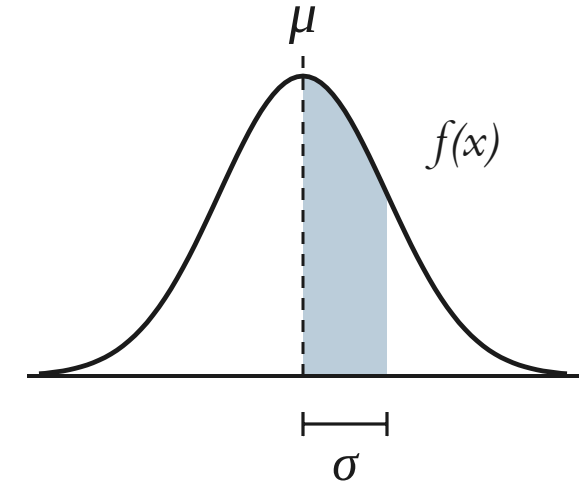
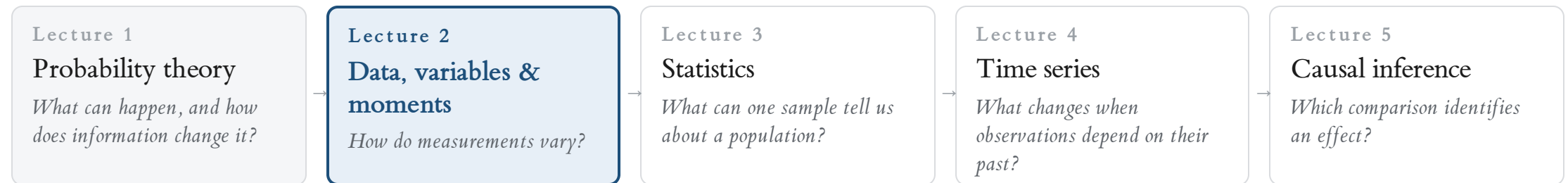


Data, variables & moments

How do measurements vary, and how do we summarise and relate them?



1 Describe → 2 Summarise → 3 Relate



OPENING QUESTION

Ten exam scripts: what mark should you expect from one drawn at random?

30, 30, 30

three scripts

60, 60

two scripts

80, 80, 80

three scripts

90, 100

one each

PUZZLE 1

(a) What mark should you expect? Does any script have it?

PUZZLE 2

(b) Women's four scripts average 72.5, men's six average 58.33. Is the class mean halfway, 65.42?

WHY START HERE

A national average wage built from group averages raises the same question: which weights, and why?

LECTURE 2 · THE ROUTE

Describe the variation, summarise it, then relate it to another variable

PART 1

Describe: from records to distributions

How likely is each value, each interval, each threshold?

random variable PMF density
CDF quantile

Notes 2.1–2.4



PART 2

Summarise: centre, spread and the standard normal

Where is a distribution centred, how spread out is it, and how far from typical is a given value?

expectation variance
standard deviation standardise Φ

Notes 2.5–2.7



PART 3

Relate: joint, conditional and association

What does knowing one variable tell us about another?

joint marginal conditional mean
covariance correlation

Notes 2.8–2.10

BY THE END OF THIS LECTURE YOU CAN

- ✓ Classify a variable by type, support and structure, and read probabilities and quantiles from a PMF, density or CDF.
- ✓ Compute and interpret a mean, variance and standard deviation, using linearity and the change-of-units rules.
- ✓ Standardise a normal variable to find probabilities, quantiles and central intervals from one table.
- ✓ Derive marginal and conditional distributions and means from a joint table, and tell independence from zero correlation.

PART 1 OF 3

Describe: from records to distributions

How likely is each value, each interval, each threshold?

KEY TERMS

random variable

PMF

density

CDF

quantile

[Notes 2.1–2.4 ↗](#)

WHY IT MATTERS

Economic data mix codes, counts and amounts, recorded as cross-sections, time series and panels; the type decides which calculations mean anything. A distribution then answers questions such as what share of incomes lies at or below £x.

IN THIS PART

- 2.1 Measurement, support and data structure
- 2.2 A PMF and the binomial
- ▶ Animation: why a density can exceed 1 while probability is area
- 2.3 Densities: probability is area
- 2.4 The CDF, intervals and quantiles

2.1 | WHAT IS BEING MEASURED

Know what a number measures before calculating with it

DEFINITION

Random variable

$X : \Omega \rightarrow \mathbb{R}$ gives each outcome a number.
Its **support** is the set of values the model permits.

COMMON MISTAKE

Health coded poor 0, fair 1, good 2 is ordered only: “2 is twice as healthy as 1” has no meaning.

Type	Meaning	Example
Nominal	labels only	employed 1, unemployed 0
Ordinal	order, not distance	poor, fair, good
Discrete	countable quantities	people in a household
Continuous	a continuum, modelled	height, a lifetime

Structure: a cross-section X_i , a time series X_t , a panel X_{it} . GDP for 20 countries over 5 years with two records missing is an unbalanced panel of 98 records: rows, not independent evidence.

TAKEAWAY The mean of a 0/1 employment indicator is the employment rate; the mean of eye-colour codes is meaningless.

2.2 | MASSES ON VALUES

A probability mass function puts probability on values; a binomial counts every sequence

DEFINITION

Probability mass function (PMF)

$p_X(x) = P(X = x)$, with $p_X(x) \geq 0$ and $\sum_x p_X(x) = 1$.

WORKED EXAMPLE *ten exam scripts, one drawn at random*

Marks 30, 30, 30, 60, 60, 80, 80, 80, 90, 100:

x	30	60	80	90	100
$p_X(x)$	0.3	0.2	0.3	0.1	0.1
$P(X \geq 80) = 0.5; P(X = 70) = 0.$					

KEY FORMULA

RESULT

(1)

Binomial distribution

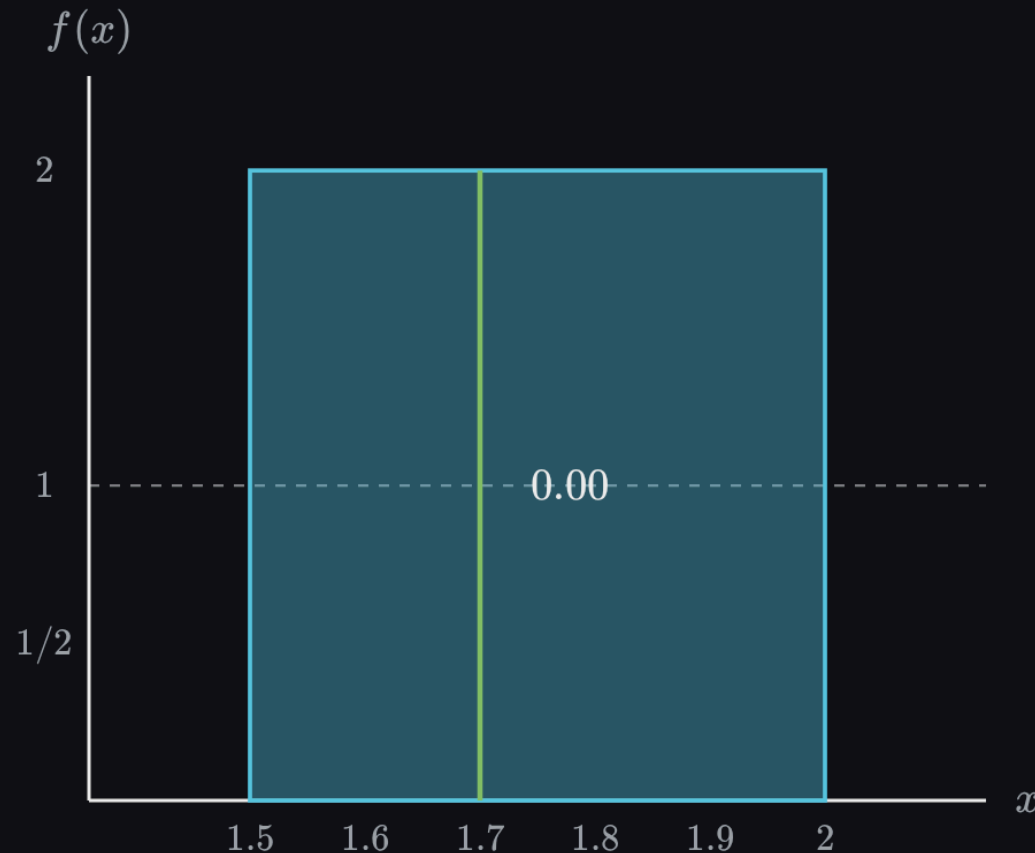
$$P(Y = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

IN WORDS $\binom{N}{k}$ sequences with k successes, each with probability $p^k (1 - p)^{N-k}$.

NEEDS N independent trials with a common $p = P(\text{success})$.

Two heads in three fair tosses: HHT, HTH, THH, so $3/8$, not $1/8$. Five in ten: $252/1024 \approx 0.246$.

Can a density be greater than 1?



width $\times$ height = area

$$U(-1, 1): 2 \times 1/2 = 1$$

$$U(0, 1): 1 \times 1 = 1$$

$$U(1.5, 2): 1/2 \times 2 = 1$$

height 2 is a density, not a probability

$$P(1.6 < X \leq 1.9) = 0.3 \times 2 = 0.6$$

$$P(X = 1.7) = 0, \text{ yet } f(1.7) = 2$$

Probability is area, not height.

2.3 | AREA UNDER A DENSITY

For a continuous variable, probability is area under a density

KEY FORMULA DEFINITION

(2)

Probability from a density

$$P(a < X \leq b) = \int_a^b f_X(x) dx$$

IN WORDS the area under the curve between a and b ; $f_X \geq 0$ with total area 1.

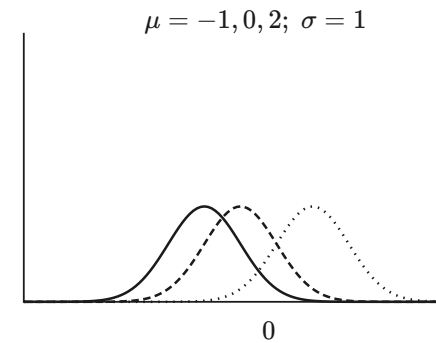
NEEDS X has a density; then $P(X = x) = 0$, so the endpoints do not matter.

COMMON MISTAKE

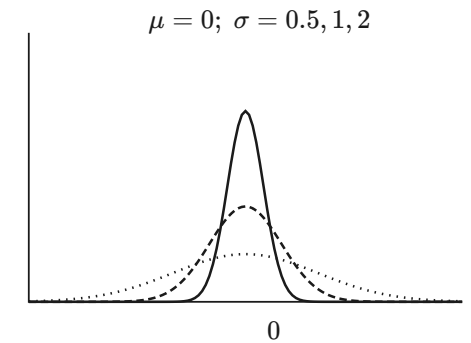
A density of 2 is not a probability of 2: it is probability per unit. Clip intervals to the support: for $U(-1, 1)$, $P(-2 < X \leq 0.5) = 1.5 \times \frac{1}{2} = 0.75$, not 1.25.

Shift location or change spread

Same spread



Same centre



A **density** family: $N(\mu, \sigma^2)$ shifts with μ and widens with σ ; the area stays 1.

2.4 | ACCUMULATING PROBABILITY

The CDF accumulates probability: subtract for intervals, read backwards for quantiles

KEY FORMULA DEFINITION

(3)

Cumulative distribution function

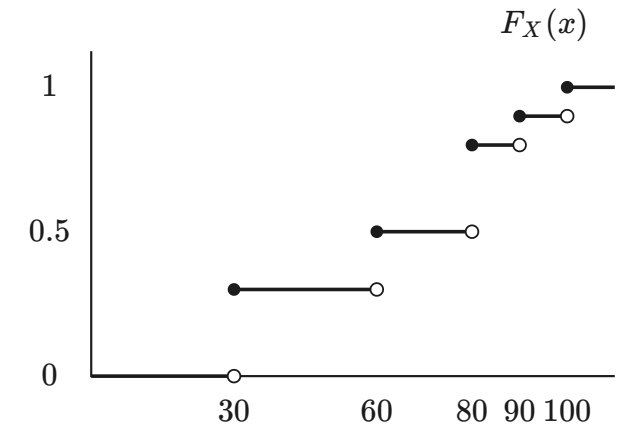
$$F_X(x) = P(X \leq x), \quad P(a < X \leq b) = F_X(b) - F_X(a)$$

IN WORDS the probability up to a threshold, for every real x ; a **quantile** reverses it: $Q_X(p)$ is the lowest x with $F_X(x) \geq p$.

COMMON MISTAKE

For the marks, $P(30 < X \leq 80) = 0.8 - 0.3 = 0.5$ but
 $P(30 \leq X \leq 80) = 0.8$: the endpoint includes three scripts.

A jump adds the probability at its mark



The marks' staircase: 0.3, 0.5, 0.8, 0.9, 1. Median $Q_X(0.5) = 60$; the usual sample median, 70, is a different convention.

TAKEAWAY A PMF or density spreads probability 1; the CDF accumulates it, so intervals are differences of F and quantiles are F read backwards.

PART 2 OF 3

Summarise: centre, spread and the standard normal

Where is a distribution centred, how spread out is it, and how far from typical is a given value?

KEY TERMS

expectation

variance

standard deviation

standardise

 Φ [Notes 2.5–2.7 ↗](#)

SO FAR a distribution puts probability on values or areas; the CDF accumulates it and quantiles read it backwards.

WHY IT MATTERS

Means and standard deviations are the first summary of wages, prices and returns. An expected payoff is not the likely payoff, and the standard normal scale is the yardstick for every interval and test in Lecture 3.

IN THIS PART

- 2.5 The mean is a balance point
- 2.5 Linear operations pass through expectations
- ▶ Animation: why the variance squares deviations
- 2.6 Variance and standard deviation
- ▶ Animation: standardising keeps the event's area
- 2.7 Standardise, look up Φ , translate back

2.5 | THE PROBABILITY-WEIGHTED BALANCE

The mean is a probability-weighted balance point, not a prediction

KEY FORMULA DEFINITION

(4)

Expected value

$$E[X] = \sum_x x p_X(x) \quad \text{or} \quad \int x f_X(x) dx$$

IN WORDS each value weighted by its probability; also called the mean.

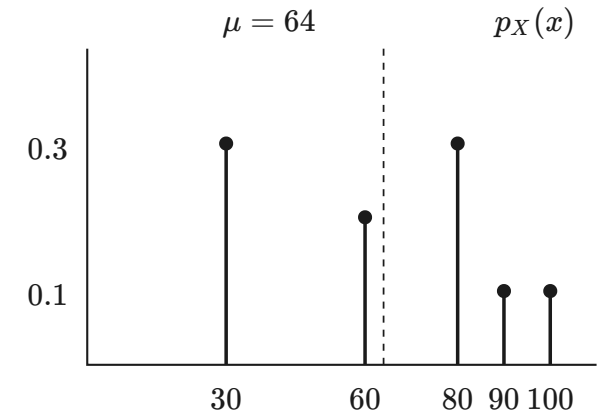
NEEDS $E[|X|] < \infty$.

WORKED EXAMPLE *the ten scripts*

$$E[X] = 30(0.3) + 60(0.2) + 80(0.3) + 90(0.1) + 100(0.1) = 64$$

No script has mark 64, and the most likely marks are 30 and 80.

The mean need not be a supported value


WHY ECONOMISTS CARE

A payoff of 0 with probability 0.9 or 100 with probability 0.1 has mean 10 but mode 0.

The **expectation** of a 0/1 indicator is a proportion.

2.5 | TRANSFORMATIONS

Linear operations pass through expectations; nonlinear ones do not

RESULT *linearity*

$$E[a + bX] = a + bE[X], \quad E[X + Y] = E[X] + E[Y]$$

No independence needed. But $E[XY] = E[X]E[Y]$ only under independence.

RESULT *Jensen's inequality*

For convex g , $E[g(X)] \geq g(E[X])$; for concave g the reverse. For the marks: $E[X^2] = 4720$ but $(E[X])^2 = 4096$.

For the marks: $E[5 + 0.8X] = 5 + 0.8(64) = 56.2$.

COMMON MISTAKE

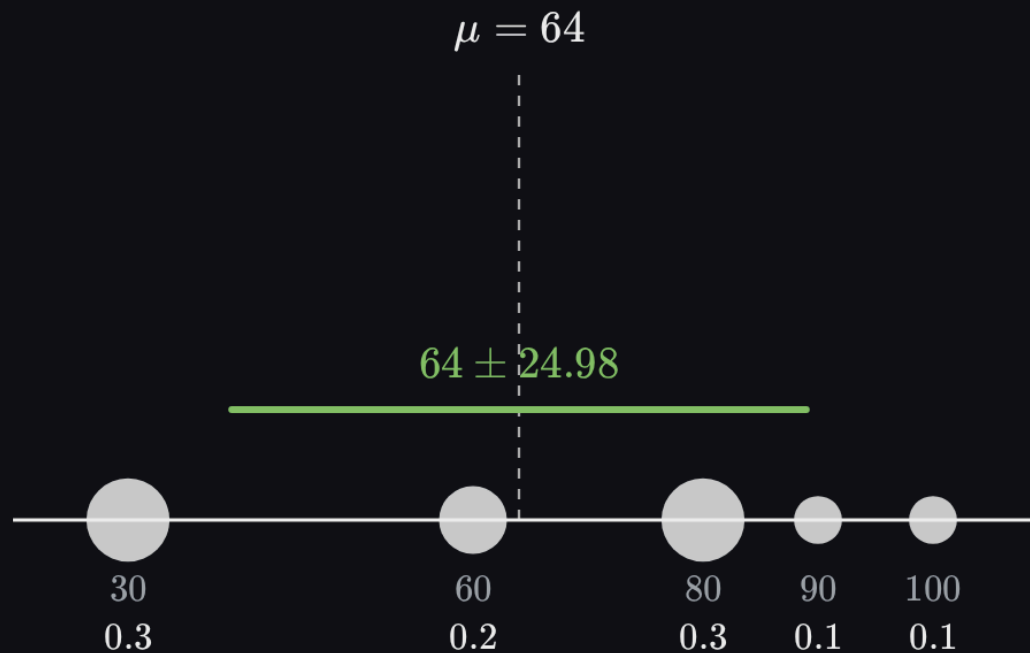
The mean of log wages is not the log of the mean wage, and averaging ratios is not the ratio of averages:

$$E[X/Y] = 3/4 \text{ while}$$

$$E[X]/E[Y] = 2/3 \text{ in the notes' two-point example.}$$

TAKEAWAY Changing units is safe; logs, squares and ratios are not. The gap $4720 - 4096$ is the variance.

Why square the deviations?

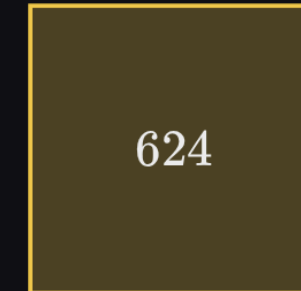


$\mu = 64$: the balance point

deviations: $-34, -4, 16, 26, 36$

weighted, they cancel: $E[X - \mu] = 0$

squares: $346.8 + 3.2 + 76.8 + 67.6 + 129.6$



$$\text{Var}(X) = 624 \text{ marks}^2$$

$$\text{SD}(X) = \sqrt{624} \approx 24.98 \text{ marks}$$

2.6 | SPREAD

Variance averages squared deviations; the standard deviation returns to the original units

KEY FORMULA DEFINITION

(5)

Variance

$$\text{Var}(X) = E[(X - \mu)^2] = E[X^2] - \mu^2$$

IN WORDS the average squared distance from the mean; the **standard deviation** is its square root.

RESULT *changing units*

$$\text{Var}(a + bX) = b^2 \text{Var}(X), \quad \text{SD}(a + bX) = |b| \text{SD}(X)$$

WORKED EXAMPLE *the ten scripts*

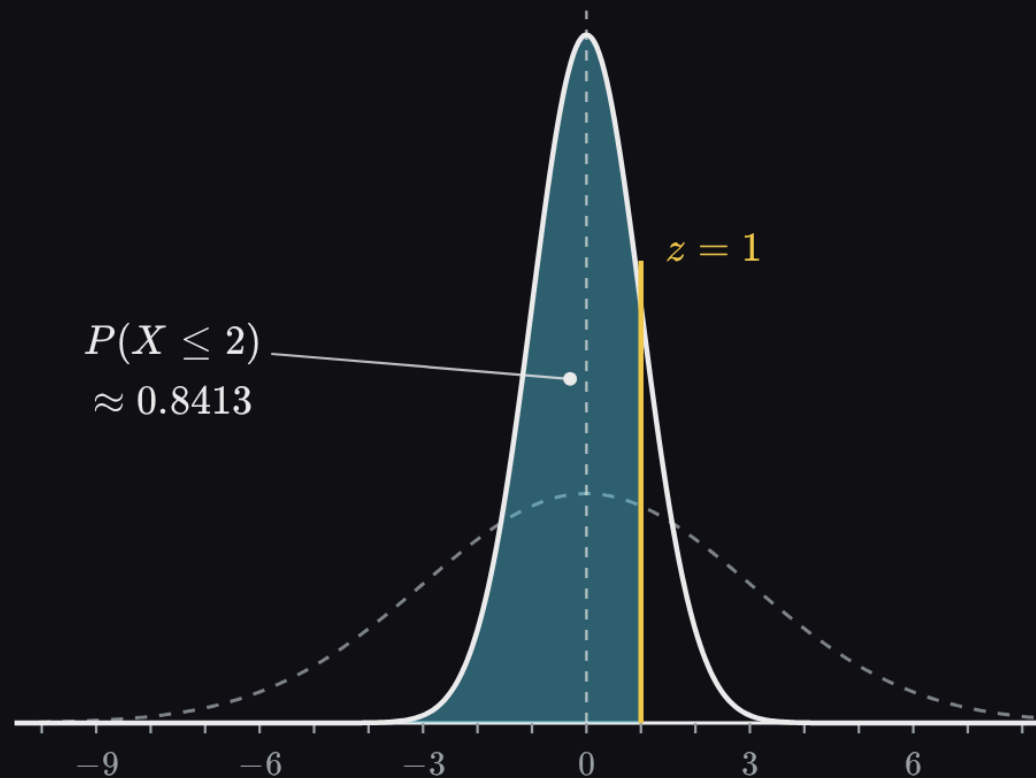
$\text{Var}(X) = 4720 - 4096 = 624$ marks², so $\text{SD} = \sqrt{624} \approx 24.98$ marks. For $10 - 2X$: mean -118 , $\text{SD} \approx 49.96$.

DEFINITION

A 0/1 variable

Mean p , **variance** $p(1 - p)$: the mean fixes the spread, at most $\frac{1}{4}$. An average of N independent indicators has variance $p(1 - p)/N$.

Standardising: what changes, and what does not?



$$X \sim N(-1, 9): \sigma = \sqrt{9} = 3, \text{ not } 9$$

subtract $\mu = -1$: slide by $+1$

$$X - \mu \sim N(0, 9); \text{ threshold } 2 - (-1) = 3$$

divide by $\sigma = 3$: width $\div 3$, height $\times 3$

$$Z \sim N(0, 1); \text{ threshold } 3/3 = 1$$

$$P(X \leq 2) = P(Z \leq 1) = \Phi(1) \approx 0.8413$$

Same event, same area.

2.7 | ONE TABLE FOR EVERY NORMAL

One table serves every normal: standardise, look up Φ , translate back

KEY FORMULA RESULT

(6)

Normal probabilities and quantiles

$$P(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad x_p = \mu + \sigma z_p$$

IN WORDS Φ is the standard normal CDF; z_p its quantile. Divide by σ , never σ^2 .

NEEDS $X \sim N(\mu, \sigma^2)$ with $\sigma > 0$.

WORKED EXAMPLE a right tail

$X \sim N(0.2, 1.5)$, so $\sigma = \sqrt{1.5}$:

$$z = (1.6 - 0.2) / \sqrt{1.5} \approx 1.143$$

$$P(X \geq 1.6) = 1 - \Phi(1.143) \approx 0.1265$$

COMMON MISTAKE

A central 95% interval leaves 2.5% in each tail, so it uses ± 1.96 , the 97.5th percentile, not the 95th (1.645). For $N(-1, 9)$: $-1 \pm 3(1.96) = [-6.88, 4.88]$.

Selected values of $\Phi(z)$

z	-1	0	1	1.96	2
$\Phi(z)$	0.1587	0.5	0.8413	0.9750	0.9772

TAKEAWAY The mean is the balance point, the variance the average squared distance; $(x - \mu)/\sigma$ puts any normal on one table.

PART 3 OF 3

Relate: joint, conditional and association

What does knowing one variable tell us about another?

KEY TERMS

joint

marginal

conditional mean

covariance

correlation

[Notes 2.8–2.10 ↗](#)

SO FAR the mean and variance summarise one variable; for a normal model $(x - \mu)/\sigma$ puts any question on one table.

WHY IT MATTERS

Most economic questions compare groups or relate two measurements, such as earnings by gender or age. Conditional means are what regression estimates; covariance sets the spread of sums. None of these is, by itself, a causal effect.

IN THIS PART

- 2.8 Margins by adding, conditionals by normalising
- ▶ Animation: how group means combine into the class mean
- 2.9 Group averages and iterated expectation
- 2.10 Covariance and correlation
- 2.10 Zero correlation is not independence

2.8 | JOINT, MARGINAL, CONDITIONAL

From a joint table: add to get a margin, normalise a slice to condition

The class: mark Y by group X (1 women, 0 men)

Y	$X = 0$	$X = 1$	margin
30	0.3	0	0.3
60	0	0.2	0.2
80	0.2	0.1	0.3
90	0	0.1	0.1
100	0.1	0	0.1
margin	0.6	0.4	1

KEY FORMULA DEFINITION

(7)

Marginal and conditional

$$p_X(x) = \sum_y p_{X,Y}(x, y), \quad p_{Y|X}(y | x) = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

IN WORDS add over the other variable; or fix x and rescale the slice to total 1.

$P(Y = 80 | X = 1) = 0.1/0.4 = \frac{1}{4}$ but $P(Y = 80 | X = 0) = 0.2/0.6 = \frac{1}{3}$:
the marks depend on group.

COMMON MISTAKE

The numerator is the joint cell, 0.1, not the mark's margin, 0.3.

How do the group means give the class mean?



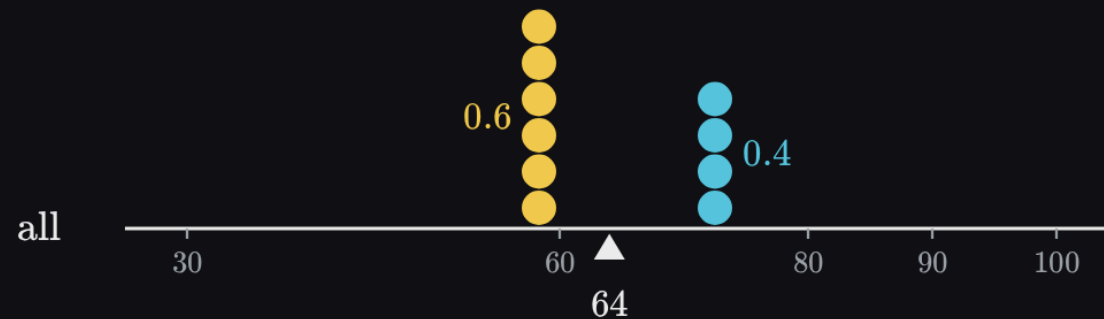
$$E[Y \mid X = 1] = 290/4 = 72.5$$

$$P(X = 1) = 4/10 = 0.4$$



$$E[Y \mid X = 0] = 350/6 \approx 58.33$$

$$P(X = 0) = 6/10 = 0.6$$



$$E[Y] = 640/10 = 64$$

$$E[E[Y \mid X]] = 0.4(72.5) + 0.6(175/3) \\ = 29 + 35 = 64$$

$$E[Y] = E[E[Y \mid X]]$$

$$\text{not } (72.5 + 58.33)/2 \approx 65.42$$

2.9 | CONDITIONAL EXPECTATION

Group means recombine into the overall mean with the group shares as weights

DEFINITION

Conditional mean

$m(x) = E[Y \mid X = x]$: the average of Y within the group $X = x$. Women: $290/4 = 72.5$. Men: $350/6 \approx 58.33$.

KEY FORMULA RESULT

Law of iterated expectation

$$E[Y] = E[E[Y \mid X]]$$

IN WORDS average within groups, then average the group means using the group probabilities.

(8)

WORKED EXAMPLE *the class*

$$E[Y] = 0.4(72.5) + 0.6 \frac{175}{3} = 29 + 35 = 64$$

Equal weights would give 65.42: the mean of a half-and-half class.

WHY ECONOMISTS CARE

A gap in conditional means, such as average earnings by recorded gender, describes an association. It does not by itself measure what changing gender would cause (Lecture 5).

2.10 | MOVING TOGETHER

Covariance measures whether deviations move together

KEY FORMULA DEFINITION

(9)

Covariance and correlation

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

IN WORDS the average product of deviations; correlation removes the units and lies in $[-1, 1]$.

RESULT *variance of a sum*

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2 \text{Cov}(X, Y)$$

Products of deviations determine the sign

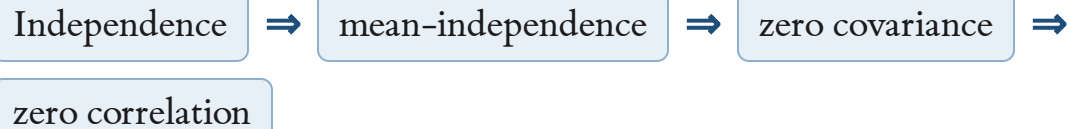
$Y - \mu_Y$	
$(-)(+) < 0$	$(+)(+) > 0$
$(-)(-) > 0$	$(+)(-) < 0$
$X - \mu_X$	

WORKED EXAMPLE *the class*

$E[XY] = 29$, so $\text{Cov} = 29 - 0.4(64) = 3.4$
and $\rho \approx 0.278$: a descriptive association.

2.10 | INDEPENDENCE AND CORRELATION

Zero correlation is not independence: the implications run one way

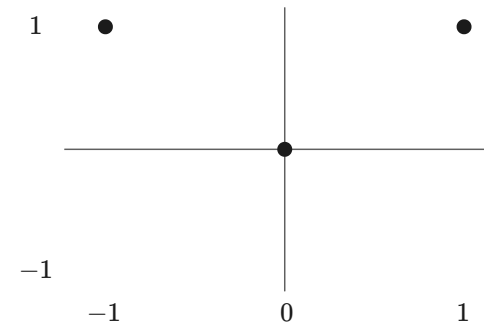


None of the reverse implications holds, and none of these is a causal statement.

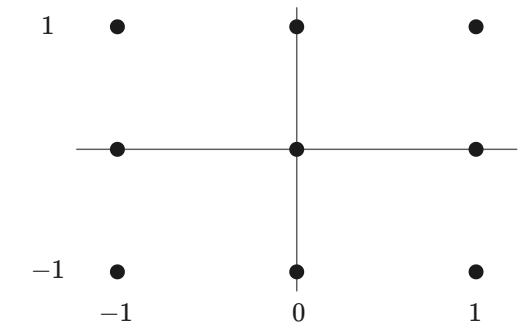
WORKED EXAMPLE $X \in \{-1, 0, 1\}$ *equally likely*, $Y = X^2$
 $E[XY] = 0 = E[X]E[Y]$, so $\text{Cov} = 0$. Yet $X = 0$ forces $Y = 0$: $P(Y = 0 \mid X = 0) = 1 \neq \frac{1}{3}$.

Zero correlation can arise from different joint models

Deterministic quadratic



Independent values



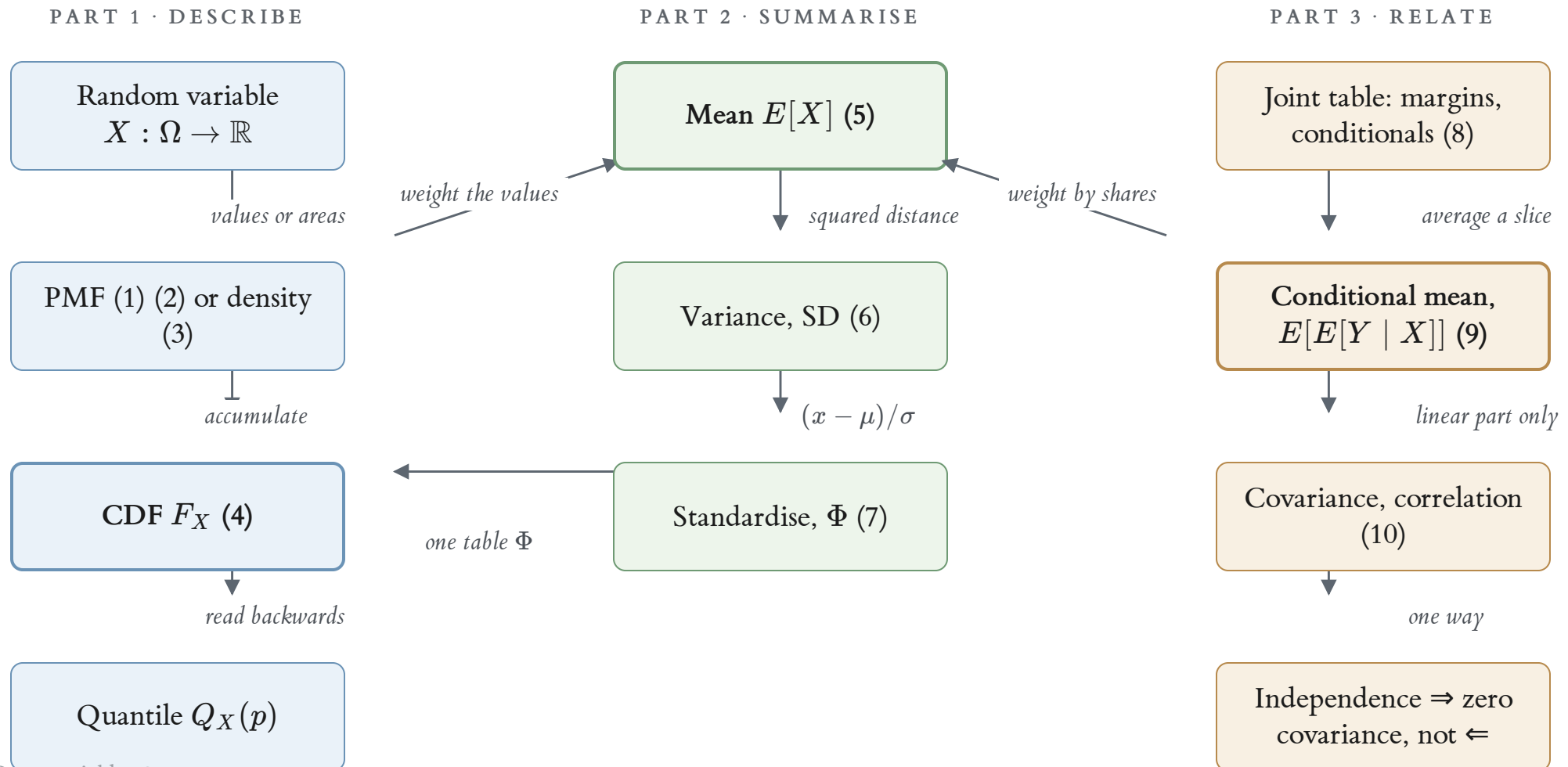
WHY IT MATTERS

“Uncorrelated” does not mean “unrelated”: correlation captures only linear association.

TAKEAWAY Margins add, conditionals normalise, group means recombine with shares; correlation is linear association only, and not a cause.

REVIEW · HOW THE IDEAS CONNECT

One probability model answers three questions: describe, summarise, relate



REVIEW · SUMMARY SHEET

What to take away: the terms and the ten formulas

KEY TERMS

1 DESCRIBE

Random variable $X : \Omega \rightarrow \mathbb{R}$: a number for each outcome.

Probability mass function (PMF) $P(X = x)$ for each value; the masses add to 1.

Density probability per unit; probabilities are areas under it.

Quantile $Q_X(p)$: the lowest threshold with $F_X \geq p$.

2 SUMMARISE

Expectation the probability-weighted average of the values.

Variance the average squared deviation; its square root, the SD, is in the original units.

Standardised variable $(X - \mu)/\sigma$: mean 0, variance 1, same shape.

3 RELATE

Joint distribution probabilities for pairs measured on the same unit.

Conditional mean the average of Y within one group.

Covariance the average product of deviations; correlation removes its units.

KEY FORMULAS

1 DESCRIBE

(1) **Binomial distribution** RESULT

$$P(Y = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$
N independent trials, common p

(2) **Probability from a density** DEFINITION

$$P(a < X \leq b) = \int_a^b f_X(x) dx$$
X has a density; $P(X = x) = 0$

(3) **Cumulative distribution function** DEFINITION

$$F_X(x) = P(X \leq x)$$
intervals: $F_X(b) - F_X(a)$; endpoints matter if discrete

2 SUMMARISE

(4) **Expected value** DEFINITION

$$E[X] = \sum_x x p_X(x) \quad \text{or} \quad \int x f_X(x) dx$$

$$E[|X|] < \infty$$

(5) **Variance** DEFINITION

$$\text{Var}(X) = E[(X - \mu)^2] = E[X^2] - \mu^2$$
a finite second moment

(6) **Normal probabilities and quantiles** RESULT

$$P(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad x_p = \mu + \sigma z_p$$

$$X \sim N(\mu, \sigma^2), \sigma > 0$$

3 RELATE

(7) **Marginal and conditional** DEFINITION

$$p_{Y|X}(y | x) = p_{X,Y}(x, y) / p_X(x)$$

$$p_X(x) > 0; \text{ margins add over the other variable}$$

(8) **Law of iterated expectation** RESULT

$$E[Y] = E[E[Y | X]]$$
Y integrable

(9) **Covariance and correlation** DEFINITION

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

$$\rho = \text{Cov} / (\sigma_X \sigma_Y) \text{ needs both SDs positive}$$

REVIEW · BACK TO THE OPENING QUESTIONS

Both answers are 64: a balance point, and a share-weighted average

RESULT *the opening questions*

(a) $E[X] = 64$: no script has it, and it is not the most likely mark.

(b) Not 65.42: each student counts once, so $0.4(72.5) + 0.6(175/3) = 64$.

REMEMBER

1. A density's height is not a probability; the mean need not be a possible value.
2. Linear operations pass through expectations; squares, logs and ratios do not.
3. Group means need group shares as weights; zero correlation is not independence.

BEFORE NEXT TIME

- Notes 2.1–2.10 with the worked solutions.
- The binomial, normal, means and association figures in the notes.
- Practise the normal table: sketch, shade, standardise.

NEXT · LECTURE 3

Statistics. We know only a sample: Φ , central intervals and $p(1 - p)/N$ return as standard errors, intervals and tests.