Section 12.2 — Linear Regression

Chris Godbout

Outline

Introduction

Definition

Definition (Least-Squares Regression Line)

The least-squares regression line is the line for which the best average variation from the data is the smallest. It is sometimes called the line of best fit. It is given by

$$\hat{y} = b_0 + b_1 x$$

Slope and Intercept

Slope

The slope of the least-squares regression line is

$$b_1 = \frac{s_y}{s_x} r$$

Slope and Intercept

Slope

The slope of the least-squares regression line is

$$b_1 = \frac{s_y}{s_x} r$$

y–intercept

The y–intercept of the least-squares regression line is

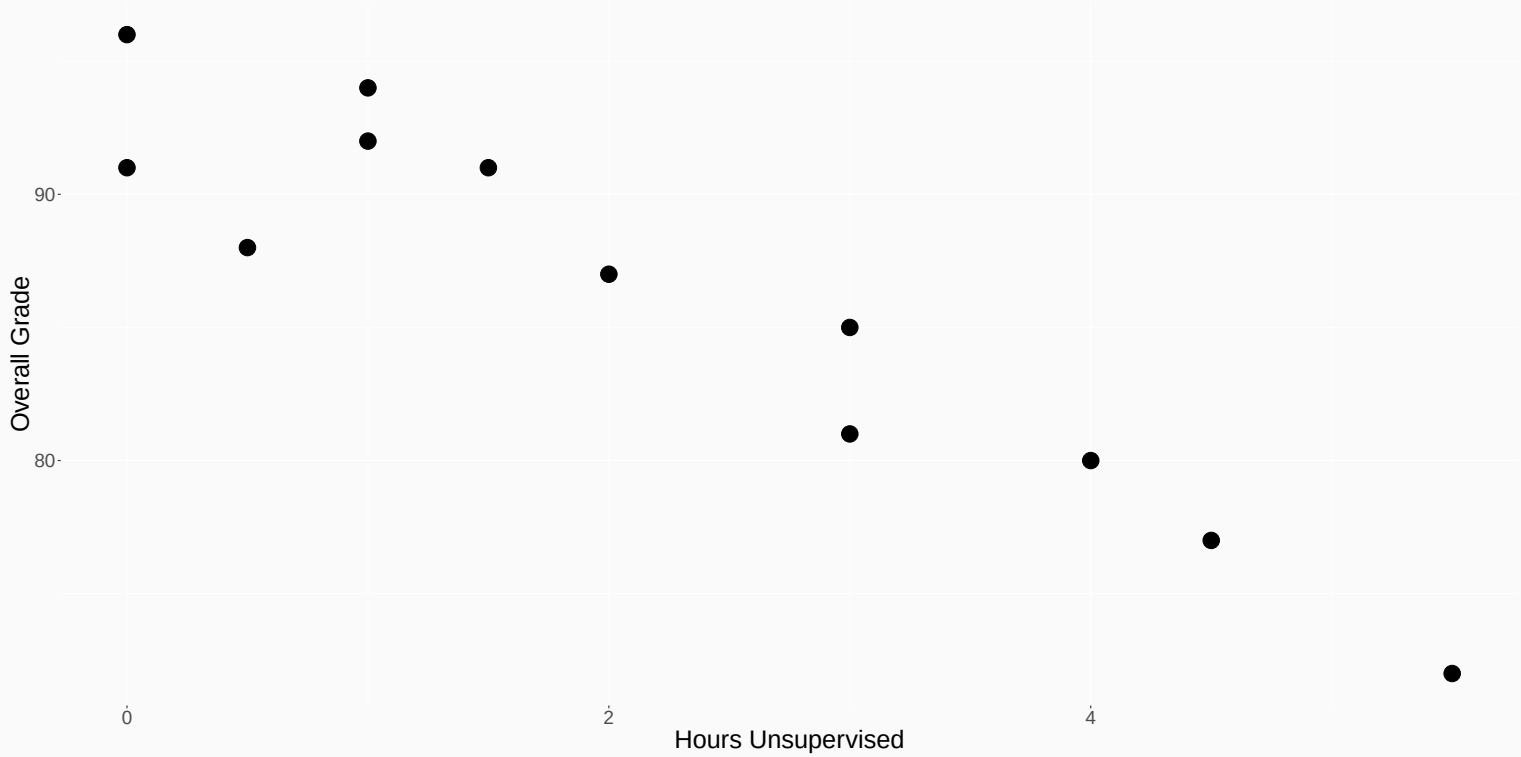
$$b_0 = \bar{y} - b_1 \bar{x}$$

Examples

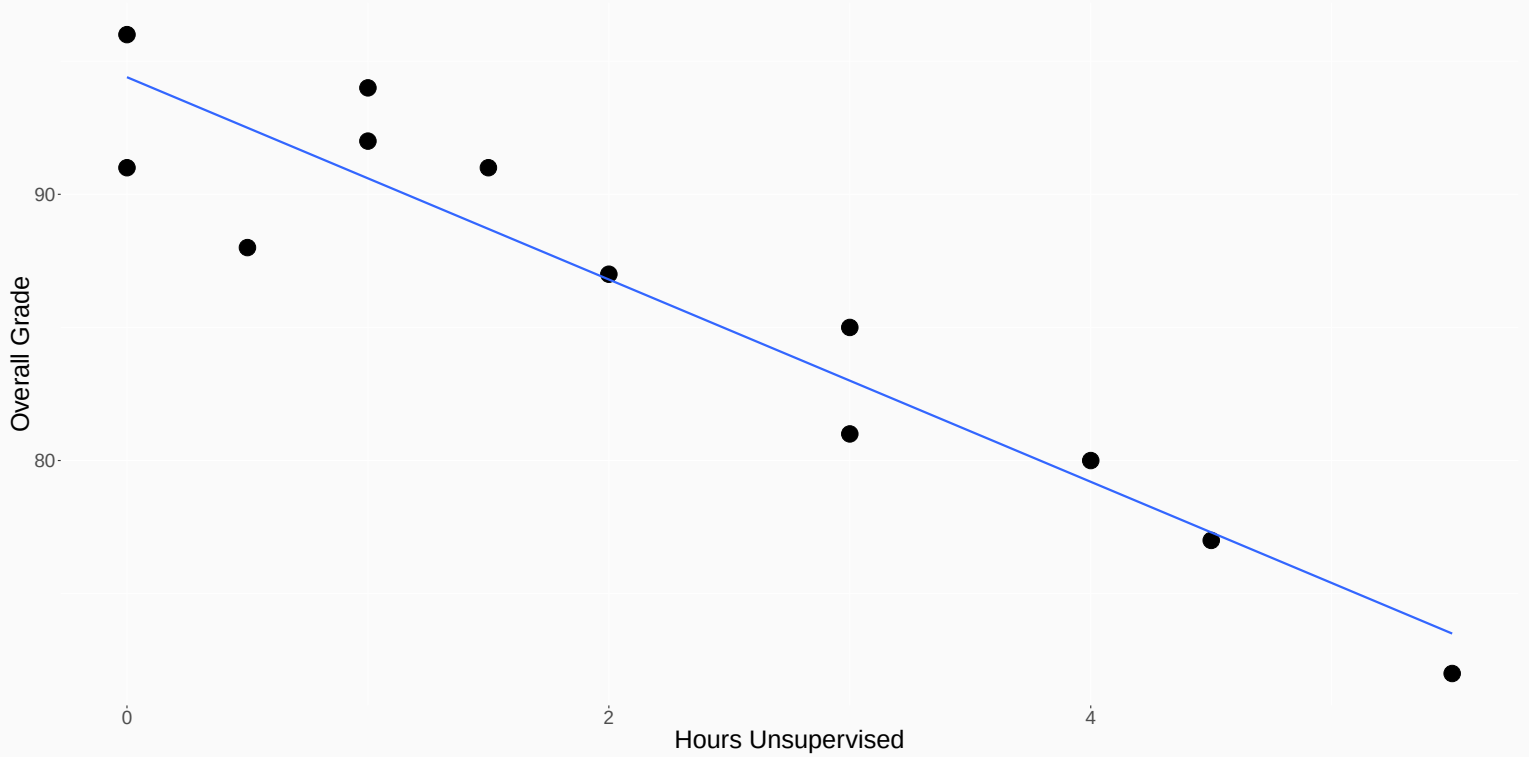
Grades

Hours Unsupervised	0	0	0.5	1.0	1.0	1.5
Overall Grade Average	96	91	88	92	94	91
Hours Unsupervised	2.0	3.0	3.0	4.0	4.5	5.6
Overall Grade Average	87	85	81	80	77	72

Grades



Grades



Grades

Grades

- What grade would you predict for a child left unsupervised 2.5 hours per day?

Grades

- What grade would you predict for a child left unsupervised 2.5 hours per day?
- What about 8 hours?

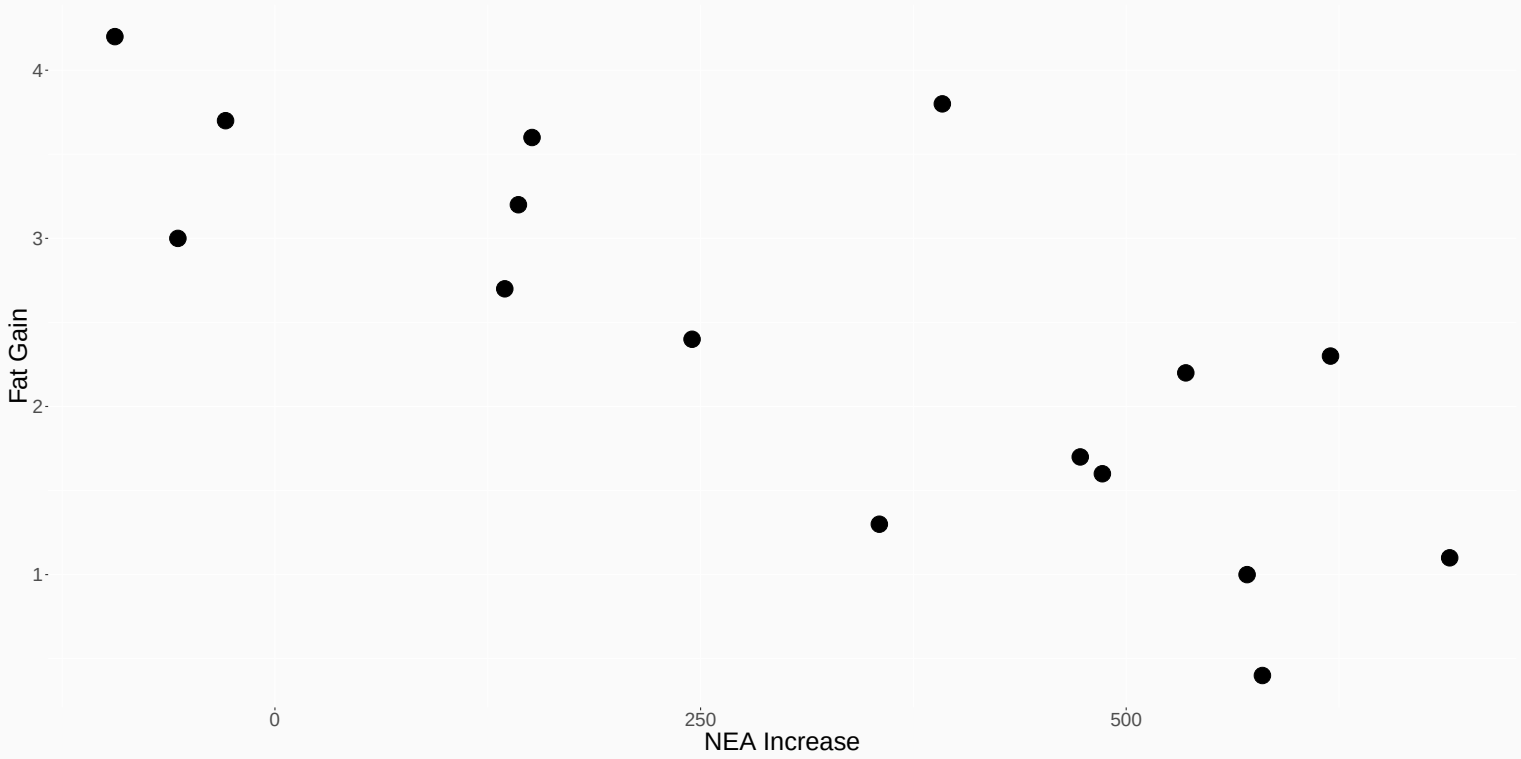
Grades

- What grade would you predict for a child left unsupervised 2.5 hours per day?
- What about 8 hours?
- Which do you think is more accurate?

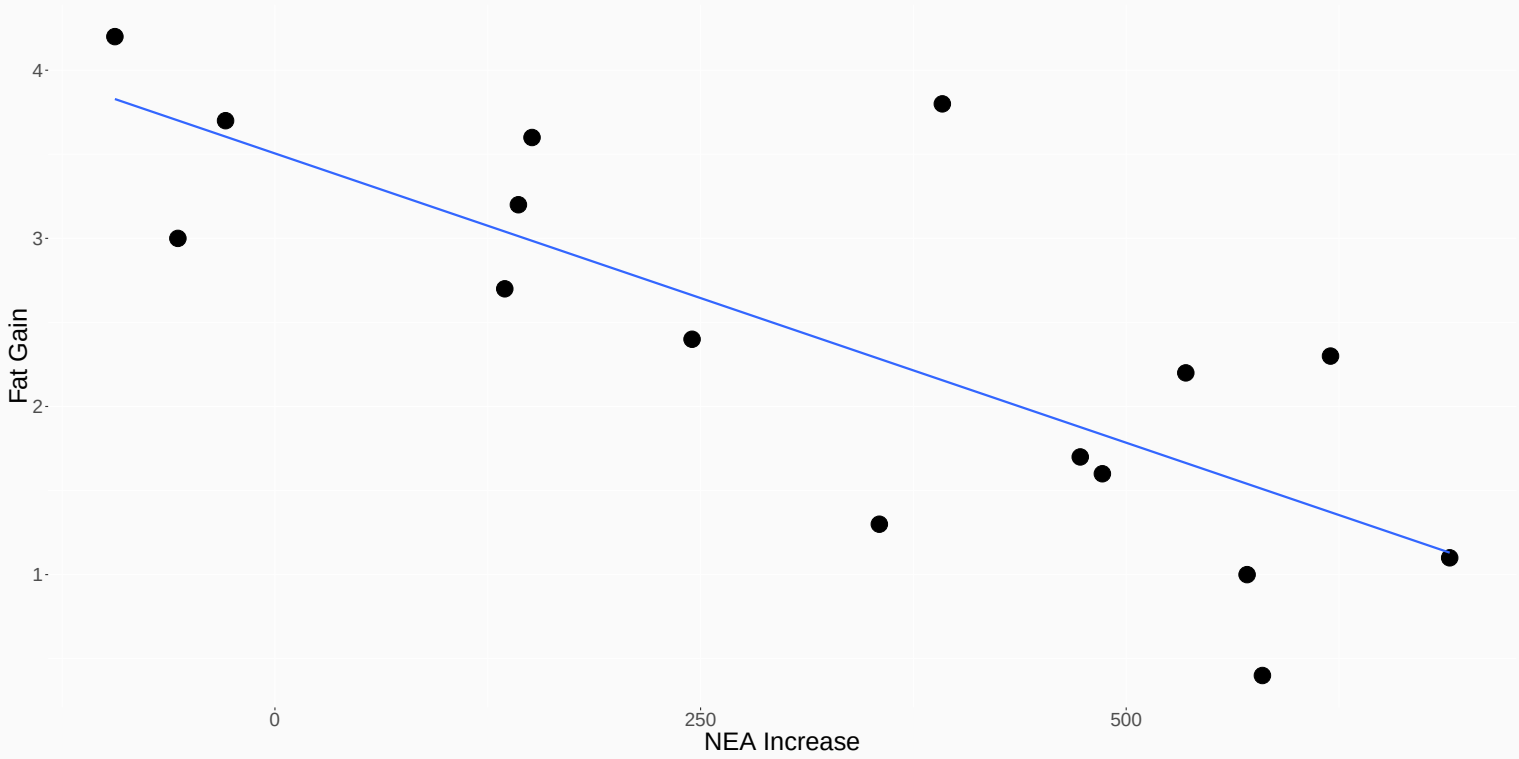
Fidgeting

NEA increase	Fat gain	NEA increase	Fat gain
-94	4.2	392	3.8
-57	3.0	473	1.7
-29	3.7	486	1.6
135	2.7	535	2.2
143	3.2	571	1.0
151	3.6	580	0.4
245	2.4	620	2.3
355	1.3	690	1.1

Fidgeting



Fidgeting



Fidgeting

Fidgeting

- How much fat gain would you expect if the non-exercise-activity increased by 50 calories?

Fidgeting

- How much fat gain would you expect if the non-exercise-activity increased by 50 calories?
- What about 1500?

Fidgeting

- How much fat gain would you expect if the non-exercise-activity increased by 50 calories?
- What about 1500?
- Which do you think is more accurate?

Warning!

Warning!

A prediction should not be made with a regression model if

Warning!

A prediction should not be made with a regression model if

- The data do not fall in a linear pattern when graphed as a scatter plot.

Warning!

A prediction should not be made with a regression model if

- The data do not fall in a linear pattern when graphed as a scatter plot.
- The correlation coefficient is not statistically significant.

Warning!

A prediction should not be made with a regression model if

- The data do not fall in a linear pattern when graphed as a scatter plot.
- The correlation coefficient is not statistically significant.
- You wish to make a prediction about a value outside the range of the sample data.

Warning!

A prediction should not be made with a regression model if

- The data do not fall in a linear pattern when graphed as a scatter plot.
- The correlation coefficient is not statistically significant.
- You wish to make a prediction about a value outside the range of the sample data.
- The population is different than that from which the sample data were drawn.