STAC67: Regression Analysis

Lecture 17

Thibault Randrianarisoa

November 11, 2026

University of Toronto Scarborough

Announcements: the final exam

The date of the final exam is now set:

Thursday 17 December, 9:00 to 12:00

Room IA 1150

It covers the material of the whole term.

Announcements: the project

The checkpoint is due Friday 13 November, 23:59. There is no submission page on Quercus or Crowdmark: send it by email.

- One email per group, to `t.randrianarisoa@utoronto.ca`, with `[STAC67] Checkpoint` and your group name in the subject (for instance `[STAC67] Checkpoint, Diamonds 2`).
- Attach the short report (PDF or HTML, at most 3 pages, code not shown) and your R Markdown file.
- The report contains your question, a descriptive analysis targeted at it, and a first multiple linear regression with its residual analysis.

Review of Week 8

- Extra sums of squares: to test q coefficients at once, compare a full and a reduced model, $F^* = \big[(SSE(R) - SSE(F))/q\big]/MSE(F)$.
- Multicollinearity: correlated predictors inflate the variances of the estimated coefficients and make them unstable, but without violating any assumption.
- A qualitative predictor with k levels enters through $k-1$ indicators, where each coefficient compares a level with the baseline.
- A quantitative predictor and an indicator: parallel lines, compared at the same X. Adjusting can change an effect, or even reverse it (Simpson's paradox).
- An interaction $D \cdot X$ lets the slopes differ too. Same slope? t test. Same line? F test. But always keep the main effects.

Exercise (from Lecture 16)

For the insurance firms, `lm(months ~ size * stock)` gives, with X the size of the firm and $D = 1$ for stock firms:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	33.8384	2.4406	13.8645	0.0000
size	-0.1015	0.0131	-7.7786	0.0000
stock	8.1313	3.6541	2.2253	0.0408
size:stock	-0.0004	0.0183	-0.0228	0.9821

1. Write down the fitted line of each type of firm.
2. Estimate the difference in mean months between a stock and a mutual firm of size 200.
3. Which model would you report, this one or the parallel-lines model, and why?

$D=0$: $\hat{y} = 33.84 - 0.10\,X$

$D=1$: $\hat{y} = 41.96 - 0.10\,X$

Exercise: solution

1. Mutual ($D = 0$): $\hat{Y} = 33.84 - 0.1015\, X$. Stock ($D = 1$): $\hat{Y} = 41.97 - 0.1019\, X$.

Exercise: solution

1. Mutual ($D = 0$): $\hat{Y} = 33.84 - 0.1015\, X$. Stock ($D = 1$): $\hat{Y} = 41.97 - 0.1019\, X$.

2. $\hat{\beta}_2 + \hat{\beta}_3 \times 200 = 8.13 - 0.0004 \times 200 = 8.05$ months.

Exercise: solution

1. Mutual ($D = 0$): $\hat{Y} = 33.84 - 0.1015\,X$. Stock ($D = 1$): $\hat{Y} = 41.97 - 0.1019\,X$.

2. $\hat{\beta}_2 + \hat{\beta}_3 \times 200 = 8.13 - 0.0004 \times 200 = 8.05$ months.

3. The parallel-lines model. In the table, the test of $H_0 : \beta_3 = 0$ (same slope) gives $t = -0.02$, $p = 0.98$: no evidence that the slopes differ, and they are almost equal (-0.1015 and -0.1019). The simpler model is easier to interpret: one slope, and one difference between the types, 8.1 months ($t = 5.52$), that holds at every size.

Example: SENIC dataset

113 hospitals in the United States, 1975-76. Response: `risk`, the estimated probability (in %) of acquiring an infection in the hospital. Predictors: the average age of the patients, the culturing ratio, the daily census, the facilities, the affiliation with a medical school, and the `region`: NE, NC, S or W (28, 32, 37, 16 hospitals).

Example: SENIC dataset

113 hospitals in the United States, 1975-76. Response: `risk`, the estimated probability (in %) of acquiring an infection in the hospital. Predictors: the average age of the patients, the culturing ratio, the daily census, the facilities, the affiliation with a medical school, and the `region`: NE, NC, S or W (28, 32, 37, 16 hospitals).

Full model: a separate regression in each region: the 5 predictors, 3 indicators and their 5×3 products, $p = 24$. Reduced model: one regression for all, $p = 6$.

```
red  <- lm(risk ~ age + culture + census + facilities +
             medschool, data = senic)
full <- lm(risk ~ (age + culture + census + facilities +
             medschool) * region, data = senic)
anova(red, full)
```

Example: SENIC dataset

113 hospitals in the United States, 1975-76. Response: `risk`, the estimated probability (in %) of acquiring an infection in the hospital. Predictors: the average age of the patients, the culturing ratio, the daily census, the facilities, the affiliation with a medical school, and the `region`: NE, NC, S or W (28, 32, 37, 16 hospitals).

Full model: a separate regression in each region: the 5 predictors, 3 indicators and their 5×3 products, $p = 24$. Reduced model: one regression for all, $p = 6$.

```
red  <- lm(risk ~ age + culture + census + facilities +
             medschool, data = senic)
full <- lm(risk ~ (age + culture + census + facilities +
             medschool) * region, data = senic)
anova(red, full)
```

$F^* = 1.20$ on 18 and 89 df, $p = 0.28$: no evidence that the regression functions differ across regions.

Polynomial regression

Polynomial regression models

or even $X_1^2 X_2$, $X_1^3 X_2^4$, ...

A polynomial regression model contains powers of a predictor, such as X^2 and X^3, and possibly products of predictors, such as $X_1 X_2$. Two uses in practice:

1. the true mean response truly is a polynomial;
2. the true mean response is unknown or complicated, and a polynomial approximates it well over the range of the data. This is the more common use.

Polynomial regression models

A polynomial regression model contains powers of a predictor, such as X^2 and X^3, and possibly products of predictors, such as $X_1 X_2$. Two uses in practice:

1. the true mean response truly is a polynomial;
2. the true mean response is unknown or complicated, and a polynomial approximates it well over the range of the data. This is the more common use.

It is still a linear model: linear in the β's, with X^2 for instance as one more column of $\boldsymbol{X}$. Least squares, the t and F tests and the intervals all apply unchanged.

Polynomial regression models

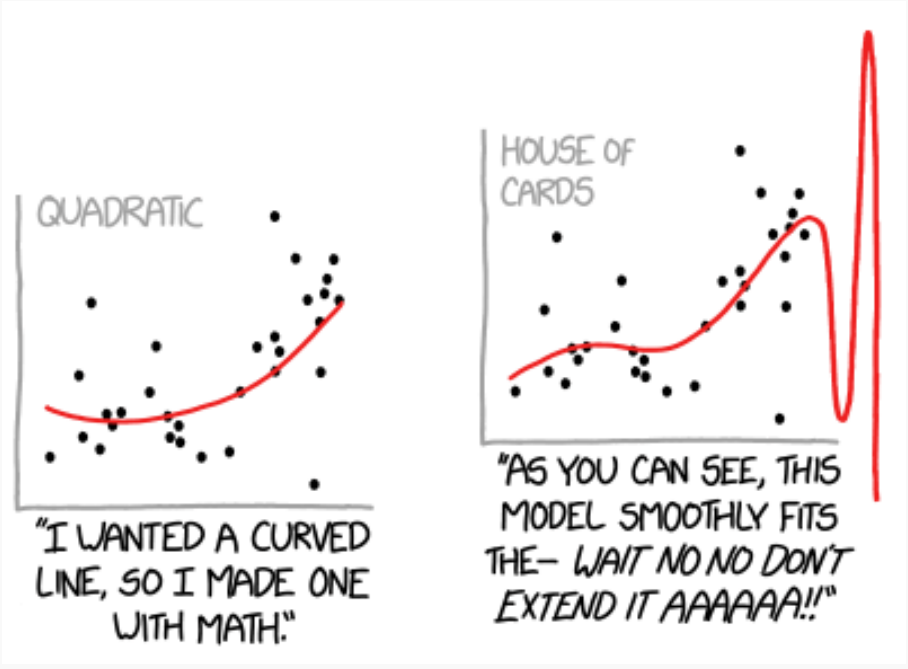
A polynomial regression model contains powers of a predictor, such as X^2 and X^3, and possibly products of predictors, such as $X_1 X_2$. Two uses in practice:

1. the true mean response truly is a polynomial;
2. the true mean response is unknown or complicated, and a polynomial approximates it well over the range of the data. This is the more common use.

It is still a linear model: linear in the β's, with X^2 for instance as one more column of $\boldsymbol{X}$. Least squares, the t and F tests and the intervals all apply unchanged.

⚠ A polynomial may fit the data at hand well, and turn in an unexpected direction just outside them: do not extrapolate.

How far can a polynomial be pushed?



Left: a curve where the data bend. Right: a high-order polynomial, close to the data inside their range and wild just outside it.

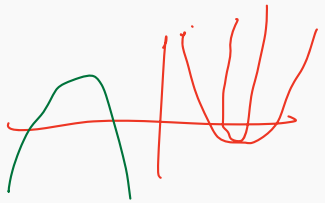
xkcd.com/2048 (CC BY-NC 2.5), two of its twelve panels

One predictor: the second-order model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \epsilon_i, \qquad x_i = X_i - \bar{X}.$$

The mean response is a parabola. The index 11 says that $x^2 = x \cdot x$. With two predictors we would also meet β_{22} and β_{12}.

One predictor: the second-order model



$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \epsilon_i, \qquad x_i = X_i - \bar{X}.$$

The mean response is a parabola. The index 11 says that $x^2 = x \cdot x$. With two predictors we would also meet β_{22} and β_{12}.

- β_{11}: gives the curvature, the parabola opens down if $\beta_{11} < 0$, up if $\beta_{11} > 0$.
- β_1: the slope at $x = 0$, that is at $X = \bar{X}$, since the derivative of $\beta_0 + \beta_1 x + \beta_{11} x^2$ is $\beta_1 + 2\beta_{11} x$.

One predictor: the second-order model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \epsilon_i, \qquad x_i = X_i - \bar{X}.$$

The mean response is a parabola. The index 11 says that $x^2 = x \cdot x$. With two predictors we would also meet β_{22} and β_{12}.

- β_{11}: gives the curvature, the parabola opens down if $\beta_{11} < 0$, up if $\beta_{11} > 0$.
- β_1: the slope at $x = 0$, that is at $X = \bar{X}$, since the derivative of $\beta_0 + \beta_1 x + \beta_{11} x^2$ is $\beta_1 + 2\beta_{11} x$.

Higher orders (x^3, ...) are rarely useful beyond the third: hard to interpret, and erratic between and beyond the data, as in the comic. Keep the lower powers of any power in the model.

Why centre?

A quick computation with $X = 1, 2, 3$, so that $\bar{X} = 2$ and $x = X - 2$:

X	1	2	3	
X^2	1	4	9	$r(X, X^2) = 0.99$
x	-1	0	1	
x^2	1	0	1	$r(x, x^2) = 0$

X^2 grows with X: the columns X and X^2 of $\boldsymbol{X}$ are almost collinear. After centring, x^2 is the same at $x = -1$ and $x = 1$: $\sum_i x_i(x_i^2 - \overline{x^2}) = -\frac{1}{3} + 0 + \frac{1}{3} = 0$.

Why centre?

A quick computation with $X = 1, 2, 3$, so that $\bar{X} = 2$ and $x = X - 2$:

X	1	2	3	
X^2	1	4	9	$r(X, X^2) = 0.99$
x	-1	0	1	
x^2	1	0	1	$r(x, x^2) = 0$

X^2 grows with X: the columns X and X^2 of $\boldsymbol{X}$ are almost collinear. After centring, x^2 is the same at $x = -1$ and $x = 1$: $\sum_i x_i(x_i^2 - \overline{x^2}) = -\frac{1}{3} + 0 + \frac{1}{3} = 0$.

It is the same model. Replace x by $X - \bar{X}$ and expand:

$$\beta_0 + \beta_1 x + \beta_{11} x^2 = (\beta_0 - \beta_1 \bar{X} + \beta_{11} \bar{X}^2) + (\beta_1 - 2\beta_{11} \bar{X})\, X + \beta_{11} X^2.$$

Every parabola in x is a parabola in X, and conversely, so least squares picks the same one: the same fitted values, residuals, SSE and R^2, and the same b_{11}.

Why centre?

A quick computation with $X = 1, 2, 3$, so that $\bar{X} = 2$ and $x = X - 2$:

X	1	2	3	
X^2	1	4	9	$r(X, X^2) = 0.99$
x	-1	0	1	
x^2	1	0	1	$r(x, x^2) = 0$

X^2 grows with X: the columns X and X^2 of $\boldsymbol{X}$ are almost collinear. After centring, x^2 is the same at $x = -1$ and $x = 1$: $\sum_i x_i(x_i^2 - \overline{x^2}) = -\frac{1}{3} + 0 + \frac{1}{3} = 0$.

It is the same model. Replace x by $X - \bar{X}$ and expand:

$$\beta_0 + \beta_1 x + \beta_{11} x^2 = (\beta_0 - \beta_1 \bar{X} + \beta_{11} \bar{X}^2) + (\beta_1 - 2\beta_{11} \bar{X})\, X + \beta_{11} X^2.$$

Every parabola in x is a parabola in X, and conversely, so least squares picks the same one: the same fitted values, residuals, SSE and R^2, and the same b_{11}.

Only b_0 and b_1 change meaning: b_1 is now the slope at $\bar{X}$, in the middle of the data, rather than at $X = 0$, which may be far outside them.

Two predictors: the second-order model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \epsilon_i.$$

The mean response is a curved surface, and $\beta_{12} x_1 x_2$ is an interaction between the two predictors.

Two predictors: the second-order model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \epsilon_i.$$

The mean response is a curved surface, and $\beta_{12} x_1 x_2$ is an interaction between the two predictors.

It is a multiple regression with $p = 6$ and

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}x_{12} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}^2 & x_{n2}^2 & x_{n1}x_{n2} \end{pmatrix}, \qquad \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}.$$

With three predictors: 3 linear terms, 3 squares and 3 products, so $p = 10$.

Two predictors: the second-order model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \epsilon_i.$$

The mean response is a curved surface, and $\beta_{12} x_1 x_2$ is an interaction between the two predictors.

It is a multiple regression with $p = 6$ and

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}x_{12} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}^2 & x_{n2}^2 & x_{n1}x_{n2} \end{pmatrix}, \qquad \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}.$$

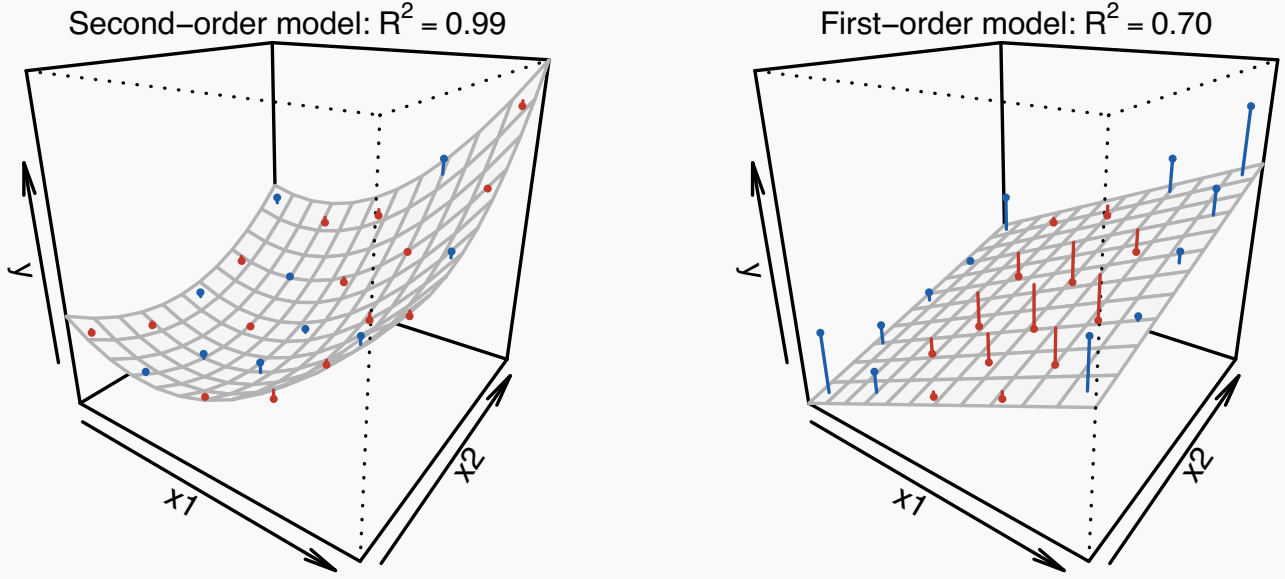
With three predictors: 3 linear terms, 3 squares and 3 products, so $p = 10$.

In R, `I()` computes a term before it enters the model:

```
lm(y ~ x1 + x2 + I(x1^2) + I(x2^2) + I(x1 * x2))
```

A curved surface against a plane

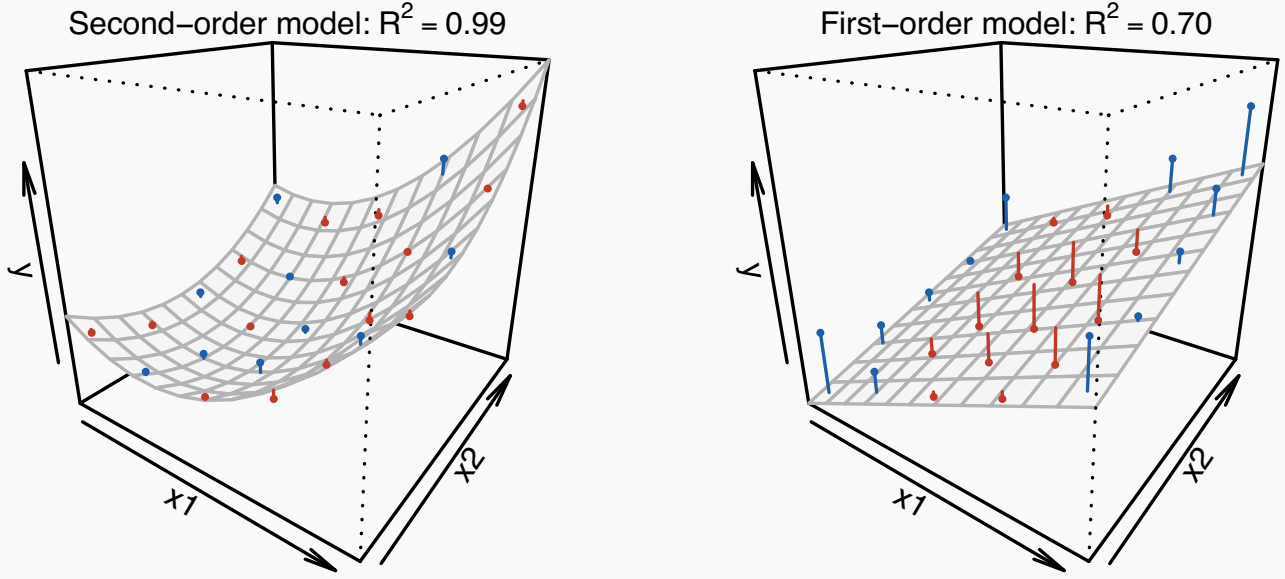
Simulated data ($n = 25$) from a bowl, and two fitted surfaces:



Blue: points above the fitted surface; red: points below.

A curved surface against a plane

Simulated data ($n = 25$) from a bowl, and two fitted surfaces:



Blue: points above the fitted surface; red: points below.

The plane $b_0 + b_1 x_1 + b_2 x_2$ has the right tilt but cannot bend: the points lie below it in the middle, above it at the edges. The squares and the product let the surface bend.

Interaction between two quantitative predictors

The product term on its own, without the squares:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon : \qquad \text{slope of } X_2 \text{ is } \beta_2 + \beta_3 X_1.$$

The effect of one predictor changes linearly with the value of the other.

Interaction between two quantitative predictors

The product term on its own, without the squares:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon : \qquad \text{slope of } X_2 \text{ is } \beta_2 + \beta_3 X_1.$$

The effect of one predictor changes linearly with the value of the other.

Acceleration data, `acc ~ wt * hp`:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.194e+01	1.023e+00	21.44	9.89e-68
wt	1.328e-03	3.379e-04	3.93	1.00e-04
hp	-1.327e-01	1.191e-02	-11.14	3.65e-25
wt:hp	1.037e-05	2.904e-06	3.57	4.01e-04

Interaction between two quantitative predictors

The product term on its own, without the squares:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon : \qquad \text{slope of } X_2 \text{ is } \beta_2 + \beta_3 X_1.$$

The effect of one predictor changes linearly with the value of the other.

Acceleration data, `acc` $\sim$ `wt * hp`:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.194e+01	1.023e+00	21.44	9.89e-68
wt	1.328e-03	3.379e-04	3.93	1.00e-04
hp	-1.327e-01	1.191e-02	-11.14	3.65e-25
wt:hp	1.037e-05	2.904e-06	3.57	4.01e-04

Slope of horsepower: $\hat{\beta}_2 + \hat{\beta}_3$ `wt`, that is

- -0.112 s per hp for a car of 2000 lb,
- -0.091 s per hp for a car of 4000 lb: power helps less in heavy cars.

Example: power cells

A preliminary experiment on a new type of power cell. Response Y: the number of discharge-charge cycles before the cell failed. Predictors: the charge rate X_1 (amperes) and the temperature X_2 (degrees Celsius), each at three levels.

Y	150	86	49	288	157	131	184	109	279	235	224
X_1	0.6	1.0	1.4	0.6	1.0	1.0	1.0	1.4	0.6	1.0	1.4
X_2	10	10	10	20	20	20	20	20	30	30	30

Example: power cells

A preliminary experiment on a new type of power cell. Response Y: the number of discharge-charge cycles before the cell failed. Predictors: the charge rate X_1 (amperes) and the temperature X_2 (degrees Celsius), each at three levels.

Y	150	86	49	288	157	131	184	109	279	235	224
X_1	0.6	1.0	1.4	0.6	1.0	1.0	1.0	1.4	0.6	1.0	1.4
X_2	10	10	10	20	20	20	20	20	30	30	30

Each of the $3 \times 3 = 9$ combinations of a charge rate and a temperature was tested on one cell, except the middle one, $(X_1, X_2) = (1.0, 20)$, which was tested on three cells: $n = 9 + 2 = 11$.

The second-order model has $p = 6$ parameters, so $n - p = 5$ degrees of freedom are left to estimate σ^2.

The second-order model on the raw predictors

`lm(Y ~ X1 + X2 + I(X1^2) + I(X2^2) + I(X1 * X2))`:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	337.7215	149.9616	2.2521	0.0741
X1	-539.5175	268.8603	-2.0067	0.1011
X2	8.9171	9.1825	0.9711	0.3761
I(X1^2)	171.2171	127.1255	1.3468	0.2359
I(X2^2)	-0.1061	0.2034	-0.5214	0.6244
I(X1 * X2)	2.8750	4.0468	0.7104	0.5092

$R^2 = 0.914$, $F^* = 10.57$ on 5 and 5 df, $p = 0.011$.

The second-order model on the raw predictors

`lm(Y ~ X1 + X2 + I(X1^2) + I(X2^2) + I(X1 * X2))`:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	337.7215	149.9616	2.2521	0.0741
X1	-539.5175	268.8603	-2.0067	0.1011
X2	8.9171	9.1825	0.9711	0.3761
I(X1^2)	171.2171	127.1255	1.3468	0.2359
I(X2^2)	-0.1061	0.2034	-0.5214	0.6244
I(X1 * X2)	2.8750	4.0468	0.7104	0.5092

$R^2 = 0.914$, $F^* = 10.57$ on 5 and 5 df, $p = 0.011$.

The model as a whole is significant, but no single coefficient is. A sign of multicollinearity.

The second-order model on the raw predictors

```
lm(Y ~ X1 + X2 + I(X1^2) + I(X2^2) + I(X1 * X2)):
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  337.7215   149.9616  2.2521   0.0741
X1          -539.5175   268.8603 -2.0067   0.1011
X2             8.9171     9.1825  0.9711   0.3761
I(X1^2)      171.2171   127.1255  1.3468   0.2359
I(X2^2)       -0.1061     0.2034 -0.5214   0.6244
I(X1 * X2)     2.8750     4.0468  0.7104   0.5092
```

$R^2 = 0.914$, $F^* = 10.57$ on 5 and 5 df, $p = 0.011$.

The model as a whole is significant, but no single coefficient is. A sign of multicollinearity.

Indeed $r(X_1, X_1^2) = 0.991$ and $r(X_2, X_2^2) = 0.986$: for X_1, the factor $1/(1 - r^2)$ of Lecture 16 is already 56.

Centring, and the centred fit

Centre at the means: $x_1 = X_1 - 1$, $x_2 = X_2 - 20$. The correlation between the linear term and its square drops from 0.991 to 0.000, for both predictors. The same model on `x1` and `x2`:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  162.8421    16.6076  9.8053   0.0002
x1          -139.5833    33.0418 -4.2245   0.0083
x2            7.5500     1.3217  5.7125   0.0023
I(x1^2)     171.2171   127.1255  1.3468   0.2359
I(x2^2)      -0.1061     0.2034 -0.5214   0.6244
I(x1 * x2)    2.8750     4.0468  0.7104   0.5092
```

Centring, and the centred fit

Centre at the means: $x_1 = X_1 - 1$, $x_2 = X_2 - 20$. The correlation between the linear term and its square drops from 0.991 to 0.000, for both predictors. The same model on `x1` and `x2`:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	162.8421	16.6076	9.8053	0.0002
x1	-139.5833	33.0418	-4.2245	0.0083
x2	7.5500	1.3217	5.7125	0.0023
I(x1^2)	171.2171	127.1255	1.3468	0.2359
I(x2^2)	-0.1061	0.2034	-0.5214	0.6244
I(x1 * x2)	2.8750	4.0468	0.7104	0.5092

- The linear terms are now clearly significant ($p = 0.008$ and 0.002): they are the slopes at the centre of the design.
- The second-order coefficients and their tests, the fitted values and $R^2 = 0.914$ are unchanged.

Centring, and the centred fit

Centre at the means: $x_1 = X_1 - 1$, $x_2 = X_2 - 20$. The correlation between the linear term and its square drops from 0.991 to 0.000, for both predictors. The same model on `x1` and `x2`:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  162.8421    16.6076  9.8053   0.0002
x1          -139.5833    33.0418 -4.2245   0.0083
x2             7.5500     1.3217  5.7125   0.0023
I(x1^2)      171.2171   127.1255  1.3468   0.2359
I(x2^2)       -0.1061     0.2034 -0.5214   0.6244
I(x1 * x2)     2.8750     4.0468  0.7104   0.5092
```

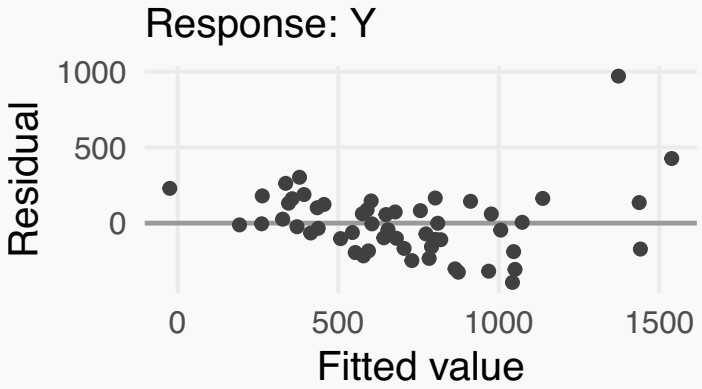
- The linear terms are now clearly significant ($p = 0.008$ and 0.002): they are the slopes at the centre of the design.
- The second-order coefficients and their tests, the fitted values and $R^2 = 0.914$ are unchanged.

Are the three second-order terms needed at all? A test with full and reduced models ($F^* = 0.78$ on 3 and 5 df): this week's problem set.

Transforming the response

When to take logs

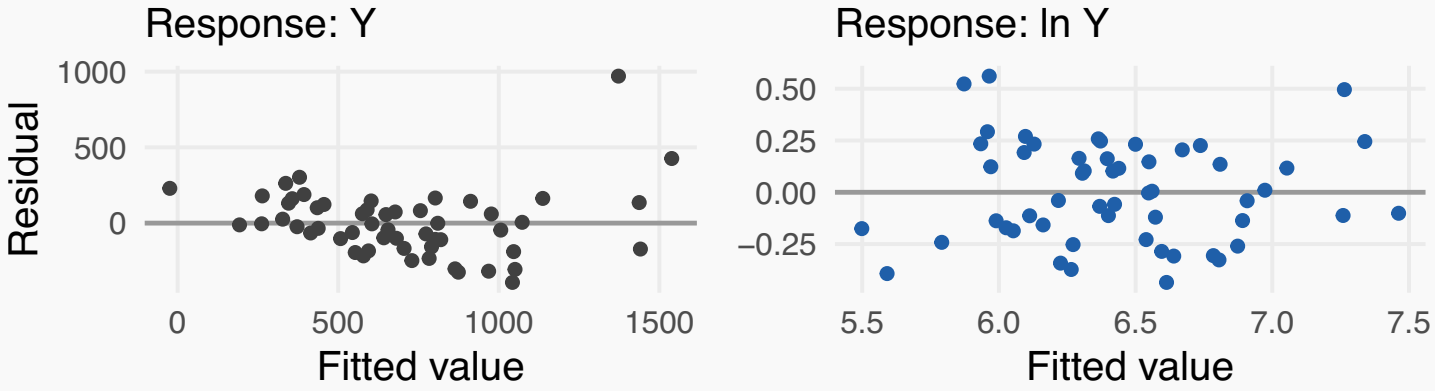
Survival time Y (days) of 54 patients after a liver operation, on four scores measured before it (X_1 to X_4). Residuals against fitted values:



On the raw scale: larger residuals at larger fitted values, and a long right tail (Shapiro-Wilk: $p = 0.0003$).

When to take logs

Survival time Y (days) of 54 patients after a liver operation, on four scores measured before it (X_1 to X_4). Residuals against fitted values:

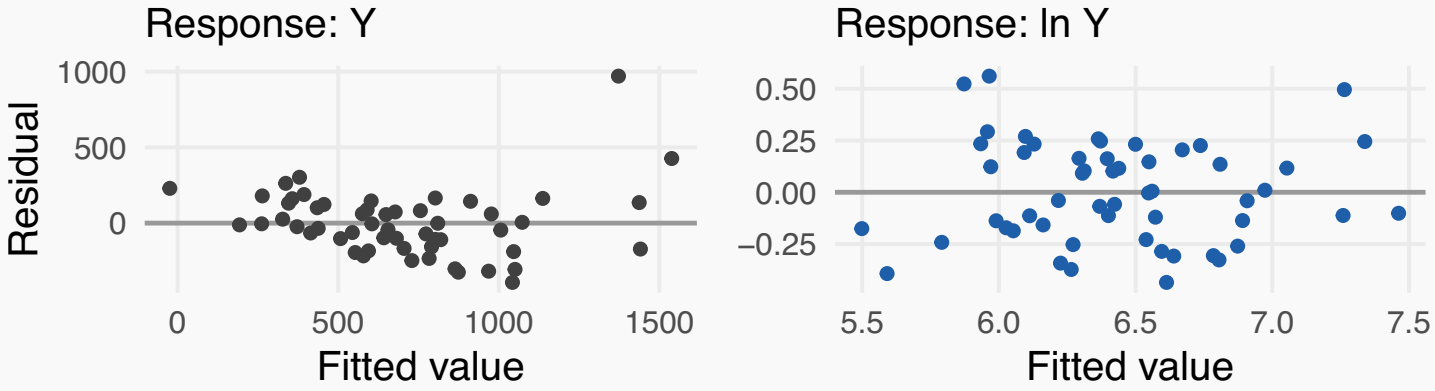


On the raw scale: larger residuals at larger fitted values, and a long right tail (Shapiro-Wilk: $p = 0.0003$).

Box-Cox: $\hat{\lambda} = 0.03$, with 95% interval $(-0.25, 0.30)$, which contains 0: take $\ln Y$. The residuals are then fine (Shapiro-Wilk: $p = 0.18$).

When to take logs

Survival time Y (days) of 54 patients after a liver operation, on four scores measured before it (X_1 to X_4). Residuals against fitted values:



On the raw scale: larger residuals at larger fitted values, and a long right tail (Shapiro-Wilk: $p = 0.0003$).

Box-Cox: $\hat{\lambda} = 0.03$, with 95% interval $(-0.25, 0.30)$, which contains 0: take $\ln Y$. The residuals are then fine (Shapiro-Wilk: $p = 0.18$).

When to try $\ln Y$: Y is positive, with a few very large values, and the residuals spread out as the fitted values grow (left panel). Typical of prices, wages, salaries and durations.

Reading a model with a log response

`lm(log(Y) ~ X1 + X2 + X3 + X4)`:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.8519	0.2663	14.4666	0.0000
X1	0.0837	0.0288	2.9042	0.0055
X2	0.0127	0.0023	5.4736	0.0000
X3	0.0156	0.0021	7.4403	0.0000
X4	0.0321	0.0515	0.6229	0.5363

Reading a model with a log response

`lm(log(Y) ~ X1 + X2 + X3 + X4)`:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.8519	0.2663	14.4666	0.0000
X1	0.0837	0.0288	2.9042	0.0055
X2	0.0127	0.0023	5.4736	0.0000
X3	0.0156	0.0021	7.4403	0.0000
X4	0.0321	0.0515	0.6229	0.5363

One more unit of X_k, the others held fixed, adds b_k to $\ln Y$: it multiplies Y by e^{b_k}, a change of $100\,(e^{b_k} - 1)\,\%$.

- Enzyme score: $e^{0.0156} = 1.016$, survival 1.6% longer per point; 95% interval from $e^{0.0114}$ to $e^{0.0198}$: 1.1% to 2.0%.
- Blood clotting score: $e^{0.0837} = 1.087$, 8.7% per point. When b_k is small, $100\,(e^{b_k} - 1) \approx 100\,b_k$.

Reading a model with a log response

`lm(log(Y) ~ X1 + X2 + X3 + X4)`:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.8519	0.2663	14.4666	0.0000
X1	0.0837	0.0288	2.9042	0.0055
X2	0.0127	0.0023	5.4736	0.0000
X3	0.0156	0.0021	7.4403	0.0000
X4	0.0321	0.0515	0.6229	0.5363

One more unit of X_k, the others held fixed, adds b_k to $\ln Y$: it multiplies Y by e^{b_k}, a change of $100\,(e^{b_k} - 1)\,\%$.

- Enzyme score: $e^{0.0156} = 1.016$, survival 1.6% longer per point; 95% interval from $e^{0.0114}$ to $e^{0.0198}$: 1.1% to 2.0%.
- Blood clotting score: $e^{0.0837} = 1.087$, 8.7% per point. When b_k is small, $100\,(e^{b_k} - 1) \approx 100\,b_k$.

What is multiplied is the median of Y (Assignment 2, Question 6): $e^{\hat{Y}'}$, with $\hat{Y}'$ the fitted value of $\ln Y$, estimates the median survival time, not the mean.