Regression Rush: Linear Regression in R (Beginner → Intermediate)

Univariate • Bivariate • Multiple Regression in 4 Weeks

- **Duration:** 4 weeks (Beginner → Intermediate)
- **Tools:** R, `lm()`, tidyverse, ggplot2 (plus optional diagnostic packages)
- **Core Promise:** Learn to *fit, interpret, diagnose, and report* linear regression models responsibly.
- **Prerequisites:** 10+2 algebra; no advanced math required; R experience not required.

1. Course Description

This course is a hands-on introduction to linear regression using R, designed for learners moving from absolute basics to a practical intermediate workflow. We begin with univariate regression (one predictor), expand to bivariate and multiple regression, and learn model diagnostics, interpretation, and communication using real datasets.

2. Motivation

Linear regression is foundational in statistics, data science, and machine learning. Beyond fitting a line, this course emphasizes: (i) what model coefficients mean, (ii) how to quantify uncertainty, (iii) how to test whether a model is trustworthy using diagnostics, and (iv) how to present results responsibly.

3. Learning Outcomes

By the end of the course, students will be able to:

1. Fit univariate, bivariate, and multiple linear regression models in R using `lm()`.
2. Interpret coefficients (including "holding other variables constant" in multiple regression).
3. Compute and explain confidence intervals and prediction intervals.
4. Diagnose model assumptions using residual plots and influence measures.
5. Handle categorical predictors with factors (dummy coding intuition).
6. Compare models using adjusted R^2 and AIC (intro) and justify a final model.
7. Produce a clean, reproducible regression report (plots + narrative + code).

4. Prerequisites

- Basic algebra and comfort reading graphs (10+2 level).
- Basic computer literacy (files, folders, installing software or using Colab).
- No prior statistics or R programming required.

5. Teaching & Learning Pedagogy

Each week follows a cycle:

Concept → Visual intuition → R implementation → Interpretation drills → Mini-lab → Weekly deliverable

Emphasis is placed on interpretation and diagnostics (the skills that make regression useful in practice).

6. Software & Course Resources

- **Required:** R (latest stable), RStudio *or* Google Colab (R runtime).
- **Core packages:** `tidyverse`, `ggplot2`, `broom`
- **Optional diagnostics:** `car` (VIF), `performance` (checks)
- **Course materials:** slides/notes, coding notebooks, weekly practice sets.

7. Assessment Plan

Component	Weight
Weekly Quizzes (concept checks)	20%
Weekly Deliverables (notebooks/reports)	40%
Participation / Practice Labs	10%
Final Capstone Project	30%

Submission Format

Unless specified otherwise, submissions should be:

- a PDF export of the notebook/report, and
- the source file (`.R`, `.Rmd`, or Colab notebook link/export).

Feedback & Rubric (Suggested)

Weekly deliverables are graded on:

- **Correctness** (model fitting, computations),
- **Interpretation** (clear meaning of coefficients/intervals),
- **Diagnostics** (appropriate plots + conclusions),
- **Communication** (clean narrative, plots labeled, reproducible code).

8. Weekly Schedule (Modules + Deliverables)

Topics

- R basics, data frames, importing data
- Scatterplots; correlation intuition
- Univariate regression: `lm(y ~x)`
- Residual intuition; basic diagnostic plots

Weekly Deliverable 1 (Report): EDA + fitted line + slope/intercept interpretation + simple predictions.

Topics

- Bivariate regression: `lm(y ~x1 + x2)`
- Confidence vs prediction intervals; p-values responsibly
- Residual plots; communicating uncertainty

Weekly Deliverable 2 (Inference Notebook): Coefficients + CI + PI + residual plot interpretations.

Topics

- Multiple regression; "holding others constant"
- Factors/dummy coding intuition; categorical predictors
- Multicollinearity; VIF (intro)
- Model comparison: adjusted R^2 + AIC (intro)

Weekly Deliverable 3 (Model Comparison Brief): Build 2–3 models, compare, justify final choice, interpret factors.

Topics

- Assumptions and diagnostics: linearity, variance, normality (for inference)
- Outliers, leverage, influence; Cook's distance
- Transformations, interactions, polynomial terms (intro)
- Reporting: narrative + plots + limitations

Weekly Deliverable 4 (Refinement Notebook): Diagnose + improve one model + final summary + limitations.

9. Capstone Project

Goal: Choose a dataset, pose a question, build a regression model, diagnose assumptions, refine the model, and communicate results clearly.

Required Sections (Checklist)

1. Problem statement + target variable
2. EDA (minimum 3 meaningful plots)
3. Baseline model + interpretation
4. Diagnostics + issues found
5. Improved model + comparison
6. Predictions with intervals
7. Limitations + next steps

Recommended Capstone Dataset Idea

Ames Housing (House Prices): rich numeric + categorical predictors, realistic challenges, meaningful prediction target (`SalePrice` or `log(SalePrice)`). Suggested question: *"Which features explain house prices, and how well can we predict price with a trustworthy linear model?"*

10. Course Policies (Editable)

Attendance & Participation

- Participation is encouraged; labs are skill-building and count toward the participation grade.
- If you miss a session, review the notebook and complete the practice tasks.

11. Books & References

- James, Witten, Hastie, Tibshirani — *An Introduction to Statistical Learning (ISLR).*
- Wickham & Grolemund — *R for Data Science.*

- Fox & Weisberg — *An R Companion to Applied Regression.*
- Faraway — *Linear Models with R.*
- Weisberg — *Applied Linear Regression.*

End of Syllabus.