Regression and Linear Models (37252)

Lecture 7 - "Non-linear" Linear Regression

Lecturer: Joanna Wang
Notes adopted from Scott Alexander

School of Mathematical and Physical Sciences, UTS

Autumn 2024

Acknowledgement

These notes are based, in part, on earlier versions prepared by Dr Ed Lidums and Prof. James Brown.

7. Lecture outline

Topics:

- "non-linear" linear regression
- transforming variables
  - linear v. linear
  - linear v. squared
  - linear v. square-root
  - linear v. exponential
  - linear v. log
  - exponential v. exponential
  - "log v. log
  - Box-Cox transform

7. Lecture outline

Topics continued:

- transforming independent variable
  - linear model
  - quadratic model 1
  - quadratic model 2
  - quadratic model 3
  - "polynomial" linear regression
- transforming dependent variable
  - linear model
  - log model

See Chapters 12 and 13 of Draper and Smith (1998).

7. “Non-linear” linear regression

The bulk of our exposure to linear regression has been in the context of fitting lines (simple regression) or planes (multiple regression) to data.

Of course, such models are only useful when there is a linear relationship between the dependent and independent variables.

However, often the relationship between dependent and independent variables is not linear

As we saw in lecture 2, this does not necessarily preclude the use of linear regression – it is sometimes possible to transform the data in a manner that reveals linear relationships between the transformed variables, allowing a linear model to be fit using regression.

In this lecture we look more closely at this issue and various forms of “non-linear” linear regression.

7. “Non-linear” linear regression

All the techniques here involve transformation of dependent or independent variables.

In the first part of the lecture we revisit our “toy” example from Lecture 2 and look at a variety of general situations that can approached by transforming the dependent, independent or both variables before fitting a simple linear regression model.

We then look more closely at transformations of the independent variable in the context of **quadratic linear regression**, a particular form of **polynomial linear regression**. The example we investigate involves extending a simple regression model into a multiple regression model by the addition of the predictor x^2.

We finish with a detailed example of transforming the dependent variable, using a log-transform to reduce non-constant variance.

7. Transforming variables

Recall the toy example from Lecture 2 where we we generated sample data (X_i, Y_i), $i \in \{1, \ldots, n\}$, according to the rule

$$Y_i = 1 + 2X_i + \epsilon_i$$

with $n = 100$, $X_i = \frac{4}{n} i$ and $\epsilon_i \sim \mathrm{N}(0, 1)$.

(Actually, the standard deviation has been increased from $\frac{1}{2}$ to 1 in the examples presented here.)

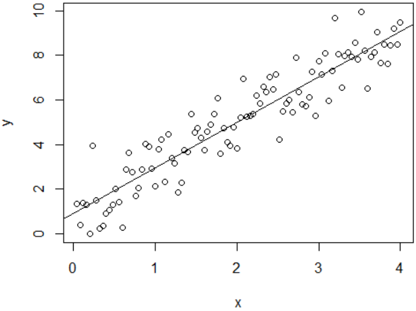
In this section we show some typical situations requiring transformation of dependent, independent or both variable(s).

(Data is available in "toyData.csv" on Canvas.)

7. Transforming variables – linear v. linear

For reference, we begin with an ideal situation, a "linear v. linear relationship".

Scatter plot of data with simple linear regression line

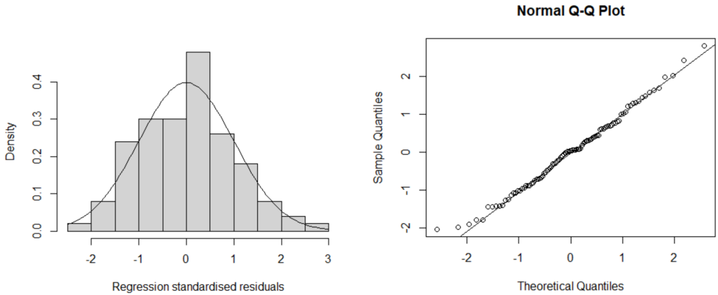


This is a perfect situation for the simple linear model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

7. Transforming variables – linear v. linear

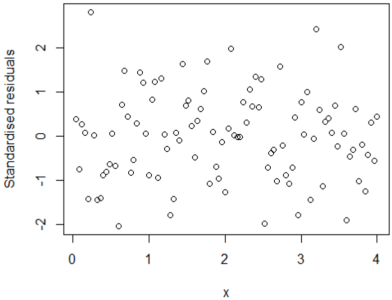
Standardised residual histogram and QQ-Plots for simple linear model

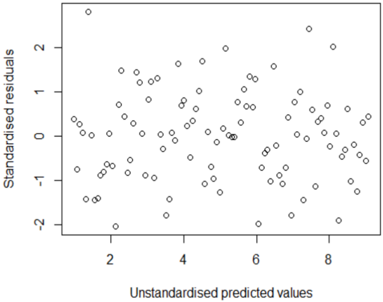


As expected, these plots show no deviations from assumptions ...

7. Transforming variables – linear v. linear

Standardised residual scatter plots for simple linear model



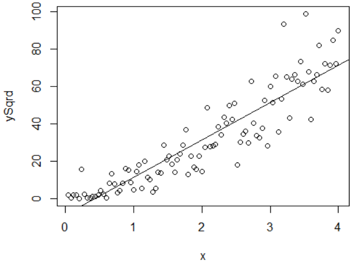


... and nor do the scatter plots show deviations from assumptions.

7. Transforming variables – linear v. squared

In this scenario, the dependent variable data has been squared.

Scatter plot of data with simple linear regression line



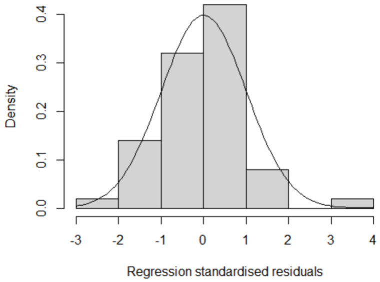
In this case, either the dependent or independent variable may be transformed to fit the model

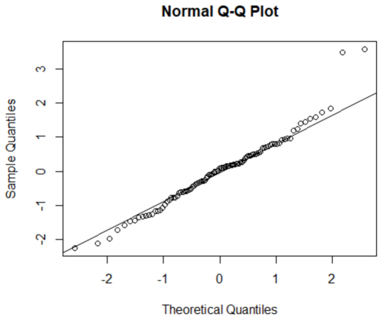
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x^2 \quad \text{or} \quad \sqrt{\hat{Y}} = \hat{\beta}_0 + \hat{\beta}_1 x,$$

choosing the former if the Y_i sample data takes negative values.

7. Transforming variables – linear v. squared

Standardised residual histogram and QQ-Plots for simple linear model

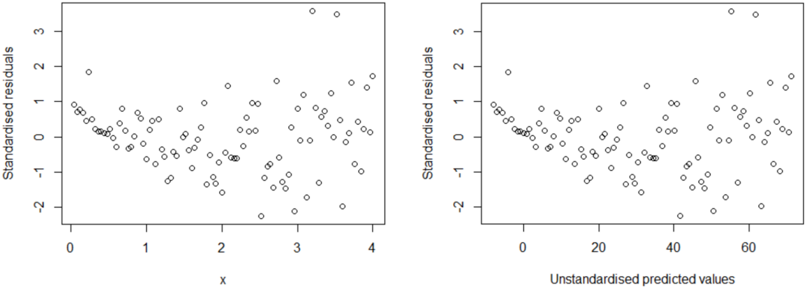




If the data is not transformed and a simple linear model fitted, then the histogram shows outliers and the QQ-Plot departure from $N(0,1)$ ...

7. Transforming variables – linear v. squared

Standardised residual scatter plots for simple linear model

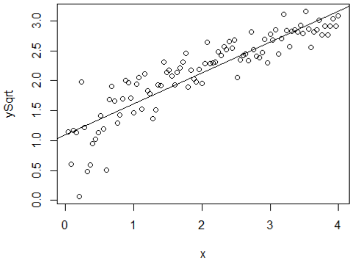


... and the curvature in the sample data now manifests itself in the residuals, which also show non-constant variance.

7. Transforming variables – linear v. square-root

In this scenario, the square-root of the dependent has been taken.

Scatter plot of data with simple linear regression line



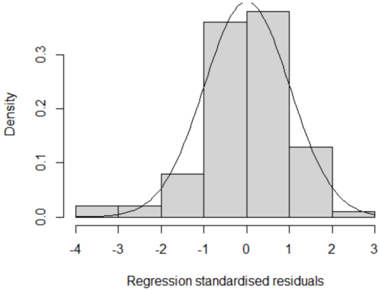
In this case, either the dependent or independent variable may be transformed to fit the model

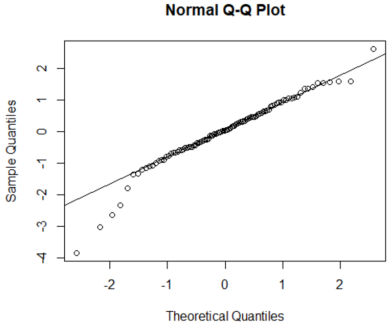
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \sqrt{x} \quad \text{or} \quad \hat{Y}^2 = \hat{\beta}_0 + \hat{\beta}_1 x,$$

choosing the latter if the X_i sample data takes negative values.

7. Transforming variables – linear v. square-root

Standardised residual histogram and QQ-Plots for simple linear model

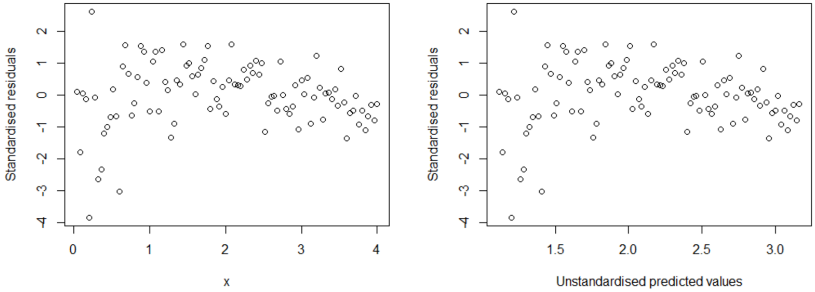




If the data is not transformed and a simple linear model fitted, then the histogram shows outliers and the QQ-Plot departure from $N(0, 1)$ ...

7. Transforming variables – linear v. square-root

Standardised residual scatter plots for simple linear model

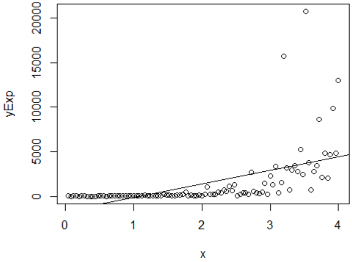


… and the curvature in the sample data now manifests itself in the residuals, which also show non-constant variance.

7. Transforming variables – linear v. exponential

In this scenario, the exponential of the dependent has been taken.

Scatter plot of data with simple linear regression line



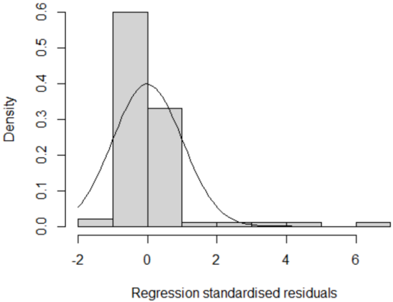
In this case, either the dependent or independent variable may be transformed to fit the model

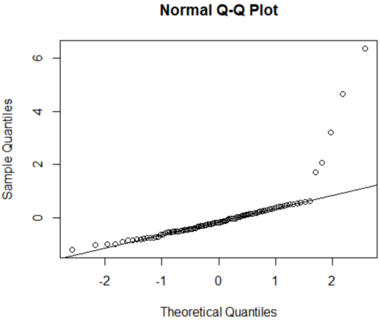
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 e^x \quad \text{or} \quad \log(\hat{Y}) = \hat{\beta}_0 + \hat{\beta}_1 x,$$

choosing the former if the Y_i sample data takes negative values.

7. Transforming variables – linear v. exponential

Standardised residual histogram and QQ-Plots for simple linear model

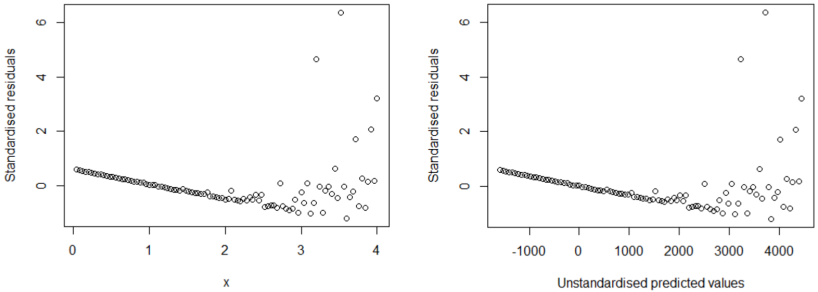




If the data is not transformed and a simple linear model fitted, then the histogram and the QQ-Plot show gross departure from $N(0, 1)$ ...

7. Transforming variables – linear v. exponential

Standardised residual scatter plots for simple linear model

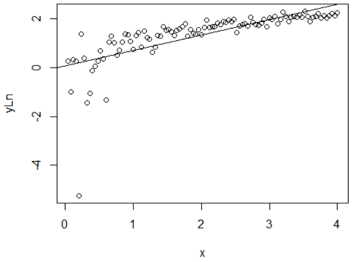


... and the scatter plots show serial correlation and non-constant variance.

7. Transforming variables – linear v. log

In this scenario, the log of the dependent has been taken.

Scatter plot of data with simple linear regression line

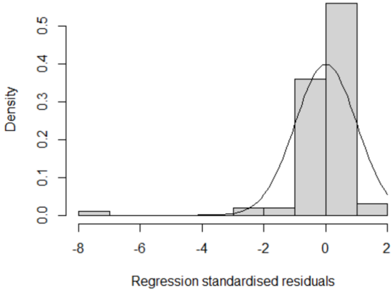


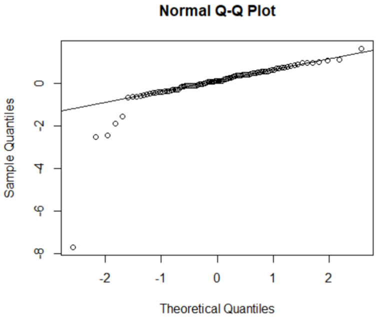
In this case, either the dependent or independent variable may be transformed to fit the model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \log(x) \quad \text{or} \quad e^{\hat{Y}} = \hat{\beta}_0 + \hat{\beta}_1 x,$$

choosing the latter if the X_i sample data takes negative values.

Standardised residual histogram and QQ-Plots for simple linear model

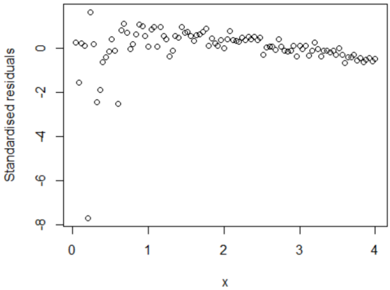


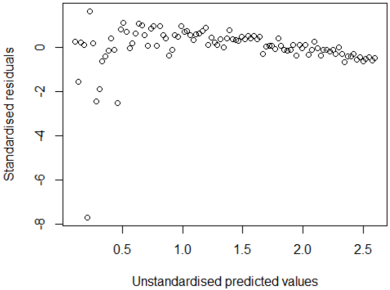


If the data is not transformed and a simple linear model fitted, then the histogram and the QQ-Plot show gross departure from $N(0, 1)$ ...

7. Transforming variables – linear v. log

Standardised residual scatter plots for simple linear model



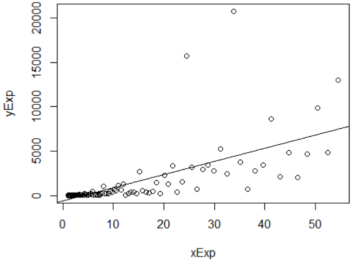


... and the scatter plots show serial correlation and non-constant variance.

7. Transforming variables – exponential v. exponential

In this scenario, the exponential of both variables has been taken.

Scatter plot of data with simple linear regression line



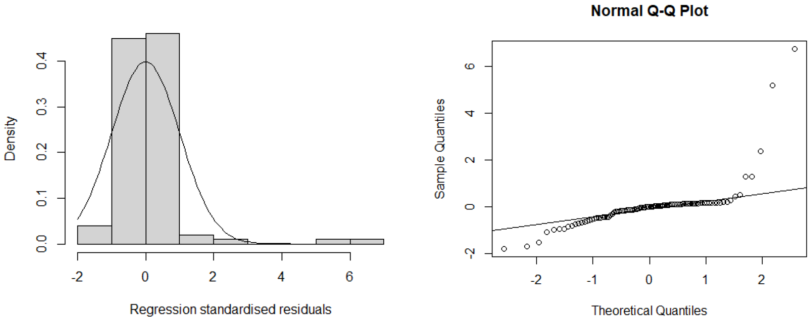
In this case, both variables should be transformed to fit the model

$$\log(\hat{Y}) = \hat{\beta}_0 + \hat{\beta}_1 \log(x),$$

which is only viable if the (X_i, Y_i) sample data does not take negative values.

7. Transforming variables – exponential v. exponential

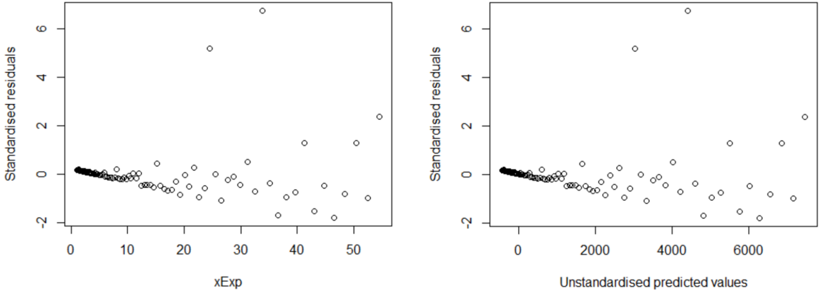
Standardised residual histogram and QQ-Plots for simple linear model



If the data is not transformed and a simple linear model fitted, then the histogram and the QQ-Plot show gross departure from $N(0, 1)$ ...

7. Transforming variables – exponential v. exponential

Standardised residual scatter plots for simple linear model

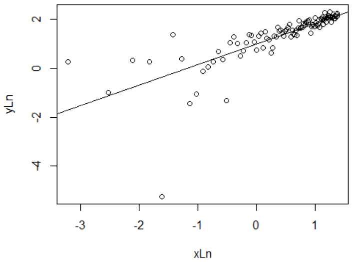


... and the scatter plots show serial correlation and non-constant variance.

7. Transforming variables – log v. log

In this scenario, the log of both variables has been taken.

Scatter plot of data with simple linear regression line

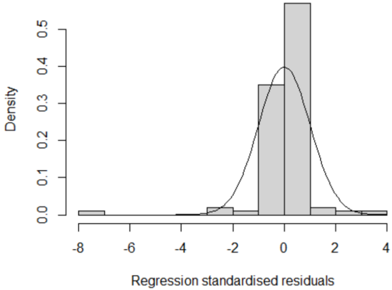


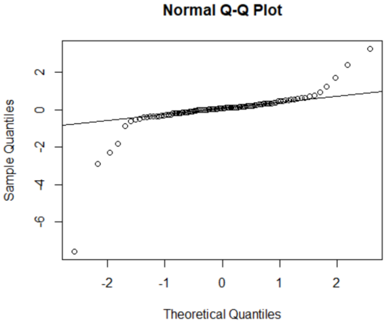
In this case, both variables should be transformed to fit the model

$$e^{\hat{Y}} = \hat{\beta}_0 + \hat{\beta}_1 e^{x}.$$

7. Transforming variables – log v. log

Standardised residual histogram and QQ-Plots for simple linear model

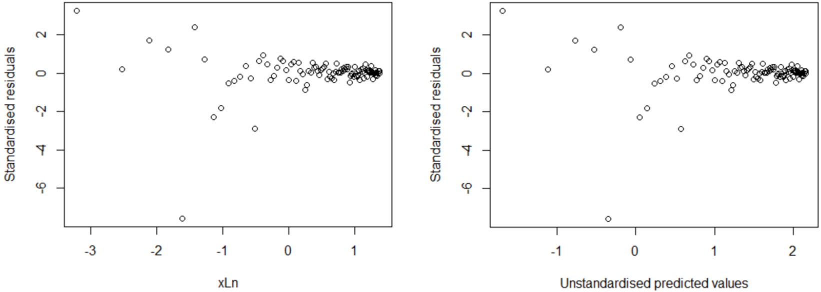




If the data is not transformed and a simple linear model fitted, then the histogram and the QQ-Plot show gross departure from $N(0, 1)$ ...

7. Transforming variables – log v. log

Standardised residual scatter plots for simple linear model



... and the scatter plots show serial correlation and non-constant variance.

7. Transforming variables – Box-Cox transform

Some of the transformations of the dependent variables we have just seen are special cases of the **Box-Cox transform**, which in its simplest form can be described as

$$Y_i^{(\lambda)} = \begin{cases} \frac{Y_i^\lambda}{\lambda}, & \lambda \neq 0 \\ \ln(Y_i), & \lambda = 0 \end{cases}.$$

The Box-Cox transform takes non-normal data and attempts to transform it so that it behaves as if normal.

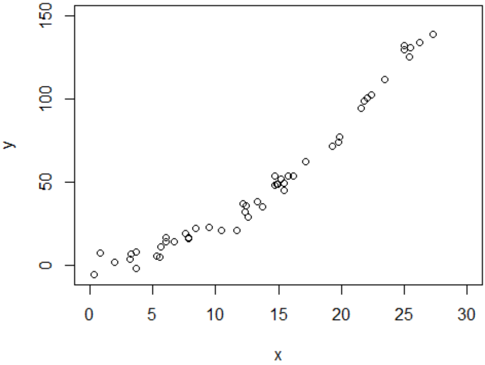
It is clear that for many choices of λ, the Box-Cox transform can only be applied to non-negative dependent data Y_i.

We don't have time time to go into details in this course.

7. Transforming independent variable

Now we look more closely at transforming the independent variables.

Suppose we have some sample data of (X_i, Y_i) pairs to which we wish to fit a regression model. Below is a scatter plot.



(Data is available in "quad.csv" on Canvas.)

7. Transforming independent variable – linear model

Ignoring what appears to be a small degree of curvature, we proceed and fit the simple linear model

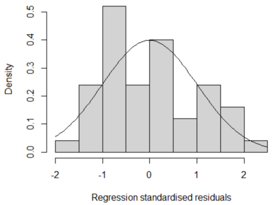
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x. \tag{1}$$

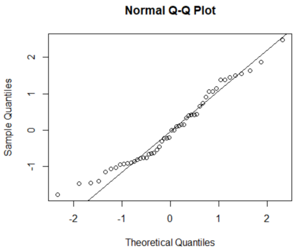
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -23.3254     3.1845  -7.325 2.34e-09 ***
x             5.4191     0.2071  26.170  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 11 on 48 degrees of freedom
Multiple R-squared:  0.9345,Adjusted R-squared:  0.9331
F-statistic: 684.9 on 1 and 48 DF,  p-value: < 2.2e-16
```

The output shows both parameters to be statistically-significant and that the model captures over 90% of the variation in the sample data.

7. Transforming independent variable – linear model

We move on to a visual inspection of the residuals.



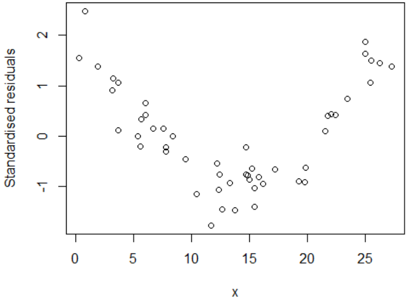


These plots don't look too bad with no gross departures from normality evident, an observation confirmed by a formal test of normality.

```
> shapiro.test(mod8$residuals)
Shapiro-Wilk normality test

data:  mod8$residuals
W = 0.96506, p-value = 0.1448
```

7. Transforming independent variable – linear model

Looking for serial correlation and/or nonconstant variance we also view the scatter plot of standardised residuals against standardised predicted values.



This serial correlation is revealed through visual inspection of the standardised residuals when plotted against standardised fitted values.

By ignoring curvature in the sample data and mis-specifying a linear model, the curvature now manifests itself in the residuals.

7. Transforming independent variable – quadratic model 1

There are a variety of ways to proceed from this point.

One is to attempt the capture of the curvature through the addition of the quadratic term x^2 to the model, the square of the independent variable x.

Specifically, we fit the **quadratic linear regression model**

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 \tag{2}$$

which is nothing more than a multiple regression model with predictors x and x^2.

This is a perfectly legitimate model as the independent variables are not linearly dependent.

However, we should expect high multicollinearity, given the close relationship between x and x^2.

7. Transforming independent variable – quadratic model 1

R provides the following information for the quadratic model (2).

```
> quad$x2 <- quad$x^2
> mod9 <- lm(y ~ x + x2,  data = quad)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.27885    1.92855   0.663     0.51
x            0.39800    0.31344   1.270     0.21
x2           0.18085    0.01092  16.568   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 4.251 on 47 degrees of freedom
Multiple R-squared:  0.9904,Adjusted R-squared:   0.99
F-statistic:  2431 on 2 and 47 DF,  p-value: < 2.2e-16
> vif(mod9)
       x       x2
15.34707 15.34707
```

We immediately spot a problem with this model: the variance inflation factor warns us of excessive multicollinearity between x and x^2 (variance inflation factor is greater than threshold value of ten).

CONCLUSION: parameter estimates may be unreliable.

7. Transforming independent variable – quadratic model 2

Removing the linear term x gives the pure quadratic model

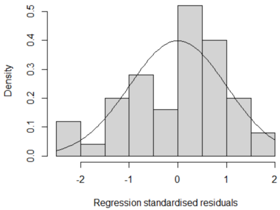
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_2 x^2. \tag{3}$$

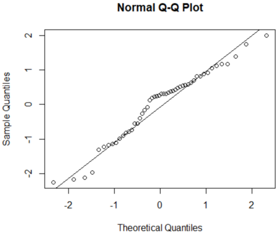
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.450015   0.897632   3.843 0.000356 ***
x2          0.194249   0.002804  69.276  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.278 on 48 degrees of freedom
Multiple R-squared:  0.9901,Adjusted R-squared:  0.9899
F-statistic:  4799 on 1 and 48 DF,  p-value: < 2.2e-16
```

Note the change in $S_{\hat{\beta}_2}$, the standard error of $\hat{\beta}_2$, falling from 0.011 to 0.003 and hence tighter 95% CI for $\hat{\beta}_2$.

7. Transforming independent variable – quadratic model 2

We inspect the histogram and QQ-Plot of the residuals.



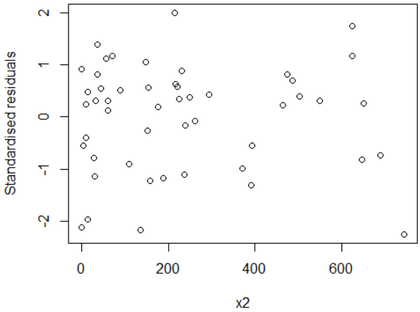


These plots aren't the most reassuring and a formal statistical test shows the residuals just squeak by without rejection of the null hypothesis of normality (0.05 significance level, 0.077 p-value).

```
Shapiro-Wilk normality test

data:  mod10$residuals
W = 0.95841, p-value = 0.07622
```

7. Transforming independent variable – quadratic model 2

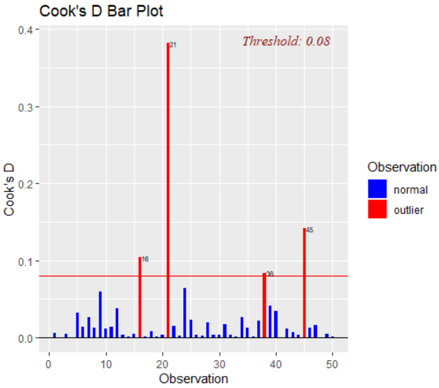
We also check the scatter plot of standardised residuals.



Certainly an improvement over the first model.

7. Transforming independent variable – quadratic model 2

Perhaps the model fit can be further improved by removing any influential points?



A box plot shows sample point 21 to be a point of influence, with Cook's D well in excess of the threshold value $\frac{4}{n-m-1} = \frac{1}{12}$ or $\frac{4}{n}$=0.08

CONCLUSION: regression model oversensitive to point 21.

7. Transforming independent variable – quadratic model 3

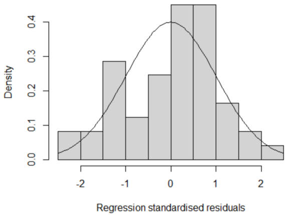
Filtering out record number 21 from the data set we again fit the model.

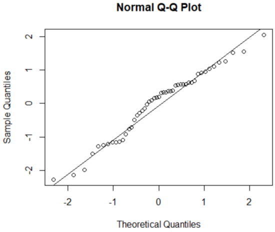
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_2 x^2.$$

```
> quad_filtered <- quad[-21, ]
> mod11 <- lm(y ~ x2, data = quad_filtered)
> summary(mod11)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.123186   0.869112   3.594 0.000779 ***
x2          0.196504   0.002846  69.044  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.088 on 47 degrees of freedom
Multiple R-squared:  0.9902,Adjusted R-squared:   0.99
F-statistic:  4767 on 1 and 47 DF,  p-value: < 2.2e-16
```

7. Transforming independent variable – quadratic model 3

Inspecting the histogram and QQ-Plot of the residuals.



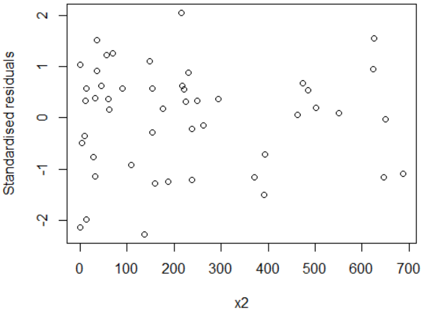


Removal of the point of influence has improved the behaviour of the residuals to a small degree, with the p-value associated with the normality test now at 0.187.

```
Shapiro-Wilk normality test

data:  mod11$residuals
W = 0.96723, p-value = 0.1874
```

7. Transforming independent variable – quadratic model 3

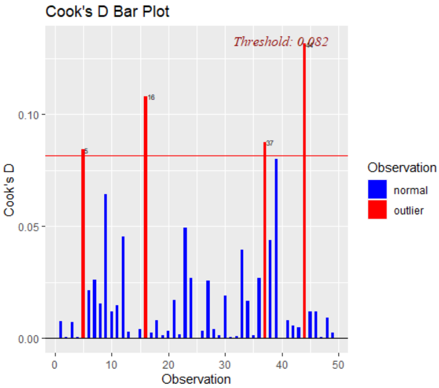
Inspecting the scatter plot of standardised residuals.



Looking pretty good.

7. Transforming independent variable – quadratic model 3

We again check for points of influence.



There are a few points above the threshold value $\frac{4}{n-m-1}$ or $\frac{4}{n} = 0.082$, and it way be worth removing points 44 and 16 and repeating the analysis – we don't purse this here.

CONCLUSION: we have found a viable model.

7. Transforming ind. variable – “polynomial” regression

Quadratic linear regression is a special case of **polynomial linear regression**, which itself is just a form of multiple regression where the independent variables are powers of x.

Generally, this type of regression involves fitting models of the form

$$\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^{m} \hat{\beta}_j x^j.$$

As the preceding examples demonstrate, given the independent variables $x, x^2, \ldots, x^m$ can be expected to be highly correlated, we need to be wary of excessive multicollinearity.

On the other hand, over the last few lectures we have seen that linear regression is quite a robust modelling technique and can handle high levels of correlation between independent variables before multicollinearity becomes an issue.

7. Transforming ind. variable – "polynomial" regression

We can also go further and include categorical variables.

For instance, to include a categorical variable defined for M categories we would choose a reference category, encode $M - 1$ dummy variables and fit the model

$$\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^{m} \hat{\beta}_j x^j + \sum_{j=1}^{M-1} \hat{\gamma}_j z_j.$$

(See Lecture 6 for details).

The important point is that we perform the usual analysis:

- check residuals for normality, constant variance and serial correlation
- check parameters for significance using F and/or T-tests
- use Cook's D or DFITS to identify influential points, removing them where necessary and re-fitting the model
- where more than one viable model survives the analysis, choose that with the highest R^2_{adj}.

7. Transforming dependent variable

Sometimes it makes more sense to apply transformations to the dependent variable.

This is especially the case for multiple regression, where instead of potentially having to apply different transformations to *m* independent variables, we can (hopefully) apply a single transformation to the dependent variable.

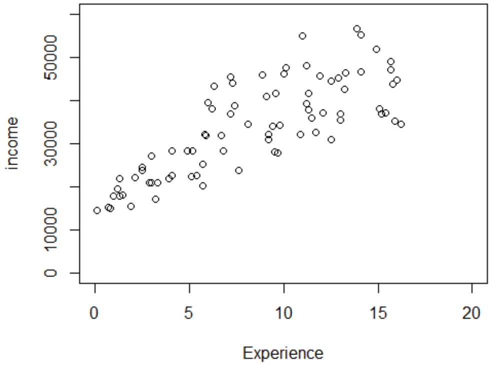
This technique is especially well-suited to the case of **heteroskedastic** residuals, i.e. residuals with non-constant variance $\sigma_i^2 \propto f(Y_i)$, with f some transformation applied to the sample dependent data Y_i.

We have already seen some examples situations with the toy examples explored in the first part of the lecture.

We now go into more detail using some real data.

7. Transforming dependent variable – linear model

Suppose we have some sample data of $(experience_i, income_i)$ pairs to which we wish to fit a regression model. A scatter plot is below.



(Data is available in “income.csv” on Canvas.)

7. Transforming dependent variable – linear model

We begin by fitting the model

$$\widehat{income} = \hat{\beta}_0 + \hat{\beta}_1 experience.$$

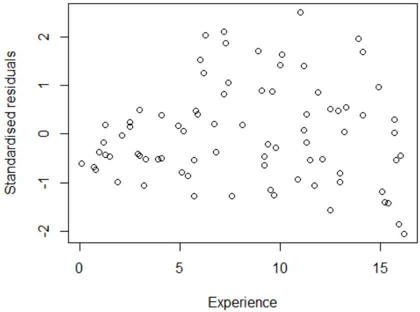
R provides the following information.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  18292.7     1541.7   11.87   <2e-16 ***
experience    1835.0      160.4   11.44   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6711 on 78 degrees of freedom
Multiple R-squared:  0.6267,Adjusted R-squared:  0.6219
F-statistic: 130.9 on 1 and 78 DF,  p-value: < 2.2e-16
```

The output shows both parameters to be statistically-significant with the model capturing just over 60% of the variation in the sample data.

7. Transforming dependent variable – linear model

To see what the issue is here we inspect the standardised residuals when plotted against the idnependerce variable.



The plot shows strong evidence of increasing variance.

CONCLUSION: the modelling assumptions are violated.

7. Transforming dependent variable – log model

We consider a **log transformation of the dependent variable** and fit the model

$$\log(\widehat{income}) = \hat{\beta}_0 + \hat{\beta}_1 experience.$$

*ASIDE: note that if we take the exponential of both sides of the fitted model equation we obtain the model in **original units***

$$\widehat{income} = e^{\hat{\beta}_0} e^{\hat{\beta}_1 experience}.$$

*This is an example of a **multiplicative model**.*

7. Transforming dependent variable – log model

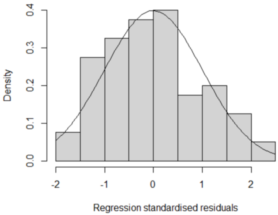
R provides the following output.

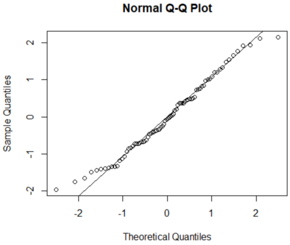
```
> income$ln_income <- log(income$income)
> mod13 <- lm(ln_income ~ experience, data = income)
> summary(mod13)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 9.853621   0.047369  208.02   <2e-16 ***
experience  0.061198   0.004927   12.42   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2062 on 78 degrees of freedom
Multiple R-squared:  0.6642,Adjusted R-squared:  0.6599
F-statistic: 154.3 on 1 and 78 DF,  p-value: < 2.2e-16
```

The fit of this model has improved, as shown by R^2_{adj} increasing from 0.627 to 0.660.

7. Transforming dependent variable – log model

Has the behaviour of the residuals improved?



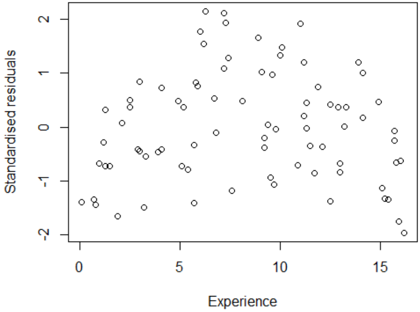



```
Shapiro-Wilk normality test

data:  mod13$residuals
W = 0.97893, p-value = 0.2093
```

The formal test of normality indicates the assumption of normal residuals cannot be rejected (p-value 0.209).

7. Transforming dependent variable – log model

To check if increasing variance of the residuals remains a problem we inspect the standardised residuals plotted against the independence variable.



The log transformation of the dependent variable has definitely improved matters, although a hint of negative curvature can still be perceived in the scatter plot.

References I

Draper, N. R. and Smith, H. (1998). *Applied regression analysis*. Wiley-Interscience, Somerset, US.