Describing Associations, Covariance, Correlation, and Causality

Lecture 4

Reading: Chapter 6 &
SW11 (Readings page in Quercus)

1

Interest Rates and Inflation

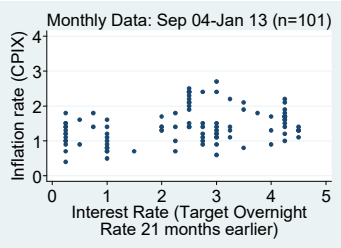
"At the heart of Canada's monetary policy framework is the inflation-control target. The target for inflation is the 2 per cent midpoint of a control range of 1 to 3 per cent. Inflation is measured as the year-over-year rate of increase in the total consumer price index (CPI). The Bank also monitors a set of "core" inflation measures, including the CPIX which strips out eight of the most volatile CPI components. The Bank carries out monetary policy through changes in its policy interest rate—the Target for the Overnight Rate. Monetary policy actions (changes in the policy rate) take time—usually between six and eight quarters—to work their way through the economy and to have their full effect on inflation."

http://www.bankofcanada.ca/wp-content/uploads/2010/11/monetary_policy.pdf

2

date	CPIX	lag_tr
2013-01	0.5	1
2012-12	0.7	1
2012-11	0.8	1
2012-10	0.9	1
2012-09	0.7	1
2012-08	1	1
2012-07	1.2	1
2012-06	1.6	1
2012-05	1.4	1
2012-04	1.8	0.75
2012-03	1.4	0.75
2012-02	1.6	0.5
2012-01	1.5	0.25
…		
2004-09	1	2.75

Data & Scatter Diagram



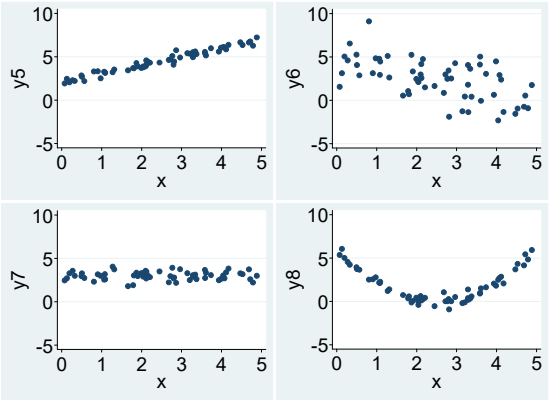
Which kind of data are these?

3

Qualitative Assessments

- Use scatter diagram to <u>qualitatively</u> assess relationship between two variables:
  - positive linear relationship
  - negative linear relationship
  - non-linear relationship
  - no relationship
  - strong relationship
  - weak relationship

4



Which qualitative assessments should we make?

5

Which qualitative assessments should we make?

The Economist, Feb. 8, 2018, "The great experiment: What will result from America's strangely timed fiscal stimulus?" https://www.economist.com/united-states/2018/02/08/what-will-result-from-americas-strangely-timed-fiscal-stimulus

6

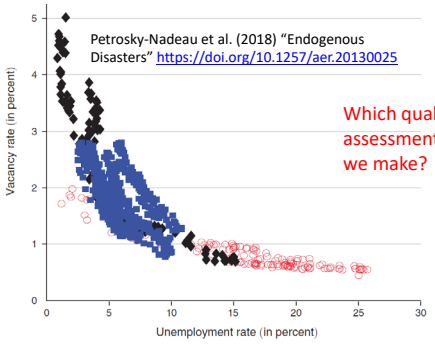


FIGURE 1. THE US BEVERIDGE CURVE, APRIL 1929–DECEMBER 2013

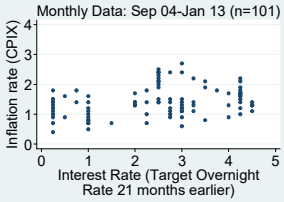
Notes: The April 1929–December 1939 observations are in red circles, the January 1940–December 1950 observations are in black diamonds, and the remaining observations are in blue squares.
Source: The unemployment and vacancy rate series from Petrosky-Nadeau and Zhang (2013)

7

Qualitative then Quantitative

- Strength depends on scatter & slope
- Statistics quantify strength
- For linear cases: covariance, correlation, R^2, OLS (slope)

Which qualitative assessments should we make? Why scattered?



8

Covariance

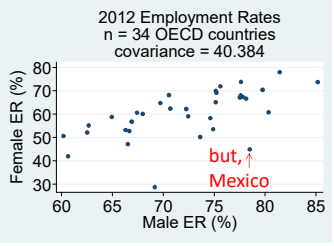
- Covariance: How two variables vary with respect to each other
  - co: with, together, joint
  - variance: vary about mean
- $\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)}{N}$
- $s_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{n-1}$

Units of measurement?

- Zero covariance: no linear relationship
- Positive covariance: when X big Y tends to be big; when X small Y tends to be small
- Negative covariance: when X is big Y tends to be small & v.v.

9

Formula and Intuition



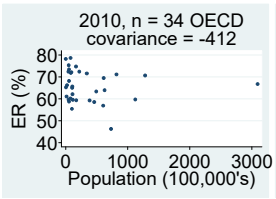
$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{n-1}$$

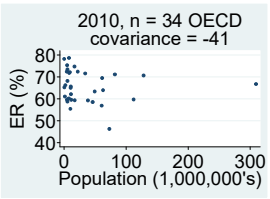
Approximate mean of X? Y?

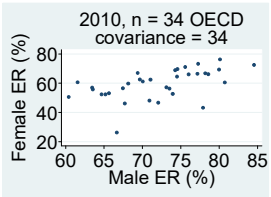
If x_i and y_i are both above average (e.g. Iceland with ER of 78% for females & 81% for males)?

If x_i and y_i are both below average (e.g. Greece with ER of 42% for females & 61% for males)?

10







Which is strongest relationship?

Covariance can only indicate the direction of a linear relationship: nothing about strength

11

Coefficient of Correlation

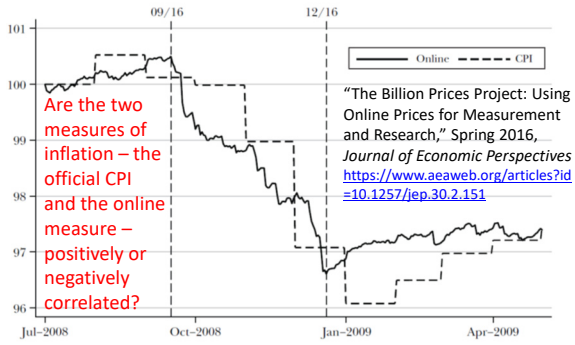
Parameter "rho": $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ Statistic: $r = \frac{s_{xy}}{s_x s_y}$

- Measures strength of a *linear relationship* between *two* variables: values from -1 to 1
  - What is σ_{xy}? What is s_x? What about sign?
  - What are the units of measurement?
  - Value near -1→ strong neg. linear relationship
  - Value near 1→ strong pos. linear relationship
  - Value near 0→ no linear relationship

12

Figure 3

US Consumer Price Index around the Bankruptcy of Lehman Brothers

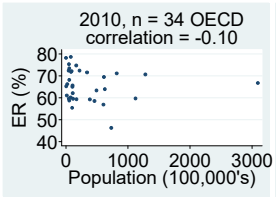


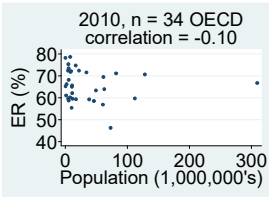
Are the two measures of inflation – the official CPI and the online measure – positively or negatively correlated?

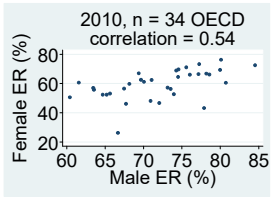
"The Billion Prices Project: Using Online Prices for Measurement and Research," Spring 2016, *Journal of Economic Perspectives* https://www.aeaweb.org/articles?id=10.1257/jep.30.2.151

Source: Authors using online price index computed by PriceStats and the Consumer Price Index from the US Bureau of Labor Statistics.

13

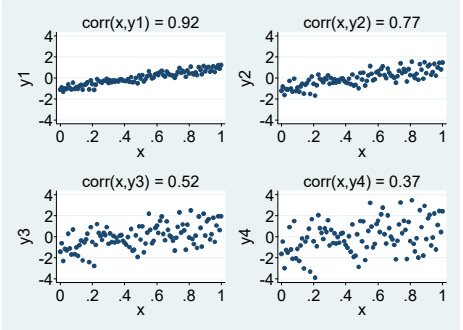






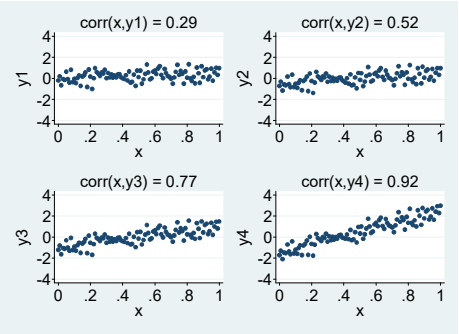
Which is strongest relationship?

14



Which concept do these graphs illustrate?

15

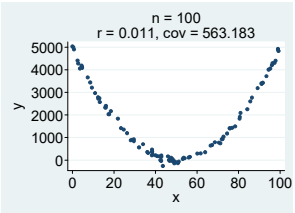


Which concept do these graphs illustrate?

16

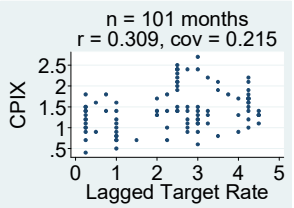
Correlation: Use *Only* to Measure Strength of *Linear* Relationships

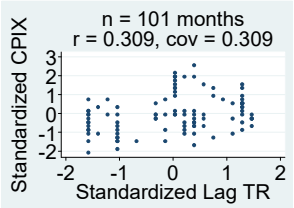
- Obtain nonsense if use on non-linear relationships
- No relationship versus no linear relationship: not the same thing
- "Association" versus "Correlation": not the same thing



17

Linking Lecture & Textbook





Once you standardize, the covariance is mathematically equal to the correlation.

18

```
correlate debt_pct_gdp_2009 emp_rate_2009
fem_emp_rate_2009 male_emp_rate_2009;

(obs=34)                    Correlation Matrix
             | deb~2009 emp~2009 fem~2009 mal~2009
-------------+------------------------------------
debt_pc~2009 |   1.0000
emp_rat~2009 |  -0.0666   1.0000
fem_emp~2009 |  -0.1009   0.9515   1.0000
male_em~2009 |   0.0092   0.8024   0.5806   1.0000

correlate debt_pct_gdp_2009 emp_rate_2009
fem_emp_rate_2009 male_emp_rate_2009, covariance;

(obs=34)                 Variance-Covariance Matrix
             | deb~2009 emp~2009 fem~2009 mal~2009
-------------+------------------------------------
debt_pc~2009 |  1368.71
emp_rat~2009 | -18.7978  58.2477
fem_emp~2009 | -41.0458  79.8299  120.838
male_em~2009 |  2.00417  35.9529  37.4713  34.4689
```

What is s.d. of debt %?

Source: OECD website, n = 34 OECD members


```
correlate debt_pct_gdp_2009 emp_rate_2009 fem_emp_rate_2009
male_emp_rate_2009 if country~="Japan"

(obs=33)
             | deb~2009 emp~2009 fem~2009 mal~2009
-------------+------------------------------------
debt_pc~2009 |   1.0000
emp_rat~2009 |  -0.1690   1.0000
fem_emp~2009 |  -0.1352   0.9559   1.0000
male_em~2009 |  -0.1937   0.8062   0.5975   1.0000

correlate debt_pct_gdp_2009 emp_rate_2009 fem_emp_rate_2009
male_emp_rate_2009 if country~="Japan", covariance;

(obs=33)
             | deb~2009 emp~2009 fem~2009 mal~2009
-------------+------------------------------------
debt_pc~2009 |  856.244
emp_rat~2009 |  -38.133  59.4349
fem_emp~2009 | -44.1728  82.2623  124.609
male_em~2009 | -32.7324  35.9015  38.5267   33.365
```

Why is 856.244 so much smaller?

Source: OECD website, n = 34 OECD members

Research Question

- Research question: Inquires about the causal relationship among variables
- Example: What is the effect of lecture attendance on learning?

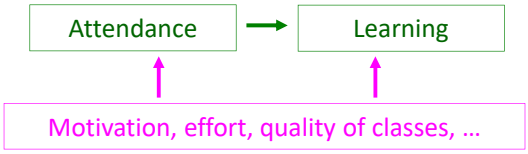
  - X variable ("explanatory"): Attendance
  - Y variable ("dependent"): Learning
  - How big is the effect? (skinny or thick arrow?)

Observational Data

- Observational Data: What has actually happened to agents (people, countries, firms, etc) where all variables are likely affected by choices/behaviors of agents and unobserved variables that affect *both* the dependent and independent variable
  - Unobserved variables: Not in your data and affect *both* your x and y variable
    - aka lurking, confounding, or omitted variables

22

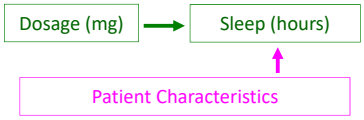
Observational Data



23

Experimental Data

- Experimental Data: Data collected in an experimental setting where the values of the x variable are set by the researcher
  - Researchers usually randomly set values for explanatory variable(s) and see reactions



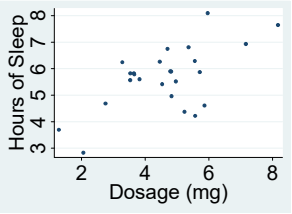
What is key difference from observational data?

24

Experimental Data: Drug Trial

i	Dosage (mg) x_i	Hours of Sleep y_i
1	5.9	4.6
2	3.5	5.8
3	7.2	6.9
4	3.6	5.8
…	…	…
25	8.2	7.6

X-bar = 4.61 Y-bar = 5.66
s_x = 1.51 s_y = 1.18
s_{xy} = 1.09



How do patient characteristics factor into this graph?

Can we infer causality?

25

Exogenous, Endogenous & Endogeneity Bias

- Exogenous: X variable not associated with factors that also affect Y (no lurking variables)
- Endogenous: X variable is associated with factors that also affect Y (lurking variables)
  - Endogeneity bias: Observed correlation driven by unobserved variables
    - Correlation either overstates or understates any truly causal relationship between the variables
    - Spurious correlation: false/misleading correlation

26

"Leveling Up: Early Results from a Randomized Evaluation of Post-Secondary Aid"

Abstract: Does financial aid increase college attendance and completion? Selection bias and the high implicit tax rates imposed by overlapping aid programs make this question difficult to answer. This paper reports initial findings from a randomized evaluation of a large privately-funded scholarship program for applicants to Nebraska's public colleges and universities. Our research design answers the challenges of aid evaluation with random assignment of aid offers and a strong first stage for aid received: randomly assigned aid offers increased aid received markedly. This in turn appears to have boosted enrollment and persistence, while also shifting many applicants from two- to four-year schools. *cont'd next slide…*

Research question? Observational or experimental data? Bias?

Source: Angrist et al (2014) http://www.nber.org/papers/w20800.pdf

Note: To access NBER papers, use a U of T computer.

27

"Leveling Up: Early Results from a Randomized Evaluation of Post-Secondary Aid"

Abstract cont'd: Awards offered to nonwhite applicants, to those with relatively low academic achievement, and to applicants who targeted less-selective four-year programs (as measured by admissions rates) generated the largest gains in enrollment and persistence, while effects were much smaller for applicants predicted to have stronger post-secondary outcomes in the absence of treatment. Thus, awards enabled groups with historically-low college attendance to level up, largely equalizing enrollment and persistence rates with traditionally college-bound peers, particularly at four-year programs. Awards offered to prospective community college students had little effect on college enrollment or the type of college attended.

This part discusses *interaction effects*, which we will study in Chapter 21/Lecture 22 (part of multiple regression analysis)

28

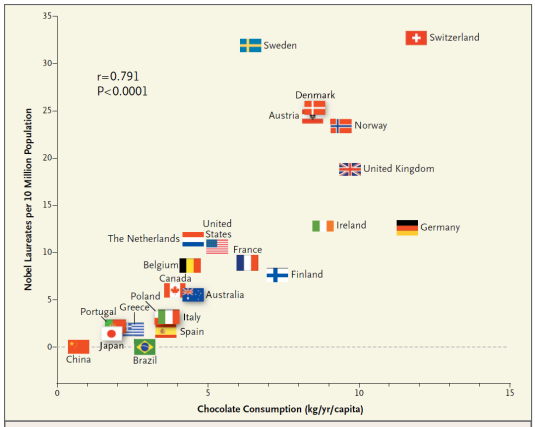


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

29