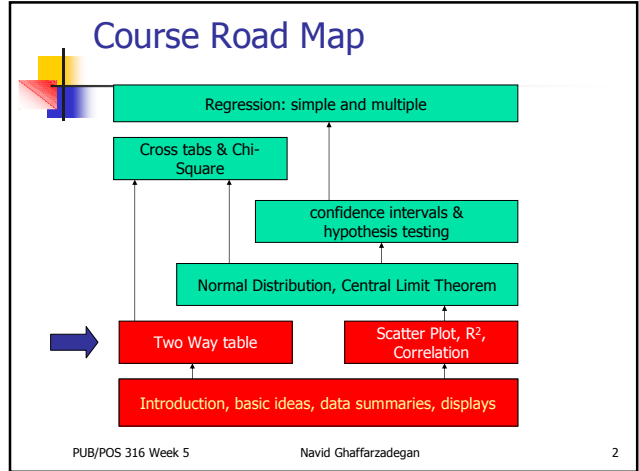


**PUB – POS 316**  
**Week 5**

---

**Two-way tables**

Navid Ghaffarzadegan  
 navidg@gmail.com  
 Last updated - sep 24, 09



### Agenda

- Two-way Table
  - Data in categories
  - Graphs
  
- Correlation vs Causation
  - Review of Correlation
  - Two examples

PUB/POS 316 Week 5      Navid Ghaffarzadegan      3

### Data in categories

- What is "data in categories?"
  - Two Variables are categorical:
    - X: Men or Women      Y: Yes or No

	Gender	
Frequent of binge drinker	Men	Women
Yes	1630	1684
No	5550	8232

- How should we analyze this data?
- Joint Distribution: dist. of the whole data
- Conditional distribution: dist. only for one condition (e.g., men)
- Marginal distribution: ratio of the row or column to total.

PUB/POS 316 Week 5      Navid Ghaffarzadegan      4

## Data in categories

Frequent of binge drinker	Gender		total
	Men	Women	
Yes	1630	1684	3314
No	5550	8232	13782
Total	7180	9916	17096

Frequent of binge drinker	Gender	
	Men	Women
Yes	0.095344	0.0985026
No	0.324637	0.4815161

e.g., "Men-Yes" cell,  
 $1630/17096=0.095$

Frequent of binge drinker	Gender	
	Men	Women
Yes	0.227019	0.1698265
No	0.772981	0.8301735

e.g., "Men-Yes" cell  
 $1630/7180=0.227$

- Marginal distribution: e.g., **marginal distribution** for gender:  $7180/17096$  and  $9916/17096$

PUB/POS 316 Week 5

Navid Ghaffarzadegan

5

## Data in categories

- What is "data in categories?"
  - Two Variables are categorical
    - Each Variable can have more than two categories.
      - Categories of age vs. Categories of education

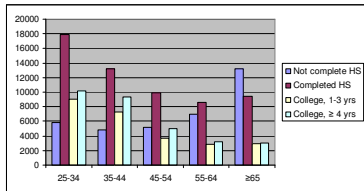
Data (1000's)	25-34	35-44	45-54	55-64	≥65	Total
Not complete HS	5836	4841	5230	7024	13183	36114
Completed HS	17889	13200	9860	8580	9412	58941
College, 1-3 yrs	9069	7309	3698	2793	2915	25784
College, ≥ 4 yrs	10174	9332	5008	3246	3018	30778
Total	42968	34682	23796	21643	28528	151617

PUB/POS 316 Week 5

Navid Ghaffarzadegan

6

## Graphs



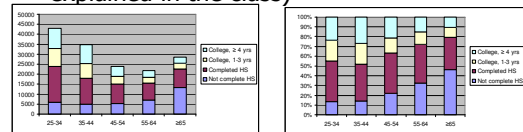
PUB/POS 316 Week 5

Navid Ghaffarzadegan

7

## Graphs

In a similar way draw other graphs. (As explained in the class)



■ Stacked bar charts

■ percentage bar charts

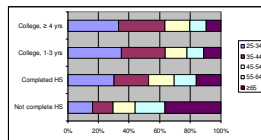
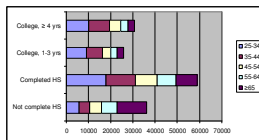
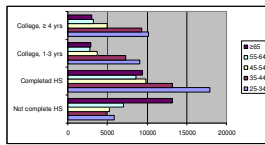
- In a similar way draw other graphs where  $x$  is NOT age BUT education.

PUB/POS 316 Week 5

Navid Ghaffarzadegan

8

## Graphs



- Many times new graphs give new insights.

PUB/POS 316 Week 5

Navid Ghaffarzadegan

9

## Graphs

What kind of graph can we draw?

- Bar charts
- Stacked bar charts
- Percents bar charts

PUB/POS 316 Week 5

Navid Ghaffarzadegan

10

## Agenda

### Two-way Table

- Data in categories
- Graphs

### Correlation vs Causation

- Review of Correlation
- Two examples

PUB/POS 316 Week 5

Navid Ghaffarzadegan

11

## Correlation vs Causation

- A question:
  - How should we help students to perform better?
- How do you measure performance?
- Method?
  - survey?
  - Experiment?
  - Secondary data?

PUB/POS 316 Week 5

Navid Ghaffarzadegan

12

## Correlation vs Causation – student performance

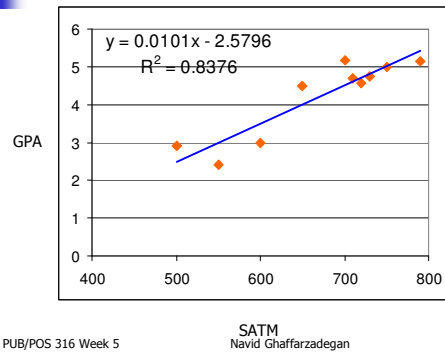
Observation	GPA	Male/Female	SATM	Hours of Study (per day)
1	2.4	M	550	1
2	2.91	F	500	1.5
3	4.5	F	650	4.6
4	5.16	M	700	5.8
5	5.14	F	790	7
6	4.75	F	730	8.2
7	4.58	M	720	9.4
8	4.7	M	710	5.5
9	5	M	750	6
10	3	F	600	4

PUB/POS 316 Week 5

Navid Ghaffarzadegan

13

## Correlation vs Causation – student performance



PUB/POS 316 Week 5

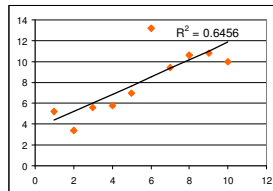
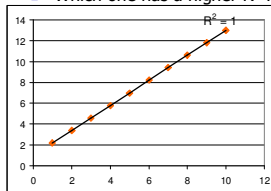
SATM  
Navid Ghaffarzadegan

14

## $R^2$ – A quick reminder

What was  $R^2$  ?

- Try to understand and keep in mind the technical definition of  $R^2$ . (the fraction of ...)... Remind yourself. Practice.
- What is the maximum and minimum possible value for  $R^2$  ?
  - $0 \leq R^2 \leq 1$
- Which one has a higher  $R^2$  ?



- What is correlation ( $r$ )? 1) take sqrt root of  $R^2$ , 2) take care of the sign, based on the slop of the line..

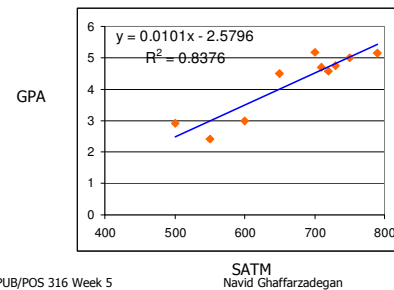
PUB/POS 316 Week 5

Navid Ghaffarzadegan

15

## Back to our problem..

What can we conclude? Can we say higher SATM causes higher GPA?



PUB/POS 316 Week 5

SATM  
Navid Ghaffarzadegan

16

**AS MORE ICE CREAM IS EATEN . . . THE CRIME RATE GOES UP**

- The local police chief in a town observes:
- As ice cream consumption increases, crime rates increases.
  - Even a scatter plot supports it!
  - Even R<sup>2</sup> is high!!
  - Even correlation is high!
    - So let's stop selling ice cream??
    - Let's arrest whoever eats more ice cream!
  - The outside temperature is what they both have in common.
    - Warm temperature → more windows left open
    - Warm temperature → more ice cream

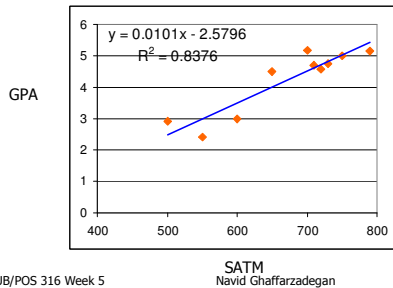
**AS MORE ICE CREAM IS EATEN . . . THE CRIME RATE GOES UP**

- The local police chief in a town observes:
- As ice cream consumption increases, crime rates increases.
  - Even a scatter plot supports it!
  - Even R<sup>2</sup> is high!!
  - Even correlation is high!
    - So let's stop selling ice cream??
    - Let's arrest whoever eats more ice cream!
  - The outside temperature is what they both have in common.
    - Warm temperature → more windows left open
    - Warm temperature → more ice cream

Correlation is NOT causation

**Back to our problem.. Student performance**

What can we conclude? Can we say higher SATM causes higher GPA?

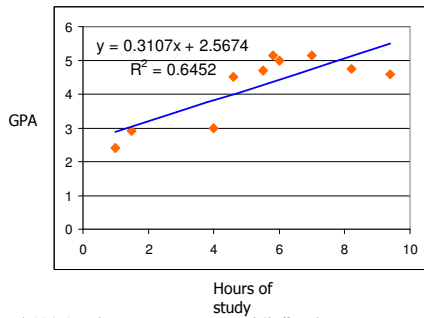


**Correlation vs Causation – student performance**

Observation	GPA	Male/Female	SATM	Hours of Study (per day)
1	2.4	M	550	1
2	2.91	F	500	1.5
3	4.5	F	650	4.6
4	5.16	M	700	5.8
5	5.14	F	790	7
6	4.75	F	730	8.2
7	4.58	M	720	9.4
8	4.7	M	710	5.5
9	5	M	750	6
10	3	F	600	4

## Correlation vs Causation – student performance

What can we conclude?



PUB/POS 316 Week 5

Navid Ghaffarzadegan

21

## Correlation vs Causation – student performance

two events that occur together are many times claimed to have a cause-and-effect relationship.

- Sometimes, there is NO causal relation between those events (higher SATM and higher GPA).
- They can be both influenced by a lurking variable. (temperature in the ice cream example)
- Sometimes, there is a causal relation (hours of study and higher GPA)
- So, we should be very careful in evaluating results of a scatter plot.
  - Correlation is not causation

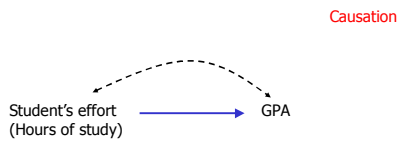
PUB/POS 316 Week 5

Navid Ghaffarzadegan

22

## Correlation vs Causation

Three types of association:



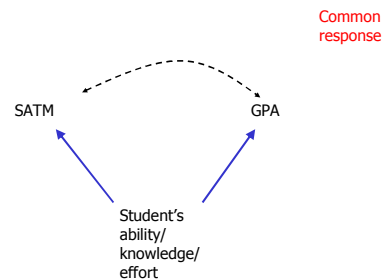
PUB/POS 316 Week 5

Navid Ghaffarzadegan

23

## Correlation vs Causation

Three types of association:



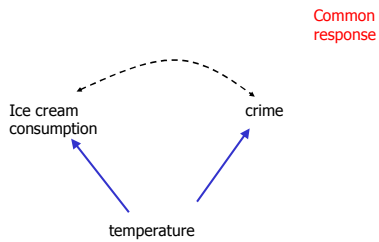
PUB/POS 316 Week 5

Navid Ghaffarzadegan

24

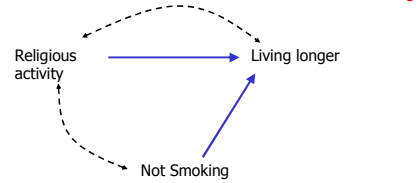
## Correlation vs Causation

Three types of association:



## Correlation vs Causation

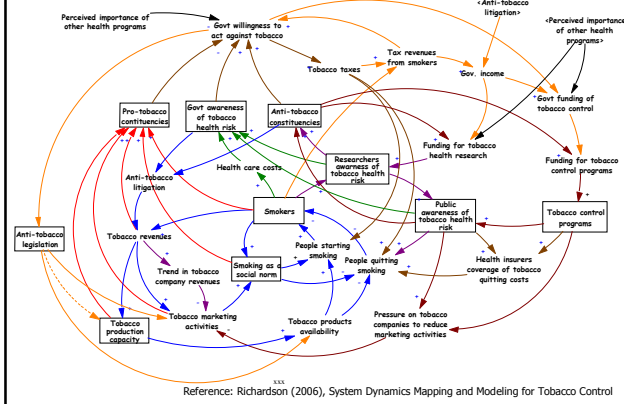
Three types of association: Confounding – Two variables effect cannot be distinguished.



How does the real world look like?

## The real world!

This is still a model of the real world ☺



## Take home message

That big loopi model is important and we should learn about it in future...

BUT NOT NOW.

Our today take home message:

***Correlation is NOT causation***

The Ice-cream story