Random Variables and Probability Distributions

26 Aug 2026 | Chapter 2

ECO 351 – Introduction to Econometrics

Shi Qiu – Fall 2026

Recall Lecture 1: read a raw difference carefully

Class size and test scores

Lecture 1 asked whether the same population of students would score differently in smaller classes. Each student is actually observed in only one class-size assignment.

$$\tau = \mathbb{E}[Y(1) - Y(0)]$$

Your task

What is missing from an ordinary comparison of small and large classes?

Work in this order: Name the unobserved outcome first; then ask what design can approximate it.

Why the raw comparison may not be causal

What is going on in the example?

The missing outcome is the score the same student would have received under the other class-size assignment. Random assignment makes treated and control groups comparable in expectation, so one group's average can estimate the other group's counterfactual mean.

Bridge to today's lecture

Today we need probability language to describe the uncertainty that remains when data are drawn from a population.

A random variable records one uncertain outcome

Define the outcome

Let M be the number of times a computer crashes while a student completes a term paper. The possible values are $0, 1, 2, 3$, and 4.

Then assign probability

A probability distribution gives $Pr(M = m)$ for every possible value m. Those probabilities must add to one.

The random variable is the numerical outcome. The distribution describes the uncertainty surrounding that outcome.

A probability distribution answers questions about events

Number of crashes, m	0	1	2	3	4
$Pr(M = m)$	.80	.10	.06	.03	.01

An exact outcome

$Pr(M = 0) = .80$: no computer crash occurs.

An event with several outcomes

$Pr(M = 1 \text{ or } M = 2) = .10 + .06 = .16$.

Mutually exclusive outcomes can be added. Here, a student has a 16% chance of exactly one or two crashes.

The CDF answers an "at most" question

m	0	1	2	3	4
$Pr(M = m)$	.80	.10	.06	.03	.01
$F_M(m) = Pr(M \leq m)$	.80	.90	.96	.99	1.00

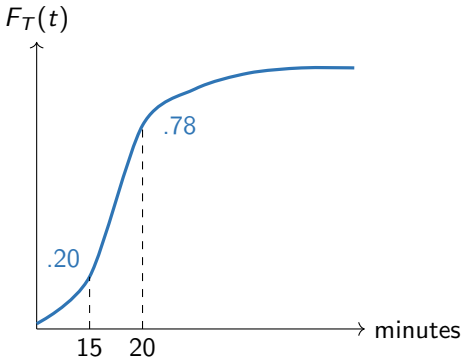
Read the second row cumulatively

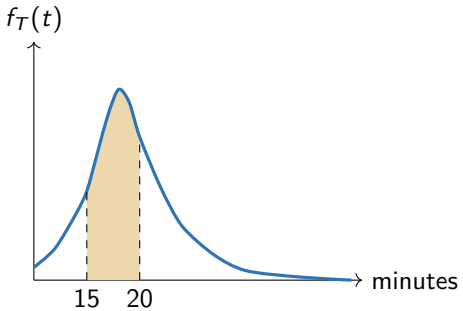
$F_M(1) = Pr(M \leq 1) = Pr(M = 0) + Pr(M = 1) = .80 + .10 = .90.$

Do not replace “at most” with “exactly”

$Pr(M = 1) = .10$, whereas $Pr(M \leq 1) = .90$. The CDF accumulates every outcome at or below the cutoff.

For commuting time, probability belongs to an interval





Two ways to read the same interval probability

$Pr(15 < T \leq 20) = F_T(20) - F_T(15) = .78 - .20 = .58$. On the density graph, the same .58 is the shaded area from 15 to 20 minutes.

Stata: describe the observed distribution before modeling it

Before running the code: `summarize` will report the sample's center and spread. The histogram will show how the observed values are distributed across intervals.

```
sysuse auto, clear
summarize mpg, detail
histogram mpg, percent
```

Read each output for its own purpose

The detailed summary reports the sample mean, standard deviation, percentiles, and extremes of fuel economy. The histogram reports the percentage of cars in each mileage interval. Neither output, by itself, establishes a population distribution or a causal relationship.

Expected value is a probability-weighted average

m	0	1	2	3	4
$Pr(M = m)$	.80	.10	.06	.03	.01

Calculation

$$E(M) = \sum_m mPr(M = m)$$

$$= 0(.80) + 1(.10) + 2(.06) + 3(.03) + 4(.01) = .35$$

Over many comparable term papers, the average number of crashes is .35. It does *not* say that one particular computer will crash .35 times.

Variance and standard deviation measure a different feature

Variance: average squared distance from the mean

$$Var(M) = \sum_m (m - E(M))^2 Pr(M = m) = .6475$$

Standard deviation: return to the original units

$$SD(M) = \sqrt{Var(M)} = \sqrt{.6475} \approx .80 \text{ crashes}$$

The mean, .35 crashes, describes the center of the distribution. The standard deviation, .80 crashes, describes how dispersed outcomes are around that center. Variance is measured in squared crashes, so standard deviation is usually easier to interpret.

A linear policy can change average income and its spread

After-tax income

Suppose pre-tax income X is random and a tax-and-grant policy gives

$$Y = 2000 + .80X.$$

What follows

$$E(Y) = 2000 + .80E(X)$$

$$SD(Y) = .80SD(X)$$

The grant shifts every income by 2000, so it shifts the mean but not dispersion. Keeping 80% of pre-tax income scales both the mean component from X and the standard deviation by .80.

Practice: calculate center and spread from a different distribution

A campus tutoring center receives $Y = 0$, 2, or 4 walk-in students in a half-hour period with probabilities .25, .50, and .25, respectively.

Work for 90 seconds

Calculate $E(Y)$ and $Var(Y)$. Then write one sentence that distinguishes the expected number of students from the standard deviation.

Use the table of possible values and probabilities. Do not turn the expected value into a claim about every half-hour period.

Solution: the average and the spread answer different questions

Expected number of walk-ins

$$E(Y) = 0(.25) + 2(.50) + 4(.25) = 2.$$

Variance and standard deviation

$$Var(Y) = (0-2)^2(.25) + (2-2)^2(.50) + (4-2)^2(.25) = 2, \qquad SD(Y) = \sqrt{2}.$$

The center is 2 walk-ins per half hour. The standard deviation, about 1.41 walk-ins, says that realized counts typically vary substantially around that center.

A joint distribution records two outcomes at the same time

Let $R = 1$ when it rains and $L = 1$ when a commute lasts at least 20 minutes.

	Rain, $R = 1$	No rain, $R = 0$	Total
Long commute, $L = 1$	.15	.07	.22
Short commute, $L = 0$	.15	.63	.78
Total	.30	.70	1.00

Each interior cell is a joint probability. For example, $Pr(L = 1, R = 1) = .15$: on 15% of days, it rains and the commute is long.

A marginal distribution ignores the other variable

	Rain, $R = 1$	No rain, $R = 0$	Total
Long commute, $L = 1$	.15	.07	.22
Short commute, $L = 0$	.15	.63	.78
Total	.30	.70	1.00

Sum across the values of the variable being ignored

$$Pr(L = 1) = Pr(L = 1, R = 1) + Pr(L = 1, R = 0) = .15 + .07 = .22.$$

The marginal probability of a long commute is 22%, regardless of weather.

Conditional probability restricts attention to an observed group

Rainy days only	Probability
Long commute	.15
Short commute	.15
All rainy days	.30

Within the rainy-day group

$$Pr(L = 1 \mid R = 1) = \frac{Pr(L = 1, R = 1)}{Pr(R = 1)} = \frac{.15}{.30} = .50.$$

The denominator is no longer every day. It is the 30% of days on which rain was observed. Among those days, one half have a long commute.

Conditional means use the same logic to summarize a subgroup

	$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	$E(M \mid A)$
Old computer, $A = 0$	.70	.13	.10	.05	.02	.56
New computer, $A = 1$	.90	.07	.02	.01	.00	.14

Calculate a mean within each condition

$$E(M \mid A = a) = \sum_m mPr(M = m \mid A = a).$$

The expected number of crashes is .56 for an old computer and .14 for a new one. Later, regression will model the conditional mean $E(Y \mid X = x)$ using observed data.

Practice: read a conditional probability as a within-group percentage

	College degree	No college degree	Total
Employed	.38	.30	.68
Not employed	.12	.20	.32
Total	.50	.50	1.00

Work for 90 seconds

Calculate $Pr(\text{employed} \mid \text{college degree})$. Compare it with the unconditional probability of employment, and identify the denominator in each calculation.

Solution: conditioning changes the reference group

Among those with a college degree

$$Pr(\text{employed} \mid \text{college degree}) = \frac{.38}{.50} = .76.$$

Among all individuals

$$Pr(\text{employed}) = .68.$$

The conditional probability uses only the college-degree group in its denominator; the unconditional probability uses the full population. The difference, $.76 - .68$, is descriptive. It is not yet evidence of the causal effect of earning a degree.

Keep the probability objects distinct

Question	Probability object
What are the possible values of one uncertain outcome?	Distribution: $Pr(Y = y)$
What is the probability below a cutoff?	CDF: $F_Y(y) = Pr(Y \leq y)$
What is the long-run center or spread?	$E(Y)$ and $SD(Y)$
How does the distribution change after X is known?	Conditional distribution: $Pr(Y = y \mid X = x)$ and $E(Y \mid X = x)$

Closing check

Rain raises the long-commute probability from .22 overall to .50 among rainy days. Which object produces .50, and why is its denominator .30 rather than 1?

Next class: reference distributions, including the normal distribution and the probabilities calculated from it.