

ANSWER KEY

Statistics, Probability & Power BI - Evaluation (Actual Paper)

32 questions · 9 topics · correct answers with explanations

CodeWithMunnaX

TOPICS

1. Probability & Bayes' Theorem Q1, Q14, Q17, Q29
2. Descriptive Statistics & Data Types Q9, Q11, Q19, Q20
3. Distributions & Sampling Q10, Q15, Q21, Q23
4. Hypothesis Testing Q3, Q5, Q13, Q26, Q28, Q30
5. Model Evaluation Metrics Q4, Q24
6. Power BI & DAX Q6, Q7, Q12, Q16, Q18, Q22
7. Data Visualisation & Storytelling Q8, Q27
8. Seaborn & Plotly Q2, Q25
9. Case Studies (Subjective) Q31, Q32

TOPIC 1 Probability & Bayes' Theorem

4 Qs

Q1

Multiple select

A supermarket studies 500 baskets: 200 contain milk, 150 contain tea and 90 contain both. Which of the following statements are correct? (Select all that apply.)

Note: This is a Multiple Select Question (MSQ). The answer is marked incorrect if any correct option is missed or any incorrect option is selected.

- A** Given that a basket contains milk, the probability that it also contains tea is 0.45 ✓ Correct
- B** The probability that a basket contains both milk and tea is 0.18, that is 90 out of 500 ✓ Correct
- C** $P(\text{tea} \mid \text{milk})$ equals $P(\text{milk} \mid \text{tea})$, because both use the same joint count of 90
- D** Bayes' theorem gives $P(\text{milk} \mid \text{tea}) = 0.45 \times 0.40 \div 0.30 = 0.60$ for a basket that contains tea ✓ Correct

EXPLANATION

The joint probability is $90/500 = 0.18$. Given milk, the probability of tea is $P(\text{tea} \mid \text{milk}) = 90/200 = 0.45$, using the milk baskets as the new total. Bayes' theorem reverses the conditioning: $P(\text{milk} \mid \text{tea}) = P(\text{tea} \mid \text{milk}) \times P(\text{milk}) / P(\text{tea}) = 0.45 \times 0.40 / 0.30 = 0.60$, which matches the direct count $90/150 = 0.60$. The two

conditional probabilities are not equal (0.45 versus 0.60) because they divide the same joint count by different totals.

Q14

Single choice

Late-refund complaints have three possible root causes. Historical data gives each cause's prior probability and the probability of receiving a late-refund complaint under that cause. Payment gateway: prior 0.50, likelihood 0.20. Bank settlement: prior 0.10, likelihood 0.60. Courier: prior 0.40, likelihood 0.30. A new late-refund complaint arrives, and the total probability of such a complaint is not available. Which cause has the highest posterior probability, and why can it be ranked without that total?

- ☐ A Bank settlement: it has the largest likelihood of 0.60, and the likelihood alone decides which cause is most probable
- ☐ B Payment gateway: it has the largest prior of 0.50, and the prior alone decides which cause is most probable
- ☒ C Courier: $0.40 \times 0.30 = 0.12$ is the largest prior $\times$ likelihood, and the shared denominator does not change the ranking **✓ Correct**
- ☐ D None can be ranked: the posteriors cannot be compared until the total probability of the complaint is known

EXPLANATION

In its comparison form, Bayes' theorem is $P(A|B) \propto P(B|A) \times P(A)$: the denominator $P(B)$ is the same for every cause, so ranking needs only prior $\times$ likelihood. Payment gateway: $0.50 \times 0.20 = 0.10$. Bank settlement: $0.10 \times 0.60 = 0.06$. Courier: $0.40 \times 0.30 = 0.12$, the largest. Ranking by the prior alone would favour the payment gateway and ranking by the likelihood alone would favour the bank settlement, but each ignores half of the formula, and the missing total does not block the comparison because it scales all three posteriors equally.

Q17

Single choice

In a college, 40% of all students attend the weekend coding workshop (event B), and 12% of all students both attend the workshop and secure a placement offer (event $A \cap B$). A student is chosen at random and is known to have attended the workshop. What is the probability that this student secured a placement offer?

- ☒ A 0.30 **✓ Correct**
- ☐ B 0.12
- ☐ C 0.52
- ☐ D 0.048

EXPLANATION

The question asks for a conditional probability: $P(A|B) = P(A \cap B) / P(B) = 0.12 / 0.40 = 0.30$. The value 0.12 is the joint probability itself, without dividing by $P(B)$; 0.52 adds the two probabilities; and 0.048 multiplies them. None of those is the proportion of workshop attendees who secured an offer.

Q29

Single choice

A bank's fraud model raises alerts. Let A be the event 'a transaction is fraudulent' and B the event 'a transaction is flagged'. The bank knows $P(A) = 0.02$, $P(B|A) = 0.90$ and $P(B) = 0.10$. What is the probability that a flagged transaction is actually fraudulent, $P(A|B)$?

A 0.90**B** 0.18

✓ Correct

C 0.018**D** 0.20**EXPLANATION**

By Bayes' theorem, $P(A|B) = P(B|A) \times P(A) / P(B) = 0.90 \times 0.02 / 0.10 = 0.018 / 0.10 = 0.18$. The value 0.90 is the likelihood $P(B|A)$, a different quantity from the posterior $P(A|B)$; 0.018 is the numerator alone, before dividing by the evidence $P(B)$; and 0.20 divides the prior by the evidence and skips the likelihood.

TOPIC 2 Descriptive Statistics & Data Types

4 Qs

Q9

Multiple select

A data scientist works with three variables: (i) the number of defective units in a batch of 40, (ii) the time in minutes that an order takes to arrive, and (iii) the height of a customer. Which of the following statements are correct? (Select all that apply.)

Note: This is a Multiple Select Question (MSQ). The answer is marked incorrect if any correct option is missed or any incorrect option is selected.

A The probabilities for (i) can be listed as a PMF, and they add up to 1 across all possible counts

✓ Correct

B For (ii), the probability of exactly 23.5 minutes is zero, so ranges such as 20 to 25 minutes are used

✓ Correct

C For (iii), the area under the PDF across the whole range of heights adds up to exactly 1

✓ Correct

D For (i), the height of a PDF at 5 defective units gives the probability of exactly 5 defects

EXPLANATION

A count such as the number of defective units is a discrete variable: each possible count has its own probability, listed in a PMF, and all of them sum to 1. Arrival time is a continuous variable, so the probability of any single exact value is zero and probabilities are calculated for ranges. Height is also continuous, so its PDF describes probability density, and the area under the PDF over the entire range is 1. A PDF is not used for a discrete count, and even for a continuous variable the height of the PDF at a point is a density, not a probability.

Q11

Single choice

Meera scored 78 in a statistics test whose scores are approximately normally distributed with mean 66 and standard deviation 8. Arjun scored 84 in a coding test whose scores are approximately normally distributed with mean 70 and standard deviation 10. Which student performed better relative to their own test once each score is standardised?

A Arjun, since his raw score of 84 is higher than her raw score of 78

B Arjun, since his score is 14 marks above his test mean against her 12 marks

C Meera, since her z-score of 1.5 exceeds Arjun's z-score of 1.4

✓ Correct

D Neither, since both scores lie within two standard deviations of their means

EXPLANATION

A z-score standardises a score with $z = (x - \mu) / \sigma$. Meera: $(78 - 66) / 8 = 1.5$. Arjun: $(84 - 70) / 10 = 1.4$. On the standardised scale 1.5 is higher than 1.4, so Meera performed slightly better relative to her own test. Comparing raw scores (84 versus 78) or raw distances from the mean (14 versus 12) ignores that the two tests have different spreads (10 versus 8), and both scores lying within two standard deviations does not make the z-scores equal.

Q19

Single choice

A campus canteen records how every bill is settled: UPI, debit card, credit card or cash. These are nominal categories with no natural order, and the owner wants a single figure for the "typical" payment method. Which measure of central tendency can be applied to this kind of data?

A Standard deviation, the typical spread around the mean

- B** Mean, the sum of the values divided by their count
- C** Median, the middle value of the ascending-ordered data
- D** Mode, the most frequently occurring category ✓ **Correct**

EXPLANATION

Payment method is nominal data, so only a measure that works on categories can be used. The mode simply reports the most frequently occurring category and is the only measure of central tendency that applies to nominal data. The mean needs numeric values that can be added and divided, the median needs the values arranged in ascending order, and the standard deviation is a measure of spread rather than a measure of central tendency.

Q20

Single choice

A sample of $n = 6$ measurements has a mean of 20. Five of the six deviations from the mean, $(x_i - \bar{x})$, are $-3, +5, -1, +2$ and $+4$; the sixth measurement was lost. What is the sample variance of all six measurements?

- A** 17.33
- B** 20.80 ✓ **Correct**
- C** 11.00
- D** 26.00

EXPLANATION

Deviations from the mean always sum to zero, so the missing deviation is $-(-3 + 5 - 1 + 2 + 4) = -7$. The squared deviations are 9, 25, 1, 4, 16 and 49, which total 104. Sample variance divides by $n - 1 = 5$, because only five deviations are free to vary once the mean is fixed: $104 / 5 = 20.80$. Dividing by $n = 6$ gives 17.33, which is the population formula; ignoring the lost deviation gives $55 / 5 = 11.00$; and dividing by 4 gives 26.00, which removes one degree of freedom too many.

TOPIC 3 Distributions & Sampling

4 Qs

Q10

Multiple select

A survey estimates a population mean from simple random samples, and the population standard deviation is 40. The team raises the sample size from 100 to 400. Which of the following statements are correct? (Select all that apply.)

Note: This is a Multiple Select Question (MSQ). The answer is marked incorrect if any correct option is missed or any incorrect option is selected.

- ☐ A The population standard deviation falls to 20, because a larger sample reduces the variability of the population
- ☐ B The standard error falls to a quarter, because it is inversely proportional to the sample size
- ☒ C The standard error falls from $40/\sqrt{100} = 4$ to $40/\sqrt{400} = 2$, so quadrupling the sample size halves it ✓ **Correct**
- ☒ D Sample means will typically lie closer to the population mean, so the estimate becomes more precise ✓ **Correct**

EXPLANATION

The standard error is $\sigma/\sqrt{n}$. With $\sigma = 40$ it is $40/\sqrt{100} = 4$ at $n = 100$ and $40/\sqrt{400} = 2$ at $n = 400$, so quadrupling the sample size halves it, and a smaller standard error means sample means cluster closer to the population mean. The population standard deviation is a property of the population and does not change with the sample size. The standard error varies with the square root of n , not with n itself, so it is halved rather than cut to a quarter.

Q15

Single choice

Order values on an e-commerce app are strongly right-skewed, with a population mean of INR 800 and a population standard deviation of INR 300. An analyst repeatedly draws random samples of 225 orders and records each sample's mean. Which statement about these sample means is most accurate?

- ☒ A They are approximately normal around INR 800, and about 95% fall between INR 760 and INR 840 ✓ **Correct**
- ☐ B They stay right-skewed like the orders, and about 95% fall between INR 200 and INR 1,400
- ☐ C They are approximately normal around INR 800, and about 68% fall between INR 500 and INR 1,100
- ☐ D They are approximately normal around INR 800, and about 95% fall between INR 780 and INR 820

EXPLANATION

For large samples the sample means are approximately normal whatever the shape of the population, centred on the population mean of 800. Their spread is the standard error $\sigma/\sqrt{n} = 300/\sqrt{225} = 300/15 = 20$. About 95% of a normal distribution lies within two standard deviations, so about 95% of sample means fall within $800 \pm 2 \times 20$, that is 760 to 840. Staying right-skewed contradicts the theorem; a spread of 300 is the

population's standard deviation, not the standard error; and 780 to 820 covers only one standard error either side, about 68% rather than 95%.

Q21

Single choice

A helpdesk logs an average of 6 support tickets per hour. The manager wants the probability that exactly 9 tickets arrive in a given hour. The number of trials and a per-trial probability of success are not known. Which distribution and formula fit this need?

- A** Poisson, using $e^{(-\lambda)} \times \lambda^x \div x!$ with $\lambda = 9$ and $x = 6$
- B** Binomial, using $nCr \times p^r \times (1-p)^{(n-r)}$ with $n = 6$ and $r = 9$
- C** Normal, taking the PDF value at $x = 9$ as $P(\text{exactly } 9 \text{ tickets})$
- D** Poisson, using $e^{(-\lambda)} \times \lambda^x \div x!$ with $\lambda = 6$ and $x = 9$

✓ Correct

EXPLANATION

Only the average count per hour is known ($\lambda = 6$) and the value of interest is a specific count ($x = 9$), which is exactly the situation for the Poisson PMF $e^{(-\lambda)} \times \lambda^x \div x!$. The binomial PMF needs a fixed number of trials n and a constant success probability p , neither of which is available, and $n = 6$ trials could never produce 9 successes. A normal PDF gives a probability density rather than the probability of an exact value, and ticket counts are discrete. Swapping the roles ($\lambda = 9$, $x = 6$) treats the count being asked about as the known average.

Q23

Single choice

A university has 65% undergraduate and 35% postgraduate students. A survey team wants a sample of 400 students whose programme mix matches the university's mix by design rather than by luck. Which plan achieves this?

- A** Stratified sampling: draw 260 undergraduates and 140 postgraduates
- B** Simple random sampling: give every student the same chance regardless of programme
- C** Systematic sampling: pick every k -th student from an alphabetical list of names
- D** Convenience sampling: survey the first 400 students who reach the library

✓ Correct

EXPLANATION

Stratified sampling takes samples from each group in proportion to its representation, so 65% of 400 = 260 undergraduates and 35% of 400 = 140 postgraduates fix the mix by design. Simple random sampling gives every student an equal chance, so the programme mix would match only on average and by luck. Systematic sampling picks every k-th student by a fixed interval from a list that carries no information about programme, so the mix is left to chance, and convenience sampling is biased toward whoever is easiest to reach.

TOPIC 4 Hypothesis Testing

6 Qs

Q3

Multiple select

For each situation below, a statistical test is proposed. Which pairings are appropriate? (Select all that apply.)

Note: This is a Multiple Select Question (MSQ). The answer is marked incorrect if any correct option is missed or any incorrect option is selected.

- ☐ A The same 30 riders' average delivery times on weekdays versus weekends: two-sample independent T-test
- ☐ B Whether return status (returned or kept) is associated with city, both categorical variables: one-way ANOVA
- ☒ C Average order value on the old app versus the new app for different customers: two-sample independent T-test ✓ **Correct**
- ☒ D The same 30 riders' delivery times before and after a route-planning workshop: dependent (paired) T-test ✓ **Correct**

EXPLANATION

Two different groups of customers call for the two-sample independent T-test, while the same riders measured twice, before and after a workshop or on weekdays and then on weekends, are dependent samples and call for the paired T-test rather than an independent one. Association between two categorical variables, return status and city, is tested with a chi-square test, not ANOVA, which compares means across groups.

Q5

Multiple select

An analyst tests whether average marks improved this year (H_1 : average marks increased; H_0 : average marks have not increased). The test returns a P-value of 0.03. Which conclusions are justified? (Select all that apply.)

Note: This is a Multiple Select Question (MSQ). The answer is marked incorrect if any correct option is missed or any incorrect option is selected.

- ☒ A At a 1% level of significance, the decision is to fail to reject H_0 because 0.03 is not below 0.01 ✓ **Correct**

B At a 5% level of significance, the decision is to reject H_0 because 0.03 is below 0.05

✓ Correct

C The P-value of 0.03 means there is a 3% probability that H_0 is true, so H_0 is unlikely

D Rejecting H_0 here is evidence against H_0 , but it does not by itself prove that marks improved

✓ Correct

EXPLANATION

The decision rule is to reject H_0 when the P-value is below the level of significance. 0.03 is below 5%, so H_0 is rejected at that level, and 0.03 is not below 1%, so H_0 is not rejected at that level. A small P-value is evidence against H_0 but does not prove the alternate hypothesis. The P-value is the probability of results as extreme as those observed assuming H_0 is true; it is not the probability that H_0 itself is true.

Q13

Single choice

A dairy labels its packets 500 g. A consumer-rights inspector wants evidence that packets are being under-filled, meaning the average fill weight is below 500 g, and will test this with a random sample of packets. How should the hypotheses and the test be set up?

A $H_0: \mu \leq 500$ g and $H_1: \mu > 500$ g, tested one-tailed with the whole rejection region in the upper tail

B $H_0: \mu < 500$ g and $H_1: \mu \geq 500$ g, tested one-tailed with the whole rejection region in the lower tail

C $H_0: \mu = 500$ g and $H_1: \mu \neq 500$ g, tested two-tailed with 2.5% of the rejection region in each tail

D $H_0: \mu \geq 500$ g and $H_1: \mu < 500$ g, tested one-tailed with the whole rejection region in the lower tail

✓ Correct

EXPLANATION

The inspector's claim, that the mean fill weight is below 500 g, is what must be demonstrated, so it goes in the alternate hypothesis $H_1: \mu < 500$ g and everything else goes in $H_0: \mu \geq 500$ g. Because H_1 specifies a direction, the test is one-tailed and the whole rejection region lies in the lower tail. Putting the claim in H_0 reverses the burden of evidence; a two-tailed test with $H_1: \mu \neq 500$ g would also react to over-filling, which is not the claim; and $H_1: \mu > 500$ g tests the opposite direction.

Q26

Single choice

A logistics firm claims its trucks average 18 km per litre. An auditor tests a random sample of 36 trucks and finds a sample mean of 17.2 km per litre with a sample standard deviation of 2.4 km per litre; the population standard deviation is unknown. The test is two-tailed at $\alpha = 0.05$, and the t-table gives critical values of ± 2.030 for 35 degrees of freedom. What is the correct conclusion?

A $t = -2.0$ lies inside ± 2.030 , so H_0 is not rejected; the sample does not contradict the claim

✓ Correct

B $t = -2.0$ lies beyond the critical value ± 1.96 , so H_0 is rejected and the claim is shown to be false

C $t = -2.0$ is negative, so the sample mean is below the claimed mean and H_0 must be rejected

D $t = -2.0$ lies inside ± 2.030 , so H_0 is proven true and the claim is fully confirmed by the sample

EXPLANATION

$t = (\bar{x} - \mu) / (s/\sqrt{n}) = (17.2 - 18) / (2.4/\sqrt{36}) = -0.8 / 0.4 = -2.0$. With 35 degrees of freedom the two-tailed 5% critical values are ± 2.030 , and -2.0 lies between them, so H_0 is not rejected. Comparing with ± 1.96 wrongly uses the Z critical value although the population standard deviation is unknown; the sign of t does not decide the outcome, only whether it falls in the rejection region; and failing to reject H_0 never proves it true.

Q28

Single choice

A pharma team tests H_0 : 'the new tablet does not lower blood pressure' against H_1 : 'the tablet lowers blood pressure' at $\alpha = 0.01$. The trial returns a P-value of 0.03. In reality the tablet does lower blood pressure. Which statement correctly describes the decision and its consequence?

A H_0 is not rejected; since the P-value exceeds α , this is a correct decision and no error occurred

B H_0 is not rejected; since H_0 is actually false, this is a Type II error, whose probability is β

✓ Correct

C H_0 is rejected; since the P-value is small, this is a correct decision with power $1 - \beta$

D H_0 is not rejected; since H_0 is actually false, this is a Type I error, whose probability is α

EXPLANATION

The P-value of 0.03 is not below $\alpha = 0.01$, so the decision rule says fail to reject H_0 . But H_0 is actually false because the tablet works, and failing to reject a false null hypothesis is a Type II error, whose probability is β . A decision is not automatically correct because it follows the rule, since the truth matters too; H_0 is not rejected at all, because 0.03 is greater than 0.01; and a Type I error means rejecting a true H_0 , which is not what happened here.

Q30

Single choice

A quick-commerce analyst wants to know whether mean delivery time differs across 5 warehouses. A colleague proposes running a two-sample T-test for every pair of warehouses. How many T-tests would that take, and which

single test avoids the problem of running so many comparisons?

- ☐ A 20 T-tests; a paired T-test compares all five means in a single test
- ☐ B 5 T-tests; one-way ANOVA compares all five means in a single test
- ☐ C 10 T-tests; a chi-square test compares all five mean delivery times at once
- ☒ D 10 T-tests; one-way ANOVA compares all five means in a single test ✓ Correct

EXPLANATION

Every pair of the 5 warehouses needs its own test: $5C2 = 5! / (2! \times 3!) = 10$. Running 10 separate tests multiplies the chances of a false positive, so a one-way ANOVA, which is based on the F statistic, compares all the means at once. Five tests would give only one comparison per warehouse; a chi-square test relates two categorical variables rather than comparing means; and 20 counts each pair twice, while a paired T-test needs the same subjects measured twice and cannot compare five groups.

TOPIC 5 Model Evaluation Metrics

2 Qs

Q4

Multiple select

A cancer-screening model is tested on 200 patients and produces TP = 16, FP = 24, FN = 4 and TN = 156. Which of the following statements are correct? (Select all that apply.)

Note: This is a Multiple Select Question (MSQ). The answer is marked incorrect if any correct option is missed or any incorrect option is selected.

- ☒ A Precision is 40%, so most of the model's positive predictions are false alarms ✓ Correct
- ☒ B Recall is 80%, so 4 of the 20 patients who have the disease are missed ✓ Correct
- ☐ C Accuracy is 86%, so the model's positive predictions can be trusted
- ☒ D The 24 false positives are Type I errors and the 4 false negatives are Type II errors ✓ Correct

EXPLANATION

Recall = $TP / (TP + FN) = 16 / 20 = 80\%$, so the 4 false negatives are patients with the disease who were missed. Precision = $TP / (TP + FP) = 16 / 40 = 40\%$, so 24 of the 40 positive predictions (60%) are false alarms. A false positive rejects a true null hypothesis (Type I error) and a false negative fails to reject a false one (Type II

error). Accuracy = $(16 + 156)/200 = 86\%$ is arithmetically right, but with only 20 diseased patients out of 200 it can look strong while precision is just 40%, so it does not show that positive predictions are trustworthy.

Q24

Single choice

A bank's loan-default model must not let defaulters slip through, because a missed defaulter costs at least ten times as much as a wrongly rejected applicant. Two candidate models are tested on the same 1,000 applicants, of whom 100 actually default. Model X flags 240 applicants as likely defaulters, and 96 of them do default. Model Y flags 110 applicants, and 77 of them do default. Which model should the bank prefer, and why?

- ☐ A Model Y, because its accuracy of 94.4% is higher, and accuracy alone decides which model is better
- ☐ B Model Y, because its precision of 70% means fewer false alarms, and that matters most when defaulters are costly
- ☒ C Model X, because its recall of 96% means fewer missed defaulters, which are the costlier errors ✓ Correct
- ☐ D Neither, because precision and recall cannot be compared across two different models, only accuracy can

EXPLANATION

A missed defaulter is a false negative (a Type II error), and recall = $TP/(TP + FN)$ measures how many of the actual defaulters the model catches. Model X: $TP = 96$ and $FN = 100 - 96 = 4$, so recall = 96%; $FP = 240 - 96 = 144$, so precision = $96/240 = 40\%$. Model Y: $TP = 77$ and $FN = 23$, so recall = 77%; $FP = 110 - 77 = 33$, so precision = $77/110 = 70\%$. Weighting each missed defaulter at 10 units and each false alarm at 1 unit gives a total cost of $4 \times 10 + 144 = 184$ for Model X against $23 \times 10 + 33 = 263$ for Model Y, and Model X stays cheaper for any cost ratio above about 5.8, so "at least ten times" settles the choice: Model X's much higher recall is preferred despite its lower precision, which only means more false alarms, the cheaper error here. Accuracy is misleading on imbalanced data: Model Y scores $(77 + 867)/1,000 = 94.4\%$ against Model X's $(96 + 756)/1,000 = 85.2\%$, yet Model Y lets 23 defaulters through instead of 4. Precision and recall can be compared across models tested on the same data.

TOPIC 6 Power BI & DAX

6 Qs

Q6

Multiple select

A BI dashboard shows total Sales for each Postal Code, with the bars split by Order Year. Which of the following classifications are correct? (Select all that apply.)

Note: This is a Multiple Select Question (MSQ). The answer is marked incorrect if any correct option is missed or any incorrect option is selected.

- A** Order Year is a dimension here, because it splits the data into groups on the chart ✓ **Correct**
- B** Sales is a measure, because it is aggregated with an operation such as a sum or an average ✓ **Correct**
- C** Postal Code must be treated as a measure, because its values are numbers that can be summed
- D** Postal Code is a dimension even though its values look numeric, because it categorises the rows ✓ **Correct**

EXPLANATION

Dimensions slice and organise the data and become the chart's categories; measures are numeric values that get aggregated with operations such as sum, average or count. Sales is aggregated, so it is a measure. Postal Code is used to categorise rows, so it is a dimension even though its values look numeric, and Order Year is a dimension when it splits the bars into groups; context decides the role. Summing postal codes would be meaningless, so treating Postal Code as a measure is wrong.

Q7

Multiple select

A team builds a sales report in Power BI. Which of the following steps or definitions are correct? (Select all that apply.)

Note: This is a Multiple Select Question (MSQ). The answer is marked incorrect if any correct option is missed or any incorrect option is selected.

- A** Define total orders by summing the Quantity column, because every unit sold counts as one order
- B** Relate the sales table to a calendar table through the date column so that year, quarter and month can slice the data ✓ **Correct**
- C** Define profit margin percentage as profit divided by total sales, and format it as a percentage ✓ **Correct**
- D** Add slicers for region, product and salesperson so that the report's visuals respond to the selected values ✓ **Correct**

EXPLANATION

A calendar table related to the sales table by date supplies year, quarter and month for slicing. Slicers for region, product and salesperson filter the report, and its visuals respond to the selected values. Profit margin percentage is profit divided by total sales, formatted as a percentage. Total orders is a count of distinct orders, not a sum of the Quantity column, because one order can contain many units.

Q12

Single choice

A retail chain's daily-revenue dashboard is opened by about 300 managers every morning. Its source is a large transactional warehouse that also runs the billing system, and the managers are happy with data that is at most an hour old. Which connection choice best fits, and why?

- ☐ A Live connection: every dashboard view should always show the very latest data, whatever the load on the warehouse
- ☒ B Extract with a scheduled refresh: an hourly lag is acceptable, and it shields the warehouse from constant queries ✓ **Correct**
- ☐ C Live connection: it renders faster than pre-computed snapshots and keeps every number current to the second
- ☐ D Extract with a scheduled refresh: it queries the source in real time on every dashboard view, so managers see the newest rows

EXPLANATION

The managers accept data that is up to an hour old and the source also serves billing, so the default choice is an extract: data is downloaded and pre-computed at scheduled intervals, dashboards render quickly from stored snapshots, and the warehouse is protected from 300 people triggering live queries. A live connection queries the source on every view, which is slower and loads the infrastructure, so it is kept for real-time needs; claiming that live renders faster reverses the trade-off; and an extract does not query the source on each view.

Q16

Single choice

An Orders table has one row per order, and its Customer_ID column reads C1, C2, C1, C3, C2, C1 in row order. Two DAX measures are defined:

```
Count of Orders = COUNTROWS(Orders)
Distinct Customers = DISTINCTCOUNT(Orders[Customer_ID])
```

A slicer then restricts the report to customers C1 and C3 only. What values do the two measures show?

- ☐ A 2 and 4
- ☐ B 4 and 4
- ☒ C 4 and 2 ✓ **Correct**
- ☐ D 6 and 3

EXPLANATION

After the slicer keeps only customers C1 and C3, the remaining Customer_ID values are C1, C1, C3, C1: four rows. COUNTROWS(Orders) counts rows, so Count of Orders is 4. DISTINCTCOUNT(Orders[Customer_ID]) counts unique values, C1 and C3, so Distinct Customers is 2. The pair 6 and 3 ignores the slicer, 4 and 4 treats DISTINCTCOUNT like COUNTROWS, and 2 and 4 swaps the two measures.

Q18

Single choice

In a Power BI sales table, a calculated column named Profit is defined as $\text{Quantity} \times \text{Unit Price} - \text{Discount} - \text{Shipping Cost}$, where Discount and Shipping Cost are order-level amounts. All amounts are in INR. One order has Quantity = 6, Unit Price = 250, Discount = 150 and Shipping Cost = 100. What value does the Profit column show for this row?

A 1,750**B** 1,250

✓ Correct

C 1,350**D** 1,400**EXPLANATION**

$\text{Quantity} \times \text{Unit Price} = 6 \times 250 = 1,500$. Subtracting the discount of 150 and the shipping cost of 100 gives $1,500 - 150 - 100 = 1,250$. The value 1,750 adds the two deductions instead of subtracting them, 1,350 subtracts only the discount, and 1,400 subtracts only the shipping cost.

Q22

Single choice

In a Power BI model, the Customers table stores IDs such as 'CUST-001' while the Orders table stores the same customers as 1, 2, 3, so a relationship between them cannot match. Separately, management wants a total-profit figure calculated row by row as $\text{units sold} \times (\text{revenue per unit} - \text{cost per unit})$ and reflecting whichever filters the report user applies. How should each need be handled?

A Do both in DAX, since Power Query cannot change the values inside a column and can only rename columns**B** Reformat the IDs with a DAX measure after loading, and compute the profit in Power Query before loading**C** Reformat the IDs in Power Query before loading, and write the profit as a DAX SUMX measure after loading

✓ Correct

- D** Do both in Power Query, since DAX cannot perform any row-by-row calculation and only aggregates whole columns

EXPLANATION

Power Query cleans and reshapes data before it is loaded, so turning 'CUST-001' into 1 (for example with find-and-replace or by splitting the column) belongs there and lets the tables join, because matching depends on the values rather than the column names. DAX works after the data is loaded, and SUMX performs a row-by-row calculation before summing, which suits a profit measure that responds to report filters. The reverse assignment misuses both tools, Power Query can find-and-replace and split column values, and DAX SUMX does iterate over rows.

TOPIC 7 Data Visualisation & Storytelling

2 Qs

Q8

Multiple select

A public-facing report chart uses a red-green rainbow palette, 3D bars, heavy gridlines and a legend mapping eight colours to eight bars. Which changes follow sound data-visualisation practice for clarity and accessibility? (Select all that apply.)

Note: This is a Multiple Select Question (MSQ). The answer is marked incorrect if any correct option is missed or any incorrect option is selected.

- A** Replace the 3D bars with flat bars and label each bar directly instead of relying on the legend ✓ **Correct**
- B** Start the value axis just below the smallest bar so that the differences between bars look larger
- C** Use gray for the context bars and a single accent colour for the one bar that carries the message ✓ **Correct**
- D** Add patterns or symbols alongside colour and avoid red-green pairings, so colour-blind readers can tell bars apart ✓ **Correct**

EXPLANATION

3D graphs are hard to read on a 2D surface, and labelling bars directly is preferred to a legend that makes readers juggle between legend and graph. Gray for context with a single accent colour builds a visual hierarchy and directs attention to the message. About 8% of people are colour-blind, so meaning should never rely on colour alone and red-green pairings should be avoided. Starting the axis near the smallest bar distorts the perception of magnitude, since axes should start at zero.

Q27

Single choice

A data analyst must show the C-suite that monthly churn rose from 4.1% to 6.2% after a pricing change. Which chart design best follows sound visual-storytelling practice for this audience?

- ☐ A A full table of every month with gridlines and small text, and a topic title such as 'Churn Data 2025'
- ☐ B A 3D column chart of twelve business metrics, with gridlines and a legend, and a topic title such as 'Monthly Metrics'
- ☐ C A rainbow-coloured line chart of eight product lines, with a legend and a topic title such as 'Churn by Product Line'
- ☒ D One line chart with post-change months in red and earlier months in gray, and a headline title stating the rise from 4.1% to 6.2% ✓ **Correct**

EXPLANATION

Executives need the conclusion and the key number, with a title that states the 'so what'. One line chart that uses an accent colour for the months that matter and gray for context, under a headline title, delivers a single message that can be grasped in about ten seconds. The 3D chart with gridlines and a legend adds clutter and dilutes the message; the rainbow palette gives every series equal weight so nothing stands out; and a full table under a topic title forces the reader to hunt for the message.

TOPIC 8 Seaborn & Plotly

2 Qs

Q2

Multiple select

Which of the following statements about Seaborn's multi-panel plots are correct? (Select all that apply.)

Note: This is a Multiple Select Question (MSQ). The answer is marked incorrect if any correct option is missed or any incorrect option is selected.

- ☒ A In a pair plot, the diagonal panels show the distribution of each variable rather than scatter plots ✓ **Correct**
- ☒ B Adding `hue=` to a pair plot colours the points by a categorical variable, adding a further dimension ✓ **Correct**
- ☐ C A facet grid finds the relationships among all numeric columns automatically, so no splitting variable is chosen
- ☐ D With 100 or more categories a facet grid stays easy to read, because every category simply gets its own panel

EXPLANATION

A pair plot draws every numeric variable against every other numeric variable, and its diagonal shows each variable's own distribution; hue= colours the points by a categorical variable, which adds a dimension. A facet grid is the opposite: the analyst must choose which categorical variable to split on with col= or row=, and it compares one relationship across categories rather than discovering all relationships. With 100 or more categories a facet grid produces far too many panels, which is exactly its trade-off.

Q25

Single choice

An analyst holds a DataFrame `df` with five numeric columns and two categorical columns, `region` (3 distinct values) and `plan` (2 distinct values), and every region-plan combination occurs in the data. She runs:

```
sns.pairplot(df[["orders", "basket", "delivery_min", "returns", "rating"]], corner=True)
sns.relplot(data=df, x="orders", y="delivery_min", col="region", row="plan")
```

How many panels (cells of the plot grids) do the two calls draw in total?

A 21

✓ Correct

B 31**C** 16**D** 20**EXPLANATION**

The pair plot lays out a 5×5 grid, but `corner=True` removes the redundant mirrored panels above the diagonal, leaving the 5 diagonal panels (each variable's own distribution) and the 10 below it: 15 panels. The `relplot` with `col=region` and `row=plan` draws one panel per combination: $3 \times 2 = 6$. The total is $15 + 6 = 21$. Forgetting `corner=True` gives $25 + 6 = 31$; dropping the diagonal gives $10 + 6 = 16$; and adding the rows and columns ($3 + 2 = 5$) instead of multiplying them gives $15 + 5 = 20$.

TOPIC 9 Case Studies (Subjective)

2 Qs

Q31

Subjective

Problem Statement

Kaveri Fresh, a grocery chain, runs 12 stores in three regions and two formats (Express and Mega). The regional head wants to see how four store metrics relate to one another, whether the picture changes by region and format, and then

to present one clear result to the executive team. The analytics team works in Python with a DataFrame `df` built from the data below. Seaborn and Plotly Express are free, open-source libraries; no account or API key is needed.

Dataset

store	region	format	daily_orders	avg_basket_inr	delivery_min	return_pct
Indiranagar	South	Express	820	410	21	2.4
Adyar	South	Express	760	455	24	2.9
Whitefield	South	Mega	360	1180	46	6.1
Velachery	South	Mega	310	1290	52	6.8
Andheri	West	Express	900	380	19	2.1
Baner	West	Express	690	480	27	3.3
Vashi	West	Mega	410	1050	41	5.4
Kothrud	West	Mega	280	1340	55	7.2
Dwarka	North	Express	640	505	29	3.6
Noida	North	Express	710	440	25	3.0
Gurugram	North	Mega	390	1120	44	5.8
Rohini	North	Mega	300	1250	50	6.5

Ready-to-run initializer (identical values):

```
import pandas as pd

df = pd.DataFrame({
    "store": ["Indiranagar", "Adyar", "Whitefield", "Velachery", "Andheri", "Baner", "Vashi",
              "Kothrud", "Dwarka", "Noida", "Gurugram", "Rohini"],
    "region": ["South", "South", "South", "South", "West", "West", "West", "West", "North",
              "North", "North", "North"],
    "format": ["Express", "Express", "Mega", "Mega", "Express", "Express", "Mega", "Mega",
              "Express", "Express", "Mega", "Mega"],
    "daily_orders": [820, 760, 360, 310, 900, 690, 410, 280, 640, 710, 390, 300],
    "avg_basket_inr": [410, 455, 1180, 1290, 380, 480, 1050, 1340, 505, 440, 1120, 1250],
    "delivery_min": [21, 24, 46, 52, 19, 27, 41, 55, 29, 25, 44, 50],
    "return_pct": [2.4, 2.9, 6.1, 6.8, 2.1, 3.3, 5.4, 7.2, 3.6, 3.0, 5.8, 6.5],
})
```

Tasks

Task 1 (Themed pair plot): Set Seaborn's theme to a white background with light grid lines and a presentation-sized context. Then write code for one figure showing every pairwise relationship among the four numeric metrics, with the two store formats in different colours and without the redundant mirrored panels above the diagonal. State how many panels the grid contains and what the diagonal panels show.

Task 2 (Regional view): Write code that plots `daily_orders` (x-axis) against `delivery_min` (y-axis) as a scatter, with a separate panel for each region and the points coloured by store format. State how many panels appear, and what the panels show about Express versus Mega stores in each region.

Task 3 (Interactive figure): Write code for an interactive scatter for the executive review: `avg_basket_inr` on the x-axis and `delivery_min` on the y-axis, points coloured by store format and sized by `daily_orders`, with the store name shown when the viewer hovers over a point. Say in one sentence what interactivity adds over a static image.

Task 4 (Audience and message): The executives give each chart about ten seconds. (a) Write one quantified, data-supported sentence as the message of the executive chart, and a headline title for it. (b) Give three design choices for that chart, each tied to a design principle. (c) Say which of your figures from Tasks 1 to 3 suits the analysts and which suits the executives, with one reason for each.

Submission Format

Paste one code block covering Tasks 1 to 3 in the input box, followed by your written answers to the other parts of Tasks 1 to 4.

Note: You may use any IDE (online: Google Colab, Jupyter Notebook; offline: VS Code, PyCharm, or similar) to write, run, and test your code. You must **disable all AI-based extensions** (GitHub Copilot, Tabnine, Codeium, etc.) before starting. AI tools of any kind are not permitted and will lead to disqualification. **Submit your complete code as a copy-pasted snippet in the input box — do not submit a URL, notebook link, GitHub repository, or Google Drive file.**

MODEL ANSWER

Question Explanation

This question tests whether a student can turn a multi-variable dataset into the right Seaborn and Plotly figures for two audiences: a pair plot with `hue` and `corner=True` for exploration, a `relplot` facet grid with `col=` for the regional comparison, and an interactive `px.scatter` with `colour`, `size` and `hover` information for the executive review. It then tests the design reasoning behind the executive chart: one quantified message, a headline title, gray for context with a single accent colour, reduced clutter and colour-blind-safe encoding.

Short Explanation

Set a presentation theme with `sns.set_style` and `sns.set_context`, then draw `sns.pairplot(df, hue="format", corner=True)`: a 4×4 grid reduced to 10 panels (4 diagonal distributions plus 6 below the diagonal). Use `sns.relplot(..., col="region", hue="format")` for three regional panels, and `px.scatter` with `color`, `size` and `hover_name` for the interactive executive chart. The data support one message: Express stores deliver in about 24 minutes on average against about 48 for Mega stores, in every region.

Long Explanation

Task 1. `sns.set_style("whitegrid")` gives a white background with light grid lines and `sns.set_context("talk")` scales fonts and line widths for a presentation. `sns.pairplot` plots every numeric column against every other one; the four numeric metrics (`daily_orders`, `avg_basket_inr`, `delivery_min`, `return_pct`) give a 4×4 grid of 16 panels. `hue="format"` colours the two formats, and `corner=True` removes the redundant mirrored panels above the diagonal, leaving the 4 diagonal panels plus the 6 below it: 10 panels. The diagonal panels show each variable's own distribution, not scatter plots.

Task 2. `sns.relplot` draws a scatter on a facet grid; `col="region"` splits it into one panel per region, so there are 3 panels, and `hue="format"` keeps the format colours. In every region the Express stores sit at higher daily orders (640 to 900) and lower delivery times (19 to 29 minutes) than the Mega stores (280 to 410 orders and 41 to 55 minutes), so the pattern is not an artefact of one region.

Task 3. `px.scatter` with `x="avg_basket_inr"`, `y="delivery_min"`, `color="format"`, `size="daily_orders"` and `hover_name="store"` gives an interactive bubble chart. Interactivity adds hover information, zoom, pan and the ability to toggle series without cluttering the static plot.

Task 4. The average delivery time is 24.2 minutes for the six Express stores and 48.0 minutes for the six Mega stores, about half, and this holds in all three regions. A one-sentence message with a headline title: "Express stores deliver in about 24 minutes, half the 48 minutes of Mega stores, in every region." Design choices for the executive chart: (1) gray for the Mega points and one accent colour for the Express points, so colour contrast directs attention to the message; (2) label the two groups directly and remove the legend, heavy gridlines and borders, following the data-ink principle; (3) do not rely on colour alone, by using different marker symbols and a colour-blind-safe pair rather than red and green. The pair plot and the facet grid suit the analysts, who can read many relationships and supporting evidence; the executives need a single message they can read in ten seconds, so a single-message chart such as the annotated scatter suits them.

Model Solution

```
import seaborn as sns
import plotly.express as px

# Task 1: presentation theme, then the pair plot without mirrored panels
sns.set_style("whitegrid")
sns.set_context("talk")
sns.pairplot(df, hue="format", corner=True)

# Task 2: one panel per region, points coloured by store format
sns.relplot(data=df, x="daily_orders", y="delivery_min", col="region", hue="format")

# Task 3: interactive scatter for the executive review
fig = px.scatter(
    df,
    x="avg_basket_inr",
    y="delivery_min",
    color="format",
    size="daily_orders",
    hover_name="store",
)
fig.show()
```

Written answers

- **Task 1:** The pair plot grid has 10 panels: 4 diagonal panels plus 6 below the diagonal, because `corner=True` removes the mirrored upper triangle. The diagonal panels show each variable's own distribution rather than scatter plots.
- **Task 2:** Three panels appear, one per region. In every region the Express stores have higher daily orders and lower delivery times than the Mega stores.
- **Task 3:** Interactivity adds hover details, zoom, pan and toggling of series without cluttering the static plot.
- **Task 4(a):** Message: "Express stores deliver in about 24 minutes on average, half the roughly 48 minutes of Mega stores, in every region." Headline title: "Express stores deliver twice as fast as Mega stores: 24 versus 48 minutes."
- **Task 4(b):** Gray for the Mega points and one accent colour for the Express points; direct labels instead of a legend, with gridlines and borders removed; symbols or patterns plus a colour-blind-safe pair, avoiding red-green combinations.
- **Task 4(c):** The pair plot and facet grid suit the analysts because they expose many relationships at once. The single-message annotated (or interactive) scatter suits the executives, who need one takeaway in about ten seconds.

Q32

Subjective

Problem Statement

NutriBox, a meal-kit start-up, advertises that its orders reach customers in 30 minutes on average. An auditor suspects that deliveries are actually slower than advertised and records the delivery times (in minutes) of a random sample of 25 orders. The sample mean is $\bar{x} = 32.85$ minutes and the sample standard deviation is $s = 7.5$ minutes. The population standard deviation is unknown, and the auditor works at a 5% level of significance ($\alpha = 0.05$).

Critical values from the t-table for 24 degrees of freedom: **1.711** for a one-tailed test at $\alpha = 0.05$, and **± 2.064** for a two-tailed test at $\alpha = 0.05$.

Tasks

Task 1 (Frame the test): State H_0 and H_1 for the auditor's suspicion. Say whether the test is one-tailed or two-tailed and where the rejection region lies, and justify why a T-test rather than a Z-test is the right choice here.

Task 2 (Decide): Compute the degrees of freedom, the standard error and the t statistic. Using the appropriate critical value, state your decision and what it means for NutriBox in plain business words. Then state the decision you would have reached with the two-tailed critical value instead, and explain why the two decisions differ.

Task 3 (Errors): In the context of this audit, describe a Type I error and a Type II error in business terms, and state the consequence of each. Say which of the two has probability α and what $1 - \beta$ represents.

Task 4 (Choose the test): Name the most suitable test, with a one-line reason, for each situation: (i) the delivery times of the same 25 riders are compared before and after a route-planning workshop; (ii) delivery times are compared across NutriBox's six city hubs (also state how many pairwise two-sample T-tests this would otherwise need); (iii) the auditor asks whether a late delivery (yes or no) is associated with the city.

Task 5 (Late-order model): NutriBox also builds a model that flags orders likely to be late. On 400 test orders, 60 are actually late; the model flags 50 orders, and 40 of the flagged orders are actually late. Build the confusion matrix (TP, FP, FN, TN), compute the accuracy, precision and recall, say which error type (Type I or Type II) the false positives correspond to, and explain why accuracy alone gives a misleading picture of this model.

Submission Format

Write your complete answer as numbered responses to Tasks 1 to 5 in the input box provided. Show every calculation step.

Note: Write your answer in clear, concise English. Bullet points and structured sections are encouraged. Do not paste external links or screenshots — write the full response in the input box provided. AI tools of any kind are not permitted and will lead to disqualification.

MODEL ANSWER

Question Explanation

This question tests the full one-sample hypothesis-testing workflow on a claim that has a direction: framing H_0 and H_1 , choosing a T-test when the population standard deviation is unknown, computing the t statistic from summary statistics, reading the right critical value for one-tailed versus two-tailed tests, and interpreting the decision. It then checks the two error types, the choice of test for related situations (paired, more than two groups, two categorical variables), and the link between hypothesis-testing errors and the confusion matrix and its metrics.

Short Explanation

$H_0: \mu \leq 30$ and $H_1: \mu > 30$ make this a one-tailed T-test with the whole 5% rejection region in the upper tail. With $df = 24$, $SE = 7.5/\sqrt{25} = 1.5$ and $t = (32.85 - 30)/1.5 = 1.9$, which exceeds the one-tailed critical value 1.711, so H_0 is rejected: there is evidence that deliveries take longer than 30 minutes. With the two-tailed cut-offs of ± 2.064 , 1.9 would fall inside the acceptance region and H_0 would not be rejected. A Type I error (probability α) wrongly rejects a true H_0 ; a Type II error (probability β) misses a real slowdown. The late-order model has $TP = 40$, $FP = 10$, $FN = 20$, $TN = 330$, so accuracy 92.5%, precision 80% and recall about 66.7%.

Long Explanation

Task 1. The auditor's suspicion is that deliveries are slower than advertised, so it belongs in the alternate hypothesis: $H_1: \mu > 30$ minutes. Everything else goes in $H_0: \mu \leq 30$ minutes (writing $H_0: \mu = 30$ states the same status quo). Because H_1 specifies a direction, the test is one-tailed and the entire 5% rejection region sits in the upper tail. The population standard deviation is unknown and is estimated by the sample standard deviation $s = 7.5$, so a T-test is used rather than a Z-test; the T-distribution allows for the extra uncertainty of estimating σ from only 25 orders.

Task 2. Degrees of freedom = $n - 1 = 24$. Standard error = $s/\sqrt{n} = 7.5/\sqrt{25} = 7.5/5 = 1.5$. Then $t = (\bar{x} - \mu)/(s/\sqrt{n}) = (32.85 - 30)/1.5 = 2.85/1.5 = 1.9$. The one-tailed critical value for 24 degrees of freedom at $\alpha = 0.05$ is 1.711. Because $1.9 > 1.711$, the statistic lies in the rejection region and H_0 is rejected: at the 5% level the sample gives evidence, not proof, that NutriBox's average delivery time is longer than the advertised 30 minutes. With the two-tailed critical values ± 2.064 , the statistic 1.9 lies between -2.064 and $+2.064$, so H_0 would not be rejected. The decisions differ because a two-tailed test splits the 5% into 2.5% per tail, pushing the upper cut-off out from 1.711 to 2.064, while the one-tailed test puts all 5% in the upper tail where the auditor's suspicion points.

Task 3. A Type I error is rejecting H_0 when it is true: concluding that NutriBox is slower than 30 minutes when its true average is within the claim, which could bring a wrongful penalty or a needless operational overhaul. A Type II error is failing to reject H_0 when it is false: concluding nothing is wrong when deliveries really are slower, so customers keep receiving late orders and the inflated claim stands. The probability of a Type I error is $\alpha = 0.05$. The probability of a Type II error is β , and $1 - \beta$ is the power of the test, the probability of detecting a real slowdown.

Task 4. (i) The same 25 riders measured before and after form dependent samples, so a dependent (paired) T-test applies. (ii) Six hubs call for a one-way ANOVA, based on the F statistic; comparing every pair of six hubs would need $6C2 = 15$ two-sample T-tests, and ANOVA replaces them with one test that avoids inflating the false-positive risk. (iii) Late delivery (yes or no) and city are two categorical variables, so the chi-square test of association applies.

Task 5. $TP = 40$ (flagged and late), $FP = 50 - 40 = 10$ (flagged but not late), $FN = 60 - 40 = 20$ (late but not flagged) and $TN = 400 - 40 - 10 - 20 = 330$. Accuracy = $(40 + 330)/400 = 92.5\%$; precision = $40/(40 + 10) = 80\%$; recall = $40/(40 + 20) \approx 66.7\%$. The false positives are Type I errors and the false negatives are Type II errors. Accuracy misleads because only 60 of the 400 orders are late: a model that flagged nothing would already score $340/400 = 85\%$, so 92.5% looks strong while recall shows that 20 of the 60 late orders, one in three, are still missed.

Model Solution

1. Framing. $H_0: \mu \leq 30$ min (status quo). $H_1: \mu > 30$ min (the auditor's suspicion). One-tailed test, whole 5% rejection region in the upper tail. T-test, because σ is unknown and $s = 7.5$ is used in its place.

2. Decision.

- $df = 25 - 1 = 24$

- $SE = 7.5 / \sqrt{25} = 1.5$
- $t = (32.85 - 30) / 1.5 = 1.9$
- One-tailed critical value = 1.711; $1.9 > 1.711$, so reject H_0 . At the 5% level there is evidence that the average delivery time exceeds 30 minutes.
- With the two-tailed critical value ± 2.064 , 1.9 lies inside the interval, so the decision would be to fail to reject H_0 . The two-tailed test splits 5% into 2.5% per tail, which moves the upper cut-off out to 2.064; the one-tailed test keeps all 5% in the upper tail.

3. Errors.

- Type I: concluding deliveries are slower than 30 minutes when they are not (rejecting a true H_0); consequence: a wrongful penalty or an unnecessary overhaul. Probability = $\alpha = 0.05$.
- Type II: failing to detect a real slowdown (not rejecting a false H_0); consequence: customers keep getting late orders and the claim stands. Probability = β , and power = $1 - \beta$.

4. Tests. (i) Dependent (paired) T-test, because the same riders are measured twice. (ii) One-way ANOVA; 6C2 = 15 pairwise T-tests would otherwise be needed. (iii) Chi-square test of association between two categorical variables.

5. Late-order model.

	Actually late	Actually on time
Flagged late	TP = 40	FP = 10
Not flagged	FN = 20	TN = 330

Accuracy = $370/400 = 92.5\%$; precision = $40/50 = 80\%$; recall = $40/60 \approx 66.7\%$. False positives are Type I errors (false negatives are Type II). Accuracy misleads because late orders are only 60 of 400: predicting no late orders would score 85%, and the 92.5% hides that a third of the late orders are missed.