ALLAMA IQBAL OPEN UNIVERSITY

USMAN NOTES
www.usmannotes.com

Name:

Registration No.:

Course Code: 804

Program: MSc Economics

Semester: Autumn 2026

ALLAMA IQBAL OPEN UNIVERSITY, ISLAMABAD

Course: Statistics for Economists (804) • Level: MSc Economics

Semester: Autumn 2026

Assignment No. 1 (Units 1–4)

SOLVED ASSIGNMENT GUIDE

(For Study & Reference)

Note: This guide is prepared for study and reference purposes. Students are advised to read the relevant units of the course book, understand the concepts, and then write the assignment in their own words. Total marks of the assignment are 100; each question carries the marks shown with it.

USMAN NOTES www.usmannotes.com

Question 1: Subjective Probability and Symmetric Probability (20 Marks)

Explain the following: (a) Subjective Probability (10); (b) Symmetric Probability (10).

Introduction

Probability is the language in which economics deals with uncertainty. Almost every economic decision — whether a farmer deciding how much wheat to sow, a firm deciding how much stock to hold, a bank deciding whether to approve a loan, or a government forecasting next year's tax revenue — is taken before the outcome is known, so it must be based on some judgement about how likely different outcomes are. Statistics for economists therefore begins by asking a deceptively simple question: what exactly do we mean when we say that an event has a certain probability? There is no single answer. Over three centuries, mathematicians, philosophers and economists have developed several interpretations of probability, each suited to a different kind of problem. The classical interpretation works when outcomes are symmetric and equally likely, as in coins and dice. The relative-frequency interpretation works when an experiment can be repeated many times under identical conditions, as in quality control or

insurance. But many of the most important economic events — a devaluation next year, a new product succeeding, a war affecting oil prices — can neither be broken into equally likely cases nor repeated as identical trials. For such events we use the subjective interpretation. This answer explains subjective probability in detail, and then explains symmetric (classical) probability, the oldest interpretation, on which the equally-likely-cases calculations of elementary statistics rest. Understanding both, and the contrast between them, is essential before an economist can use probability responsibly.

Part (a): Subjective Probability

Subjective probability is the degree of belief, confidence or conviction that a particular person holds about the occurrence of an event, expressed as a number between 0 and 1. It is called subjective because it belongs to a subject — a person — rather than to the objective physical set-up of an experiment. Two equally well-informed people may honestly assign different probabilities to the same event, for example to the event that the State Bank of Pakistan will cut the policy rate at its next meeting, because they weigh the available evidence, their experience and their reading of the situation differently. Neither is necessarily wrong; each number faithfully reports that person's state of information and judgement at that moment.

The interpretation was developed formally by Frank Ramsey (1926), Bruno de Finetti (1937) and Leonard Savage (1954), who showed that subjective probabilities are not mere guesses. If a person's judgements are coherent — that is, if they do not contradict one another in ways that would allow a bookmaker to make a guaranteed profit against them, the so-called Dutch-book condition — then those judgements must obey the ordinary axioms of probability: probabilities lie between 0 and 1, the probability of a certain event is 1, and the probability of one of two

mutually exclusive events is the sum of their separate probabilities. Subjective probability is therefore genuine probability in the mathematical sense; what is personal is the input, not the rules for combining it.

A practical way of measuring a subjective probability is through betting behaviour or its business equivalent. If a manager is just willing to pay Rs 30 for a contract that pays Rs 100 if a new branch becomes profitable within a year and unwilling to pay Rs 35, we infer that his subjective probability of success is roughly 0.30. Modern forecasting practice uses the same idea in structured forms: expert elicitation, prediction markets, scenario probabilities in project appraisal, and the prior distributions of Bayesian statistics are all, at bottom, disciplined subjective probabilities. In Bayesian inference the subjective element is made fully explicit: the analyst states a prior probability, combines it with sample evidence through Bayes' theorem, and obtains a posterior probability that can be used for decisions.

Subjective probability is indispensable in economics for at least four reasons. First, most strategic economic events are unique and non-repeatable: a firm enters a particular market once, a budget is passed once, a currency is devalued on a particular date. No long-run frequency exists for such events, so the frequency interpretation is silent about them, yet decisions about them cannot wait. Second, economic agents demonstrably behave as if they hold such beliefs: insurance purchases, portfolio choices, inventory orders and wage bargains all reveal implicit probabilities. Expected-utility theory, the standard model of choice under uncertainty, takes subjective probabilities as one of its two basic ingredients, alongside utilities. Third, subjective probability allows all available information — theory, casual observation, expert knowledge, and partial data — to be used, instead of discarding everything except

formal sample frequencies. Fourth, it gives a rational way of updating: as new evidence arrives, beliefs are revised by Bayes' rule rather than by whim.

The interpretation also has clear limitations, which an economist should state honestly. Because it is personal, it cannot be verified by experiment in the way a frequency can; a stated probability of 0.70 for rain tomorrow is not refuted by either rain or sunshine. Different experts may differ widely, and human judgement is subject to well-documented biases — overconfidence, anchoring on the first number mentioned, and the availability bias that makes vivid events seem more likely than dull ones. For this reason, good practice disciplines subjective assessment: probabilities for a complete set of mutually exclusive outcomes are required to sum to one, assessors are asked for reasons and base rates, several experts are combined, and forecasts are scored afterwards for calibration. Used in this disciplined way, subjective probability is not an excuse for arbitrariness but the only coherent language available for one-off economic decisions.

Part (b): Symmetric Probability

Symmetric probability — better known as the classical or Laplacean definition — applies when the possible outcomes of an experiment are symmetric, that is, when there is no reason to expect any one of them rather than another, so that they may be treated as equally likely. If a random experiment has n mutually exclusive, exhaustive and equally likely outcomes, and m of them are favourable to an event A, then P(A) = m / n. The word symmetric describes the structure of the experiment itself: a fair coin is physically symmetric between its two faces, a fair die is a symmetric cube with six identical faces, and a well-shuffled pack is symmetric

among its 52 cards. Because of this symmetry, we can assign the probability before any trial is performed, which is why the classical probability is also called a priori probability.

The standard illustrations follow immediately. A fair coin gives P(head) = 1/2. A fair die gives P(six) = 1/6 and P(even number) = 3/6 = 1/2. A card drawn from a shuffled pack gives P(ace) = 4/52 = 1/13. The same reasoning underlies the coin and digit problems solved in Question 2 of this assignment: three fair coins generate 2 × 2 × 2 = 8 symmetric outcomes, and the permutations of three distinct digits generate 3! = 6 symmetric arrangements, so every probability there is a count of favourable cases divided by 8 or by 6 respectively. Combinatorics — permutations and combinations — is the counting machinery that the classical definition needs once experiments become complex, as in lottery problems, sampling without replacement, and the hypergeometric situations that arise when an inspector draws a sample from a finite lot.

The classical definition has real strengths. It is precise, objective and reproducible: everyone who accepts the symmetry of the set-up must accept the same numbers. It built the first successful mathematical theory of probability in the hands of Pascal, Fermat and Laplace, and it remains exactly correct for the deliberately randomised devices on which statistical sampling theory is built — random number tables, lottery draws and randomised experiments are engineered to be symmetric, so that classical probabilities apply to them by construction. Random sampling from a population, the foundation of the inference methods used later in this course, is nothing but an artificial symmetry imposed by the sampler: every unit is given an equal chance of selection.

Its limitations are equally clear. First, the definition is in a sense circular: "equally likely" itself means "equally probable", so probability is defined in terms of probability unless symmetry is understood physically, as a property of the coin or die. Second, most economic outcomes are not symmetric. There is no symmetry between a borrower repaying and defaulting, between a harvest being good and bad, or between the possible growth rates of GDP next year; the number of "cases" is arbitrary and they cannot plausibly be counted as equal. Third, the definition fails when the number of outcomes is infinite, and it can be manipulated by re-describing the same situation with a different set of cases. Where symmetry fails, the economist must turn to frequencies where repetition is possible, and to subjective probability where it is not.

A Comparison of the Two Interpretations

Table 1: Subjective and symmetric (classical) probability compared

Feature	Subjective probability	Symmetric (classical) probability
Basis	Personal degree of belief, disciplined by coherence	Physical symmetry of equally likely outcomes
Formula / source	Elicited from judgement, betting equivalents or priors	P(A) = favourable cases / total cases
Suitable events	Unique, non-repeatable events (policy changes, new ventures)	Coins, dice, cards, randomised sampling
Objectivity	Differs between persons; depends on information	Same for everyone who accepts the symmetry
Verification	Checked for calibration over many forecasts	Checked against long-run frequencies
Main weakness	Bias and overconfidence in judgement	Inapplicable when outcomes are not equally likely

The two interpretations are complements, not rivals. A complete statistician uses symmetric probability wherever genuine or constructed symmetry exists, frequency evidence wherever repetition is available, and subjective probability — openly stated and coherently combined — for the large remainder of economic life where a decision must be made on a unique event. What the axioms of probability provide is the common grammar that lets all three be calculated with, compared and combined.

Conclusion

To conclude, subjective probability is a coherent numerical measure of a person's belief about an uncertain event, governed by the same axioms as any other probability but personal in its inputs; it is unavoidable, and immensely useful, for the one-off decisions that dominate economic life, provided it is elicited carefully and checked for coherence and calibration. Symmetric probability is the classical definition P(A) = m/n, valid when the outcomes of an experiment are symmetric and equally likely, as with fair coins, dice, cards and randomised samples; it is exact and objective within that domain, but the domain is narrow and most economic events fall outside it. The wise economist knows which interpretation a problem calls for — and says so — before reaching for a number.

Question 2: Sample Spaces, Events and Partitions (20 Marks)

(a) List the sample space of three coins tossed simultaneously. Define events A: all coins are the same; B: precisely 1 head; C: at least 2 heads. Evaluate P(A) + P(B) + P(C). Do these events form a partition? (10) (b) A three-digit number is formed by arranging the digits 1, 5, and 6 in a random order. (i) List the sample space. (ii) Find the probability of getting a number larger than 250. (iii) What is the probability that an even number is obtained? (10)

Introduction

Every probability calculation begins with the sample space: the complete list of outcomes that a random experiment can produce. If the sample space is listed correctly and the outcomes are equally likely — as they are for fair coins and for random arrangements of distinct digits — then every question reduces to careful counting: the probability of an event is the number of outcomes inside it divided by the number of outcomes in the whole space. This question practises that discipline twice, once with coins and once with digits, and then asks a subtler conceptual question: whether three given events form a partition of the sample space. A partition is a family of events that are pairwise mutually exclusive and jointly exhaustive — their probabilities must then sum exactly to one — so the question tests whether the student can check both conditions rather than merely compute.

Part (a): Three Coins Tossed Simultaneously

Each coin can fall head (H) or tail (T), so three coins produce $2 \times 2 \times 2 = 8$ outcomes. Tossing the coins simultaneously rather than one after another makes no difference to the list, provided we keep track of which outcome is which — for example by imagining the coins are of different denominations, or by reading them in a fixed left-to-right order. The sample space is S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}, and because the coins are fair, each of the eight outcomes has probability 1/8. Event A, “all coins are the same”, contains the two outcomes in which the coins agree completely: A = {HHH, TTT}, so P(A) = 2/8 = 0.25. Event B, “precisely 1 head”, contains the three outcomes in which the single head falls in the first, second or third position: B = {HTT, THT, TTH}, so P(B) = 3/8 = 0.375. Event C, “at least 2 heads”, means two heads or three heads: C = {HHT, HTH, THH, HHH}, four outcomes, so P(C) = 4/8 = 0.50.

Adding the three probabilities gives P(A) + P(B) + P(C) = 0.25 + 0.375 + 0.50 = 1.125, or in fractions 2/8 + 3/8 + 4/8 = 9/8. The sum exceeds one, which is already a strong hint about the partition question. For a family of events to form a partition of S, two conditions must both hold: (i) the events must be pairwise mutually exclusive — no outcome may belong to two of them; and (ii) they must be exhaustive — every outcome of S must belong to at least one of them, so their union is S. Events A, B and C fail the first condition. The outcome HHH belongs both to A (all coins the same) and to C (at least two heads), so A ∩ C = {HHH} ≠ ∅; the events are not mutually exclusive. That single overlap is exactly why the probabilities sum to 9/8 rather than 8/8: the outcome HHH, worth 1/8, has been counted twice. (Exhaustiveness also deserves a check: running through the list, TTT is in A, the three one-head outcomes are in B, and the three two-head outcomes together with HHH are in C, so the union A ∪ B ∪ C does cover all eight outcomes of S. It is the exclusivity condition alone that fails.) Hence A, B and C do NOT form a partition of the sample space.

Table 1: Events for three coins and their probabilities

Event	Outcomes	Number of outcomes	Probability
A: all coins the same	HHH, TTT	2	2/8 = 0.250
B: precisely 1 head	HTT, THT, TTH	3	3/8 = 0.375
C: at least 2 heads	HHT, HTH, THH, HHH	4	4/8 = 0.500
Sum	— (HHH counted twice)	9 counts	9/8 = 1.125

For contrast, a genuine partition of the same sample space is given by the four events “no head” {TTT}, “exactly 1 head” {HTT, THT, TTH}, “exactly 2 heads” {HHT, HTH, THH} and “exactly 3 heads” {HHH}: they are mutually exclusive, they exhaust S, and their probabilities 1/8 + 3/8 +

3/8 + 1/8 sum exactly to one, as the law of total probability requires. Partitions matter in practice because they are the scaffolding of the law of total probability and of Bayes' theorem: an event of interest is broken across the cells of a partition, its probability is computed cell by cell, and the pieces are added — a procedure that is only valid because no outcome is counted twice or omitted.

Part (b): Three-Digit Numbers from the Digits 1, 5 and 6

(i) Arranging three distinct digits in a random order gives 3! = 3 × 2 × 1 = 6 equally likely numbers. The sample space is S = {156, 165, 516, 561, 615, 651}. Each number has probability 1/6. (ii) A number larger than 550 must have hundreds digit 5 or 6, because any number beginning with 1 is at most 165, while every number beginning with 5 or 6 exceeds 500. The favourable outcomes are {516, 561, 615, 651} — four of them — so the required probability is 4/6 = 2/3 ≈ 0.667. (iii) A number is even exactly when its units digit is even. Among the digits 1, 5 and 6 only 6 is even, so the number is even exactly when 6 occupies the units place: the favourable outcomes are {156, 516}, two of them, and the probability is 2/6 = 1/3 ≈ 0.333. Both answers can be cross-checked by symmetry: by symmetry each digit is equally likely to stand in the hundreds place, so P(hundreds digit is 5 or 6) = 2/3, matching (ii); and each digit is equally likely to stand in the units place, so P(units digit is 6) = 1/3, matching (iii). Direct counting and the symmetry check agree, as they must.

Two lessons from this elementary exercise carry through the whole subject. First, "random order" is a symmetry assumption: it is what licenses the classical formula P = favourable/total, and when arrangements are not equally likely the counting formula fails. Second, most errors in probability are listing or counting errors, not conceptual ones; writing the sample space out in

full, as done here, is slow but safe, and for larger problems the same discipline is preserved by systematic counting rules — the multiplication rule, permutations and combinations.

Conclusion

In summary: for three coins, S has 8 outcomes; P(A) = 2/8, P(B) = 3/8 and P(C) = 4/8, summing to 9/8 = 1.125, and because HHH lies in both A and C the events are not mutually exclusive, so they do not form a partition. For the digits 1, 5 and 6, S = {156, 165, 516, 561, 615, 651}; the probability of a number larger than 250 is 4/6 = 2/3, and the probability of an even number is 2/6 = 1/3. The exercise shows why the two partition conditions — exclusivity and exhaustiveness — must be checked separately, and why a probability sum different from one is a diagnostic signal that an outcome has been double-counted or missed.

Question 3: Joint Probability, Independence and Covariance (20 Marks)

Describe the joint probability function with the help of a suitable example. Suppose X and Y have the following joint distribution: p(1,1) = 0.30, p(1,2) = 0.10, p(2,1) = 0.20, p(2,2) = 0.05, p(3,1) = 0.30, p(3,2) = 0.05. (a) Find p(x) and p(y); then by verifying that p(x)p(y) = p(x, y) confirm that X and Y are independent. (10) (b) What is σXY (the covariance of X and Y)? (10)

Introduction

Economic variables rarely move alone. Income and consumption, price and quantity demanded, education and earnings, rainfall and crop yield: the interesting questions are almost always about two or more variables together — how they are distributed jointly, whether knowing one tells us anything about the other, and how strongly they move together. The joint probability function is the tool that describes two random variables at once, and from it flow the marginal

distributions, the test for independence, and the covariance, all of which this question exercises on one small distribution. Working a tiny example completely by hand is worthwhile precisely because every formula of bivariate analysis appears in it in its simplest form.

The Joint Probability Function

For two discrete random variables X and Y, the joint probability function p(x, y) assigns to every pair of values the probability that both occur together: p(x, y) = P(X = x and Y = y). It must satisfy two conditions: every value lies between 0 and 1, and the probabilities over all pairs sum to exactly 1, because one of the pairs must occur. A joint distribution is usually written as a two-way table, with the values of X down the side, the values of Y across the top, and the joint probabilities in the cells. From the joint function we recover what each variable does on its own by summing across the other variable: the marginal probability function of X is p(x) = Σy p(x, y) (a row sum), and the marginal of Y is p(y) = Σx p(x, y) (a column sum). Two variables are independent when the joint probability factorises for every pair: p(x, y) = p(x) p(y) for all x and y. Independence means that learning the value of one variable changes nothing about the probabilities of the other — the strongest possible statement of "no relationship". If factorisation fails for even one pair, the variables are dependent.

A suitable example makes the definition concrete. Suppose a small training institute records, for each trainee, X = the number of computer courses completed (1, 2 or 3) and Y = whether the trainee found related employment within six months, coded Y = 1 for "not yet" and Y = 2 for "yes". A survey of past trainees, expressed in proportions, might give exactly the distribution of this question: 30 percent of trainees completed one course and had not found work, 10 percent completed one course and had found work, and so on. The joint function then answers joint

questions at a glance — the probability that a trainee completed three courses and found work is p(3, 2) = 0.05 — and, after the marginals are computed, it lets us test whether employment is independent of the number of courses, a question of obvious interest to the institute.

Part (a): Marginals and the Independence Test

Summing each row of the table gives the marginal distribution of X: p(1) = 0.30 + 0.10 = 0.40; p(2) = 0.20 + 0.05 = 0.25; p(3) = 0.30 + 0.05 = 0.35. The row probabilities sum to 0.40 + 0.25 + 0.35 = 1.00, as they must. Summing each column gives the marginal distribution of Y: p(y = 1) = 0.30 + 0.20 + 0.30 = 0.80; p(y = 2) = 0.10 + 0.05 + 0.05 = 0.20, again summing to 1.00. The full table with its margins is shown below.

Table 1: Joint distribution of X and Y with marginal distributions

x \ y	y = 1	y = 2	p(x) — row total
1	0.30	0.10	0.40
2	0.20	0.05	0.25
3	0.30	0.05	0.35
p(y) — column total	0.80	0.20	1.00

Now test the factorisation p(x) p(y) = p(x, y) cell by cell. For (x, y) = (1, 1): p(1) p(1) = 0.40 × 0.80 = 0.32, but p(1, 1) = 0.30 — not equal. For (1, 2): 0.40 × 0.20 = 0.08, but p(1, 2) = 0.10 — not equal. For (2, 1): 0.25 × 0.80 = 0.20, and p(2, 1) = 0.20 — equal. For (2, 2): 0.25 × 0.20 = 0.05, and p(2, 2) = 0.05 — equal. For (3, 1): 0.35 × 0.80 = 0.28, but p(3, 1) = 0.30 — not equal. For (3, 2): 0.35 × 0.20 = 0.07, but p(3, 2) = 0.05 — not equal. The factorisation holds for the middle row only and fails for four of the six cells. The honest conclusion is therefore the opposite of the one the question anticipates: X and Y are NOT independent; they are dependent. A single failing cell

would already be enough to prove dependence, because independence is an all-cells requirement. The discrepancy is instructive rather than embarrassing: it is exactly the kind of check a statistician must perform rather than assume, and the pattern of the failures is informative — relative to independence, probability has shifted away from the (1, 1) and (3, 2) cells towards (1, 2) and (3, 1), the first hint of the negative association measured by the covariance in part (b). In the institute example, the dependence means the number of courses completed is associated with employment status, although the direction and size of that association need the covariance, not just the independence test, to describe.

Table 2: Independence check — p(x)p(y) against p(x, y)

(x, y)	p(x) p(y)	p(x, y)	Equal?
(1, 1)	0.32	0.30	No
(1, 2)	0.08	0.10	No
(2, 1)	0.20	0.20	Yes
(2, 2)	0.05	0.05	Yes
(3, 1)	0.28	0.30	No
(3, 2)	0.07	0.05	No

Part (b): The Covariance σXY

The covariance measures the direction and strength of the linear co-movement of X and Y: σXY = E(XY) − E(X) E(Y), where each expectation is computed from the distributions above. First, E(X) = 1(0.40) + 2(0.25) + 3(0.35) = 0.40 + 0.50 + 1.05 = 1.95. Second, E(Y) = 1(0.80) + 2(0.20) = 0.80 + 0.40 = 1.20. Third, E(XY) is computed from the joint distribution by multiplying each pair's product by its probability: E(XY) = (1)(1)(0.30) + (1)(2)(0.10) + (2)(1)(0.20) + (2)(2)(0.05) +

(3)(1)(0.30) + (3)(2)(0.05) = 0.30 + 0.20 + 0.40 + 0.20 + 0.90 + 0.30 = 2.30. Therefore σXY = 2.30 − (1.95)(1.20) = 2.30 − 2.34 = − 0.04.

The result ties the whole question together. Independence would have forced the covariance to be exactly zero, because under independence E(XY) = E(X)E(Y). The covariance is not zero — it is − 0.04 — which independently confirms the dependence found in part (a); the two calculations are consistent witnesses. Its sign is negative: larger values of X tend slightly to go with smaller values of Y. Its size is small: the co-movement is weak, which is why four cells of the independence table miss equality only by 0.02. Two warnings complete the picture. First, the converse of the covariance test is false: a zero covariance does not prove independence, since variables can be strongly related in a curved pattern with zero covariance; only the full factorisation test of part (a) settles independence. Second, covariance is measured in the product of the units of X and Y and so cannot be compared across problems; dividing by the two standard deviations gives the correlation coefficient, the unit-free version developed later in the course. For the record, $Var(X) = E(X^2) - [E(X)]^2 = [1(0.40) + 4(0.25) + 9(0.35)] - 1.95^2 = 4.55 - 3.8025 = 0.7475$, and $Var(Y) = [1(0.80) + 4(0.20)] - 1.20^2 = 1.60 - 1.44 = 0.16$, so the correlation is $-0.04 / \sqrt{(0.7475 \times 0.16)} \approx -0.116$, a weak negative correlation — fully consistent with a small negative covariance and with mild, not dramatic, dependence.

Conclusion

To sum up: the joint probability function p(x, y) describes two variables together, its row and column sums give the marginals p(x) = (0.40, 0.25, 0.35) and p(y) = (0.80, 0.20), and the factorisation test shows that p(x)p(y) ≠ p(x, y) in four of the six cells, so X and Y are dependent, not independent as the question presupposed — a result the student should report as found,

not as expected. The covariance confirms it from a second direction: σXY = E(XY) − E(X)E(Y) = 2.30 − 2.34 = − 0.04, a small negative value indicating weak negative association. The question is a model of good statistical practice: compute the marginals, test the claim cell by cell, and let the arithmetic overrule the presumption.

Question 4: The Central Limit Theorem: Probability for Sample Means (20 Marks)

(a) Describe the central limit theorem with an example. (12) (b) The weights of packages filled by a machine are normally distributed about a mean of 25 ounces, with a standard deviation of one ounce. What is the probability that n packages from the machine will have an average weight of less than 28 ounces if n = 1, 4, 16, 64? (08)

Introduction

The central limit theorem (CLT) is routinely called the most important result in statistics, and for an economist it earns the title. Almost all inference in this course — confidence intervals, hypothesis tests, regression standard errors — works not because economic data are normally distributed (they usually are not: incomes are skewed, price changes have fat tails, counts are discrete) but because averages of data behave as if they were normal. The CLT is the theorem that guarantees this. Part (a) states it carefully and illustrates it; part (b) applies its sampling-distribution machinery to a filling machine whose output is already normal, which isolates the other great lesson of the theorem: the standard error shrinks like 1/√n, so averages concentrate around the mean far faster than individual observations do.

Part (a): Statement and Meaning of the Central Limit Theorem

Statement. Let X1, X2, …, Xn be a random sample of size n from any population with mean μ and finite variance σ². Then, as n grows large, the distribution of the sample mean $\bar{X}$ is approximately normal with mean μ and variance σ²/n, whatever the shape of the parent population. Equivalently, the standardised mean $Z = (\bar{X} - \mu) / (\sigma/\sqrt{n})$ tends to the standard normal distribution N(0, 1). Three features of the statement deserve emphasis. First, the theorem is about the mean (or sum) of the sample, not about the observations themselves: collecting more data does not make individual incomes normal, it makes averages of incomes approximately normal. Second, the centre and spread of the approximating distribution are exact, not asymptotic: $E(\bar{X}) = \mu$ for every n, and the standard error of the mean is $\sigma/\sqrt{n}$ for every n; only the normal shape is the large-sample gift of the theorem. Third, "large" is relative: for roughly symmetric parent populations n = 10 may suffice, the conventional rule of thumb is n ≥ 30, and for strongly skewed populations such as incomes or insurance claims a few hundred may be needed before the approximation is trustworthy in the tails.

An example makes the mechanism visible. A single fair die has a flat (uniform) distribution over {1, 2, 3, 4, 5, 6} with mean 3.5 and standard deviation about 1.71 — as far from a bell curve as a distribution can be. Roll two dice and average them, and the distribution of the average is already triangular, piled up in the middle because there are many ways to average 3.5 and only one way to average 1 or 6. Average ten dice and the distribution is a smooth bell centred on 3.5 with standard deviation $1.71/\sqrt{10} \approx 0.54$; average a hundred and the bell is indistinguishable from normal, with standard deviation 0.171. Nothing about the dice changed; averaging itself manufactures normality, because extreme averages require almost all the draws to be extreme

in the same direction at once, and that coincidence becomes vanishingly rare as n grows while moderate draws cancel each other's errors. An economic example runs in parallel: individual household incomes in Pakistan are heavily right-skewed, but the mean income of a random sample of 400 households is, by the CLT, approximately normal around the population mean with standard error σ/20, which is exactly why a survey such as the Household Integrated Economic Survey can attach a normal-based margin of error to its estimated mean income.

The theorem's consequences organise the rest of statistics. It justifies z- and t-tests and normal confidence intervals for means from non-normal data; it explains why least-squares estimates in large samples are treated as approximately normal even when regression errors are not; it underlies the sampling theory of proportions, since a proportion is a mean of 0–1 variables; and it explains the familiar 1/√n law by which quadrupling a sample only halves the standard error — the arithmetic behind the cost of precision in every survey budget. Its limits must also be respected: the population variance must be finite, the observations must be (at least approximately) independent or only weakly dependent, and the approximation concerns the centre of the distribution sooner than the extreme tails.

Part (b): Average Weight of n Packages

Here the parent population is normal with μ = 25 ounces and σ = 1 ounce, so the sample mean is exactly normal for every n, with standard error σ/√n = 1/√n, and we need $P(\bar{X} < 28) = P(Z < (28 - 25) / (1/\sqrt{n})) = P(Z < 3\sqrt{n})$. For n = 1 the standard error is 1, Z = 3.00, and P(Z < 3.00) = 0.9987: even a single package is almost certainly below 28 ounces, since 28 lies three full standard deviations above the mean. For n = 4 the standard error is 1/2 = 0.50, Z = 3/0.50 = 6.00, and P(Z < 6) ≈ 1.0000 (the shortfall is about one in a billion). For n = 16 the standard error is 0.25, Z = 12,

and the probability is 1.0000 to any practical accuracy. For n = 64 the standard error is 0.125, Z = 24, and again the probability is effectively 1. The pattern is the 1/√n law in action: the target 28 ounces does not move, but the sampling distribution squeezes tighter and tighter around 25, so the fixed point 28 retreats from 3 standard errors away to 6, 12 and 24 standard errors away, and the probability of falling short of it races to certainty.

Table 1: P(X̄ < 28) for n = 1, 4, 16, 64 (μ = 25, σ = 1)

n	Standard error σ/√n	Z = 3√n	P(X̄ < 28)
1	1.000	3.00	0.9987
4	0.500	6.00	≈ 1.0000
16	0.250	12.00	≈ 1.0000
64	0.125	24.00	≈ 1.0000

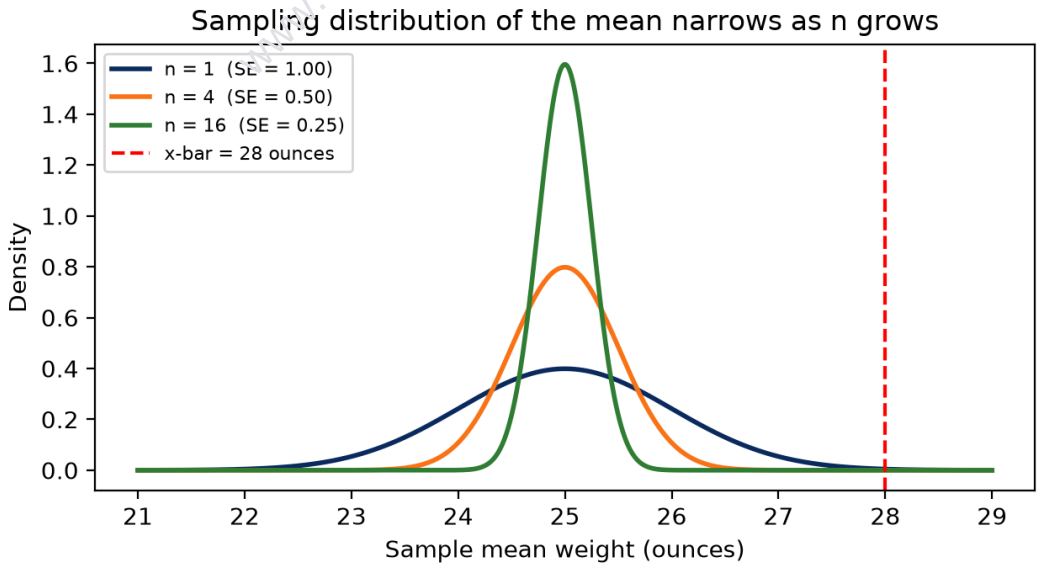


Figure 1: The sampling distribution of the mean concentrates around μ = 25 as n increases; the fixed value 28 moves ever further into the right tail.

A practical reading of the numbers is worth adding. A quality manager who checks single packages will very occasionally (about 13 times in 10,000) find one of 28 ounces or more, pure

chance from a correctly set machine; but the average of even four packages will essentially never reach 28 unless the machine has genuinely drifted. That asymmetry — noisy individuals, reliable averages — is why statistical process control charts plot sample means rather than individual items, and it is the CLT, plus the exact normality here, that quantifies the gain.

Conclusion

The central limit theorem says that sample means from any population with finite variance approach a normal distribution centred at μ with standard error σ/√n; the dice example shows averaging itself creating the bell shape, and the survey example shows why the result is the working foundation of economic inference. In the filling-machine application, $P(\bar{X} < 28)$ equals 0.9987 for n = 1 and is effectively 1 for n = 4, 16 and 64, because the standard error falls from 1 to 0.5, 0.25 and 0.125 and the standardised distance of 28 ounces grows from 3 to 24. Averages are dramatically more predictable than individuals — that is the theorem's practical message.

Question 5: Mean and Variance of a Sample Distribution (20 Marks)

On a certain Monday noon, a check was made of five different computer rooms in campus residence units. The number of students using computers in the five units was 20, 12, 10, 8, and 11, respectively. (a) Find the average number of users per room. (10) (b) Find the variance of this sample distribution. (10)

Introduction

Before probability can describe data, the data must be summarised. Two numbers carry most of the story of any batch of observations: a measure of location, saying where the batch is centred, and a measure of dispersion, saying how spread out it is. This question asks for the two

foundational summaries — the arithmetic mean and the variance — for a tiny data set of five observations, the numbers of students found using computers in five campus residence rooms at Monday noon: 20, 12, 10, 8 and 11. Because the data set is so small, every step can be shown in full, and the answer uses the opportunity to explain what each statistic means, how the two conventions for variance (dividing by n and dividing by n − 1) differ and why, and what the numbers would tell a campus IT manager.

Part (a): The Average Number of Users per Room

The arithmetic mean is the sum of the observations divided by their number: $\bar{x} = \Sigma x / n$. Here $\Sigma x = 20 + 12 + 10 + 8 + 11 = 61$ and $n = 5$, so $\bar{x} = 61 / 5 = 12.2$. The average number of users per room is 12.2 students. The figure is not a possible count for any single room — no room held 12.2 students — and that is normal for a mean: it is a balance point, the value each room would show if the 61 students were redistributed equally, and the point about which the deviations sum to zero. That balancing property is easily checked: the deviations from 12.2 are 7.8, − 0.2, − 2.2, − 4.2 and − 1.2, which indeed add to 0.0. The mean also has the least-squares property that becomes central later in the course: it is the single number that minimises the sum of squared deviations from the data, which is why it pairs naturally with the variance, itself an average of squared deviations. Sorted, the data are 8, 10, 11, 12, 20, so the median is 11; the mean (12.2) exceeds the median because the single large value 20 pulls it upward — a first, gentle illustration that the mean uses every value and is therefore sensitive to extremes, a strength when all values carry information and a weakness when an outlier is an error or an exceptional case.

Part (b): The Variance of the Sample Distribution

The variance measures the average squared distance of the observations from their mean. Computing deviation by deviation: for 20, the deviation is 7.8 and its square is 60.84; for 12, − 0.2 and 0.04; for 10, − 2.2 and 4.84; for 8, − 4.2 and 17.64; for 11, − 1.2 and 1.44. The sum of squared deviations is $\Sigma(x - \bar{x})^2 = 60.84 + 0.04 + 4.84 + 17.64 + 1.44 = 84.80$. The same total is obtained by the computing formula $\Sigma x^2 - (\Sigma x)^2/n$: $\Sigma x^2 = 400 + 144 + 100 + 64 + 121 = 829$, and $(\Sigma x)^2/n = 61^2/5 = 3721/5 = 744.2$, giving $829 - 744.2 = 84.8$, which confirms the arithmetic.

Table 1: Deviation table for the five rooms ($\bar{x} = 12.2$)

Room	x (users)	$x - \bar{x}$	$(x - \bar{x})^2$
1	20	7.8	60.84
2	12	− 0.2	0.04
3	10	− 2.2	4.84
4	8	− 4.2	17.64
5	11	− 1.2	1.44
Total	61	0.0	84.80

Two divisions are now possible, and the distinction is one of the most important small print items in statistics. If the five rooms are treated as the complete population of interest — the whole story is these five rooms on this Monday — the (population) variance is $\sigma^2 = 84.80 / 5 = 16.96$, with standard deviation $\sigma = 4.12$ users. But the question calls this a sample distribution, and that word governs the answer: the five rooms are a sample from the larger population of room-hours the campus cares about, and the sample variance uses the divisor n − 1: $s^2 = 84.80 / (5 - 1) = 84.80 / 4 = 21.20$, with sample standard deviation $s = \sqrt{21.20} \approx 4.60$ users. The n − 1 divisor (Bessel's correction) is used because the deviations are measured from the sample

mean, which was itself fitted to these very data and therefore sits closer to them than the true population mean does; dividing by n would systematically understate the population variance, and dividing by n − 1 exactly removes that bias, making s^2 an unbiased estimator of σ^2. One degree of freedom was spent estimating the mean, leaving $n - 1 = 4$ for the variance. The answer to report is therefore: sample variance $s^2 = 21.20$ (squared users), sample standard deviation $s \approx 4.60$ users; the corresponding population form values, if the five rooms were the entire population, would be 16.96 and 4.12.

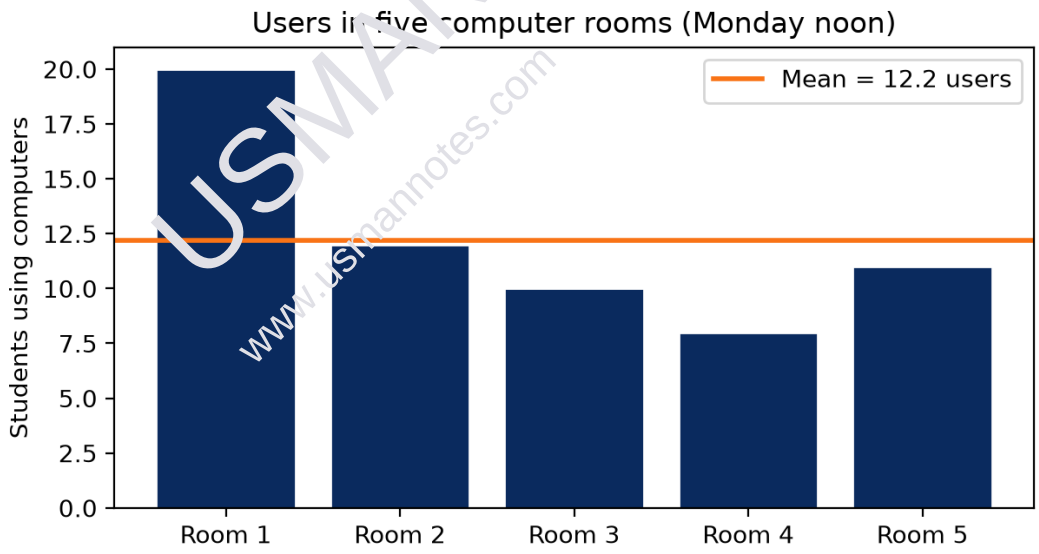


Figure 1: Users per room and their mean, 12.2. Room 1 (20 users) accounts for most of the variation.

Interpretation completes the exercise. A standard deviation of about 4.6 users on a mean of 12.2 is a coefficient of variation of roughly 38 percent: usage differs a lot from room to room at the same moment, driven overwhelmingly by Room 1, whose squared deviation (60.84) is nearly 72 percent of the total 84.80. For the IT manager the message is that capacity is unevenly loaded — one room is crowded while another sits half empty — and averages alone would have hidden that. A single noon check on one Monday is, of course, only a snapshot: the

same rooms checked on other days would give a different mean and variance, and that sample-to-sample variation is precisely what the sampling theory of Questions 3 and 4, and the inference units that follow, are built to handle.

Conclusion

The average number of users per room is $\bar{x} = 61/5 = 12.2$ students. The sum of squared deviations about that mean is 84.80, so the variance of this sample distribution is $s^2 = 84.80/4 = 21.20$, a standard deviation of about 4.60 students (16.96 and 4.12 respectively if the five rooms are treated as the whole population). Together the two numbers summarise the snapshot: typically about a dozen students per room, but with wide room-to-room variation dominated by one busy room – the location-and-dispersion pair in its simplest working form.

USMAN NOTES www.usmannotes.com