Interpreting data from Scatter plot - TARSOG

Mana Maths

From Pattern to a Contextual Conclusion

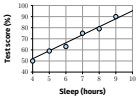
Key idea

State what tends to happen to the **response variable** as the explanatory variable changes. Use “tends to”, because points vary.

Conclusion

Students who sleep longer **tend to** have higher test scores in this sample. The strong positive association supports comparison, not a causal claim.

Conclusion = context + direction + uncertainty



Common mistake

“Sleep causes higher scores” is not justified by a scatter plot alone.

Line of Best Fit: Estimate, Then Interpret

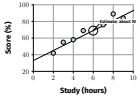
Key idea

A sensible fit follows the centre of a roughly linear cloud, with points balanced above and below. Read across from x, up to the line, then across to y.

Interpolation

At 6 hours, the line predicts about **70%**. This is interpolation because 6 is inside the observed range 2–9 hours. A reasonable graph-reading range is 68–72%.

A prediction is on the line, not necessarily a point



Interpolation Is Safer Than Extrapolation

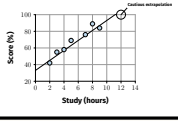
Key idea

Interpolation predicts within the observed *x*-range. **Extrapolation** extends beyond it, where the pattern may change. The farther outside, the weaker the evidence.

Cautious wording

A prediction at 12 hours extends three hours beyond the data. It may be unrealistic because scores cannot keep rising past 100%.

The fitted line is a model, not a guarantee



Variation, Outliers, and Groups Limit Conclusions

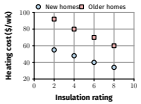
Key idea

Vertical gaps from the line show prediction error. Outliers can pull a fit; groups can make one overall line misleading. Investigate the grouping variable before combining data.

Lurking variable

Older and newer houses form separate groups. House age may influence both insulation rating and heating cost, so insulation alone is not a proven cause.

Separate groups may need separate models



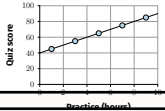
Common mistake

A single overall line can describe neither group well, even when it looks convincing.

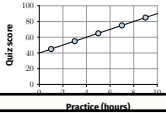
Start Tasks

Identify the explanatory and response variables, then read the trend and association from a scatter plot.

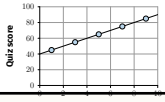
1. Which variable is explanatory (independent) and which is response (dependent) here?



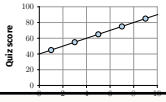
2. Describe the association (positive/negative) and the trend (linear/non-linear) shown.



3. Using the line of best fit, estimate the quiz score for 6 hours of practice.



4. Is a prediction at 6 hours interpolation or extrapolation? Explain why.



Start Tasks — Answers

1. Which variable is explanatory (independent) and which is response (dependent) here?

Solution

Practice (hours) is explanatory; Quiz score is response

2. Describe the association (positive/negative) and the trend (linear/non-linear) shown.

Solution

Positive association; linear trend

3. Using the line of best fit, estimate the quiz score for 6 hours of practice.

Solution

Score $= 5(6) + 40 = 70$

4. Is a prediction at 6 hours interpolation or extrapolation? Explain why.

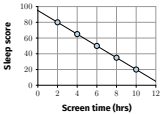
Solution

Interpolation – 6 hrs is inside the data range (1-9)

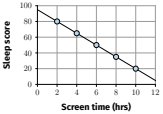
Start Tasks

Read a negative-association scatter plot, then decide whether a further prediction is safe to make.

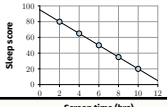
5. Which variable is explanatory and which is response in this graph?



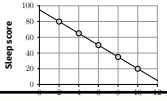
6. Describe the association and the trend shown.



7. Using the line of best fit, estimate the sleep score for 6 hours of screen time.



8. A prediction is made for 11 hours. Is this interpolation or extrapolation? Explain.



Start Tasks — Answers

5. Which variable is explanatory and which is response in this graph?

Solution

Screen time is explanatory; Sleep score is response

6. Describe the association and the trend shown.

Solution

Negative association; linear trend

7. Using the line of best fit, estimate the sleep score for 6 hours of screen time.

Solution

Score $= 95 - 7.5(6) = 50$

8. A prediction is made for 11 hours. Is this interpolation or extrapolation? Explain.

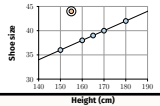
Solution

Extrapolation – 11 hrs is outside the data range (2-10)

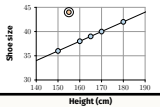
Build Tasks

Judge relationship strength, then quantify an outlier's gap from the line of best fit.

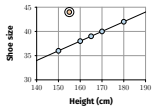
1. Judge the strength (weak/moderate/strong) of this relationship, ignoring the outlier.



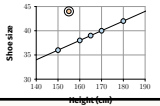
2. Identify the outlier's coordinates. By how much does it exceed the line's prediction at that height?



3. Predict the shoe size at 175 cm. Is this interpolation or extrapolation?



4. Predict the shoe size at 195 cm. Why might this prediction be unreliable?



Build Tasks — Answers

1. Judge the strength (weak/moderate/strong) of this relationship, ignoring the outlier.

Solution

Strong – the 5 main points lie on a line

2. Identify the outlier's coordinates. By how much does it exceed the line's prediction at that height?

Solution

$(155, 44)$, predicted 37. Exceeds by 7

3. Predict the shoe size at 175 cm. Is this interpolation or extrapolation?

Solution

Size $= 0.2(175) + 6 = 41$. Interpolation (within 150-180)

4. Predict the shoe size at 195 cm. Why might this prediction be unreliable?

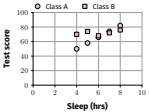
Solution

Size $= 0.2(195) + 6 = 45$. Extrapolation – beyond 180cm

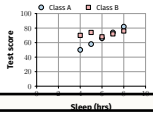
Build Tasks

Compare two keyed groups on one scatter plot.

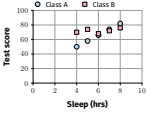
5. What does the key tell you about the grouping shown here?



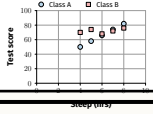
6. Which class shows the stronger association between sleep and score? Justify your answer.



7. Estimate Class A's test score for 6.5 hours of sleep, using their trend.



8. Explain why one combined line of best fit through both classes would be misleading.



Build Tasks — Answers

5. What does the key tell you about the grouping shown here?

Solution

Circles = Class A, squares = Class B – two groups plotted together

6. Which class shows the stronger association between sleep and score? Justify your answer.

Solution

Class A – its points sit exactly on a line; Class B is scattered

7. Estimate Class A's test score for 6.5 hours of sleep, using their trend.

Solution

Score $= 8(6.5) + 18 = 70$

8. Explain why one combined line of best fit through both classes would be misleading.

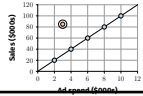
Solution

Their trends differ – one line hides both

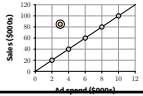
Push Tasks

Critique a flawed claim about the line of best fit, using a viral-outlier dataset.

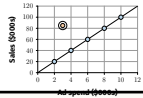
1. Describe the trend, association, and strength of the main cluster, ignoring the outlier.



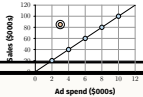
2. Find the outlier's coordinates. By how much do its actual sales exceed the trend's prediction?



3. "Every extra $1000 spent always adds exactly $10,000 in sales." Give two reasons this is wrong.



4. Would you trust this line's prediction for a $30,000 spend? Justify, referring to the data range shown.



Push Tasks — Answers

1. Describe the trend, association, and strength of the main cluster, ignoring the outlier.

Solution

Linear, positive, strong – points lie on a line

2. Find the outlier's coordinates. By how much do its actual sales exceed the trend's prediction?

Solution

$(3, 85)$, predicted 30. Exceeds by 55

3. "Every extra $1000 spent always adds exactly $10,000 in sales." Give two reasons this is wrong.

Solution

(1) Only 5 points; (2) the outlier already breaks it

4. Would you trust this line's prediction for a $30,000 spend? Justify, referring to the data range shown.

Solution

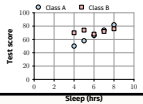
No – $30k is outside the $10k data range

Push Tasks

Critique an explanatory/response mix-up, then quantify and compare the two groups' trends.

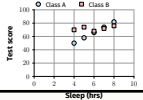
5. A student swaps the axes, then calls sleep the response variable. Explain why this is wrong.

6. Calculate the % increase in Class A's predicted score from 4 to 8 hours of sleep.



7. Class B's points scatter more than Class A's. What does this mean for trusting a Class B prediction?

8. Predict Class A's score at 7.5 hours sleep. Is this interpolation or extrapolation?



Push Tasks — Answers

5. A student swaps the axes, then calls sleep the response variable. Explain why this is wrong.

Solution

Sleep predicts score, so sleep is explanatory (x-axis) and score is response (y-axis)

6. Calculate the % increase in Class A's predicted score from 4 to 8 hours of sleep.

Solution

Scores: $50 \to 82$ $(+32)$.

% increase

$= 32/50 = 64\%$

7. Class B's points scatter more than Class A's. What does this mean for trusting a Class B prediction?

Solution

More scatter $=$ weaker fit – Class B predictions are less reliable

8. Predict Class A's score at 7.5 hours sleep. Is this interpolation or extrapolation?

Solution

Score $= 8(7.5) + 18 = 78$.

Interpolation (within 4-8 hrs)