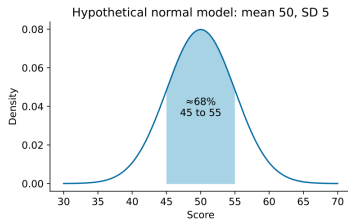


Data displays, normal distributions and association

Original skill preparation; versioned source scope

Method and conditions



- A frequency counts repeated observations.
- Ordered data; frequency; mean and median.

Worked method 1

Read and reason

A frequency counts repeated observations. Values 1, 3, 7 with frequencies 2, 1, 1 represent four observations: 1, 1, 3, 7. $\bar{x} = \sum fx / \sum f = (2(1) + 1(3) + 1(7))/4 = 3$. The median is $(1 + 3)/2 = 2$; it need not equal the mean.

Original context 1

Read and reason

Task context

Read and reason

Each task below states its own complete data. Values may differ from the worked example.

Attempt 1.1

Read and reason

A normal model has mean 80 and standard deviation 6. Give approximate central 68% and 95% intervals. Every score is changed to $z = 2x + 5$; give the transformed mean and standard deviation.

Reasoning 1.1

Read and reason

Under the stated normal model, about 68% lie in 74–86 and 95% in 68–92.

Transformed mean is $2(80) + 5 = 165$ and standard deviation $2(6) = 12$. These are model proportions, not guarantees for a finite sample.

Attempt 2.1

Read and reason

A fitted model is $y = 4x + 3$. At $x = 5$ observed $y = 20$. Find predicted output and residual. Can a positive slope alone prove causation?

Reasoning 2.1

Read and reason

Prediction is $4(5) + 3 = 23$. Residual $e = y_{obs} - y_{pred} = 20 - 23 = -3$. A positive slope gives association under the fit; confounding or selection can explain it without a direct cause.

Attempt 3.1

Read and reason

Data are 2,4,4,6,9,11,12,16. Using medians of the lower and upper halves, find the five-number summary and interquartile range. Which display preserves this summary: a box plot, a category bar chart or a scatterplot?

Reasoning 3.1

Read and reason

Median $Q_2 = (6 + 9)/2 = 7.5$. Lower half 2,4,4,6 has $Q_1 = 4$; upper half 9,11,12,16 has $Q_3 = 11.5$. Five-number summary is (2, 4, 7.5, 11.5, 16) and $IQR = Q_3 - Q_1 = 7.5$. A box plot displays these values. A category bar chart compares categorical counts; a scatterplot needs paired quantitative measurements.

Reasoning 3.2

Read and reason

Neither directly displays this univariate five-number summary. The stated quartile convention prevents disagreement between other valid conventions.

Attempt 4.1

Read and reason

A hypothetical histogram has equal-width bins $[0,10)$, $[10,20)$, $[20,30]$ with frequencies 2,6,2. State the proportion below 20 and whether the exact mean is determined. A scatterplot of hours and output rises with a curved trend and one high outlier: explain why a straight line joining every point is unsuitable.

Reasoning 4.1

Read and reason

Below 20 there are $2+6=8$ of 10 observations, so 80%. The exact observations within bins are unknown, so no exact mean follows; midpoint estimates must be labelled estimates. A scatterplot represents pairs, and a fitted relation should describe its main shape with the unusual point considered. Joining every observation invents a path between separate cases and ignores the curved trend.

Reasoning 4.2

Read and reason

Inspect actual axis scales and the outlier before fitting or extrapolating.

Transfer and limits

Read and reason

Ordered data; frequency; mean and median.

Check yourself

Read and reason

Close the worked reasoning. Explain why the method fits the final transfer task, reconstruct its conclusion, and state a condition that would change your answer. Then check the full reasoning in the reference.