6.5 Sums of Squares and ANOVA

We look at an alternative test, the analysis of variance (ANOVA) test for the slope parameter, $H_0 : m = 0$, of the simple linear model,

$$Y = b + mX + \epsilon,$$

where, in particular, ϵ is $N(0, \sigma^2)$, where the ANOVA table is

Source	Sum Of Squares	Degrees of Freedom	Mean Squares
Regression	$SS_{Reg} = \sum(\hat{y}_i - \bar{y})^2$	1	$MS_{Reg} = \frac{SS_{Reg}}{1}$
Residual	$SS_{Res} = \sum(y_i - \hat{y}_i)^2$	n - 2	$MS_{Res} = \frac{SS_{Res}}{n-2}$
Total	$SS_{Tot} = \sum(y_i - \bar{y})^2$	n - 1	

where

$$f = \frac{MS_{Reg}}{MS_{Res}},$$

with corresponding critical value $f_\alpha(1, n-2)$. Related to this, the average of the y

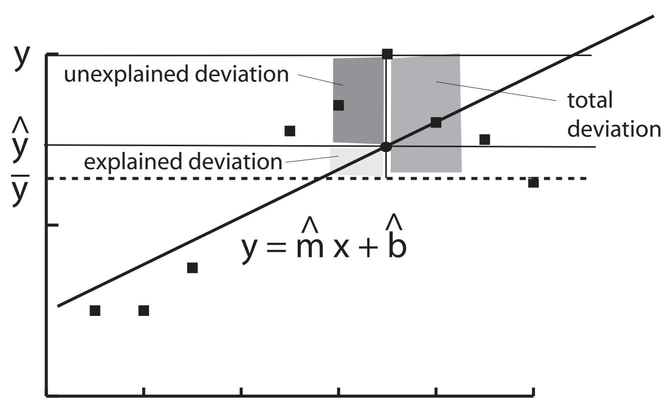

Figure 6.13: Types of deviation

variable, $\bar{y}$, is a kind of baseline and since

$$\underbrace{(y - \bar{y})}_{\text{total deviation}} = \underbrace{(\hat{y} - \bar{y})}_{\text{explained deviation}} + \underbrace{(y - \hat{y})}_{\text{unexplained deviation}},$$

then taking sum of squares over all data points,

$$\underbrace{\sum(y - \bar{y})^2}_{\text{total variation}} = \underbrace{\sum(\hat{y} - \bar{y})^2}_{\text{explained variation}} + \underbrace{\sum(y - \hat{y})^2}_{\text{unexplained variation}}$$

and so

$$r^2 = \frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{SS_{Tot} - SS_{Res}}{SS_{Tot}} = \frac{SS_{Reg}}{SS_{Tot}} = \frac{\text{explained variation}}{\text{total variation}},$$

the *coefficient of determination*, is a measure of the proportion of the total variation in the y-values from $\bar{y}$ explained by the regression equation.

Exercise 6.5 (Sums of Squares and ANOVA)

1. *ANOVA of slope m using test statistic: reading ability vs brightness.*

illumination, x	1	2	3	4	5	6	7	8	9	10
ability to read, y	70	70	75	88	91	94	100	92	90	85

Use the ANOVA procedure to test if the slope m is zero at $\alpha = 0.05$, compare test statistic with critical value; also, find r^2.

(a) *Statement.*

i. $H_0 : m = 0$ versus $H_1 :\ m > 0$.
ii. $H_0 : m = 0$ versus $H_1 :\ m < 0$.
iii. $H_0 : m = 0$ versus $H_1 :\ m \neq 0$.

(b) *Test.* the ANOVA table is given by,

Source	Sum Of Squares	Degrees of Freedom	Mean Squares
Regression	482.4	1	482.4
Residual	490.1	8	61.3
Total	972.5	9	

and so the test statistic is

$$f = \frac{MS_{Reg}}{MS_{Res}} = \frac{482.4}{61.3} \approx$$

(i) **6.88** (ii) **7.88** (iii) **8.88**.
and the critical value at $\alpha = 0.05$, with 1 and 8 df, is
(i) **5.32** (ii) **6.32** (ii) **7.32**

```
brightness <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
reading.ability <- c(70, 70, 75, 88, 91, 94, 100, 92, 90, 85)
linear.regression.ANOVA(brightness, reading.ability, 0.05)

                        SS df               MS                F
Regression 482.427272727273  1 482.427272727273 7.87519477628553
Residual   490.072727272727  8 61.2590909090909
Total                 972.5  9

   intercept        slope          r^2 F crit value  F test stat       p value
    72.20000      2.41818      0.49607      5.31766      7.87519       0.02297
```

(c) *Conclusion.*
Since test statistic = 7.88 > critical value = 5.32,
(i) **do not reject** (ii) **reject** null $H_0 : m = 0$.
Data indicates population slope
(i) **equals** (ii) **does not equal** (iii) **greater than** zero (0).
In other words, reading ability
(i) **is** (ii) **is not** associated with brightness.

(d) *Coefficient of Determination.*

$$r^2 =$$

(i) **0.49** (ii) **0.50** (iii) **0.51**
in other words, regression explains
(i) **49%** (ii) **50%** (iii) **51%**
of the total variation in the scatterplot

(e) *Other statistics.* The degrees of freedom for the regression are (always) 1 and for the residual are $n - 2 = 10 - 2 = 8$. Also,

$$SS_{Reg} =$$

(i) **482.4** (ii) **582.4** (iii) **682.4**

$$SS_{Res} =$$

(i) **682.4** (ii) **882.4** (iii) **972.5**

2. *ANOVA of slope m with p-value: reading ability vs brightness.*

illumination, x	1	2	3	4	5	6	7	8	9	10
ability to read, y	70	70	75	88	91	94	100	92	90	85

Use the ANOVA procedure to test if the slope m is zero; compare p-value with level of significance at $\alpha = 0.05$.

(a) *Statement.*

i. $H_0 : m = 0$ versus $H_1 : m > 0$.
ii. $H_0 : m = 0$ versus $H_1 : m < 0$.
iii. $H_0 : m = 0$ versus $H_1 : m \neq 0$.

(b) *Test.* Since the test statistic is $F = 7.88$, the p–value, with 1 and $n - 2 = 10 - 2 = 8$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 7.88)$$

which equals (i) **0.00** (ii) **0.022** (iii) **0.043**.
The level of significance is 0.05.

(c) *Conclusion.* Since p–value, 0.022, is smaller than level of significance, 0.05, we (i) **fail to reject** (ii) **reject** null hypothesis the slope m is zero.

(d) *Comment.* Conclusions reached here using F–distribution with the ANOVA procedure are (i) **the same as** (ii) **different from** the conclusions reached previously using the t–distribution.

3. *ANOVA of slope m using test statistic: response vs drug dosage.* The responses of fifteen different patients are measured for one drug at three dosage levels (in mg).

10 mg	20 mg	30 mg
5.90	5.51	5.01
5.92	5.50	5.00
5.91	5.50	4.99
5.89	5.49	4.98
5.88	5.50	5.02
$\bar{x}_1 \approx 5.90$	$\bar{x}_2 \approx 5.50$	$\bar{x}_3 \approx 5.00$

Use the ANOVA procedure to test if the slope m is zero at $\alpha = 0.05$, compare test statistic with critical value; also, find r^2.

(a) *Statement.*

i. $H_0 : m = 0$ versus $H_1 : m > 0$.
ii. $H_0 : m = 0$ versus $H_1 : m < 0$.
iii. $H_0 : m = 0$ versus $H_1 : m \neq 0$.

(b) *Test.* the ANOVA table is given by,

Source	Sum Of Squares	Degrees of Freedom	Mean Squares
Regression	2.025	1	2.025
Residual	0.0105	13	0.00081
Total	2.0355	14	

and so the test statistic is

$$f = \frac{MS_{Reg}}{MS_{Res}} \approx \frac{2.025}{0.00081} \approx$$

(i) **2299.2** (ii) **2399.2** (iii) **2499.2**.
and the critical value at $\alpha = 0.05$, with 1 and 13 df, is
(i) **4.67** (ii) **6.32** (ii) **7.32**

```
dosage <- c(10, 10, 10, 10, 10, 20, 20, 20, 20, 20, 30, 30, 30, 30, 30)
response <- c(5.90, 5.92, 5.91, 5.89, 5.88, 5.51, 5.50, 5.50, 5.49, 5.50, 5.01, 5.00, 4.99, 4.98, 5.02)
linear.regression.ANOVA(dosage, response, 0.05)
```

```
                            SS df                   MS                F
Regression               2.025  1                2.025 2499.20886075947
Residual   0.0105333333333334 13 0.000810256410256419
Total        2.03553333333333 14

   intercept        slope          r^2 F crit value  F test stat      p value
   6.367e+00   -4.500e-02    9.948e-01    4.667e+00    2.499e+03    2.220e-16
```

(c) *Conclusion.*
Since test statistic = 2499.2 > critical value = 4.67,
(i) **do not reject** (ii) **reject** null $H_0 : m = 0$.
Data indicates population slope
(i) **equals** (ii) **does not equal** (iii) **greater than** zero (0).
In other words, response
(i) **is** (ii) **is not** associated with dosage.

(d) *Coefficient of Determination.*

$$r^2 =$$

(i) **0.09** (ii) **0.10** (iii) **0.99**
in other words, regression explains
(i) **9%** (ii) **10%** (iii) **99%**
of the total variation in the scatterplot

(e) *Comparing ANOVA of linear regression with ANOVA of means.*
Recall, fifteen different patients, chosen at random, subjected to three *different* drugs. Test if at least one of the three mean patient responses (notice, all the same as above) to drug is different at $\alpha = 0.05$.

drug 1	drug 2	drug 3
5.90	5.51	5.01
5.92	5.50	5.00
5.91	5.50	4.99
5.89	5.49	4.98
5.88	5.50	5.02
$\bar{x}_1 \approx 5.90$	$\bar{x}_2 \approx 5.50$	$\bar{x}_3 \approx 5.00$

The ANOVA test of means is

- $H_0 : m = 0$ versus $H_1 : m \neq 0$,
- H_0 : means same vs H_1 : at least one of the means different,

(i) **the same** (ii) **different from** the ANOVA test of linear regression.

The ANOVA of means table is

Source	Sum Of Squares	Degrees of Freedom	Mean Squares
Treatment	2.033	2	1.0167
Residual	0.0022	12	0.00018
Total	2.0355	14	

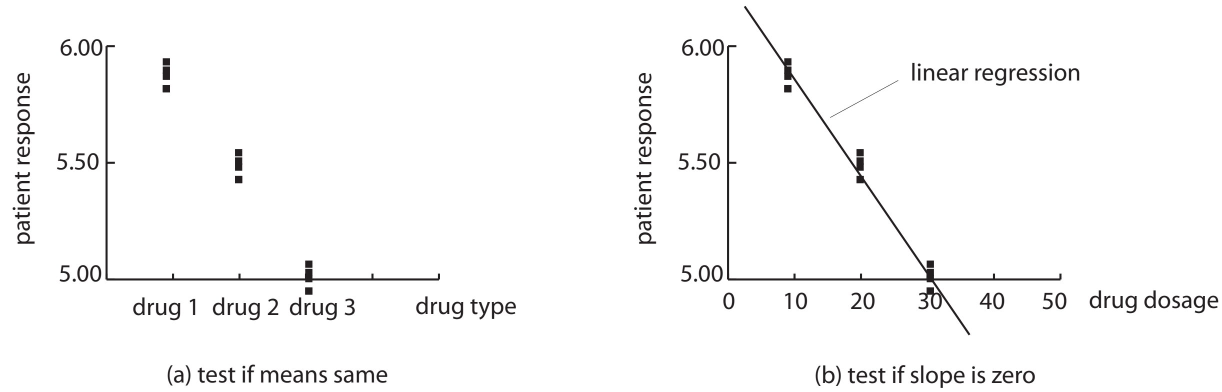

Figure 6.14: ANOVA of means vs ANOVA of slope

where

$$f = \frac{MS_{Reg}}{MS_{Res}} \approx \frac{1.0167}{0.00018} \approx 5648$$

(i) **the same** (ii) **different from** the ANOVA table of linear regression.

The ANOVA of means requires
(i) **fewer** (ii) **more**
assumptions than ANOVA of linear regression.

6.6 Nonlinear Regression

Scatterplots of nonlinear data can be fit with hypothesized (guessed) nonlinear equations using different methods. The method described in this text involves converting a nonlinear equation to a linear equation form where the original nonlinear parameters and variables (data) have been transformed to conform to this linear form. A least-squares regression performed on this created linear equation form results in estimates of the transformed parameters which can then be un-transformed to give estimates of the original nonlinear parameters. Furthermore, the coefficient of determination, r^2 of the linear model to the transformed data is used to measure the "fit" of the nonlinear model to the original data. Four nonlinear models are considered.

description	nonlinear model	linear transformation	variable transformed
logarithmic	$e^{\frac{y}{b}} = xe^{\frac{a}{b}}$	$y = a + b\ln x$	x only
exponential	$y = ae^{bx}$	$\ln y = \ln a + bx$	y only
power	$y = ax^b$	$\ln y = \ln a + b\ln x$	both x and y
logistic	$y = \frac{L}{1+e^{a+bx}}$	$\ln\left(\frac{L-y}{y}\right) = a + bx$	y only, for binary data

Exercise 6.6 (Nonlinear Regression)

1. *Linearize nonlinear models of "data" derived from mathematical functions.*
   Let $y = 75 - 2x^2$ then complete the following table.

x	1	2	3	4	5
x^2	1	4	9	______	______
y	73	67	57	______	______

Nonlinear function $y = 75 - 2x^2$ is linearized by transforming (i) $\boldsymbol{x}$ (i) $\boldsymbol{y}$ axis.

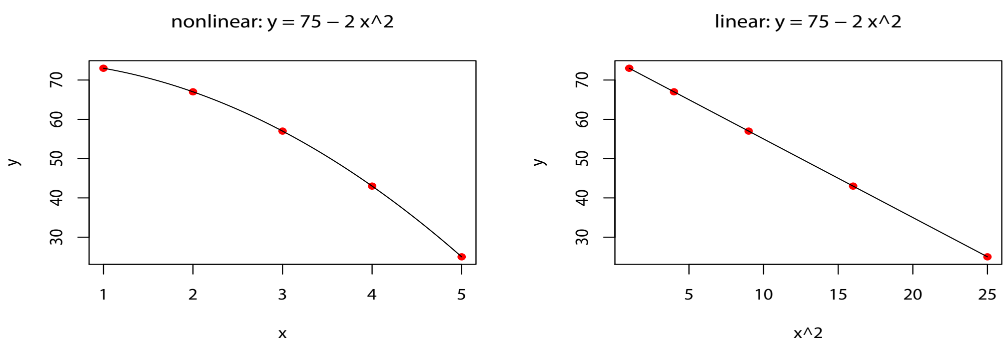

Figure 6.15: Nonlinear and linear version of $y = 75 - 2x^2$

Using the 5 (x, y) data points, regress y on x^2 (rather than x), and "discover" intercept (i) $\mathbf{-2}$ (i) **75**, slope (i) $\mathbf{-2}$ (i) **75** and $r^2 =$ (i) **0** (i) **1** because these points (i) **perfectly** (ii) **imperfectly** fit linearized model $y = 75 - 2x^2$. Typically, linear models (i) **do** (ii) **do not** *perfectly* fit *sampled* (x, y) data.

```
x <- c(1, 2, 3, 4, 5)
y <- c(73, 67, 57, 43, 25)
linear.regression.ANOVA(x^2, y, 0.05)
```

```
             SS df   MS   F
Regression 1496  1 1496 Inf
Residual      0  3    0
Total      1496  4

   intercept        slope          r^2 F crit value  F test stat       p value
       75.00        -2.00         1.00        10.13          Inf          0.00
```

2. *Nonlinear models of data: reading ability vs brightness.*

illumination, x	1	2	3	4	5	6	7	8	9	10
ability to read, y	70	70	75	88	91	94	100	92	90	85

Apply various nonlinear models to the data, predict reading ability at $x = 7.5$, measure fit of each model by calculating r^2 of linearized versions of the nonlinear regressions.

```
brightness <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
reading.ability <- c(70, 70, 75, 88, 91, 94, 100, 92, 90, 85)
```

(a) *Original linear model.* Least-squares linear model is

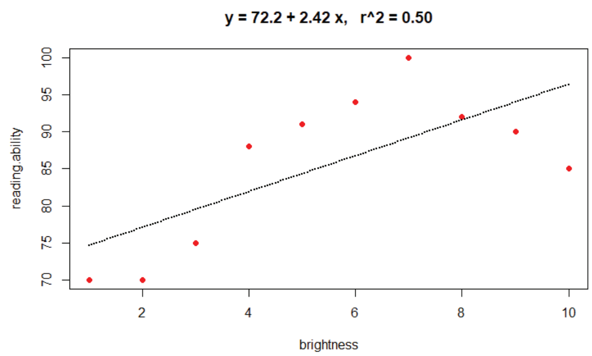

Figure 6.16: Linear model, no transformation

i. $y = 68.091 + 11.526 \ln x$
ii. $y = 72.2 + 2.42x$
iii. $y = 68.091 - 11.526 \ln x$

and, at $x = 7.5$ for example, $\hat{y} = 72.2 + 2.42(7.5) \approx$
(i) **90.17** (ii) **91.31** (iii) **91.34** (iv) **92.55**

but because $r^2 =$ (i) **0.50** (ii) **0.52** (iii) **0.66** (iv) **0.69**,
only 50% of variation is explained by linear regression and so prediction at $x = 7.5$ is (i) **poor** (ii) **good**.

```
linear.regression.predict(brightness, reading.ability, x.zero=7.5)
plot(d,pch=16,col="red",xlab="brightness",ylab="reading.ability",
     main="y = 72.2 + 2.42 x,   r^2 = 0.50") # original, linear model
x0 <- seq(1,10,0.05)
y0 <- 72.2 + 2.42 * x0
points(x0,y0,pch=16,cex=0.2,col="black")
r2 <- cor(x,y)^2; r2
```

```
   intercept        slope            x y.predict(x)
   72.200000     2.418182     7.500000    90.336364
> r2 <- cor(x,y)^2; r2
[1] 0.4960692
```

(b) *Nonlinear logarithmic model.*

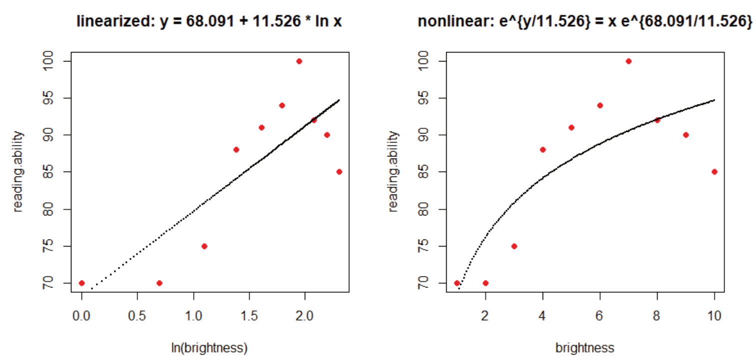

Figure 6.17: Logarithmic transformation

```
nonlinear.regression(brightness, reading.ability, 1, "logarithmic")

     transformation  trans.intercept, a        intercept, a             slope, b                    r^2
      "logarithmic"    "68.091307394593"   "68.091307394593"  "11.5255674599614"  "0.660562267926854"
```

To fit the nonlinear logarithmic model

$$e^{\frac{y}{b}} = xe^{\frac{a}{b}}$$

to the data, first convert (if possible) to a linear equation:

$$\begin{aligned} \frac{y}{b} &= \ln x + \frac{a}{b}, \quad \text{take ln on both sides} \\ y &= b \ln x + a \quad \text{multiple both sides by } b \end{aligned}$$

then take a least-squares approximation of this linear transformation,

i. $y = 68.091 + 11.526 \ln x$
ii. $\ln y = 4.276 + 0.030x$
iii. $\ln y = 4.226 + 0.143 \ln x$
iv. $\ln \left(\frac{101-y}{y} \right) = -0.961 - 0.191x$

where $r^2 =$ (i) **0.27** (ii) **0.52** (iii) **0.66** (iv) **0.69**

whereas the logarithmic regression itself is

i. $y = \frac{101}{1+e^{-0.961-0.191x}}$
ii. $e^{\frac{y}{11.526}} = xe^{\frac{68.091}{11.526}}$
iii. $y = 72.005e^{0.030x}$
iv. $y = 68.460x^{0.143}$

and, at $x = 7.5$, $e^{\frac{\hat{y}}{11.526}} = 7.5e^{\frac{68.091}{11.526}}$ or $\hat{y} = 68.091 + 11.526 \ln(7.5) \approx$
(i) **90.17** (ii) **91.31** (iii) **91.32** (iv) **92.55**

(c) *Nonlinear exponential model.*

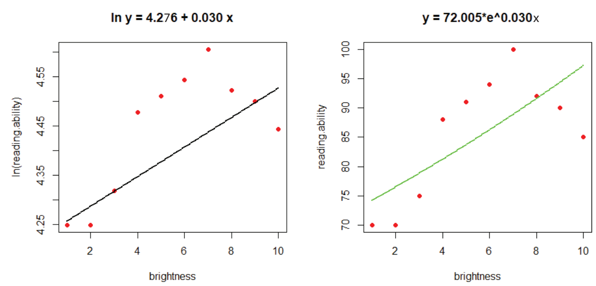

Figure 6.18: Exponential transformation

```
nonlinear.regression(brightness, reading.ability, 1, "exponential")

      transformation   trans.intercept, a         intercept, a               slope, b                        r^2
       "exponential"    "4.2767375112164"   "72.0051404219156" "0.0299638959744328"  "0.518078387957388"
```

To fit the nonlinear exponential model

$$y = ae^{bx}$$

to the data, first convert to a linear equation:

$$\ln y = \ln a + bx, \quad \text{take ln on both sides}$$

then take a least-squares approximation of this linear transformation,

i. $y = 68.091 + 11.526 \ln x$
ii. $\ln y = 4.276 + 0.030x$
iii. $\ln y = 4.226 + 0.143 \ln x$
iv. $\ln \left(\frac{101-y}{y} \right) = -0.961 - 0.191x$

where $r^2 =$ (i) **0.27** (ii) **0.52** (iii) **0.66** (iv) **0.69**

whereas the exponential regression itself is

i. $y = \frac{101}{1+e^{-0.961-0.191x}}$
ii. $e^{\frac{y}{11.526}} = 7.5e^{\frac{68.091}{11.526}}$
iii. $y = 72.005e^{0.030x}$
iv. $y = 68.460x^{0.143}$

and, at $x = 7.5$, $\hat{y} = 72.005(e)^{0.030(7.5)} \approx$
(i) **90.17** (ii) **91.31** (iii) **91.32** (iv) **92.55**

(d) *Nonlinear power model.*

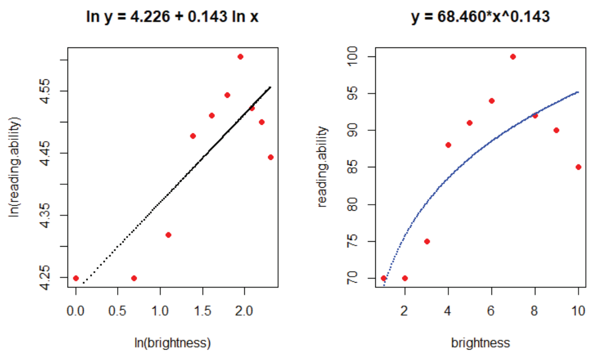

Figure 6.19: Power transformation

```
nonlinear.regression(brightness, reading.ability, 1, "power")

     transformation  trans.intercept, a          intercept, a              slope, b                  r^2
           "power"  "4.22624256172365"  "68.4595158951469" "0.142538729202824" "0.687209998444701"
```

To fit the nonlinear power model

$$y = ax^b$$

to the data, first convert to a linear equation:

$$\ln y = \ln a + b \ln x, \quad \text{take ln on both sides}$$

then take a least-squares approximation of this linear transformation,

i. $y = 68.091 + 11.526 \ln x$
ii. $\ln y = 4.276 + 0.030x$
iii. $\ln y = 4.226 + 0.143 \ln x$
iv. $\ln \left(\frac{101-y}{y} \right) = -0.961 - 0.191x$

where $r^2 =$ (i) **0.27** (ii) **0.52** (iii) **0.66** (iv) **0.69**

whereas the power regression itself is

i. $y = \frac{101}{1+e^{-0.961-0.191x}}$
ii. $e^{\frac{y}{11.526}} = 7.5e^{\frac{68.091}{11.526}}$
iii. $y = 72.005e^{0.030x}$
iv. $y = 68.460x^{0.143}$

and, at $x = 7.5$, $\hat{y} = 68.4607.5^{0.143} \approx$
(i) **90.17** (ii) **91.31** (iii) **91.32** (iv) **92.55**

(e) *Nonlinear logistic model.*

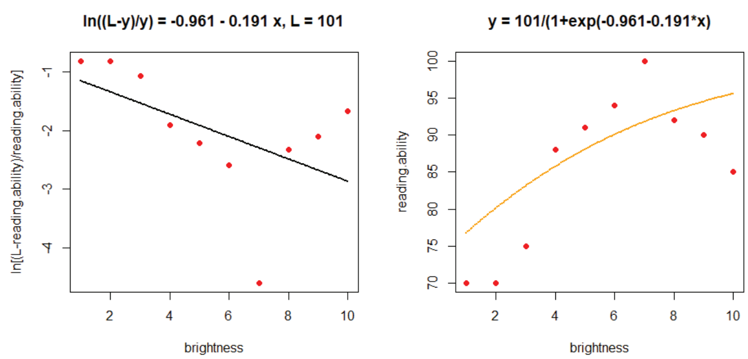

Figure 6.20: Logistic transformation

```
nonlinear.regression(brightness, reading.ability, 101, "logistic")

      transformation   trans.intercept, a          intercept, a              slope, b                    r^2
          "logistic" "-0.960644676603185" "-0.960644676603185"  "-0.19094033646998"  "0.270637267048632"
```

To fit the nonlinear logistic model where maximum $L = 101 > 100$,

$$y = \frac{L}{1 + e^{a+bx}},$$

to the data, first convert to a linear equation:

$$\begin{aligned} 1 + e^{a+bx} &= \frac{L}{y}, \\ e^{a+bx} &= \frac{L}{y} - 1 = \frac{L}{y} - \frac{y}{y} = \frac{L-y}{y}, \\ a + bx &= \ln\left(\frac{L-y}{y}\right), \end{aligned}$$

then take a least-squares approximation of this linear transformation,

i. $y = 68.091 + 11.526 \ln x$
ii. $\ln y = 4.276 + 0.030x$
iii. $\ln y = 4.226 + 0.143 \ln x$
iv. $\ln\left(\frac{101-y}{y}\right) = -0.961 - 0.191x$

where $r^2 =$ (i) **0.27** (ii) **0.52** (iii) **0.66** (iv) **0.69**

whereas the logistic regression itself is

i. $y = \frac{101}{1+e^{-0.961-0.191x}}$

ii. $e^{\frac{y}{11.526}} = 7.5e^{\frac{68.091}{11.526}}$
iii. $y = 72.005e^{0.030x}$
iv. $y = 68.460x^{0.143}$

and, at $x = 7.5$, $\hat{y} = \frac{101}{1+e^{-0.961-0.191(7.5)}} \approx$
(i) **90.17** (ii) **91.31** (iii) **91.32** (iv) **92.55**

(f) *Best nonlinear transformation.*

regression	r^2
linear	0.50
logarithmic	0.66
exponential	0.52
power	0.69
logistic	0.27

Comparing graphs and r^2, the best-fitting regression is
(i) **linear** (ii) **logarithmic** (iii) **exponential** (iv) **power** (v) **logistic**

whereas the worst-fitting regression is
(i) **linear** (ii) **logarithmic** (iii) **exponential** (iv) **power** (v) **logistic**

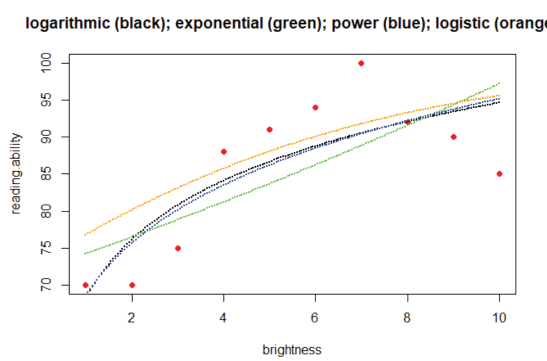

Figure 6.21: Comparing nonlinear transformations

(g) *Why do nonlinear model involve natural log and exponential functions?*
The nonlinear models given here use the natural log, "ln", or exponential, "exp", because not only do they "bend" the regression to fit the data better but also the important normal probability distribution, $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(1/2)[(x-\mu)/\sigma]^2}$ is defined with the exponential function. Consequently, it becomes easier to perform inference on the nonlinear regression which often requires normal assumptions.
(i) **True** (ii) **False**

3. *Logistic regression for binary data.*
   Reconsider the reading ability and brightness example, but, this time, subjects in a study were able to read, indicated by a "0.9", or not, indicated by a "0.1".

brightness, x	9	7	11	16	21	19	23	29	31	33
ability to read, y	0.1	0.1	0.1	0.1	0.1	0.9	0.9	0.9	0.9	0.9

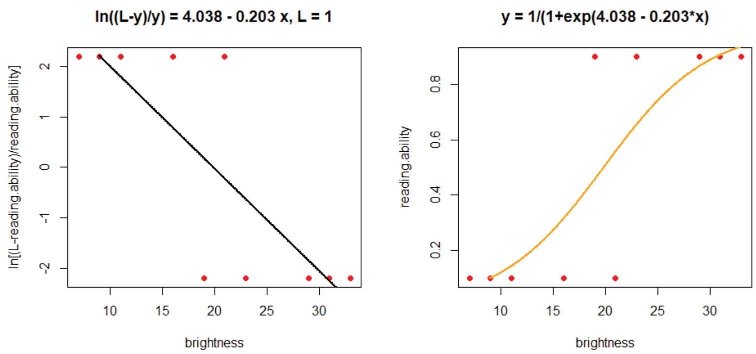

Figure 6.22: Logistic transformation for binary data

```
x <- c(9, 7, 11, 16, 21, 19, 23, 29, 31, 33)
y <- c(0.1, 0.1, 0.1, 0.1, 0.1, 0.9, 0.9, 0.9, 0.9, 0.9)
nonlinear.regression(x, y, 1, "logistic")
```

```
     transformation   trans.intercept, a        intercept, a             slope, b                  r^2
         "logistic"   "4.03753232581395"   "4.03753232581395" "-0.202891071648942"  "0.655611913122643"
```

Least-squares approximation of linear transformation of logistic model

(a) $y = 68.091 + 11.526 \ln x$

(b) $\ln y = 4.226 + 0.030x$

(c) $\ln y = 4.226 + 0.143 \ln x$

(d) $\ln\left(\frac{1-y}{y}\right) = 4.038 - 0.203x$

where $r^2 =$ (i) **0.27** (ii) **0.52** (iii) **0.66** (iv) **0.69**

whereas the logistic regression itself is

(a) $y = \frac{1}{1+e^{4.038-0.203x}}$

(b) $e^{\frac{y}{11.526}} = 7.5e^{\frac{68.091}{11.526}}$

(c) $y = 72.005e^{0.030x}$

(d) $y = 68.460x^{0.143}$

and, at $x = 12$, $\hat{y} = \frac{1}{1+e^{4.038-0.203(12)}} \approx$
(i) **0.17** (ii) **0.56** (iii) **0.78** (iv) **0.88**

and, at $x = 24$, $\hat{y} = \frac{1}{1+e^{4.038-0.203(24)}} \approx$
(i) **0.17** (ii) **0.56** (iii) **0.70** (iv) **0.88**

6.7 Multiple Regression

The multiple linear regression population model $y_i = b + mx_1 + m_2x_2 + \cdots + m_kx_k + \epsilon_i$, is estimated by sample linear regression function,

$$\hat{y} = \hat{m}_0 + \hat{m}_1x_1 + \hat{m}_2x_2 + \cdots + \hat{m}_kx_k,$$

where standard error residual, s_e, is

$$s_e = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-k-1}} = \sqrt{\frac{SS_{Res}}{n-k-1}}$$

where k is number of predictors, n is sample size, degrees of freedom is df $= n-k-1$ and where scatter is assumed linear, points are independent (sampled at random) and residuals, ϵ_i, are normal with equal variance. Overall test-statistic F for whether all slopes, m_j, $j = 1, ..., k$, of regression model $y = b + bx_1 + m_2x_2 + \cdots + m_kx_k + \epsilon_i$ are zero is

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{MS_{Reg}}{MS_{Res}},$$

where *multiple coefficient of determination* is

$$R^2 = \frac{SS_{Reg}}{SS_{Tot}} = 1 - \frac{SS_{Res}}{SS_{Tot}},$$

where *regression sum of squares* $SS_{Reg} = \sum(\hat{y} - \bar{y})^2$ and where *total sum of squares* $SS_{Tot} = SS_{Reg} + SS_{Res}$. Also, test statistic and CI for each *individual* slope, m_j, of regression model is

$$t_{n-k-1} = \frac{\hat{m}_j - m_j}{SE(\hat{m}_j)}, \qquad \hat{m}_j \pm t^*_{\frac{\alpha}{2},n-k-1} \times SE(\hat{m}_j)$$

and *adjusted* (for number of parameters) multiple coefficient of determination R^2_{adj} is

$$R^2_{adj} = 1 - (1-R^2)\frac{n-1}{n-k-1} = 1 - \frac{SS_{Res}/(n-k-1)}{SS_{Tot}/(n-1)}.$$

With regard to assumptions for inference, scatter is assumed linear, points are independent (sampled at random) and residuals, ϵ_i, are normal with equal variance. Also, critical value $F^*_{\alpha;k,n-k-1}$ is associated with given confidence level and $(k, n-k-1)$ degrees of freedom and critical value $t^*_{\frac{\alpha}{2},n-k-1}$ is associated with given confidence level and $n-k-1$ degrees of freedom.

Exercise 6.7 (Multiple Regression)

1. *Different models: reading ability, noise and brightness.*

brightness, x_1	9	7	11	16	21	19	23	29	31	33
noise, x_2	100	93	85	76	61	58	46	32	24	12
ability to read, y	40	50	64	73	86	97	104	113	123	130

```
brightness <- c(9, 7, 11, 16, 21, 19, 23, 29, 31, 33)
noise <- c(100, 93, 85, 76, 61, 58, 46, 32, 24, 12)
reading.ability <- c(40, 50, 64, 73, 86, 97, 104, 113, 123, 130)
d <- data.frame(brightness, noise, reading.ability)
```

(a) Linear regression reading ability versus brightness (alone) is

i. $\hat{y} = 23.5 + 3.24x_1$
ii. $\hat{y} = 147.4 - 1.01x_2$
iii. $\hat{y} = 164.0 - 0.44x_1 - 1.15x_2$

Reading ability increases 3.24 units per unit increase brightness.

```
lm(reading.ability ~ brightness,d)

(Intercept)   brightness
      23.53         3.24
```

Linear regression of reading ability versus noise (alone) is

i. $\hat{y} = 23.5 + 3.24x_1$
ii. $\hat{y} = 147.4 - 1.01x_2$
iii. $\hat{y} = 164.0 - 0.44x_1 - 1.15x_2$

On average, reading ability decreases 1.01 units per unit increase noise.

```
lm(reading.ability ~ noise,d)

(Intercept)        noise
    147.392       -1.012
```

Figure shows two (simple) linear regressions,
each with (i) **one** (ii) **two** (iii) **three** predictor(s).

```
par(mfrow=c(1,2))
plot(brightness,reading.ability, pch=16,col="red",xlab="Brightness, x1",ylab="Reading Ability, y")
model.reading <- lm(reading.ability~brightness); model.reading; abline(model.reading,col="black")
plot(noise, reading.ability, pch=16,col="red",xlab="Noise, x2",ylab="Reading Ability, y")
model.reading <- lm(reading.ability~noise); model.reading; abline(model.reading,col="black")
par(mfrow=c(1,1))
```

(b) The *multiple linear regression* is given by,

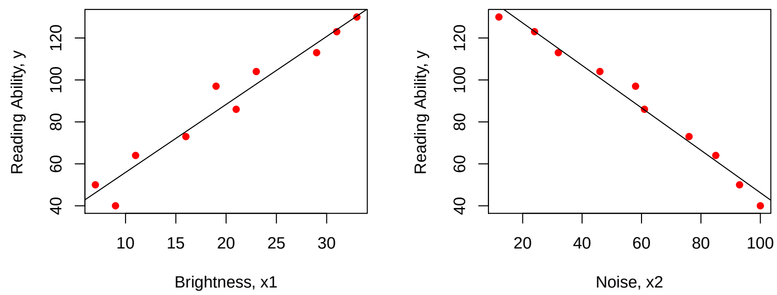

Figure 6.23: Scatter plots and two simple linear regressions

i. $\hat{y} = 23.5 + 3.24x_1$
ii. $\hat{y} = 147.4 - 1.01x_2$
iii. $\hat{y} = 164.0 - 0.44x_1 - 1.15x_2$

The y–intercept of this line, b, is (i) **164.0** (ii) $\mathbf{-0.44}$ (iii) $\mathbf{-1.15}$.
The *slope* in the x_1 direction, $\hat{m}_1$, is (i) **164.0** (ii) $\mathbf{-0.44}$ (iii) $\mathbf{-1.15}$.
The *slope* in the x_2 direction, $\hat{m}_2$, is (i) **164.0** (ii) $\mathbf{-0.44}$ (iii) $\mathbf{-1.15}$.

```
lm(reading~brightness + noise)

Coefficients:
(Intercept)   brightness        noise
   164.0466      -0.4416      -1.1458
```

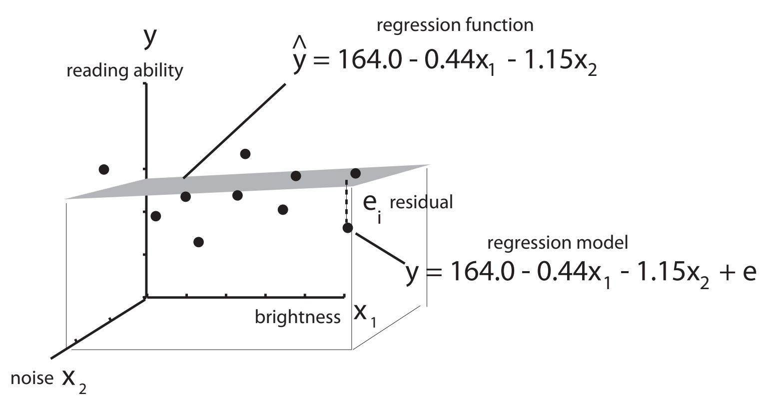

Figure 6.24: Scatter plot and multiple regression

Multiple regression has (i) **one** (ii) **two** (iii) **three** predictors.
The multiple regression is (i) **linear** (ii) **quadratic** in the x_i.
There are (i) **10** (ii) **20** (iii) **30** data points.
One data point is $(x_1, x_2, \hat{y}) =$ (i) $\mathbf{(19, 58)}$ (ii) $\mathbf{(19, 58, 97)}$ (iii) $\mathbf{(58, 97)}$.
Data point $(x_1, x_2, y) = (19, 58, 97)$ means

i. for brightness 19, the reading ability is 97.
ii. for noise level 58, the reading ability is 97.
iii. for brightness 19 and a noise level 58, the reading ability is 97.

(c) Coefficient estimate $\hat{m}_1 = -0.44$ means, on average, reading ability decreases 0.44 units per unit increase brightness, *after accounting for noise level*. This is the (i) **same** (ii) **different** from simple linear case where $\hat{m}_1 = 3.24$. At any given noise level, the reading ability (i) **worsens** (ii) **improves** per unit increase in brightness: accounting for noise converts a previously positive association into a negative association between reading ability and brightness.

(d) Coefficient estimate $\hat{m}_2 = -1.15$ means, on average, reading ability decreases 1.15 units per unit increase noise, *after accounting for brightness*. This is (i) **the same** (ii) **different** from simple linear case where $\hat{m}_2 = -1.01$.

(e) The predicted value of the reading ability at $(x_1, x_2) = (19, 58)$, is
$\hat{y} = 164.0 - 0.44(19) - 1.15(58) \approx$ (i) **83.52** (ii) **84.79** (iii) **88.94**.
Draw a *vertical* line which passes through (19,58) on the "(x_1, x_2)" plane. Now draw an *horizontal* line which passes through the point where the solid regression plane and the previously drawn vertical line intersect. This horizontal line will intersect the "reading ability" axis at 88.94.

(f) At level $(x_1, x_2) = (19, 58)$, $\hat{y} = 88.94$. The difference between this value and the *observed* value, $y = 97$ (look at the table of the data above) is called the *residual* (residual) and is given by
$e_i = y_i - \hat{y}_i = 97 - 88.94 =$ (i) **6.1** (ii) **7.2** (iii) **8.3**.

(g) If we were to draw the residual (residual) for $(x_1, x_2, y) = (19, 58, 97)$ on the scatter plot, we would

i. draw line parallel to the regression plane.
ii. draw a line connecting the point (19, 58) to the point (58, 97).
iii. draw a line connecting observed point (19, 58, 97) to expected point (19, 58, 88.94) on the regression plane.

(h) There are (i) **1** (ii) **5** (ii) **10** residuals.

(i) Predicted value of reading ability at $(x_1, x_2) = (2, 3)$, is
$\hat{y} = 164.0 - 0.44(2) - 1.15(3) \approx$ (i) **134.52** (ii) **159.67** (iii) **167.94**.
In this case, since $(x_1, x_2) = (2, 3)$ is outside the range of data, the predicted value, $\hat{y} \approx 159.67$, is most likely a (i) **poor** (ii) **good** estimate of reading ability.

(j) In this case, we *assume* the effect of x_1 on $\hat{y}$ does not depend on x_2. This is also true of x_2. In other words, x_1 and x_2 do not interact with one another. The model is said to be (i) **additive** (ii) **interactive**.

(k) If we sampled at random another ten individuals, we would get (i) **the same** (ii) **different** scatter plot of points. The data is a example of a (i) **sample** (ii) **population**.

2. *Choosing the best model: reading ability, noise and brightness.*

brightness, x_1	9	7	11	16	21	19	23	29	31	33
noise, x_2	100	93	85	76	61	58	46	32	24	12
ability to read, y	40	50	64	73	86	97	104	113	123	130

```
brightness <- c(9, 7, 11, 16, 21, 19, 23, 29, 31, 33)
noise <- c(100, 93, 85, 76, 61, 58, 46, 32, 24, 12)
reading.ability <- c(40, 50, 64, 73, 86, 97, 104, 113, 123, 130)
d <- data.frame(brightness, noise, reading.ability)
```

(a) Identify all possible models for this data from the following.

i. $\hat{y} = \bar{y} = 88$

ii. $\hat{y} = 23.5 + 3.24x_1$

iii. $\hat{y} = 147.4 - 1.01x_2$

iv. $\hat{y} = 164.0 - 0.44x_1 - 1.15x_2$

```
lm(reading.ability ~ 1,d)
lm(reading.ability ~ brightness,d)
lm(reading.ability ~ noise,d)
lm(reading.ability ~ brightness + noise,d)

lm(formula = reading.ability ~ 1, data = d)

Coefficients:
(Intercept)
         88
lm(formula = reading.ability ~ brightness, data = d)

Coefficients:
(Intercept)   brightness
      23.53         3.24
lm(formula = reading.ability ~ noise, data = d)

Coefficients:
(Intercept)        noise
    147.392       -1.012
lm(formula = reading.ability ~ brightness + noise, data = d)

Coefficients:
(Intercept)   brightness        noise
   164.0466      -0.4416      -1.1458
```

(b) *Assess fit of model 1: reading ability regressed on intercept,* $\hat{y} = b = \bar{y} = 88$.

A. Is intercept $b = \bar{y} = 88$ *significant?*
Is $b = \bar{y} = 88$ a better predictor of reading ability than $b = 0$?
Statement.

i. $H_0 : b = 0$ versus $H_1 : b > 0$
ii. $H_0 : b = 0$ versus $H_1 : b < 0$
iii. $H_0 : b = 0$ versus $H_1 : b \neq 0$

Test. Chance $|t = 9.053|$ or more, *if* $b = 0$, is
p–value $= 2 \cdot P(t \geq 9.053) \approx$ (i) **0.00** (ii) **0.01** (iii) **0.11**
level of significance $\alpha =$ (i) **0.01** (ii) **0.05** (iii) **0.10**.

Conclusion. Since p–value $= 0.00 < \alpha = 0.05$,
(i) **do not reject** (ii) **reject** null $H_0 : b = 0$.
data indicates intercept, $b = \bar{y} = 88$
(i) **smaller than** (ii) **equals** (iii) **does not equal** zero (0)
so, yes, $b = \bar{y} = 88$ is significant; that is, it is a better predictor than $b = 0$ of reading ability.

B. Is residual standard error, s_e, small?
If s_e is small, the data is close to the model $\hat{y} = b = \bar{y} = 88$.

$s_e =$ (i) **10.74** (ii) **20.74** (iii) **30.74**

which is may or may not be "large" (since there is nothing to compare this number against) but it turns out to be large and so the data is (i) **close to** (ii) **far away from** the model $\hat{y} = \bar{y} = 88$, so this measure indicates the model does not fit the data very well.

```
lm(reading.ability ~ 1,d) # one possible model

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   88.000      9.721   9.053 8.14e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.74 on 9 degrees of freedom
```

(c) *Model 2: reading ability regressed on brightness only,* $\hat{y} = 23.5 + 3.24x_1$.

A. Is intercept $b = \bar{y} = 23.5$ significant?
Since p–value $= 0.004 < \alpha = 0.05$,
(i) **do not reject** (ii) **reject** null $H_0 : b = 0$.
data indicates intercept, $b = \bar{y} = 23.5$
(i) **smaller than** (ii) **equals** (iii) **does not equal** zero (0)
so, yes, $b = \bar{y} = 23.5$ is significant

B. Is slope $m_1 = 3.24$ significant?
Since p–value $= 0.000 < \alpha = 0.05$,
(i) **do not reject** (ii) **reject** null $H_0 : m_1 = 0$.
data indicates slope $m_1 = 3.24$
(i) **smaller than** (ii) **equals** (iii) **does not equal** zero (0)
so, yes, $m_1 = 3.24$ is significant
in fact, "more" significant than intercept b because of smaller p-value.

C. Is residual standard error, s_e, small?
$s_e =$ (i) **10.74** (ii) **7.37** (iii) **30.74**

which is smaller than s_e for model 1, so data is
(i) **closer to** (ii) **farther away from** model 1 than model 2.

D. Are R^2 and R^2_{adj} large?
If both are large, large proportion of data variation described by model.
$R^2 =$ (i) **0.94** (ii) **0.95** (iii) **0.96**
$R^2_{adj} =$ (i) **0.94** (ii) **0.95** (iii) **0.96**
which are both large, so
(i) **large** (ii) **small** proportion of variation described by model 2.

```
lm(reading.ability ~ brightness,d)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  23.5301     5.7758   4.074  0.00356 **
brightness    3.2397     0.2656  12.198 1.89e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.365 on 8 degrees of freedom
Multiple R-squared:  0.949,       Adjusted R-squared:  0.9426
F-statistic: 148.8 on 1 and 8 DF,  p-value: 1.893e-06
```

(d) *Model 3: reading ability regressed on noise only, $\hat{y} = 147.4 - 1.01x_2$.*

A. Is intercept $b = \bar{y} = 147.39$ significant?
Since p–value $= 0.00 < \alpha = 0.05$,
(i) **do not reject** (ii) **reject** null $H_0 : b = 0$.
data indicates intercept, $b = \bar{y} = 147.39$
(i) **smaller than** (ii) **equals** (iii) **does not equal** zero (0)
so, yes, $b = \bar{y} = 147.39$ is significant

B. Is slope $m_2 = -1.01$ significant?
Since p–value $= 0.00 < \alpha = 0.05$,
(i) **do not reject** (ii) **reject** null $H_0 : m_2 = 0$.
data indicates slope $m_2 = -1.01$
(i) **smaller than** (ii) **equals** (iii) **does not equal** zero (0)
so, yes, $m_2 = -1.01$ is significant
but "less" significant than intercept b because of larger p-value.

C. Is residual standard error, s_e, small?
$s_e =$ (i) **4.65** (ii) **7.37** (iii) **30.74**
which is smaller than s_e for model 2, so data is
(i) **closer to** (ii) **farther away from** model 3 than model 2.

D. Are R^2 and R^2_{adj} large?
R^2 is *always* larger than R^2_{adj} because latter (more fairly) adjusts smaller for more parameters
$R^2 =$ (i) **0.94** (ii) **0.97** (iii) **0.98**
$R^2_{adj} =$ (i) **0.94** (ii) **0.97** (iii) **0.98**
which are both large, so
(i) **large** (ii) **small** proportion of variation described by model 3.

```
summary(lm(reading.ability ~ noise,d))

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 147.39173    3.36402   43.81 8.12e-11 ***
noise        -1.01178    0.05154  -19.63 4.72e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.65 on 8 degrees of freedom
Multiple R-squared:  0.9797,       Adjusted R-squared:  0.9771
F-statistic: 385.3 on 1 and 8 DF,  p-value: 4.719e-08
```

(e) *Model 4: both brightness and noise,* $\hat{y} = 164.0 - 0.44x_1 - 1.15x_2$.

A. Is intercept $b = \bar{y} = 164.05$ *significant?*
Since p–value $= 0.006 < \alpha = 0.05$,
(i) **do not reject** (ii) **reject** null $H_0 : b = 0$.
data indicates intercept, $b = \bar{y} = 164.05$
(i) **smaller than** (ii) **equals** (iii) **does not equal** zero (0)
so, yes, $b = \bar{y} = 164.05$ is significant

B. Is slope $m_1 = -0.44$ *significant?*
Since p–value $= 0.71 > \alpha = 0.05$,
(i) **do not reject** (ii) **reject** null $H_0 : m_1 = 0$.
data indicates slope $m_1 = -0.44$
(i) **smaller than** (ii) **equals** (iii) **does not equal** zero (0)
so $m_1 = -0.44$ is *not* significant
which is strange because it was, possible interaction with m_2?

C. Is slope $m_2 = -1.15$ *significant?*
Since p–value $= 0.01 < \alpha = 0.05$,
(i) **do not reject** (ii) **reject** null $H_0 : m_2 = 0$.
data indicates slope $m_2 = -1.15$
(i) **smaller than** (ii) **equals** (iii) **does not equal** zero (0)
so, yes, $m_2 = -1.15$ is significant
but "less" significant than intercept b because of larger p-value.

D. Is residual standard error, s_e*, small?*
$s_e =$ (i) **4.92** (ii) **7.37** (iii) **30.74**
which is smaller than s_e for model 3, so data is
(i) **closer to** (ii) **farther away from** model 4 than model 3.

E. Are R^2 and R^2_{adj} large?
$R^2 =$ (i) **0.94** (ii) **0.97** (iii) **0.98**
$R^2_{adj} =$ (i) **0.94** (ii) **0.97** (iii) **0.98**
which are both large, so
(i) **large** (ii) **small** proportion of variation described by model 3.

F. Is F large?
If F is large, then at least one slope is *not* zero.
Statement.

i. $H_0 : m_1 = m_2 = 0$ versus $H_1 : m < 0, m_2 > 0$
ii. $H_0 : m_1 = m_2 = 0$ versus H_1 : at least one $m_i \neq 0$, $i = 1, 2$
iii. $H_0 : m_1 = m_2 \neq 0$ versus $H_1 : m_1 = m_2 = 0$

Test. Chance $F = 172.4$ or more, if $m_1 = m_2 = 0$, is
p–value $= P(F \geq 172.4) \approx$ (i) **0.00** (ii) **0.01** (iii) **0.11**
level of significance $\alpha =$ (i) **0.01** (ii) **0.05** (iii) **0.10**.

Conclusion. Since p–value $= 0.00 < \alpha = 0.05$,
(i) **do not reject** (ii) **reject** null $H_0 : m_1 = m_2 = 0$.
data indicates
(i) **all slopes zero** (ii) **at least one slope not zero**
so, yes, F is large; model 4 is a good "overall" fit of data.

```
summary(lm(reading.ability ~ brightness + noise,d))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 164.0466    42.6464   3.847  0.00632 **
brightness   -0.4416     1.1267  -0.392  0.70679
noise        -1.1458     0.3463  -3.308  0.01297 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.917 on 7 degrees of freedom
Multiple R-squared:  0.9801,       Adjusted R-squared:  0.9744
F-statistic: 172.4 on 2 and 7 DF,  p-value: 1.112e-06
```

(f) *Summary of models.*

Variables	R^2	R^2_{adj}	F p-value	s_e
1. intercept	na	na	na	30.74
2. brightness	0.949	0.943	0.00	7.37
3. noise	0.980	0.977	0.00	4.65
4. brightness, noise	0.980	0.974	0.00	4.92

The model which best fits the model is (i) **1** (ii) **2** (iii) **3** (iv) **4**
although all except the intercept model are very good fitting models.

3. *Check model 4 assumptions using residuals.*

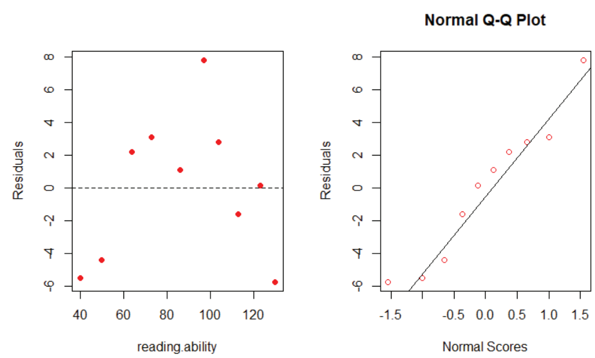

Figure 6.25: Check assumptions using residuals

```
residuals <- resid(lm(reading.ability ~ brightness + noise,d)); residuals
par(mfrow=c(1,2))
plot(reading.ability, residuals, pch=16, col="red", xlab="reading.ability", ylab="Residuals")
abline(h=0,lty=2,col="black")
qqnorm(residuals, col="red", ylab="Residuals", xlab="Normal Scores")
qqline(residuals) # Q-Q (normal probability plot) of residuals check for normality
par(mfrow=c(1,1))
```

(a) *Linearity assumption/condition?*
According to either scatter diagram or residual plot,
there (i) **is a** (ii) **is no** pattern (around line): points are curved.

(b) *Independence assumption?*
Subjects act (i) **independently** (ii) **dependently** of one another.

(c) *Constant (equal) variance condition?*
According to residual plot, residuals vary -6 and 8 over entire range of brightness; that is, data variance is (i) **constant** (ii) **variable**.

(d) *Nearly normal condition?*
Normal probability plot indicates residuals
(i) **normal** (ii) **not normal** because plot more or less straight.

4. *Nonlinear Model 5: brightness² predictor added to brightness predictor.*
Fill in missing values.

brightness, x_1	9	7	11	16	21	19	23	29	31	33
brightness², x_1^2	81	49	121	256	441	361	529	841	____	____
ability to read, y	40	50	64	73	86	97	104	113	123	130

Model 5 is

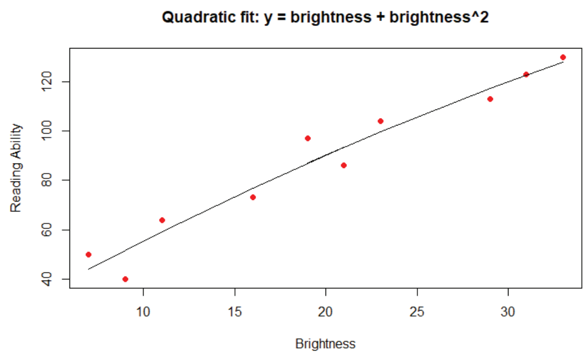

Figure 6.26: Model 5: reading ability = brightness + brightness^2

(a) $\hat{\boldsymbol{y}} = \mathbf{72.20} + \mathbf{2.4}\boldsymbol{x_1}$

(b) $\hat{\boldsymbol{y}} = \mathbf{15.299} + \mathbf{4.257}\boldsymbol{x_1} - \mathbf{0.025}\boldsymbol{x_1^2}$

(c) $\hat{\boldsymbol{y}} = \mathbf{79.10} + \mathbf{2.42}\boldsymbol{x_1} - \mathbf{0.84}(\boldsymbol{x_2} - \bar{\boldsymbol{x}}_2)^2$

Compare model 5 with other models, by filling in the blanks:

Variables	R^2	R^2_{adj}	F p-value	s_e
1. intercept	na	na	na	30.74
2. brightness	0.949	0.943	0.00	7.37
3. noise	0.980	0.977	0.00	4.65
4. brightness, noise	0.980	0.974	0.00	4.92
5. brightness, brightness^2	______	______	______	______

Model 5 (i) **is** (ii) **is not** as good as other models.
Brightness, brightness^2 (i) **dependent on** (ii) **independent of** one another which is fine if predicting reading ability but problematic if interpreting model, trying to figure out "how much" brightness relative to brightness^2 influence reading ability

```
brightness2 <- brightness^2; brightness2 # quadratic predictor
model.reading2 <- lm(reading.ability~brightness + brightness2); summary(model.reading2) # quadratic fit
plot(brightness, reading.ability, pch=16, col="red", xlab="Brightness",
     ylab="Reading Ability", main="Quadratic fit: y = brightness + brightness^2")
x <- brightness; y <- predict(model.reading2,list(brightness=x)); lines(x,y,col="black") # quadratic plot
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.29871   13.64003   1.122   0.2990
brightness   4.25682    1.53935   2.765   0.0279 *
brightness2 -0.02540    0.03781  -0.672   0.5234
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 7.631 on 7 degrees of freedom
Multiple R-squared:  0.9521,        Adjusted R-squared:  0.9384
F-statistic: 69.51 on 2 and 7 DF,  p-value: 2.412e-05
```

5. *Sum of squares and ANOVA*

brightness, x_1	9	7	11	16	21	19	23	29	31	33
noise, x_2	100	93	85	76	61	58	46	32	24	12
ability to read, y	40	50	64	73	86	97	104	113	123	130

```
anova(lm(reading.ability~brightness+noise)) # sum of squares
summary(lm(reading.ability~brightness+noise)) # summary of fit statistics
```

```
Analysis of Variance Table

Response: reading.ability
           Df Sum Sq Mean Sq F value    Pr(>F)
brightness  1 8070.1  8070.1 333.757 3.645e-07 ***
noise       1  264.7   264.7  10.946   0.01297 *
Residuals   7  169.3    24.2
#################################################
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 164.0466    42.6464   3.847  0.00632 **
brightness   -0.4416     1.1267  -0.392  0.70679
noise        -1.1458     0.3463  -3.308  0.01297 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.917 on 7 degrees of freedom
Multiple R-squared:  0.9801,        Adjusted R-squared:  0.9744
F-statistic: 172.4 on 2 and 7 DF,  p-value: 1.112e-06
```

(a) $SS_{Res} =$ (i) **169.3** (ii) **264.7** (iii) **8070.1**

(b) $SS_{Reg} = 8070.1 + 264.7 =$ (i) **169.3** (ii) **264.7** (iii) **8334.8**

(c) $SS_{Tot} = SS_{Res} + SS_{Reg} =$ (i) **169.3** (ii) **8334.8** (iii) **8504.1**

(d) $MS_{Res} = \frac{SS_{Res}}{n-k-1} = \frac{169.2}{10-2-1} =$ (i) **24.2** (ii) **264.7**

(e) $MS_{Reg} = \frac{SS_{Reg}}{k} = \frac{8334.8}{2} =$ (i) **24.2** (ii) **4167.4** (iii) **8070.1**

(f) $\frac{SS_{Reg}}{SS_{Tot}} = \frac{8334.8}{8504.1} =$ (i) $\boldsymbol{R^2}$ (ii) $\boldsymbol{R^2_{adj}}$ (iii) $\boldsymbol{s_e}$ (with some round-off)

(g) $1 - \frac{SS_{Res}/(n-k-1)}{SS_{Tot}/(n-1)} = 1 - \frac{169.3/(10-2-1)}{8504.1/(10-1)} =$ (i) $\boldsymbol{R^2}$ (ii) $\boldsymbol{R^2_{adj}}$ (iii) $\boldsymbol{s_e}$

(h) $\frac{MS_{Reg}}{MS_{Res}} = \frac{4167.4}{24.2} =$ (i) $\boldsymbol{R^2}$ (ii) $\boldsymbol{t}$-statistic (iii) $\boldsymbol{F}$-statistic

(i) $\sqrt{\frac{SS_{Res}}{n-k-1}} = \sqrt{\frac{169.3}{10-2-1}} =$ (i) $\boldsymbol{R^2}$ (ii) $\boldsymbol{R^2_{adj}}$ (iii) $\boldsymbol{s_e}$

6. *Matrix approach to simple linear regression: reading ability vs brightness.*

illumination, x	1	2	3	4	5	6	7	8	9	10
ability to read, y	70	70	75	88	91	94	100	92	90	85

Use the matrix approach to find the linear regression equation; since

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \\ 7 & 1 \\ 8 & 1 \\ 9 & 1 \\ 10 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 70 \\ 70 \\ 75 \\ 88 \\ 91 \\ 94 \\ 100 \\ 92 \\ 90 \\ 85 \end{bmatrix}, \quad \text{so} \quad \mathbf{m} = \left(\mathbf{A}^{\mathbf{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{b} =$$

$$\text{(i)} \begin{bmatrix} 3.878 \\ 25.8 \end{bmatrix}, \quad \text{(ii)} \begin{bmatrix} 2.418 \\ 72.2 \end{bmatrix}, \quad \text{(iii)} \begin{bmatrix} 72.2 \\ 2.418 \end{bmatrix}, \quad \text{(iv)} \begin{bmatrix} 3.878 \\ 25.8 \end{bmatrix},$$

then the simple linear regression is

$\hat{y} = 3.878x + 25.8$

$\hat{y} = 25.8x + 3.878$

$\hat{y} = 72.2 + 2.418x$

$\hat{y} = -3.878x - 25.8$

```
brightness <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
intercept = c(rep(1,10))
reading.ability <- c(70, 70, 75, 88, 91, 94, 100, 92, 90, 85)
A <- as.matrix (cbind(brightness,intercept))
b <- reading.ability
solve (t(A)%*%A)%*%t(A)%*%b
```

```
                [,1]
brightness  2.418182
intercept  72.200000
```