Unit 5 · Patterns and Bivariate Data Cheat Sheet

Unit 5 focus: scatter plots, association and lines of best fit

1 KEY VOCABULARY

bivariate data	Pairs of values for two variables.	(hours, score)
scatter plot	A graph of the pairs as points.	see section 3
association	A pattern between the two variables.	more study, higher score
outlier	A point far from the pattern.	a 95 after 1 hour
cluster	A group of points close together.	scores in the 80s

2 MAKING A SCATTER PLOT

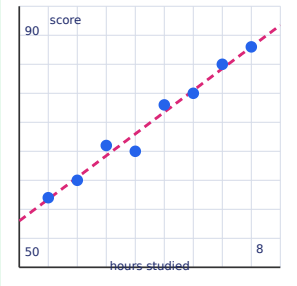
- Put the explanatory variable on the x-axis (hours) and the response on the y-axis (score).
- Choose equal steps on each axis; an axis may start above 0 (use a break ≀).
- Plot each pair as one point; do **not** connect the dots.
- Title the graph and label both axes with units.

3 LINE OF BEST FIT

Draw a line close to most points, with about as many above as below. Find its equation from two points on the line.

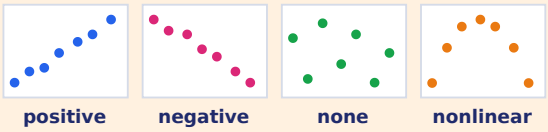
Example: The line passes through $(0, 58)$ and $(8, 88)$.
Slope: $\frac{88-58}{8-0} = 3.75$
Equation: $y = 3.75x + 58$

Answer: $y = 3.75x + 58$



4 TYPES OF ASSOCIATION

Positive: both increase. **Negative**: one increases as the other decreases. **No association**: no pattern. **Linear** points follow a line; **nonlinear** points follow a curve. **Strong** = points close to the pattern.



5 USING THE LINE TO PREDICT

Substitute into the equation. The slope is the change in y per unit of x; the intercept is the value at $x = 0$.

Example: Predict the score after 6.5 hours with $y = 3.75x + 58$.
Substitute: $3.75(6.5) + 58 = 24.375 + 58$

Answer: about 82 points (+3.75 points per hour)

6 QUICK REMINDERS & COMMON MISTAKES

✓ Do remember

- A line of best fit does not have to touch any point.
- Predictions far outside the data are unreliable.
- Association does not prove that one thing causes the other.

✘ Common mistakes

- Connecting the dots: draw one straight line
- Forcing the line through the first and last point: balance the points
- Calling a curve "linear": check the shape

Unit 5 · Patterns and Bivariate Data Cheat Sheet

More practice: two-way tables, relative frequencies and making predictions

1 TWO-WAY TABLES

	Plays sport	No sport	Total
Has a pet	18	12	30
No pet	14	6	20
Total	32	18	50

Rows and columns add to the totals. The bottom-right cell is the grand total.

2 FILLING IN A TABLE

Use the totals: each missing cell is a total minus the known cells.

Example: 50 students, 30 have a pet, 32 play a sport, 18 do both.
Pet, no sport: $30 - 18 = 12$
Sport, no pet: $32 - 18 = 14$
Neither: $50 - 18 - 12 - 14 = 6$

Answer: the table in section 1

3 RELATIVE FREQUENCY

Question	Divide by
What fraction of **all** students have a pet and play?	grand total: $\frac{18}{50} = 36\%$
Of pet owners, what fraction play?	row total: $\frac{18}{30} = 60\%$
Of non-pet owners, what fraction play?	row total: $\frac{14}{20} = 70\%$

4 IS THERE AN ASSOCIATION?

Compare row (or column) percentages. Very different percentages suggest an association.

Example: 60% of pet owners play a sport; 70% of non-pet owners play.
Compare: 60% vs. 70%: a 10-point difference

Answer: a weak association (the percents are close)

5 READING SLOPE AND INTERCEPT

$y = 3.75x + 58$, x = hours, y = score: the slope means...	**+3.75 points per hour**
...and the y-intercept means...	**about 58 with no study**
A line $y = -2x + 40$ for minutes of phone use vs. pages read: association?	**negative**
Data points far from the line?	**outliers**

6 CHECK YOURSELF

- Are the axes labeled with equal steps?
- Does my line have about the same number of points above and below it?
- Did I divide by the correct total (grand, row or column)?
- Did I describe direction, form and strength of the association?