Statistics, the memo sheet.

D8A.ACADEMY/RESOURCES/STATS-FORMULAS

CENTER

mean = Σx / n	pulled by outliers
median	the middle, robust
mode	most frequent

SPREAD

std σ	avg distance to mean
IQR = Q3 - Q1	the middle 50%
range = max - min	fragile, one outlier

OUTLIERS

z = (x - μ) / σ	|z| > 3 = suspicious
x < Q1 - 1.5·IQR	boxplot fence
log(x)	tames right skew

RELATIONSHIPS

r ∈ [-1, 1]	direction + strength
r^2	variance explained
r = 0.9	← still not causation.

TESTING

H_0	assume no effect
p-value	P(data this odd | H_0)
α = 0.05	false-alarm budget
95% CI	the plausible range

RULES OF THUMB

n ≥ 30	CLT starts helping
±1σ ≈ 68% · ±2σ ≈ 95%	normal quick math
median vs mean far apart	skewed data

WORTH MEMORIZING

01 The whole sheet, in pandas

Every number on this page is one line of code away.

```
x = df["revenue"]
x.mean(), x.median(), x.std()
q1, q3 = x.quantile([.25, .75])
fence = q1 - 1.5 * (q3 - q1)
z = (x - x.mean()) / x.std()
outliers = x[z.abs() > 3]
df[["ads", "revenue"]].corr()
```

→ Compute all of them before trusting any chart. Two minutes, every dataset.

02 Trap: what a p-value is not

The most misquoted number in data. Three readings, two of them wrong.

P = 0.03 MEANS	VERDICT
3% chance of data this extreme if there is truly no effect	correct
97% chance our hypothesis is right	wrong
the effect is big enough to matter	wrong: that is effect size

→ And checking the test every day until it dips under 0.05 is called peeking. It manufactures significance.