

- Residual analysis.
- Detect the violations of the pre-assumptions of the

Residual Plot: Residual vs. Fitted line.

LINER Principles.

- Equal variance assumption.
- Fun-out
- Fun-in
- Non-linear

- Linear
- Independent \rightarrow Good fit if it appears as "random pattern".
- Normal
- Equal variance.

Remedial Measures

- Fix the violations by applying chosen transformations.
- Either be applied to Y or X or both variables.

Skills in JMP

- Relationship between "Estimates", "Std Error", and "t Ratio".

$$\frac{\text{"Estimates"}}{\text{"Std Error"}} = \text{"t Ratio"}$$

Common transformations

Transformations	on dep. var.	on ind. var.
Square root	\sqrt{Y}	\sqrt{X}
Log	$\log(Y)$	$\log(X)$
Inverse	$1/Y$	$1/X$

- Residual Normal Q-Q plot.
- To see if all points are on the fitted line.

- Residual by Estimation Plot.
- Fun-out
- Not Equal Variance.
- Non-linear.

Box-Cox transformation

Some remarks

- Those are generally called "power transformations".

$$T(Y) = \begin{cases} \frac{Y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(Y) & \text{if } \lambda = 0 \end{cases}$$

Detection of outliers

- Choose a range of lambda.

- "Unusual observations"
- Outlying or extreme
- Not consistent to the rest of the part of the model.
- "But unusual"

- Influence.

- Leverage
- X range

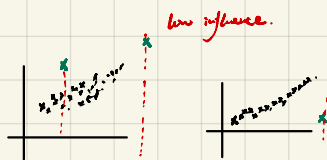
Cook's Distance

(in a least-

- A measure of influence of a data point. (spikes regression analysis).

- Observations with Cook's distance $D_i > 1$ are high influence.

- "Tolerance of the influence".



High leverage.
and High influence.

Remedial Measures (Continued)



Choose the variable with wider range.

Remedy Heteroscedastic Variance.

- Always give a transformation of the dependent variable Y .
- Frequently: \sqrt{Y} or $\log(Y)$
- Because non-constant variance is a property of Y .
- We want a transformation that improves the R-Squared value.

Example: Fit an experiment model:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$