ECON 101 | FALL 2025

From a Scatterplot to a Regression Line

Selected RStudio/Quarto exercise | Regression coefficients, units, and interpretation

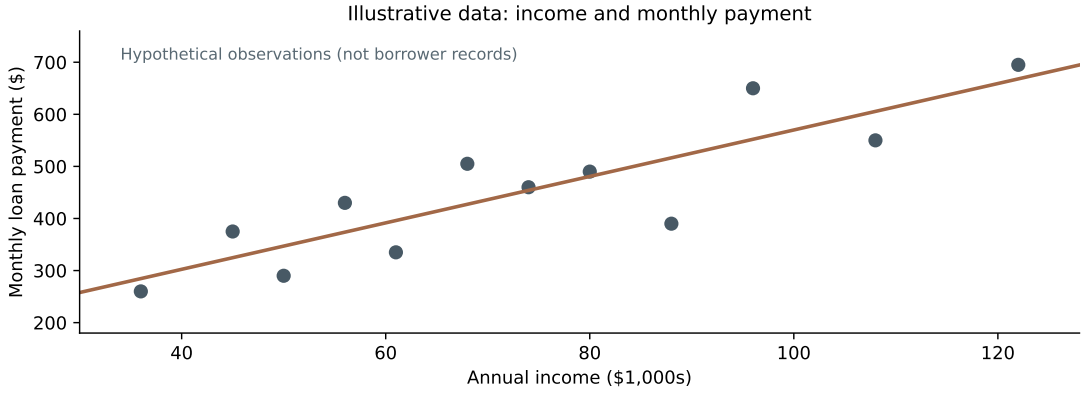
A fitted regression line summarizes an association between two variables. The goal of this exercise is not just to run `lm()`, but to interpret what a slope measures and what changes when we change the units of a variable.

The observations below are hypothetical loan examples created for this self-contained teaching sample, not actual borrowers.

1. Estimate a simple regression

Let y be a monthly loan payment in dollars and let x be annual income in thousands of dollars. Enter the data and fit a line in RStudio.

```
loans <- data.frame(
  income_k = c(36,45,50,56,61,68,74,80,88,96,108,122),
  payment = c(260,375,290,430,335,505,460,490,390,650,550,695)
)
fit_k <- lm(payment ~ income_k, data = loans)
round(coef(fit_k), 3)
```



The fitted line is approximately

$$\widehat{\text{payment}} = 124.03 + 4.46\,\text{income_k}, \qquad R^2 \approx 0.755.$$

Interpretation. In these illustrative data, $1,000 higher annual income is associated with about $4.46 higher monthly payment, on average in the fitted line. This is a descriptive association, not evidence of a causal effect.

2. Same model, different units

Suppose we measure income in dollars rather than thousands of dollars. The observations do not change; only the scale of x changes.

```
loans$income_dollars <- loans$income_k * 1000
fit_d <- lm(payment ~ income_dollars, data = loans)
round(coef(fit_d), 6)
```

The slope with income in dollars is approximately 0.004459. The slope with income in thousands is approximately 4.458888:

$$\hat{\beta}_{1,\text{dollars}} = \frac{\hat{\beta}_{1,\text{thousands}}}{1000}.$$

These coefficients describe the same relationship. One unit of income in thousands is $\$1{,}000$; one unit of income in dollars is $\$1$. The fitted values, residuals, and R^2 remain unchanged.

```
# Compare fitted values from both specifications
all.equal(as.numeric(fitted(fit_k)),
          as.numeric(fitted(fit_d)))
```

Question. If the slope is larger when income is measured in thousands, does this mean income has a stronger association with monthly payment?
Answer. No. The numerical slope depends on the units of measurement. The fit and the underlying association are unchanged.

3. Make the plot in R

```
plot(loans$income_k, loans$payment,
     xlab = "Annual income ($1,000s)",
     ylab = "Monthly loan payment ($)", pch = 19)
abline(fit_k, col = "steelblue", lwd = 2)
```

A plot helps check whether a straight line is a useful summary, how much the points scatter around it, and whether unusual observations merit closer attention.

4. Questions for students

1. **Units**. Interpret the slope when income is measured in dollars. Compare your interpretation with the slope when income is measured in thousands.
2. **Prediction**. Use the fitted equation to predict the monthly payment for an observation with annual income of $\$75{,}000$. Would the prediction change if you used `fit_d` instead?
3. **Model fit**. What does $R^2 \approx 0.755$ tell you about the fitted values? What does it *not* tell you?
4. **Causality**. Could loan amount, interest rate, or the borrower's credit history affect monthly payments? Why does their omission matter when interpreting the income coefficient?

Teaching role: ECON 101, Statistics for Economics, Fall 2025 (Teaching Assistant). This portfolio example is independently prepared using topics from the course labs; it is not a reproduction of the course instructor's original assignments.