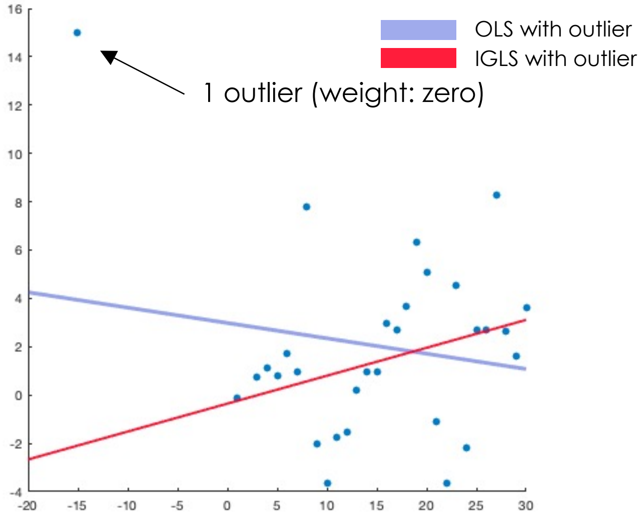
```
[beta, stats] = robustfit([ones(1,30); datax2]', datay2,'bisquare', [], 'off');
h3 = refline([beta(2) beta(1)]);
h3.Color = [1 0 0.1 0.9];
h3.LineWidth = 2;
```

MATLAB

Data with linear relationship (n=30)



Same data, with one outlier



Principles of fMRI (Lindquist and Wager)

Robust regression: Summary

- OLS technique are hugely influenced by outliers
- Robust regression is a good way to minimize **the influence of outliers** when assumptions and data cannot be checked for every test performed
  - In fMRI cases, separate regression models are typically fit for each of 30,000-200,000 voxels in the brain
  - It is especially substantial benefits in neuroimaging analysis (see Wager et al., 2005)

If you need to more understand the robust regression in fMRI, module 14 of the Principles of fMRI (part 2) will be great materials with recommend readings

Wager, T. D., Keller, M. C., Lacey, S. C., & Jonides, J. (2005). Increased sensitivity in neuroimaging analyses using robust regression. Neuroimage, 26(1), 99-113.