Unit 12 · Statistics and Data Modeling Cheat Sheet

Unit 12 focus: center, spread, box plots and outliers

1 KEY VOCABULARY

mean	Add the values, divide by how many.	$2, 4, 9$: 5
median	The middle value of the ordered data.	$2, 4, 9$: 4
mode	The value that appears most often.	$1, 3, 3$: 3
range	Maximum minus minimum.	$9 - 2 = 7$
outlier	A value far away from the rest of the data.	$2, 3, 4, 40$

2 SPREAD: MAD

The mean absolute deviation is the average distance from the mean.

Example: Find the MAD of $2, 4, 6, 8$.
Mean: $20 \div 4 = 5$
Distances: $3, 1, 1, 3$
Average: $8 \div 4 = 2$

Answer: MAD $= 2$

3 QUARTILES AND IQR

Q1 is the median of the lower half, Q3 the median of the upper half. IQR = Q3 − Q1.

Example: $2, 4, 5, 7, 8, 10, 12, 15$
Halves: $2, 4, 5, 7$ and $8, 10, 12, 15$
Quartiles: $Q_1 = 4.5,\ Q_3 = 11$
IQR: $11 - 4.5 = 6.5$

Answer: IQR $= 6.5$

4 BOX PLOTS

- Five-number summary: **min, Q1, median, Q3, max**.
- The box goes from Q1 to Q3; the line inside is the median.
- The whiskers reach the min and max.
- Each part (whisker or half-box) holds about 25% of the data.
- A longer box means more spread in the middle half.

5 OUTLIERS

A value is an outlier if it is below Q1 − 1.5·IQR or above Q3 + 1.5·IQR.

Example: $Q_1 = 10,\ Q_3 = 18$. Is 31 an outlier?
IQR: $18 - 10 = 8$
Fences: $10 - 12 = -2$ and $18 + 12 = 30$

Answer: 31 > 30: yes

6 QUICK REMINDERS & COMMON MISTAKES

✓ Do remember

- Put the data in order before finding the median or quartiles.
- Outliers pull the mean, not the median.
- Use median and IQR for skewed data; mean and MAD for symmetric data.

✘ Common mistakes

- median of unordered data: order it first!
- two middle values: average them
- range: is max − min, not the biggest number

Unit 12 · Statistics and Data Modeling Cheat Sheet

More practice: two-way tables, scatter plots, lines of fit and residuals

1 TWO-WAY TABLES

	Dog	No dog	Total
8th	12	13	25
9th	6	19	25
Total	18	32	50

Joint relative frequency: one cell ÷ grand total, e.g. 12/50 = 24%.
Marginal: a row or column total ÷ grand total, e.g. 18/50 = 36%.
Conditional: one cell ÷ ITS row (or column) total, e.g. 6/25 = 24% of 9th graders have a dog.

2 CORRELATION

Type	Looks like	Example
positive	rises left to right	study time vs. score
negative	falls left to right	car age vs. value
none	no pattern	shoe size vs. birthday
r near ±1	strong (tight line)	$r = 0.95$
r near 0	weak or none	$r = 0.1$

Correlation does NOT mean one thing causes the other.

3 LINE OF FIT

Use the line to predict: substitute x and find y. The slope is the rate; the y-intercept is the starting value.

Example: $y = 2.5x + 10$. Predict y at $x = 8$.
Substitute: $2.5(8) + 10$
Simplify: $20 + 10 = 30$
Answer: $y = 30$

4 RESIDUALS

Residual = actual − predicted. Positive: the point is above the line. Negative: below the line.

Example: Model $y = 3x + 4$; at $x = 9$, actual $y = 34$.
Predicted: $3(9) + 4 = 31$
Residual: $34 - 31 = 3$
Answer: 3 **(above the line)**

5 SHAPE OF DATA

Shape	Tail	Mean vs. median
symmetric	none: both sides match	mean ≈ median
skewed right	long tail to the right	mean > median
skewed left	long tail to the left	mean < median

6 CHECK YOURSELF

- Did I order the data first?
- Did I divide by the right total in the two-way table?
- Residual: actual minus predicted (not the other way)?
- Does my prediction make sense in the context?