

Statistics recipe sheet

Four tests of one association, and what each one answers.

Test	What it answers	How to run it
Correlation and regression	How strong and in what direction	Pearson correlation and ordinary least squares on the primary dataset
Bootstrap confidence interval	How precisely the slope is known	Resample the pairs with replacement many times and take the interval of the slopes
Randomisation null	How often chance produces this	Shuffle the pairing ten thousand times and count how often the value is reached
Leave one out	Whether one point or target drives it	Remove each target in turn and recompute the correlation

Reference results from the worked study

Analysis	Result
Pocket volume against molecular weight, thirty five compounds	r about 0.876, r squared about 0.768
Bootstrap slope	about 0.217 Daltons per cubic Angstrom, interval about 0.166 to 0.264
Lipophilicity	r about 0.689, r squared about 0.475
Per drug structures, eleven compounds	r about 0.914, r squared about 0.834
Independent pocket tool	r about 0.562, r squared about 0.316
Randomisation of ten thousand iterations	no iteration reached the observed value
Leave one out	r stayed between about 0.862 and 0.889

Reporting rules

- Give the effect with its precision rather than a probability value alone.
- Show the scatter plot with every point visible and coloured by group.
- Put the weakest analysis in the main text beside the strongest.
- State what the association does not license, namely causation, precise prediction and transfer to other systems.

Based on the ERAS Africa Molecular Docking using Python course, 2026, and the study "Target Protein Pocket Volume and Antiretroviral Drug Physicochemical Properties" (Hailekiros, 2026) with its reproduction guide. Tool documentation from RDKit, AutoDock Vina, Meeko, RCSB PDB, PubChem and ProteinsPlus. Free to copy and adapt with attribution.