

# Student Homework Sheet — Stage 10a: Modeling: Linear Regression

---

## Assignment (what to deliver)

Create a notebook `notebooks/modeling_regression_<team>.ipynb` that:

1. Fits a linear regression model on your **project dataset** (or use the starter synthetic data to prototype).
2. Plots and interprets **residuals** to assess assumptions: linearity, independence, homoscedasticity, normality.
3. Reports **R<sup>2</sup> and RMSE** and explains whether you trust the model.
4. (Optional stretch) Add a transformed feature (e.g.,  $x^2$  or interaction) and explain why this is still linear regression.

**Chain statement:** *In the lecture, we learned to fit a baseline model and diagnose assumptions with residuals. Now, you will adapt that workflow to your own dataset to evaluate trust and usefulness.*

## Step-by-Step Instructions

1. Load your cleaned dataset (or generate synthetic data from the starter).
2. Split into train/test (e.g., 80/20).
3. Fit `sklearn.linear_model.LinearRegression`.
4. Compute predictions, residuals,  $R^2$ , and RMSE.
5. Plot residuals vs fitted, histogram, QQ; optionally residuals vs a key predictor and residual lag-1.
6. Write markdown interpreting each assumption.
7. (Optional) Add a simple transformation (e.g., square of a predictor); refit and compare diagnostics.
8. Conclude: state whether you trust this model and why.

## Grading Rubric (100 pts)

- (20) **Correct model fit:** data split, fit, predict, metrics.
- (30) **Diagnostics & plots:** residual vs fitted, histogram, QQ (and at least one more check).
- (30) **Interpretation quality:** clear commentary on assumptions; explanation vs prediction distinction.
- (10) **Metrics discussion:**  $R^2$  and RMSE connected to usefulness.
- (10) **Reproducibility & clarity:** clean code, organized markdown, seed set or documented randomness.

**Stretch credit (+1):** sensible transformation (e.g., polynomial term) with improved diagnostics and correct explanation.

## Example Submission Expectations

- Notebook contains: code cells + plots + markdown interpretation.
- Figures are readable (axes labeled, titles).
- Conclusion paragraph summarizes trust level and next steps.

## Submission

- File: `notebooks/modeling_regression_<team>.ipynb`
- Due: next class session
- Commit and push to your project repo.

## Collaboration

- Discuss concepts with classmates, but write your own code and interpretations. Cite any borrowed snippets.