Lecture 3:

Linear Models

Dr. Roman V Belavkin

Middlesex University

BIS4435

Introduction to Modelling

Linear Functions - Lines, Planes and Hyperplanes

Overview of Linear Models

Creating a Linear Model

Correlation and Dependency

Advantages and Limitations of Linear Models

WHAT IS A MODEL?

- The word '*model*' comes from a Latin word meaning '*small*', and usually we mean some small representation of the real object (e.g. a model of a house, a car or an aircraft)

$$\text{Reality} > \text{Model}$$

WHAT IS A MODEL?

- The word '*model*' comes from a Latin word meaning '*small*', and usually we mean some small representation of the real object (e.g. a model of a house, a car or an aircraft)

$$\text{Reality} > \text{Model}$$

- Some models can be larger in size (e.g. a model of an atom). We shall mean 'smaller' in terms of complexity or uncertainty

$$\text{Uncertainty(Reality)} > \text{Uncertainty(Model)}$$

WHAT IS A MODEL?

- The word '*model*' comes from a Latin word meaning '*small*', and usually we mean some small representation of the real object (e.g. a model of a house, a car or an aircraft)

$$\text{Reality} > \text{Model}$$

- Some models can be larger in size (e.g. a model of an atom). We shall mean 'smaller' in terms of complexity or uncertainty

$$\text{Uncertainty(Reality)} > \text{Uncertainty(Model)}$$

- The remaining part is called the *error* of a model. Thus

$$\text{Reality} = \text{Model} + \text{Error}$$

THE KEY QUESTION

- A system (reality) consists of objects that interact and depend on each other.

THE KEY QUESTION

- A system (reality) consists of objects that interact and depend on each other.
- We represent these objects by variables, say, x and y

THE KEY QUESTION

- A system (reality) consists of objects that interact and depend on each other.
- We represent these objects by variables, say, x and y
- Earlier, we have seen how joint probability can reveal the dependency between events $(P(x, y) \neq P(x)P(y))$

THE KEY QUESTION

- A system (reality) consists of objects that interact and depend on each other.
- We represent these objects by variables, say, x and y
- Earlier, we have seen how joint probability can reveal the dependency between events ($P(x, y) \neq P(x)P(y)$)
- What dependency is it exactly?

THE KEY QUESTION

- A system (reality) consists of objects that interact and depend on each other.
- We represent these objects by variables, say, x and y
- Earlier, we have seen how joint probability can reveal the dependency between events $(P(x, y) \neq P(x)P(y))$
- What dependency is it exactly?
- Can we find a mapping (a function) between x and y that will represent this dependency?

$$y = f(x)$$

THE KEY QUESTION

- A system (reality) consists of objects that interact and depend on each other.
- We represent these objects by variables, say, x and y
- Earlier, we have seen how joint probability can reveal the dependency between events ($P(x, y) \neq P(x)P(y)$)
- What dependency is it exactly?
- Can we find a mapping (a function) between x and y that will represent this dependency?

$$y = f(x)$$

Example

If $f(x) = \frac{x}{2}$, what is y for $x = 4$, $x = 6$, $x = 1000$?

DATA-DRIVEN MODELS

Case:	Age	Gender	Income ($K p.m.)	Expenses ($K p.m.)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

- Data is a ‘footprint’ of reality.

DATA-DRIVEN MODELS

Case:	Age	Gender	Income ($K p.m.)	Expenses ($K p.m.)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

- Data is a ‘footprint’ of reality.
- Does the credit score depend on a person’s income?

DATA-DRIVEN MODELS

Case:	Age	Gender	Income ($K p.m.)	Expenses ($K p.m.)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

- Data is a 'footprint' of reality.
- Does the credit score depend on a person's income?
- Can we find a function $f(\cdot)$ such that

$$\text{Credit score} = f(\text{Income, Expenses, Age, Gender}, \ldots)$$

DATA-DRIVEN MODELS

Case:	Age	Gender	Income ($K p.m.)	Expenses ($K p.m.)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

- Data is a 'footprint' of reality.
- Does the credit score depend on a person's income?
- Can we find a function $f(\cdot)$ such that

$$\text{Credit score} = f(\text{Income, Expenses, Age, Gender}, \ldots)$$

- Data-driven modelling is a search for such functions that represent the dependencies between different variables.

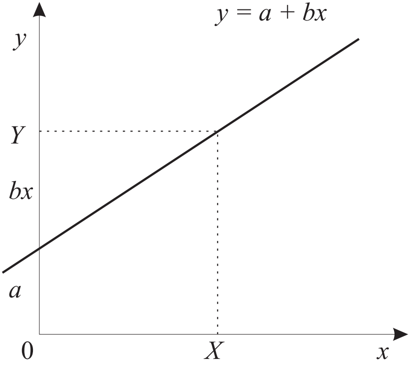
LINEAR FUNCTIONS

- There are many kinds of functions $y = f(x)$. The simplest are the linear functions.

LINEAR FUNCTIONS

- There are many kinds of functions $y = f(x)$. The simplest are the linear functions.
- A line on xy-plane is defined by the linear function

$$y = a + b\,x$$

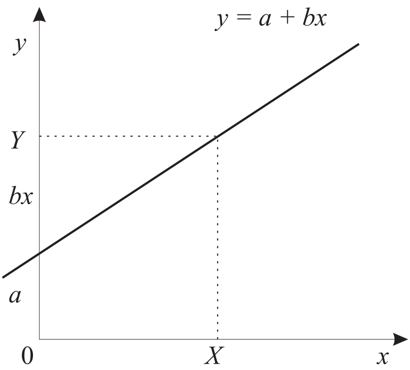


where a is the **intercept**, and b is called the **slope** of the line, and together that are the prameters of the model.

LINEAR FUNCTIONS

- There are many kinds of functions $y = f(x)$. The simplest are the linear functions.
- A line on xy-plane is defined by the linear function

$$y = a + b\,x$$



where a is the **intercept**, and b is called the **slope** of the line, and together that are the prameters of the model.

Example

$$y = 2 + 3\,x \qquad (a = 2\,,\ \ b = 3)$$

What value is y when $x = 4$?

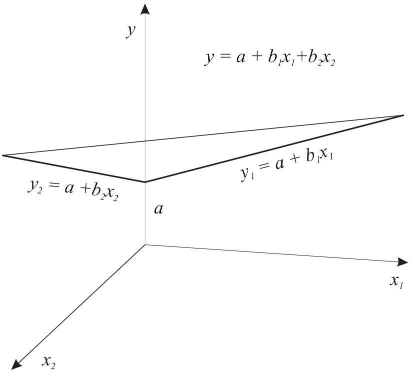
PLANES and HYPERPLANES

- Variable y may depend on several variables x_1, x_2,...

PLANES and HYPERPLANES

- Variable y may depend on several variables x_1, x_2,...
- A linear function $y = f(x_1, x_2)$ of two variables describes a plane, which we can plot on an x, y, z (or x_1, x_2, y) chart.

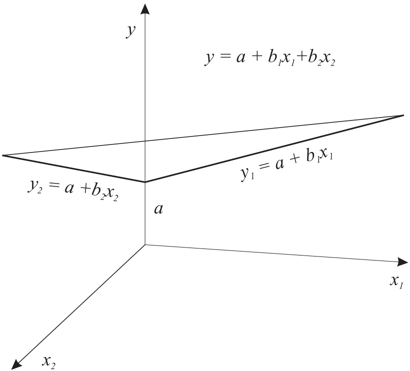
$$y = a + b_1\,x_1 + b_2\,x_2$$



PLANES and HYPERPLANES

- Variable y may depend on several variables x_1, x_2,...
- A linear function $y = f(x_1, x_2)$ of two variables describes a plane, which we can plot on an x, y, z (or x_1, x_2, y) chart.

$$y = a + b_1\, x_1 + b_2\, x_2$$



A linear function of m variables $y = f(x_1, \ldots, x_m)$ defines a hyperplane in an $m+1$ dimensional space

$$y = a + b_1\, x_1 + \cdots + b_m\, x_m$$

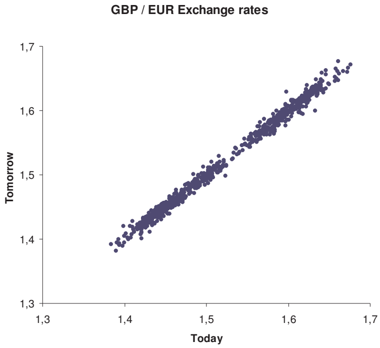
with one intercept and m slopes (there are $m+1$ parameters).

LINEAR FORECAST

- A forecast can be made using a function relating past to the future.
- In the coursework, we can denote the exchange rate of today by x and the rate of tomorrow by y.

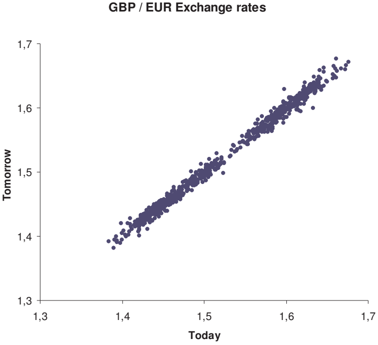
LINEAR FORECAST

- A forecast can be made using a function relating past to the future.
- In the coursework, we can denote the exchange rate of today by x and the rate of tomorrow by y.



LINEAR FORECAST

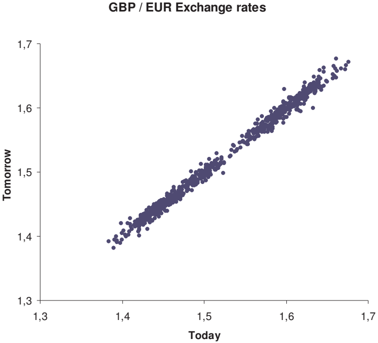
- A forecast can be made using a function relating past to the future.
- In the coursework, we can denote the exchange rate of today by x and the rate of tomorrow by y.



- Does the relation $y = f(x)$ look linear?

LINEAR FORECAST

- A forecast can be made using a function relating past to the future.
- In the coursework, we can denote the exchange rate of today by x and the rate of tomorrow by y.



- Does the relation $y = f(x)$ look linear?
- Can you suggest values for the intercept (a) and the slope (b)?

$$y \approx f(x) = a + b\,x$$

LINEAR FORECAST (cont.)

- It would be more useful to predict the *differences* between two consequitive values, or the returns Δy

$$\Delta y_i = y_i - y_{i-1}$$

LINEAR FORECAST (cont.)

- It would be more useful to predict the *differences* between two consequitive values, or the returns Δy

$$\Delta y_i = y_i - y_{i-1}$$

LINEAR FORECAST (cont.)

- It would be more useful to predict the *differences* between two consequitive values, or the returns Δy

$$\Delta y_i = y_i - y_{i-1}$$

- Does this relation look linear?

LINEAR FORECAST (cont.)

- It would be more useful to predict the *differences* between two consequitive values, or the returns Δy

$$\Delta y_i = y_i - y_{i-1}$$

- Does this relation look linear?
- Do you think from the chart that the return of tomorrow depend on the return of today?

THE COST OF THE ERROR

- Which function is the ‘best’ or the optimal model?

THE COST OF THE ERROR

- Which function is the ‘best’ or the optimal model?
- Recall that the model is only an approximation of the reality, and there is also the error, and a good model should have a small error

$$y = f(x) + \text{Error}$$

THE COST OF THE ERROR

- Which function is the 'best' or the optimal model?
- Recall that the model is only an approximation of the reality, and there is also the error, and a good model should have a small error

$$y = f(x) + \text{Error}$$

- The way we measure the error is defined by the cost function

$$c(y, f(x)) = \text{Error}$$

THE COST OF THE ERROR

- Which function is the 'best' or the optimal model?
- Recall that the model is only an approximation of the reality, and there is also the error, and a good model should have a small error

$$y = f(x) + \text{Error}$$

- The way we measure the error is defined by the cost function

$$c(y, f(x)) = \text{Error}$$

Example

The absolute cost (aka the absolute deviation):

$$c(y, f(x)) = |y - f(x)|$$

The quadratic cost (aka the squared deviation):

$$c(y, f(x)) = |y - f(x)|^2$$

THE COST OF THE ERROR

- Which function is the 'best' or the optimal model?
- Recall that the model is only an approximation of the reality, and there is also the error, and a good model should have a small error

$$y = f(x) + \text{Error}$$

- The way we measure the error is defined by the cost function

$$c(y, f(x)) = \text{Error}$$

- The choice of $c(y, f(x))$ determines, which model is the best.

Example

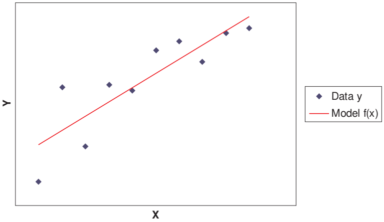
The absolute cost (aka the absolute deviation):

$$c(y, f(x)) = |y - f(x)|$$

The quadratic cost (aka the squared deviation):

$$c(y, f(x)) = |y - f(x)|^2$$

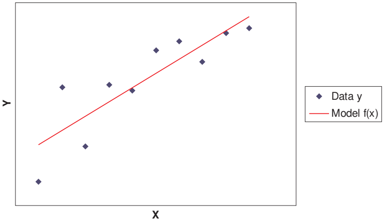
THE MEAN-SQUARE REGRESSION



THE MEAN-SQUARE REGRESSION

- It is known that if $c(y, f(x)) = |y - f(x)|^2$ (quadratic cost function), then on average the optimal model is the conditional expectation

$$y \approx E\{y \mid x\}$$

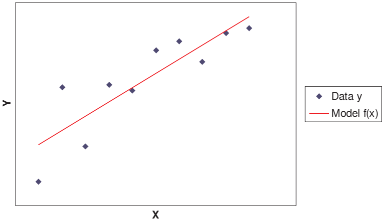


THE MEAN-SQUARE REGRESSION

- It is known that if $c(y, f(x)) = |y - f(x)|^2$ (quadratic cost function), then on average the optimal model is the conditional expectation

$$y \approx E\{y \mid x\}$$

- If we assume that $E\{y \mid x\}$ is a linear function $y = a + b\,x$, then we only need to find the intercept (a) and the slope (b) of a line that minimises the quadratic error on average.

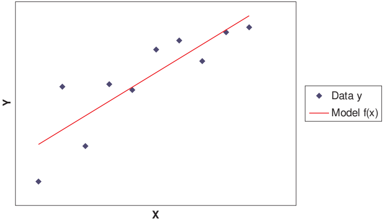


THE MEAN-SQUARE REGRESSION

- It is known that if $c(y, f(x)) = |y - f(x)|^2$ (quadratic cost function), then on average the optimal model is the conditional expectation

$$y \approx E\{y \mid x\}$$

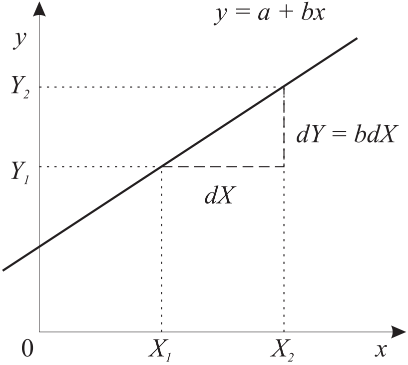
- If we assume that $E\{y \mid x\}$ is a linear function $y = a + b\,x$, then we only need to find the intercept (a) and the slope (b) of a line that minimises the quadratic error on average.



The full name of this kind of model is linear mean-square regression

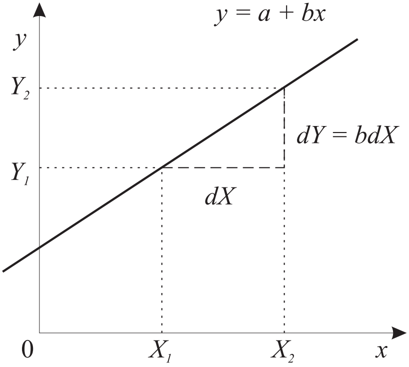
A LINE THROUGH TWO POINTS

- Given two points (X_1, Y_1) and (X_2, Y_2), find a and b.
- Two points uniquely define the line.



A LINE THROUGH TWO POINTS

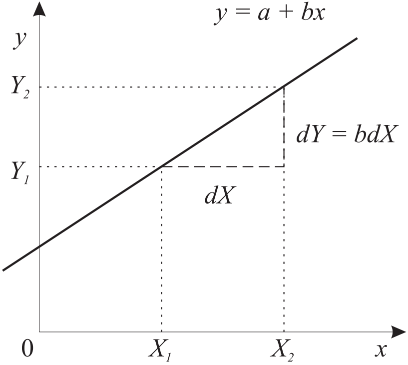
- Given two points (X_1, Y_1) and (X_2, Y_2), find a and b.
- Two points uniquely define the line.



$$b = \frac{dY}{dX} = \frac{Y_2 - Y_1}{X_2 - X_1}$$

A LINE THROUGH TWO POINTS

- Given two points (X_1, Y_1) and (X_2, Y_2), find a and b.
- Two points uniquely define the line.

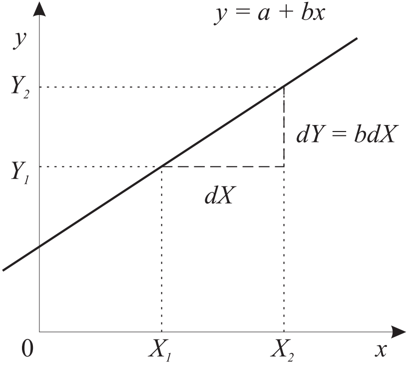


$$b = \frac{dY}{dX} = \frac{Y_2 - Y_1}{X_2 - X_1}$$

$$a = Y_1 - b\,X_1$$

A LINE THROUGH TWO POINTS

- Given two points (X_1, Y_1) and (X_2, Y_2), find a and b.
- Two points uniquely define the line.



$$b = \frac{dY}{dX} = \frac{Y_2 - Y_1}{X_2 - X_1}$$

$$a = Y_1 - b X_1$$

We can also write the equation for the line as

$$y = Y_1 + b(x - X_1)$$

A LINE THROUGH SEVERAL POINTS

x	y
X_1	Y_1
X_2	Y_2
$\vdots$	$\vdots$
X_n	Y_n

A LINE THROUGH SEVERAL POINTS

- For several points, the slope (b) is computed using the measures of dispersion (covariance and variance)

x	y
X_1	Y_1
X_2	Y_2
$\vdots$	$\vdots$
X_n	Y_n

A LINE THROUGH SEVERAL POINTS

- For several points, the slope (b) is computed using the measures of dispersion (covariance and variance)
- For the intercept (a), we use the fact that the line must go through the centre of gravity $(E\{x\}, E\{y\})$.

x	y
X_1	Y_1
X_2	Y_2
$\vdots$	$\vdots$
X_n	Y_n

A LINE THROUGH SEVERAL POINTS

- For several points, the slope (b) is computed using the measures of dispersion (covariance and variance)
- For the intercept (a), we use the fact that the line must go through the centre of gravity $(E\{x\}, E\{y\})$.

x	y
X_1	Y_1
X_2	Y_2
$\vdots$	$\vdots$
X_n	Y_n

$$b = \frac{Cov(x,y)}{Var(x)}$$

A LINE THROUGH SEVERAL POINTS

- For several points, the slope (b) is computed using the measures of dispersion (covariance and variance)
- For the intercept (a), we use the fact that the line must go through the centre of gravity $(E\{x\}, E\{y\})$.

x	y
X_1	Y_1
X_2	Y_2
$\vdots$	$\vdots$
X_n	Y_n

$$b = \frac{Cov(x,y)}{Var(x)}$$

$$a = E\{y\} - b\,E\{x\}$$

A LINE THROUGH SEVERAL POINTS

- For several points, the slope (b) is computed using the measures of dispersion (covariance and variance)
- For the intercept (a), we use the fact that the line must go through the centre of gravity $(E\{x\}, E\{y\})$.

x	y
X_1	Y_1
X_2	Y_2
$\vdots$	$\vdots$
X_n	Y_n

$$b = \frac{Cov(x,y)}{Var(x)}$$

$$a = E\{y\} - b\,E\{x\}$$

And the equation for the line is

$$y = E\{y\} + b\,[x - E\{x\}]$$

A SIMPLE MODEL FOR CREDIT SCORE

Monthly Income ($ K)	Credit Score
2	3
1	1
6	5
3	4

A SIMPLE MODEL FOR CREDIT SCORE

Monthly Income (\$ K)	Credit Score
2	3
1	1
6	5
3	4

- Denote by x the income and by y the credit score.

A SIMPLE MODEL FOR CREDIT SCORE

Monthly Income (\$ K)	Credit Score
2	3
1	1
6	5
3	4

- Denote by x the income and by y the credit score.
- Construct a linear model $y \approx \color{red}{a} + \color{blue}{b}\,x$

A SIMPLE MODEL FOR CREDIT SCORE

Monthly Income ($ K)	Credit Score
2	3
1	1
6	5
3	4

- Denote by x the income and by y the credit score.
- Construct a linear model $y \approx a + b\,x$
- We need to find slope (b) and intercept (a) from the data

SOLUTION

$$E\{x\} = (2+1+6+3)/4 = 3$$

SOLUTION

$$
\begin{aligned}
E\{x\} &= (2+1+6+3)/4 = 3 \\
E\{y\} &= (3+1+5+4)/4 = 3,25
\end{aligned}
$$

SOLUTION

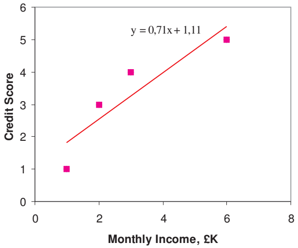
$$\begin{aligned} E\{x\} &= (2+1+6+3)/4 = 3 \\ E\{y\} &= (3+1+5+4)/4 = 3,25 \\ Cov(x,y) &= [(2-3)(3-3,25)+\cdots+(3-3)(4-3,25)]/4 = 2,5 \end{aligned}$$

SOLUTION

$$\begin{aligned}
E\{x\} &= (2+1+6+3)/4 = 3 \\
E\{y\} &= (3+1+5+4)/4 = 3,25 \\
Cov(x,y) &= [(2-3)(3-3,25)+\cdots+(3-3)(4-3,25)]/4 = 2,5 \\
Var(x) &= [(2-3)^2+\cdots+(3-3)^2]/4 = 3,5
\end{aligned}$$

SOLUTION

$$
\begin{aligned}
E\{x\} &= (2+1+6+3)/4 = 3 \\
E\{y\} &= (3+1+5+4)/4 = 3,25 \\
Cov(x,y) &= [(2-3)(3-3,25)+\cdots+(3-3)(4-3,25)]/4 = 2,5 \\
Var(x) &= [(2-3)^2+\cdots+(3-3)^2]/4 = 3,5
\end{aligned}
$$



$$
\begin{aligned}
{\color{blue}b} &= \frac{Cov(x,y)}{Var(x)} = 0,71 \\
{\color{red}a} &= E\{y\} - {\color{blue}b}\,E\{x\} = 1,11
\end{aligned}
$$

MULTIPLE LINEAR REGRESSION

x_1	x_2	$\cdots$	x_m	y
X_{11}	X_{12}	$\cdots$	X_{1m}	Y_1
X_{21}	X_{22}	$\cdots$	X_{2m}	Y_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
X_{n1}	X_{n2}	$\cdots$	X_{nm}	Y_n

MULTIPLE LINEAR REGRESSION

x_1	x_2	$\cdots$	x_m	y
X_{11}	X_{12}	$\cdots$	X_{1m}	Y_1
X_{21}	X_{22}	$\cdots$	X_{2m}	Y_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
X_{n1}	X_{n2}	$\cdots$	X_{nm}	Y_n

- Here, y depends not on one, but on several variables

$$y \approx f(x_1, \ldots, x_m) = a + b_1 x_1 + \cdots + b_m x_m$$

MULTIPLE LINEAR REGRESSION

x_1	x_2	$\cdots$	x_m	y
X_{11}	X_{12}	$\cdots$	X_{1m}	Y_1
X_{21}	X_{22}	$\cdots$	X_{2m}	Y_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
X_{n1}	X_{n2}	$\cdots$	X_{nm}	Y_n

- Here, y depends not on one, but on several variables

$$y \approx f(x_1, \dots, x_m) = a + b_1\, x_1 + \cdots + b_m\, x_m$$

- Thus, we need to find one intercept a and m **regression coefficients** b_1, b_2, …, b_m ('slopes')

A MORE COMPLEX CREDIT SCORE MODEL

Monthly Income ($ K)	Monthly Expenses ($ K)	Credit Score
2	1	3
1	2	1
6	2	5
3	1	4

A MORE COMPLEX CREDIT SCORE MODEL

Monthly Income (\$ K)	Monthly Expenses (\$ K)	Credit Score
2	1	3
1	2	1
6	2	5
3	1	4

- Denote by x_1 the income, by x_2 expenses and by y the credit score.

A MORE COMPLEX CREDIT SCORE MODEL

Monthly Income (\$ K)	Monthly Expenses (\$ K)	Credit Score
2	1	3
1	2	1
6	2	5
3	1	4

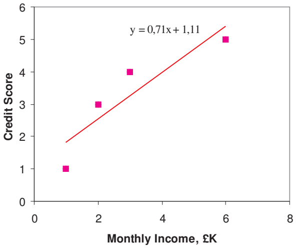
- Denote by x_1 the income, by x_2 expenses and by y the credit score.
- Construct a linear model $y \approx a + b_1 x_1 + b_2 x_2$

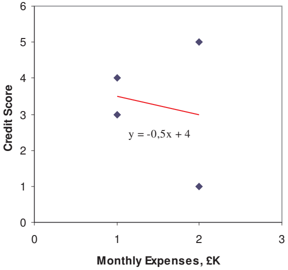
A MORE COMPLEX CREDIT SCORE MODEL

Monthly Income ($ K)	Monthly Expenses ($ K)	Credit Score
2	1	3
1	2	1
6	2	5
3	1	4

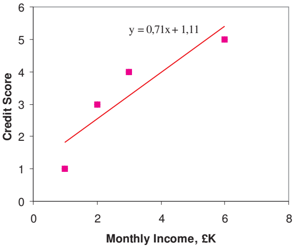
- Denote by x_1 the income, by x_2 expenses and by y the credit score.
- Construct a linear model $y \approx a + b_1 x_1 + b_2 x_2$
- We need to find two slopes (b_1, b_2) and one intercept (a)

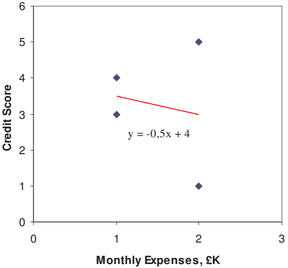
APPROXIMATE SOLUTION





APPROXIMATE SOLUTION

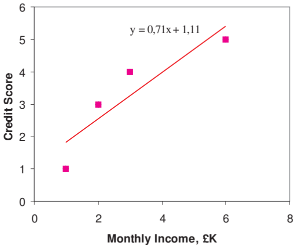


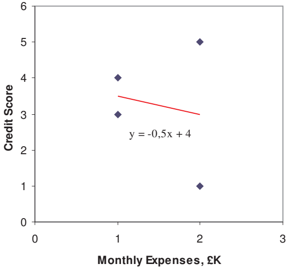


$$b_1 = \frac{Cov(x_1, y)}{Var(x_1)} = 0,71$$

$$b_2 = \frac{Cov(x_2, y)}{Var(x_2)} = -0,5$$

APPROXIMATE SOLUTION





$$b_1 = \frac{Cov(x_1, y)}{Var(x_1)} = 0,71 \qquad b_2 = \frac{Cov(x_2, y)}{Var(x_2)} = -0,5$$

$$a = E\{y\} - b_1\, E\{x_1\} - b_2\, E\{x_2\} = 1,86$$

$$f(x_1, x_2) = 1,86 + 0,71\, x_1 - 0,5\, x_2$$

SLOPE AND CORRELATION

- Recall that correlation is

$$Corr(x,y) = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$$

SLOPE AND CORRELATION

- Recall that correlation is

$$Corr(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}}$$

- Thus, we can compute the slope as

$$b = \frac{Cov(x, y)}{Var(x)} = Corr(x, y)\sqrt{\frac{Var(y)}{Var(x)}}$$

SLOPE AND CORRELATION

- Recall that correlation is

$$Corr(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}}$$

- Thus, we can compute the slope as

$$b = \frac{Cov(x, y)}{Var(x)} = Corr(x, y)\sqrt{\frac{Var(y)}{Var(x)}}$$

- Positive correlation means positive slope $b > 0$

SLOPE AND CORRELATION

- Recall that correlation is

$$Corr(x,y) = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$$

- Thus, we can compute the slope as

$$b = \frac{Cov(x,y)}{Var(x)} = Corr(x,y)\sqrt{\frac{Var(y)}{Var(x)}}$$

- Positive correlation means positive slope $b > 0$
- Negative correlation means negative slope $b < 0$ (anticorrelated)

SLOPE AND CORRELATION

- Recall that correlation is

$$Corr(x,y) = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$$

- Thus, we can compute the slope as

$$b = \frac{Cov(x,y)}{Var(x)} = Corr(x,y)\sqrt{\frac{Var(y)}{Var(x)}}$$

- Positive correlation means positive slope $b > 0$
- Negative correlation means negative slope $b < 0$ (anticorrelated)
- Zero correlation means zero slope $b = 0$ (uncorrelated)

CORRELATION AND DEPENDENCY

- Correlation indicates the degree of a linear dependency

CORRELATION AND DEPENDENCY

- Correlation indicates the degree of a linear dependency
- Zero correlation means there is no linear dependency

CORRELATION AND DEPENDENCY

- Correlation indicates the degree of a linear dependency
- Zero correlation means there is no linear dependency
- Zero correlation, however, does not imply independence

$$P(x, y) = P(x)\, P(y)$$

CORRELATION AND DEPENDENCY

- Correlation indicates the degree of a linear dependency
- Zero correlation means there is no linear dependency
- Zero correlation, however, does not imply independence

$$P(x, y) = P(x)\,P(y)$$

- Here, the correlation is 0,053
- There still can be some nonlinear dependency

CORRELATION IS NOT CAUSATION

- There is a positive correlation between sales of ice-cream and shark attacks. Does this mean that ice-cream causes shark attacks?

CORRELATION IS NOT CAUSATION

- There is a positive correlation between sales of ice-cream and shark attacks. Does this mean that ice-cream causes shark attacks?
- It is a common fallacy to conclude a causal relation based on correlation

CORRELATION IS NOT CAUSATION

- There is a positive correlation between sales of ice-cream and shark attacks. Does this mean that ice-cream causes shark attacks?
- It is a common fallacy to conclude a causal relation based on correlation
- Often, correlation between x and y can be because they both depend on (or caused by) a third variable z (e.g. both ice-cream sales and shark attacks increase in the summer season)

ADVANTAGES OF LINEAR MODELS

- Given data, they are easy to implement

ADVANTAGES OF LINEAR MODELS

- Given data, they are easy to implement
- Multiple linear mean-square regression is a standard feature of many analytical tools

ADVANTAGES OF LINEAR MODELS

- Given data, they are easy to implement
- Multiple linear mean-square regression is a standard feature of many analytical tools
- If there is a strong linearity in the data, then the mean-square regression can always find the optimal model

ADVANTAGES OF LINEAR MODELS

- Given data, they are easy to implement
- Multiple linear mean-square regression is a standard feature of many analytical tools
- If there is a strong linearity in the data, then the mean-square regression can always find the optimal model
- Such a model can be used to explain and understand the dependencies in data (i.e. using slopes or correlations)

ADVANTAGES OF LINEAR MODELS

- Given data, they are easy to implement
- Multiple linear mean-square regression is a standard feature of many analytical tools
- If there is a strong linearity in the data, then the mean-square regression can always find the optimal model
- Such a model can be used to explain and understand the dependencies in data (i.e. using slopes or correlations)
- The model can be used for prediction and 'what-if' analysis.

LIMITATIONS OF LINEAR MODELS

- There can be no significant linear dependency

LIMITATIONS OF LINEAR MODELS

- There can be no significant linear dependency
- Linear models cannot account for nonlinear effects

LIMITATIONS OF LINEAR MODELS

- There can be no significant linear dependency
- Linear models cannot account for nonlinear effects
- Mean-square error (quadratic cost) is very sensitive to outliers (unusual cases)

LIMITATIONS OF LINEAR MODELS

- There can be no significant linear dependency
- Linear models cannot account for nonlinear effects
- Mean-square error (quadratic cost) is very sensitive to outliers (unusual cases)
- It is much more difficult to find linear models optimising non-quadratic cost functions (e.g. an absolute error $|y - f(x)|$)

SUMMARY

- Models are simplified representations of reality

SUMMARY

- Models are simplified representations of reality
- The unexplained part of reality results in an error of the model

SUMMARY

- Models are simplified representations of reality
- The unexplained part of reality results in an error of the model
- Linear functions, defining lines, planes and hyperplanes, can be used to construct the simplest data-driven models

SUMMARY

- Models are simplified representations of reality
- The unexplained part of reality results in an error of the model
- Linear functions, defining lines, planes and hyperplanes, can be used to construct the simplest data-driven models
- Linear mean-square regression is a standard method of computing such models

SUMMARY

- Models are simplified representations of reality
- The unexplained part of reality results in an error of the model
- Linear functions, defining lines, planes and hyperplanes, can be used to construct the simplest data-driven models
- Linear mean-square regression is a standard method of computing such models
- Linear models can reveal linear dependencies in data