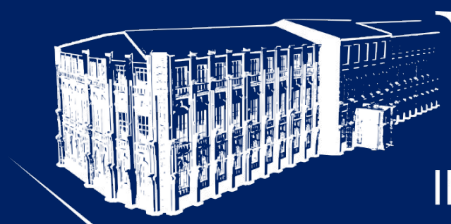

WESTMINSTER

INTERNATIONAL UNIVERSITY IN TASHKENT

An Accredited Institution of the University of Westminster (UK)

LECTURE 10
REGRESSION AND TIME SERIES

Saidgozi Saydumarov
Sherzodbek Safarov
QM Module Leaders
ssaydumarov@wiut.uz
s.safarov@wiut.uz

Office hours: by appointment

Lecture outline

- Quick review of regression
- Simple linear regression as conditional mean
- Using regression for estimation
- Using regression for trend in time series

Quick review

Regression *is a technique for determining the statistical relationship between two or more variables where a change in a dependent variable is associated with, and depends on, a change in one or more independent variables.*

$$y = \alpha + \beta x + \epsilon$$

To calculate coefficients of the regression equation, we need to find the descriptive statistics (average, standard deviation, covariance) of the data first.

Quick review

Regression example: Student mark vs absence

The following data about the average mark for a student and the number of hours the student was absent was collected from a group of 24 students.

We would like to see whether the number of hours a student was absent affects the marks that a student gets.

Student mark	Number of hours absent	Student mark	Number of hours absent
61	2	53	4
56	3	67	1
74	0	58	3
57	4	65	1
65	0	25	12
75	0	67	2
73	0	74	0
75	0	73	0
70	1	70	1
68	0	63	1
42	7	50	5
67	0	62	3

Quick review

Regression example: Student mark vs absence

Average

- Student mark: 62.9
- Hours absent per week: 2.1

Standard deviation

- Student mark: 11.8
- Hours absent per week: 2.8

Covariance: -32.3

$$Beta = \frac{-32.3}{2.8^2} = -4$$

$$Alpha = 62.9 + 4 * 2.1 = 71.2$$

Student mark	Number of hours absent	Student mark	Number of hours absent
61	2	53	4
56	3	67	1
74	0	58	3
57	4	65	1
65	0	25	12
75	0	67	2
73	0	74	0
75	0	73	0
70	1	70	1
68	0	63	1
42	7	50	5
67	0	62	3

$$Student\ mark = 71.2 - 4 * Hours\ absent$$

Regression as conditional mean

- What is the average mark of students that were never absent?

One approach is to calculate the mean from the sample of students:

$$= \frac{74 + 65 + 75 + 73 + 75 + 68 + 67 + 74 + 73}{9} = 71.5$$

But since we estimated the regression equation, we can use that too:

$$Student\ mark = \alpha + \beta * 0 = 71.2$$

Student mark	Number of hours absent	Student mark	Number of hours absent
61	2	53	4
56	3	67	1
74	0	58	3
57	4	65	1
65	0	25	12
75	0	67	2
73	0	74	0
75	0	73	0
70	1	70	1
68	0	63	1
42	7	50	5
67	0	62	3

So we could say:

$$Average\ student\ mark = 71.2 - 4 * Average\ hours\ absent$$

Regression as conditional mean

- What is the average mark of students that were absent for an hour on average?

We can calculate it from the sample data directly:

$$= \frac{70 + 65 + 67 + 70 + 63}{5} = 67$$

Or we can use the regression equation:

$$Student\ mark = \alpha + \beta * 1 = 71.2 - 4 * 1 = 67.2$$

The regression equation gives the mean conditional on a value for X. Here, it is the number of hours absent.

Student mark	Number of hours absent	Student mark	Number of hours absent
61	2	53	4
56	3	67	1
74	0	58	3
57	4	65	1
65	0	25	12
75	0	67	2
73	0	74	0
75	0	73	0
70	1	70	1
68	0	63	1
42	7	50	5
67	0	62	3

Using regression for estimation

- What is the average mark of students that were absent for 10 hours on average?

Note that we do not have any observations for a student who was absent for 10 hours.

But we can use the regression equation:

$$Avg\ student\ mark = 71.2 - 4 * 10 = 31.2$$

Student mark	Number of hours absent	Student mark	Number of hours absent
61	2	53	4
56	3	67	1
74	0	58	3
57	4	65	1
65	0	25	12
75	0	67	2
73	0	74	0
75	0	73	0
70	1	70	1
68	0	63	1
42	7	50	5
67	0	62	3

Using regression for estimation

Similarly, we can find the conditional mean for any number of hours absent:

Hours absent = 7

Average mark = 71.2 – 4 * 7 = 43.2

We can even estimate the student mark for number of hours absent outside the range of our data. For example:

Hours absent = 15

Average mark = 71.2 – 4 * 15 = 11.2

Student mark	Number of hours absent	Student mark	Number of hours absent
61	2	53	4
56	3	67	1
74	0	58	3
57	4	65	1
65	0	25	12
75	0	67	2
73	0	74	0
75	0	73	0
70	1	70	1
68	0	63	1
42	7	50	5
67	0	62	3

However, caution must be taken when estimating using out of sample ranges.

What happens when X = 18?

Using regression for trend line

In the previous week, we looked at forecasting time series data.

That included calculating the Centered Moving Average to get the trend component:

Year	Quarter	Number of visitors	CMA
2013	I	8,604	
	II	6,556	
	III	3,824	7,365
	IV	9,462	7,832
2014	I	10,628	8,179
	II	8,275	8,511
	III	4,881	
	IV	11,054	

Using regression for trend line

Since the trend is the long term changes in the data, it needs to be smooth and straight as possible. The regression equation suits this purpose very well.

$$Trend = \alpha + \beta * time$$

Time	CMA
1	
2	
3	7,365
4	7,832
5	8,179
6	8,511
7	
8	

Using regression for trend line

- $Trend = \alpha + \beta * time$

Average

Time = 4.5

Trend = 7,972

Variance

Time = 86 / 4 – 4.5^2 = 1.25

Covariance

=145,383 / 4 – 4.5 * 7,972 = 473.1

Beta

= 473.1 / 1.25 = 378.5

Alpha

= 7,972 – 378.5 * 4.5 = 6,268.4

Time	CMA (trend)	$Time^2$	Trend*Time
1			
2			
3	7,365	9	22,094
4	7,832	16	31,330
5	8,179	25	40,897
6	8,511	36	51,063
7			
8			
SUM		86	145,383

Using regression for trend line

- Forecasting the trend component for Q1 2015:

Time = 9

$$Trend = 6{,}268.4 + 378.5 * 9$$
$$= 9674.9$$

Year	Quarter	Time	Number of visitors	CMA
2013	I	1	8,604	
	II	2	6,556	
	III	3	3,824	7,365
	IV	4	9,462	7,832
2014	I	5	10,628	8,179
	II	6	8,275	8,511
	III	7	4,881	
	IV	8	11,054	

Jon Curwin and Roger Slater. *"Quantitative methods for Business Decisions,"* Chapters 15 and 17.