Mathematics 231

Lecture 6

Liam O'Brien

Announcements

- Reading
  - Today — M&M 2.1 — 83-94
    M&M 2.2 — 101-104
  - Next class — M&M 2.3 — 108-121

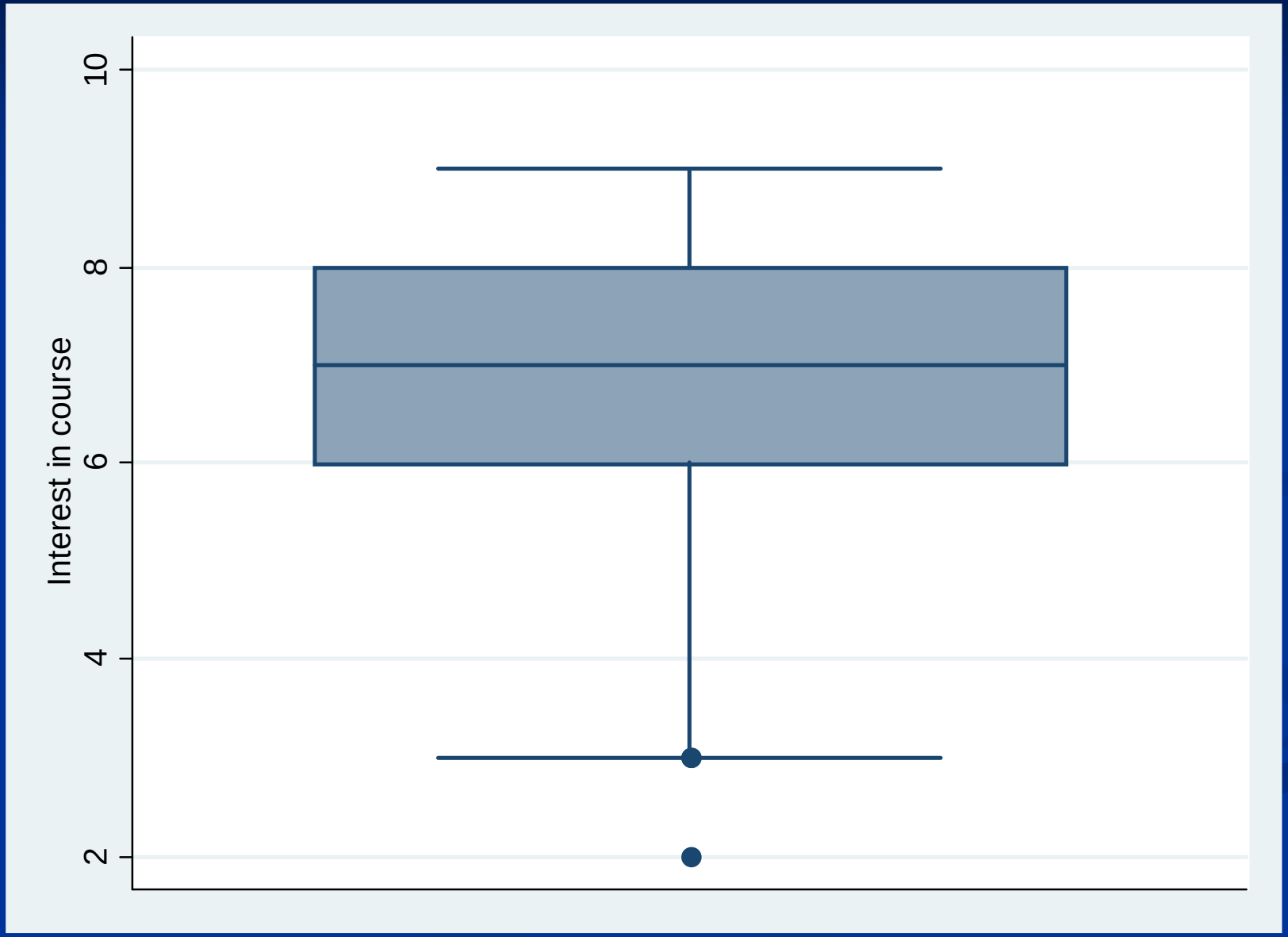
Methods to Check Normality

- There are many methods that have been developed to check normality.
- Some methods are more sensitive to departures from normality than others.
- Histograms and boxplots provide a graphical method for checking the basic shape and spread of a distribution.
- One more sensitive and common method for checking the assumption of normality is called a normal quantile plot.

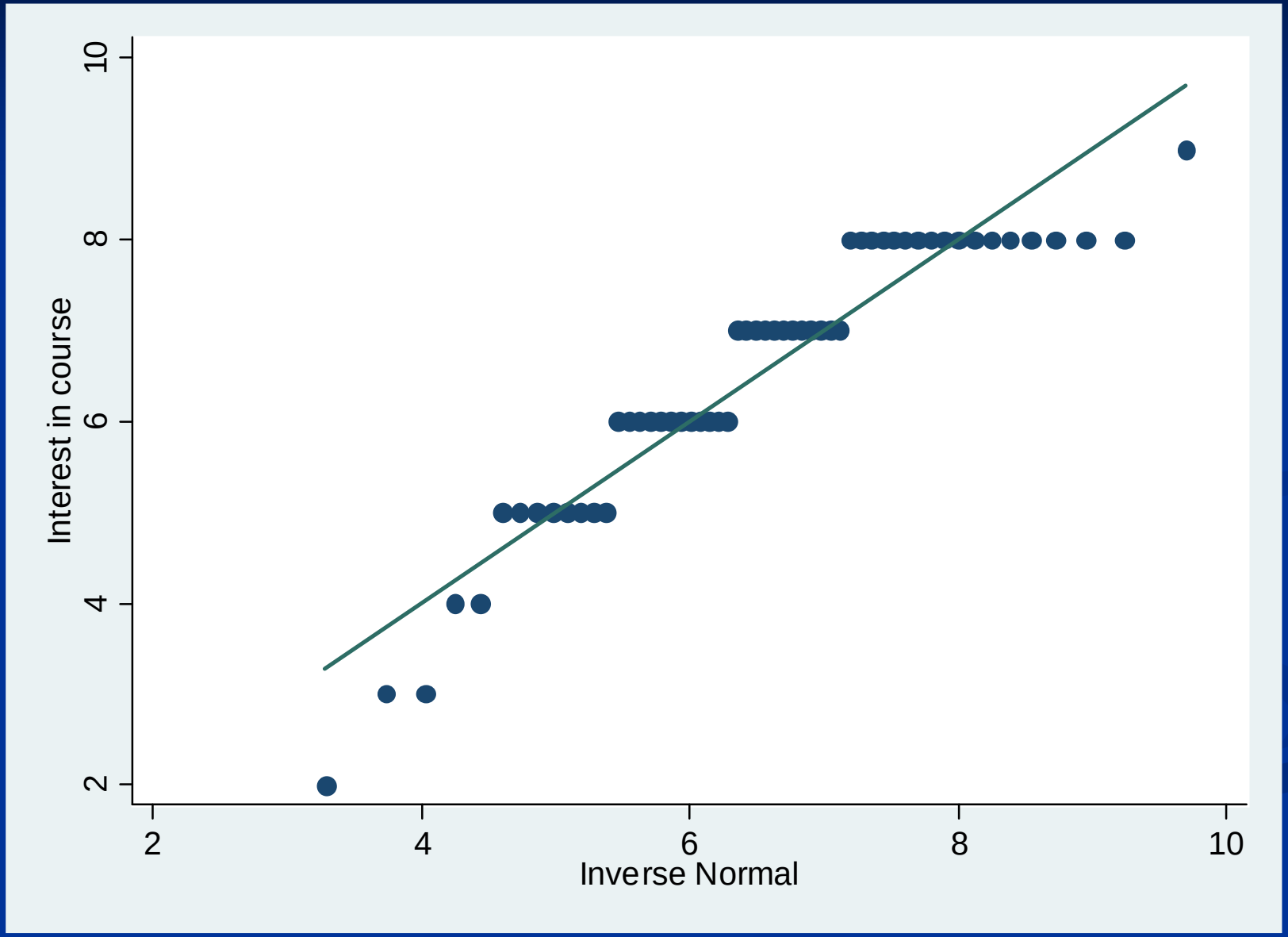
Normal Quantile Plots

- Cannot be constructed by hand, but Stata can do them.
- If you were to construct them by hand, you would first order the data.
- Then you would construct a z-score for each point.
- Plot the ordered x versus the z-scores.
- If the data are normal, then they will fall on a straight line.

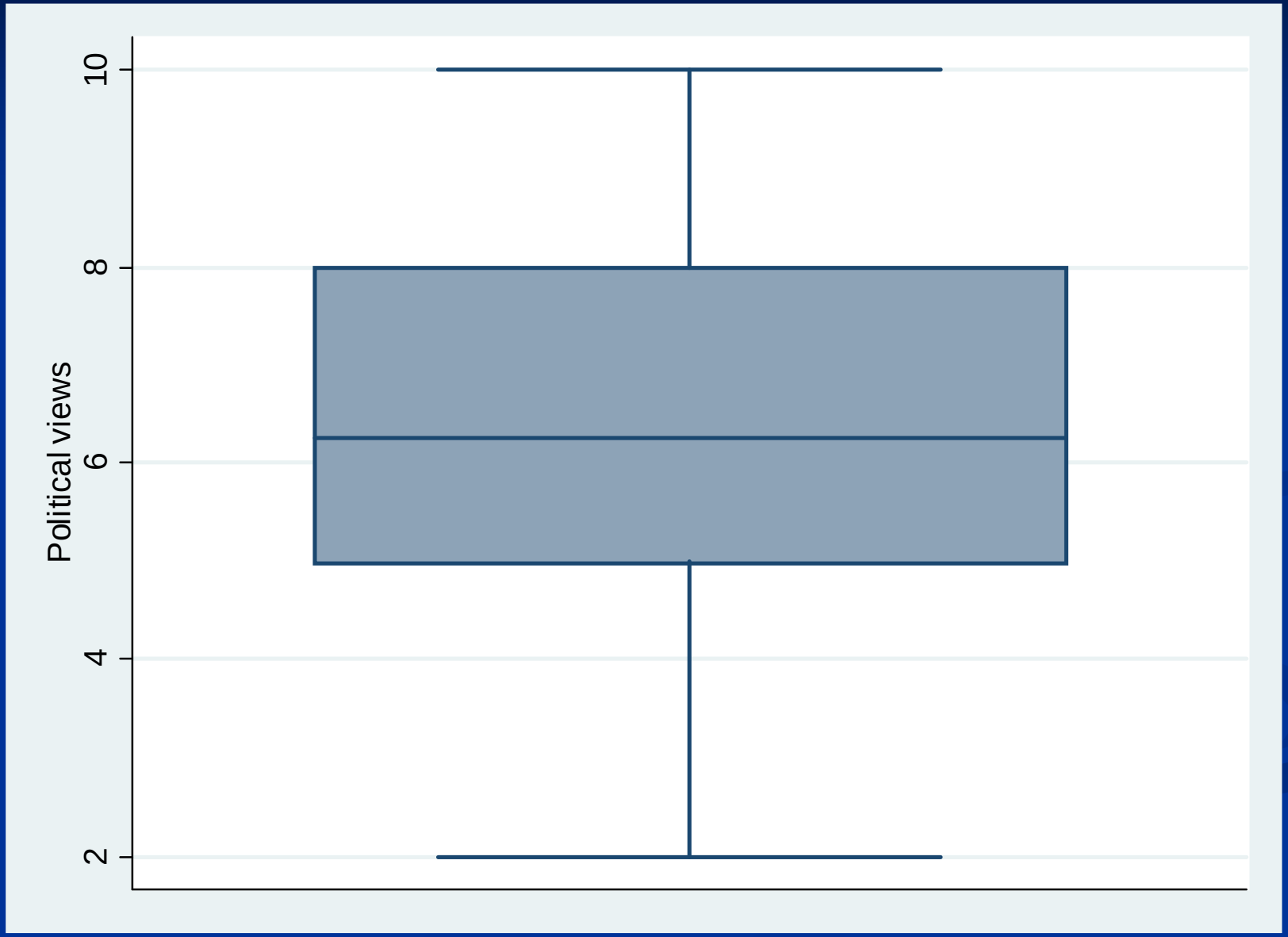
Interest Question



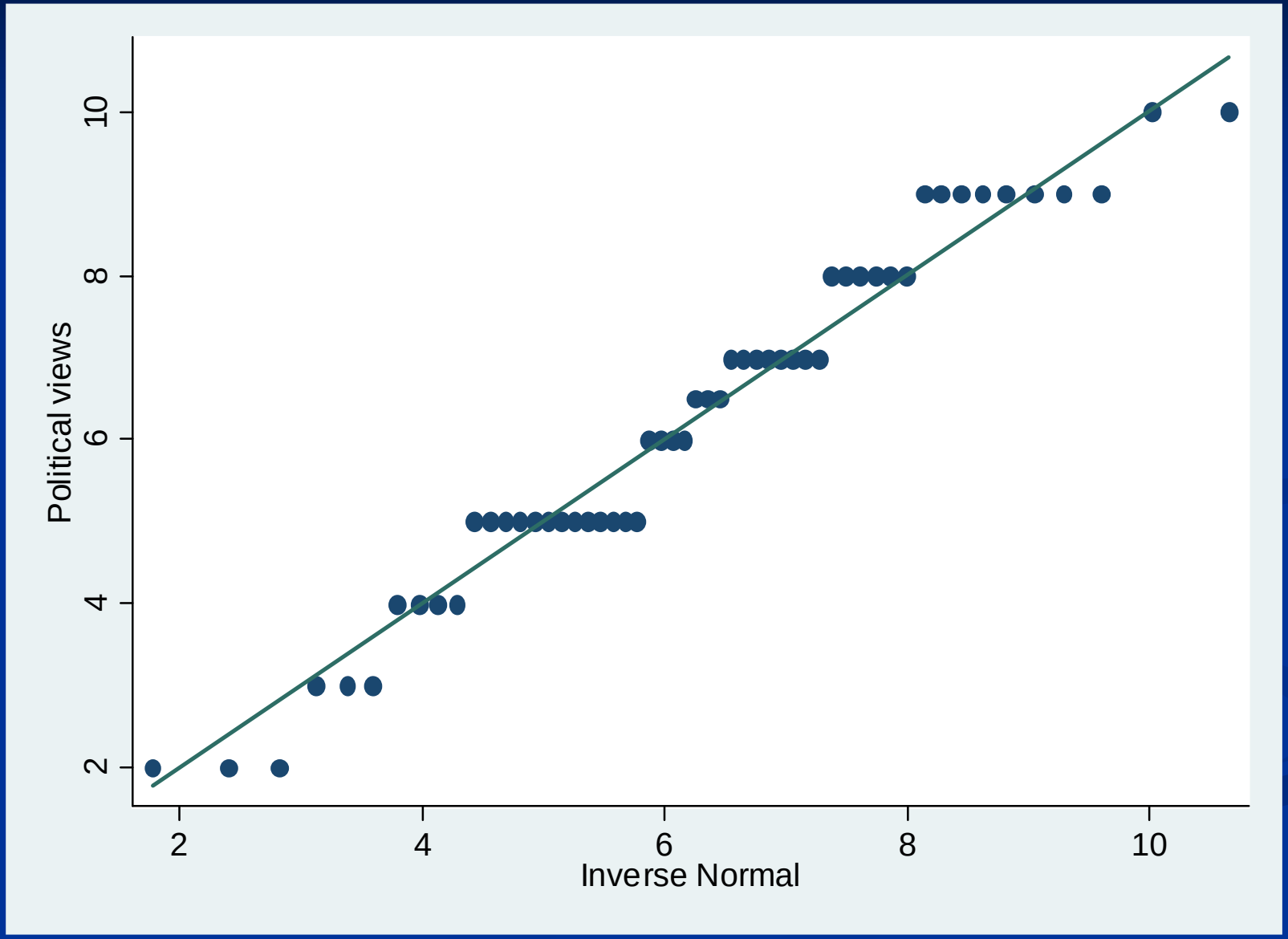
Normal Quantile Plot for Interest



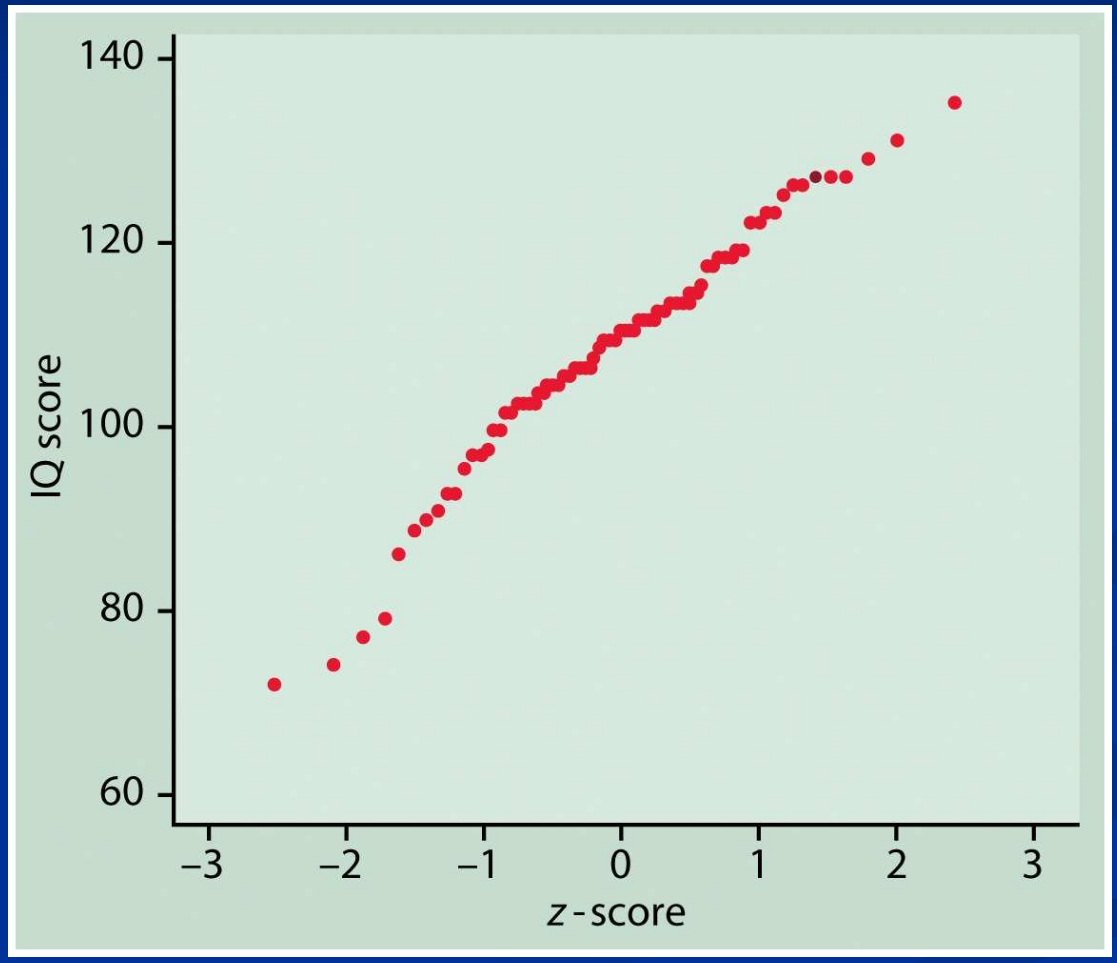
Politics Question



Quantile Plot for Politics

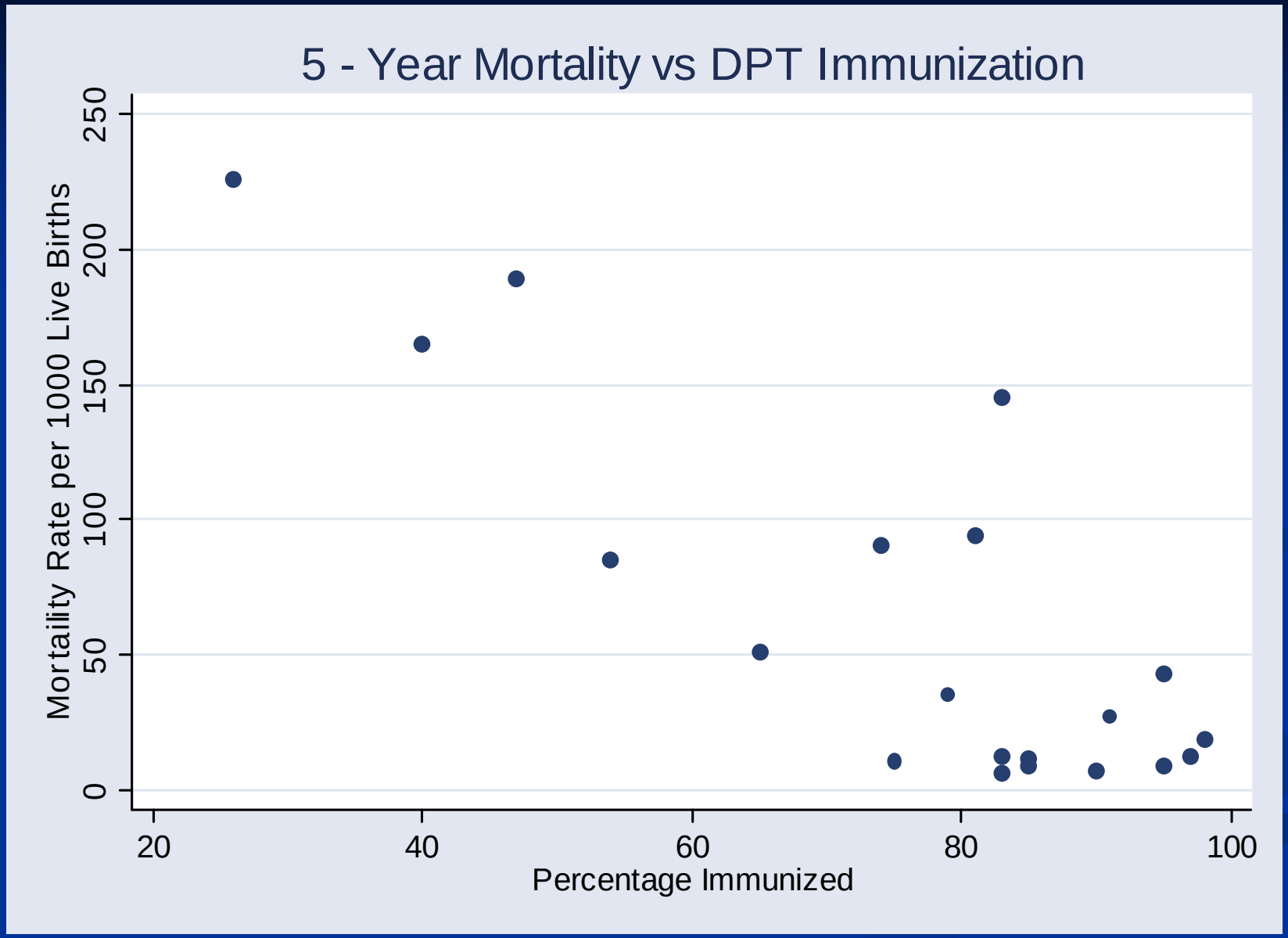


Quantile Plot for IQ Scores



Linear Relationships and Correlation

- Distinction between categorical and quantitative variables
- Linear relationship between two variables
- Scatter plot
- Correlation


5 - Year Mortality vs DPT Immunization
Mortaility Rate per 1000 Live Births
0
50
100
150
200
250
20
40
60
80
100
Percentage Immunized

Positive and Negative Association

- **Positive Association:** High values of one variable tend to accompany high values of the other variable, and low values of one variable tend to accompany low values of the other variable.
- **Negative Association:** High values of one variable tend to accompany low values of the other variable and vice versa.
- What type of association do we see in the immunization data?

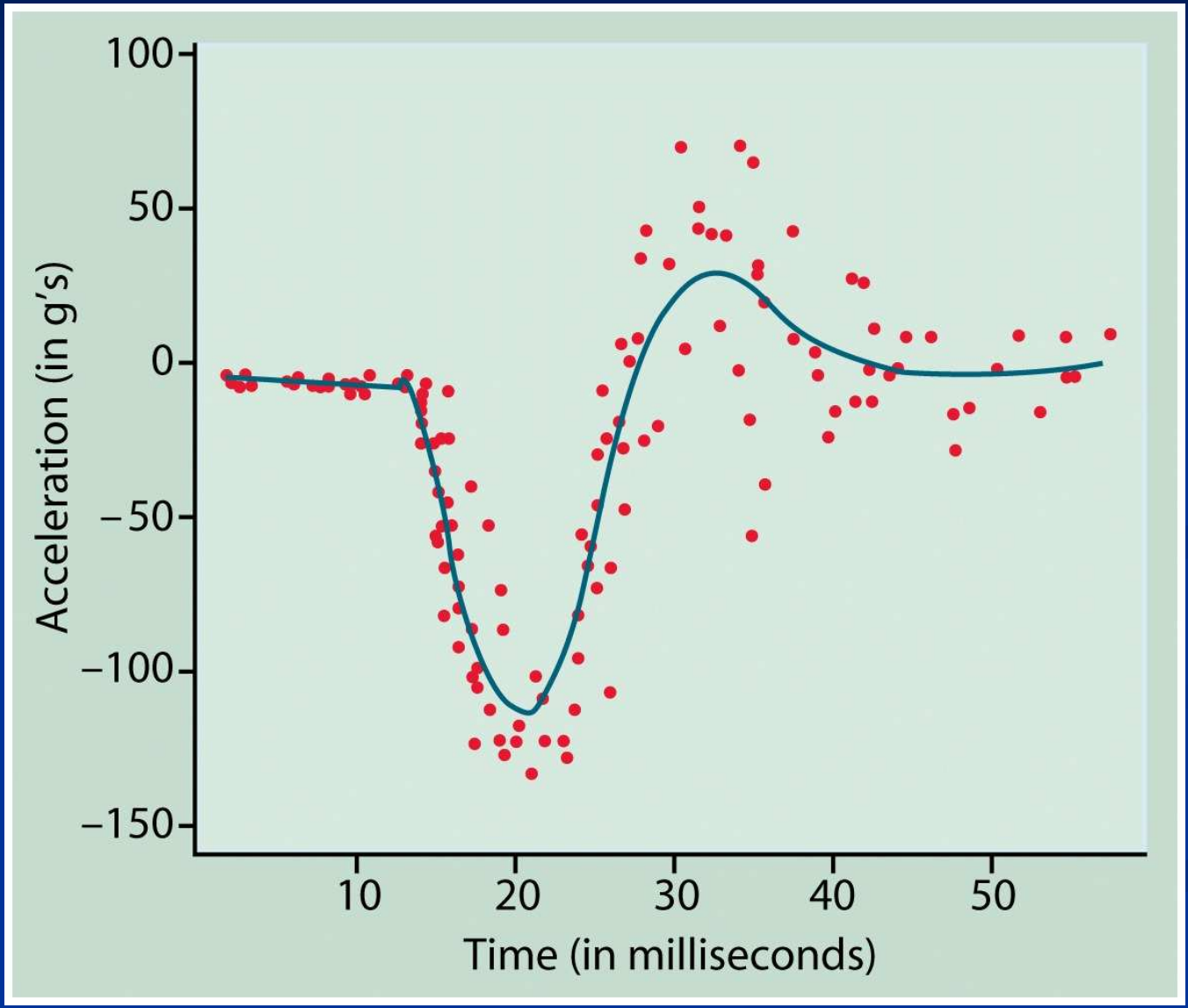
Caveat

- **Association does NOT imply causation!!!**
- Association between two variables does not necessarily mean that one variable causes a change in the other.

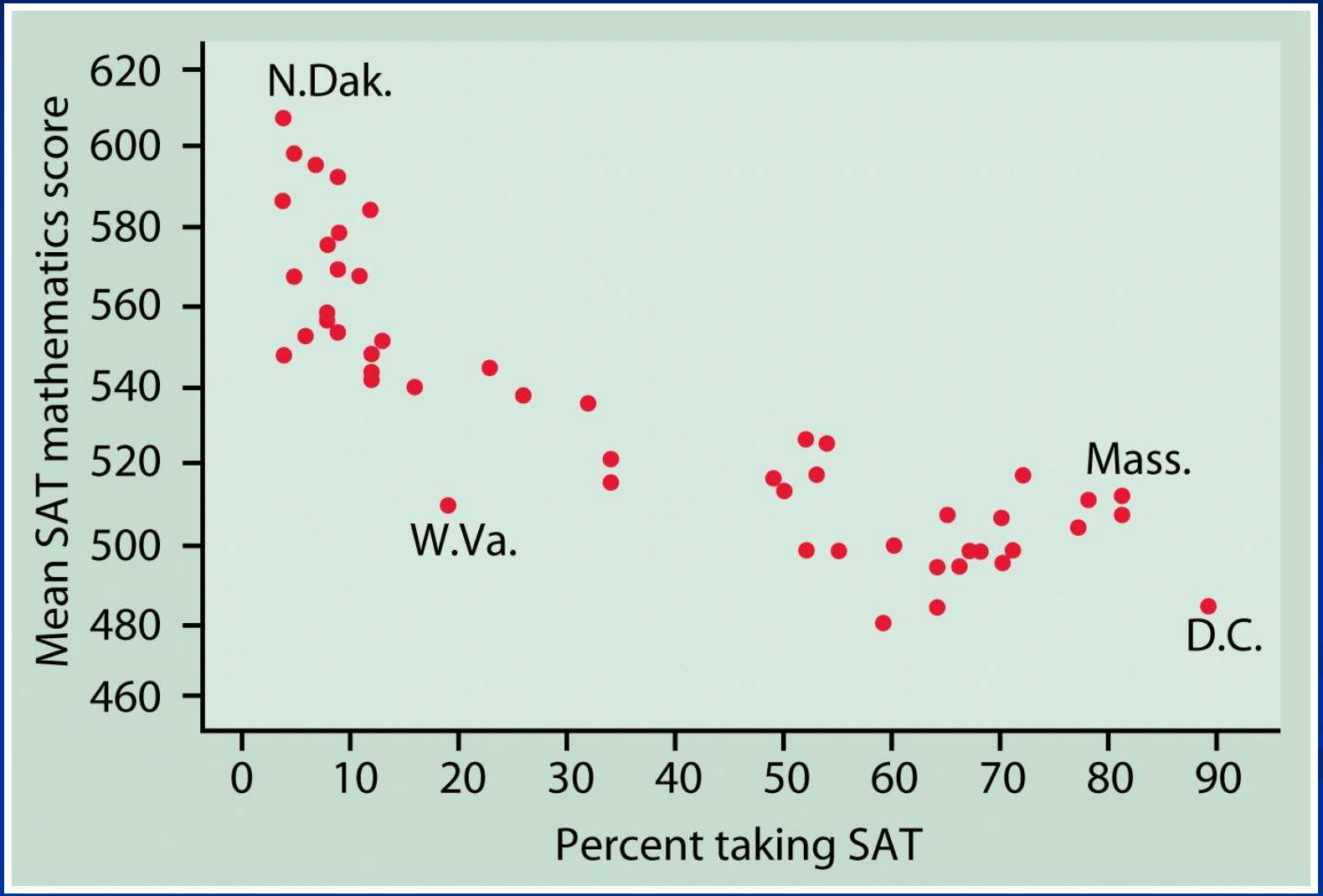
Linear Relationship

- Do the immunization data show a linear relationship?
- Could you imagine easily fitting a straight line to these data?
- Do you think that line would fit well?

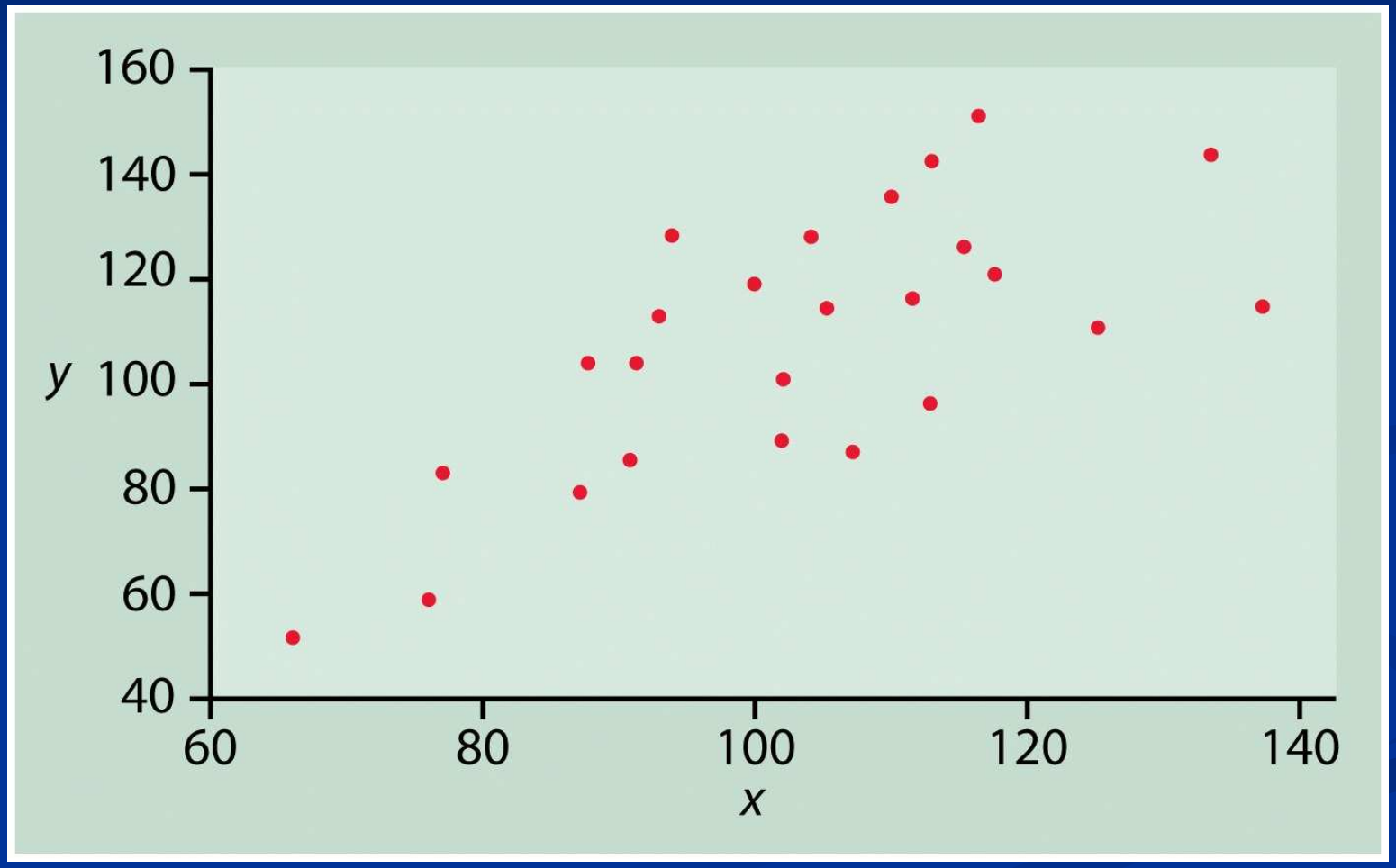
Non-Linear Relationship



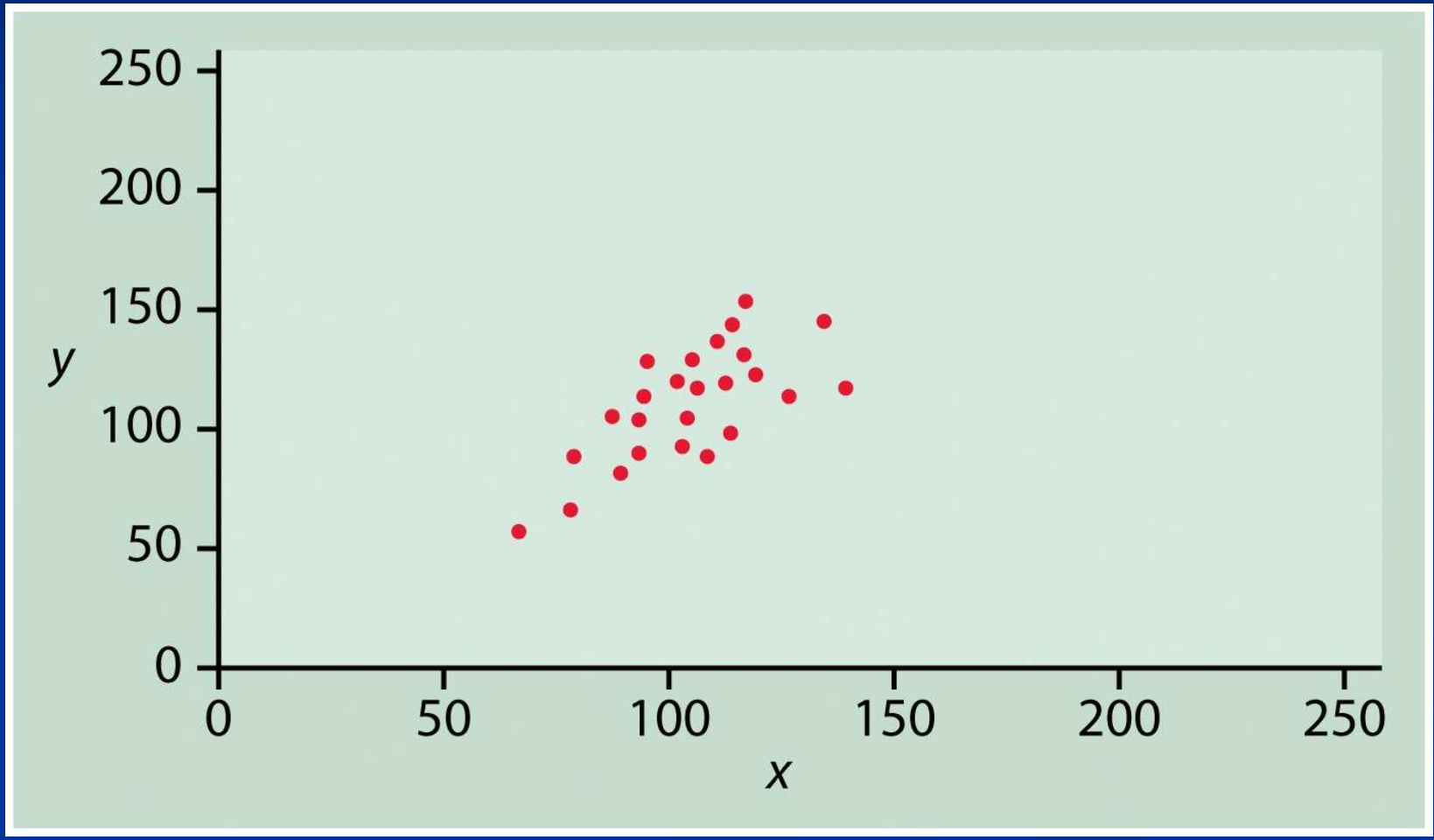
Linear Relationship



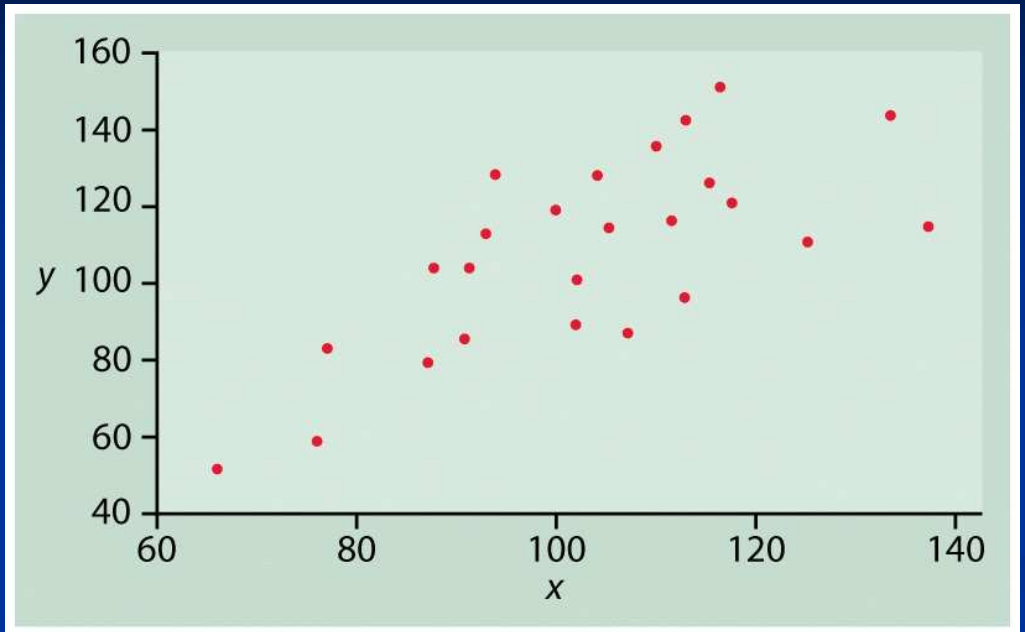
Measuring Linear Relationships

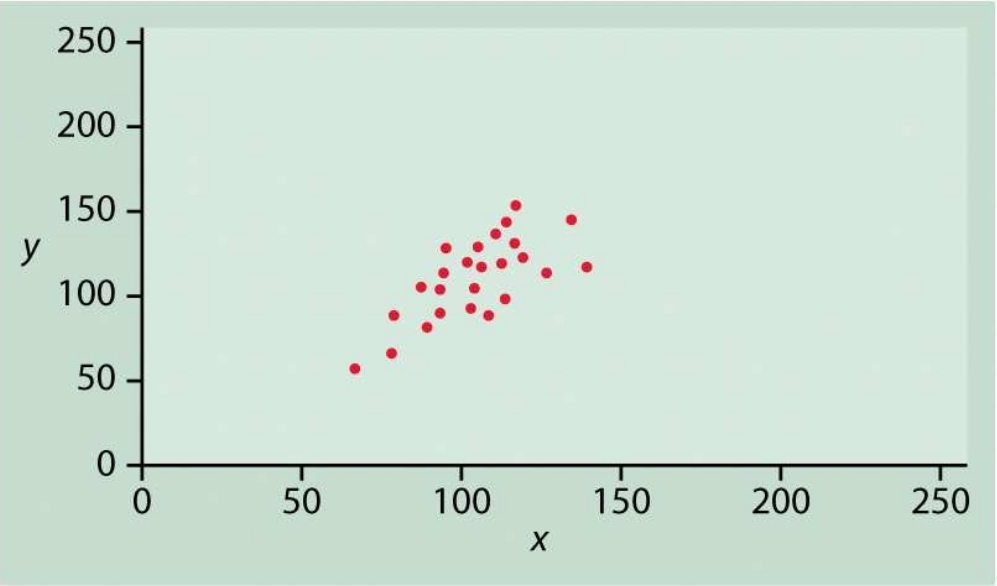


Measuring Linear Relationships



Measuring Linear Relationships





Correlation Coefficient

- **Correlation Coefficient**, denoted as ***r*** (or ρ), measures the strength of *linear* association between two quantitative variables.
- **Properties:**
  - *r* measures linear association
  - *r* always falls between -1 and +1
  - $r = 1$ or -1 only if observations lie exactly along a straight line

Correlation Coefficient: Definition

- Suppose we have a list of *n* pairs of observations:

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n y_n)$$

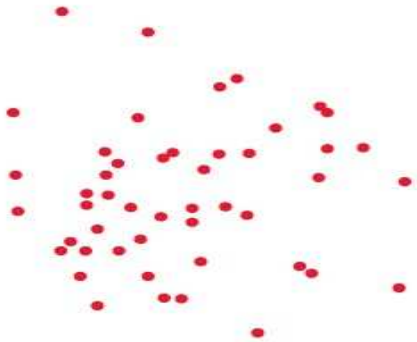
Correlation of X and Y is given by,

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

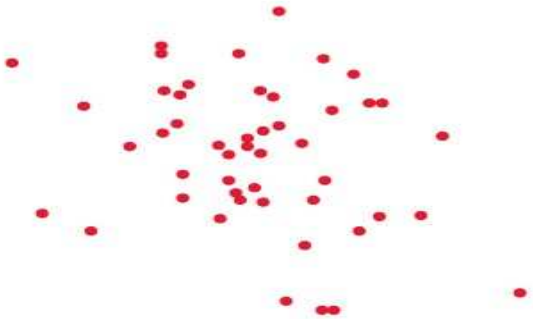
where s_x and s_y are the SDs of X and Y.

Correlation Coefficient

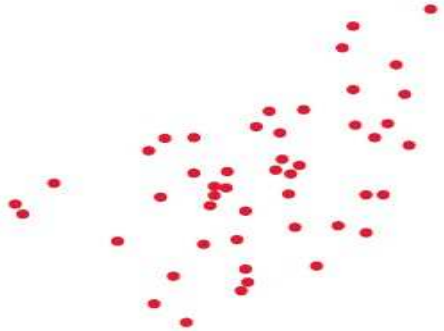
- Additional Properties:
  - Sign if *r* indicates whether there is positive (+) or negative (-) association.
  - Absolute value of *r* measures the strength of the linear relationship.
  - *r* is unaltered by changes in the units of measurement of X and/or Y.
  - *r* has no units of measurement, e.g., $r = 0.8$ is not twice as strong as $r = 0.4$.



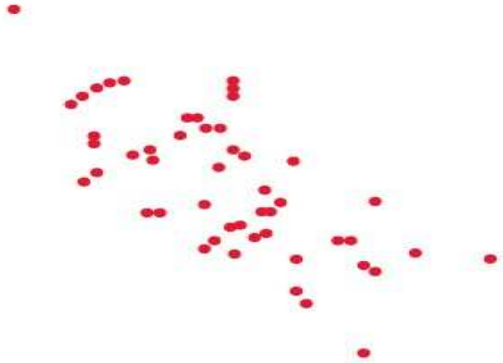
Correlation $r = 0$



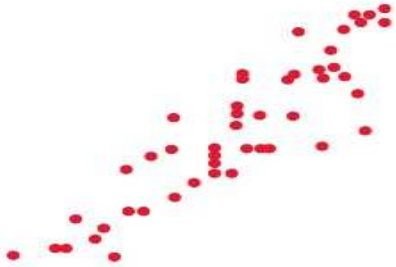
Correlation $r = -0.3$



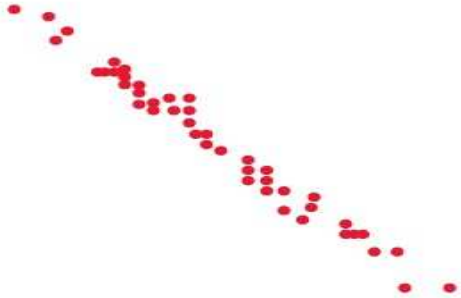
Correlation $r = 0.5$



Correlation $r = -0.7$



Correlation $r = 0.9$



Correlation $r = -0.99$

Example: Mortality

- Under 5-mortality rate per 1,000 live births versus percent immunized against DPT.

```
. correlate immune mortality
(obs=20)

             |   immune mortal~y
-------------+------------------
      immune |   1.0000
   mortality |  -0.8291   1.0000
```

Example: Mortality

- With $r = -0.83$, there is evidence of a strong negative association between the under-5 mortality rate and percent of children immunized against DPT.
- Note: The correlation does not allow us to estimate how much the under-5 mortality rate would decrease if a country were to increase the percentage of children immunized by 10%.

Correlation measures LINEAR relationships