



Homework 2

Simple Linear Regression

Due: Saturday, October 17, 2026

Required work: 90–120 minutes

Total: 100 points

Required: 20 / 30 / 50 points

Read a graph, explore a regression in Colab, then learn and apply the formula by hand. Question 4 is optional extra credit (score multiplier 1.05). Short presentation examples for Questions 1 and 2 are in [HW2_Examples.pdf](#). No confidence intervals, hypothesis tests, or multiple regression are required.

Question 1: Explain the provided graph — 20 points

The graph below shows **simulated teaching data** for delivery distance and delivery time. The dots are eight observed deliveries; the straight line is their least-squares fitted line. Use the graph and small regression table to explain the relationship in your own words. You do not need Colab or a new analysis for this question.



Term	Coefficient
Intercept	9.857
Distance (km)	2.143

- What does the graph suggest about distance and delivery time? Explain the slope using its units.
- What is the difference between a dot and the line? Choose a dot above or below the line and explain what that position means.
- What could a delivery planner use this line for, and what should they avoid concluding from it? Explain one limitation visible in the graph or implied by these data.

Question 2: Upload, prompt, and fit a line in Colab — 30 points

Find a CSV dataset that interests you, or download **one** of the original files below. Choose one numerical X and one numerical Y. Analyze only one dataset for this question.

Original CSV	A row represents	Suggested X	Suggested Y
<code>tips.csv</code>	A restaurant bill	<code>total_bill</code> (dollars)	<code>tip</code> (dollars)
<code>mpg.csv</code>	A car model	<code>weight</code> (pounds)	<code>mpg</code> (miles per gallon)
<code>penguins.csv</code>	A penguin	<code>flipper_length_mm</code> (mm)	<code>body_mass_g</code> (grams)

The CSV files listed above are ready to download from the homework page. Choose one, save it on your computer, and upload it to Colab. If you find your own data, include its source link.

Create a **blank Google Colab notebook** and upload the raw CSV yourself. Write your own prompts to Colab's AI assistant to read the file, fit a simple linear regression with an intercept, and display a regression table and a scatterplot with the fitted straight line. Request the intercept, slope, observation count, and R-squared. Ask for clearly labeled axes with units and no confidence band.

Include screenshots of the uploaded filename, your actual prompts, regression table, and graph. Explain in your own words:

- What question are you asking, what does one row mean, and why did you choose X and Y?
- Why did you write your prompts that way? What did the AI do to turn your uploaded data into the table and line? Explain any follow-up you needed.
- How do the slope, intercept, and line relate to the fitted-line idea in Question 1 and the formula you will investigate in Question 3? Explain what the slope and R-squared mean for your data.
- Choose one specific statement in the AI's explanation. Do you agree, partly agree, or disagree? Refer to your table or graph to explain why. A correct statement is acceptable if you explain why it is correct.

Colab starting steps

1. Download and, if necessary, unzip your CSV. Open [Google Colab](#), sign in, and create a new notebook.
2. Open the Gemini / AI chat using its spark icon. Attach the CSV through the chat's upload control and check the filename. Alternatively, upload through the Files panel and tell the AI the filename and runtime path shown there.
3. Write your own prompts. First check that the AI can read the correct file and show its columns; then ask for the analysis you need.
4. Review and run its proposed analysis or generated code cells to obtain actual output. Check the selected variables and labels, and follow up if something is wrong. Take screenshots of the results.

No ready-made regression notebook is supplied. If AI chat is unavailable, contact the instructor for help. No paid upgrade is required.

Question 3: How can I calculate a best-fit line myself? — 50 points

Open learning question: If I want a formula for the best-fit straight line, rather than asking a computer to draw it for me, how can I calculate that line? How does the formula work, and why does it make sense?

Learn through your own prompts

Use an AI assistant of your choice to learn the formula and method for a simple least-squares best-fit line with an intercept. Start with your own question and continue the conversation until you understand how to use the formula. Ask follow-up questions about anything you do not understand. There is no prescribed wording or required number of prompts.

Show the conversation that explains how you learned: include screenshots or a transcript of your actual prompts and the relevant AI replies. Explain **why you asked those questions**, what you learned from the replies, and how a follow-up helped resolve confusion or check your understanding. A pasted conversation by itself is not an answer. Write the formula you learned and explain its symbols and steps in your own words. You do not need to derive it using calculus.

Apply your understanding by hand

Use these **five simulated teaching observations**. X is the number of workers on a shift; Y is the number of orders processed per hour.

Observation	X: workers	Y: orders per hour
A	1	4
B	2	4
C	3	7
D	4	8
E	5	12

Handwrite your calculation of the slope, intercept, and fitted equation. Show your intermediate arithmetic so someone can follow how you used the formula. Explain how your calculation corresponds to the method you learned in the AI conversation.

On paper, draw and label the axes, plot all five observations, and **use a ruler to draw your calculated fitted straight line**. Show how you calculated two points on the line to position it. Do not connect the five observed dots to one another. A basic calculator may be used for arithmetic, but do not use AI or a regression program to calculate these five observations or produce this graph for you.

Explain and evaluate the method

In your own words, explain what you think “best fit” means and what the formula is trying to achieve. Drawing on the class discussion, why is this a relatively reasonable way to choose a line? What problem does it solve that a less careful way of choosing or judging a line might create? Do you think this method is always reasonable? Explain at least one limitation or situation where you would question it, and give your reason. Also explain what your hand-drawn line says about these workers and orders.

What to submit

Submit a **Word document or clear scans/photos of handwritten answers**, with your AI learning conversation, actual prompts, screenshots, and your own explanations. **Question 3’s calculations and ruler-drawn graph must be handwritten**, even if the rest is typed. Include answers to Questions 1–3; add Question 4 if you choose the extra credit. Copied AI explanations or screenshots alone are incomplete. No code or notebook submission is required.

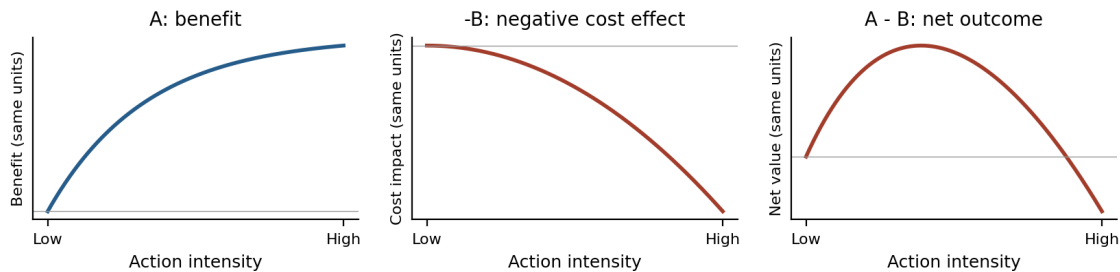
Colab guidance: [official Colab FAQ](#).

Question 4: One action, two sides — optional extra credit

Reward: a satisfactory answer multiplies your score on Questions 1–3 by **1.05** (80 becomes 84). Allow about 10–15 extra minutes. No calculations, code, or detailed research are needed.

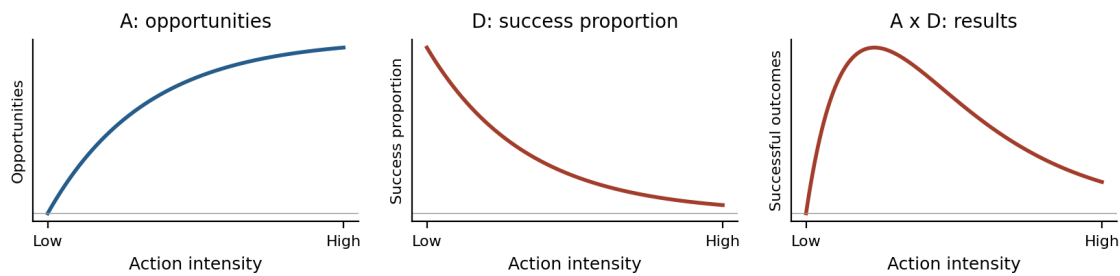
Sometimes doing more of something helps at first but becomes harmful when we do too much. The diagrams show two simple ways this can happen.

A – B: benefit minus cost



Example: more training can help workers perform better, but training also costs money and takes time away from work. A is the benefit, B is the cost, and $A - B$ is what remains. The middle graph shows $-B$, the negative effect of the cost: a downward curve means the cost is becoming more harmful. Subtraction makes sense when benefit and cost use the same units, such as dollars.

$A \times D$: opportunities times a success proportion



Example: more advertising may reach more customers (A), but the share who buy (D) may fall as we reach less interested people. Customers reached $\times$ the share who buy gives the number of buyers. That is why we multiply these two quantities.

Your answer

Give **one management example for $A - B$** and **one for $A \times D$** . For each, briefly explain:

- What are we doing more of? What are its two sides?
- What do the two quantities mean, and why should we subtract or multiply them?
- Why might the result improve at first but worsen when we do too much?

Finish with one or two sentences about why this helps a manager think beyond “more is always better.” Use your own words. You do not need to prove the curve or discuss advanced statistics.

Keep in mind: two opposite effects do not always give a visible inverted U. One may be much stronger than the other, so the result keeps rising or falling. It may look nearly flat, or the peak may be outside the range we see. This is why we check actual data instead of relying only on a story. These diagrams are illustrations, not observed data.