

## Chapter 9: bivariate statistics

### 1) What is bivariate statistics?

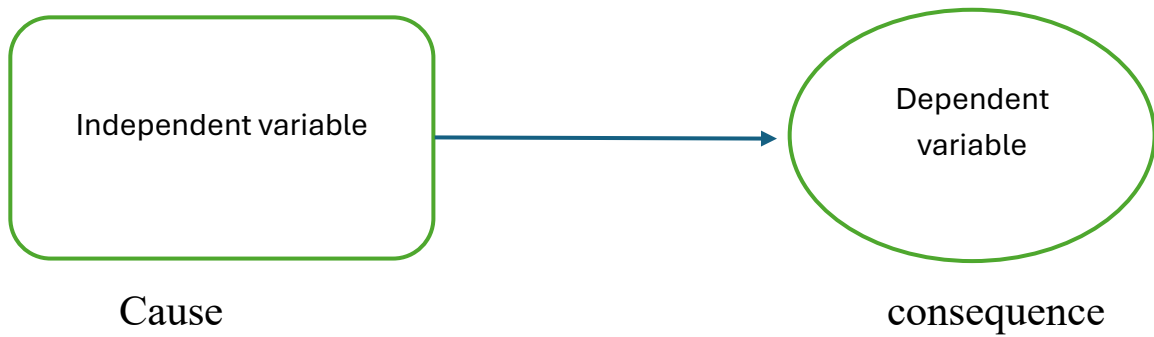
bivariate statistics refers to relationships between two variables, it handles two aspects of statistical analysis: the strength of the bivariate association, and the significance of the bivariate relationship (p value).

When two variables are both discrete variables, bivariate statistics use cross tab and chi-square (p value)

When independent variable is the dummy, while dependent variable is interval/ratio, bivariate statistics uses two-sample t test.

When the independent variable is discrete (groups > 2), and the dependent variable is interval/ratio, we use ANOVA (Analysis of Variance).

## 2) Conceptualizing relationship between two variables



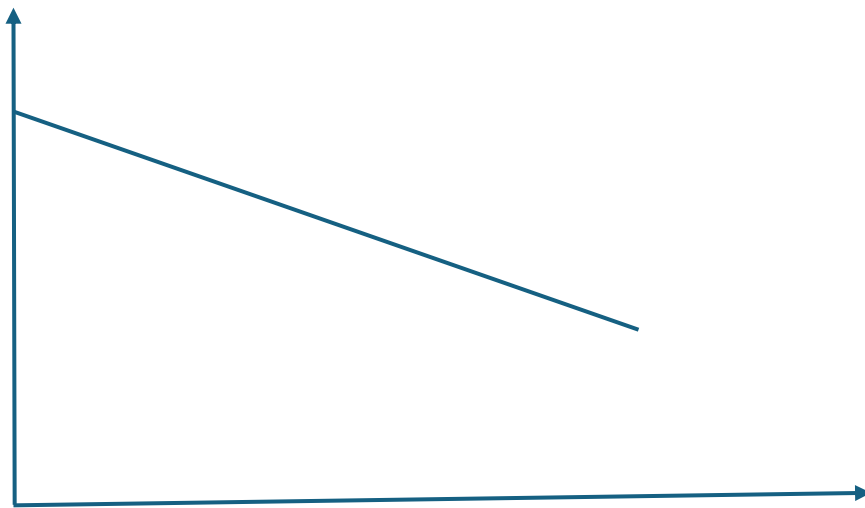
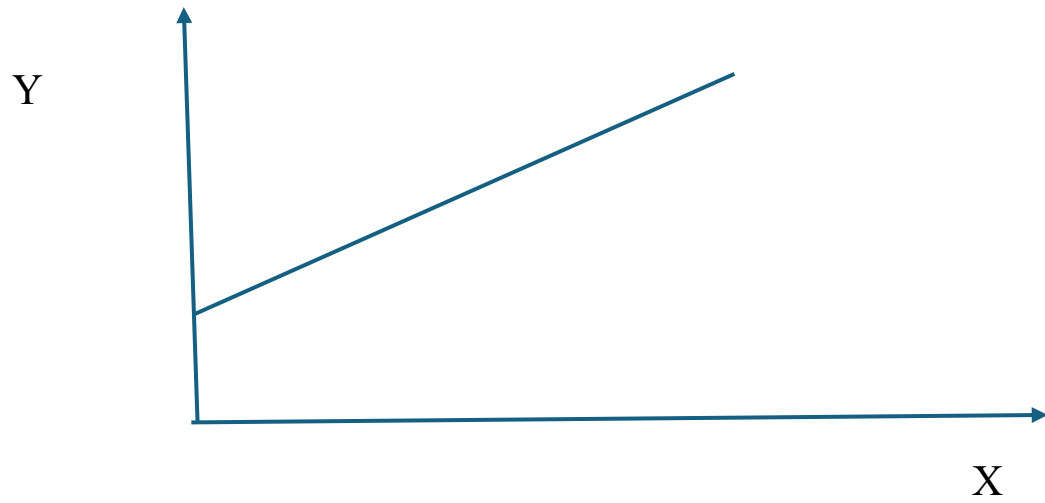
A) LSAT -> Law school admission

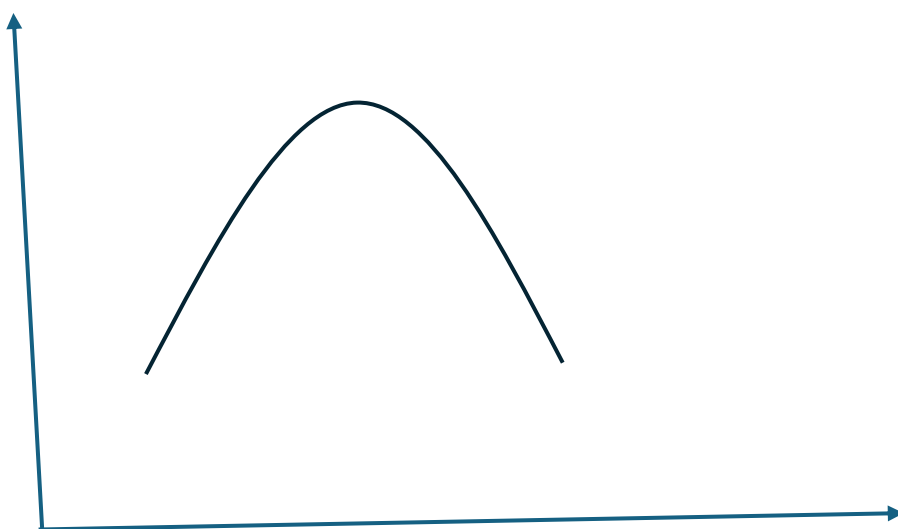
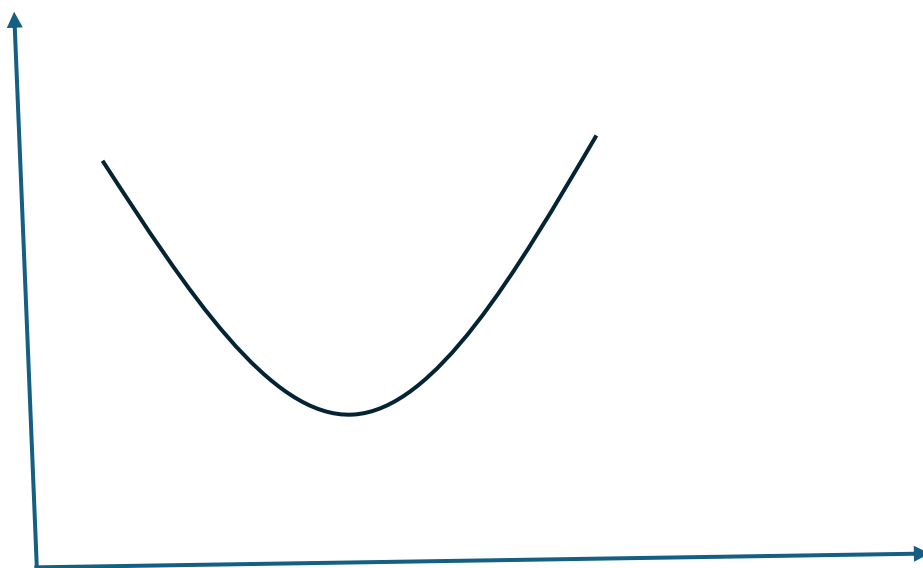
B) Race -> income

C) Gender -> income

D) Attendance -> grade

### 3) Direction of bivariate statistics





#### 4) Converting from raw data to crosstab

| Case # | Gender (m/w) | Type of cars (pickup truck/other types) |
|--------|--------------|-----------------------------------------|
| 1      | M            | Pickup                                  |
| 2      | W            | Other                                   |
| 3      | W            | Other                                   |
| 4      | M            | Pickup                                  |
| 5      | W            | Other                                   |
| 6      | W            | Other                                   |
| 7      | W            | Pickup                                  |
| 8      | M            | Other                                   |
| 9      | W            | Pickup                                  |
| 10     | W            | Other                                   |
| 11     | M            | Other                                   |

#### Crosstab between gender and type of cars

|                | Men        | Women        | RM     |
|----------------|------------|--------------|--------|
| Pickup         | 2<br>(50%) | 2<br>(28.6%) | 4      |
| Other vehicles | 2<br>(50%) | 5<br>(71.4%) | 7      |
| CM             | 4          | 7            | N = 11 |

- 1) Always arrange independent variable groups across different columns
- 2) Always arrange dependent variable groups across different row

- 3) RM (Row Margins) are the summation of the cell frequencies across different columns within a given row
- 4) CM (Column Margins) are the summation of the cell frequencies across different rows within a given column
- 5) Cell percentage is calculated with cell frequencies divided by its CM