Unit 14 · Data & Regression Cheat Sheet

Unit 14 focus: summarize data and read scatterplots and correlation

1 DATA WORDS

Term	Meaning / quick example
Mean	Sum ÷ count.
Median	Middle value in order.
Outlier	A value far from the rest.
IQR	Q3 − Q1: spread of the middle half.

2 MEAN OR MEDIAN?

Outliers pull the mean; the median resists.

Example: Salaries \$42k, \$45k, \$48k, \$51k, \$350k.
Median: \$48,000
Mean: $536{,}000 \div 5 = 107{,}200$

Answer: The median (\$48,000) is more typical

3 IQR

Order the data; Q1 and Q3 are the medians of the lower and upper halves.

Example: 12, 15, 18, 20, 24, 27, 31, 40
Q1, Q3: $Q_1 = 16.5,\ Q_3 = 29$
IQR: $29 - 16.5 = 12.5$

Answer: 12.5

4 SCORE NEEDED

Target total = target mean × number of scores.

Example: 78, 85, 90; want a mean of 85 after 4.
Target total: $85(4) = 340$
Needed: $340 - 253 = 87$

Answer: 87

5 READING R

r	Describe as
−0.86	strong negative
0.52	moderate positive
0.12	weak positive

Strong |r| ≥ 0.7; moderate 0.4–0.7; weak below 0.4. Sign gives direction.

6 CORRELATION IS NOT CAUSATION

- Look for a lurking variable (heat → ice cream and sunburns).
- Ask whether the cause could run the other way.
- Only a randomized experiment shows cause and effect.

Unit 14 · Data & Regression Cheat Sheet

More practice: predict with lines of best fit, residuals and honest graphs

1 LINE OF BEST FIT

$\hat{y} = mx + b$ — the hat means "predicted."
Slope m: predicted change in y for each 1 unit of x.

$\hat{y} = -145x + 26{,}500$ (price vs. thousands of miles): each 1,000 miles → \$145 lower predicted price.

2 PREDICT AND SOLVE

Substitute x to predict; set ŷ to a value to solve for x.

Example: $\hat{y} = -145x + 26{,}500$
x = 60: $-145(60) + 26{,}500 = 17{,}800$
ŷ = \$20,700: $x = (20{,}700 - 26{,}500) \div (-145) = 40$

Answer: \$17,800; 40 thousand miles

3 RESIDUAL

$\text{residual} = \text{actual} - \text{predicted}$

Example: Predicted \$17,800; listed \$16,200.
Residual: $16{,}200 - 17{,}800 = -1{,}600$
Meaning: below the line

Answer: −\$1,600: priced below the model

4 TRUNCATED AXIS

Compare what the bars show with the real difference.

Example: Ratings 52 and 58; axis starts at 50.
Looks: $(58 - 50) \div (52 - 50) = 4$ times as tall
Actually: $(58 - 52) \div 52 \approx 11.5\%$ higher

Answer: A modest gap looks huge

5 TRUST A MODEL?

- Is x inside the range of the data? (no extrapolation)
- Do residuals look random, not curved?
- Is the sample large and fair?
- Do predictions make sense (no scores above 100)?

6 REMINDERS & MISTAKES

✓ Do remember

- Sort before finding a median.
- Residual = actual − predicted.

✗ Common mistakes

- Strong r means causation: look for a lurking variable.
- Negative r is weak: the sign is direction, |r| is strength.
- Predict far outside the data: extrapolation is risky.
- Bars may start anywhere: bar graphs should start at 0.