Remedial Measures
Normal simple linear regression model:

\[ y_i = \beta_0 + \beta_1 x_i + \epsilon_i \]

\[ x_i, i = 1...n \] fixed (or condition on)
\[ \epsilon_i, i = 1...n \] random errors s.t.
\[ \epsilon_i \sim N(0, \sigma^2), \forall i \]

independent

Graphical diagnostics and/or diagnostic tests may indicate substantial departures from it...

To remedy:

- Adopt more complex models or fitting algorithms
- alternatively, **transform** X and/or Y in a way that makes the model appropriate for the transformed data.
**Non-linearity of the regression function:** Could be addressed transforming either $X$ or $Y$. However, if residuals are reasonably normal and constant variance, it is preferable to work with *transformations of $X$.*

Fit instead:

$$y_i = \beta_0 + \beta_1 \tilde{x}_i + \varepsilon_i$$

Possibly better: fit a *more complex model* that includes both a linear term, and term(s) to capture curvature; e.g. second order polynomial

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$
**Non-constant error variance**: if mean function is reasonably linear, address working with *variance stabilizing transformations of $Y$*.  

Fit instead:

$$\tilde{y}_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

the transformation blows up small values and shrinks large ones; the fan shapes become milder.

Non-constant error variance undermines the optimality of Least Squares estimates (Gauss-Markov Theorem)

Alternative: Re-establish optimality by fitting the original model with **Weighted Least Squares** (larger weights for points in regions of small variance, smaller weights for points in regions of large variance).  

\[ \tilde{y} = \sqrt{y} \]

\[ \tilde{y} = \log(y) \]
Example: Bears data

Fitted Line Plot
Weight = -441.4 + 10.34 Length

Lack of fit test
Possible curvature in variable Length  (P-Value = 0.000 )
Possible lack of fit at outer X-values (P-Value = 0.000)
Overall lack of fit test is significant at P = 0.000

---
evidence of both non-linear mean function and non constant error variance
(issues seldom occur alone!)

---
… need to act on both fronts, checking how things change one step at a time…
dealing with non-linearity: **use a quadratic model**

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$$

most of the curvature accounted for (although there may still be some)

still prominent non-constant error variance
dealing with non-constant error variance: use a variance stabilizing response transfo

\[
\log(y_i) = \beta_0 + \beta_1 x_i + \varepsilon_i
\]

non-constant error variance removed, and with log(y) we don't need the x^2 anymore!

\[
\log(y_i) = \beta_0 + \beta_1 x_i + \varepsilon_i
\]

THIS MODEL WORKS FINE!
Other departures

- **Non-normality of errors**: often variance stabilizing transformations of the response also make residuals more consistent with an iid Gaussian sample.

- **Omission of important predictors**: use more complex models that include them.

- **Non-independent errors**: use more complex models so that errors about them might indeed be reasonably independent, or model first differences, or use models designed to handle dependent errors (e.g. see time series literature).

- **Outlying observations**: if also influential, can remove, or not remove but down-weight using again a Weighted Least Square algorithm to fit the current model.
Box-Cox transformations of the response

Instead of selecting a transformation “by eye”, select an optimal power transformation

\[
\tilde{y} = y^\lambda = \begin{cases} 
1/y^2 \\
1/y \\
1/\sqrt{y} \\
\ln(y) \\
\sqrt{y} \\
y \\
y^2 \\
e tc... 
\end{cases}
\]

- natural log = power 0 by convention
- no transfo = power 1

Fit the model

\[
y_i^\lambda = \beta_0 + \beta_1 x_i + \varepsilon_i
\]

using Gaussian Maximum Likelihood, which simultaneously fits \(\beta_0, \beta_1, \sigma^2, \lambda\).

the selected power provides the best approximation to a linear regression function with constant variance, independent Gaussian errors.
**In MINITAB**

Stat > Control Charts > Box-Cox

Select “All observations for a chart are in one column”

Enter the response Y

Set “Subgroup size” equal sample size n

Will find the power that makes $y_i^\lambda$, $i=1\ldots n$ as close as possible to a sample from a Gaussian distribution.

*(NOTE: this is **NOT** what we need, but can be useful)*