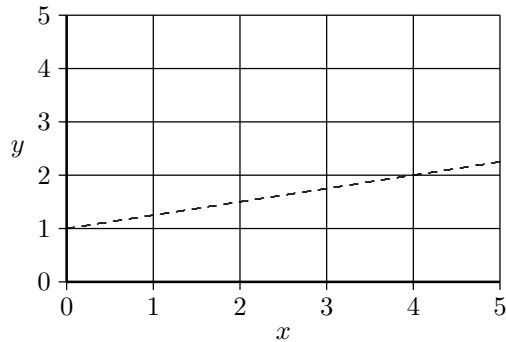


REVIEW OF STRAIGHT LINES

- For the dashed line below:

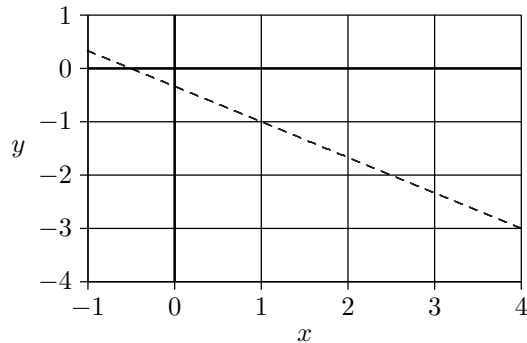


- slope = rate at which y increases, per unit increase in x
 $= \frac{\text{rise}}{\text{run}} = \underline{\hspace{2cm}}$.

- y -intercept = $\underline{\hspace{2cm}}$.

- Equation of the line is: $y = \underline{\hspace{2cm}} + \underline{\hspace{2cm}} \times x$.

- For the dashed line below:



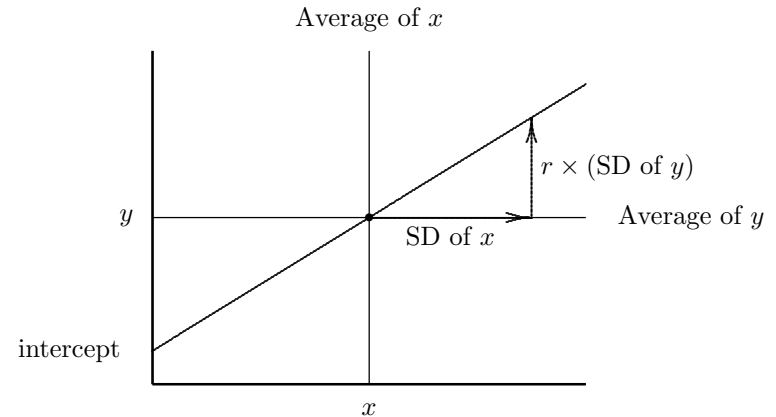
- slope = $\underline{\hspace{2cm}}$.

- y -intercept = $\underline{\hspace{2cm}}$.

- Equation of the line is: $y = \underline{\hspace{2cm}} + \underline{\hspace{2cm}} \times x$.

THE SLOPE, INTERCEPT, AND EQUATION OF A REGRESSION LINE

- For the regression line of y on x below:



- slope = $\underline{\hspace{2cm}}$.

- equation: $y = \underline{\hspace{2cm}} + \underline{\hspace{2cm}} \times x$.

- intercept = $\underline{\hspace{2cm}} - \underline{\hspace{2cm}} \times \underline{\hspace{2cm}}$.

- What is the equation useful for?

- Making regression predictions for several x s, by substituting each of them in turn into the equation.

- What is the statistical interpretation of the slope?

- The slope is the *average change in y , per unit change in x* .

- What is the statistical interpretation of the intercept?

- The intercept is the regression estimate for y , when x is 0.

- In 1988, the relationship between income and education for white men age 35–54, with full time jobs, is summarized as follows (figures slightly rounded):

average education \approx 13 years, SD \approx 3 years
 average income \approx \$34,000, SD \approx \$21,000, $r \approx 3/7$

- What is the regression equation for income on education?
 - slope:
 - intercept:
 - equation:
- On average, an extra year of education is worth _____ .
 - On average, how much more do people with 13 years of education earn than people with 12 years of education?
 - If you took a group of people with just a high school education and gave them all an extra year of education, would their average income go up \$3,000? Explain.
- For a man with no education, the regression line predicts an income of _____ .
- Predict the income of a man with 16 years of education:

- For the men aged 24-34 in the HANES sample, the regression equation for predicting height from education is

$$\text{predicted height} = (0.25'' \text{ per year}) \times (\text{education}) + 66.75''.$$

- Predict the height of a man with:
 - 12 years of education;
 - 16 years of education.
- Does going to college increase a man's height? Explain.
- With an observational study, the slope and intercept of the regression line are only descriptive statistics. They say how the average value of one variable is related to values of another variable, in the population being observed. The slope cannot be relied on to predict how y would respond if the investigator changed the value of x .

THE METHOD OF LEAST SQUARES

What is the *method of least squares* for fitting a straight line to the points in a scatter plot?

- For each point, use the line under consideration to predict y from x , and compute the prediction error = actual y – predicted y .
- Measure the lack of fit of the line to the data by the root-mean-square of the prediction errors.
- Move the line around so as to minimize the r.m.s. error. The best fitting line is called the *least squares line*.
- What is the relationship between the least squares line and the regression line?
 - They are one and the same.
- When does one use the term “least squares line”, and when “regression line”?
 - Investigators tend to talk about “least squares” when the main goal is to estimate the parameters of the line — the slope and/or intercept. Investigators tend to talk about regression when they are interested in how one variable varies with respect to another.

- Suppose you wanted to predict some response variable y from two explanatory variables x_1 and x_2 , using a so-called *multiple regression equation* of the form

$$y = a + b \times x_1 + c \times x_2.$$

What criterion might you use to choose the coefficients a , b , and c ?

- Least squares: find the a , b , and c that minimize the root-mean-square of the prediction errors.
- What is the interpretation of the coefficient b in the multiple regression equation above?
 - b is the average change in y per unit change in x_1 , when x_2 is held fixed.