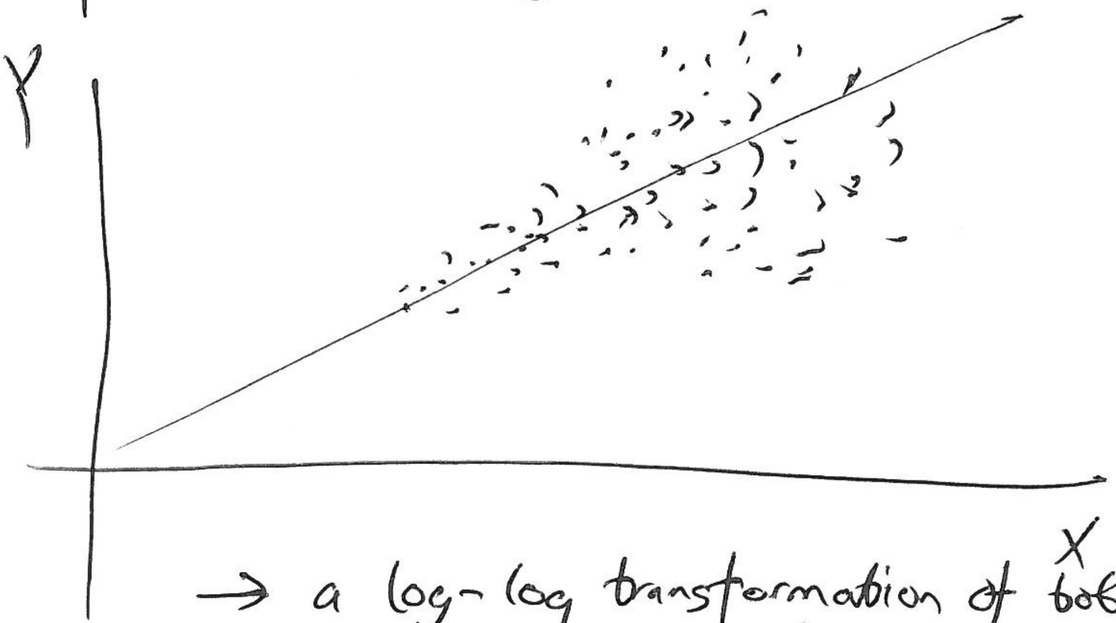
STAT2008/4038/6038 Regression Modelling 29/3/2017

Transformations

In fitting regression models, we can experiment with changes of scale (transformations) for the variables involved and then assume the regression model holds on the transformed, rather the untransformed scale.

The transformations we typically use are ones that are typically monotonic, increasing & are considered to be linearising &/or variance-stabilising

A middle range example of a variance-stabilising transformation is the log() transformation



→ a log-log transformation of both X & Y often works in this situation (see the "brains" example in tomorrow's lecture)

NB: the default for log() in R is base e, ie natural logarithms to the base e (ln())

not common logarithms $\log_{10}$ in R log(_, base=10) or log10()

or binary logarithms $\log_2$ ← in R log(_, base = 2)

Transformations contd

A log transformation can also be "linearising" in that it make a multiplicative non-linear relationship on the untransformed scale into a linear relationship on the transformed scale

$$Y = c\,X^{d}$$

exponent

constant

$$\begin{aligned} \ln(Y) &= \ln(c X^{d}) \\ &= \ln(c) + \ln(X^{d}) \\ &= \ln(c) + d\,\ln(X) \end{aligned}$$

So, let $Y^* = \ln(Y)$, $X^* = \ln(X)$, $\beta_0 = \ln(c)$ & $\beta_1 = d$

$$Y^* = \beta_0 + \beta_1 . X^*$$

This model is <u>linear</u>

(ie linear in the coefficients $\beta_0, \beta_1, \beta_2, \beta_3, .. \beta_k$ not necessarily linear in the original variables X, Y)

Finally log() is just a mid-range example of transformations such a sqrt(), -ve inverse $(-\frac{1}{x})$ that may help.