Chapter 2. Forecasting

Session 1. One-Variable Forecasting

Deduce the next number:

1, 2, 3…

1, 3, 5…

1, 1, 2, 3, 5, 8…

Why can we deduce the next number? Information from existing numbers.

If we have a time frame, we should take advantage of **previous information**.

Those numbers share a strict math pattern, but usually we do not have such a beautiful pattern.

Stephen Curry's Average Points

Age	PTS
21	17.5
22	18.6
23	14.7
24	22.9
25	24.0
26	23.8
27	30.1
28	25.3
29	27.6

Can you predict his average point next year?

The main advantage of prediction by a time frame is: the environment will not change largely, especially in a short time. Next year, we know Stephen Curry will still be a great player. Usually, the most recent information will be the most useful.

Assumption: things will not change a lot.

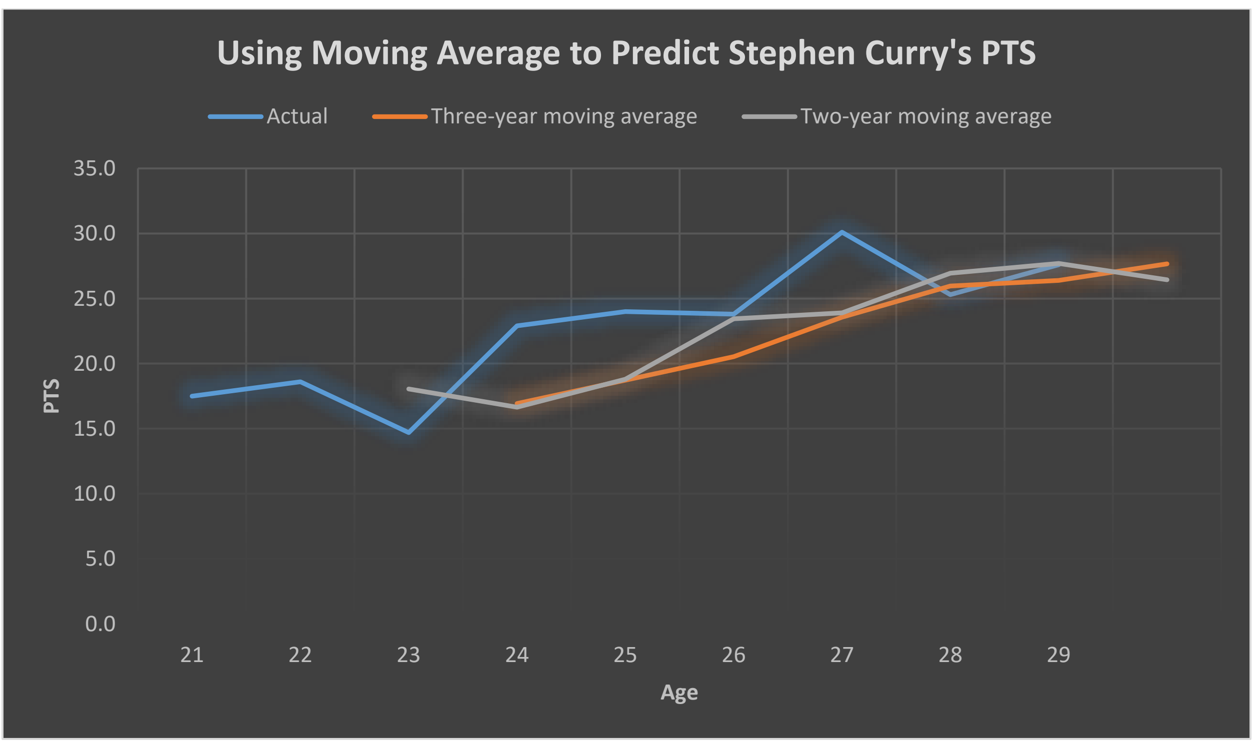
Method 1: Directly use data from last period.

However, we still have a lot of information left. Maybe we can use information from last two years, or three years.

S1.1 Moving Average

Average several periods of data

Actual	Three-year moving average	Two-year moving average
17.5		
18.6		
14.7		18.1
22.9	16.9	16.7
24.0	18.7	18.8
23.8	20.5	23.5
30.1	23.6	23.9
25.3	26.0	27.0
27.6	26.4	27.7
	27.7	26.5



S1.2 Weighted Moving Average

Adjusts moving average method to more closely reflect data fluctuations

E.g.

Month	Weight	Data
August	17%	130
September	33%	110
October	50%	90
November		?

November forecast

$0.17 \times 130 + 0.33 \times 110 + 0.50 \times 90 = 103.4$

S1.3 Exponential Smoothing

$$F_{t+1} = \alpha D_t + (1-\alpha)F_t$$

$$F_{t+1} = \alpha D_t + (1-\alpha)F_t = \alpha D_t + (1-\alpha)[\alpha D_{t-1} + (1-\alpha)F_{t-1}] = \cdots$$

$$= \alpha D_t + \alpha(1-\alpha)D_{t-1} + \alpha(1-\alpha)^2 D_{t-2} + \cdots + \alpha(1-\alpha)^{t-2}D_2 + (1-\alpha)^{t-1}F_2$$

$$= \alpha D_t + \alpha(1-\alpha)D_{t-1} + \alpha(1-\alpha)^2 D_{t-2} + \cdots + \alpha(1-\alpha)^{t-2}D_2 + (1-\alpha)^{t-1}D_1$$

where

F_{t+1} : forecast for next period

D_t: actual value for present period

F_t: previously determined forecast for present period

α: weighting factor, smoothing constant

E.g.

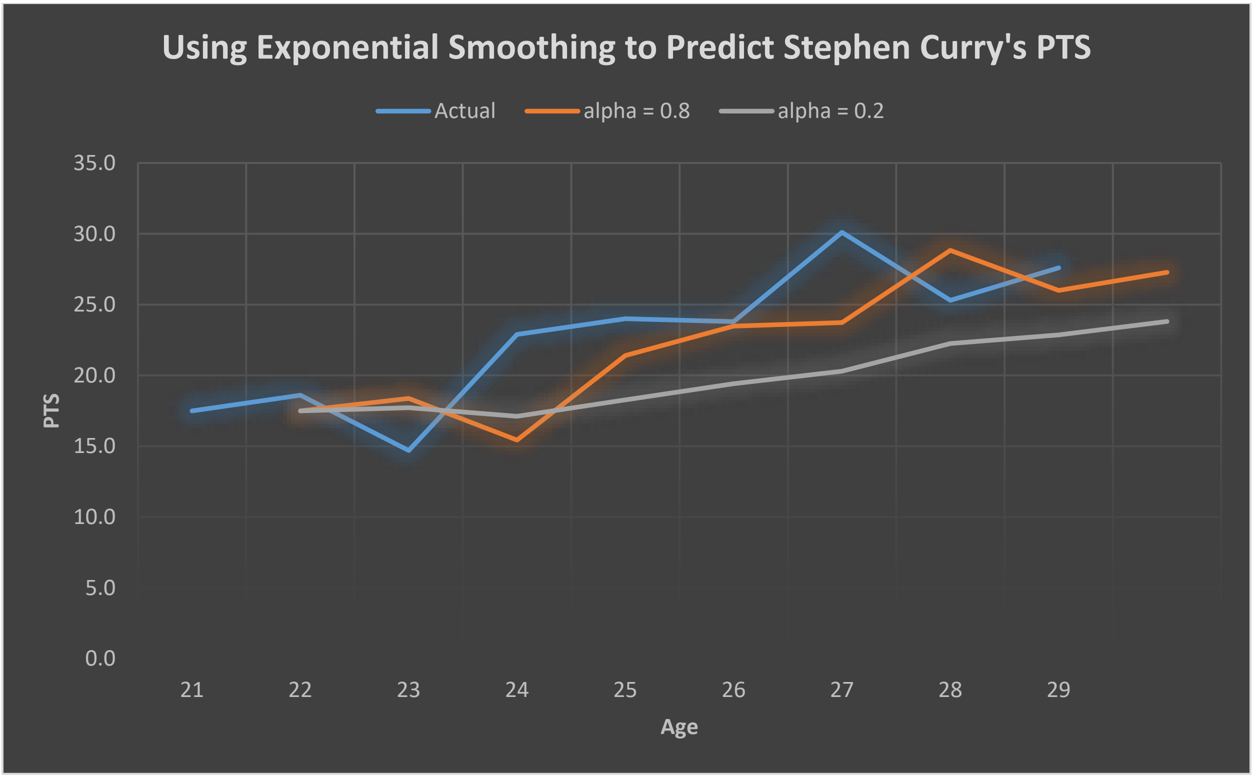
Age	Actual	alpha = 0.8	alpha = 0.2
21	17.5		
22	18.6	17.5	17.5
23	14.7	18.4	17.7
24	22.9	15.4	17.1
25	24.0	21.4	18.3
26	23.8	23.5	19.4
27	30.1	23.7	20.3
28	25.3	28.8	22.3
29	27.6	26.0	22.9
30		27.3	23.8

Remark:

For exponential smoothing, we consider the second period of forecasting value equal to the actual value of the first period.

If $\alpha = 0.8$, then

$$F_{30} = 0.8 \times D_{29} + (1 - 0.8) \times F_{29} = 0.8 \times 27.6 + 0.2 \times 26.0 = 27.3$$



<u>Compare moving average and exponential smoothing:</u>

Exponential smoothing takes into account all past data, whereas moving average only takes into account k past data points.

Whereas in the simple moving average the past observations are weighted equally (or weighted moving average, the past observations are weighted in several arbitrary numbers), exponential functions are used to assign exponentially decreasing weights over time. - Rule of thumb

From experience, exponential smoothing works better.

Session 2. Two-Variable Forecasting

S2.1 Linear Trend Line

$y = a + bx$

where

a: intercept (at period 0); b: slope of the line; x: the time period; y: forecast for demand for period x

$b = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$

$a = \bar{y} - b\bar{x}$

where

n: number of periods; $\bar{x}$: mean of the *x* values; $\bar{y}$: mean of the *y* values

Example

	x(period)	y(demand)	xy	x^2
	1	37	37	1
	2	40	80	4
	3	41	123	9
	4	37	148	16
	5	45	225	25
	6	50	300	36
	7	43	301	49
	8	47	376	64
	9	56	504	81
	10	52	520	100
	11	55	605	121
	12	54	648	144
sum	78 ($\sum x$)	557 ($\sum y$)	3867 ($\sum xy$)	650 ($\sum x^2$)
average	6.5($\bar{x}$)	46.42 ($\bar{y}$)		

$$\mathrm{b} = \frac{3867 - 12 \times 6.5 \times 46.42}{650 - 12 \times 6.5^2} = 1.72$$

$$a = 46.42 - 1.72 \times 6.5 = 35.2$$

Linear trend line:

$$y = 35.2 + 1.72x$$

Forecast for period 13

$$y = 35.2 + 1.72 \times 13 = 57.56$$

Remark:

If we add a constant to all the “x”, suppose the period series become: “2001, 2002, … , 2012”, will the predicted value of 2013 different from the predicted value of 13?

Regression

The term "regression" was coined by Francis Galton in the 19th century to describe a biological phenomenon. The phenomenon was that the heights of descendants of tall ancestors tend to regress down towards a normal average (a phenomenon also known as regression toward the mean).

The “x” does not have to be period.

In a regression model
“x”: independent variable (predictor variable, explanatory variable)

“y”: dependent variable (response variable, explained variable)

An example in excel

number of reviews	sales
322	142
880	286
527	223
829	255
564	210
697	251
531	202
356	178
462	172
331	171

Advantage of Regression for forecasting:

1 Very flexible: give any input “x”, we will have a forecasting “y” (it is not appropriate to input an “x” significantly out of the original range of “x”)

2 Using all the information contained in the data

3 No missing forecasting values

Disadvantage:

1 Vulnerable to extreme point (outliers)

2 Strong assumption: what if not linear?

S2.2 Evaluation of Forecasting

Effect of Smoothing Constant

$$0 \leq \alpha \leq 1$$

If $\alpha = 0.2$, then $F_{t+1} = 0.2D_t + 0.8F_t$

If $\alpha = 0$, then $F_{t+1} = 0 \times D_t + 1 \times F_t = F_t$, which means forecast does not reflect recent data

If $\alpha = 1$, then $F_{t+1} = 1 \times D_t + 0 \times F_t = F_t$, which means forecast based only on most recent data

How to choose the best α?

Evaluation tools for forecasting, i.e. how to evaluate the forecasting result?

A Special Type of Time Series Data

Trend: gradual, long-term up or down movement

When there is a trend, it means most recent data is more useful for forecasting. Therefore, we need to choose a short period moving average or a large α for exponential smoothing.

Mean Absolute Deviation (MAD)

$$MAD = \frac{\sum |D_t - F_t|}{n}$$

where

t: the period number

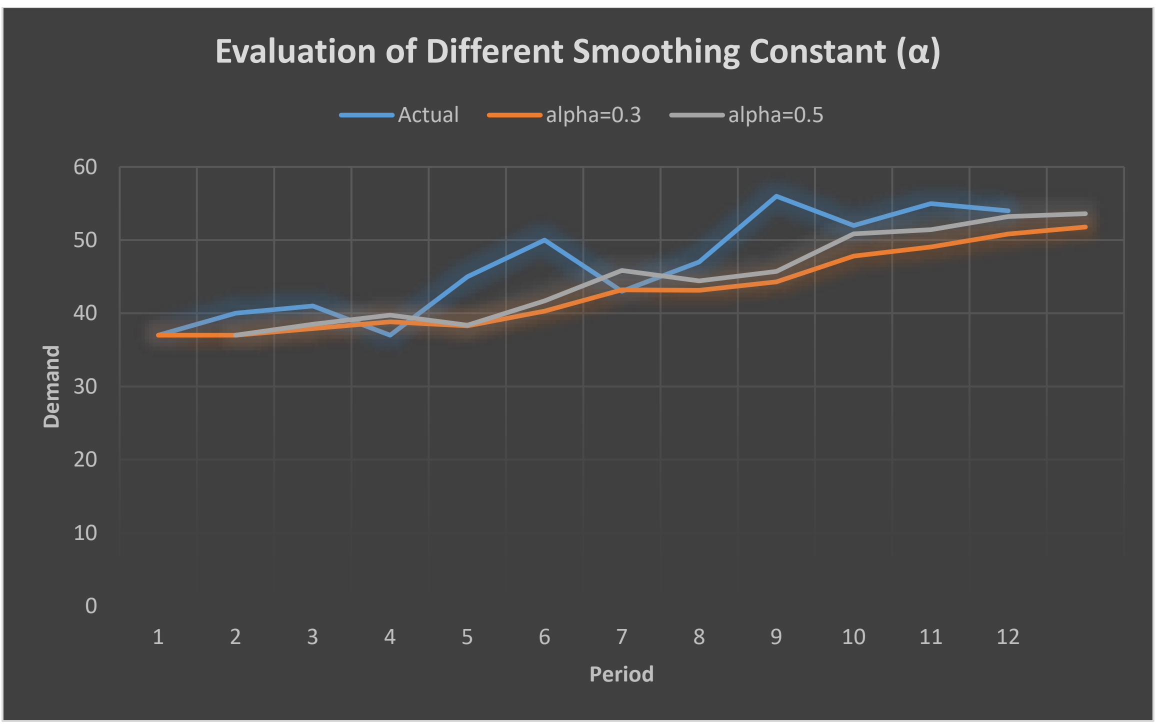
D_t: actual value in period t

F_t: the forecast for period t

n: the total number of periods

||: the absolute value

Period	D_t	F_t (α =0.3)	$(D_t - F_t)$	$\lvert D_t - F_t \rvert$	F_t (α =0.5)	$\lvert D_t - F_t \rvert$
1	37					
2	40	37.00	3.00	3.00	37.00	3.00
3	41	37.90	3.10	3.10	38.50	2.50
4	37	38.83	-1.83	1.83	39.75	2.75
5	45	38.28	6.72	6.72	38.38	6.63
6	50	40.29	9.69	9.69	41.69	8.31
7	43	43.20	-0.20	0.20	45.84	2.84
8	47	43.14	3.86	3.86	44.42	2.58
9	56	44.30	11.70	11.70	45.71	10.29
10	52	47.81	4.19	4.19	50.86	1.14
11	55	49.06	5.94	5.94	51.43	3.57
12	54	50.84	3.15	3.15	53.21	0.79
		51.79	49.31	53.39		44.40



Mean-Squared Error

$$MSE = \frac{\sum (D_t - F_t)^2}{n}$$

Period	D_t	F_t (α =0.3)	$(D_t - F_t)^2$	F_t (α =0.5)	$(D_t - F_t)^2$
1	37				
2	40	37.00	9.00	37.00	9.00
3	41	37.90	9.61	38.50	6.25
4	37	38.83	3.35	39.75	7.56
5	45	38.28	45.16	38.38	43.89
6	50	40.29	93.90	41.69	69.10
7	43	43.20	0.04	45.84	8.09
8	47	43.14	14.90	44.42	6.65
9	56	44.30	136.89	45.71	105.86
10	52	47.81	17.56	50.86	1.31
11	55	49.06	35.28	51.43	12.76
12	54	50.84	9.92	53.21	0.62
		51.79	375.61	53.61	271.09

$Coefficient\ of\ determination\ R^2$

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2} = 1 - \frac{\frac{\sum(y-\hat{y})^2}{n}}{\frac{\sum(y-\bar{y})^2}{n}} = 1 - \frac{MSE}{Var(Y)}$$

number of reviews (x)	sales (y)	$\hat{y}$	$(y-\hat{y})^2$
322	142	159.46	308.86
880	286	280.75	24.62
527	223	204.02	347.58
829	255	269.67	218.19
564	210	212.06	3.09
697	251	240.97	95.38
531	202	204.89	8.27
356	178	166.85	126.32
462	172	189.89	317.34
331	171	161.42	85.23
	$\sum(y-\bar{y})^2 = 18263.31$		$\sum(y-\bar{y})^2 = 1534.89$

$$R^2 = 1 - \frac{1534.89}{18263.31} = 0.92$$

Session 3. Seasonal Adjustment

Seasonal variations: Regular upward or downward movements in a time series that tie to recurring events.

E.g. coal, fuel oil during cold winter months

gold clubs or sunscreen in summer

$$\text{Seasonal Index} = \frac{Average\ Period\ Demand}{Total\ Average\ Quantity}$$

Step1: Compute each year total quantity, forecast next year's total quantity by the methods we know

Step2: Create a new column named average period quantity:

Compute the average period quantity for the past several years

Step3: Find the seasonal index

Step4: Use the forecasting total quantity in Step1 and seasonal index in Step3 to predict seasonal quantity in next year.

Demand of turkeys (Unit: 1000)

	Quarter				
Year	1	2	3	4	Total
1999	12.6	8.6	6.3	17.5	45.0
2000	14.1	10.3	7.5	18.2	50.1
2001	15.3	10.6	8.1	19.6	53.6
Total	42.0	29.5	21.9	55.3	148.7

$q_1 = \frac{D_1}{\sum D} = \frac{42}{148.7} = 0.28$

$q_2 = \frac{D_2}{\sum D} = \frac{29.5}{148.7} = 0.20$

$q_3 = \frac{D_3}{\sum D} = \frac{21.9}{148.7} = 0.15$

$q_4 = \frac{D_4}{\sum D} = \frac{55.3}{148.7} = 0.37$

	Quarter				
Year	1	2	3	4	Total
1999	12.6	8.6	6.3	17.5	45.0
2000	14.1	10.3	7.5	18.2	50.1
2001	15.3	10.6	8.1	19.6	53.6
Total	42.0	29.5	21.9	55.3	148.7
q_i	0.28	0.20	0.15	0.37	

First, we need calculate the total demand of 2002 using regression.

Year	Total
1999	45.0
2000	50.1
2001	53.6
2002	?

Year	Total
1	45.0
2	50.1
3	53.6
4	?

$y = 40.97 + 4.30x$

If $x = 4$, then $y = 40.97 + 4.30 \times 4 = 58.17$

Now calculate demand of each quarter of 2002:

$$\mathrm{F}_{2002}^{1} = q_1 \times F_{2002} = 0.28 \times 58.17 = 16.28$$

$$\mathrm{F}_{2002}^{2} = q_2 \times F_{2002} = 0.20 \times 58.17 = 11.63$$

$$\mathrm{F}_{2002}^{3} = q_3 \times F_{2002} = 0.15 \times 58.17 = 8.73$$

$$\mathrm{F}_{2002}^{4} = q_4 \times F_{2002} = 0.37 \times 58.17 = 21.53$$

Multiple Regression

1. Partial Derivative

2. x1 increases by one unit, y will increase/decrease by beta1 unit, holding other factors constant.

3. R-squared still valid. Prefer adjusted R-squared.

Extension Causal Inference

Correlation and Causality

Counterfactual: parallel world/universe

A claim, hypothesis, or other belief that is contrary to the facts.

This medicine is very effective, because after taking it, my headache is gone.

She gets a very good job offer, because she graduates from Harvard.

Counterfactual

Very important!

How should we utilize the huge amount of data?

We do get a lot of information from data analysis, but data can also mislead us because data itself cannot tell causality.

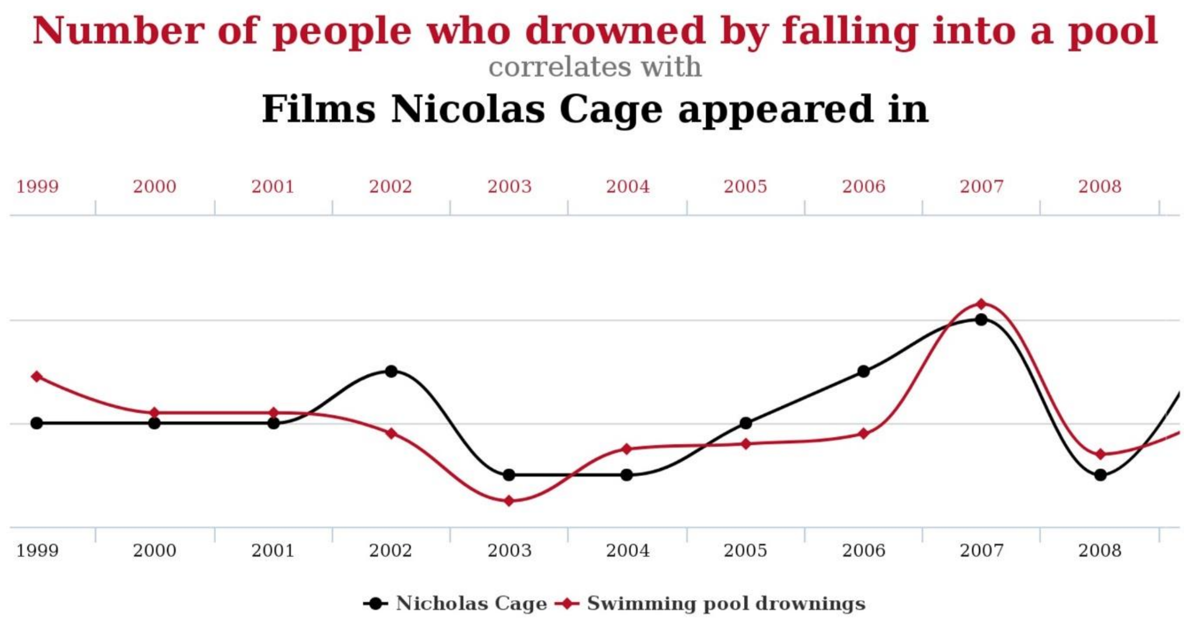
Causal Modeling with Linear Regression

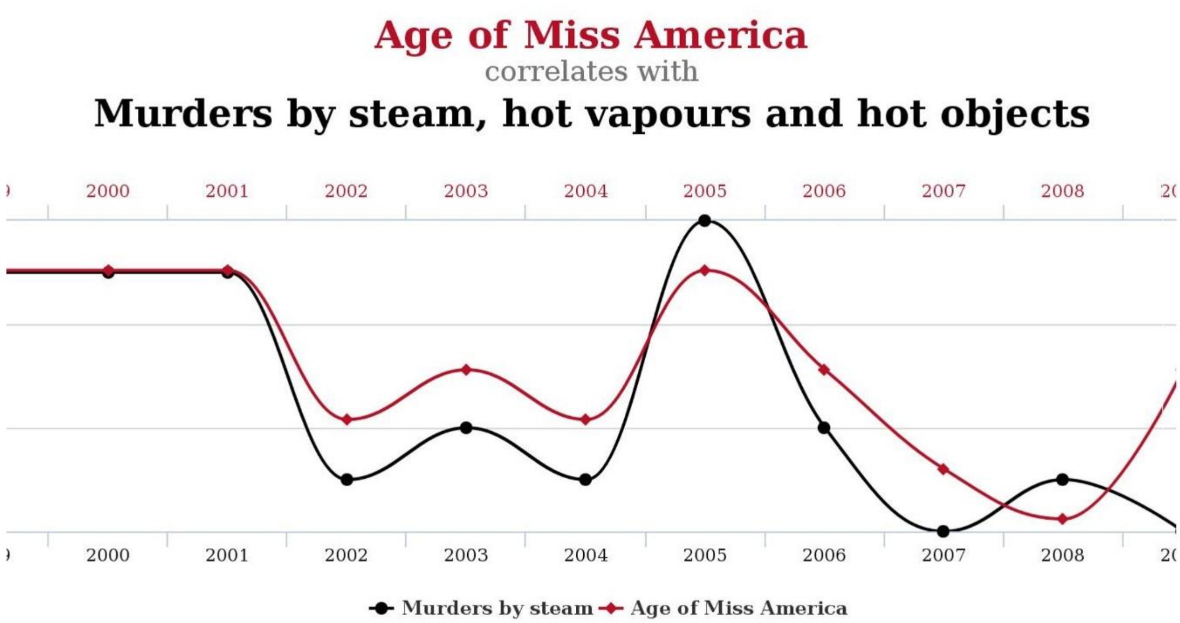
1. Regression is not likely to disclose a causal relationship.

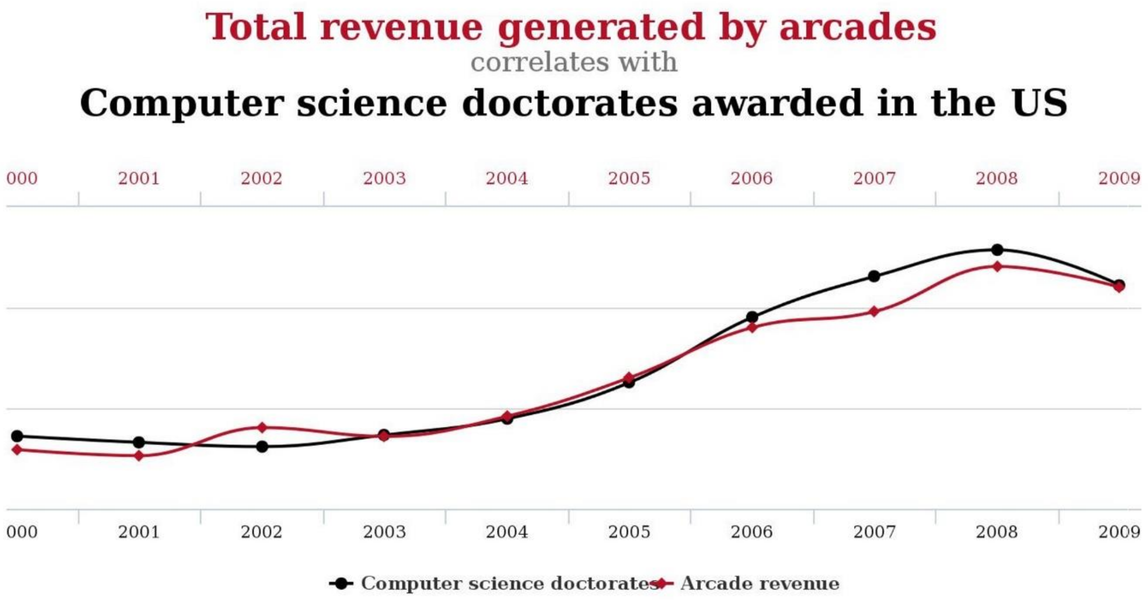
E.g.

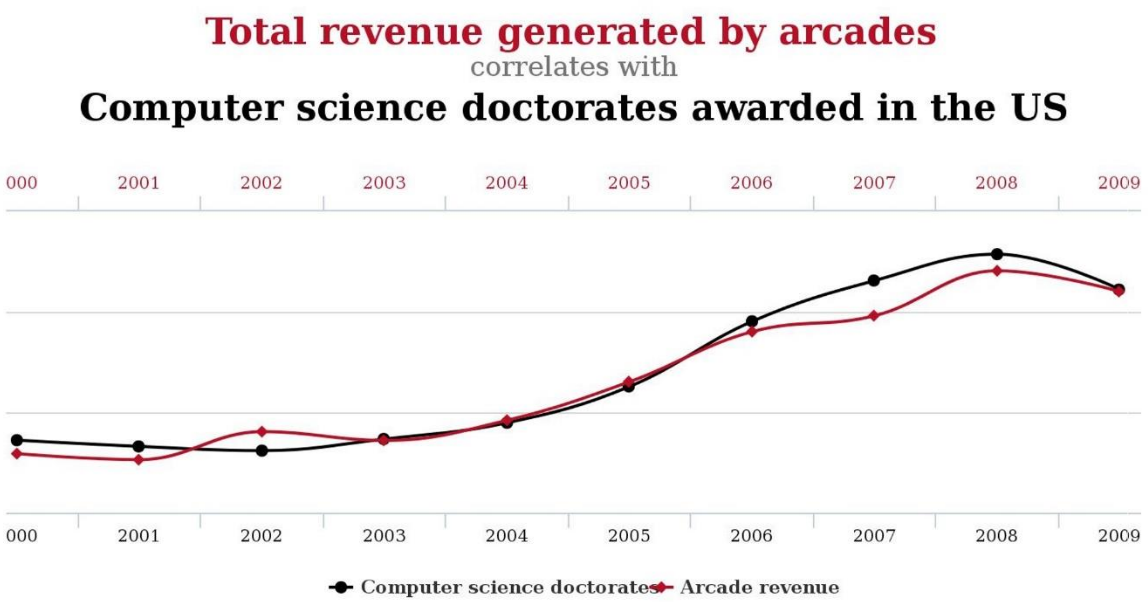
2. Some common mistakes about causality. As long as you know the magic of correlation, you can cheat people who do not understand the difference between causality and correlation. You want to make a big news.

Some examples with apparent logic mistakes:


Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in
1999 2000 2001 2002 2003 2004 2005 2006 2007 2008
1999 2000 2001 2002 2003 2004 2005 2006 2007 2008
Nicholas Cage
Swimming pool drownings


Age of Miss America
correlates with
Murders by steam, hot vapours and hot objects
2000 2001 2002 2003 2004 2005 2006 2007 2008
2000 2001 2002 2003 2004 2005 2006 2007 2008
Murders by steam
Age of Miss America





How can we use data to identify causality?

By experiment!

Question: suppose you are considering whether to get an MBA, and you are wondering whether MBA can increase your salary.

You did a survey, and find the average salary of MBA is lower than non-MBA, so you decide not to take an MBA.

What’s wrong here?

This medicine is very effective, because after taking it, my headache is gone.

She gets a very good job offer, because she graduates from Harvard.

Can education increase income?