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. 2 (Units 5–9)

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: Properties of the OLS Estimator; Confidence Interval for a Difference of Proportions (20 Marks)

(a) Describe the properties of OLS estimator. (10) (b) A sample of 300 employees was interviewed for opinions on their attitude concerning a new bonus plan being proposed by the management. The results, summarized separately for males and females, showed that 52 percent of the 180 male and 55 percent of the 120 female favoured it. Calculate a 90 percent confidence interval for the true opinion difference. (10)

Introduction

The second half of this course moves from describing data to learning from data: estimating the parameters of an economic relationship and judging how much confidence those estimates deserve. Ordinary least squares (OLS) is the workhorse estimator of econometrics, and part (a) asks what properties justify the trust economists place in it — unbiasedness, efficiency, consistency and the sampling distribution that makes testing possible. Part (b) turns from regression to a plainer inference problem of the same family: two sample percentages, one for male and one for female employees, must be converted into a confidence interval for the true

difference of opinion about a proposed bonus plan. The two parts belong together: a point estimate without its sampling properties, and without an interval around it, is only half an answer.

Part (a): Properties of the OLS Estimator

Set the scene. In the linear regression model $Y = \alpha + \beta X + e$, OLS chooses the intercept and slope that minimise the sum of squared residuals, giving $b = \Sigma(x - \bar{x})(y - \bar{y}) / \Sigma(x - \bar{x})^2$ and $a = \bar{y} - b\,\bar{x}$. Because the sample is random, a and b are random variables with sampling distributions, and the "properties" of the estimator are statements about those distributions under the classical assumptions: the model is linear and correctly specified; X is non-stochastic (or independent of the error); the error has zero mean, $E(e) = 0$; its variance is constant, $Var(e) = \sigma^2$ (homoscedasticity); errors are uncorrelated with each other (no autocorrelation); and, for exact small-sample testing, errors are normally distributed. The properties are as follows.

- Linearity. The OLS estimators are linear functions of the observed Y values: $b = \Sigma\, w_i Y_i$ with weights $w_i = (x_i - \bar{x}) / \Sigma(x - \bar{x})^2$ that depend only on the X's. Linearity makes the sampling distribution of b tractable and is one leg of the Gauss–Markov result.
- Unbiasedness. $E(a) = \alpha$ and $E(b) = \beta$: over repeated samples, OLS estimates are centred on the true parameter values, with no systematic tendency to overstate or understate them. Unbiasedness flows chiefly from $E(e) = 0$ and correct specification; omit a relevant correlated variable, and unbiasedness is the first property lost.
- Efficiency (minimum variance) — the Gauss–Markov theorem. Among all linear unbiased estimators, OLS has the smallest variance: $Var(b) = \sigma^2 / \Sigma(x - \bar{x})^2$. OLS is therefore BLUE — Best Linear Unbiased Estimator. No other linear unbiased rule extracts more information

from the same sample; rival estimators either are biased or scatter more widely around the truth.

- Consistency. As $n \to \infty$, a and b converge in probability to α and β: the variance of b shrinks toward zero as the spread of X and the sample size grow, so a large sample pins the parameters down as precisely as desired. Consistency is the large-sample insurance policy that survives even when some finite-sample properties need qualification.
- A known sampling distribution, hence testability. With normally distributed errors, b is exactly normally distributed in samples of any size, $b \sim N(\beta, \sigma^2/\Sigma(x - \bar{x})^2)$, and the standardised statistic follows the t-distribution when σ^2 is estimated by the residual variance $s^2 = \Sigma\hat{e}^2/(n - 2)$. This is what licences t-tests, F-tests and confidence intervals for regression coefficients — OLS does not merely produce numbers, it produces numbers with standard errors attached.
- Sufficiency of fit measures and minimum-variance residuals. The OLS residuals sum to zero and are uncorrelated with X by construction, the fitted line passes through the point of means $(\bar{x}, \bar{y})$, and the decomposition of total variation into explained and residual parts (and hence R^2) holds exactly — bookkeeping properties that make OLS output fully interpretable.

The properties are conditional, and naming the failure modes is part of describing OLS honestly. Heteroscedastic or autocorrelated errors leave OLS unbiased but destroy its efficiency and, more dangerously, invalidate the usual standard errors, so significance is misjudged (the remedies are robust standard errors or GLS). Measurement error in X and simultaneity — X correlated with e — destroy unbiasedness and consistency themselves, which is why instrumental-variables methods exist. Multicollinearity breaks no assumption but inflates Var(b)

through a small effective spread of independent X-variation. The professional summary is: under the classical assumptions OLS is BLUE, consistent, and exactly testable; the diagnostic craft of econometrics consists of checking how badly a real data set bends those assumptions.

Part (b): 90% Confidence Interval for the Difference in Opinion

Let p1 be the true proportion of male employees favouring the plan and p2 the true proportion of females. The samples give $\hat{p}1 = 0.52$ from $n1 = 180$ males and $\hat{p}2 = 0.55$ from $n2 = 120$ females, so the point estimate of the difference (males minus females) is $\hat{p}1 - \hat{p}2 = 0.52 - 0.55 = -0.03$: in the sample, support among females is 3 percentage points higher. Both samples are large and both np and n(1 − p) comfortably exceed 5 in each group, so the sampling distribution of $\hat{p}1 - \hat{p}2$ is well approximated by the normal distribution (each proportion is a mean of 0–1 responses, and the central limit theorem applies). Its standard error uses each group's own sample proportion: $SE = \sqrt{[\ \hat{p}1(1-\hat{p}1)/n1 + \hat{p}2(1-\hat{p}2)/n2\ ]} = \sqrt{[\ (0.52)(0.48)/180 + (0.55)(0.45)/120\ ]} = \sqrt{[\ 0.001387 + 0.002063\ ]} = \sqrt{0.003449} = 0.0587$. For 90 percent confidence the critical value is $z = 1.645$, so the margin of error is $1.645 \times 0.0587 = 0.0966$. The interval is therefore -0.03 ± 0.0966, that is, from −0.1266 to +0.0666. In percentage-point language: the true male-minus-female difference in support lies between about −12.7 and +6.7 percentage points, with 90 percent confidence.

Table 1: Working for the confidence interval

Quantity	Males	Females
Sample size	n1 = 180	n2 = 120
Favouring the plan	$\hat{p}1 = 0.52$	$\hat{p}2 = 0.55$
p(1 − p)/n	0.001387	0.002063
Difference (M − F)	$\hat{p}1 - \hat{p}2 = -0.03$	SE = 0.0587

90% interval	$-0.03 \pm 1.645(0.0587)$	$= (-0.127, +0.067)$

The managerial reading is immediate: the interval contains zero, so the survey provides no reliable evidence that male and female employees actually differ on the bonus plan; the observed 3-point gap is well within sampling noise at this sample size. Management should read the result as "both groups are roughly evenly divided, at about half in favour" rather than as "women support the plan more". If a precise estimate of the gap were genuinely needed, the interval's width (± 9.7 points) shows the price: halving the margin of error would require roughly four times the interviews, since the width shrinks only with $1/\sqrt{n}$.

Conclusion

OLS earns its place through its properties: linear, unbiased, minimum-variance among linear unbiased estimators by the Gauss–Markov theorem, consistent, and equipped with an exact normal (t) sampling theory under the classical assumptions — properties that hold conditionally and fail in instructive ways when the assumptions fail. Applied interval estimation gives the second answer: the 90 percent confidence interval for the true male-minus-female difference in support for the bonus plan is $(-0.127, +0.067)$, which includes zero, so no genuine gender difference in opinion is established by this survey.

Question 2: Testing Hypotheses about Means (20 Marks)

(a) Two sets of 50 elementary school children were taught to read by two different methods. After the instruction was over, a reading test gave the following results: $\bar{x}1 = 78.4$, $\bar{x}2 = 75.3$, $s1 = 6$, $s2 = 8$. Test the hypothesis that $\mu1 = \mu2$. (10) (b) Given $\bar{x} = 84$, $\sigma^2 = 13$, and $n = 100$, test the null hypothesis that $\mu = 86$. (10)

Introduction

A hypothesis test is a disciplined wager: assume for the sake of argument that nothing is going on — the null hypothesis — and ask whether the data would be surprising under that assumption. If they would be very surprising, the assumption is rejected in favour of an alternative. This question applies the wager twice: first to the comparison of two independent group means, where the natural question of educational research — does the new reading method actually work better? — is at stake; and then to a single mean with known population variance, the clean textbook case of the z-test. In both parts the five-step ritual is followed explicitly: state the hypotheses, choose the test statistic and significance level, compute, compare with the critical value, and conclude in the language of the original problem. Unless otherwise stated, tests are two-sided at the 5 percent level, with the 1 percent verdict also reported where it differs in spirit.

USMAN NOTES www.usmannotes.com

Part (a): Comparing Two Reading Methods

The hypotheses are H0: μ1 = μ2 (the two methods produce the same mean reading score in the population) against H1: μ1 ≠ μ2 (the means differ). Each group has n1 = n2 = 50 children, large enough for the central limit theorem to make the sample means approximately normal even if individual scores are not. The observed difference is $\bar{x}1 - \bar{x}2 = 78.4 - 75.3 = 3.1$ marks in favour of method 1. The standard error of that difference, treating the groups as independent samples, is $SE = \sqrt{(s1^2/n1 + s2^2/n2)} = \sqrt{(36/50 + 64/50)} = \sqrt{(0.72 + 1.28)} = \sqrt{2} = 1.414$. The test statistic is therefore $t = 3.1 / 1.414 = 2.19$ (with samples this size, the z and t critical values are nearly identical; the t-distribution with about 98 degrees of freedom gives the two-sided 5 percent critical value 1.98, and the normal value is 1.96). Since the calculated 2.19 exceeds the

5 percent critical value, we reject H0 at the 5 percent level: the difference of 3.1 marks is statistically significant, and the evidence supports the conclusion that method 1 produces a genuinely higher mean reading score than method 2. At the stricter 1 percent level the critical value is about 2.63, which 2.19 does not exceed, so at 1 percent we would not reject H0; the result is significant at 5 percent but not at 1 percent, with a two-sided p-value of roughly 0.03. That borderline character should be reported, not hidden: the study is good evidence for a real, modest advantage of method 1, not overwhelming proof. Practical significance also deserves a sentence — 3.1 marks on a test whose within-group standard deviation is 6–8 marks is an effect of roughly 0.4 of a standard deviation, educationally meaningful if the method costs no more to deliver.

Part (b): Testing a Single Mean with Known Variance

Here H0: $\mu = 86$ against H1: $\mu \neq 86$, with $\bar{x} = 84$ from $n = 100$ and known population variance $\sigma^2 = 13$, so $\sigma = \sqrt{13} = 3.606$. Because the population variance is known and the sample is large, the test statistic is the z-statistic $z = (\bar{x} - \mu_0) / (\sigma/\sqrt{n}) = (84 - 86) / (3.606/10) = -2 / 0.3606 = -5.55$. The two-sided critical values are 1.96 at 5 percent and 2.576 at 1 percent. The calculated $|z| = 5.55$ dwarfs both, and the p-value is far below 0.001. We therefore reject H0 decisively at both levels: the sample mean of 84 is not compatible with a population mean of 86 given this small variance and large sample. The intuition matches the arithmetic: the standard error of the mean is only 0.36, so the observed mean sits five and a half standard errors below the hypothesised value — a distance chance alone essentially never produces. Notice how the two parts of the question illustrate the two levers of every test: the signal (the size of the discrepancy, 3.1 marks and 2 units respectively) and the noise (the standard error, 1.414 and

0.361). Part (a)'s larger signal was only borderline because its noise was also large; part (b)'s smaller signal is conclusive because its noise is tiny.

A word on design completes part (a). A test's ability to detect a real difference — its power — rises with the size of the difference, falls with the variability of scores, and rises with sample size through the standard error's 1/√n law. The reading study, with 50 children per method and within-group standard deviations of 6–8 marks, was only just powerful enough to certify a 3.1-mark advantage at the 5 percent level; a smaller true advantage, or a smaller sample, would have left the same educational question officially "unresolved". Planning a study therefore means choosing n in advance from the smallest effect worth detecting, not discovering afterwards that the sample was too small to answer the question posed.

Two cautions belong in a complete answer. First, "reject H0" never means "prove H1 with certainty": every test carries a Type I error probability equal to its significance level, and every non-rejection carries the possibility of a Type II error when the effect is real but the sample too small — part (a)'s verdict would flip at 1 percent precisely because the study is near the edge of its resolving power. Second, statistical significance is not importance: the test answers "is the difference distinguishable from zero?", while the decision-maker must still ask "is it worth acting on?", using the effect size, the costs, and the stakes.

Conclusion

Part (a): t = 2.19 exceeds the 5 percent critical value (≈ 1.98), so H0: μ1 = μ2 is rejected at the 5 percent level (though not at 1 percent): the first reading method yields a significantly higher mean score. Part (b): z = − 5.55, far beyond any conventional critical value, so H0: μ = 86 is rejected decisively. Together the two tests show that significance is always a ratio —

discrepancy divided by standard error — and that honest reporting includes the level at which the verdict was reached.

Question 3: Least-Squares Regression and Correlation — Goats and Leather Jackets (20 Marks)

Suppose we have data of goats X (in million) and production of leather jackets (in thousands) for a certain region of Pakistan, as follows: 2020: Y = 4, X = 3; 2021: Y = 5, X = 4; 2022: Y = 6, X = 5; 2023: Y = 10, X = 5. (a) Find the estimating equation for Y = a + bX + e, by the method of least squares. (12) (b) Find the coefficient of correlation between Y and X. (08)

Introduction

Regression analysis asks how one variable moves, on average, with another; correlation asks how tightly the two move together. This question gives four years of regional data linking the goat population X (millions) — the raw-material base of the leather trade — to the production of leather jackets Y (thousands), and asks for the least-squares estimating equation and the correlation coefficient. Four observations make the arithmetic fully transparent, which is the point: every sum, every cross-product and the final check can be displayed. The small sample is also a standing warning, taken up at the end, about how little trust a line fitted to four points deserves for forecasting, however exactly it is computed.

Part (a): The Least-Squares Estimating Equation

The least-squares criterion chooses a and b to minimise the sum of squared vertical residuals, $\Sigma(Y - a - bX)^2$. The solution is $b = [\ \Sigma XY - (\Sigma X)(\Sigma Y)/n\ ] / [\ \Sigma X^2 - (\Sigma X)^2/n\ ]$ and $a = \bar{Y} - b\,\bar{X}$. Building the sums from the data: $\Sigma X = 3 + 4 + 5 + 5 = 17$, so $\bar{X} = 4.25$; $\Sigma Y = 4 + 5 + 6 + 10 = 25$, so $\bar{Y} = 6.25$; $\Sigma XY = (3)(4) + (4)(5) + (5)(6) + (5)(10) = 12 + 20 + 30 + 50 = 112$; $\Sigma X^2 = 9 + 16 + 25 + 25 = 75$; $\Sigma Y^2 =$

16 + 25 + 36 + 100 = 177. The numerator of b is 112 − (17)(25)/4 = 112 − 106.25 = 5.75, and the denominator is 75 − $(17)^2/4$ = 75 − 72.25 = 2.75, so b = 5.75 / 2.75 = 2.0909 ≈ 2.09. Then a = 6.25 − (2.0909)(4.25) = 6.25 − 8.886 = − 2.636 ≈ − 2.64. The estimating equation is therefore $\hat{Y}$ = − 2.64 + 2.09 X.

Table 1: Least-squares working table

Year	X (goats, million)	Y (jackets, 000)	XY	X^2	Y^2
2020	3	4	12	9	16
2021	4	5	20	16	25
2022	5	6	30	25	36
2023	5	10	50	25	100
Total (n = 4)	17	25	112	75	177

USMAN NOTES www.usmannotes.com

Interpretation: the slope b = 2.09 says that, over these four years, an additional one million goats in the region is associated with about 2.09 thousand additional leather jackets produced — the expected response of output to herd size along the fitted line. The intercept a = − 2.64 is the line's value at X = 0; it has no economic meaning here (a region with no goats produces no jackets, not a negative quantity), because X = 0 lies far outside the observed range of 3–5 million, and it serves only to position the line. The equation passes through the point of means, as every OLS line must: at X = 4.25 it gives $\hat{Y}$ = − 2.636 + 2.0909(4.25) = 6.25 = $\bar{Y}$, a useful arithmetic check. Fitted values and residuals complete the verification: for X = 3, $\hat{Y}$ = 3.64 (residual +0.36); X = 4, $\hat{Y}$ = 5.73 (residual − 0.73); X = 5 (2022), $\hat{Y}$ = 7.82 (residual − 1.82); X = 5 (2023), $\hat{Y}$ = 7.82 (residual +2.18). The residuals sum to zero (0.36 − 0.73 − 1.82 + 2.18 ≈ 0), confirming the least-squares bookkeeping. Note what the residuals reveal: the two years with

identical X = 5 differ in Y by 4 thousand jackets, variation the line cannot explain — goat numbers alone do not determine jacket output, which also depends on hide quality, prices, tannery capacity and demand.

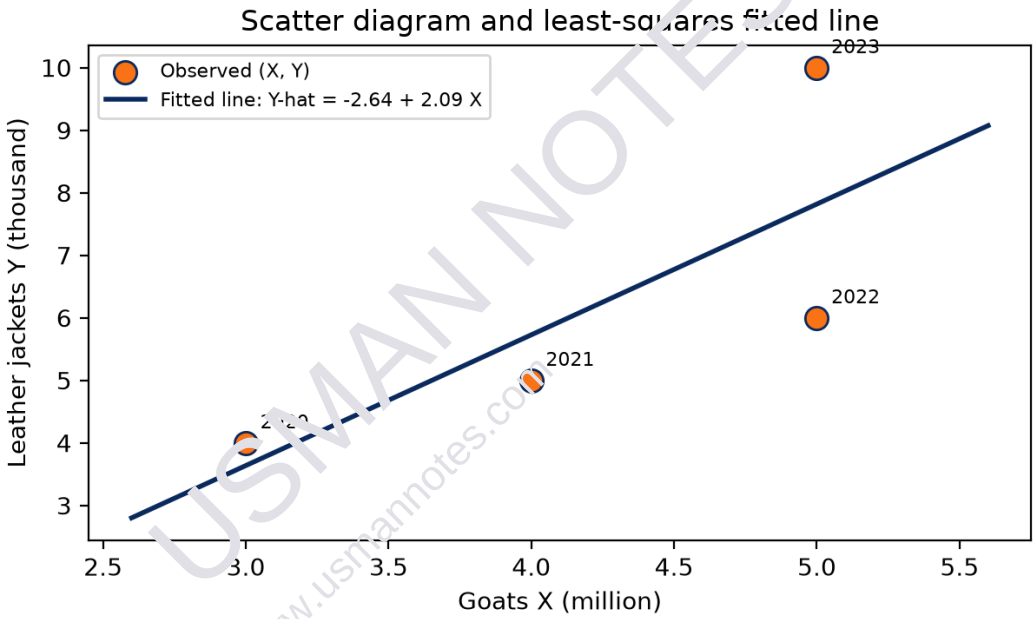


Figure 1: Scatter of jacket production against goat numbers with the fitted least-squares line $\hat{Y} = -2.64 + 2.09X$.

Part (b): The Coefficient of Correlation

The (Pearson) coefficient of correlation is $r = [\ \Sigma XY - (\Sigma X)(\Sigma Y)/n\ ] / \sqrt{\{\ [\ \Sigma X^2 - (\Sigma X)^2/n\ ]\ [\ \Sigma Y^2 - (\Sigma Y)^2/n\ ]\ \}}$. The numerator is the same 5.75 found above. The Y-term in the denominator is $\Sigma Y^2 - (\Sigma Y)^2/n = 177 - 625/4 = 177 - 156.25 = 20.75$, and the X-term is 2.75, so the denominator is $\sqrt{(2.75 \times 20.75)} = \sqrt{57.0625} = 7.554$. Hence $r = 5.75 / 7.554 = 0.761 \approx 0.76$. The correlation is positive and fairly strong: goat numbers and jacket production moved together quite closely over these four years, in the same direction. Its square, $r^2 = 0.58$, says that about 58 percent of the variation in jacket production across the four years is accounted for by the linear association with goat numbers, leaving 42 percent to other factors and chance — the

coefficient of determination of Question 5 in miniature. Consistency checks bind the two parts together: the slope and correlation obey b = r (sY/sX), and indeed with sX = √(2.75/4) = 0.829 and sY = √(20.75/4) = 2.278, r (sY/sX) = 0.761 × 2.747 = 2.09 = b, exactly as theory requires.

The necessary cautions are unusually important with n = 4. A correlation of 0.76 computed from four points is fragile: with so few observations it is not statistically distinguishable from zero at the 5 percent level (the critical value of r with 2 degrees of freedom is about 0.95), and a single year — the 2023 jump in Y — does much of the work. Correlation also measures linear association only, and association is not causation: both series trend upward over time, so part of the 0.76 may reflect a common time trend (growing demand, expanding industry) rather than a direct herd-to-jacket mechanism. The honest conclusion is descriptive: in this region and period, bigger goat populations went with bigger jacket output, moderately tightly, along the line $\hat{Y} = -2.64 + 2.09X$.

Conclusion

The least-squares estimating equation is $\hat{Y} = -2.64 + 2.09X$: each additional million goats is associated with roughly 2.09 thousand more leather jackets, with the line verified through the point of means and residuals summing to zero. The coefficient of correlation is r = + 0.76 (r^2 = 0.58), a fairly strong positive linear association — reported with the warning that four observations cannot support strong inferential claims, and that association, here as always, is not by itself causation.

Question 4: Confidence Intervals as a General Technique; Testing a Sample of Digits (20 Marks)

(a) Explain the concept of the confidence interval as a general technique. (10) (b) At a 1% significance level, test the hypothesis that the following sample is drawn from a population of random digits: 2, 1, 2, 0, 8, 6, 1, 2, 4, 5, 1, 2, 4, 8, 4, 4, 3, 0, 2. Suppose that the alternative hypothesis is that there is a bias towards small digits. (10) [Note: in the printed booklet the list begins with a comma and no legible first digit; the test below uses the 19 legible digits exactly as printed.]

Introduction

Point estimates pretend to a precision they do not possess. A sample mean of 84 does not mean the population mean is 84; it means 84 is the best single guess from this sample, and a different sample would have given a different guess. The confidence interval is the profession's honest answer to that problem: it reports a range, together with the long-run reliability of the method that produced the range. Part (a) develops the concept in general; part (b) shows hypothesis testing — the confidence interval's close cousin — applied to a forensic little problem, deciding whether a string of digits behaves like genuinely random digits or betrays a bias towards small values.

Part (a): The Confidence Interval as a General Technique

A confidence interval for an unknown parameter θ is a pair of statistics (L, U), computed from the sample, such that the procedure generating them captures the true θ in a stated proportion of repeated applications: $P(L \le \theta \le U) = 1 - \alpha$, where $100(1 - \alpha)$ percent is the confidence level — conventionally 90, 95 or 99 percent. The general recipe has four ingredients. First, a point

estimator $\hat{\theta}$ with a known (exact or approximate) sampling distribution. Second, a pivotal quantity — a function of estimator and parameter whose distribution does not depend on unknown parameters, such as $Z = (\hat{\theta} - \theta)/SE$ for large samples or the t-statistic in small normal samples. Third, critical values from that distribution that trap probability $1 - \alpha$ in the centre, for example ± 1.96 for a 95 percent normal interval. Fourth, algebraic rearrangement of the probability statement so that θ stands alone in the middle: the inequalities $P(-1.96 \le (\bar{x} - \mu)/SE \le 1.96) = 0.95$ become $P(\bar{x} - 1.96\ SE \le \mu \le \bar{x} + 1.96\ SE) = 0.95$, and the interval $\bar{x} \pm 1.96$ SE is read off. Every textbook interval is this one manoeuvre dressed for a different parameter: μ with known or unknown variance, a proportion $\hat{p} \pm z\sqrt{\hat{p}(1 - \hat{p})/n}$, a difference of means or of proportions (as computed in Question 1 of this assignment), a regression coefficient $b \pm t \cdot SE(b)$, even a variance, where the pivot is chi-square rather than normal.

The interpretation must be stated with care, because it is the most commonly garbled sentence in applied statistics. The parameter is fixed and the interval is random: before sampling, the procedure has probability $1 - \alpha$ of producing an interval that covers θ; after a particular interval, say (81.2, 86.8), has been computed, μ either is or is not inside it, and no probability attaches to that single event. “95 percent confident” is therefore shorthand for “produced by a method that succeeds 95 percent of the time”, not “the probability is 0.95 that μ lies here” — although in everyday reporting, and exactly in Bayesian credible intervals, the looser reading is what audiences take away. What governs an interval’s usefulness is its width, and width obeys clear laws: it shrinks with the square root of sample size, grows with the population variability, and grows with the confidence demanded (99 percent intervals are wider than 95 percent ones from the same data). The technique thus makes the price of precision explicit and tradeable at

the design stage: a survey planner can solve the width formula for the n that delivers a required margin of error before spending a rupee on fieldwork.

Two further virtues make the confidence interval a general technique rather than a formula. It is dual to hypothesis testing: the 100(1 − α) percent interval contains exactly those hypothesised parameter values that a two-sided test at level α would not reject, so one computation serves both purposes — and the interval is strictly more informative, because it shows not only whether a value such as "no difference" is compatible with the data but also which other values are, and with what margin. And it disciplines conclusions in policy and business reporting: an estimated effect of + 2 with interval (− 6, + 10) tells the decision-maker, at a glance, that the data are consistent with harm, nothing, and substantial benefit alike — a truth a bare point estimate conceals. Its limitations are the assumptions beneath it: random sampling, the stated distributional model, and freedom from non-sampling errors such as non-response and mis-measurement, against which no width formula protects.

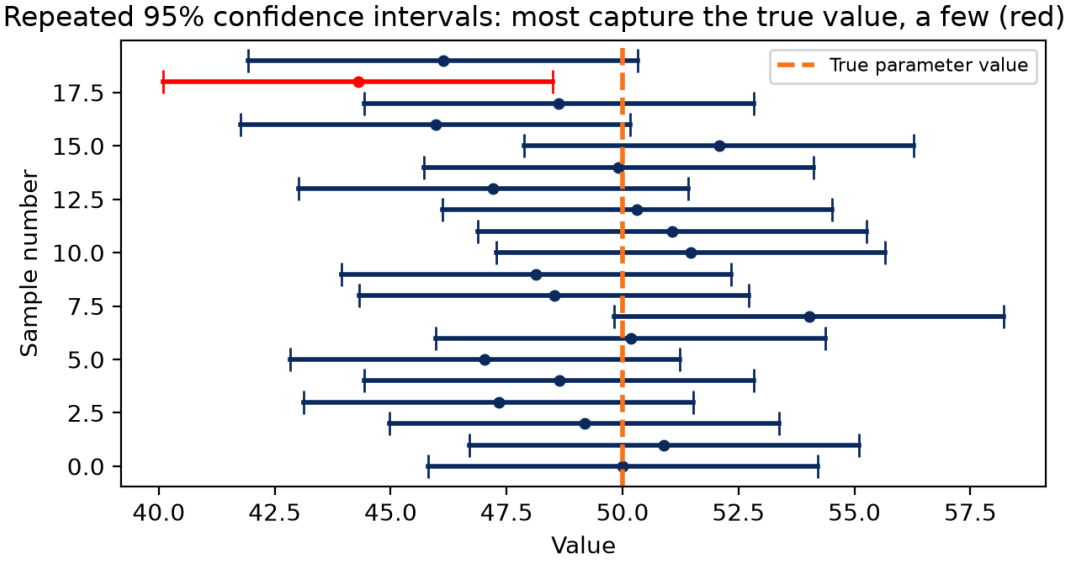


Figure 1: Twenty simulated 95% confidence intervals for the same true value: the method, not any single interval, carries the 95% guarantee.

Part (b): Are the Digits Random? A Test for Bias towards Small Digits

If digits are genuinely random — drawn uniformly from 0 to 9, as in a random number table — each digit has probability 1/10, the population mean of the digits is $\mu = (0 + 1 + \ldots + 9)/10 = 4.5$, and the population variance is $\sigma^2 = (0^2 + 1^2 + \ldots + 9^2)/10 - 4.5^2 = 28.5 - 20.25 = 8.25$, so $\sigma = 2.872$. The hypotheses translate the story exactly: H0: $\mu = 4.5$ (the sample comes from a random-digit population) against the one-sided alternative H1: $\mu < 4.5$ (bias towards small digits), at $\alpha = 0.01$, for which the one-sided critical value is $z = -2.326$. The printed sample (19 legible digits, as noted under the question) is 2, 1, 2, 0, 8, 6, 1, 2, 4, 5, 1, 2, 4, 8, 4, 4, 3, 0, 2. Their sum is $\Sigma x = 59$, so the sample mean is $\bar{x} = 59/19 = 3.105$ — visibly below 4.5, and a supporting count shows why: 15 of the 19 digits are 4 or less, and only 4 exceed 4. The standard error of the mean is $\sigma/\sqrt{n} = 2.872/\sqrt{19} = 0.659$, giving the test statistic $z = (3.105 - 4.5) / 0.659 = -1.395 / 0.659 = -2.12$.

Table 1: Frequency of the sampled digits

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	2	3	5	1	4	1	1	0	2	0
Expected under H0	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9

Decision: the calculated $z = -2.12$ lies short of the 1 percent critical value -2.326 (the one-sided p-value is about 0.017). We therefore do NOT reject H0 at the 1 percent level: although the sample leans unmistakably towards small digits — a mean of 3.11 against an expected 4.5 — the evidence, from only 19 digits, is not strong enough to convict at the strict 1 percent

standard. The same data WOULD reject H0 at the 5 percent level (critical value − 1.645), and the answer should say both things, because that is the truth about the evidence: suggestive of bias, decisive at 5 percent, inconclusive at 1 percent. This is not indecision but the logic of testing working as designed: the 1 percent level demands stronger proof before crying fraud, accepting a larger risk of missing a real bias in a small sample. A larger sample of digits from the same source would settle the matter, since the standard error shrinks with √ n while a real bias in the mean would persist.

Conclusion

The confidence interval generalises one idea — estimator ± critical value × standard error, derived by inverting a probability statement about a pivot — to every parameter of economic interest, and its correct interpretation is about the long-run reliability of the method. Applied testing then shows the machinery from the other side: for the digit sample, $\bar{x} = 3.105$, $z = -2.12$, short of the 1 percent critical value − 2.326, so randomness is not rejected at the 1 percent level despite a visible tilt towards small digits that would be significant at 5 percent.

Question 5: Four Key Terms — Maximum Likelihood, R^2, Least Squares, Normal Distribution (20 Marks)

Explain the following terms: (a) Maximum likelihood method; (b) Coefficient of determination; (c) Method of least squares; (d) Normal Distribution. (5 + 5 + 5 + 5)

Introduction

Four terms, four pillars of the course. The method of least squares and the maximum likelihood method are the two great general principles for estimating parameters from data; the coefficient of determination is the standard measure of how well a fitted regression explains its

dependent variable; and the normal distribution is the probability model on which classical inference rests — directly, for errors and measurements, and indirectly, through the central limit theorem, for almost everything else. Each term is explained below with its definition, its logic, a small worked illustration, and its place in economic statistics.

(a) Maximum Likelihood Method

The maximum likelihood method chooses, as the estimate of an unknown parameter, the value that makes the actually observed sample most probable — most "likely". Formally, if the observations x1, …, xn are drawn from a distribution with density (or probability function) f(x; θ), the likelihood function is $L(\theta) = \prod_i f(x_i; \theta)$, the joint probability of the sample regarded as a function of the parameter. The maximum likelihood estimator (MLE) is the value $\hat{\theta}$ that maximises L(θ), usually found by maximising the log-likelihood ln L(θ) and solving $d \ln L/d\theta = 0$. A miniature example fixes the idea: if 7 of 10 inspected items are defective and the defect probability p is unknown, the likelihood is proportional to $p^7(1 - p)^3$, maximised at $\hat{p} = 0.7$ — the sample proportion, exactly as intuition demands. Similarly, the MLE of a normal mean is the sample mean $\bar{x}$, and of the variance is $\Sigma(x - \bar{x})^2/n$. The method's attractions are its strong large-sample properties: under standard regularity conditions the MLE is consistent, asymptotically normal, and asymptotically efficient — it achieves the smallest possible variance among consistent estimators in large samples, the Cramér–Rao bound — which is why it is the default engine behind logit, probit, Poisson and other models of modern applied econometrics where least squares is not natural. Its caveats: it requires a full distributional assumption for the data (a wrong distribution can mislead it), the maximisation may need numerical methods, and some

MLEs are biased in small samples — the normal-variance MLE divides by n, not n − 1, and so understates σ^2 slightly in small samples even while remaining consistent.

(b) Coefficient of Determination (R^2)

The coefficient of determination, R^2, measures the proportion of the total variation in the dependent variable that is explained by the regression. Total variation decomposes exactly for OLS: $\Sigma(Y - \bar{Y})^2 = \Sigma(\hat{Y} - \bar{Y})^2 + \Sigma(Y - \hat{Y})^2$, that is, TSS = ESS + RSS (total = explained + residual sum of squares), and R^2 = ESS/TSS = 1 − RSS/TSS. It lies between 0 and 1: R^2 = 1 means the points lie exactly on the fitted line and the model explains everything; R^2 = 0 means the regressors explain nothing beyond the mean. In simple regression R^2 is literally the square of the correlation coefficient — in Question 3, r = 0.76 and R^2 = 0.58, so 58 percent of the variation in jacket output was associated with goat numbers. R^2 is the most quoted single number in regression output because it speaks plain language ("the model accounts for X percent of the variation"), but it must be read with three warnings. It never falls when regressors are added, even useless ones, which is why the adjusted R^2, which penalises extra regressors by their degrees of freedom, is preferred for comparing models. A high R^2 does not prove a good or causal model — trending nonsense variables can produce R^2 above 0.9 — and a low R^2 does not condemn a model, since cross-section regressions on noisy individual data routinely and usefully live at 0.1–0.3. Finally, R^2 values are not comparable across different dependent variables or different samples. Used as a descriptive measure of fit within one regression, it is excellent; used as a league table of truth, it deceives.

(c) Method of Least Squares

The method of least squares fits a relationship by minimising the sum of the squared discrepancies between observed and fitted values. For the line Y = a + bX, it chooses a and b to minimise S = Σ(Yi − a − bXi)²; setting the partial derivatives to zero yields the normal equations, whose solution is b = Σ(x − x̄)(y − ȳ)/Σ(x − x̄)² and a = ȳ − b x̄, as applied in Question 3. Squaring, rather than using absolute or signed deviations, is the heart of the method: signed deviations cancel and would let wild errors hide; squaring penalises large errors disproportionately, yields a unique solution with clean algebra, connects directly to variance (the minimum value of the criterion about a single point is attained at the mean, and the minimised average is the variance), and — through the Gauss–Markov theorem of Question 1 — delivers estimators that are BLUE under the classical assumptions. The same principle extends far beyond one line: multiple regression, polynomial fits, and the estimation of any model whose parameters enter linearly all reduce to least squares, and even where maximum likelihood is the chosen principle, least squares often coincides with it (for normal errors, the two give identical slope and intercept). Its weaknesses mirror its strengths: because errors are squared, outliers drag the line towards themselves, which motivates robust and least-absolute-deviation alternatives; and the method supplies estimates mechanically whether or not the underlying assumptions — linearity, exogeneity, constant error variance — hold, so the craft lies in diagnosis after fitting, not in the fitting itself.

(d) Normal Distribution

The normal (Gaussian) distribution is the continuous bell-shaped distribution with density f(x) = (1/(σ√(2π))) exp[− (x − μ)²/(2σ²)], fully determined by its mean μ (which locates the centre) and

its standard deviation σ (which sets the spread); the standard normal has μ = 0 and σ = 1 and is tabulated as the z-distribution. Its properties define an ideal of well-behaved variation: it is symmetric about μ, with mean, median and mode all equal; it is unimodal and bell-shaped, tapering towards ± ∞ without ever touching the axis; and its area obeys the empirical rule — about 68.27 percent of values lie within one standard deviation of the mean, 95.45 percent within two, and 99.73 percent within three — so that a value more than three standard deviations out is a genuine rarity demanding explanation (recall Question 4 of Assignment No. 1, where 28 ounces lay exactly three standard deviations above the mean of 25, with probability 0.9987 of falling below it). Linear combinations of normal variables are normal; standardising any normal variable, $z = (x - \mu)/\sigma$, converts it to the standard normal from which all probabilities are read; and the central limit theorem guarantees that sample means approach normality whatever the parent population, which is why this one distribution services confidence intervals, z- and t-tests, control charts and regression inference across the whole course. Many natural and economic magnitudes — heights, measurement errors, examination marks, package weights — are approximately normal because each is the accumulated result of many small independent influences, the very mechanism the CLT formalises. Its limits are equally familiar: financial returns show fat tails, incomes are skewed, and counts are discrete, so normality is a model to be checked, not a fact of nature; but as a benchmark and as the limiting law of averages, it has no rival.

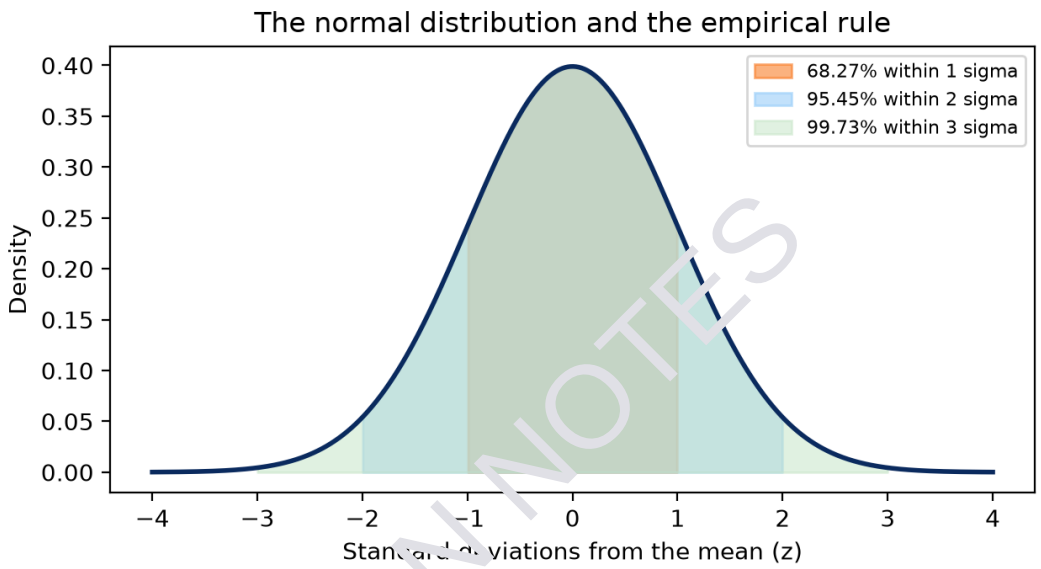


Figure 1: The normal curve and the 68–95–99.7 (empirical) rule.

Conclusion

Together the four terms describe a complete workflow: least squares and maximum likelihood are rival-and-partner principles for estimating a model — the first minimising squared errors with minimal assumptions, the second maximising the probability of the data within a stated distribution and earning the strongest large-sample guarantees; the coefficient of determination then summarises the fitted model's explanatory success as a proportion of variation explained; and the normal distribution supplies the probability law that turns estimates into tests and intervals, exactly for normal data and approximately, via the central limit theorem, for nearly everything else. An economist who can define all four precisely — and state the caveat attached to each — holds the conceptual core of statistical inference.