Review of Probability and Statistics I

Zhenhao Gong

University of Connecticut

This course is designed to be:

1. Introductory
2. Leading by interesting questions and applications
3. Less math, useful, and fun!

Most important:

Feel free to ask any questions!

Enjoy! ☺

Review

- What is Econometrics?
- Why study Econometrics?
- What is causal effects? How to measure it?
- What is randomized controlled experiment?
- What are the two main sources of data in econometrics?
- What are the three main types of data sets in econometrics?

Goal

Reviews the **core** ideas of the theory of probability and statistics that are needed to understand regression analysis and econometrics

- The probability framework for statistical inference
- Estimation
- Hypothesis Testing
- Confidence Intervals

Empirical problem

Class size and educational output:

- Policy question: What is the **quantitative** effect of reducing class size on student achievement?
- Specifically, what is the effect on test scores of reducing class size by one student per class? by 6 students per class?
- We must use data to find out (is there any way to answer this without data?).

Data set:
All K-6 and K-8 California school districts (n = 420)
(www.cde.ca.gov)

Useful variables:

- 5^{th} grade test scores (Stanford-9 achievement test, combined math and reading), district average
- Student-teacher ratio (STR) = number of students in the district divided by number full-time equivalent teachers

Initial look

What does this table show?

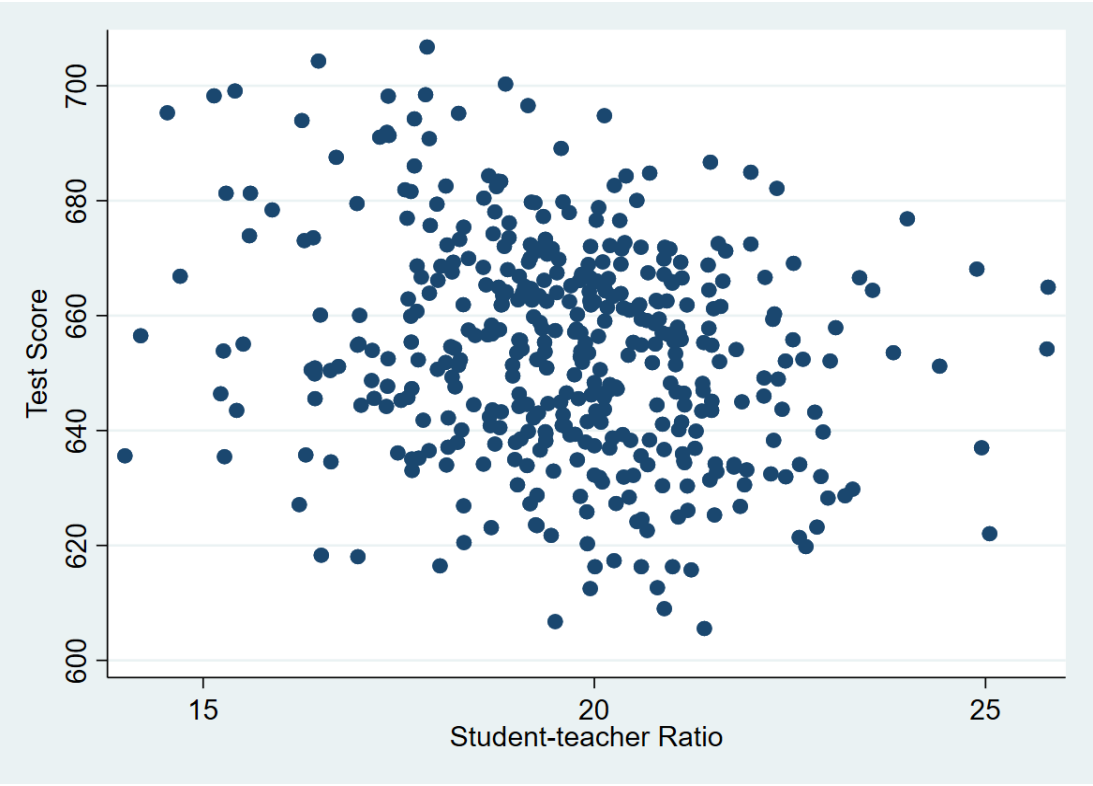
Summary of test scores in the data

```
. sum testscr, detail

                          testscr
-------------------------------------------------------------
      Percentiles      Smallest
 1%       612.65          605.55
 5%       623.15          606.75
10%      630.375             609       Obs                 420
25%          640           612.5       Sum of Wgt.         420

50%       654.45                       Mean           654.1565
                         Largest       Std. Dev.      19.05335
75%      666.675           699.1
90%        679.1           700.3       Variance       363.0301
95%        685.5           704.3       Skewness       .0916151
99%       698.45          706.75       Kurtosis       2.745712
```

Initial look

What does this figure show?

Scatterplot of test score v.s. student-teacher ratio



Key Question

Do districts with smaller classes have higher test scores?

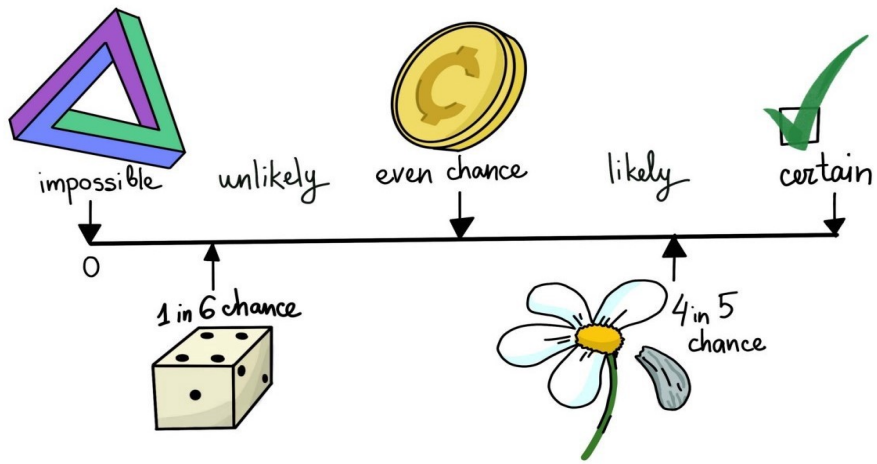
We need to get some numerical evidence on whether districts with low STRs have higher test scores – but how?

Steps

We can get the numerical evidence from the data set by following steps:

1. Compare average test scores in districts with low STRs to those with high STRs (**"estimation"**)
2. Test the "null" hypothesis that the mean test scores in the two types of districts are the same, against the "alternative" hypothesis that they differ (**"hypothesis testing"**)
3. Estimate an interval for the difference in the mean test scores, high v.s. low STR districts (**"confidence interval"**)

Review of Probability



Most aspects of the world around us have an element of **randomness**:

- the gender of the next new person you meet
- the number of times your computer will crash while you are writing a term paper
- the change of the stock market price
- the district average test score or the district STR

In each of these examples, there is something not yet known that is eventually revealed.

The theory of probability provides mathematical tools for quantifying and describing this randomness.

Outcomes:

- The mutually exclusive potential results of a random process.

Probability of an outcomes:

- the proportion of the time that the outcome occurs in the long run.

Population (Sample space):

- the group or collection of all possible entities of interest (school districts)

Event:

- a subset of the sample space, that is, an event is a set of one or more outcomes.

Random variable Y:

- numerical summary of a random outcome in the population (district test average sore, district STR)

Population distribution of Y:

- the probabilities of different values of Y that occur in the population (when Y is **discrete**): $\mathbf{Pr}(Y = 1)$
- the probabilities of intervals of these values (when Y is **continuous**): $\mathbf{Pr}(640 \leq Y \leq 660)$

Cumulative probability distribution (c.d.f.) of Y:

- the probability that the random variable is less than or equal to a particular value $\mathbf{Pr}(Y \leq 3)$

TABLE 2.1 Probability of Your Computer Crashing *M* Times

	Outcome (number of crashes)				
	0	1	2	3	4
Probability distribution	0.80	0.10	0.06	0.03	0.01
Cumulative probability distribution	0.80	0.90	0.96	0.99	1.00

(Graph?)

Example: Bernoulli

An important special case of a discrete random variable is when the random variable is binary, that is, the outcomes are 0 or 1.

A binary random variable is called a **Bernoulli random variable**, and its probability distribution is called the **Bernoulli distribution**.

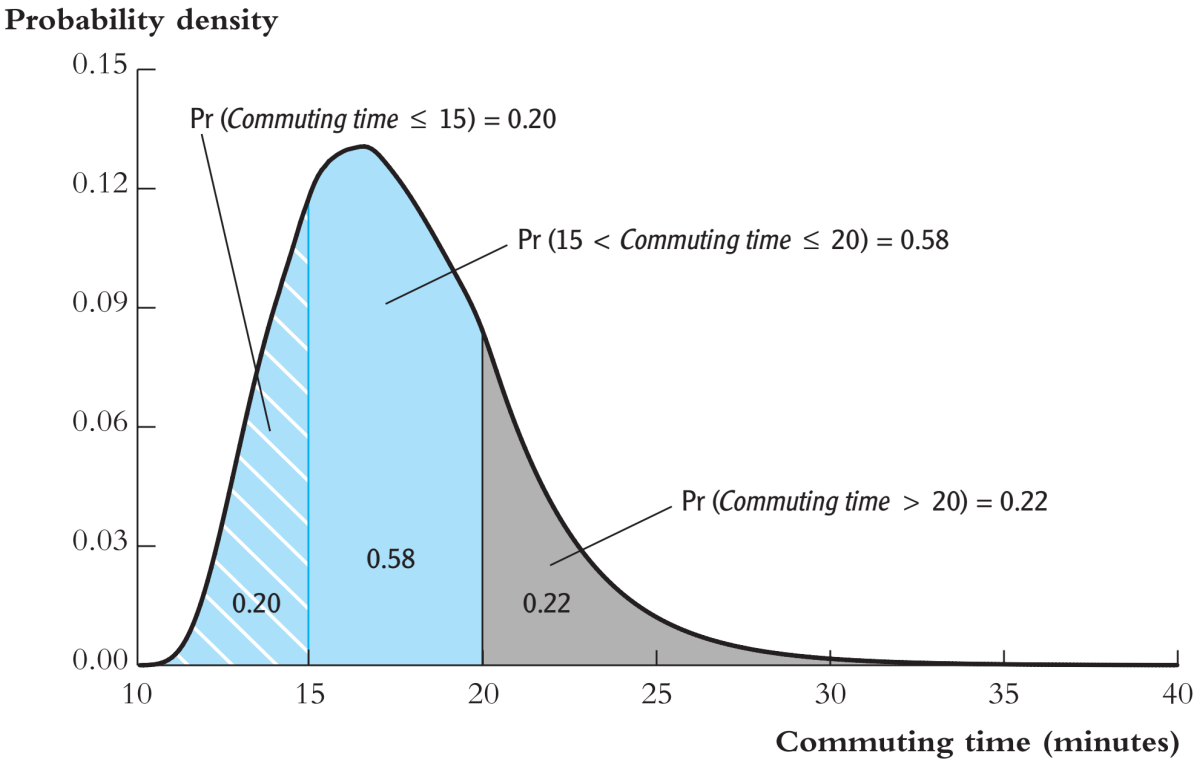
Example: let G be the gender of the next new person you meet, then we can express the outcome of G and their probability as

$$G = \begin{cases} 1 \text{ with probability } p \\ 0 \text{ with probability } 1-p \end{cases}$$

where p is the probability of the next new person you meet being a woman.

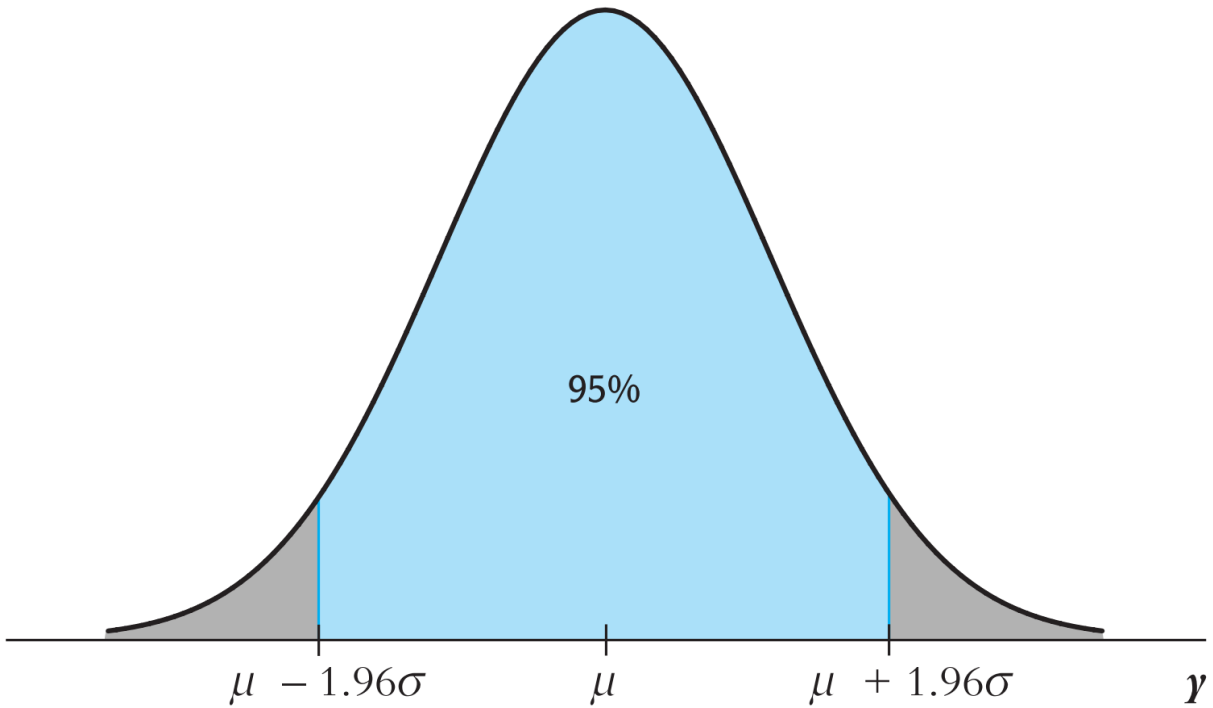
Probability density function (p.d.f.) of Y:

- the area under the probability density function between any two points is the probability that the random variable falls between those two points. (connection with c.d.f.?)



Example: Normal

The **normal** probability density function of Y with mean μ_Y and variance σ_Y^2 is a bell-shaped curve, centered at μ_Y. The area under the normal p.d.f. between $\mu_Y - 1.96\sigma_Y$ and $\mu_Y + 1.96\sigma_Y$ is 0.95. The normal distribution is denoted $N(\mu_Y, \sigma_Y^2)$.



Standard normal

The **standard normal distribution** is the normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ and is denoted N(0, 1).

Random variables that have a N(0, 1) distribution are often denoted Z, and its corresponding c.d.f. is denoted by the Greek letter Φ; $P(Z \leq c) = \Phi(c)$.

Suppose Y is distributed $N(\mu_Y, \sigma_Y^2)$. Then Y is standardized by subtracting its mean and dividing by its standard deviation, that is, by computing $Z = (Y - \mu_Y)/\sigma_Y$. (Calculation)

For the **population** distribution of Y:

mean = expected value (expectation) of Y
= the first moment of Y
= $E(Y) = \mu_Y$

variance = the second moment of Y
= measure of the spread of the distribution
= $E(Y - \mu_Y)^2 = \sigma_Y^2$

standard deviation = $\sqrt{\text{variance}} = \sigma_Y$

(Properties and Examples)

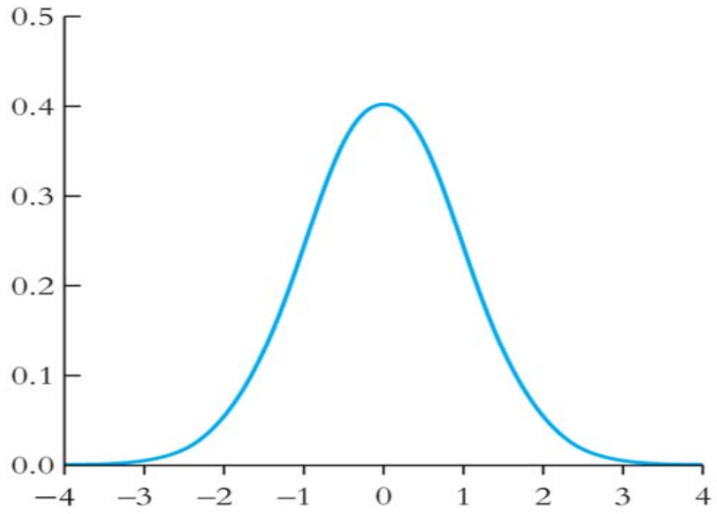
Skewness and kurtosis

Skewness:

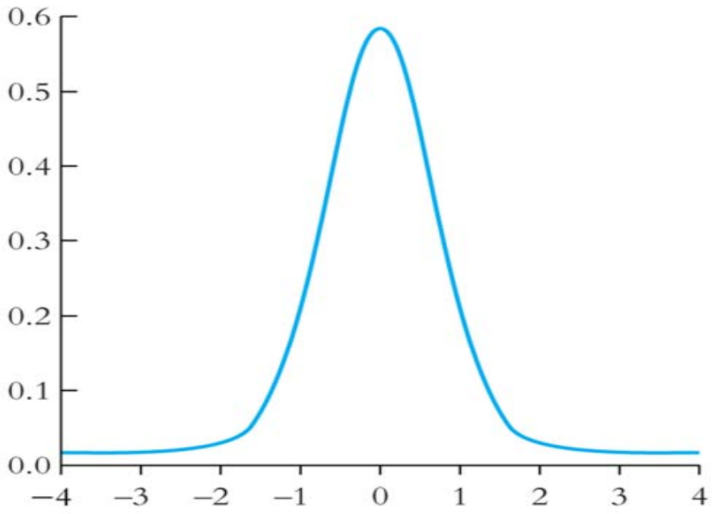
- measure of asymmetry of a distribution
- skewness $= 0$: distribution is symmetric
- skewness $> (<) 0$: distribution has long right (left) tail

kurtosis:

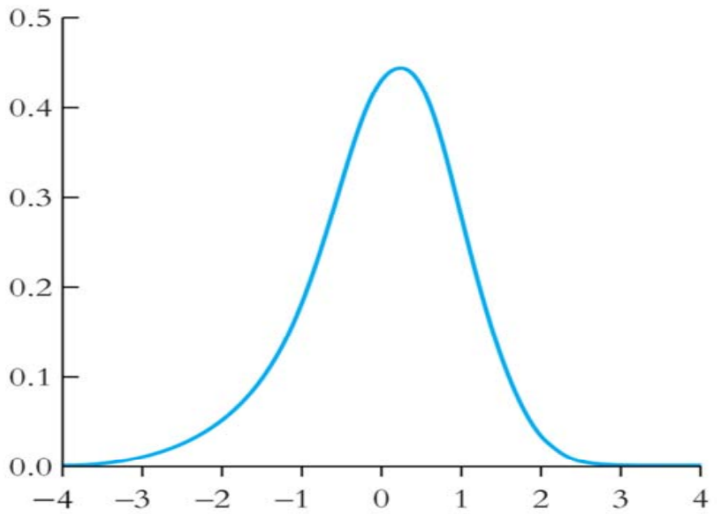
- measure of probability of large values
- kurtosis $= 3$: normal distribution
- kurtosis > 3 : heavy tails ("leptokurtotic")



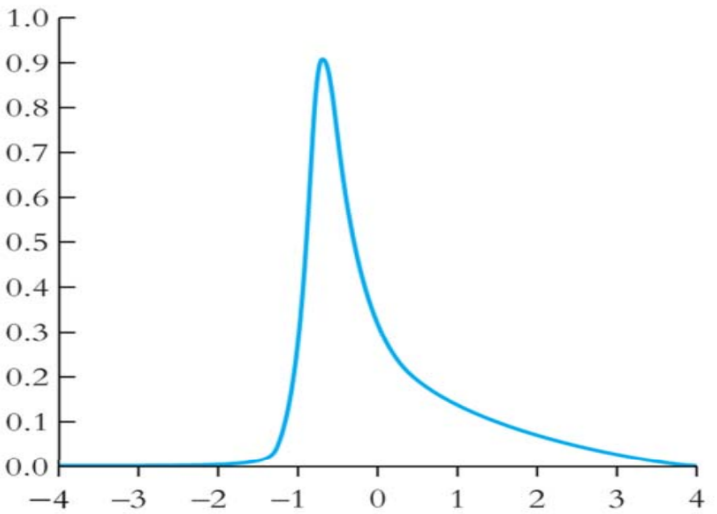
(a) Skewness = 0, kurtosis = 3



(b) Skewness = 0, kurtosis = 20



(c) Skewness = −0.1, kurtosis = 5



(d) Skewness = 0.6, kurtosis = 5

Random variables X (district average score) and Y (STR) have a **joint distribution** (at least two random variables).

The **covariance** between X and Y is

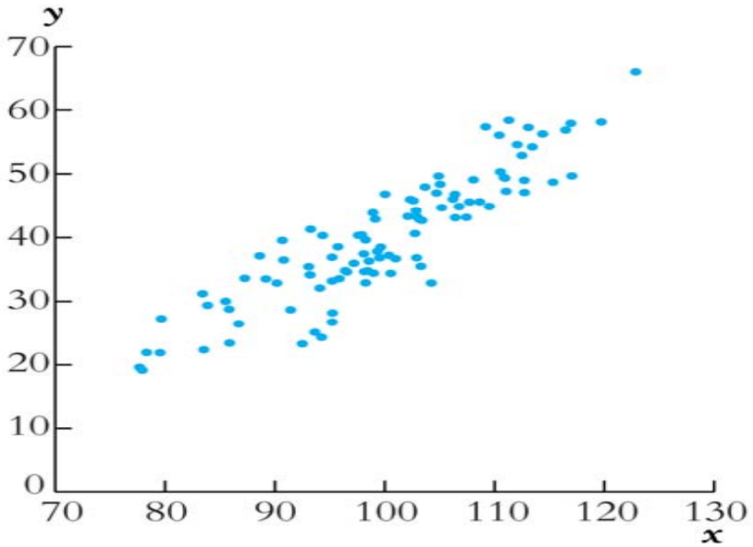
$$\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY}.$$

- measure of the extent to which two random variables X and Y move together.
- $\text{cov}(X, Y) > (<) 0$ means a positive (negative) relation between X and Y.
- If X and Y are independently distributed, then $\text{cov}(X, Y) = 0$.

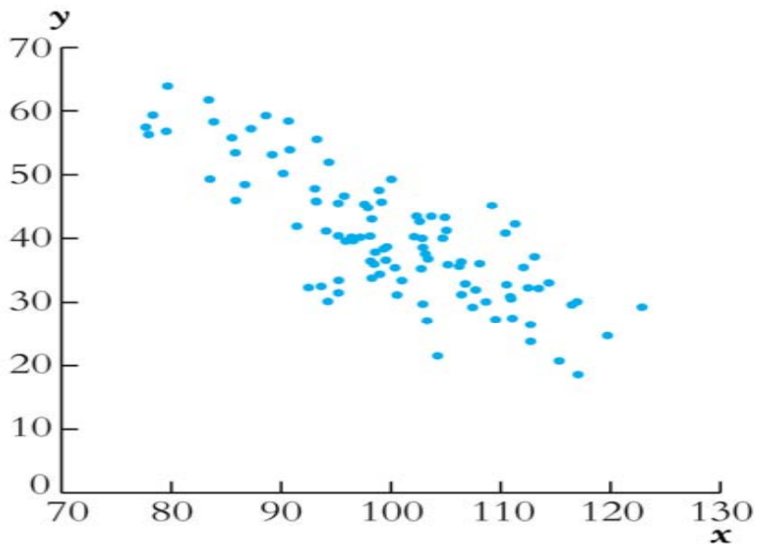
The **correlation coefficient** of X and Y is defined in terms of the covariance:

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = r_{XY}.$$

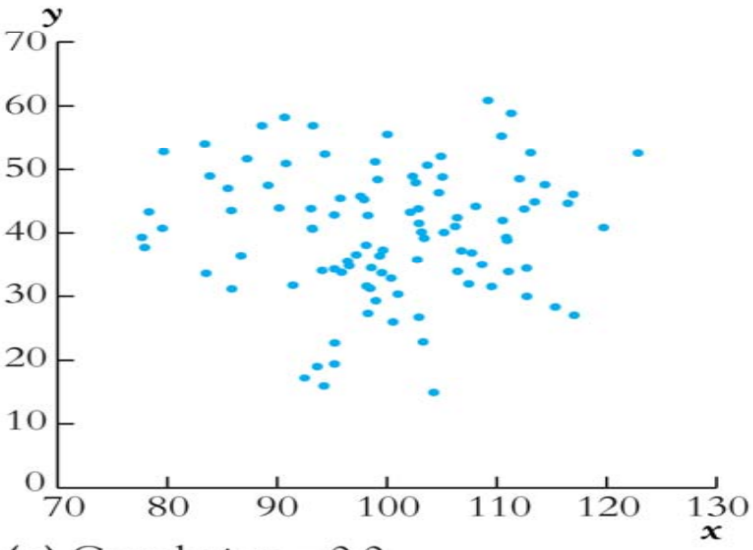
- $-1 \leq \text{corr}(X, Y) \leq 1$.
- $\text{corr}(X, Y) = 1$: perfect positive linear association
- $\text{corr}(X, Y) = -1$: perfect negative linear association
- $\text{corr}(X, Y) = 0$: no linear association



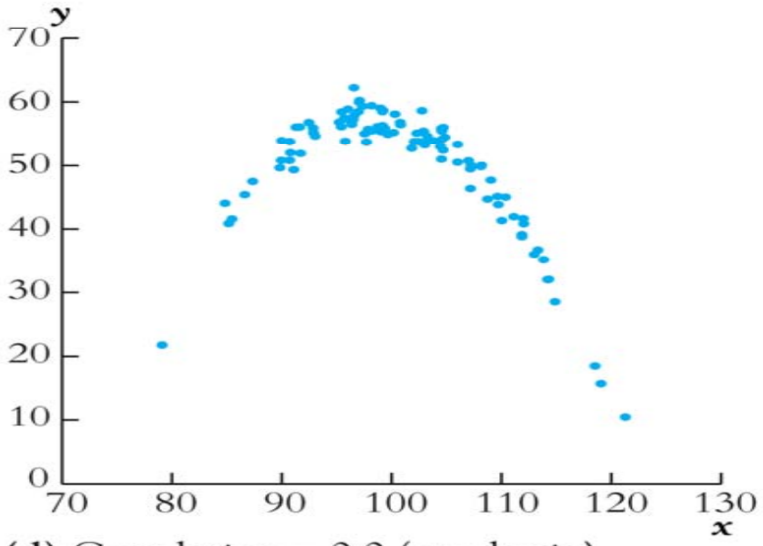
(a) Correlation = +0.9



(b) Correlation = −0.8

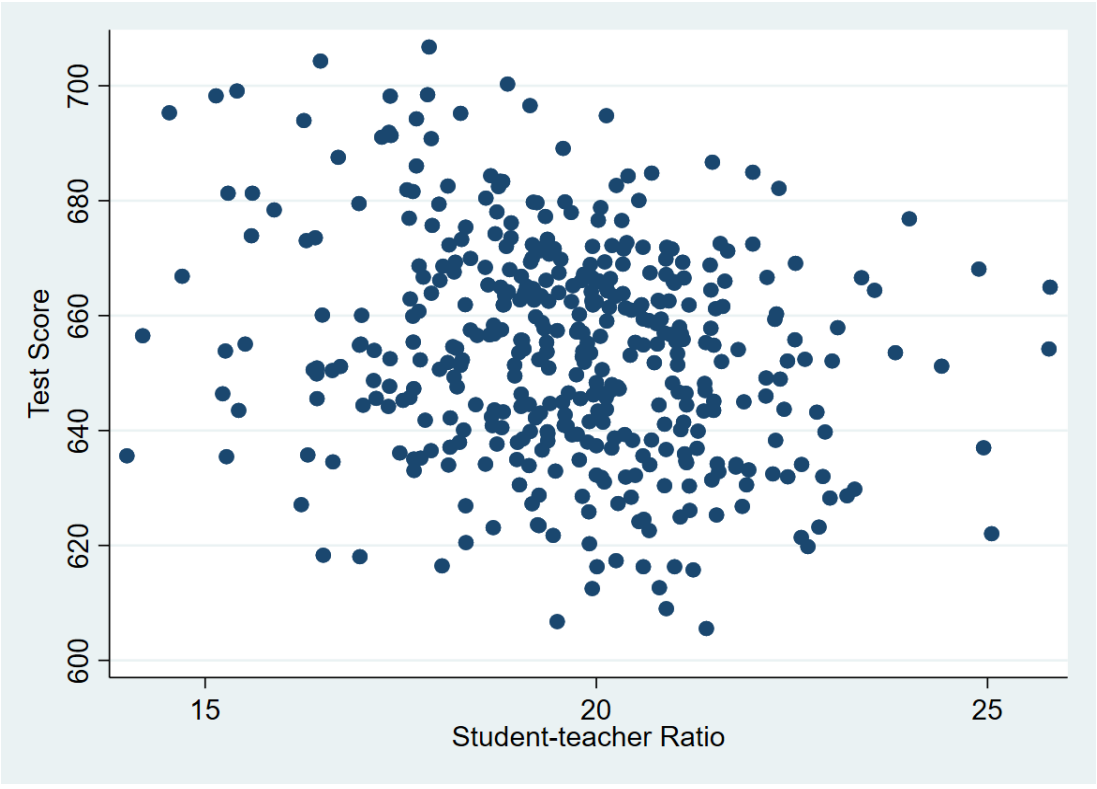


(c) Correlation = 0.0



(d) Correlation = 0.0 (quadratic)

The correlation of X (STR) and Y (test score) is -0.23!
Scatterplot of test score v.s. student-teacher ratio



Conditional distribution of Y:

- The distribution of Y, given value(s) of some other random variable, X
  Example: the distribution of test scores, given STR < 20.

Conditional mean (expectation) of Y:

- The mean of conditional distribution: $E(Y|X)$ (important!)
  Examples:
  1. the mean of test scores among districts with small class sizes: $E(Y = \text{Test score}|X = \text{STR} < 20)$;
  2. the mean wage of all female workers $E(Y = \text{wages}|X = \text{gender})$.

  (Calculation)

Sampling distribution

- Distribution of a sample of data drawn **randomly** from a population: $Y_1, \cdots, Y_n$. (Why we need to do this?)
- Under simple random sampling, $Y_1, \cdots, Y_n$ are independently and identically distributed **(i.i.d.)**

This framework allows rigorous statistical inferences about moments of population distributions, using a sample of data from that population.

We use **the sampling distribution of** $\bar{Y}$ to do the statistical inference. So this is the **key** concept.

- $\bar{Y} = (Y_1 + Y_2 + \cdots + Y_n)/n$ is a random variable
- The distribution of $\bar{Y}$ over different possible samples of size n is called the sampling distribution of $\bar{Y}$
- The mean and variance of Y are the mean and variance of its sampling distribution, $E(\bar{Y})$ and $\text{Var}(\bar{Y})$

Question:
What are the differences from the population distribution of Y and its corresponding moments $E(Y)$ and $\text{Var}(Y)$?

Recall goal

Reviews the **core** ideas of the theory of probability and statistics that are needed to understand regression analysis and econometrics

- The probability framework for statistical inference ✓
- Estimation
- Hypothesis Testing
- Confidence Intervals

Review of Statistics

Read: Stock & Watson, Chapter 3.