TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI

# Statistical Algebraic Model Fitting
# MATHematics MAGnification™

Dr. Ralph deLaubenfels

TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI

Teacher-Scholar Institute

Columbus, Ohio

2020

## STATISTICAL ALGEBRAIC MODEL FITTING MAGNIFICATION

This is one of a series of very short books on math, statistics, and physics called "Math Magnifications." The "magnification" refers to focusing on a particular topic that is pivotal in or emblematic of mathematics.

## OUTLINE

This Magnification addresses data that we expect, at least on average, to fit a certain algebraic model.

One example is a random variable, call it $Y$, that is, on average, a function of another random variable, call it $X$. The ultimate goal is to use measurements of $X$ to predict $Y$. This is sometimes called *regression.*

Another example is the class of problems with the acronym ANOVA ("analysis of variance"), with the scientifically and politically important null hypothesis of populations being equal, on average.

After putting our measurements of $Y$, denoted $y_1, y_2, \ldots, y_n$, into an ordered $n$-tuple

$$\vec{y} \equiv (y_1, y_2, \ldots, y_n),$$

we will formulate problems and solutions in an intuitive geometric way, as finding the best approximation of $\vec{y}$ from a desired model subset of ordered $n$-tuples by dropping a perpendicular onto said model. See 4.2, 4.3, and 9.4.

The decomposition of $Y$ into two parts, one part explained by the model and the another part involving other factors, including random noise, is represented by a right triangle; see APP.13 in the Appendix, for the most general picture.

We will give detailed exposition and problems primarily (Chapters I through VIII) for *linear regression*, meaning that $Y$ is, on average, a linear function of $X$. That is, the expected value of $Y$, denoted $E(Y)$, equals

$$\beta_0 + \beta_1 x,$$

for some fixed numbers $\beta_0$ and $\beta_1$.

The line

$$y = \beta_0 + \beta_1 x$$

is then called the *true* or *population regression line* (see Definition 1.2); this terminology, in particular the use of the depressing-sounding word "regression," will be explained in 1.5.

We will perform statistical inference (meaning both confidence intervals and hypothesis tests, as in [5]) on both $\beta_0$ and $\beta_1$, the y-intercept and slope, respectively, of the true regression line, in Chapters VI and VII. A third parameter, the *variance* of $Y$, denoted $\sigma^2$, will also be of interest, giving us a clue about how close $Y$ is to its true regression line.

Statistical inference on a parameter begins with an estimator of said parameter. For the parameters $\beta_0, \beta_1$, and $\sigma^2$ of linear regression, the choice of estimator is much more challenging and surprising than the choice of estimators for population mean and proportion, as in [5]. See 3.1 and 4.5.

The data for regression is sets of ordered pairs $(x, y)$, that is, sets of points in the Cartesian plane, corresponding to $X$ and $Y$ being measured in pairs. For linear regression we will estimate the true regression line, hence both $\beta_0$ and $\beta_1$ simultaneously, by choosing the line

$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$

that is "closest" to the data. This closest line will be called the *estimated regression line* or *least-squares line* for the data.

The meaning of "closest" and "least squares" will be made clear in Definitions 2.3.

The numbers $\hat{\beta}_0$ and $\hat{\beta}_1$, calculated from the data, will be *least-squares estimators* of $\beta_0$ and $\beta_1$ respectively.

Regarding any set of points in the Cartesian plane, we will introduce in Chapter V a number denoted $r$, the *sample correlation coefficient*, that measures how close said set of points is to a line. In linear regression, $r$ will measure the proportion of $Y$'s activity that is due to activity of $X$ that is transmitted to $Y$ with the (true) regression lines.

The formulas needed for linear regression are summarized in Chapter VIII.

Chapter IX will give an overview of how the same ideas for linear regression may be applied to other polynomial regression and ANOVA.

Besides our usual favorite algebra reference [8], we assume the reader has read [5], or a similar introductory exposition of probability and statistics.

For those who would like to see proofs of the results in Chapters I through IX, we have an extensive Appendix between Chapter IX and the Homework. The Appendix requires some knowledge of linear algebra, although we sketch much of what is needed in the Appendix.

We will adopt the following convention in this Magnification. Any number coming from a probability table will be stated as being *equal* to the number we want, even though it is almost always only an approximation.

We also follow the usual custom of upper-case letters being random variables, lower-case letters being measurements of a random variable; e.g., a measurement of the random variable $X$ is denoted $x$ and a sequence of $n$ measurements of $X$ might be denoted $x_k, k = 1, 2, 3, \ldots x_n$.

## Chapter I. Simple Linear Regression Model.

**Definition 1.1. Bivariate data** is measurements $\{(x_k, y_k)\}_k$ of two random variables $X$ and $Y$ in pairs. For example, for each fixed $k$, $x_k$ and $y_k$ might be measured at the same time, or at the same place.

In practice, the values of $x_k$ are easy to measure, control, or predict or are specified in advance, while $Y$ is what we care about. Our goal is then to use $X$ to get information about $Y$.

Here are some examples.

$X$ might be baldness and $Y$ might be irritability; more specifically, for $1 \leq k \leq n$, $x_k$ could be the baldness of the $k^{th}$ person, $y_k$ the irritability of the $k^{th}$ person.

$X$ could be water for irrigation, $Y$ could be future crop yield.

$X$ could be cicada chirp volume, $Y$ temperature; using $X$ to predict $Y$ would be bug-based meteorology.

The measurement $x_k$ could be electricity use last year in the $k^{th}$ house ($1 \leq k \leq n$), $y_k$ the same thing this year.

$X$ could be length of stride, $Y$ height (popular with Sherlock Holmes when measuring distance between successive footprints).

$X$ could be age, $Y$ weight.

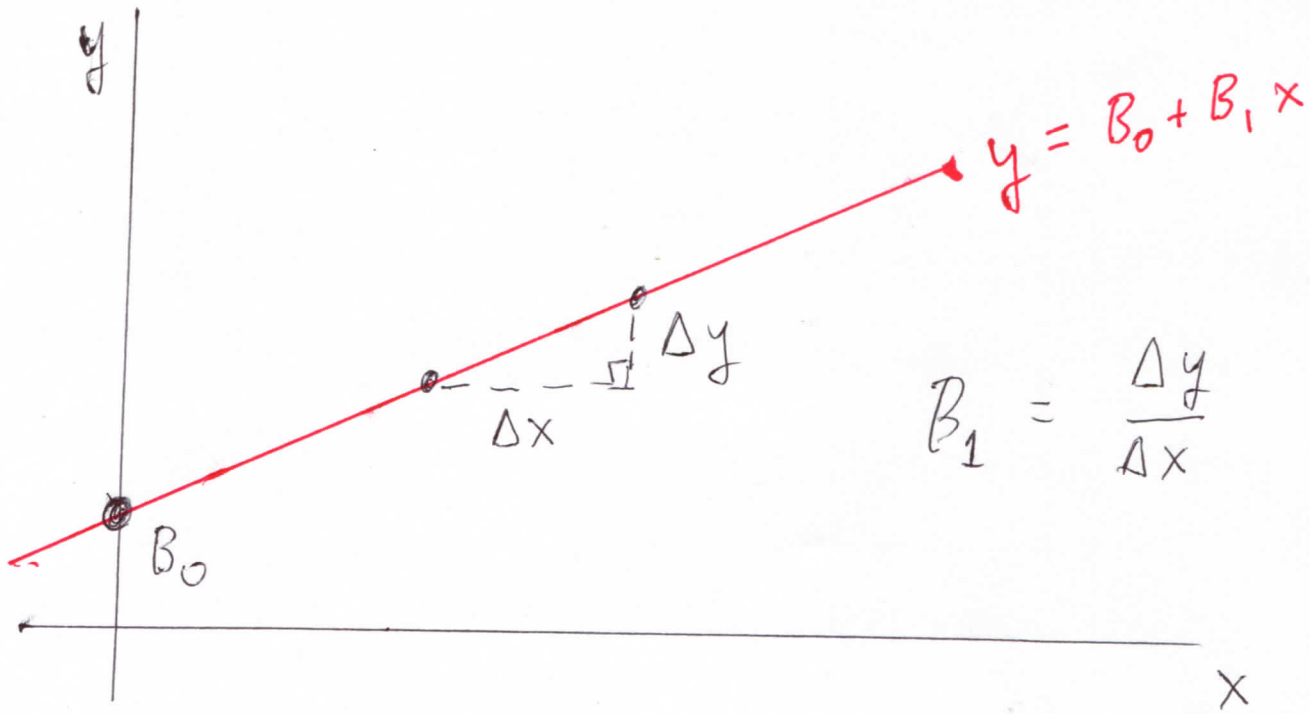
$X$ could be expected high temperature for the day, $Y$ could be ice cream sales on said day.

$X$ is called the **predictor**, or **explanatory**, or **independent** variable, $Y$ the **dependent** or **response** variable.

We are interested, in Chapters I through VIII of this Magnification, in the simplest relationship between random variables $X$ and $Y$, hence between their measurements $x$ and $y$ in our bivariate data, a *linear* relationship. In a world without randomness or uncertainty, hence without statistics, this would have the form

$$y = \beta_0 + \beta_1 x,$$

for some fixed numbers $\beta_0$ and $\beta_1$. The lack of randomness in this model earns it the description of *deterministic*.



In practice, this linear relationship is confused by random errors in measurement or by the presence of other variables besides $x$ that affect $Y$. Here is the most popular probabilistic version of a linear relationship between $x$ and $Y$

**Definition 1.2.** Given fixed numbers $\beta_0, \beta_1$, and $\sigma$, the **Simple Linear Regression Model** is

$$Y = \beta_0 + \beta_1 x + \mathcal{E}.$$

$x$ will be specified measurements and $\mathcal{E}$ is a normal random variable with mean $E(\mathcal{E}) = 0$ and variance $V(\mathcal{E}) = \sigma^2$, hence standard deviation $\sigma$.

The random variable $\mathcal{E}$ or its measurements $\epsilon$ is called **random error**, **random deviation**, or **noise**. Note that, for each fixed $x$, $Y$ is also a normal random variable, with variance $\sigma^2$ and expectation

$$E(Y) = \beta_0 + \beta_1 x.$$

The line $y = \beta_0 + \beta_1 x$ is called the **true** or **population regression line**.

Arguably $Y$ should be written $Y|x$, or in some fashion its dependence on $x$ should be made clear, but we will usually use the customary simplification of just $Y$.

**Example 1.3.** Suppose

$$Y = -2 + 3x + \mathcal{E},$$

for some $\mathcal{E}$ as in Definition 1.2, and we measure the following set of ordered pairs $(x, y)$: $\{(0, -1), (1, 0), (3, 5)\}$.

On the next page we draw the true regression line $y = -2 + 3x$, along with the values of

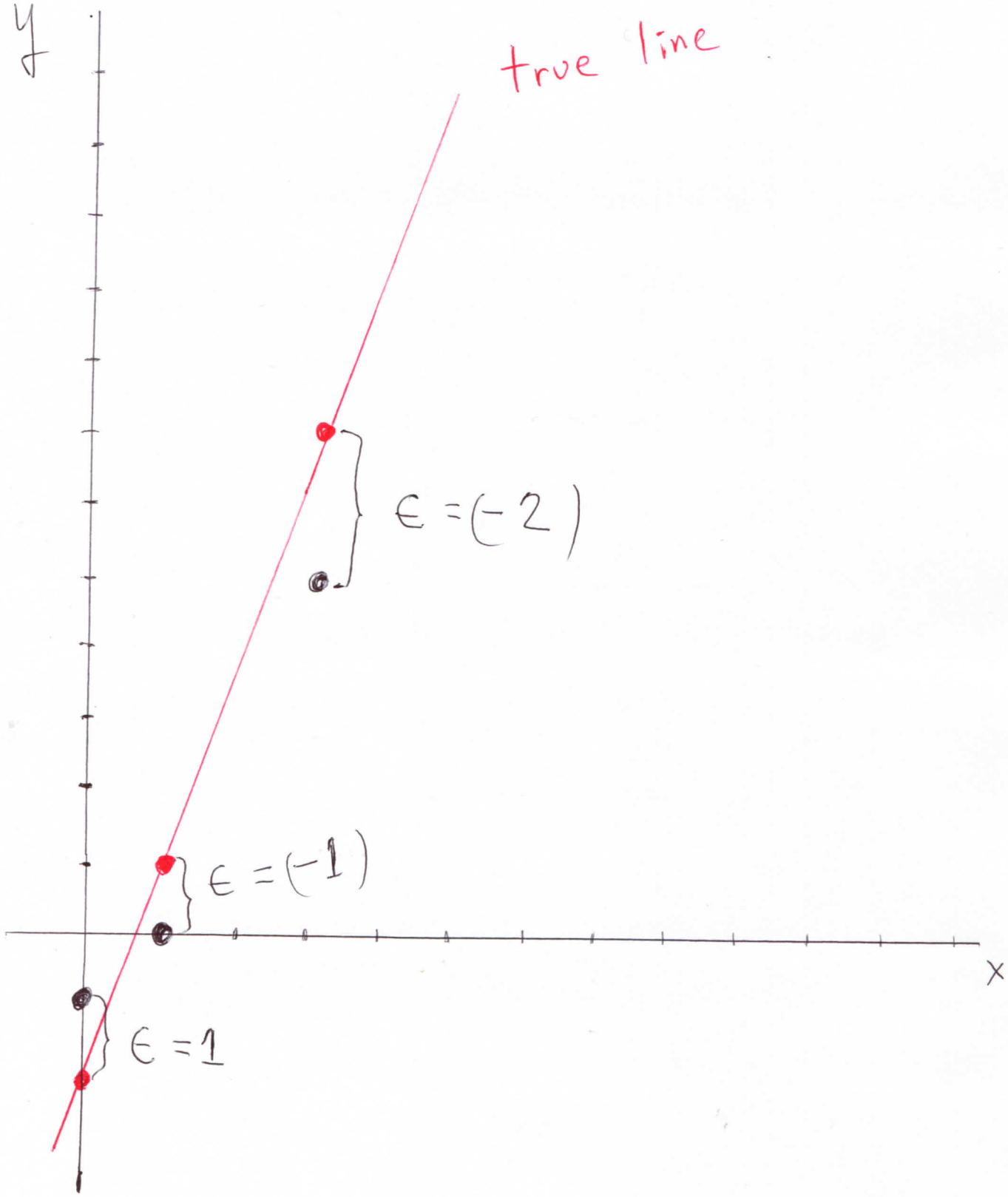
$$\epsilon = y - (-2 + 3x)$$

that distort our line.

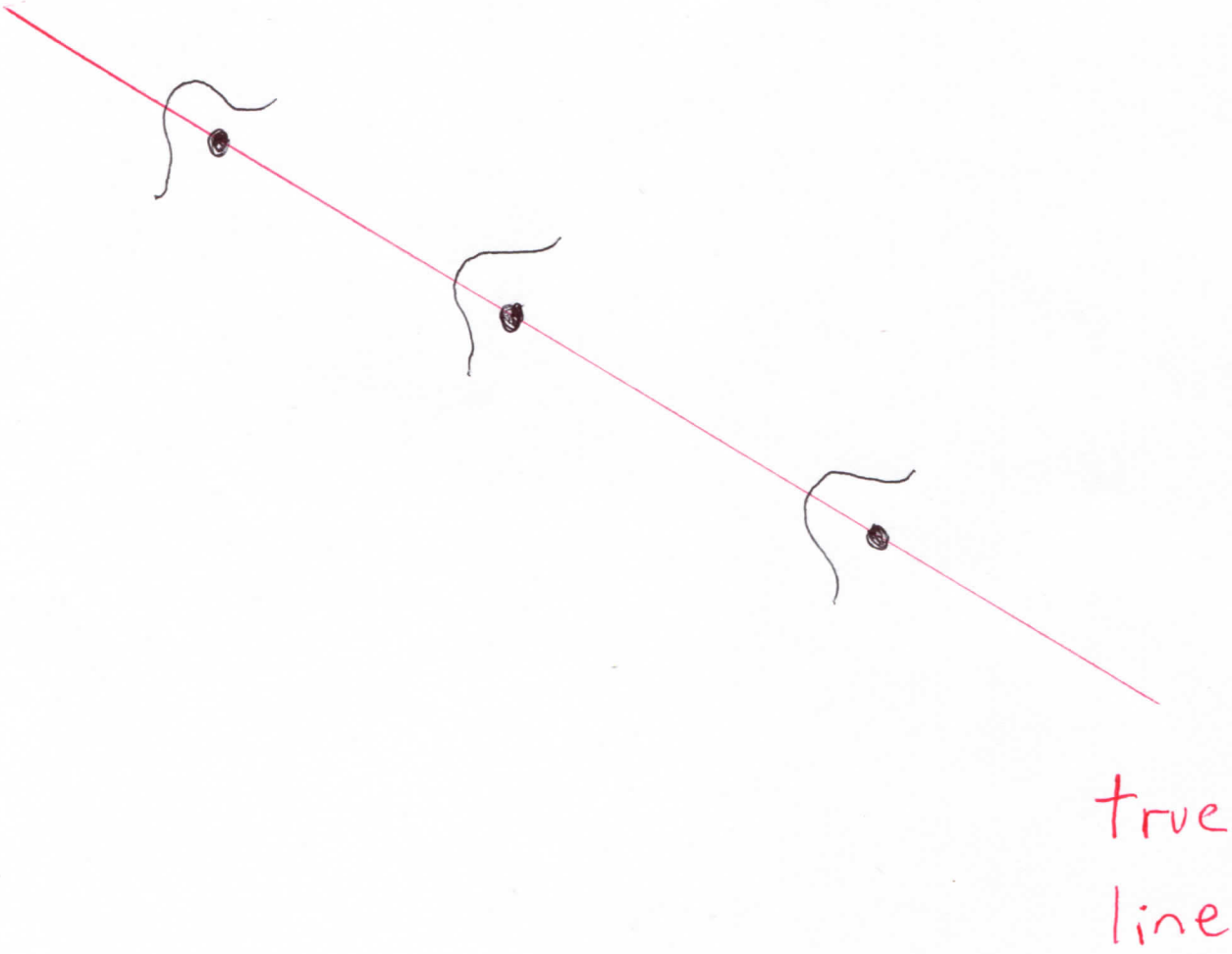
First let's organize our data:

| $x$ | 0 | 1 | 3 |
|---|---|---|---|
| $y$ | $-1$ | 0 | 5 |
| $(-2 + 3x)$ | $-2$ | 1 | 7 |
| $\epsilon$ | 1 | $-1$ | $-2$ |

We draw the true regression line $y = -2 + 3x$ in red, with large black dots for the measured ordered pairs $(x, y)$, and black brackets for $\epsilon = y - (-2 + 3x)$, taking us vertically from the red line to the black dots.


y
true line
$\epsilon = (-2)$
$\epsilon = (-1)$
$\epsilon = 1$
X

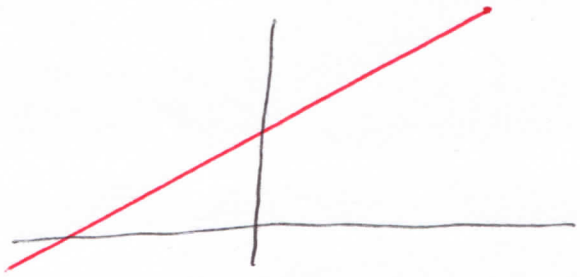
In general, think of our measured ordered pairs $(x, y)$ as beginning with a straight line, then getting "fuzzed" by $\epsilon$, which attaches a bell curve to each point on the line.



The larger $\sigma^2$, the variance of $\mathcal{E}$ is, the more fuzzing is likely.

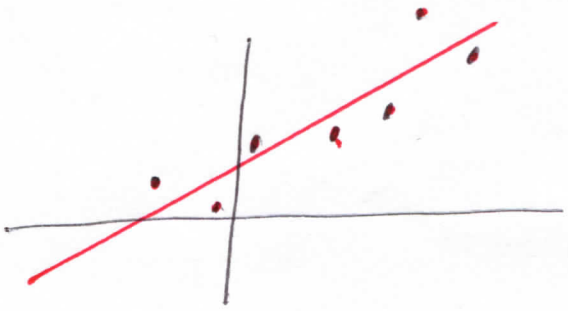
no $\mathcal{E}$ :

($\sigma^2 = 0$)



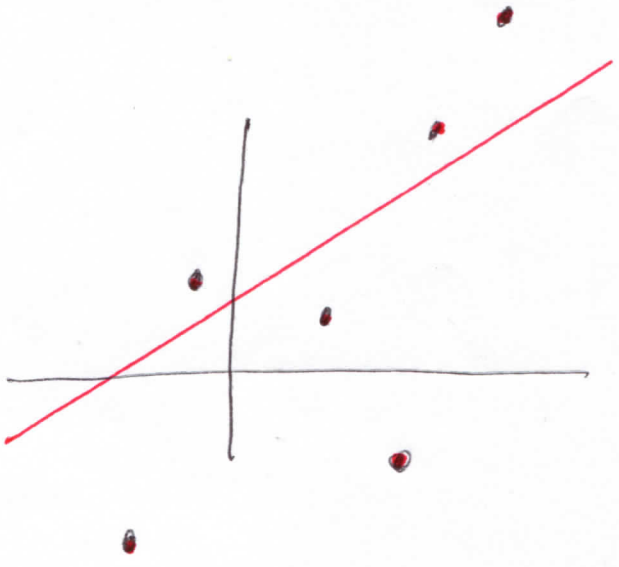
$\mathcal{E}$ "skinny"

($\sigma^2$ small)



$\mathcal{E}$ "fat"

($\sigma^2$ large)



**Examples 1.4.** Suppose weight, in pounds, denoted $Y$, is related to age, in years after birth, denoted x, by the Simple Linear Regression Model

$$Y = 7 + 9x + \mathcal{E}.$$

Notice that there need to be restrictions on $x$, to make this model believable. The independent variable $x$ must be positive, and probably less than or equal to about 12.

The true regression line $Y = 7 + 9x$ is saying that, on average, a newborn weighs 7 pounds, and children gain 9 pounds per year.

(a) What is the expected weight of a five-year-old child?

(b) How much do you expect weight to change in three years?

(c) Suppose the standard deviation $\sigma$ of $\mathcal{E}$ is twelve. What's the probability that an eight year old weighs more than 100 pounds?

**Answers.** (a) This is $E(Y|x = 5) = 7 + 9 \times 5 = 52$ pounds.

(b) Since expected weight gain per year is 9 pounds, we expect $3 \times 9 = 27$ pounds of weight change in three years.

(c) Abbreviating $Y$ for $(Y|x = 8)$, since $Y$ is normal we change $Y$ to the standard normal $Z$:

$$P(Y > 100) = P\left(Z > \frac{100 - E(Y)}{\sigma}\right) = P\left(Z > \frac{100 - (7 + 9 \times 8)}{12}\right) = P\left(Z > \frac{100 - 79}{12}\right)$$
$$= P(Z > 1.75) = 0.0401,$$

from the $Z$ tables at the end of this Magnification.

**Terminology Remarks 1.5.** The term "regression" in Definition 1.2 is due to Francis Galton in the late 1800s. He considered bivariate data as in Definition 1.1, with $X$ a father's height, $Y$ the son's height. He noticed that $|y - \overline{y}| < |x - \overline{x}|$, on average; this is *regression to the mean* as generations pass. Informally, tall fathers have tall sons, but not *as* tall, on average.

See [9, Chapter 8] or [7, Chapter 16], for extensive discussion of the origins of regression.

The "simple" in Definition 1.2 refers to there being only one independent variable.

## Chapter II. Assumptions, Goals, and Terminology.

**Assumptions 2.1.** For Chapters II–VIII of this Magnification, the set of ordered pairs

$$\{(x_k, y_k) \mid k = 1, 2, 3, \ldots, n\} = \{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$$

will be bivariate data as in Definition 1.1, with $Y$ and $x$ satisfying the Simple Linear Regression Model in Definition 1.2.

**Goals 2.2.** We wish to perform statistical inference on the parameters $\beta_0$, $\beta_1$, and $\sigma^2$, from Definition 1.2. This begins with choosing estimators (see [3, Definitions 19]) of said parameters.

The choice of an estimator for a parameter sometimes seems obvious or inevitable; for example, estimating the population mean with the sample mean or the population proportion with the sample proportion (see [3, Definitions 6]). Estimators for the parameters in Definition 1.2 are not so clear.

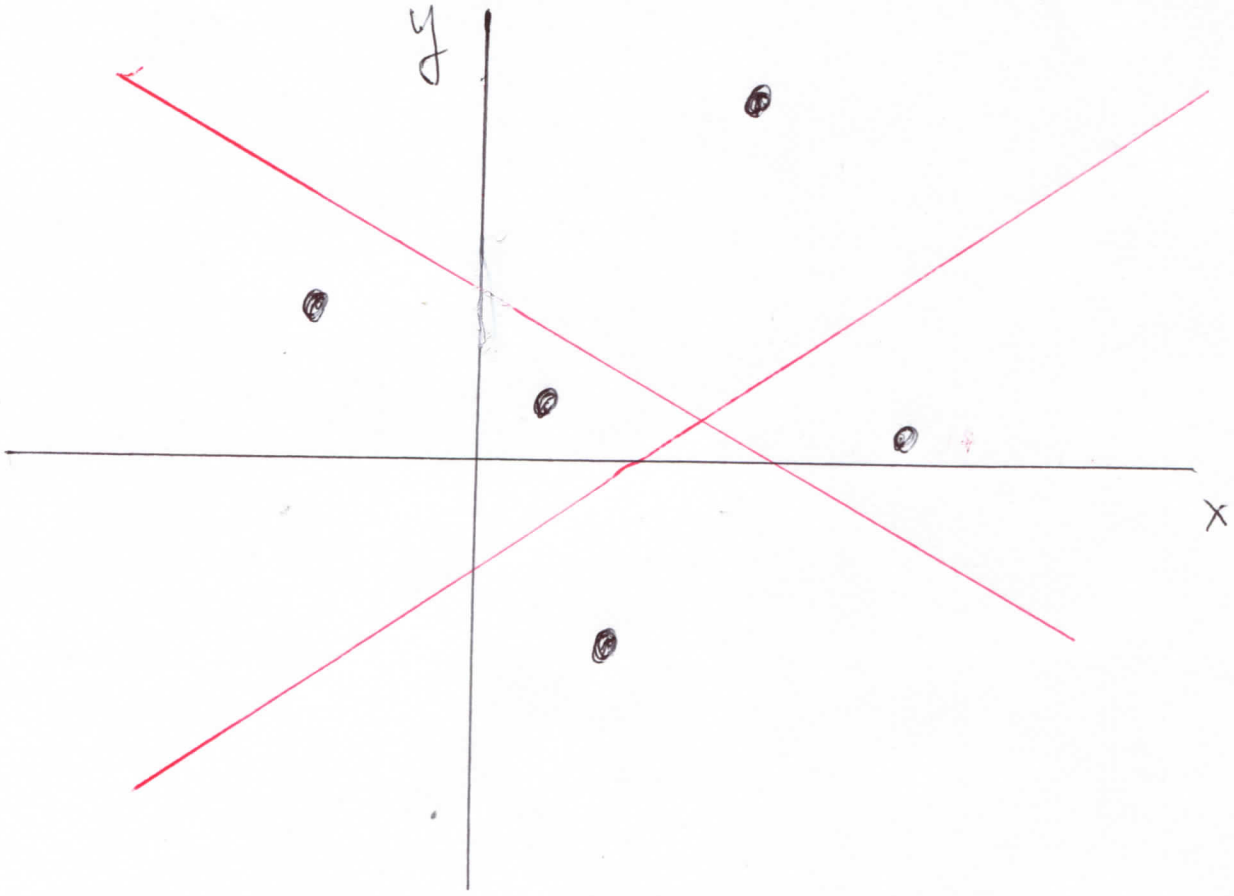
The estimator for $\sigma^2$ we will introduce later (Definition 4.5). Choosing estimators, call them $\hat{\beta}_0$ and $\hat{\beta}_1$, of $\beta_0$ and $\beta_1$, respectively, is equivalent to choosing a line

$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$

in the $(x, y)$ plane that approximates the true population regression line $y = \beta_0 + \beta_1 x$ of Definition 1.2.

We want the line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ that is "closest" to the bivariate data of Assumptions 2.1. That word "closest" really *needs* quotation marks, because there are so many things it could mean.

For example, in the drawing below of two lines, in red, and bivariate data, represented by black dots, which of the two lines is "closest" to the dots? The question is rhetorical, until we define "closeness," of a set of dots to a line.



The measure of "closeness" that works out the best, for linear algebra and Pythagorean theorem reasons (the reader familiar with vectors should see Theorems APP.10 and APP.11 in the Appendix), is sum of squares of vertical displacements.

**Definitions 2.3.** The **least-squares estimators** of $\beta_0$ and $\beta_1$ in the Simple Linear Regression Model Definition 1.2, denoted $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively, is the pair of numbers whose corresponding line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ minimizes the sum of squares of vertical displacements

$$SSV(b_0, b_1) \equiv \sum_{k=1}^{n} [y_k - (b_0 + b_1 x_k)]^2$$

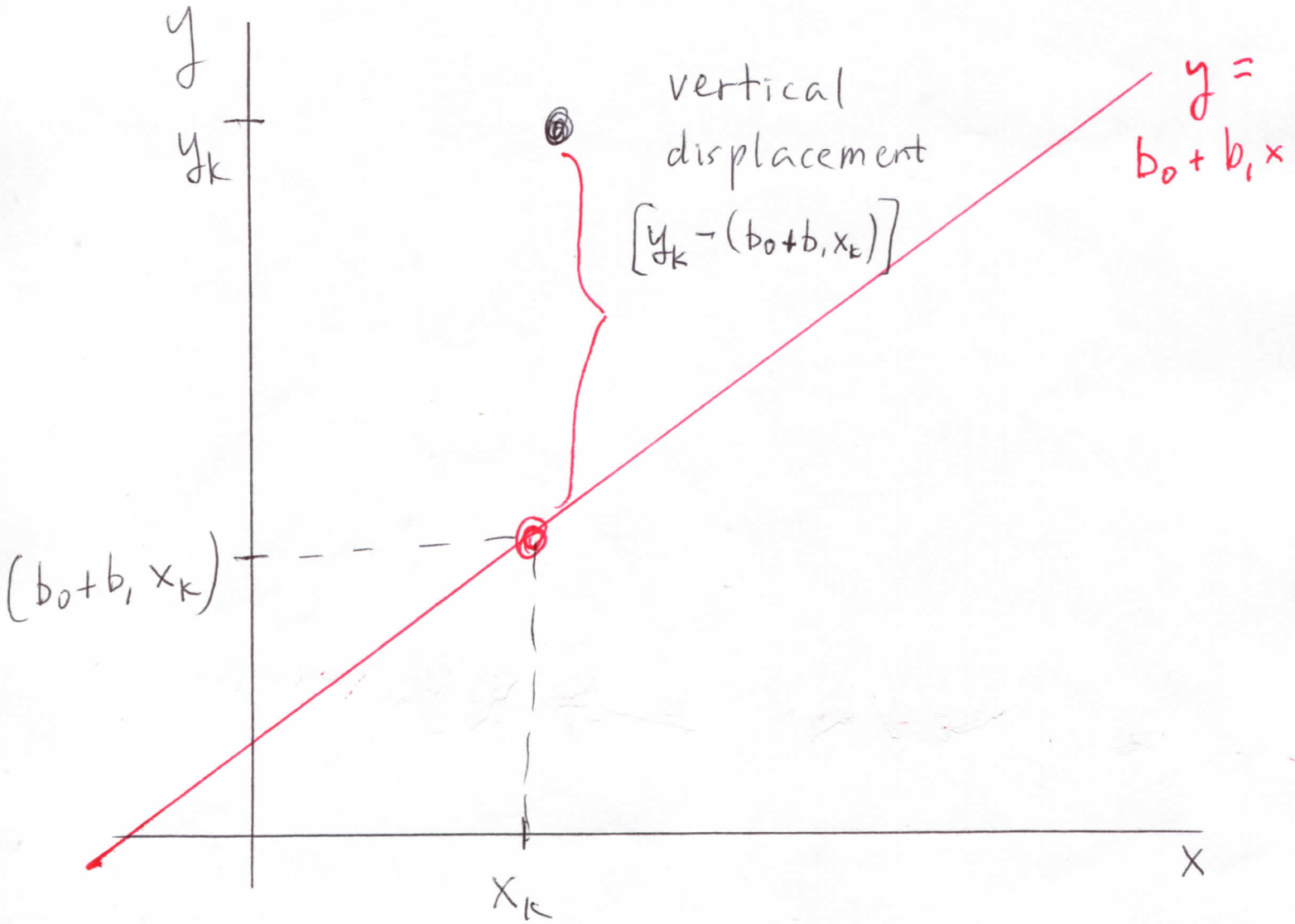
from the bivariate data to the line (see the drawing below); that is,

$$SSV(\hat{\beta}_0, \hat{\beta}_1) \leq SSV(b_0, b_1)$$

for all real numbers $b_0, b_1$.

The initials $SSV$ stand for "Sum of Squares of Vertical" (displacements). That minimum value $SSV(\hat{\beta}_0, \hat{\beta}_1)$ is given a special name: $SSE$, for "Sum of Squares of Errors," or "error sum of squares," in Definitions 4.2.

The line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ is then the **least-squares line** or **estimated regression line** for the bivariate data of 2.1.



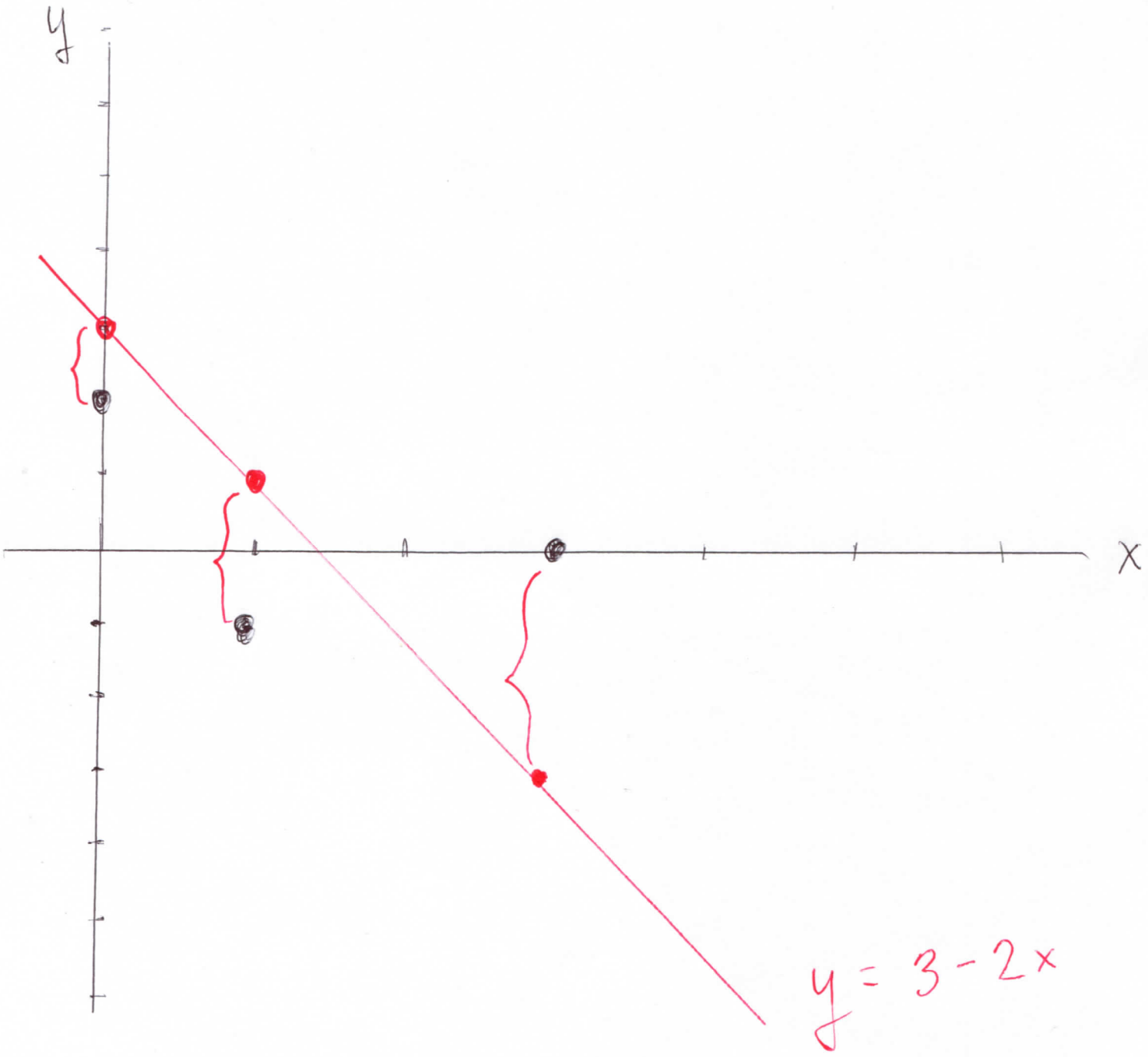
**Examples 2.4.** (a) For the bivariate data

$$\{(x_1, y_1) \equiv (0, 2), (x_2, y_2) \equiv (1, -1), (x_3, y_3) \equiv (3, 0)\}$$

how close is the line $y = 3 - 2x$, measured by the sum of the squares of vertical displacements from the bivariate data to the line, as in 2.3? That is, what is $SSV(3, -2)$ (see Definitions 2.3)?

**Answer.** Let's calculate $SSV(3, -2)$, from 2.3:

$$[y_1 - (3 - 2x_1)]^2 + [y_2 - (3 - 2x_2)]^2 + [y_3 - (3 - 2x_3)]^2 = [2 - (3 - 2 \times 0)]^2 + [(-1) - (3 - 2 \times 1)]^2 + [0 - (3 - 2 \times 3)]^2$$
$$= [2 - 3]^2 + [(-1) - 1]^2 + [0 - (-3)]^2 = (-1)^2 + (-2)^2 + 3^2 = 14.$$

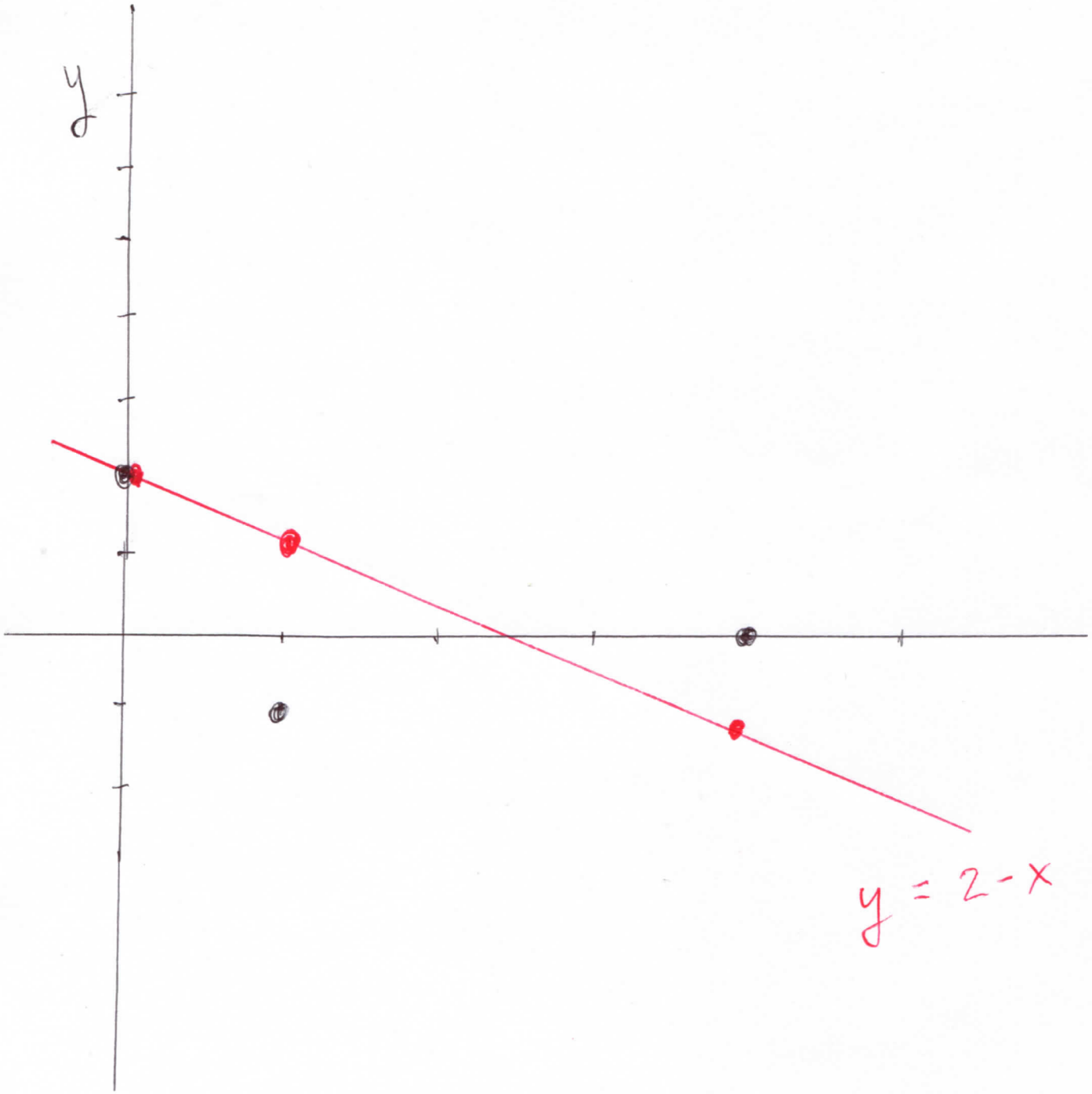


(b) Same question as (a), for $y = 2 - x$.

**Answer.** As with (a), calculate

$$SSV(2, -1) = [2 - 2]^2 + [(-1) - 1]^2 + [0 - (-1)]^2 = 0^2 + (-2)^2 + 1^2 = 5.$$

Since 5 is less than 14, the line $y = 2 - x$ is closer to the data then the line $y = 3 - 2x$, as measured using Definitions 2.3.



**Terminology 2.5.** Formulas for these estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, not to mention future statistical inference on said estimators, are most easily described using the following terminology.

For any pair of ordered $n$-tuples $\vec{w} \equiv (w_1, w_2, \ldots, w_n), \vec{z} \equiv (z_1, z_2, \ldots, z_n)$, define

$$S_{\vec{w},\vec{z}} \equiv \sum_{k=1}^{n} (w_k - \overline{w})(z_k - \overline{z});$$

recall from [3, Definitions 6] that the **sample means** $\overline{w}$ and $\overline{z}$ are

$$\overline{w} \equiv \frac{1}{n} \sum_{k=1}^{n} w_k \quad \text{and} \quad \overline{z} \equiv \frac{1}{n} \sum_{k=1}^{n} z_k.$$

**Remark 2.6.** The dubious (this is usually a good attribute) reader might be asking why we have squaring in Definitions 2.3; that is, why not minimize

$$\sum_{k=1}^{n} | \, [y_k - (b_0 + b_1 x_k)] \, |,$$

the sum of vertical distances from the bivariate data to the line?

One could ask a similar question about descriptive statistics; given data $x_1, x_2, \ldots x_n$, why do people usually work with sample variance, involving

$$\sum_{k=1}^{n} (x_k - \overline{x})^2,$$

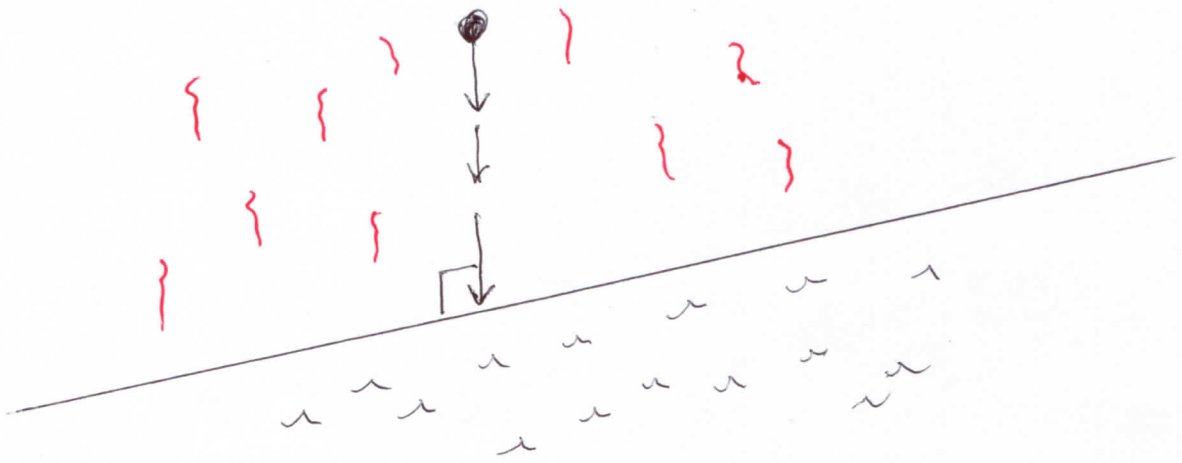
instead of absolute deviation, involving

$$\sum_{k=1}^{n} |x_k - \overline{x}|?$$

We can't help but mention here a Gary Larson "Far Side" cartoon involving $E = mc^2$ versus $E = mc^3, E = mc^4$, etc.; note that the desired exponent is "2."

A very indirect clue for our predilection for squaring is the *Pythagorean theorem*: the sum of the lengths of the sides does *not* equal the length of the hypotenuse, but the sum of the *squares* of the lengths of the sides equals the square of the length of the hypotenuse.

Stated very informally, dealing with squares instead of absolute values often gives us a notion of being *perpendicular*, which gives us an intuitive way to minimize things we don't like. See APP.9 and APP.10 in the Appendix, [1], [2, 6.14, page 412], and the drawing below, where the large dot is a person whose feet are burning on hot sand (with wriggly red lines rising above it), who wants to reach the ocean by as short a path as possible..



## Chapter III. Calculating Least-Squares Estimators and Lines

Here is how we will calculate the least-squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ from Definitions 2.3.

**Theorem 3.1.** $\hat{\beta}_1 = \frac{S_{\vec{x},\vec{y}}}{S_{\vec{x},\vec{x}}}$ and $\hat{\beta}_0 = \overline{y} - \overline{x}\hat{\beta}_1$.

**Proof:** For those readers familiar with vectors, see Examples APP.12(a). □

**Corollary 3.2.** The estimated regression line, also known as least-squares line, for the data in Assumptions 2.1 is

$$y = \left(\overline{y} - \overline{x}\hat{\beta}_1\right) + \hat{\beta}_1 x = \overline{y} + \hat{\beta}_1(x - \overline{x}), \quad \text{where} \quad \hat{\beta}_1 = \frac{S_{\vec{x},\vec{y}}}{S_{\vec{x},\vec{x}}}.$$

Before we do an example, here is a convenient formula for $S_{\vec{w},\vec{z}}$ in Terminology 2.5.

**Proposition 3.3 ("computational formula").**

$$S_{\vec{w},\vec{z}} = \sum_{k=1}^{n} w_k z_k - \frac{1}{n}\left(\sum_{k=1}^{n} w_k\right)\left(\sum_{k=1}^{n} z_k\right).$$

**Proof:**

$$\sum_{k=1}^{n}(w_k - \overline{w})(z_k - \overline{z}) = \sum_{k=1}^{n}[w_k z_k - (\overline{w})z_k - w_k(\overline{z}) + (\overline{w})(\overline{z})]$$

$$= \left(\sum_{k=1}^{n} w_k z_k\right) - \left((\overline{w})\sum_{k=1}^{n} z_k\right) - \left((\overline{z})\sum_{k=1}^{n} w_k\right) + (n(\overline{w})(\overline{z}))$$

$$= \left(\sum_{k=1}^{n} w_k z_k\right) - (\overline{w})(n\overline{z}) - (\overline{z})(n\overline{w}) + n(\overline{w})(\overline{z}) = \left(\sum_{k=1}^{n} w_k z_k\right) - n(\overline{w})(\overline{z})$$

$$= \left(\sum_{k=1}^{n} w_k z_k\right) - \frac{1}{n}\left(\sum_{k=1}^{n} w_k\right)\left(\sum_{k=1}^{n} z_k\right).$$

□

**Example 3.4.** Find the least-squares line for the data

$$\{(0,2),(3,0),(2,6),(3,-2)\}.$$

We recommend setting up the following table.

| $k$ | $x_k$ | $y_k$ | $x_k^2$ | $x_k y_k$ |
|---|---|---|---|---|
| 1 | 0 | 2 | 0 | 0 |
| 2 | 3 | 0 | 9 | 0 |
| 3 | 2 | 6 | 4 | 12 |
| 4 | 3 | −2 | 9 | −6 |
| $\text{sum}_k \equiv \sum_{k=1}^{4}$ | 8 | 6 | 22 | 6 |

Everything we need is in that final row of sums:

$$S_{\vec{x},\vec{y}} = \sum_k x_k y_k - \frac{1}{n}(\sum_k x_k)(\sum_k y_k) = 6 - \frac{1}{4}(8)(6) = -6;$$

$$S_{\vec{x},\vec{x}} = \sum_k x_k^2 - \frac{1}{n}(\sum_k x_k)^2 = 22 - \frac{1}{4}(8)^2 = 6;$$

$$\overline{x} = \frac{1}{n}(\sum_k x_k) = \frac{8}{4} = 2; \quad \overline{y} = \frac{1}{n}(\sum_k y_k) = \frac{6}{4} = 1.5,$$

thus

$$\hat{\beta}_1 = \frac{S_{\vec{x},\vec{y}}}{S_{\vec{x},\vec{x}}} = \frac{-6}{6} = -1,$$

and

$$\hat{\beta}_0 = \overline{y} - \overline{x}\hat{\beta}_1 = 1.5 - 2(-1) = 3.5,$$

so that the equation of our least-squares line is

$$y = 3.5 - x.$$

**Example 3.5.** I believe that the number of decades a bridge lasts, as a function of the average temperature in degrees Celsius during construction, satisfies the Simple Linear Regression Model. Writing y for the number of decades a bridge lasts and x for the temperature in degrees Celsius during construction, I collect data on 20 bridges

$$(x_1, y_1), (x_2, y_2), ..., (x_{20}, y_{20}),$$

and get the following summaries:

$$\sum_j x_j = -2, \quad \sum_j x_j^2 = 5, \quad \sum_j y_j = 30, \quad \sum_j y_j^2 = 57, \quad \sum_j x_j y_j = 3.$$

Get the least-squares estimators of $\beta_0$ and $\beta_1$, in the Simple Linear Regression Model, and the least squares line or estimated regression line.

**Answer.** Using Proposition 3.3,

$$S_{\vec{x},\vec{y}} = \sum_k x_k y_k - \frac{1}{n}(\sum_k x_k)(\sum_k y_k) = 3 - \frac{1}{20}(-2)(30) = 6;$$

$$S_{\vec{x},\vec{x}} = \sum_k x_k^2 - \frac{1}{n}(\sum_k x_k)^2 = 5 - \frac{1}{20}(-2)^2 = 4.8.$$

We also need

$$\overline{y} = \frac{1}{n}\sum_k y_k = \frac{1}{20}(30) = 1.5 \quad \text{and} \quad \overline{x} = \frac{1}{n}\sum_k x_k = \frac{1}{20}(-2) = -0.1.$$

Now, by Theorem 3.1, we have

$$\hat{\beta}_1 = \frac{6}{4.8} = 1.25,$$

$$\hat{\beta}_0 = 1.5 - (-0.1)(1.25) = 1.625,$$

so our estimated regression line is

$$y = 1.625 + (1.25)x.$$

## Chapter IV. More Alphabet Soup

The strings of letters and numbers $SSV$ (Definitions 2.3) and $S_{\vec{w},\vec{z}}$ (Terminology 2.5) are already pretty soupy. To discuss the value of our least-squares line

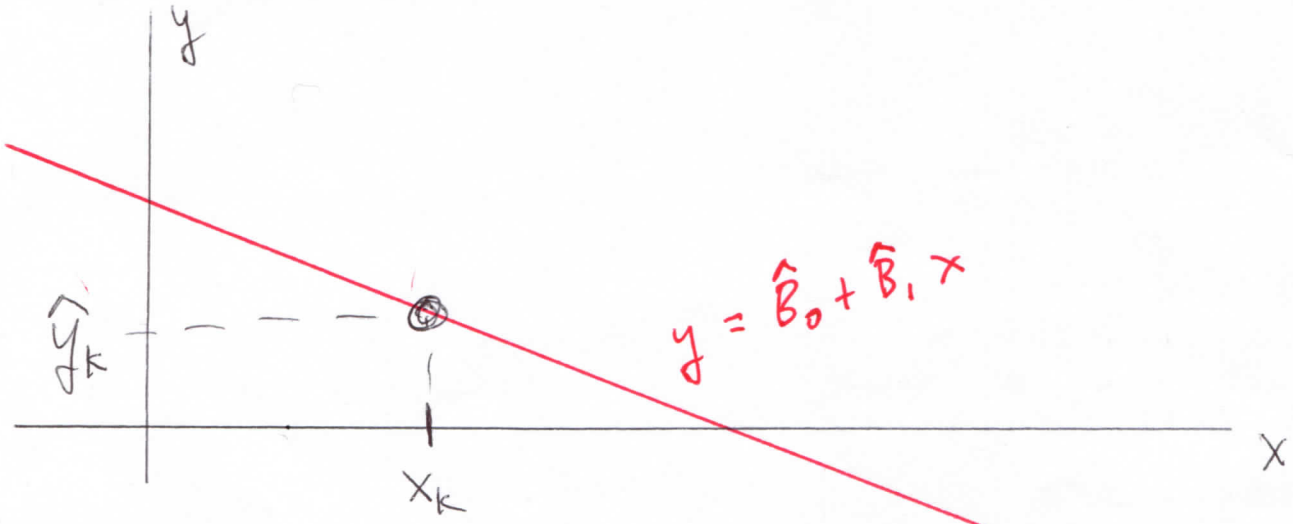
$$y = \hat{\beta}_0 + \hat{\beta}_1 x,$$

defined in 2.3 and calculated in Chapter III, and perform statistical inference on the parameters associated with said line, we need more terminology.

**Definitions 4.1.** The **fitted** or **predicted** values are

$$\hat{y_k} \equiv \hat{\beta}_0 + \hat{\beta}_1 x_k = \overline{y} + \hat{\beta}_1 (x_k - \overline{x}) \ \ (k = 1, 2, \ldots, n),$$

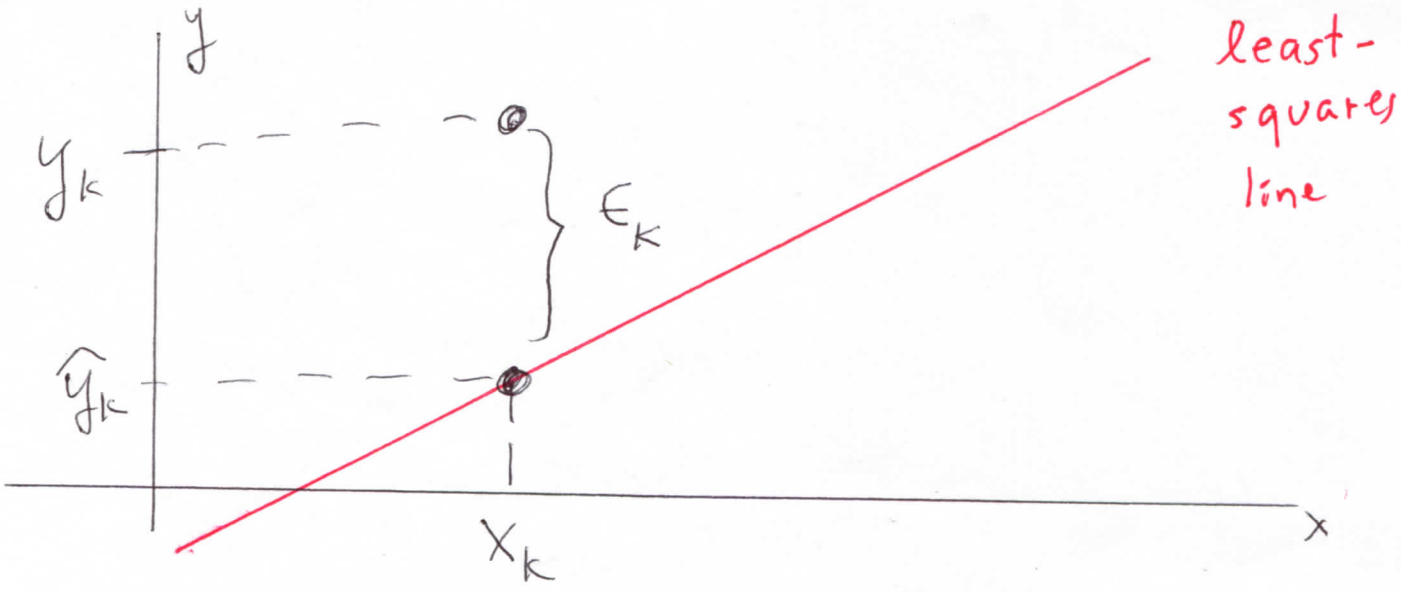
the $y$ values we obtain when plugging the specified $x$ values into our least-squares line; see Assumptions 2.1 and Corollary 3.2.




We should worry about how close $\hat{y_k}$ is to the measured value of $y_k$ when $x = x_k$, as in 2.3. For $k = 1, 2, \ldots, n$, define the $k^{th}$ **residual**

$$\epsilon_k \equiv (y_k - \hat{y_k});$$

this is the vertical displacement of the ordered pair $(x_k, y_k)$ from the least-squares line, as in 2.3; the Greek letter $\epsilon$ stands for "error."



**Definitions 4.2.** To describe the relationship of our least-squares line to the bivariate data of 2.1, in particular, its ability to explain the variability of $y_k, k = 1, 2, 3, \ldots, n$, define the **total sum of squares**

$$SST \equiv \sum_{k=1}^{n} (y_k - \overline{y})^2 \sim \text{“observed variation in } y\text{”},$$

the **regression sum of squares**

$$SSR \equiv \sum_{k=1}^{n} (\hat{y_k} - \overline{y})^2 \sim \text{“observed variation in } y \text{ explained by } x \text{ and the linear model”},$$

and the **error sum of squares**

$$SSE \equiv \sum_{k=1}^{n} (y_k - \hat{y_k})^2 \equiv \sum_{k=1}^{n} \epsilon_k^2 \sim \text{“observed variation in } y \text{ not explained by } x \text{ and the linear model”},$$

Note that $SSE = SSV(\hat{\beta}_0, \hat{\beta}_1)$, from Definitions 2.3.

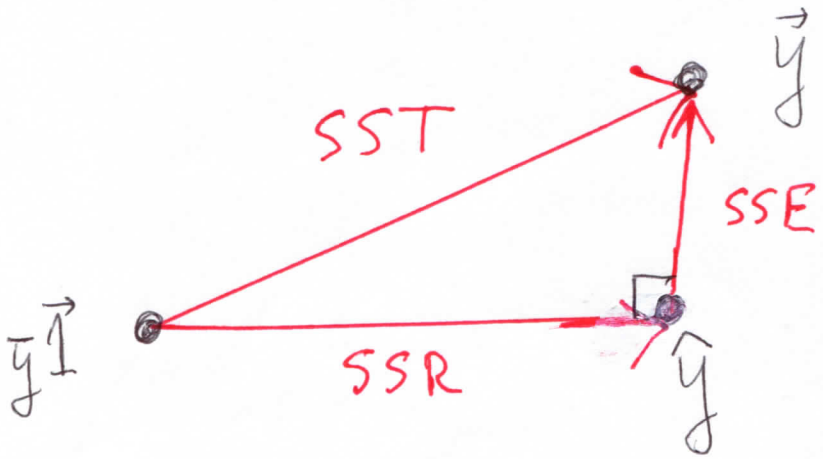
To get a very helpful and suggestive picture of all these sums of squares, put together our data and our fitted values into ordered $n$-tuples

$$\vec{y} \equiv (y_1, y_2, \ldots, y_n), \quad \hat{y} \equiv (\hat{y_1}, \hat{y_2}, \ldots, \hat{y_n}), \quad \text{and} \quad \overline{y}\vec{1} \equiv (\overline{y}, \overline{y}, \ldots, \overline{y});$$

our sums of squares are then seen as length squared of arrows (known as *vectors*; see the Appendix, especially Definitions APP.1) from one of the $n$ tuples just defined to another, in the right triangle below.

**Regression Picture 4.3.** The right-angle symbol in the picture below may be taken as metaphorical for now; APP.4 in the Appendix defines two vectors being perpendicular, in possibly more than two dimensions.

If we accept the picture below, then a Pythagorean theorem (see APP.5 in the Appendix) suggests that SST = SSR + SSE. We will sharpen this result, and Picture 4.3, in the next chapter.




**Example 4.4.** Let's get SST, SSR, and SSE for the data in Example 3.4.

Put four more columns on the table we constructed in Example 3.4 and use the fact, from 3.4 calculations, that $\overline{y} = 1.5$.

| $k$ | $x_k$ | $y_k$ | $x_k^2$ | $x_k y_k$ | $y_k^2$ | $\hat{y_k} = (3.5 - x_k)$ | $(\hat{y_k} - \overline{y})$ | $\epsilon_k = (y_k - \hat{y_k})$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 2 | 0 | 0 | 4 | 3.5 | 2 | −1.5 |
| 2 | 3 | 0 | 9 | 0 | 0 | 0.5 | −1 | −0.5 |
| 3 | 2 | 6 | 4 | 12 | 36 | 1.5 | 0 | 4.5 |
| 4 | 3 | −2 | 9 | −6 | 4 | 0.5 | −1 | −2.5 |
| $\text{sum}_k \equiv \sum_{k=1}^4$ | 8 | 6 | 22 | 6 | 44 | 6 | 0 | 0 |

Note that the fitted values $\hat{y_k}$ are in the seventh column and the residuals $\epsilon_k$ are in the ninth (last) column.

Staring at the sums in the table gives us

$$SST = S_{\vec{y},\vec{y}} = \sum_k y_k^2 - \frac{1}{n}(\sum_k y_k)^2 = 44 - \frac{1}{4}(6)^2 = 35;$$

$$SSR = \sum_{k=1}^{n} (\hat{y_k} - \overline{y})^2 = 2^2 + (-1)^2 + 0^2 + (-1)^2 = 6;$$

and

$$SSE \equiv \sum_{k=1}^{n} (y_k - \hat{y_k})^2 = (-1.5)^2 + (-0.5)^2 + (4.5)^2 + (-2.5)^2 = 29.$$

Please note that $SSR + SSE = 6 + 29 = 35 = SST$. Note also that each of the last two columns add up to zero.

We will redo Example 4.4 in an easier way (see Example 5.5), using *correlation* (Definition 5.1).

**Definition 4.5.** We are now in a position to define our favorite estimator of $\sigma$ from the Simple Linear Regression Model 1.2:

$$s \equiv \hat{\sigma} \equiv \sqrt{\frac{SSE}{(n-2)}}.$$

In the previous Example 4.4, $s$, the preferred estimator of $\sigma$, equals $\sqrt{\frac{29}{4-2}} = \sqrt{14.5}$.

As an informal rationale for $s^2$, notice that both $\sigma^2$ and $SSE$ are measuring the deviation from the desired linear model; see the drawings in Example 1.3 just before Examples 1.4 for $\sigma^2$ and Regression Picture 4.3 for $SSE$.

For the division by $(n-2)$ in the definition of

$$s^2 = \frac{SSE}{(n-2)},$$

we can only make vague statements about *free variables* (also known as *dimension* in linear algebra) in 4.2 and 4.3: $SSR$ has 2 free variables, the numbers $b_0$ and $b_1$ in the linear model, while $SST$ has $n$ free variables, in the raw data $y_1, y_2, \ldots, y_n$, leaving $(n-2)$ free variables for $SSE$.

It is also the case that $E(s^2)$, the expected value of $s^2$, is $\sigma^2$, the parameter $s^2$ is meant to estimate. This highly desirable property of an estimator—having expected value equal to the parameter being estimated—is called being *unbiased*.

**Example 4.6.** Suppose we do regression on data

$$\{(1,5),(0,2),(1,4),(0,2),(3,0)\}.$$

and get an estimated regression line of $y = 3.27 - (0.667)x$.

(a) Get the fitted values (for $x = 0, 1, 3$).

(b) Get the (five) residuals.

(c) Get $SSE$.

(d) Get $s^2$, our favorite estimator of $\sigma^2$, in the Simple Linear Regression Model.

**Answers.** (a) Plugging $x$ into the estimated regression line and writing "fit" for "fitted value":

| $x$ | 0 | 1 | 3 |
|---|---|---|---|
| fit | $(3.27 - 0.667 \times 0)$ | $(3.27 - 0.667 \times 1)$ | $(3.27 - 0.667 \times 3)$ |

or

| $x$ | 0 | 1 | 3 |
|---|---|---|---|
| fit | 3.27 | 2.603 | 1.269 |

Denoting

$$(x_1, y_1) \equiv (1,5), (x_2, y_2) \equiv (0,2), (x_3, y_3) \equiv (1,4), (x_4, y_4) \equiv (0,2), (x_5, y_5) \equiv (3,0),$$

these are fitted values

$$\hat{y_1} = 2.603, \hat{y_2} = 3.27, \hat{y_3} = 2.603, \hat{y_4} = 3.27, \hat{y_5} = 1.269.$$

(b)

$$\epsilon_1 \equiv (y_1 - \hat{y_1}) = (5 - 2.603) = 2.397, \epsilon_2 \equiv (y_2 - \hat{y_2}) = (2 - 3.27) = -1.27, \epsilon_3 \equiv (y_3 - \hat{y_3}) = (4 - 2.603) = 1.397,$$
$$\epsilon_4 \equiv (y_4 - \hat{y_4}) = (2 - 3.27) = -1.27, \epsilon_5 \equiv (y_5 - \hat{y_5}) = (0 - 1.269) = -1.269.$$

(c) $SSE$ equals

$$\sum_{k=1}^{5} (\epsilon_k)^2 = (2.397)^2 + (-1.27)^2 + (1.397)^2 + (-1.27)^2 + (-1.269)^2 \sim 12.533.$$

(d)

$$s^2 = \frac{1}{(n-2)} SSE = \frac{1}{(5-2)} \left[ (2.397)^2 + (-1.27)^2 + (1.397)^2 + (-1.27)^2 + (-1.269)^2 \right] \sim 4.178.$$

## Chapter V. Correlation

The quantities $SSE$ and $SSR$, introduced in Definitions 4.2 and illustrated in Regression Picture 4.3, represent our first attempt to describe how well our least-squares line

$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$

fits the bivariate data in 2.1. Our intuition is that large $SSR$ means a good fit, while large $SSE$ means a poor fit.

Our objection to this first attempt is that measurement techniques can artificially change both $SSE$ and $SSR$. For example, they are unstable under changes of units, e.g., changing dollars to cents would multiply all data by 100.

To standardize, we could divide both $SSE$ and $SSR$ by $SST$. This turns out to be something surprisingly simple and of great independent interest; see Theorem 5.3.

**Definitions 5.1.** The **sample correlation coefficient** for bivariate data as in 2.1 is

$$r \equiv \frac{S_{\vec{x},\vec{y}}}{\sqrt{S_{\vec{x},\vec{x}} S_{\vec{y},\vec{y}}}}.$$

$r^2$ is called the **coefficient of determination**.

When $r$ appears, we will assume both $S_{\vec{x},\vec{x}}$ and $S_{\vec{y},\vec{y}}$ are nonzero.

**Example 5.2.** Get the sample correlation coefficient and coefficient of determination for the data $\{(-1,3),(0,1),(2,0)\}$.

**Answer.** We like to organize:

| $k$ | $x_k$ | $y_k$ | $x_k^2$ | $y_k^2$ | $x_k y_k$ |
|---|---|---|---|---|---|
| 1 | $-1$ | 3 | 1 | 9 | $-3$ |
| 2 | 0 | 1 | 0 | 1 | 0 |
| 3 | 2 | 0 | 4 | 0 | 0 |
| $\text{sum}_k$ | 1 | 4 | 5 | 10 | $-3$ |

Using Proposition 3.3,

$$S_{\vec{x},\vec{x}} = \sum_k x_k^2 - \frac{1}{n}\left(\sum_k x_k\right)^2 = 5 - \frac{1}{3}(1)^2 = \frac{14}{3},$$

$$S_{\vec{y},\vec{y}} = \sum_k y_k^2 - \frac{1}{n}\left(\sum_k y_k\right)^2 = 10 - \frac{1}{3}(4)^2 = \frac{14}{3},$$

$$S_{\vec{x},\vec{y}} = \sum_k x_k y_k - \frac{1}{n}\left(\sum_k x_k\right)\left(\sum_k y_k\right) = -3 - \frac{1}{3}(1)(4) = -\frac{13}{3},$$

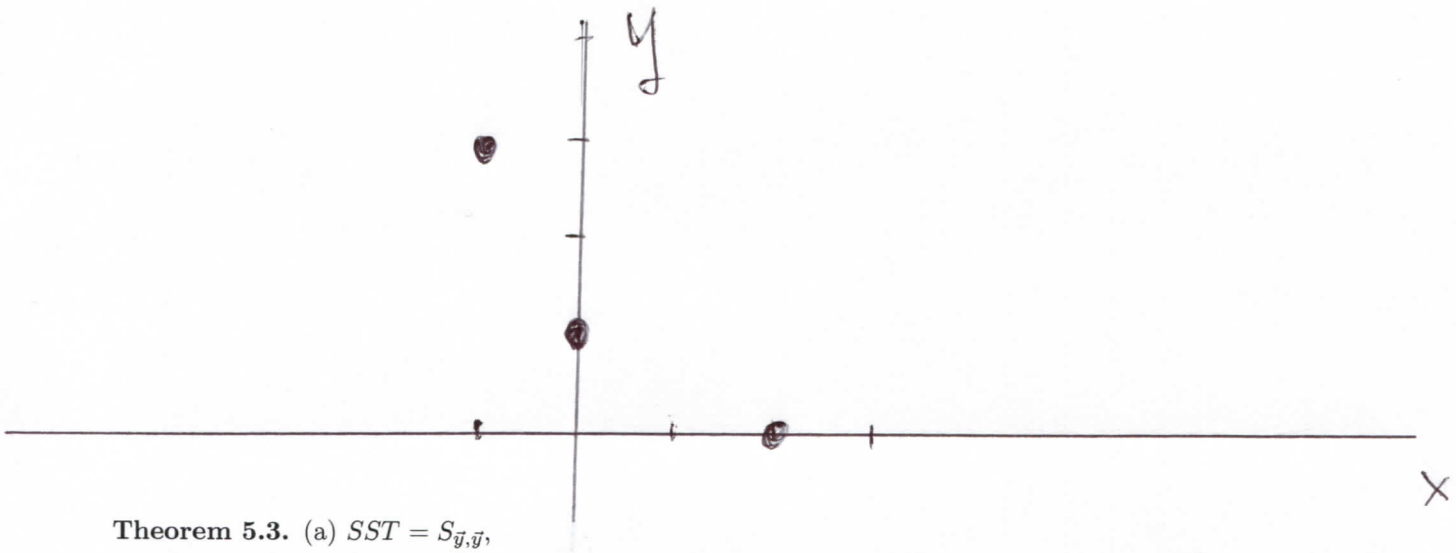
so that

$$r = \frac{-\frac{13}{3}}{\sqrt{(\frac{14}{3})(\frac{14}{3})}} = -\frac{13}{14} \sim -0.929$$

and

$$r^2 = \left(-\frac{13}{14}\right)^2 = \frac{169}{196} \sim 0.862.$$

The fact that $|r|$, hence $r^2$, is close to 1 reflects the fact that our data, drawn below, is very close to a straight line. Notice also that the $y$ values of the data decrease as the $x$ values increase. This is equivalent to $r$ being negative, which in turn is equivalent to the slope $\hat{\beta}_1$ of the least-squares line being negative (see 5.8 and 5.9).




**Theorem 5.3.** (a) $SST = S_{\vec{y},\vec{y}}$,

(b) $SSR = r^2 S_{\vec{y},\vec{y}}$, and

(c) $SSE = (1 - r^2)S_{\vec{y},\vec{y}}$.

**Proof:** (a) is merely Terminology 2.5 and Definitions 4.2.

Both (b) and (c) require some calculation, using Terminology 2.5, Theorem 3.1, Corollary 3.2, and Definitions 4.1 and 4.2.

$$SSR \equiv \sum_k (\hat{y_k} - \overline{y})^2 = \sum_k \Big( (\overline{y} + \hat{\beta}_1(x_k - \overline{x})) - \overline{y} \Big)^2 = \sum_k (\hat{\beta}_1(x_k - \overline{x}))^2 = (\hat{\beta}_1)^2 \sum_k (x_k - \overline{x})^2$$

$$= \left( \frac{S_{\vec{x},\vec{y}}}{S_{\vec{x},\vec{x}}} \right)^2 S_{\vec{x},\vec{x}} = \frac{(S_{\vec{x},\vec{y}})^2}{S_{\vec{x},\vec{x}}} = \left[ \frac{(S_{\vec{x},\vec{y}})^2}{S_{\vec{x},\vec{x}} S_{\vec{y},\vec{y}}} \right] S_{\vec{y},\vec{y}} = r^2 SST,$$

giving us (b).

$$SSE \equiv \sum_k (y_k - \hat{y_k})^2 = \sum_k \Big( y_k - (\overline{y} + \hat{\beta}_1(x_k - \overline{x})) \Big)^2 = \sum_k ((y_k - \overline{y}) - \hat{\beta}_1(x_k - \overline{x}))^2$$

$$= \sum_k (y_k - \overline{y})^2 - 2\hat{\beta}_1 \sum_k (y_k - \overline{y})(x_k - \overline{x}) + \hat{\beta}_1{}^2 \sum_k (x_k - \overline{x})^2 = S_{\vec{y},\vec{y}} - 2\hat{\beta}_1 S_{\vec{x},\vec{y}} + \beta_1^2 S_{\vec{x},\vec{x}}$$

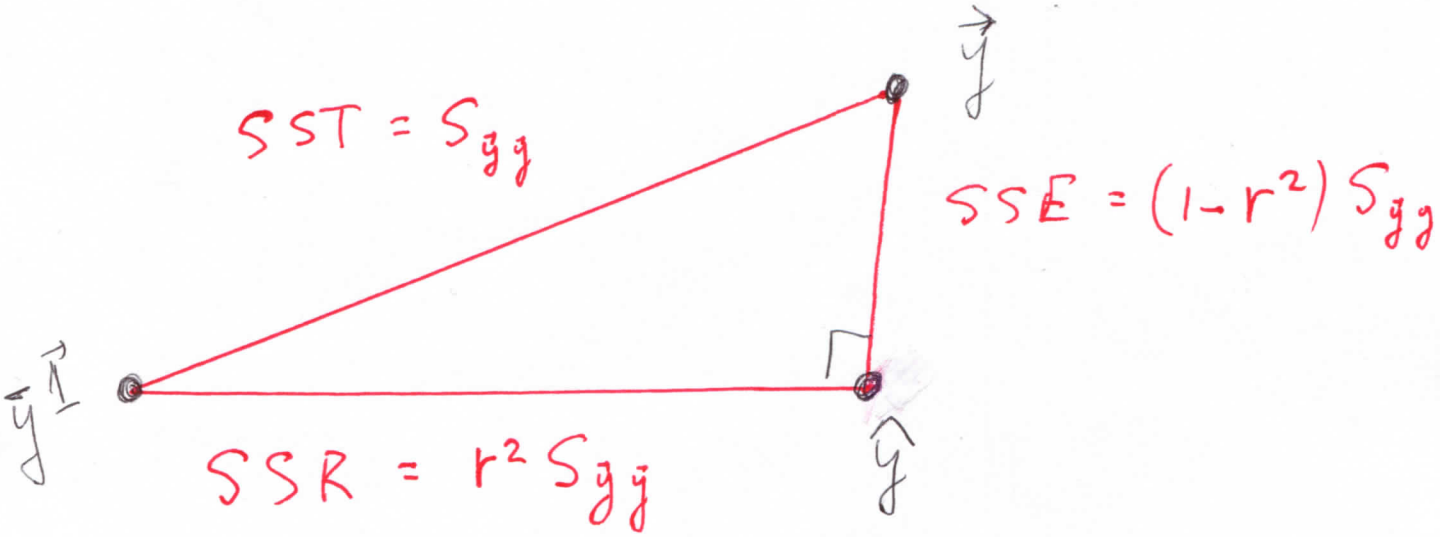
$$= S_{\vec{y},\vec{y}} - 2\left( \frac{S_{\vec{x},\vec{y}}}{S_{\vec{x},\vec{x}}} \right) S_{\vec{x},\vec{y}} + \left( \frac{S_{\vec{x},\vec{y}}}{S_{\vec{x},\vec{x}}} \right)^2 S_{\vec{x},\vec{x}} = S_{\vec{y},\vec{y}} - \frac{(S_{\vec{x},\vec{y}})^2}{S_{\vec{x},\vec{x}}} = S_{\vec{y},\vec{y}} - \left[ \frac{(S_{\vec{x},\vec{y}})^2}{S_{\vec{x},\vec{x}} S_{\vec{y},\vec{y}}} \right] S_{\vec{y},\vec{y}} = (1 - r^2) S_{\vec{y},\vec{y}} = (1 - r^2) SST,$$

giving us (c). □

In words, Theorem 5.3 is saying that $r^2 = \frac{SSR}{SST}$ measures the *proportion* of the observed variation in y explained by $x$ and the linear model while $(1-r^2) = \frac{SSE}{SST}$ measures the proportion of the observed variation in $y$ not explained by $x$ and the linear model

Compare this to $SSR$ and $SSE$ in Definitions 4.2.

**Regression Picture 5.4.** Here is a refinement of Picture 4.3, with $r^2$ inserted.




**Example 5.5.** Let's redo Example 4.4 using the sample correlation coefficient $r$.

| $k$ | $x_k$ | $y_k$ | $x_k^2$ | $x_k y_k$ | $y_k^2$ |
|---|---|---|---|---|---|
| 1 | 0 | 2 | 0 | 0 | 4 |
| 2 | 3 | 0 | 9 | 0 | 0 |
| 3 | 2 | 6 | 4 | 12 | 36 |
| 4 | 3 | −2 | 9 | −6 | 4 |
| $\text{sum}_k \equiv \sum_{k=1}^4$ | 8 | 6 | 22 | 6 | 44 |

As in Examples 3.4 and 4.4,

$$S_{\vec{x},\vec{y}} = -6, \quad S_{\vec{x},\vec{x}} = 6, \quad \text{and} \quad S_{\vec{y},\vec{y}} = 35,$$

thus

$$r = \frac{-6}{\sqrt{6 \times 35}} = -\frac{6}{\sqrt{210}} \to r^2 = \frac{36}{210} = \frac{6}{35}.$$

By Theorem 5.3,

$$SST = S_{\vec{y},\vec{y}} = 35, \quad SSR = r^2 S_{\vec{y},\vec{y}} = \frac{6}{35} \times 35 = 6 \text{ and } SSE = (1 - r^2) S_{\vec{y},\vec{y}} = (1 - \frac{6}{35})35 = 29.$$

**Terminology 5.6.** "How good a job is simple linear regression doing explaining the variation of $y$ in the data in Examples 3.4?"

This sort of goodness is measured by $r$ or $r^2$; our answer to the quoted question could be "The coefficient of determination is $\frac{6}{35}$."

For those who prefer subjective social fuzziness to explicit quantification, the answer to our question might be "Bad job, since $r^2$ is too small."

**Example 5.7.** For the data in Example 3.5:

(a) Get $r^2$, the coefficient of determination, measuring the proportion of observed variation in the number of decades a bridge lasts that can be explained by its linear relationship with the temperature during construction.

(b) Get the sample correlation coefficient.

(c) Get $SST, SSE$, and $SSR$.

(d) Get $s^2$, our favorite estimator of $\sigma^2$, and $s$, our favorite estimator of $\sigma$, in Definition 1.2.

**Answers.** We need, in addition to the calculations in 3.5,

$$S_{\vec{y},\vec{y}} = \sum_k y_k^2 - \frac{1}{n}(\sum_k y_k)^2 = 57 - \frac{1}{20}(30)^2 = 12.$$

(a) By Definitions 5.1,

$$r^2 = \frac{6^2}{(4.8)(12)} = 0.625.$$

(b)

$$r = \sqrt{0.625} \sim 0.791.$$

(c) $SST = S_{\vec{y},\vec{y}} = 12, SSE = (1-r^2)S_{\vec{y},\vec{y}} = (1-0.625)(12) = 4.5, SSR = r^2 S_{\vec{y},\vec{y}} = (0.625)(12) = 7.5$. (We could have saved work by subtracting 4.5 from 12.)

(d) See Definition 4.5.

$$s^2 = \frac{SSE}{(n-2)} = \frac{4.5}{18} = 0.25,$$

thus

$$s = \sqrt{0.25} = 0.5.$$

**Proposition 5.8: Some Properties of the Sample Correlation Coefficient.**
(a) $-1 \leq r \leq 1$.

(b) $r = \pm 1 \iff$ bivariate data is on a straight line.

(c) $\hat{\beta}_1 = 0 \iff r = 0$.

(d) $\hat{\beta}_1 > 0 \iff r > 0$.

(e) $\hat{\beta}_1 < 0 \iff r < 0$.

**Proof:** Parts (c), (d), and (e) follow from the fact, that we leave to the reader to calculate, that

$$\hat{\beta}_1 = r\sqrt{\frac{S_{\vec{y},\vec{y}}}{S_{\vec{x},\vec{x}}}}.$$

Part (a) is a special case of what is known in linear algebra as the *Cauchy inequality*; see [2, 6.26, page 428], using the terminology *dot product* (APP.4 in the Appendix of this Magnification) and *norm* (APP.1 in the Appendix of this Magnification).

For part (b), first suppose $r = \pm 1$. By Theorem 5.3, we then have

$$\sum_{k=1}^{n}(y_k - \hat{y_k})^2 \equiv SSE = 0.$$

Since each term in the sum just stated is nonnegative, each term must be zero; that is,

$$y_k = \hat{y_k}, \quad k = 1, 2, 3, \ldots, n,$$

so that $\{(x_k, y_k) \,|\, k = 1, 2, 3, \ldots, n\} = \{(x_k, \hat{y_k}) \,|\, k = 1, 2, 3, \ldots, n\}$, which lies on the least-squares line $y = \hat{\beta}_0 + \hat{\beta}_1 x$.

Conversely, suppose the bivariate data lies on a line $y = b_0 + b_1 x$, for some real $b_0$ and $b_1$. Then, for $k = 1, 2, 3, \ldots, n$, since

$$\overline{y} \equiv \frac{1}{n}\sum_{k=1}^{n} y_k = \frac{1}{n}\left[\sum_{k=1}^{n}(b_0 + b_1 x_k)\right] = \frac{1}{n}\left[nb_0 + b_1 \sum_{k=1}^{n} x_k\right] = b_0 + b_1\overline{x},$$

we have

$$(y_k - \overline{y}) = ((b_0 + b_1 x_k) - (b_0 + b_1\overline{x})) = b_1(x_k - \overline{x}),$$

thus

$$S_{\vec{x},\vec{y}} \equiv \sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y}) = b_1 \sum_{k=1}^{n}(x_k - \overline{x})^2 \equiv b_1 S_{\vec{x},\vec{x}}$$

and

$$S_{\vec{y},\vec{y}} \equiv \sum_{k=1}^{n}(y_k - \overline{y})^2 = b_1^2 \sum_{k=1}^{n}(x_k - \overline{x})^2 \equiv b_1^2 S_{\vec{x},\vec{x}},$$

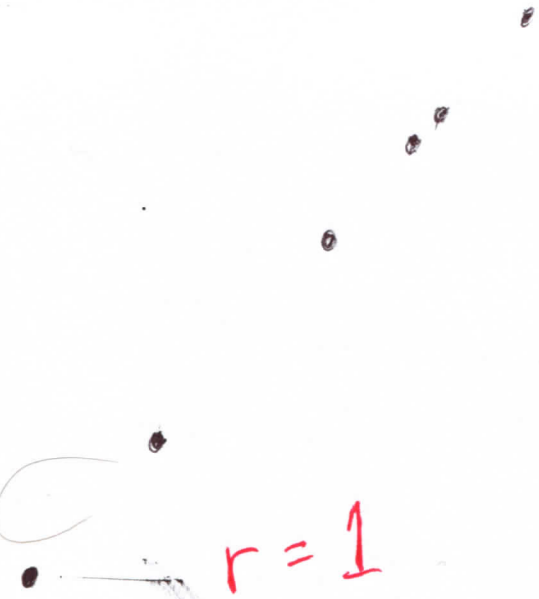
thus

$$r^2 \equiv \frac{(S_{\vec{x},\vec{y}})^2}{(S_{\vec{x},\vec{x}})(S_{\vec{y},\vec{y}})} = \frac{(b_1 S_{\vec{x},\vec{x}})^2}{(S_{\vec{x},\vec{x}})b_1^2(S_{\vec{x},\vec{x}})} = 1,$$

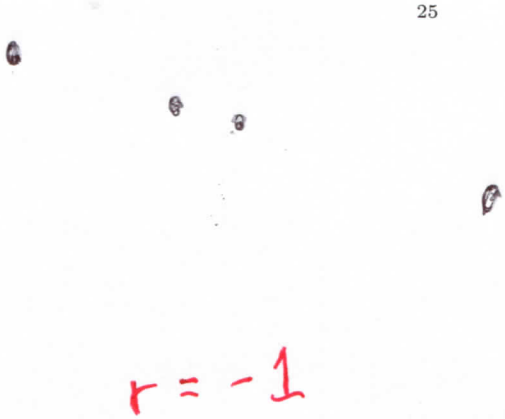
so that $r = \pm 1$. □

**Correlation Pictures and Discussion 5.9.** The proof of Proposition 5.8(b) implies that, as $|r|$ gets close to one, $(y_k - \hat{y_k})$ should get close to zero, for $k = 1, 2, \ldots, n$. In some sense, this is saying that the graph of the bivariate data, called a *scatterplot*, is getting close to a straight line, just as $|r| = 1$ implies that the scatterplot will be exactly a straight line.

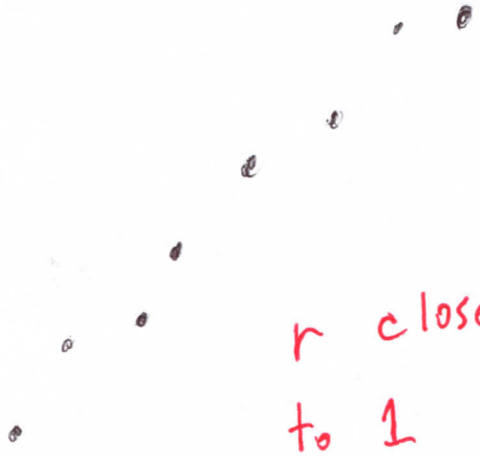
The next page presents some scatterplots, with information about $r$ from Proposition 5.8 attached.



$r = 1$



$r = -1$



r close to 1

r close to $-1$

r close to 0

r close to 0

We should mention that the intuition of the previous page is not foolproof. Consider the following two sets of bivariate data, each containing only three ordered pairs.

**Examples 5.10.(a)** For arbitrary nonzero $\epsilon$, let our bivariate data be

$$\{(-1,0),(0,3\epsilon),(1,0)\}.$$

We leave it to the reader to calculate (Proposition 3.3)

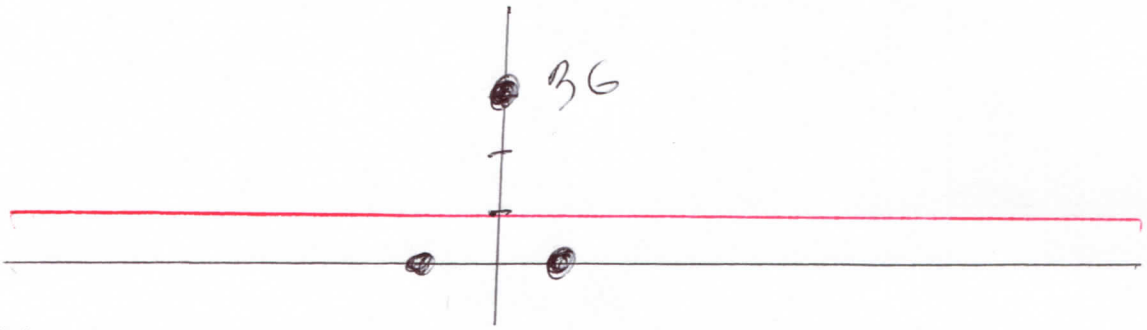
$$S_{\vec{x},\vec{y}}=0,\ \ S_{\vec{y},\vec{y}}=6\epsilon^2,\ \ S_{\vec{x},\vec{x}}=2.$$

Then (see Theorem 3.1) $\hat{\beta}_1=0$ and $\hat{\beta}_0=\epsilon$, thus our least-squares line is

$$y=\epsilon.$$

More calculation shows that $r=0=SSR$ and $SST=SSE=6\epsilon^2$.

Despite $r$ being 0, the scatterplot below (with least-squares line drawn in red) looks arbitrarily close to a straight (horizontal) line as $|\epsilon|$ gets smaller.




**(b)** Having $r=0$ is pathological (see Definition 6.4, Proposition 5.8(c) and Remark 6.7), so let's have an example with nonzero $r$.

For arbitrary nonzero $\epsilon$, let our bivariate data be

$$\{(-3,0),(1,3\epsilon),(2,0)\}.$$

Again we leave it to the reader to calculate

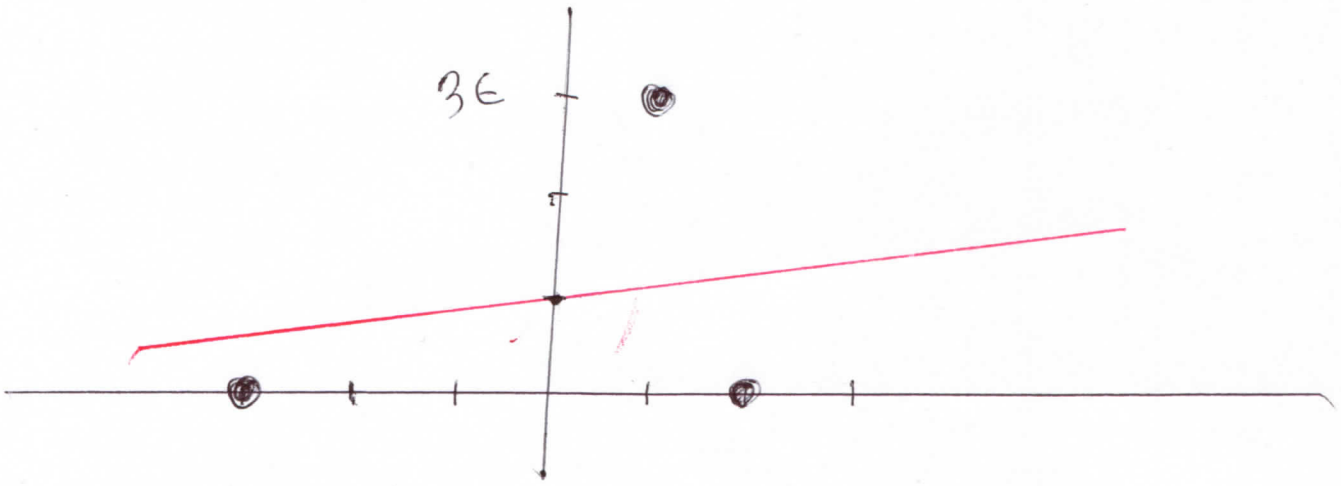
$$S_{\vec{x},\vec{y}}=3\epsilon,\ \ S_{\vec{y},\vec{y}}=6\epsilon^2,\ \ S_{\vec{x},\vec{x}}=14,\ \ \hat{\beta}_1=\frac{3\epsilon}{14},\ \text{ and }\ \hat{\beta}_0=\epsilon,$$

thus our least-squares line is

$$y=\epsilon+(\frac{3\epsilon}{14})x=\epsilon\left[1+\frac{3}{14}x\right].$$

We further leave it to the reader to calculate that $r$ equals $\frac{3}{\sqrt{84}}$ if $\epsilon>0$ and $r$ equals $-\frac{3}{\sqrt{84}}$ if $\epsilon<0$.

Below we have drawn a scatterplot, with the least-squares line drawn in red. As with (a), as $|\epsilon|$ gets close to zero, the scatterplot looks more like a straight line, yet $|r|$ does not get close to one, in fact, $|r|$ does not change as $\epsilon$ changes.



## Chapter VI. Inference on $\beta_1$, the slope of the true regression line

Assume, throughout this and the next chapter, that we are under the Simple Linear Regression Model Definition 1.2. It is time to discuss confidence intervals and hypothesis tests, involving the parameter $\beta_1$; see Definition 1.2.

**Recent History 6.1.** To motivate our activities, let's quickly summarize such inferences for a simpler parameter, the *population mean*, denoted $\mu$; see [5].

For $X$ a normal random variable with (unknown) mean $\mu$, our estimator is much less mysterious than the subject of this magnification: we estimate the population mean $\mu$ with the *sample mean* $\hat{\mu} \equiv \overline{X}$.

If the population standard deviation, call it $\sigma_X$, is known, then, denoting by $n$ the sample size,

$$\frac{\overline{X} - \mu}{\frac{\sigma_X}{\sqrt{n}}}$$

is an excellent choice for a test statistic, since it has a standard normal, denoted $Z$, distribution.

In practice, we do *not* know what $\sigma_X$ is, thus we must estimate it with the *sample standard deviation*, denoted $S_X$; see [5, Definitions 1.1]. Then our natural test statistic is

$$\frac{\overline{X} - \mu}{\frac{S_X}{\sqrt{n}}} \quad (*).$$

For $n \leq 40$, our estimator (*) no longer has a $Z$ distribution. Said estimator has what is called a *$t$ distribution, with $(n-1)$ degrees of freedom,* or *$t_{n-1}$ distribution.* A random variable with this distribution is denoted $T_{n-1}$ or sometimes $T$ for short; measurements of said random variable are denoted $t_{n-1}$ or $t$ for short.

See [5, especially 1.7–1.11] for needed properties of $t$ distributions.

Our choice (*) of test statistic drives both hypothesis testing and confidence intervals.

For the null hypothesis

$$H_0 : \mu = \mu_0,$$

we use the test statistic $\frac{\overline{X} - \mu_0}{\frac{S_X}{\sqrt{n}}}$, to either calculate P-values or to get critical values for rejection regions.

Our $(1-\alpha)100\%$ confidence intervals for $\mu$ are

$$\overline{x} \pm t_{\frac{\alpha}{2},(n-1)} \left( \frac{s_X}{\sqrt{n}} \right),$$

formed by setting (*) between $-t_{\frac{\alpha}{2},(n-1)}$ and $t_{\frac{\alpha}{2},(n-1)}$.

Note that the confidence interval has three ingredients: reading from left to right, these are the *estimate* $\overline{x}$, the *critical value* $t_{\frac{\alpha}{2},(n-1)}$ and the *estimated standard error* for $\overline{X}$, $\left( \frac{s_X}{\sqrt{n}} \right)$.

**Analogies 6.2.** Recall (Theorem 3.1) that our estimator of $\beta_1$ is

$$\hat{\beta}_1 \equiv \frac{S_{\vec{x},\vec{y}}}{S_{\vec{x},\vec{x}}}.$$

It can be shown that the expected value of $\hat{\beta}_1$ is $\beta_1$; that is, $\hat{\beta}_1$ is an *unbiased* estimator.

It can also be shown that the standard deviation of $\hat{\beta}_1$ is $\frac{\sigma}{\sqrt{S_{\vec{x},\vec{x}}}}$, where $\sigma$ is from the Simple Linear Regression Model Definition 1.2. The normality of our model now implies that

$$\frac{(\hat{\beta}_1 - \beta_1)}{\frac{\sigma}{\sqrt{S_{\vec{x},\vec{x}}}}}$$

has a $Z$ distribution.

As with inference on $\mu$, it is not realistic to assume knowledge of $\sigma$, thus we replace it with our favorite estimator (see Definitions 4.5),

$$s \equiv \hat{\sigma} \equiv \sqrt{\frac{SSE}{(n-2)}},$$

giving us the *estimated standard error* for $\hat{\beta}_1$

$$s_{\hat{\beta}_1} \equiv \frac{s}{\sqrt{S_{\vec{x},\vec{x}}}};$$

hence, analogously to (*) in 6.1, our favorite test statistic for $\beta_1$ is

$$\frac{(\hat{\beta}_1 - \beta_1)}{s_{\hat{\beta}_1}} = \frac{(\hat{\beta}_1 - \beta_1)}{\frac{s}{\sqrt{S_{\vec{x},\vec{x}}}}} \quad (**).$$

This turns out to have a $t_{n-2}$ distribution.

As with inference on $\mu$, the test statistic dictates both hypothesis testing and confidence intervals.

**(1)** For testing the null hypothesis

$$H_0 : \beta_1 = (\beta_1)_0,$$

we use the test statistic

$$T \equiv \frac{(\hat{\beta}_1 - (\beta_1)_0)}{s_{\hat{\beta}_1}},$$

with a $t_{n-2}$ distribution, to get P-values or critical values for rejection regions, as in [5].

**(2)** Our $(1-\alpha)100\%$ confidence interval for $\beta_1$ is

$$\hat{\beta}_1 \pm (t_{\frac{\alpha}{2},(n-2)}) s_{\hat{\beta}_1}.$$

Notice that we again have, from left to right, estimate, critical value and estimated standard error.

**Example 6.3.** Here is some bivariate data, for the Simple Linear Regression Model Definition 1.2.

| | | | | | |
|---|---|---|---|---|---|
| $x$ | 0 | 1 | 2 | 4 | 5 |
| $y$ | 4 | 2 | 0 | 0 | $-1$ |

(a) Get a 95% confidence interval for $\beta_1$.

**Answer.** We need $S_{\vec{x},\vec{y}}$ and $S_{\vec{x},\vec{x}}$ for $\hat{\beta}_1$. For $s_{\hat{\beta}_1}$, we need $SSE = (1-r^2)S_{\vec{y},\vec{y}}$ (Theorem 5.3).

Using the computational formula Proposition 3.3, we need

$$\sum_k x_k = 12, \sum_k y_k = 5, \sum_k x_k^2 = 46, \sum_k y_k^2 = 21, \text{and} \sum_k x_k y_k = (-3),$$

to obtain

$$S_{\vec{x},\vec{x}} = 46 - \frac{1}{5}(12)^2 = 17.2, \quad S_{\vec{y},\vec{y}} = 21 - \frac{1}{5}(5)^2 = 16, \quad S_{\vec{x},\vec{y}} = (-3) - \frac{1}{5}(12)(5) = -15,$$

giving us

$$\hat{\beta}_1 = \frac{S_{\vec{x},\vec{y}}}{S_{\vec{x},\vec{x}}} = \frac{-15}{17.2} \sim -0.872, r^2 \equiv \frac{S_{\vec{x},\vec{y}}^2}{S_{\vec{x},\vec{x}} S_{\vec{y},\vec{y}}} \sim 0.818, SSE \sim (1-0.818)16 \sim 2.92$$

and

$$s^2 = \frac{SSE}{(n-2)} \sim \frac{2.92}{(5-2)} \sim 0.97, s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{\vec{x},\vec{x}}}} \sim \sqrt{\frac{0.97}{17.2}} \sim 0.237.$$

We also need a critical value. Since $(1-\alpha) = 0.95$, $\frac{\alpha}{2} = 0.025$. From the "Critical Values" table near the end of this Magnification, prior to the References,

$$t_{\frac{\alpha}{2},n-2} = t_{0.025,3} = 3.182,$$

so, following **6.2(2)**, our confidence interval is (approximately)

$$-0.872 \pm (3.182)(0.237) \sim -0.872 \pm 0.754 = (-1.626, -0.118).$$

(b) Test $H_0 : \beta_1 = -1$ versus $H_a : \beta_1 > -1$ at significance level $\alpha = 0.1$.

**Answer.** Following 6.2(1) and using the calculations from (a), our numerical test statistic is

$$t = \frac{\hat{\beta}_1 - (-1)}{s_{\hat{\beta}_1}} \sim \frac{-0.872 - (-1)}{0.237} \sim 0.54,$$

so that, from the "t Curve Tail Areas" table near the end of this Magnification, prior to the References,

$$\text{P-value} \sim P(T_3 > 0.54) = 0.326 > 0.1 = \alpha,$$

so that we don't reject $H_0$; at significance level 0.1, the data does not support the slope of our true regression line being more than $(-1)$.

(c) Is there a "useful linear relationship" between $x$ and $y$?

**Answer.** See Definition 6.4. Our question translates into the hypothesis test

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_a : \beta_1 \neq 0.$$

More precisely, an answer of "yes" to the question in (c) is equivalent to rejecting $H_0$.

We need a P-value. Our numerical test statistic is now

$$t = \frac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}} \sim \frac{-0.872}{0.237} \sim -3.68,$$

so that

$$\text{P-value} \sim P(T_3 > 3.68) + P(T_3 < -3.68) = 2P(T_3 > 3.68) \sim 2P(T_3 > 3.7) = 2(0.017) = 0.034.$$

Since no significance level $\alpha$ is given, we should choose one of the two arbitrarily popular significance levels, $\alpha = 0.01$ or $\alpha = 0.05$.

Since P-value $\sim 0.034 \leq 0.05$, we reject (at significance level 0.05) $H_0$; at significance level 0.05, the data suggests there is a useful linear relationship between $x$ and $y$.

Since P-value $\sim 0.034 > 0.01$, we don't reject (at significance level 0.01) $H_0$; at significance level 0.01, the data does not suggest there is a useful linear relationship between $x$ and $y$.

It can be shown that the rejection of $H_0$ at significance level 0.05 is equivalent to the fact that our 95% confidence interval, from (a), does not contain 0. See [4, Theorem 5.2(c)].

**Definition 6.4.** A **useful linear relationship** between $x$ and $y$ satisfying the Simple Linear Regression Model in Definition 1.2 means that $\beta_1 \neq 0$.

"Useful" here means in terms of using $x$ to predict $y$. If $\beta_1 = 0$, then our true regression line $y = \beta_0 + \beta_1 x$ becomes $y = \beta_0$, a horizontal line. This means the value of $y$ is unaffected by the value of $x$; knowing $x$ no longer helps us to know $y$.

**Example 6.5.** See 3.5 and 5.7.

(a) Get a 99 percent confidence interval for $\beta_1$.

(b) Test, at significance level $\alpha = 0.05$, whether there is a useful linear relationship between the number of decades a bridge lasts, and the average temperature (during construction).

(c) Test, at significance level 0.05, the belief that increasing the temperature increases the lifetime of bridges (on average).

(d) Test, at significance level 0.01, the claim that increasing the temperature by one degree increases the lifetime of a bridge by more than 5 years, on average.

(e) Test, at significance level 0.01, the claim that increasing the temperature by one degree increases the lifetime of a bridge by more than 8 years, on average.

**Answers.** (a) See **6.2(2)**. Setting $(1-\alpha) = 0.99$ gives us $\frac{\alpha}{2} = 0.005$, so that

$$t_{\frac{\alpha}{2},(n-2)} = t_{0.005,18} = 2.878.$$

We also need

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{\vec{x},\vec{x}}}} = \frac{0.5}{\sqrt{4.8}},$$

so that our confidence interval is

$$\hat{\beta}_1 \pm 2.878(\frac{0.5}{\sqrt{4.8}}) \sim 1.25 \pm 0.66 = (0.59, 1.91).$$

(b) This is (see 6.4) testing

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_a : \beta_1 \neq 0$$

with test statistic

$$t = \frac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}} = \frac{1.25}{\frac{0.5}{\sqrt{4.8}}} \sim 5.48,$$

with a $t_{18}$ distribution, hence a P-value of $\sim$

$$2P(T_{18} > 5.48) \leq 2P(T_{18} > 3.9) = 2(0.001) = 0.002 \leq 0.05 = \alpha,$$

thus we reject $H_0$: the data, at significance level 0.05, suggests a useful linear relationship between the average temperature and the number of decades a bridge lasts.

(c) This is (compare to (b)) testing

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_a : \beta_1 > 0.$$

We have the same test statistic as in (b), but the P-value is now

$$P(T_{18} > 5.48) \leq P(T_{18} > 3.9) = 0.001 \leq 0.05 = \alpha,$$

thus we reject $H_0$: the data, at significance level 0.05, suggests the belief that increasing the temperature increases the lifetime of bridges (on average).

(d) This is

$$H_0 : \beta_1 = 0.5 \quad \text{versus} \quad H_a : \beta_1 > 0.5.$$

Our test statistic becomes

$$t = \frac{\hat{\beta}_1 - 0.5}{s_{\hat{\beta}_1}} = \frac{0.75}{\frac{0.5}{\sqrt{4.8}}} \sim 3.3,$$

with a $t_{18}$ distribution, hence a P-value of $\sim$

$$P(T_{18} > 3.3) = 0.002 \leq 0.01 = \alpha,$$

thus we reject $H_0$, and conclude that increasing the temperature by one degree increases the lifetime of a bridge by more than 5 years, on average, at significance level 0.01.

(e) This is

$$H_0 : \beta_1 = 0.8 \quad \text{versus} \quad H_a : \beta_1 > 0.8.$$

Argue almost identically to (d):

Our test statistic becomes

$$t = \frac{\hat{\beta}_1 - 0.8}{s_{\hat{\beta}_1}} = \frac{0.45}{\frac{0.5}{\sqrt{4.8}}} \sim 2.0,$$

with a $t_{18}$ distribution, hence a P-value of $\sim$

$$P(T_{18} > 2.0) = 0.030 > 0.01 = \alpha,$$

thus we don't reject $H_0$, and conclude that, at significance level 0.01, there is insufficient evidence to conclude that increasing the temperature by one degree increases the lifetime of a bridge by more than 8 years, on average.

**Remarks 6.6.** The test statistic we have been using for

$$H_0 : \beta_1 = 0 \quad \text{is} \quad \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}},$$

which has a $t_{n-2}$ distribution.

Another approach is to square the test statistic just mentioned. We leave it to the reader to show that

$$\left(\frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}\right)^2 = \frac{\frac{SSR}{1}}{\frac{SSE}{(n-2)}}.$$

This test statistic has what is called an $F$ distribution; the 1 and $(n-2)$ are *degrees of freedom.* Notice that, in 4.3, said test statistic is the ratio of the lengths of the legs of the right triangle, tempered by degrees of freedom. See Pictures 9.7 and, in the Appendix, APP.13, for more of this picture. See also 9.10 for other examples of where the $F$ distribution appears.

**Remark 6.7.** For any bivariate data as in Assumptions 2.1 with $\hat{\beta}_1 = 0$, said data will not suggest a useful linear relationship between $x$ and $y$, at any significance level, since the test statistic for

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_a : \beta_1 \neq 0$$

is then

$$t = \frac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}} = 0,$$

so that our P-value is

$$2P(T_{n-2} > 0) = 1.$$

## Chapter VII. Inference on $\beta_0$ and other $y$ values of the true regression line

In the last chapter we performed inference, meaning confidence intervals and hypothesis tests, on the parameter $\beta_1$ in the Simple Linear Regression Model Definition 1.2. In this section, we will do the same with the parameter $\beta_0$.

We can do much more, with no extra effort, by observing that

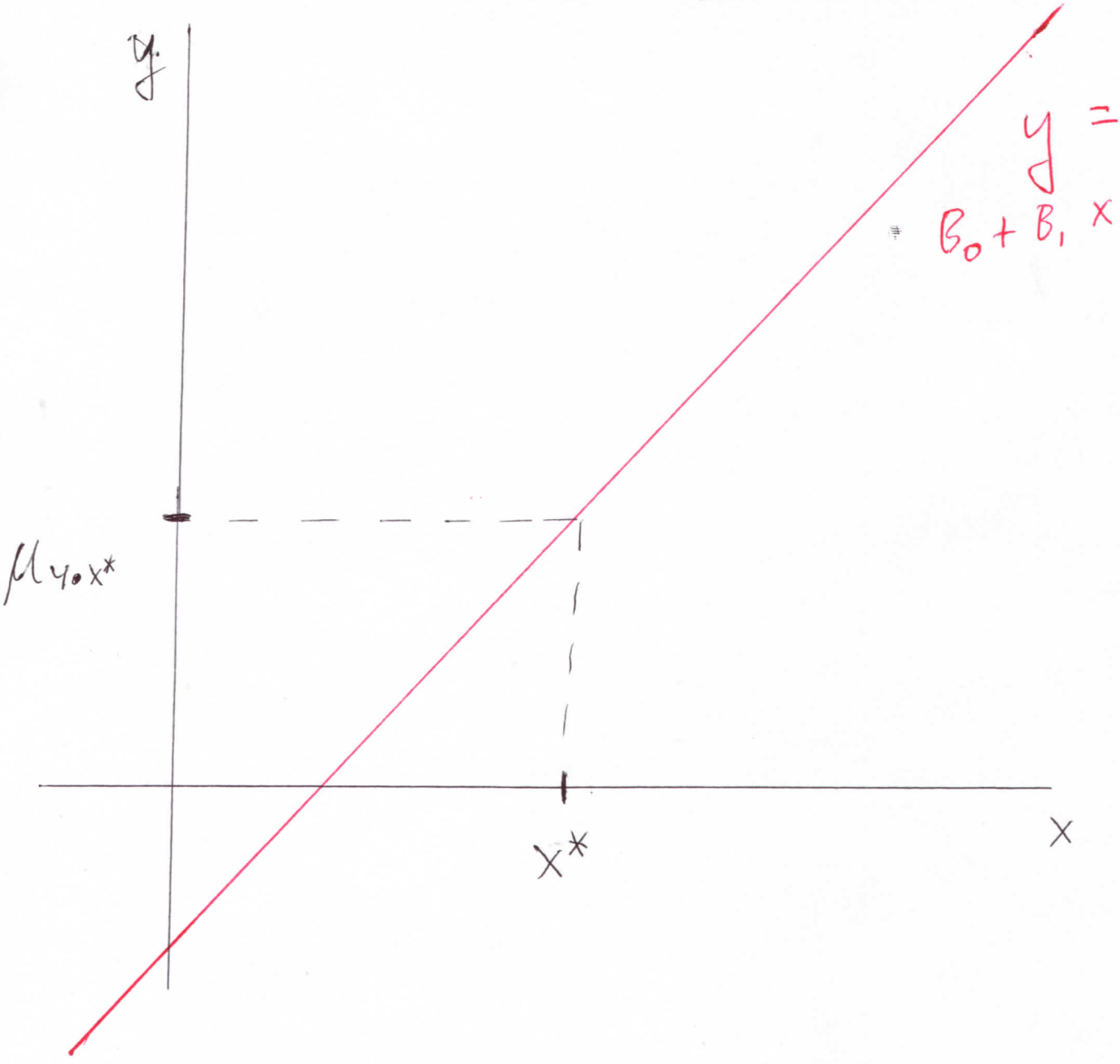
$$\beta_0 = E(Y|x = 0),$$

the expected value of $Y$ when $x = 0$.

**Definition 7.1.** For any real $x^*$, define the parameter

$$\mu_{Y \cdot x^*} \equiv (\beta_0 + \beta_1 x^*) = E(Y|x = x^*),$$

the expected value of $Y$ when $x = x^*$. This is the $y$ value of the point on the true regression line when the $x$ value is $x^*$. See the drawing below, with the true regression line drawn in red.



It seems natural to use our estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ (see 2.3 and 3.1) to estimate $\mu_{Y \cdot x^*}$.

**Definitions 7.2.** Our favorite estimator of $\mu_{Y \cdot x^*}$ is

$$\hat{\mu}_{Y \cdot x^*} \equiv \hat{\beta}_0 + \hat{\beta}_1 x^*;$$

this will often be abbreviated to $\hat{Y}$.

It can be shown that $\hat{Y}$ is normal, with expected value $E(\hat{Y}) = \mu_{Y \cdot x^*}$ and standard deviation $\sigma\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{\vec{x},\vec{x}}}}$.

As with 6.2, we replace $\sigma$ with $s$ (Definition 4.5) and define the *estimated standard error for* $\hat{Y}$ to be

$$s_{\hat{Y}} \equiv s\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{\vec{x},\vec{x}}}}.$$

Again as in 6.2, our favorite test statistic for $\mu_{Y \cdot x^*}$,

$$\frac{\hat{Y} - \mu_{Y \cdot x^*}}{s_{\hat{Y}}},$$

has a $t_{n-2}$ distribution, leading to the following outline of inference on $\mu_{Y \cdot x^*}$.

**(1)** For testing the null hypothesis

$$H_0 : \mu_{Y \cdot x^*} = (\mu_{Y \cdot x^*})_0,$$

we use the test statistic

$$T \equiv \frac{(\hat{Y} - (\mu_{Y \cdot x^*})_0)}{s_{\hat{Y}}}$$

with a $t_{n-2}$ distribution, to get P-values or critical values for rejection regions, as in [5].

**(2)** Our $(1 - \alpha)100\%$ confidence interval for $\mu_{Y \cdot x^*}$ is

$$\hat{Y} \pm (t_{\frac{\alpha}{2},(n-2)}) s_{\hat{Y}}.$$

**Example 7.3.** We will use the data and calculations in Example 6.3.

(a) Get a 95% confidence interval for $(\beta_0 + 3\beta_1)$.

**Answer.** We already (in 6.3) calculated $\hat{\beta}_1 \sim -0.872$. For $\hat{\beta}_0$, we need $\overline{y} = 1$ and $\overline{x} = 2.4$, thus

$$\hat{\beta}_0 \sim 1 - (2.4)(-0.872) \sim 3.093.$$

Here $x^* = 3$, so

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1(3) \sim 3.093 + (-0.872)(3) = 0.477$$

is our estimate of $(\beta_0 + 3\beta_1)$.

Next let's get the estimated standard error

$$s_{\hat{Y}} \equiv s\sqrt{\frac{1}{n}+\frac{(x^*-\overline{x})^2}{S_{\vec{x},\vec{x}}}} = \sqrt{0.97}\sqrt{\frac{1}{5}+\frac{(3-2.4)^2}{17.2}} \sim 0.463.$$

Finally, we need the critical value

$$t_{\frac{\alpha}{2}} = t_{0.025,3} = 3.182,$$

from 6.3.

From **7.2(2)** our $(1-\alpha)100\%$ confidence interval for $\mu_{Y\cdot x^*}$ is

$$\hat{Y} \pm (t_{\frac{\alpha}{2},(n-2)})s_{\hat{Y}} \sim 0.477 \pm (3.182)(0.463) \sim 0.477 \pm 1.473 = (-0.996, 1.950).$$

(b) Get a 95% confidence interval for $(\beta_0+\beta_1)$.

**Answer.** This is the same as (a), except that $x^* = 1$.

Now

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1(1) \sim 3.093 + (-0.872)(1) = 2.221$$

is our estimate of $(\beta_0+\beta_1)$.

We also modify

$$s_{\hat{Y}} \sim \sqrt{0.97}\sqrt{\frac{1}{5}+\frac{(1-2.4)^2}{17.2}} \sim 0.552,$$

thus our confidence interval is $\sim$

$$(2.221 \pm (3.182)(0.552) \sim 2.221 \pm 1.756 = (0.465, 3.997).$$

Notice that the confidence interval for $(\beta_0+\beta_1)$ is wider than the confidence interval for $(\beta+3\beta_1)$.

This is explained, at least technically, by the presence of $(x^*-\overline{x})^2$ in the expression for $s_{\hat{Y}}$: $x^* = 3$ is closer to $\overline{x} = 2.4$ than $x^* = 1$, thus $(x^*-\overline{x})^2$ is smaller when $x^* = 3$.

A more intuitive explanation for the widening of the confidence interval when $x^*$ is further away from $\overline{x}$ is that, in a sense that can be made precise, $x^*$ is further away from the data (see APP.15 in the Appendix), thus we have less information about quantities calculated from $x^*$; loss of information is equivalent to wider confidence intervals.

(c) Test the claim that $(\beta_0+3\beta_1) > 0$, at significance level 0.05.

**Answer.** Our hypothesis test is

$$H_0 : \mu_{Y\cdot 3} \equiv (\beta_0+3\beta_1) = 0 \text{ versus } H_a : \mu_{Y\cdot 3} \equiv (\beta_0+3\beta_1) > 0.$$

We follow **7.2(1)** with $x^* = 3, \mu_{Y\cdot 3} = 0$, $\hat{Y} \sim 0.477$ and $s_{\hat{Y}} \sim 0.463$ already calculated. Our test statistic is

$$t = \frac{\hat{Y}-0}{s_{\hat{Y}}} \sim \frac{0.477}{0.463} \sim 1.0,$$

thus our P-value is $\sim$

$$P(T_3 > 1.0) = 0.196 > 0.05 = \alpha,$$

thus we do not reject $H_0$; at significance level 0.05, the data does not suggest $(\beta_0+3\beta_1)$ is greater than 0.

**Example 7.4.** See 3.5, 5.7, and 6.5.

(a) Get a 95 percent confidence interval for $(\beta_0 + 3\beta_1)$, the true average number of decades that a bridge lasts at an average temperature (during construction) of 3 degrees Celsius.

(b) Test, at significance level $\alpha = 0.01$, the hypothesis that $(\beta_0 - \beta_1)$, the true average number of decades that a bridge lasts, at an average temperature (during construction) of $(-1)$ degrees Celsius, is less than 0.6.

**Answer.** (a) See **7.2(2)**. Here $x^* = 3$ and $\overline{x} = -0.1$. Our interval is

$$(\hat{\beta}_0 + \hat{\beta}_1(3)) \pm t_{0.025,18}(s)\sqrt{\frac{1}{n} + \frac{(3-(-0.1))^2}{S_{\vec{x},\vec{x}}}} \sim 1.625 + (1.25)(3) \pm 2.101(0.5)\sqrt{\frac{1}{20} + \frac{(3.1)^2}{4.8}}$$
$$\sim 5.375 \pm 1.505 = (3.87, 6.88).$$

(b) Our hypothesis test is

$$H_0 : (\beta_0 - \beta_1) = 0.6 \quad \text{versus} \quad H_a : (\beta_0 - \beta_1) < 0.6$$

with $x^* = -1$. We have

$$\hat{Y} \equiv (\hat{\beta}_0 - \hat{\beta}_1) = 0.375$$

and (see **7.2(1)**)

$$s_{\hat{Y}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{\vec{x},\vec{x}}}} = 0.5\sqrt{\frac{1}{20} + \frac{(-1-(-0.1))^2}{4.8}} \sim 0.234,$$

so that our test statistic is

$$t = \frac{\hat{Y} - 0.6}{s_{\hat{Y}}} \sim \frac{0.375 - 0.6}{0.234} \sim -1.0,$$

giving us a P-value of

$$P(T_{18} < -1.0) = P(T_{18} > 1.0) = 0.165 > 0.01,$$

thus we don't reject $H_0$; at significance level 0.01, the data does not suggest that the bridge lasts (on average) less than six years, at an average temperature (during construction) of $(-1)$ degrees.

## Chapter VIII. Recommended formulas for simple linear regression

**Terminology.** For any pair of ordered $n$-tuples $\vec{w} \equiv (w_1, w_2, \ldots, w_n), \vec{z} \equiv (z_1, z_2, \ldots, z_n)$, define

$$S_{\vec{w},\vec{z}} \equiv \sum_{k=1}^{n} (w_k - \overline{w})(z_k - \overline{z}), \qquad \text{where} \quad \overline{w} \equiv \frac{1}{n}\sum_{k=1}^{n} w_k \quad \text{and} \quad \overline{z} \equiv \frac{1}{n}\sum_{k=1}^{n} z_k.$$

The "computational formula" is

$$S_{\vec{w},\vec{z}} = \sum_{k=1}^{n} w_k z_k - \frac{1}{n}\left(\sum_{k=1}^{n} w_k\right)\left(\sum_{k=1}^{n} z_k\right).$$

All formulas refer to Assumptions 2.1.

$\hat{\beta}_1 = \dfrac{S_{\vec{x},\vec{y}}}{S_{\vec{x},\vec{x}}}$ is the **least-squares estimator of** $\beta_1$, $\hat{\beta}_0 = \overline{y} - \overline{x}\hat{\beta}_1$ is the **least-squares estimator of** $\beta_0$.

The **least-squares line** or **estimated regression line** is

$$y = \hat{\beta}_0 + \hat{\beta}_1 x = \overline{y} + (x - \overline{x})\hat{\beta}_1.$$

For $k = 1, 2, \ldots, n$, the **fitted values** are

$$\hat{y_k} = \hat{\beta}_0 + \hat{\beta}_1 x_k = \overline{y} + (x_k - \overline{x})\hat{\beta}_1$$

and the $k^{th}$ **residual** is

$$\epsilon_k \equiv (y_k - \hat{y_k}).$$

The **sample correlation coefficient** is

$$r = \frac{S_{\vec{x},\vec{y}}}{\sqrt{S_{\vec{x},\vec{x}} S_{\vec{y},\vec{y}}}}$$

$r^2$ is the **coefficient of determination.**

$$SST \equiv S_{\vec{y},\vec{y}}, \quad SSE \equiv \sum_k (y_k - \hat{y_k})^2 = (1 - r^2) S_{\vec{y},\vec{y}}, \quad SSR \equiv \sum_k (\hat{y_k} - \overline{y})^2 = r^2 S_{\vec{y},\vec{y}}$$

(**Total, Error, and Regression Sum of Squares**, respectively; see Regression Pictures 4.3 and 5.4)

Our favorite estimator of $\sigma$ is $s \equiv \hat{\sigma} \equiv \sqrt{\frac{SSE}{(n-2)}}$.

The estimated standard error for $\hat{\beta}_1$ is

$$s_{\hat{\beta}_1} \equiv \frac{s}{\sqrt{S_{\vec{x},\vec{x}}}}; \qquad \left(\frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}}\right) \text{ has a } t_{n-2} \text{ distribution.}$$

Denote, for $x^*$ a real number,

$$\mu_{Y \cdot x^*} \equiv E(Y | x = x^*) = \beta_0 + \beta_1 x^*, \quad \hat{Y} \equiv \hat{\mu}_{Y \cdot x^*} = \hat{\beta}_0 + \hat{\beta}_1 x^*.$$

The estimated standard error for $\hat{Y}$ is

$$s_{\hat{Y}} \equiv s\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{\vec{x},\vec{x}}}}; \qquad \left(\frac{\hat{Y} - \mu_{Y \cdot x^*}}{s_{\hat{Y}}}\right) \text{ has a } t_{n-2} \text{ distribution.}$$

## Chapter IX. Other model fitting

**Example 9.1.** Consider the bivariate data

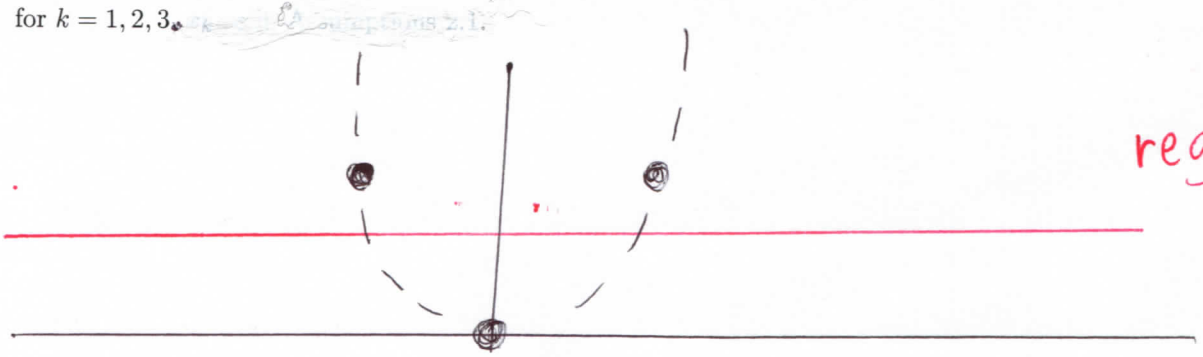
$$\{(-1,1),(0,0),(1,1)\}.$$

We will leave it to the reader to calculate that $\hat{\beta}_1 = 0 = r$. This means we are not getting a good linear fit to the data (see Definition 6.4, Proposition 5.8(c) and Remark 6.7).

Yet the data is very well behaved in terms of getting $y$ as a function of $x$;

$$y_k = x_k^2,$$

for $k = 1, 2, 3$.




This example suggests that, if we relax our standards in Definition 1.2 by allowing *quadratic* rather than merely linear functions of $x$ to equal the expected value of $Y$, we will increase our chances of success.

**Simple Quadratic Regression 9.2.** It is not hard to modify Definitions 1.2 and 2.3 to include quadratic functions $y = b_0 + b_1 x + b_2 x^2$, for $b_0, b_1, b_2$ real numbers.

A **Simple Quadratic Regression Model** (compare to Definition 1.2), for fixed numbers $\beta_0, \beta_1, \beta_2$, and $\sigma$ is

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \mathcal{E}.$$

The values of $x$ will be specified measurements and $\mathcal{E}$ is a normal random variable with mean $E(\mathcal{E}) = 0$ and variance $V(\mathcal{E}) = \sigma^2$, hence standard deviation $\sigma$.

The expected value of $Y$ is now

$$E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2.$$

For estimators of $\beta_0, \beta_1$, and $\beta_2$, as in Definitions 2.3, we want $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ so that $y = (\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2)$ minimizes the sum of squares of vertical displacements

$$SSV(b_0, b_1, b_2) \equiv \sum_{k=1}^{n} \left[y_k - (b_0 + b_1 x_k + b_2 x_k^2)\right]^2$$

from the bivariate data of Assumptions 2.1 to the parabola (see the drawing below Definitions 2.3 and replace the red line with a red parabola); that is,

$$SSV(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) \leq SSV(b_0, b_1, b_2)$$

for all real numbers $b_0, b_1, b_2$.

The parabola

$$y = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$$

is then the **least-squares parabola** or **estimated regression parabola** for the bivariate data in 2.1; $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$ are **least-squares estimators** of $\beta_0, \beta_1$, and $\beta_2$, respectively.

We do not wish to explicitly analogize Theorem 3.1 here, with direct formulas for $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$. What we *can* analogize, without excessive pain, is the systems of equations whose solutions are the desired least-squares estimators.

Here is an indirect algebraic sense in which our quadratic least-squares estimators $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$ are analogous to our linear least-squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

It can be shown (the reader with knowledge of vectors as in the Appendix should see Examples APP.12(a) and (b) in the Appendix) that the linear estimators $\hat{\beta}_0, \hat{\beta}_1$ are a solution of

$$\begin{array}{rcrcl} n\hat{\beta}_0 & + & (\sum_k x_k)\hat{\beta}_1 & = & \sum_k y_k \\ (\sum_k x_k)\hat{\beta}_0 & + & (\sum_k x_k^2)\hat{\beta}_1 & = & \sum_k x_k y_k \end{array}$$

while the quadratic estimators $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ are a solution of

$$\begin{array}{rcrcrcl} n\hat{\beta}_0 & + & (\sum_k x_k)\hat{\beta}_1 & + & (\sum_k x_k^2)\hat{\beta}_2 & = & \sum_k y_k \\ (\sum_k x_k)\hat{\beta}_0 & + & (\sum_k x_k^2)\hat{\beta}_1 & + & (\sum_k x_k^3)\hat{\beta}_2 & = & \sum_k x_k y_k \\ (\sum_k x_k^2)\hat{\beta}_0 & + & (\sum_k x_k^3)\hat{\beta}_1 & + & (\sum_k x_k^4)\hat{\beta}_2 & = & \sum_k x_k^2 y_k \end{array}$$

Solving the system of three equations directly above would give us $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$ of 9.2.

**Simple Polynomial Regression 9.3.** In 9.2, for arbitrary $N$ equal to $1, 2, 3, \ldots, b_0, b_1, b_2, b_3, \ldots, b_N$ arbitrary real numbers, replace $\left(b_0 + b_1 x + b_2 x^2\right)$ with $\left(b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots + b_N x^N\right)$.

**General Model 9.4.** Let $\mathbf{R}^n$ be the set of all ordered $n$-tuples of real numbers, also known as *vectors* $\vec{z} \equiv (z_1, z_2, \ldots, z_n)$. Let $W$ be a subspace (see Definition APP.7 in the Appendix) of $\mathbf{R}^n$ that contains the $n$-tuple $(1, 1, \ldots, 1)$ consisting entirely of 1s (denoted $\vec{1}$ in Definitions 4.2).

Given data organized as a vector $\vec{y} \equiv (y_1, y_2, \ldots, y_n)$ as in 4.3, let $\hat{y} = (\hat{y_1}, \hat{y_2}, \ldots \hat{y_n})$ be the vector in $W$ that minimizes, over $\vec{z}$ in $W$,

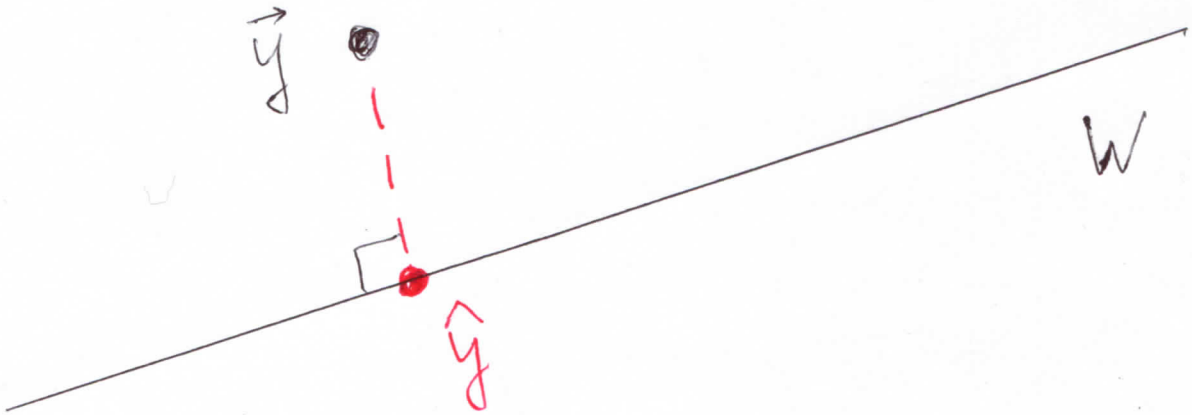
$$SSV(\vec{z}) \equiv \sum_{k=1}^{n} (y_k - z_k)^2;$$

that is, $\hat{y}$ is in $W$ and

$$SSV(\hat{y}) \leq SSV(\vec{z}),$$

for all $\vec{z}$ in $W$. In words, $\hat{y}$ is the best (least-squares) approximation of $\vec{y}$ from $W$. $W$ is the model that we are trying to fit the data to; it's where the data "should" be, in some idealized world.

In the picture below, the right-angle will be made explicit in APP.5 and APP.9 in the Appendix.



For example, with linear regression as in Definitions 2.3, $W$ would be

$$\{(a + bx_1, a + bx_2, \ldots, a + bx_n) \mid a, b \text{ real}\}.$$

When written in matrix form (see APP.8 in the Appendix), our General Model 9.4 is sometimes called the *general linear model.*

Let $SST, SSE$, and $SSR$ be as in Definitions 4.2, with $\vec{y}$ and $\hat{y}$ as in General Model 9.4; we then have the same right triangle as in Regression Picture 4.3 (this is proven, with vectors, in APP.13 in the Appendix).

We then *define* the **coefficient of multiple determination** to be

$$R^2 \equiv \frac{SSR}{SST}.$$

This is in contrast to the case of linear regression, where the coefficient of determination is defined to be

$$r^2 = \frac{(S_{\vec{x},\vec{y}})^2}{S_{\vec{x},\vec{x}} S_{\vec{y},\vec{y}}}$$

and is then (Theorem 5.3) shown to be equal to $\frac{SSR}{SST}$.

As with $r^2$, it can be shown that

$$SST = SSR + SSE,$$

which is equivalent to $(1 - R^2) = \frac{SSE}{SST}$; the reader familiar with vectors as in the Appendix should see APP.13 in the Appendix.

$R^2$ measures how close our data is to the desired model $W$: $0 \leq R^2 \leq 1$ always, and the closer $R^2$ is to 1, the closer our data is to $W$.

**Definitions 9.5. ANOVA** stands for "analysis of variance." Unfortunately for coherence, this does *not* mean the study of population variance $\sigma^2$ or sample variance $s^2$; ANOVA refers to the variance, as in differences, between means of different populations.

More specifically, let $I$ and $J$ be fixed natural numbers. Suppose, for $1 \leq i \leq I$, the $i^{th}$ population has a mean $\mu_i$. We are interested in the null hypothesis

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_I;$$

the alternative hypothesis is the negation of $H_0$:

$$H_a : \text{at least two of the means are different from each other}$$

For example, let $X_1$ be the height, in inches, of a randomly chosen person from Columbus, Ohio, $X_2$ be the height, in inches, of a randomly chosen person from Cleveland, Ohio, and $X_3$ be the height, in inches, of a randomly chosen person from Cincinnati, Ohio, The null hypothesis

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

translates as "there is no difference in average heights, of people from Columbus, Cleveland, and Cincinnati, Ohio."

This might be of interest to a basketball talent scout.

For data, we need measurements from each population, carefully indexed by population.

For $1 \leq i \leq I, 1 \leq j \leq J$, define

$$X_{ij} \equiv j^{th} \text{ measurement of } i^{th} \text{ population.}$$

In the example above regarding heights of people from Columbus (population 1), Cleveland (population 2), and Cincinnati (population 3), suppose we measure 5 people from each of those 3 cities. Then $1 \leq i \leq 3 \equiv I, 1 \leq j \leq 5 \equiv J$, and, for the values of $i$ and $j$ just stated,

$$X_{ij} = \text{ the height, in inches, of the } j^{th} \text{ person from the } i^{th} \text{ city.}$$

To make ANOVA data look more familiar, that is, of the form

$$\vec{y} \equiv (y_1, y_2, y_3, \ldots, y_n),$$

where $n$ is the number of data, note first that $n = IJ$, since our unordered data is

$$\{x_{ij} \,|\, 1 \leq i \leq I, 1 \leq j \leq J\}.$$

We will find it convenient to first list, in order, all the measurements of the first population, then all the measurements of the second population, etc.:

$$\vec{y} = (x_{11}, x_{12}, x_{13}, \ldots, x_{1J}, x_{21}, x_{22}, x_{23}, \ldots, x_{2J}, \ldots, x_{I1}, x_{I2}, x_{I3}, \ldots, x_{IJ}).$$

We need some awkward terminology.

For $1 \leq i \leq I$,

$$\overline{x_i} \equiv \frac{1}{J} \sum_{j=1}^{J} x_{ij};$$

this is the average of our measurements of the $i^{th}$ population.

$$\overline{x} \equiv \overline{y} \equiv \frac{1}{n} \sum_{k=1}^{n} y_k = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij} = \frac{1}{I} \sum_{i=1}^{I} \overline{x_i},$$

the average of the population averages; this is sometimes called the **grand mean** of all the data.

The subspace $W$ that we will try to fit our data to (see 9.4) is the set of all vectors in $\mathbf{R}^n$ ($n \equiv IJ$) such that the first $J$ coordinates are equal, the second $J$ coordinates are equal, etc. This is imagining that, within a fixed population, all measurements are the same.

The best (least squares) approximation of data $\vec{y}$ from $W$ can be shown, with vector expertise as in the Appendix (see APP.12(c) in the Appendix) to be the vector, denoted $\hat{y} \equiv (\hat{y_1}, \hat{y_2}, \ldots, \hat{y_n})$, whose first $J$ coordinates are each $\overline{x}_1$, the second $J$ coordinates are $\overline{x}_2$, etc. See Picture 9.7 on the next page.

**ANOVA Alphabet Soup 9.6.** As in Definitions 4.2, define the **total sum of squares**

$$SST \equiv \sum_{k=1}^{n} (y_k - \overline{y})^2 \equiv \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \overline{x})^2;$$

the **regression sum of squares**

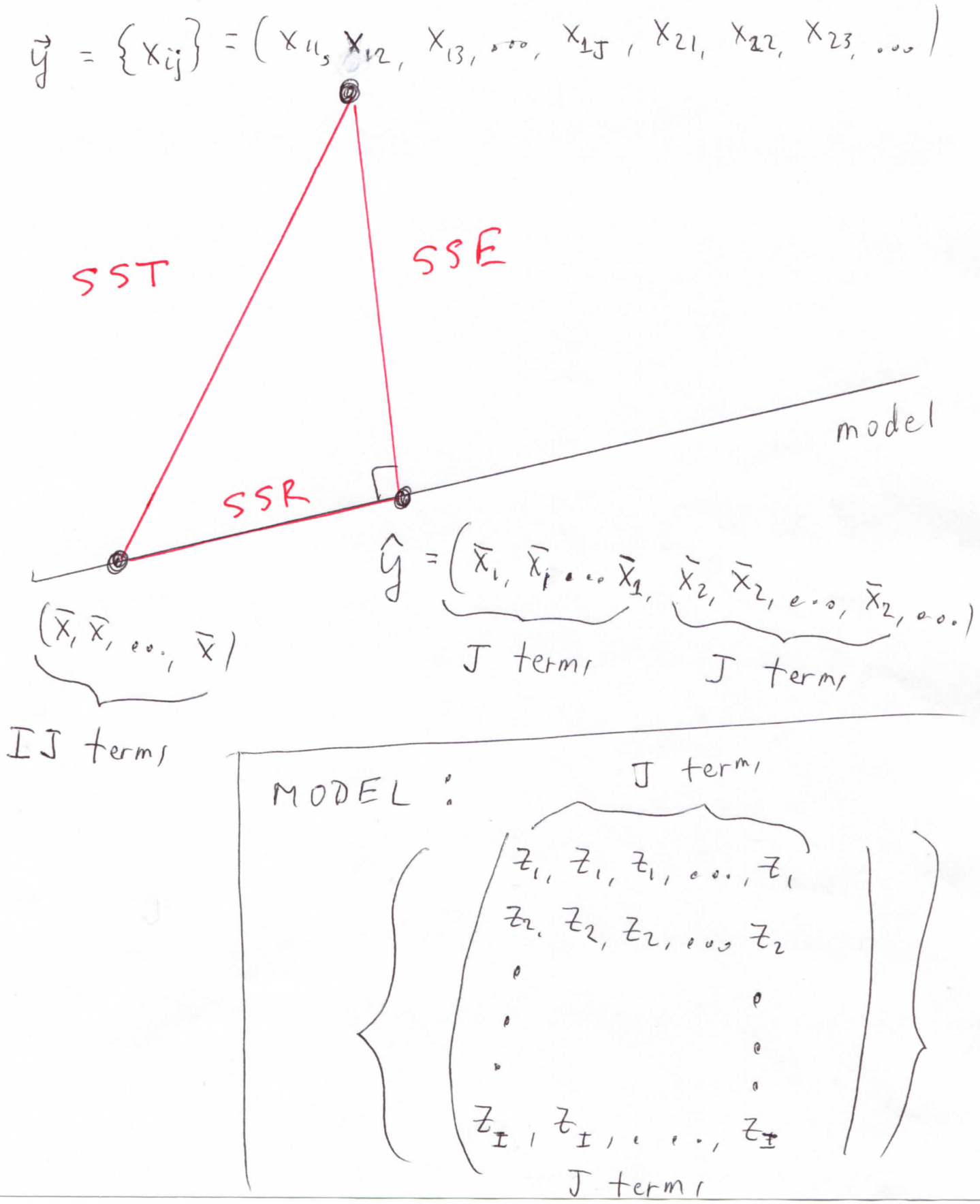
$$SSR \equiv \sum_{k=1}^{n} (\hat{y_k} - \overline{y})^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (\overline{x}_i - \overline{x})^2 = J \sum_{i=1}^{I} (\overline{x}_i - \overline{x})^2;$$

and the **error sum of squares**

$$SSE \equiv \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \overline{x}_i)^2.$$

As suggested by the picture on the next page, $SST = SSR + SSE$; compare to 4.3 and APP.13.

**ANOVA Sum of Squares Picture 9.7**



**Terminology 9.8.** Because of medical and agricultural origins, the $i^{th}$ population is sometimes called the $i^{th}$ **treatment**; $SSR$ then becomes $SSTr$. We will continue to use $SSR$, as in 9.6, as our terminology, to emphasize that ANOVA is a special case of model fitting, as in 9.4.

**Example 9.9.** Let's get back to human heights, in Columbus (population 1, abbreviated Col), Cleveland (population 2, abbreviated Cl), and Cincinnati (population 3, abbreviated Cin). Let's say we measure 5 people in each of those towns and get the following heights, in inches.

| $j$ | $x_{1j}$(Col) | $x_{2j}$(Cl) | $x_{3j}$(Cin) |
|---|---|---|---|
| 1 | 75 | 68 | 68 |
| 2 | 67 | 67 | 70 |
| 3 | 68 | 66 | 70 |
| 4 | 72 | 64 | 71 |
| 5 | 71 | 63 | 71 |
| $\sum_j$ | 353 | 328 | 350 |

Here are the averages of the measurements from each population:

$$\overline{x_1} = \frac{353}{5} = 70.6, \ \overline{x_2} = \frac{328}{5} = 65.6, \ \overline{x_3} = \frac{350}{5} = 70.$$

The grand mean is

$$\overline{x} = \frac{(70.6 + 65.6 + 70)}{3} \sim 68.73.$$

Here are the alphabet soup sums of squares. Notice that $SST = SSR + SSE$, at least up to one decimal place; it can be shown that $SST$ precisely equals $SSR + SSE$, as suggested by the Pythagorean theorem and the right triangle on the next page.

$$SST \sim (75-68.73)^2+(67-68.73)^2+(68-68.73)^2+(72-68.73)^2+(71-68.73)^2+(68-68.73)^2+(67-68.73)^2$$
$$+(66-68.73)^2+(64-68.73)^2+(63-68.73)^2+(68-68.73)^2+(70-68.73)^2+(70-68.73)^2+(71-68.73)^2+(71-68.73)^2$$
$$\sim 138.94.$$

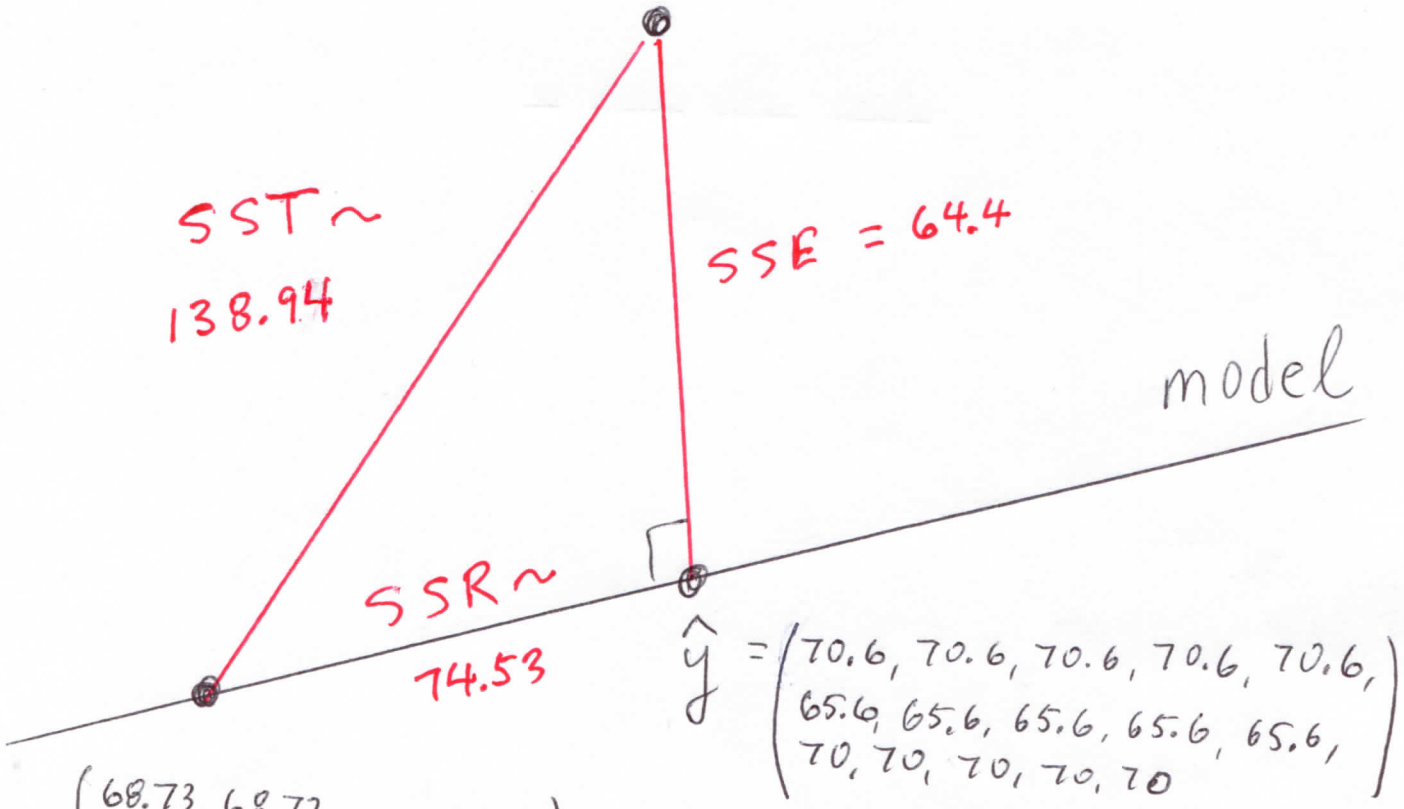
$$SSE = (75-70.6)^2+(67-70.6)^2+(68-70.6)^2+(72-70.6)^2+(71-70.6)^2+(68-65.6)^2+(67-65.6)^2$$
$$+(66-65.6)^2+(64-65.6)^2+(63-65.6)^2+(68-70)^2+(70-70)^2+(70-70)^2+(71-70)^2+(71-70)^2 = 64.4.$$

$$SSR = 5\left[(\overline{x_1} - \overline{x})^2 + (\overline{x_2} - \overline{x})^2 + (\overline{x_3} - \overline{x})^2\right] \sim 5\left[(70.6 - 68.73)^2 + (65.6 - 68.73)^2 + (70 - 68.73)^2\right] \sim 74.53.$$

See the drawing on the next page.

**ANOVA picture for Example 9.9.**

$\vec{y} = (75, 67, 68, 72, 71, 68, 67, 66, 64, 63, 68, 70, 70, 71, 71)$




MODEL:

$$\left\{ \begin{pmatrix} z_1, z_1, z_1, z_1, z_1, z_2, z_2, z_2, \\ z_2, z_2, z_3, z_3, z_3, z_3, z_3 \end{pmatrix} \right\}$$

$z_1, z_2, z_3$ real

**ANOVA Test Statistic 9.10.** For the remainder of Chapter IX, assume, for $I, J$ and other terminology as in 9.5, for $1 \leq i \leq I, 1 \leq j \leq J$, that $X_{ij}$ is normal, with mean $E(X_{ij}) = \mu_i$, variance $V(X_{ij})$ constant and $\{X_{ij} \mid 1 \leq i \leq I, 1 \leq j \leq J\}$ independent.

For the ANOVA hypothesis test at the beginning of 9.5 our test statistic is

$$f = \frac{\frac{SSR}{(I-1)}}{\frac{SSE}{I(J-1)}},$$

which turns out (we will not go into this) to have what is called an $F_{(I-1),I(J-1)}$ distribution. See 6.6 for another place where the $F$ distribution appears.

Inspection of the definitions in 9.6 tells us that $SSR$ is measuring the differences *between* populations (notice the $(\overline{x}_i - \overline{x})^2$ terms) while $SSE$ is measuring the differences *within* populations (notice the $(x_{ij} - \overline{x}_i)^2$ terms). This suggests that the ratio $\frac{SSR}{SSE}$, hence $f$, measures the *relative* difference between sample means $\overline{x_i}, 1 \leq i \leq I$, thus, as $f$ gets large, we are less inclined to believe the null hypothesis of *population* means being equal.

As with normal and $t$ distributions, we have *critical values*: if $0 < \alpha < 1$, $F_{\alpha,(I-1),I(J-1)}$ is a positive number such that

$$P\left(F_{(I-1),I(J-1)} > F_{\alpha,(I-1),I(J-1)}\right) = \alpha.$$

Since our P-value is

$$P\left(F_{(I-1),I(J-1)} > f\right),$$

we reject $H_0$, at significance level $\alpha$, if and only if $f \geq F_{\alpha,(I-1),I(J-1)}$.

**Example 9.11.** Let's perform ANOVA in Example 9.9. We have $I = 3$ and $J = 5$, so $(I-1) = 2$ and $I(J-1) = 12$. Thus our test statistic in 9.10 has an $F_{2,12}$ distribution.

Here is relevant information about critical values, from [6, Table A.9]:

| $\alpha$ | $F_{\alpha,2,12}$ |
|---|---|
| 0.1 | 2.81 |
| 0.05 | 3.89 |
| 0.01 | 6.93 |
| 0.001 | 12.97 |

From Example 9.9,

$$f \sim \frac{\frac{74.53}{2}}{\frac{64.4}{12}} \sim 6.94.$$

Since $f \geq F_{0.1,2,12}, F_{0.05,2,12}$, and $F_{0.01,2,12}$, we reject $H_0$ at significance levels $0.1, 0.05$, or $0.01$; since $f < F_{0.001,2,12}$, we do not reject $H_0$ at significance level $0.001$.

We could also have taken a more direct P-value approach.

$$\text{P-value} = P(F_{2,12} > 6.94) < P(F_{2,12} > 6.93) = 0.01,$$

while

$$\text{P-value} = P(F_{2,12} > 6.94) > P(F_{2,12} > 12.97) = 0.001,$$

thus the most we can say about our P-value, using the information given, is

$$0.001 < \text{P-value} < 0.01.$$

This tells us to reject $H_0$ at significance level greater than or equal to 0.01, and to not reject $H_0$ at significance level less than or equal to 0.001.

In words, at significance level greater than or equal to 0.01, the data suggests there is no difference in average heights of people from Columbus, Cleveland, and Cincinnati; at significance level less than or equal to 0.001, the data does not suggest there is no difference in average heights of people from that the populations of Columbus, Cleveland, and Cincinnati are equal, on average.

# APPENDIX

This Appendix will quickly summarize finite-dimensional vectors in APP.1–APP.11 and apply them to prove our results about least-squares estimators and sums of squares of error and regression (APP.12–APP.13). We also apply vectors to understanding the $(x^* - \overline{x})^2$ in the estimated standard error in Definitions 7.2; see APP.15.

For this Appendix, we assume the reader is familiar with *matrices*, as in [2, Section IA], including transpose, row $n$-vectors and column $n$-vectors, for $n = 1, 2, 3, \ldots$. Much of APP.1–APP.11 may be found in [2, Sections I.B, VI.A, VI.B, and VI.E].

**Definitions APP.1.** For $n = 1, 2, 3, \ldots$, an **$n$-vector** is an ordered $n$-tuple of real numbers

$$\vec{y} \equiv (y_1, y_2, \ldots, y_n).$$

For $1 \leq k \leq n$, $y_k$ is the $k^{th}$ **component** of $\vec{y}$.

The set of all $n$-vectors is denoted $\mathbf{R}^n$ (reads "R enn").

We have some algebra in $\mathbf{R}^n$: If $\vec{y} \equiv (y_1, y_2, \ldots, y_n)$ and $\vec{z} \equiv (z_1, z_2, \ldots, z_n)$ are $n$-vectors and $c$ is a real number, then $(\vec{y} + c\vec{z})$ is the vector

$$((y_1 + cz_1), (y_2 + cz_2), \ldots, (y_n + cz_n)).$$

In words, we add vectors and multiply vectors by numbers componentwise.

The **norm** or **magnitude** of $\vec{y} \equiv (y_1, y_2, \ldots, y_n)$ is

$$\|\vec{y}\| \equiv \sqrt{y_1^2 + y_2^2 + \cdots + y_n^2}.$$

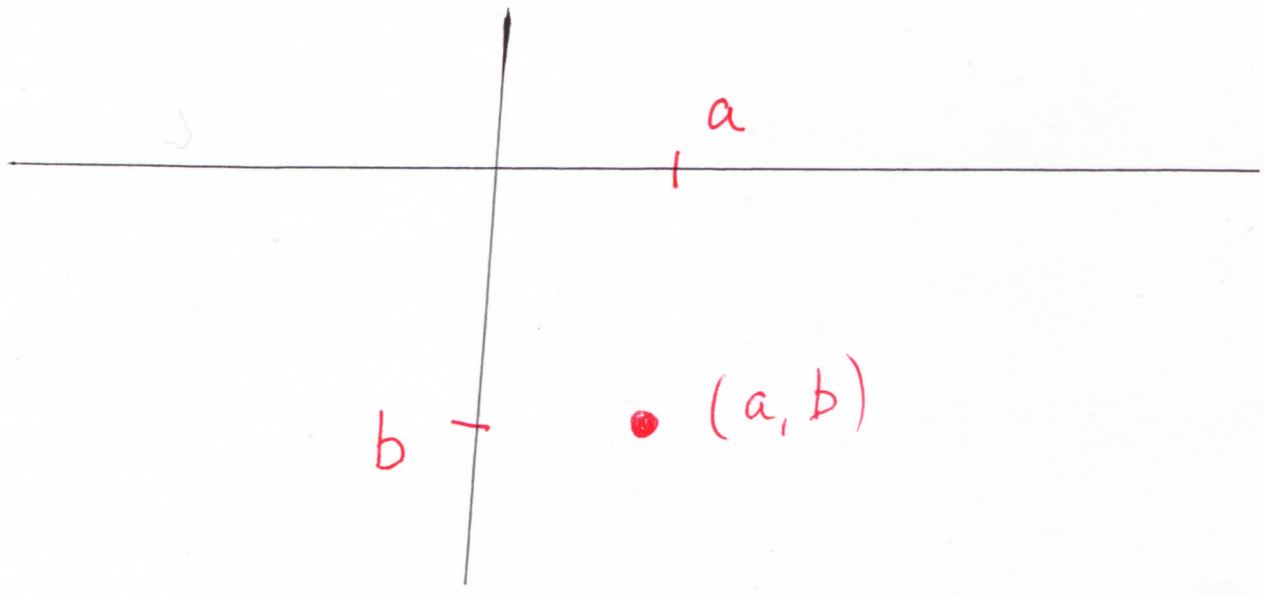
Note that *SSV*, from 9.4, is norm squared:

$$SSV(\vec{z}) \equiv \sum_{k=1}^{n} (y_k - z_k)^2 \equiv \|\vec{y} - \vec{z}\|^2,$$

which we may think of as the square of the *distance* between $\vec{z}$ and $\vec{y}$; see APP.6.

**Pictures APP.2.** When $n = 2$ we may draw pictures of $n$-vectors.

The *Cartesian plane* represents ordered pairs $(a, b)$ as points or dots; $a$ is the horizontal displacement from the origin, labeled $(0, 0)$, and $b$ is the vertical displacement from the origin.

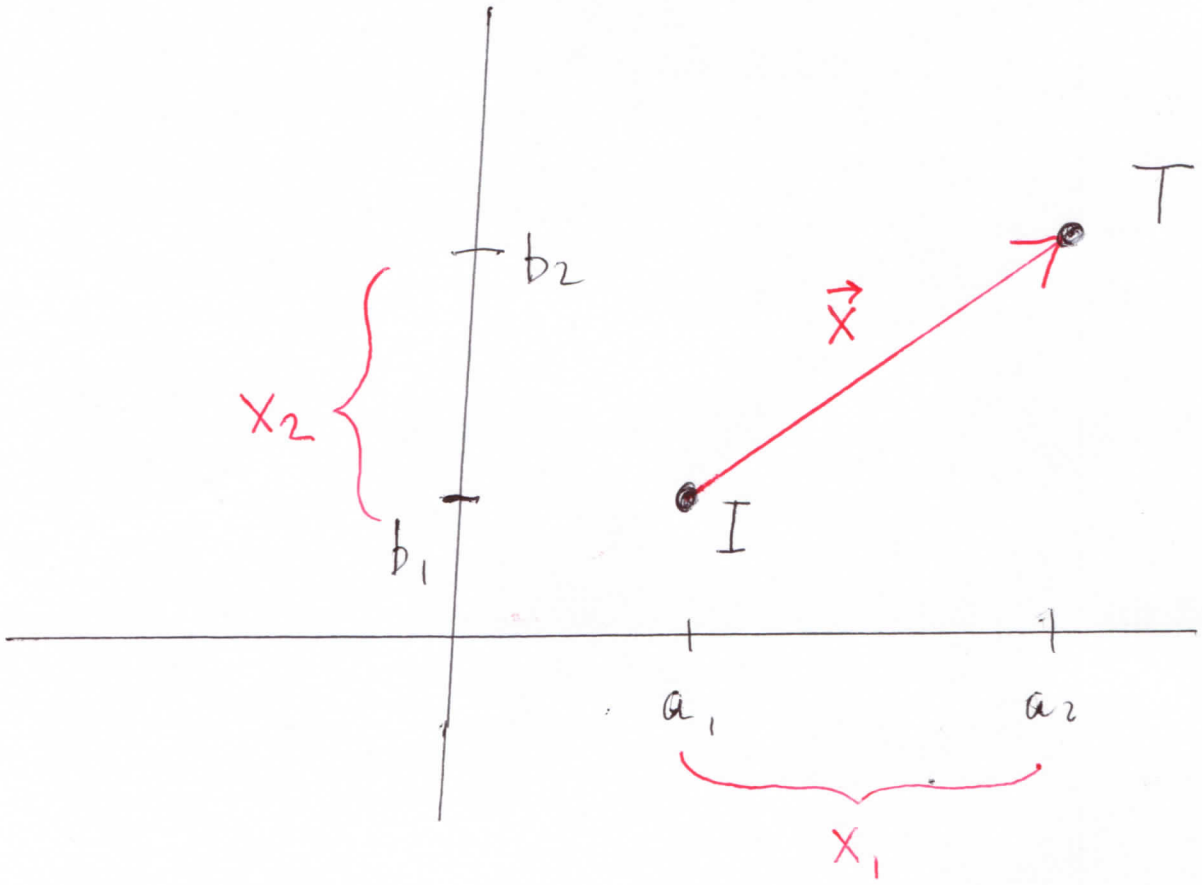


Often of more interest is to take a *pair* of points and draw an arrow from one point to the other: If $I \equiv (a_1, b_1)$ and $T \equiv (a_2, b_2)$, with

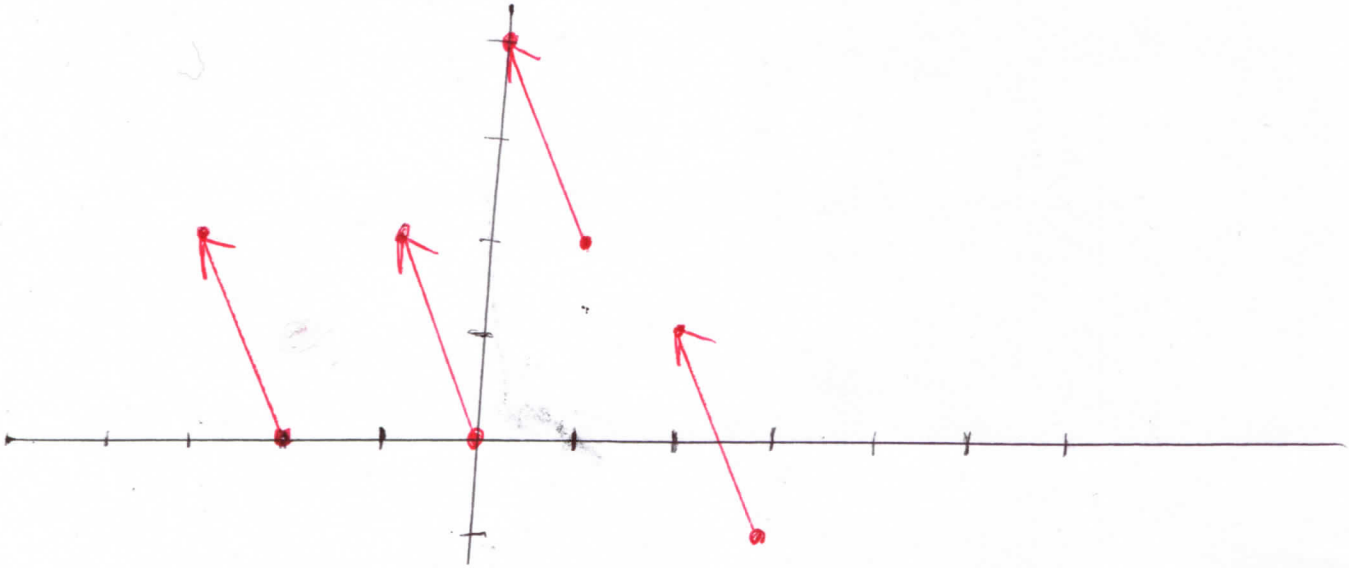
$$x_1 = (a_2 - a_1) \quad \text{and} \quad x_2 = (b_2 - b_1),$$

then the 2-vector $\vec{x} \equiv (x_1, x_2)$ is **represented by** the arrow, or directed line segment, from $I$ (the **initial point**) to $T$ (the **terminal point**).

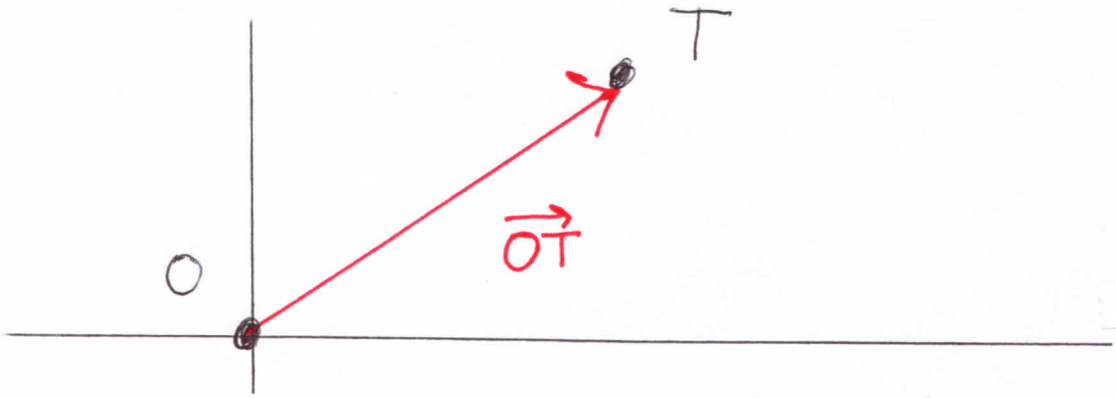
Said arrow is sometimes denoted $\overrightarrow{IT}$, and explains the arrow terminology in Definitions APP.1.




Notice that every 2-vector is represented by infinitely many arrows. Below we have drawn many arrows that represent $(-1, 2)$.



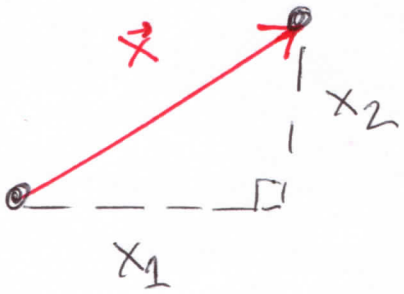
Denoting by $O$ the origin, the vector represented by $\overrightarrow{OT}$ is said to be in **standard position**. The **position vector** for a point $T$ is the vector represented by $\overrightarrow{OT}$. This describes a one-to-one correspondence between 2-vectors represented as points ($T$ below) and 2-vectors represented as arrows ($\overrightarrow{OT}$ below).




It is also worth noting that the norm of $\vec{x} = (x_1, x_2)$

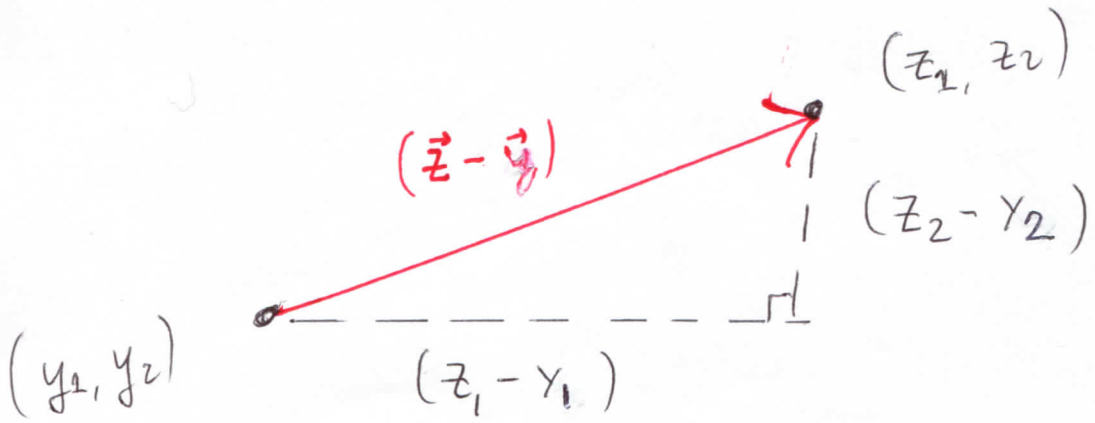
$$\|\vec{x}\| \equiv \sqrt{x_1^2 + x_2^2}$$

is the length of an arrow representing $\vec{x}$, by the Pythagorean theorem.

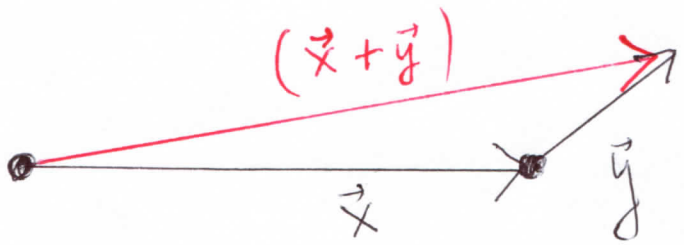



The distance between two points $\vec{y} = (y_1, y_2)$ and $\vec{z} = (z_1, z_2)$ is

$$\|(\vec{z} - \vec{y})\| = \sqrt{(z_1 - y_1)^2 + (z_2 - y_2)^2}.$$



It can also be shown that addition of 2-vectors looks like the following, in terms of arrows representing said vectors:




Multiplication by a number has two possible pictures, depending on whether the number is positive or negative.


**Calculation and Definition APP.3.** We would like an algebraic characterization of a pair of (arrows representing) 2-vectors being perpendicular.

We will leave it to the reader (or see [2, Terminology 6.5, pages 400–403]) to show that, for $\vec{x} \equiv (x_1, x_2), \vec{y} \equiv (y_1, y_2)$,

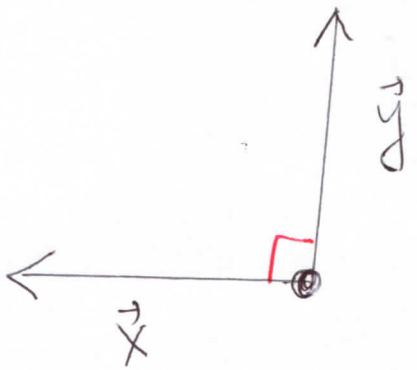
$$\|\vec{x} + \vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2(x_1y_1 + x_2y_2).$$

By the Pythagorean theorem, $\vec{x}$ is perpendicular to $\vec{y}$ if and only if that last term $(x_1y_1 + x_2y_2)$ equals zero. Thus we like to give it a name: the **dot product** or **inner product** of $\vec{x}$ and $\vec{y}$ is

$$\vec{x} \cdot \vec{y} \equiv (x_1y_1 + x_2y_2).$$

Now we may relate geometry to algebra:

**Theorem.** A pair of 2-vectors $\vec{x}$ and $\vec{y}$ are perpendicular if and only if their dot product is zero.



**Advice and Definitions APP.4.** In $\mathbf{R}^n, n = 1, 2, 3, \ldots$, we encourage the reader to *think of* $n$-vectors as if $n$ were 2, in ideas and pictures as in APP.2 and APP.3.

As with $n = 2$ in APP.2, we'd like to think of $n$-vectors as being either points or arrows.

The pictures of addition of 2-vectors and multiplication of 2-vectors by real numbers, drawn at the end of APP.2, can and should be drawn for the same operations with $n$-vectors.

For $n = 1, 2, 3, \ldots$, extend the definition of **dot product** (APP.3) to $\mathbf{R}^n$ in the most natural way:

$$(x_1, x_2, \ldots, x_n) \cdot (y_1, y_2, \ldots, y_n) \equiv \sum_{k=1}^{n} x_k y_k.$$

Motivated by the Theorem in APP.3, *define* vectors $\vec{x}$ and $\vec{y}$ in $\mathbf{R}^n$ to be **perpendicular**, also called **orthogonal**, denoted

$$\vec{x} \perp \vec{y},$$

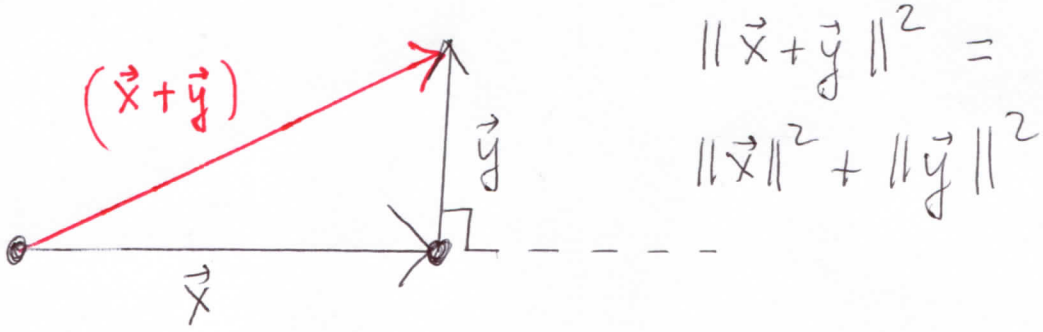
if $\vec{x} \cdot \vec{y} = 0$. The picture for the just-mentioned Theorem should be drawn here.

By virtually the same argument, the Theorem in APP.3 extends to the following.

**Pythagorean Theorem APP.5.** For $n = 1, 2, 3, \ldots$, a pair of $n$-vectors $\vec{x}$ and $\vec{y}$ are perpendicular if and only if

$$\|\vec{x} + \vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2.$$

As part of the pictorial point of view recommended, visualize this Theorem as a right triangle.



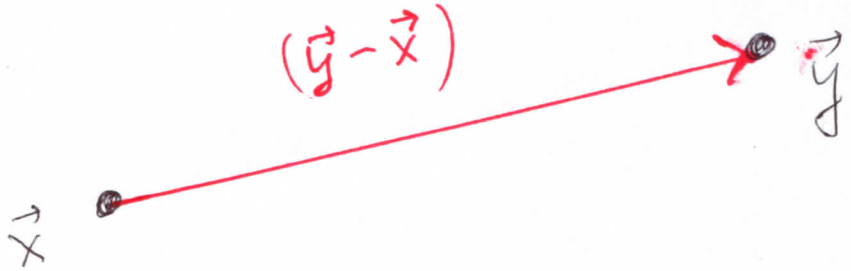

**More Advice and Definitions APP.6.** For $\vec{x} = (x_1, x_2, \ldots, x_n), \vec{y} = (y_1, y_2, \ldots, y_n)$, let's define the **vector from $\vec{x}$ to $\vec{y}$** as

$$(\vec{y} - \vec{x}),$$

which we've already defined as a special case of our vector algebra in APP.1, namely

$$((y_1 - x_1), (y_2 - x_2), \ldots, (y_n - x_n)).$$

Thinking as if $n = 2$ (see APP.2), visualize $\vec{y}$ and $\vec{x}$ as points, and $(\vec{y} - \vec{x})$ as an arrow, with initial point $\vec{x}$ and terminal point $\vec{y}$.



Again mimicking $n = 2$, define, in the picture just drawn of "points" $\vec{x}$ and $\vec{y}$, and an "arrow" $(\vec{y} - \vec{x})$, the **distance** between $\vec{x}$ and $\vec{y}$ to be

$$\|(\vec{y} - \vec{x})\| = \sqrt{((y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2)}.$$

As in the Regression Picture 4.3, we have data $y_1, y_2, \ldots, y_n$ that we put together into a vector $\vec{y} \equiv (y_1, y_2, \ldots, y_n)$. The theme of this Magnification is that we expect, at least on average, said vector to belong to a certain type of subset (our "model" that we try to fit the data to) of $\mathbf{R}^n$, that we will now describe.

**Definition APP.7.** A subset, $W$, of $\mathbf{R}^n$, is a **subspace** if it has the following two properties.

(1) If $\vec{x}$ and $\vec{y}$ are in $W$, then $(\vec{x} + \vec{y})$ is in $W$.

(2) If $\vec{x}$ is in $W$ and $c$ is a real number, then $c\vec{x}$ is in $W$.

**Terminology APP.8.** If, for some $k$ and $m$, $\vec{x}$ is a $k$-vector and $A$ is an $(m \times k)$ matrix, then $A\vec{x}$ is the $m$-vector obtained by writing $\vec{x}$ as a column $k$-vector, then performing matrix multiplication.

It can be shown that any subspace of $\mathbf{R}^n$ has the form

$$\{A\vec{x} \mid \vec{x} \text{ is in } \mathbf{R}^k\}$$

for some $k = 1, 2, 3, \ldots$, and $(n \times k)$ matrix $A$.

When a subspace $W$ is written in the matrix form just mentioned, the General Model 9.4 then becomes (see 1.2)

$$Y = A\vec{x} + \mathcal{E},$$

with $\mathcal{E}$ as in 1.2, $\vec{x}$ in $\mathbf{R}^k$, and is sometimes called the *general linear form* (see 9.4).

Given data $\vec{y}$ in $\mathbf{R}^n$ and a model, that is, a subspace $W$ of $\mathbf{R}^n$ as in 9.4, that we are trying to fit $\vec{y}$ to, our goal is to find $\vec{x}$ in $W$ that $\vec{y}$ is closest to; that is, we want

$$\|\vec{y} - \vec{x}\| \leq \|\vec{y} - \vec{w}\|,$$

for all $\vec{w}$ in $W$.

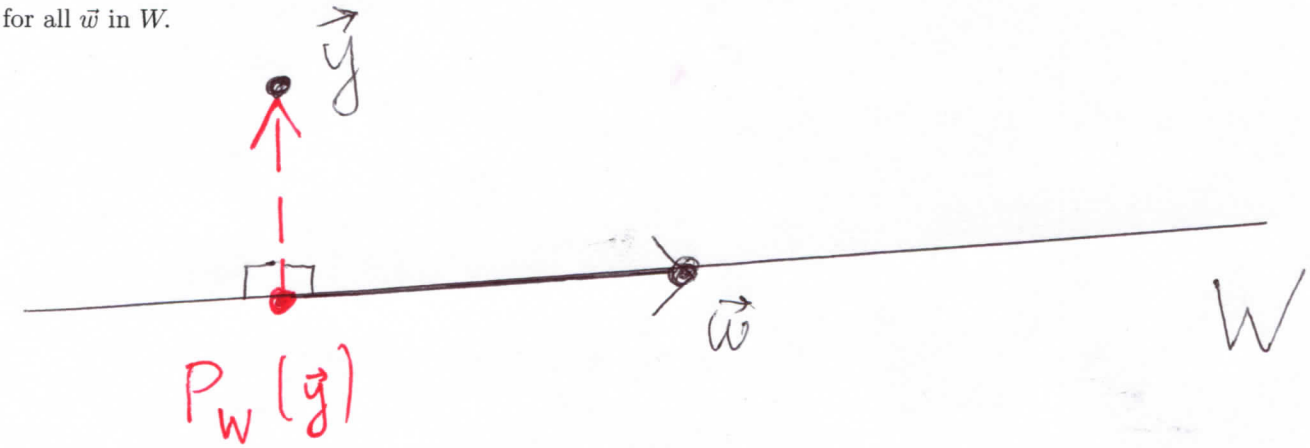
Our intuition is to "drop a perpendicular" from $\vec{y}$ onto $W$; that is, we want $\vec{w}_0$ in $W$ so that $(\vec{y} - \vec{w}_0)$ is perpendicular to all vectors in $W$.

This is given a name; see the picture after Definition APP.9.

**Definition APP.9.** (See [2, Definition 6.13, page 410].) If $W$ is a subspace of $\mathbf{R}^n$ and $\vec{y}$ is in $\mathbf{R}^n$, then the **(orthogonal) projection of $\vec{y}$ onto $W$**, denoted $P_W(\vec{y})$, is a vector in $W$ such that

$$(\vec{y} - P_W(\vec{y})) \perp \vec{w},$$

for all $\vec{w}$ in $W$.



Here is a precise statement of our "drop a perpendicular" intuition.

**Theorem APP.10.** (See [2, Theorem 6.14, page 412]) For $W$ and $\vec{y}$ as in APP.9, $P_W(\vec{y})$ is the unique best approximation (also called least-squares approximation) of $\vec{y}$ from $W$; that is,

$$\|\vec{y} - P_W(\vec{y})\| \leq \|\vec{y} - \vec{w}\|,$$

for all $\vec{w}$ in $W$.

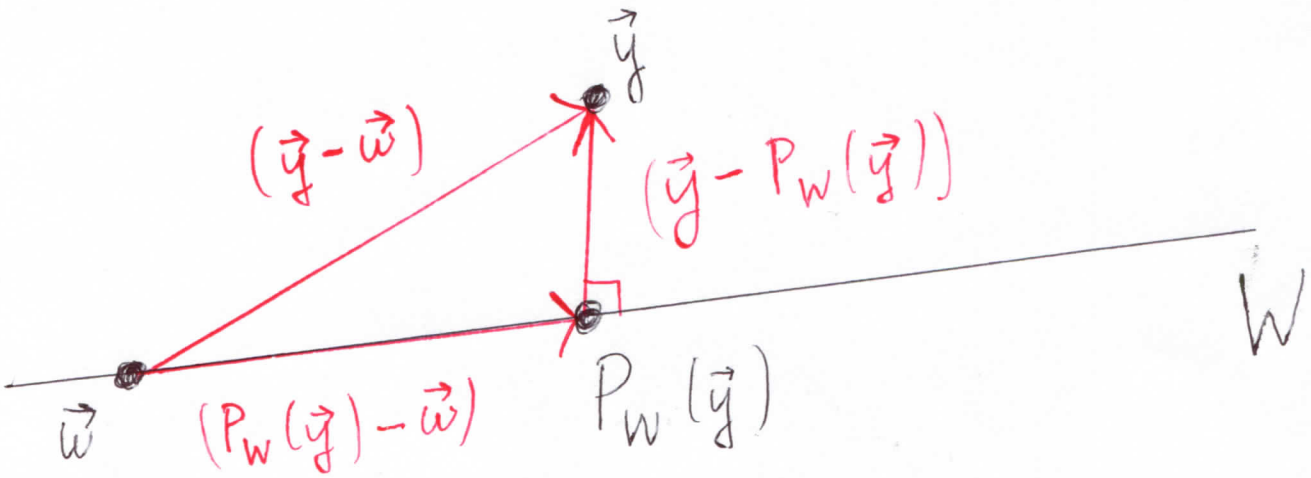
**Proof:** Fix $\vec{w}$ in $W$. Since $(\vec{y} - P_W(\vec{y})) \perp (P_W(\vec{y}) - \vec{w})$, the Pythagorean theorem APP.5 implies that

$$\|\vec{y} - \vec{w}\|^2 = \|(\vec{y} - P_W(\vec{y})) + (P_W(\vec{y}) - \vec{w})\|^2 = \|(\vec{y} - P_W(\vec{y}))\|^2 + \|(P_W(\vec{y}) - \vec{w})\|^2 \quad (*).$$

(*) clearly shows that $\|\vec{y} - \vec{w}\|^2 \geq \|\vec{y} - P_W(\vec{y})\|^2$, hence

$$\|\vec{y} - \vec{w}\| \geq \|\vec{y} - P_W(\vec{y})\|, \quad \text{for any } \vec{w} \text{ in } W,$$

thus $P_W(\vec{y})$ is a best approximation of $\vec{y}$ from $W$. Uniqueness also follows from (*), since $\|\vec{y} - \vec{w}\| = \|\vec{y} - P_W(\vec{y})\|$ then implies that $\|P_W(\vec{y}) - \vec{w}\| = 0$, which implies that $\vec{w} = P_W(\vec{y})$. □




Here is a more surprising result, where we denote by $A^T$ the *transpose* of a matrix $A$. Recall (Terminology APP.8) that all subspaces $W$ of $\mathbf{R}^n$ have the form of Theorem APP.11.

**Theorem APP.11.** (See [2, Theorem 6.55, pages 490–491]. Suppose, for some $k = 1, 2, 3, \ldots, (n \times k)$ matrix $A$,

$$W = \{A\vec{x} \,|\, \vec{x} \text{ is in } \mathbf{R}^k\}.$$

Then, for any $\vec{y}$ in $\mathbf{R}^n$, $x^*$ in $\mathbf{R}^k$, $Ax^* = P_W(\vec{y})$ if and only if $x^*$ is a solution of the **normal equations**

$$A^T A x^* = A^T \vec{y}.$$

**Proof:** See [2, Theorem 6.55, pages 490–491]. □

**Examples APP.12.** Let's use Theorems APP.10 and APP.11 to get least-squares estimators in the special cases of model fitting that we have discussed in this Magnification.

(a) **Simple Linear Regression.** The subspace we are trying to fit the data $\vec{y}$ to (see Definitions 2.3) is

$$W \equiv \{((b_0 + b_1 x_1), (b_0 + b_1 x_2), \ldots, (b_0 + b_1 x_n))\} :$$

the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are minimizing what we called, in Definitions 2.3,

$$SSV(b_0, b_1) \equiv \sum_{k=1}^{n} [y_k - (b_0 + b_1 x_k)]^2 = \|\vec{y} - ((b_0 + b_1 x_1), (b_0 + b_1 x_2), \ldots, (b_0 + b_1 x_n))\|^2.$$

Writing the vectors in $W$ as column vectors

$$\begin{bmatrix} b_0 + b_1 x_1 \\ b_0 + b_1 x_2 \\ \cdot \\ \cdot \\ \cdot \\ b_0 + b_1 x_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

we see that we are in the setting of Theorem APP.11, with

$$A \equiv \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} \quad k = 2 \quad \text{and} \quad x^* = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}.$$

Theorems APP.11 and APP.10 tell us that the least-squares approximation of $\vec{y}$ from $W$ is

$$\hat{y_k} = \hat{\beta}_0 + \hat{\beta}_1 x_k \ \ (k = 1, 2, 3, \ldots, n),$$

where $(\hat{\beta}_0, \hat{\beta}_1)$ is the solution of the normal equations

$$A^T A \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = A^T \vec{y}.$$

We leave it to the reader to perform matrix multiplication, simplifying the normal equations to

$$\begin{bmatrix} n & (\sum_{k=1}^n x_k) \\ (\sum_{k=1}^n x_k) & (\sum_{k=1}^n x_k^2) \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n y_k \\ \sum_{k=1}^n x_k y_k \end{bmatrix}$$

or

$$\begin{array}{rcrcl} n\hat{\beta}_0 & + & (\sum_k x_k)\hat{\beta}_1 & = & \sum_k y_k \\ (\sum_k x_k)\hat{\beta}_0 & + & (\sum_k x_k^2)\hat{\beta}_1 & = & \sum_k x_k y_k \end{array}$$

as in (near the end of) 9.2.

Solving these normal equations gives us Theorem 3.1.

(b) **Simple Quadratic Regression.** See 9.2. This is very similar to part (a) of these Examples, so we will only sketch the argument, relying on analogies to part (a).

The model $W$ that we are now fitting $\vec{y}$ to has vectors of the form

$$\begin{bmatrix} b_0 + b_1 x_1 + b_2 x_1^2 \\ b_0 + b_1 x_2 + b_2 x_2^2 \\ \cdot \\ \cdot \\ \cdot \\ b_0 + b_1 x_n + b_2 x_n^2 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

so we now have Theorem APP.11 with

$$A \equiv \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 \end{bmatrix} \quad k = 3 \quad \text{and} \quad x^* = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

and normal equations

$$A^T A \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = A^T \vec{y},$$

that, after much worse matrix multiplication than in (a), simplify to

$$\begin{bmatrix} n & (\sum_{k=1}^n x_k) & (\sum_{k=1}^n x_k^2) \\ (\sum_{k=1}^n x_k) & (\sum_{k=1}^n x_k^2) & (\sum_{k=1}^n x_k^3) \\ (\sum_{k=1}^n x_k^2) & (\sum_{k=1}^n x_k^3) & (\sum_{k=1}^n x_k^4) \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n y_k \\ \sum_{k=1}^n x_k y_k \\ \sum_{k=1}^n x_k^2 y_k \end{bmatrix}$$

or

$$\begin{array}{ccccccc} n\hat{\beta}_0 & + & (\sum_k x_k)\hat{\beta}_1 & + & (\sum_k x_k^2)\hat{\beta}_2 & = & \sum_k y_k \\ (\sum_k x_k)\hat{\beta}_0 & + & (\sum_k x_k^2)\hat{\beta}_1 & + & (\sum_k x_k^3)\hat{\beta}_2 & = & \sum_k x_k y_k \\ (\sum_k x_k^2)\hat{\beta}_0 & + & (\sum_k x_k^3)\hat{\beta}_1 & + & (\sum_k x_k^4)\hat{\beta}_2 & = & \sum_k x_k^2 y_k. \end{array}$$

As stated near the end of 9.2, the solution $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ of the three equations just stated is the least-squares estimator of $(\beta_0, \beta_1, \beta_2)$ and $y = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$ is the least-squares parabola for the bivariate data in Assumptions 2.1.

(c) **ANOVA.** We will do this only for the special case in Example 9.9, and leave the derivation of the general case to the reader (see the last paragraph before 9.6 and picture 9.7 on the succeeding page).

The model $W$ as in 9.4 that we are fitting the data in Example 9.9 to has the form (see "ANOVA picture for Example 9.9" on the page after the statement of Example 9.9)

$$\begin{bmatrix} z_1 \\ z_1 \\ z_1 \\ z_1 \\ z_1 \\ z_2 \\ z_2 \\ z_2 \\ z_2 \\ z_2 \\ z_2 \\ z_3 \\ z_3 \\ z_3 \\ z_3 \\ z_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

thus we are again in the setting of Theorem APP.11, with

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad k = 3, \quad \text{and} \quad x^* = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}.$$

A relatively short calculation shows that

$$A^TA = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix} \quad \text{and} \quad A^T\vec{y} = \begin{bmatrix} \sum_{k=1}^{5} y_k \\ \sum_{k=6}^{10} y_k \\ \sum_{k=11}^{15} y_k \end{bmatrix},$$

so that the normal equations become

$$\begin{bmatrix} 5z_1 \\ 5z_2 \\ 5z_3 \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{5} y_k \\ \sum_{k=6}^{10} y_k \\ \sum_{k=11}^{15} y_k \end{bmatrix},$$

easily solvable as

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{5}\sum_{k=1}^{5} y_k \\ \frac{1}{5}\sum_{k=6}^{10} y_k \\ \frac{1}{5}\sum_{k=11}^{15} y_k \end{bmatrix} = \begin{bmatrix} \overline{x_1} \\ \overline{x_2} \\ \overline{x_3} \end{bmatrix},$$

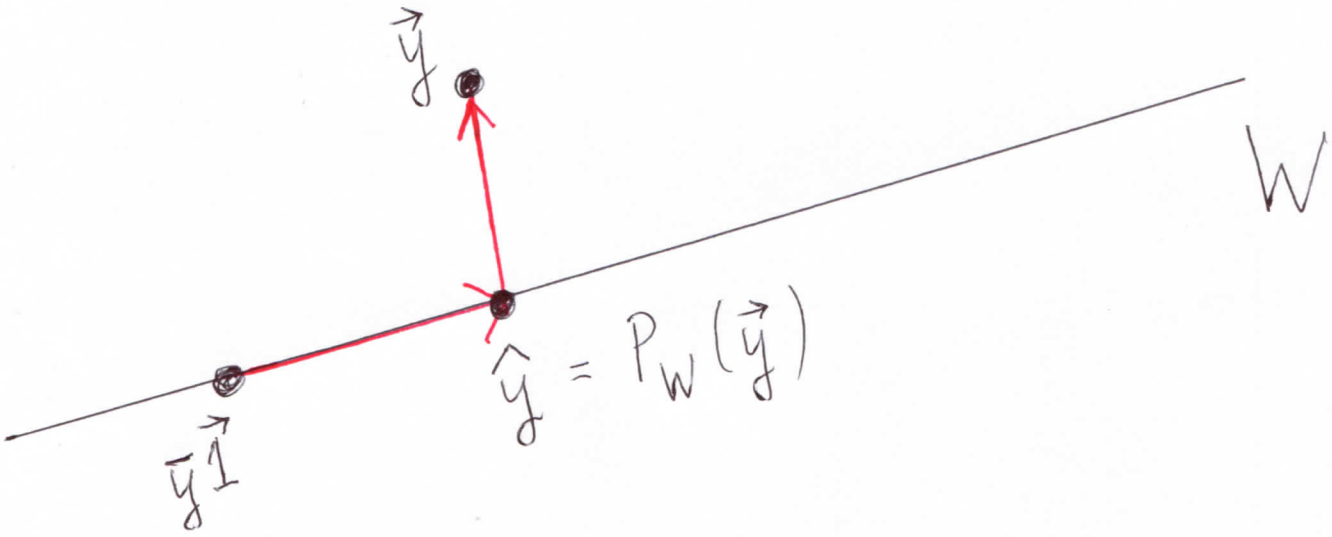
so that our least-squares approximation from $W$ is

$$A\begin{bmatrix} \overline{x_1} \\ \overline{x_2} \\ \overline{x_3} \end{bmatrix} = \begin{bmatrix} \overline{x_1} \\ \overline{x_1} \\ \overline{x_1} \\ \overline{x_1} \\ \overline{x_1} \\ \overline{x_2} \\ \overline{x_2} \\ \overline{x_2} \\ \overline{x_2} \\ \overline{x_2} \\ \overline{x_3} \\ \overline{x_3} \\ \overline{x_3} \\ \overline{x_3} \\ \overline{x_3} \end{bmatrix},$$

which, as a vector, is

$$(\overline{x_1}, \overline{x_1}, \overline{x_1}, \overline{x_1}, \overline{x_1}, \overline{x_2}, \overline{x_2}, \overline{x_2}, \overline{x_2}, \overline{x_2}, \overline{x_3}, \overline{x_3}, \overline{x_3}, \overline{x_3}, \overline{x_3})$$
$$= (70.6, 70.6, 70.6, 70.6, 70.6, 65.6, 65.6, 65.6, 65.6, 65.6, 70, 70, 70, 70, 70).$$

**Orthogonality and Pythagorean Theorem for Sums of Squares APP.13.** All our model fitting, as in 9.4, including linear regression and ANOVA, have the following picture, where $W$ is as in 9.4, $\vec{y}$ is the data and $\hat{y}$ is the fitted values, meaning the least-squares approximation of $\vec{y}$ from $W$ (see Theorem APP.10).



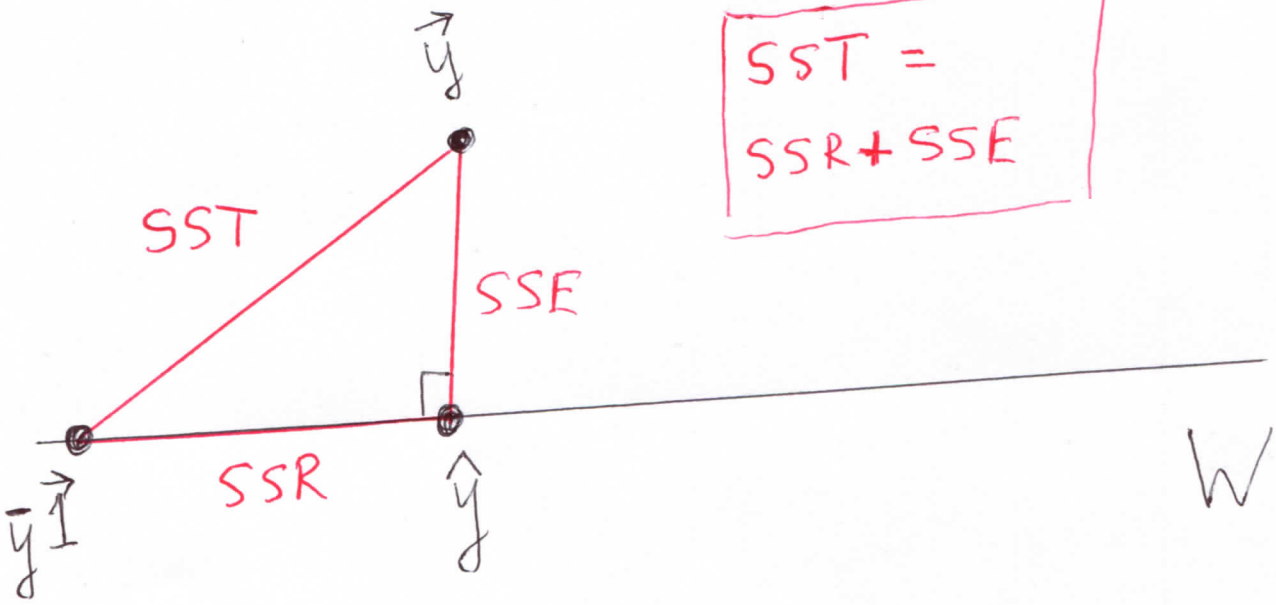

The definition of the orthogonal projection (Definition APP.9) now implies that the vectors drawn in red are perpendicular (also known as orthogonal); that is,

$$(\hat{y} - \overline{y}\vec{1}) \perp (\vec{y} - \hat{y}).$$

Since $SSR \equiv \|(\hat{y} - \overline{y}\vec{1})\|^2$ and $SSE \equiv \|\vec{y} - \hat{y}\|^2$, the Pythagorean Theorem APP.5 now implies that

$$SSR + SSE = SST \equiv \|(\hat{y} - \overline{y}\vec{1}) + (\vec{y} - \hat{y})\|^2 = \|(\vec{y} - \overline{y}\vec{1})\|^2.$$

See 4.3 and 9.7.



Our last application of linear algebra is unrelated to any other applications in this Appendix. We are motivated now by the presence of $(x^* - \overline{x})$ in the estimated standard error $s_{\hat{Y}}$ in Definitions 7.2, and the subsequent discussion in the last three paragraphs at the end of the Answer to Example 7.3(b); in particular, we would like now to state in what sense the number $x^*$ is closer to the data $x_1, x_2, \cdots, x_n$ when said number is closer to $\overline{x}$.

**Definition APP.14.** For $z_1$ and $z_2$ real numbers, we will say that $z_1$ is **closer to the data** $x_1, x_2, \ldots, x_n$ than $z_2$ if

$$\sum_{k=1}^{n}(x_k - z_1)^2 < \sum_{k=1}^{n}(x_k - z_2)^2.$$

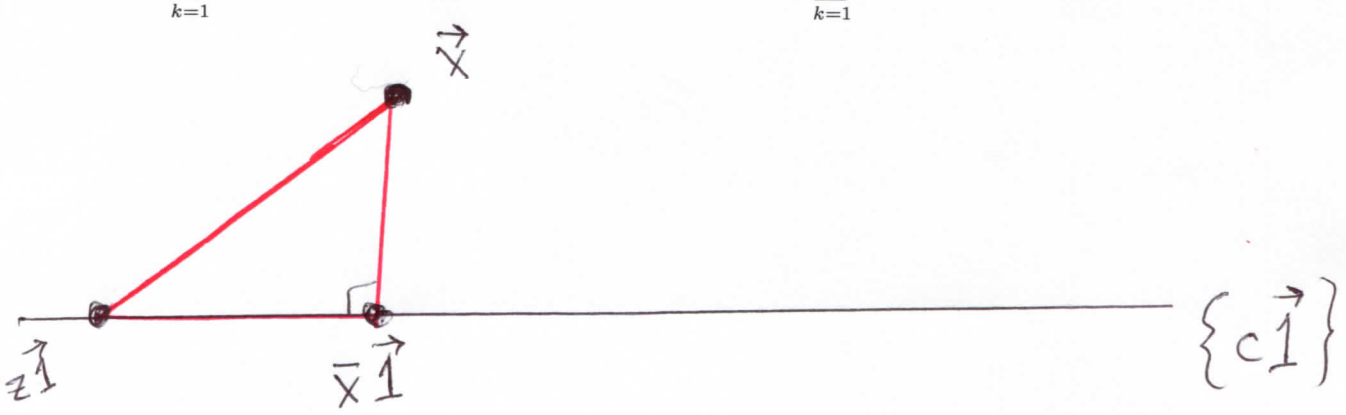
**Theorem APP.15.** $z_1$ is closer to the data, as in Definition APP.14, if and only if $z_1$ is closer to $\overline{x}$ than $z_2$; that is, $|\overline{x} - z_1| < |\overline{x} - z_2|$.

**Proof:** Let $\vec{x} \equiv (x_1, x_2, \ldots, x_n), \vec{1} \equiv (1, 1, \ldots, 1)$, the $n$-tuple whose components are all 1, and $W$ equal the subspace of real multiples of $\vec{1}$. We leave it to the reader to show that

$$\overline{x}\vec{1} = P_W(\vec{x}) \text{ (MUST SHOW that } ((\vec{x} - \overline{x}\vec{1}) \cdot c\vec{1}) = 0 \text{ for all real } c),$$

so that orthogonality and the Pythagorean theorem APP.5 imply that, for any real $z$,

$$\sum_{k=1}^{n}(x_k - z)^2 \equiv \|\vec{x} - z\vec{1}\|^2 = \|\vec{x} - \overline{x}\vec{1}\|^2 + \|\overline{x}\vec{1} - z\vec{1}\|^2 = \sum_{k=1}^{n}(x_k - \overline{x})^2 + n(\overline{x} - z)^2.$$




For real $z_1, z_2$, this implies that

$$\left[\sum_{k=1}^{n}(x_k - z_1)^2 - \sum_{k=1}^{n}(x_k - z_2)^2\right] = n\left[(\overline{x} - z_1)^2 - (\overline{x} - z_2)^2\right],$$

so that

$$\sum_{k=1}^{n}(x_k - z_1)^2 < \sum_{k=1}^{n}(x_k - z_2)^2 \quad \text{if and only if} \quad (\overline{x} - z_1)^2 < (\overline{x} - z_2)^2,$$

which is equivalent to $|\overline{x} - z_1| < |\overline{x} - z_2|$. □

**Remark APP.16.** A special case of Theorem APP.15 is the fact that

$$\sum_{k=1}^{n}(x_k - \overline{x})^2 < \sum_{k=1}^{n}(x_k - z)^2,$$

for any real $z \neq \overline{x}$; that is, $z = \overline{x}$ minimizes $\sum_{k=1}^{n}(x_k - z)^2$.

## HOMEWORK

See Chapter VIII for formulas needed in Problems 1–4.

**1.** By filling in the table below, get the least-squares line for the data

$$\{(-4,-7),(-2,0),(0,1),(0,3),(1,-1),(1,2),(2,1),(3,0),(4,1),(5,0)\}.$$

| $k$ | $x_k$ | $y_k$ | $x_k^2$ | $y_k^2$ | $x_k y_k$ |
|---|---|---|---|---|---|
| 1 | $-4$ | $-7$ | | | |
| 2 | $-2$ | 0 | | | |
| 3 | 0 | 1 | | | |
| 4 | 0 | 3 | | | |
| 5 | 1 | $-1$ | | | |
| 6 | 1 | 2 | | | |
| 7 | 2 | 1 | | | |
| 8 | 3 | 0 | | | |
| 9 | 4 | 1 | | | |
| 10 | 5 | 0 | | | |
| $\sum_k$ | | | | | |

In addition, for this data, get

(a) the sample correlation coefficient;

(b) the coefficient of determination;

(c) the sums of squares $SST$, $SSE$, and $SSR$, in Definition 4.2 and Theorem 5.3, by using the coefficient of determination from (b);

(d) $s$, our favorite estimator of $\sigma$ in the Simple Linear Regression Model 1.2.

**2.** Suppose grass growth, as a function of fertilizer, satisfies the Simple Linear Regression Model in Definition 1.2; that is, fertilizer is the predictor variable and grass growth is the response variable. We collect data on 18 plots of grass

$$\{(x_1, y_1), (x_2, y_2), \ldots (x_{18}, y_{18})\}$$

with $x$ measured in liters of fertilizer per acre and $y$ measured in centimeters per week and obtain the following summary data.

$$\sum_{k=1}^{18} x_k = 180, \sum_{k=1}^{18} x_k^2 = 2800, \sum_{k=1}^{18} y_k = 54, \sum_{k=1}^{18} y_k^2 = 170, \sum_{k=1}^{18} x_k y_k = 620.$$

(a) Find the least-squares estimators of the slope and $y$ intercept of the true regression line and the estimated regression line for the data.

(b) Find the coefficient of determination of the data.

(c) Find the sample correlation coefficient of the data.

(d) Find $SST, SSR$, and $SSE$ for the data.

(e) Get $s^2$, our favorite estimator of $\sigma^2$.

(f) Get a 99% confidence interval for the slope of the true regression line.

(g) Test, at significance level 0.01, whether there is a useful linear relationship between grass growth and fertilizer.

(h) Get a 95% confidence interval for the true number of centimeters that grass grows in a week (on average), when 20 liters of fertilizer per acre is used.

(i) Test, at significance level 0.1%, the claim that grass grows more than 3.9 centimeters in a week (on average), when 30 liters of fertilizer per acre is used.

(j) Test, at significance level 0.05, the claim that increasing the fertilizer per acre by one liter increases grass growth by more than 0.07 centimeters per week, on average.

See Example 6.5(d) or (e).

(k) Test, at significance level 0.05, the claim that increasing the fertilizer per acre by one liter increases grass growth by more than 0.06 centimeters per week, on average.

See Example 6.5(d) or (e).

**3.** Assume that happiness $Y$, as a function of pain $x$, satisfies the Simple Linear Regression Model in Definition 1.2.

We collect data

$$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\} = \{(-1, 8), (0, 5), (1, 0)), (2, 1)\}.$$

(a) Get the least-squares estimators of $\beta_0$ and $\beta_1$, in the Simple Linear Regression Model, and the least squares line, or estimated regression line.

(b) Get the coefficient of determination, measuring the proportion of observed variation in happiness that can be explained by its linear relationship with pain.

(c) Get $s^2$.

(d) Test, at significance level $\alpha = 0.1$, my belief that (on average) increasing pain decreases happiness. See Example 6.5(c).

(e) Get a 90% confidence interval for $(\beta_0 + 3\beta_1)$, the expected average happiness with a pain of 3.

(f) Test $H_0 : (\beta_0 - 2\beta_1) = 1$ versus $H_a : (\beta_0 - 2\beta_1) > 1$, at significance level 0.001.

**4.** Suppose we do regression on data

$$\{(0, 1), (1, 5), (3, 4), (0, 2), (1, 2)\}$$

and get an estimated regression line of $y = 1.97 + (0.833)x$.

(a) Get the fitted values (for $x = 0, 1, 3$).

(b) Get the (five) residuals.

(c) Get $SSE$.

(d) Get $s^2$, our favorite estimator of $\sigma^2$, in the Simple Linear Regression Model.

**5.** Given bivariate data as in Definition 1.1 find equations, analogous to those appearing at the end of 9.2, whose solutions $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ minimize the sum of squares of vertical displacements

$$SSV(b_0, b_1, b_2, b_3) \equiv \sum_{k=1}^{n} \left[y_k - (b_0 + b_1 x_k + b_2 x_k^2 + b_3 x_k^3)\right]^2;$$

that is,

$$SSV(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) \leq SSV(b_0, b_1, b_2, b_3)$$

for all real numbers $b_0, b_1, b_2, b_3$.

This may be done with matrix methods, as in Examples APP.12 in the Appendix, or by noticing the pattern in going from linear to quadratic regression, in the sets of equations at the end of 9.2, and extending said pattern to polynomials of degree three.

## HOMEWORK ANSWERS

Answers may differ because of different rounding.

**1.**

| $k$ | $x_k$ | $y_k$ | $x_k^2$ | $y_k^2$ | $x_k y_k$ |
|---|---|---|---|---|---|
| 1 | −4 | −7 | 16 | 49 | 28 |
| 2 | −2 | 0 | 4 | 0 | 0 |
| 3 | 0 | 1 | 0 | 1 | 0 |
| 4 | 0 | 3 | 0 | 9 | 0 |
| 5 | 1 | −1 | 1 | 1 | −1 |
| 6 | 1 | 2 | 1 | 4 | 2 |
| 7 | 2 | 1 | 4 | 1 | 2 |
| 8 | 3 | 0 | 9 | 0 | 0 |
| 9 | 4 | 1 | 16 | 1 | 4 |
| 10 | 5 | 0 | 25 | 0 | 0 |
| $\sum_k$ | 10 | 0 | 76 | 66 | 35 |

Using the "computational formulas"

$$S_{\vec{x},\vec{y}} = 35 - \frac{1}{10}(10)(0) = 35,\ S_{\vec{y},\vec{y}} = 66 - \frac{1}{10}(0)^2 = 66,\ S_{\vec{x},\vec{x}} = 76 - \frac{1}{10}(10)^2 = 66,\ \overline{x} = 1,\ \overline{y} = 0.$$

So let's get

$$\hat{\beta}_1 = \frac{S_{\vec{x},\vec{y}}}{S_{\vec{x},\vec{x}}} = \frac{35}{66},\ \hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = 0 - (\frac{35}{66})(1) = -\frac{35}{66},$$

thus our least-squares line is

$$y = -\frac{35}{66} + \frac{35}{66}x = \frac{35}{66}(x-1).$$

Now let's answer questions (a)–(d).

(a) $r = \frac{S_{\vec{x},\vec{y}}}{\sqrt{S_{\vec{x},\vec{x}}S_{\vec{y},\vec{y}}}} = \frac{35}{66} \sim 0.530.$

(b) $r^2 = (\frac{35}{66})^2 \sim 0.281.$

(c)

$$SST = S_{\vec{y},\vec{y}} = 66,\ SSE = (1-r^2)S_{\vec{y},\vec{y}} = (1-(\frac{35}{66})^2)66 = \frac{3,131}{66} \sim 47.440,$$

$$SSR = r^2 S_{\vec{y},\vec{y}} = (\frac{35}{66})^2 66 = \frac{1,225}{66} \sim 18.561.$$

(d) $s = \sqrt{\frac{SSE}{(n-2)}} = \sqrt{\frac{\frac{3,131}{66}}{(10-2)}} = \sqrt{\frac{3,131}{528}} \sim 2.435.$

**2.** Again we begin with "computational formulas":

$$S_{\vec{x},\vec{y}} = 620 - \frac{1}{18}(180)(54) = 80;\ \ S_{\vec{x},\vec{x}} = 2800 - \frac{1}{18}(180)^2 = 1,000;\ \ S_{\vec{y},\vec{y}} = 170 - \frac{1}{18}(54)^2 = 8;$$
$$\overline{y} = 3, \overline{x} = 10.$$

(a) $\hat{\beta}_1 = \frac{80}{1,000} = 0.08; \hat{\beta}_0 = 3 - (0.08)(10) = 2.2$; our estimated regression line is

$$y = 2.2 + (0.08)x.$$

(b) $r^2 = \frac{(80)^2}{(1,000)(8)} = 0.8.$

(c) $r = \sqrt{0.8} \sim 0.894.$

(d) $SST = S_{\vec{y},\vec{y}} = 8;\ \ SSR = 8(0.8) = 6.4;\ \ SSE = (8)(1 - 0.8) = 1.6.$

(e) $s^2 = \frac{1.6}{18-2} = 0.1.$

(f) We need

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{\vec{x},\vec{x}}}} = \sqrt{\frac{0.1}{1,000}} = 0.01.$$

We also need the critical value $t_{0.005,16} = 2.921$.

Our interval is

$$\hat{\beta}_1 \pm (2.921)(s_{\hat{\beta}_1}) = 0.08 \pm (2.921)(0.01) \sim 0.08 \pm 0.029 = (0.051, 0.109).$$

(g) This is testing (see Definition 6.4)

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_a : \beta_1 \neq 0.$$

Our statistic is

$$t = \frac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}} = \frac{0.08}{0.01} = 8,$$

thus our P-value is

$$P(|T_{16}| > 8) = 2P(T_{16} > 8) < 2P(T_{16} > 3) = 2(0.004) \leq 0.01 = \alpha,$$

thus we reject $H_0$, and conclude that there is a useful linear relationship between grass growth and fertilizer.

(h) Denote, with $x^* \equiv 20$,

$$\mu_{Y \cdot x^*} \equiv \beta_0 + \beta_1(20), \quad \hat{Y} \equiv \hat{\beta}_0 + \hat{\beta}_1(20) = 2.2 + (0.08)(20) = 3.8;$$

we also need $t_{0.025,16} = 2.120$, and

$$s_{\hat{Y}} = s\sqrt{\frac{1}{n} + \frac{(20 - \overline{x})^2}{S_{\vec{x},\vec{x}}}} = \sqrt{0.1}\sqrt{\frac{1}{18} + \frac{(20 - 10)^2}{1,000}} \sim 0.125.$$

Our confidence interval is now

$$\hat{Y} \pm t_{\frac{\alpha}{2},16} s_{\hat{Y}} \sim 3.8 \pm 2.120(0.125) = (3.535, 4.065).$$

(i) Now we have $x^* = 30$. We are testing

$$H_0 : \mu_{Y \cdot x^*} = 3.9 \quad \text{versus} \quad H_a : \mu_{Y \cdot x^*} > 3.9.$$

As with (h), define

$$\hat{Y} \equiv \hat{\beta}_0 + \hat{\beta}_1(30) = 2.2 + (0.08)(30) = 4.6, s_{\hat{Y}} = s\sqrt{\frac{1}{n} + \frac{(30-\overline{x})^2}{S_{\vec{x},\vec{x}}}} = \sqrt{0.1}\sqrt{\frac{1}{18} + \frac{(30-10)^2}{1,000}} \sim 0.213.$$

Our test statistic, with a $t_{16}$ distribution, is

$$t = \frac{\hat{Y} - 3.9}{s_{\hat{Y}}} \sim \frac{4.6 - 3.9}{0.213} \sim 3.3$$

so that our P-value is

$$P(T_{16} > 3.3) = 0.002 > 0.001,$$

thus we do not reject $H_0$; there is insufficient data to conclude that grass grows more than 3.9 centimeters per week, on average, with significance 0.1%.

(j) $H_0 : \beta_1 = 0.07$ versus $H_a : \beta_1 > 0.07$
$t = \frac{\hat{\beta}_1 - 0.07}{s_{\hat{\beta}_1}} = \frac{0.08-0.07}{0.01} = 1$, so

$$\text{P-value} = P(T_{16} > 1) = 0.166 > 0.05,$$

so we don't reject $H_0$; there is insufficient evidence to conclude that increasing the fertilizer per acre by one liter increases grass growth by more than 0.07 centimeters per week, on average.

(k) $H_0 : \beta_1 = 0.06$ versus $H_a : \beta_1 > 0.06$
$t = \frac{\hat{\beta}_1 - 0.06}{s_{\hat{\beta}_1}} = \frac{0.08-0.06}{0.01} = 2$, so

$$\text{P-value} = P(T_{16} > 2) = 0.031 \leq 0.05,$$

so we reject $H_0$; there is sufficient evidence to conclude that increasing the fertilizer per acre by one liter increases grass growth by more than 0.06 centimeters per week, on average.

**3.** Calculate, from the table below (using "computational formulas"), that

$$\overline{x} = 0.5, \quad \overline{y} = 3.5, \quad S_{xx} = 5, \quad S_{yy} = 41, \quad S_{xy} = -13.$$

| $k$ | $x_k$ | $y_k$ | $x_k^2$ | $y_k^2$ | $x_k y_k$ |
|---|---|---|---|---|---|
| 1 | $-1$ | 8 | 1 | 64 | $-8$ |
| 2 | 0 | 5 | 0 | 25 | 0 |
| 3 | 1 | 0 | 1 | 0 | 0 |
| 4 | 2 | 1 | 4 | 1 | 2 |
| $\sum_k$ | 2 | 14 | 6 | 90 | $-6$ |

(a) Our least-squares estimators are $\hat{\beta}_1 = \frac{-13}{5} = -2.6, \hat{\beta}_0 = 3.5 - (-2.6)(0.5) = 4.8$, thus our estimated regression line is

$$y = 4.8 - 2.6x.$$

(b) $r^2 = \frac{(-13)^2}{5 \times 41} = \frac{169}{205} \sim 0.824.$

(c) $SSE = (1 - r^2) S_{\vec{y},\vec{y}} = (1 - \frac{169}{205})(41) = \frac{36 \times 41}{205} = \frac{36}{5} = 7.2$, thus

$$s^2 = \frac{7.2}{4 - 2} = 3.6.$$

(d) We are testing, at significance level $\alpha = 0.1$,

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_1 : \beta_1 < 0.$$

We need

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{\vec{x},\vec{x}}}} = \sqrt{\frac{3.6}{5}} = \sqrt{0.72},$$

thus our test statistic is

$$t = \frac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}} = \frac{-2.6}{\sqrt{0.72}} \sim -3.1,$$

so our P-value is $\sim$

$$P(T_2 < -3.1) = P(T_2 > 3.1) = 0.045 \leq 0.1 \equiv \alpha,$$

so we reject $H_0$; at significance level 0.1, our data suggests that increasing pain decreases happiness.

(e) Here $x^* = 3$, so our estimator of $(\beta_0 + \beta_1 x^*)$ is

$$\hat{Y} \equiv (\hat{\beta}_0 + \hat{\beta}_1 x^*) = 4.8 + (-2.6)3 = -3,$$

and our confidence interval is

$$\hat{Y} \pm t_{0.05,2} s_{\hat{Y}} = -3 \pm 2.920 \sqrt{s^2 \left[ \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{\vec{x},\vec{x}}} \right]} = -3 \pm 2.920 \sqrt{3.6 \left[ \frac{1}{4} + \frac{(3 - 0.5)^2}{5} \right]}$$

$$= -3 \pm 2.920 \sqrt{5.4} \sim -3 \pm 6.79 = (-9.79, 3.79)$$

(f) Now we have $x^* = -2$, with inference on $(\beta_0 + \beta_1 x^*)$. Our estimator is

$$\hat{Y} \equiv \hat{\beta}_0 + \hat{\beta}_1 x^* = 4.8 + (-2.6)(-2) = 10,$$

with

$$s_{\hat{Y}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{\vec{x},\vec{x}}}} = \sqrt{3.6}\sqrt{\frac{1}{4} + \frac{(-2 - 0.5)^2}{5}} = \sqrt{5.4},$$

thus our test statistic is

$$t = \frac{10 - 1}{\sqrt{5.4}} \sim 3.9,$$

so our P-value is $\sim$

$$P(T_2 > 3.9) = 0.030 > 0.001 \equiv \alpha,$$

thus we do not reject $H_0$; the data is insufficient to conclude that $(\beta_0 - 2\beta_1) > 1$.

**4.** (a) Writing "fit" as shorthand for "fitted value," we'll mimic Example 4.6.

(a) Plugging $x$ into the estimated regression line and writing "fit" for "fitted value":

| $x$ | 0 | 1 | 3 |
|---|---|---|---|
| fit | 1.97 | 2.803 | 4.469 |

(b) Denoting

$$(x_1, y_1) \equiv (0, 1), (x_2, y_2) \equiv (1, 5), (x_3, y_3) \equiv (3, 4), (x_4, y_4) \equiv (0, 2), (x_5, y_5) \equiv (1, 2),$$

we have fitted values from (a)

$$\hat{y_1} = 1.97, \hat{y_2} = 2.803, \hat{y_3} = 4.469, \hat{y_4} = 1.97, \hat{y_5} = 2.803.$$

Here are the desired residuals:

$$\epsilon_1 \equiv (y_1 - \hat{y_1}) = (1 - 1.97) = -0.97, \epsilon_2 \equiv (y_2 - \hat{y_2}) = (5 - 2.803) = 2.197, \epsilon_3 \equiv (y_3 - \hat{y_3}) = (4 - 4.469) = -0.469,$$
$$\epsilon_4 \equiv (y_4 - \hat{y_4}) = (2 - 1.97) = 0.03, \epsilon_5 \equiv (y_5 - \hat{y_5}) = (2 - 2.803) = -0.803.$$

(c) $SSE$ equals

$$\sum_{k=1}^{5} (\epsilon_k)^2 = (-0.97)^2 + (2.197)^2 + (-0.469)^2 + (0.03)^2 + (-0.803)^2 \sim 6.633.$$

(d)

$$s^2 = \frac{1}{(n-2)} SSE = \frac{1}{(5-2)} \left[(-0.97)^2 + (2.197)^2 + (-0.469)^2 + (0.03)^2 + (-0.803)^2\right] \sim 2.211.$$

**5.**

$$\begin{array}{ccccccccc}
n\hat{\beta}_0 & + & (\sum_k x_k)\hat{\beta}_1 & + & (\sum_k x_k^2)\hat{\beta}_2 & + & (\sum_k x_k^3)\hat{\beta}_3 & = & \sum_k y_k \\
(\sum_k x_k)\hat{\beta}_0 & + & (\sum_k x_k^2)\hat{\beta}_1 & + & (\sum_k x_k^3)\hat{\beta}_2 & + & (\sum_k x_k^4)\hat{\beta}_3 & = & \sum_k x_k y_k \\
(\sum_k x_k^2)\hat{\beta}_0 & + & (\sum_k x_k^3)\hat{\beta}_1 & + & (\sum_k x_k^4)\hat{\beta}_2 & + & (\sum_k x_k^5)\hat{\beta}_3 & = & \sum_k x_k^2 y_k \\
(\sum_k x_k^3)\hat{\beta}_0 & + & (\sum_k x_k^4)\hat{\beta}_1 & + & (\sum_k x_k^5)\hat{\beta}_2 & + & (\sum_k x_k^6)\hat{\beta}_3 & = & \sum_k x_k^3 y_k
\end{array}$$

## REFERENCES

**1.** R. deLaubenfels, "The Victory of Least Squares and Orthogonality in Statistics," The Amer. Statistician 60 (2006), 315–321.

**2.** R. deLaubenfels, "Linear Algebra," or E Pluribus Unum, https://teacherscholarinstitute.com/FreeMathBooksHighschool.html (2017).

**3.** R. deLaubenfels, "Statistics Introduction Magnification," https://www.teacherscholarinstitute.com/MathMagnificationsReadyToUse.html.

**4.** R. deLaubenfels, "Statistics: Hypothesis Testing Magnification," https://www.teacherscholarinstitute.com/MathMagnificationsReadyToUse.html.

**5.** R. deLaubenfels, "Statistical Inference on Mean and Proportion Magnification," https://www.teacherscholarinstitute.com/MathMagnificationsReadyToUse.html.

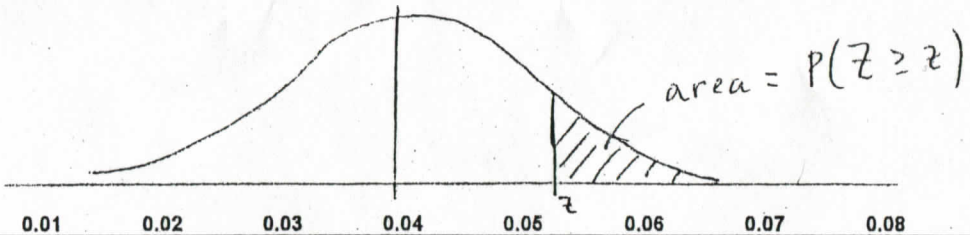
**6.** J. L. Devore, "Probability and Statistics for Engineering and the Sciences," Brooks/Cole, eighth edition, 2012.

**7.** A. Hald, "A History of Parametric Statistical Inference From Bernoulli to Fisher, 1713–1935," Springer, 2007.

**8.** J. Saxon, "Algebra 1. An Incremental Development," Second Edition, Saxon Publishers, Inc., 1990.

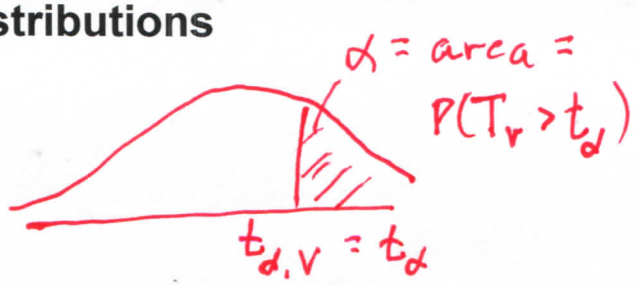
**9.** S. M. Stigler, "The History of Statistics: The Measurement of Uncertainty before 1900," the Belknap Press of Harvard University Press, Cambridge, MA, 1986.

# The Standard Normal Distribution (Areas in the Right Tail)



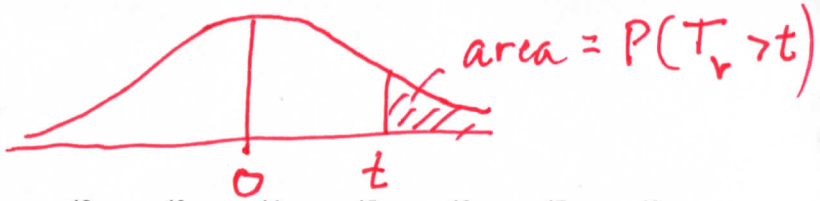

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.4960 | 0.4920 | 0.4880 | 0.4840 | 0.4801 | 0.4761 | 0.4721 | 0.4681 | 0.4641 |
| 0.1 | 0.4602 | 0.4562 | 0.4522 | 0.4483 | 0.4443 | 0.4404 | 0.4364 | 0.4325 | 0.4286 | 0.4247 |
| 0.2 | 0.4207 | 0.4168 | 0.4129 | 0.4090 | 0.4052 | 0.4013 | 0.3974 | 0.3936 | 0.3897 | 0.3859 |
| 0.3 | 0.3821 | 0.3783 | 0.3745 | 0.3707 | 0.3669 | 0.3632 | 0.3594 | 0.3557 | 0.3520 | 0.3483 |
| 0.4 | 0.3446 | 0.3409 | 0.3372 | 0.3336 | 0.3300 | 0.3264 | 0.3228 | 0.3192 | 0.3156 | 0.3121 |
| 0.5 | 0.3085 | 0.3050 | 0.3015 | 0.2981 | 0.2946 | 0.2912 | 0.2877 | 0.2843 | 0.2810 | 0.2776 |
| 0.6 | 0.2743 | 0.2709 | 0.2676 | 0.2643 | 0.2611 | 0.2578 | 0.2546 | 0.2514 | 0.2483 | 0.2451 |
| 0.7 | 0.2420 | 0.2389 | 0.2358 | 0.2327 | 0.2296 | 0.2266 | 0.2236 | 0.2206 | 0.2177 | 0.2148 |
| 0.8 | 0.2119 | 0.2090 | 0.2061 | 0.2033 | 0.2005 | 0.1977 | 0.1949 | 0.1922 | 0.1894 | 0.1867 |
| 0.9 | 0.1841 | 0.1814 | 0.1788 | 0.1762 | 0.1736 | 0.1711 | 0.1685 | 0.1660 | 0.1635 | 0.1611 |
| 1.0 | 0.1587 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 | 0.1379 |
| 1.1 | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 | 0.1170 |
| 1.2 | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 | 0.0985 |
| 1.3 | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 | 0.0823 |
| 1.4 | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0721 | 0.0708 | 0.0694 | 0.0681 |
| 1.5 | 0.0668 | 0.0655 | 0.0643 | 0.0630 | 0.0618 | 0.0606 | 0.0594 | 0.0582 | 0.0571 | 0.0559 |
| 1.6 | 0.0548 | 0.0537 | 0.0526 | 0.0516 | 0.0505 | 0.0495 | 0.0485 | 0.0475 | 0.0465 | 0.0455 |
| 1.7 | 0.0446 | 0.0436 | 0.0427 | 0.0418 | 0.0409 | 0.0401 | 0.0392 | 0.0384 | 0.0375 | 0.0367 |
| 1.8 | 0.0359 | 0.0351 | 0.0344 | 0.0336 | 0.0329 | 0.0322 | 0.0314 | 0.0307 | 0.0301 | 0.0294 |
| 1.9 | 0.0287 | 0.0281 | 0.0274 | 0.0268 | 0.0262 | 0.0256 | 0.0250 | 0.0244 | 0.0239 | 0.0233 |
| 2.0 | 0.0228 | 0.0222 | 0.0217 | 0.0212 | 0.0207 | 0.0202 | 0.0197 | 0.0192 | 0.0188 | 0.0183 |
| 2.1 | 0.0179 | 0.0174 | 0.0170 | 0.0166 | 0.0162 | 0.0158 | 0.0154 | 0.0150 | 0.0146 | 0.0143 |
| 2.2 | 0.0139 | 0.0136 | 0.0132 | 0.0129 | 0.0125 | 0.0122 | 0.0119 | 0.0116 | 0.0113 | 0.0110 |
| 2.3 | 0.0107 | 0.0104 | 0.0102 | 0.0099 | 0.0096 | 0.0094 | 0.0091 | 0.0089 | 0.0087 | 0.0084 |
| 2.4 | 0.0082 | 0.0080 | 0.0078 | 0.0075 | 0.0073 | 0.0071 | 0.0069 | 0.0068 | 0.0066 | 0.0064 |
| 2.5 | 0.0062 | 0.0060 | 0.0059 | 0.0057 | 0.0055 | 0.0054 | 0.0052 | 0.0051 | 0.0049 | 0.0048 |
| 2.6 | 0.0047 | 0.0045 | 0.0044 | 0.0043 | 0.0041 | 0.0040 | 0.0039 | 0.0038 | 0.0037 | 0.0036 |
| 2.7 | 0.0035 | 0.0034 | 0.0033 | 0.0032 | 0.0031 | 0.0030 | 0.0029 | 0.0028 | 0.0027 | 0.0026 |
| 2.8 | 0.0026 | 0.0025 | 0.0024 | 0.0023 | 0.0023 | 0.0022 | 0.0021 | 0.0021 | 0.0020 | 0.0019 |
| 2.9 | 0.0019 | 0.0018 | 0.0018 | 0.0017 | 0.0016 | 0.0016 | 0.0015 | 0.0015 | 0.0014 | 0.0014 |
| 3.0 | 0.0013 | 0.0013 | 0.0013 | 0.0012 | 0.0012 | 0.0011 | 0.0011 | 0.0011 | 0.0010 | 0.0010 |

# Critical Values for t Distributions




| ν \ α | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 0.0005 |
|---|---|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 318.310 | 636.620 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.326 | 31.598 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.213 | 12.924 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 | 4.781 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 | 4.318 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 | 4.221 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.140 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.850 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 | 3.819 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 | 3.792 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 | 3.767 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 | 3.745 |
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 | 3.725 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 | 3.707 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 | 3.690 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 | 3.674 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 | 3.659 |
| 30 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 | 3.646 |
| 32 | 1.309 | 1.694 | 2.037 | 2.449 | 2.738 | 3.365 | 3.622 |
| 34 | 1.307 | 1.691 | 2.032 | 2.441 | 2.728 | 3.348 | 3.601 |
| 36 | 1.306 | 1.688 | 2.028 | 2.434 | 2.719 | 3.333 | 3.582 |
| 38 | 1.304 | 1.686 | 2.024 | 2.429 | 2.712 | 3.319 | 3.566 |
| 40 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 | 3.551 |
| 50 | 1.299 | 1.676 | 2.009 | 2.403 | 2.678 | 3.262 | 3.496 |
| 60 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 | 3.460 |
| 120 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 | 3.373 |
| ∞ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.291 |

# t Curve Tail Areas




| | ν | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **t** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** |
| **0.0** | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 |
| **0.1** | 0.468 | 0.465 | 0.463 | 0.463 | 0.462 | 0.462 | 0.462 | 0.461 | 0.461 | 0.461 | 0.461 | 0.461 | 0.461 | 0.461 | 0.461 | 0.461 | 0.461 | 0.461 |
| **0.2** | 0.437 | 0.430 | 0.427 | 0.426 | 0.425 | 0.424 | 0.424 | 0.423 | 0.423 | 0.423 | 0.423 | 0.422 | 0.422 | 0.422 | 0.422 | 0.422 | 0.422 | 0.422 |
| **0.3** | 0.407 | 0.396 | 0.392 | 0.390 | 0.388 | 0.387 | 0.386 | 0.386 | 0.386 | 0.385 | 0.385 | 0.385 | 0.384 | 0.384 | 0.384 | 0.384 | 0.384 | 0.384 |
| **0.4** | 0.379 | 0.364 | 0.358 | 0.355 | 0.353 | 0.352 | 0.351 | 0.350 | 0.349 | 0.349 | 0.348 | 0.348 | 0.348 | 0.347 | 0.347 | 0.347 | 0.347 | 0.347 |
| **0.5** | 0.352 | 0.333 | 0.326 | 0.322 | 0.319 | 0.317 | 0.316 | 0.315 | 0.315 | 0.314 | 0.313 | 0.313 | 0.313 | 0.312 | 0.312 | 0.312 | 0.312 | 0.312 |
| **0.6** | 0.328 | 0.305 | 0.295 | 0.290 | 0.287 | 0.285 | 0.284 | 0.283 | 0.282 | 0.281 | 0.280 | 0.280 | 0.279 | 0.279 | 0.279 | 0.278 | 0.278 | 0.278 |
| **0.7** | 0.306 | 0.278 | 0.267 | 0.261 | 0.258 | 0.255 | 0.253 | 0.252 | 0.251 | 0.250 | 0.249 | 0.249 | 0.248 | 0.247 | 0.247 | 0.247 | 0.247 | 0.246 |
| **0.8** | 0.285 | 0.254 | 0.241 | 0.234 | 0.230 | 0.227 | 0.225 | 0.223 | 0.222 | 0.221 | 0.220 | 0.220 | 0.219 | 0.218 | 0.218 | 0.218 | 0.217 | 0.217 |
| **0.9** | 0.267 | 0.232 | 0.217 | 0.210 | 0.205 | 0.201 | 0.199 | 0.197 | 0.196 | 0.195 | 0.194 | 0.193 | 0.192 | 0.191 | 0.191 | 0.191 | 0.190 | 0.190 |
| **1.0** | 0.250 | 0.211 | 0.196 | 0.187 | 0.182 | 0.178 | 0.175 | 0.173 | 0.172 | 0.170 | 0.169 | 0.169 | 0.168 | 0.167 | 0.167 | 0.166 | 0.166 | 0.165 |
| **1.1** | 0.235 | 0.193 | 0.176 | 0.167 | 0.162 | 0.157 | 0.154 | 0.152 | 0.150 | 0.149 | 0.147 | 0.146 | 0.146 | 0.144 | 0.144 | 0.144 | 0.143 | 0.143 |
| **1.2** | 0.221 | 0.177 | 0.158 | 0.148 | 0.142 | 0.138 | 0.135 | 0.132 | 0.130 | 0.129 | 0.128 | 0.127 | 0.126 | 0.124 | 0.124 | 0.124 | 0.123 | 0.123 |
| **1.3** | 0.209 | 0.162 | 0.142 | 0.132 | 0.125 | 0.121 | 0.117 | 0.115 | 0.113 | 0.111 | 0.110 | 0.109 | 0.108 | 0.107 | 0.107 | 0.106 | 0.105 | 0.105 |
| **1.4** | 0.197 | 0.148 | 0.128 | 0.117 | 0.110 | 0.106 | 0.102 | 0.100 | 0.098 | 0.096 | 0.095 | 0.093 | 0.092 | 0.091 | 0.091 | 0.090 | 0.090 | 0.089 |
| **1.5** | 0.187 | 0.136 | 0.115 | 0.104 | 0.097 | 0.092 | 0.089 | 0.086 | 0.084 | 0.082 | 0.081 | 0.080 | 0.079 | 0.077 | 0.077 | 0.077 | 0.076 | 0.075 |
| **1.6** | 0.178 | 0.125 | 0.104 | 0.092 | 0.085 | 0.080 | 0.077 | 0.074 | 0.072 | 0.070 | 0.069 | 0.068 | 0.067 | 0.065 | 0.065 | 0.065 | 0.064 | 0.064 |
| **1.7** | 0.169 | 0.116 | 0.094 | 0.082 | 0.075 | 0.070 | 0.065 | 0.064 | 0.062 | 0.060 | 0.059 | 0.057 | 0.056 | 0.055 | 0.055 | 0.054 | 0.054 | 0.053 |
| **1.8** | 0.161 | 0.107 | 0.085 | 0.073 | 0.066 | 0.061 | 0.057 | 0.055 | 0.053 | 0.051 | 0.050 | 0.049 | 0.048 | 0.046 | 0.046 | 0.045 | 0.045 | 0.044 |
| **1.9** | 0.154 | 0.099 | 0.077 | 0.065 | 0.058 | 0.053 | 0.050 | 0.047 | 0.045 | 0.043 | 0.042 | 0.041 | 0.040 | 0.038 | 0.038 | 0.038 | 0.037 | 0.037 |
| **2.0** | 0.148 | 0.092 | 0.070 | 0.058 | 0.051 | 0.046 | 0.043 | 0.040 | 0.038 | 0.037 | 0.035 | 0.034 | 0.033 | 0.032 | 0.032 | 0.031 | 0.031 | 0.030 |
| **2.1** | 0.141 | 0.085 | 0.063 | 0.052 | 0.045 | 0.040 | 0.037 | 0.034 | 0.033 | 0.031 | 0.030 | 0.029 | 0.028 | 0.027 | 0.027 | 0.026 | 0.025 | 0.025 |
| **2.2** | 0.136 | 0.079 | 0.058 | 0.046 | 0.040 | 0.035 | 0.032 | 0.029 | 0.028 | 0.026 | 0.025 | 0.024 | 0.023 | 0.022 | 0.022 | 0.021 | 0.021 | 0.021 |
| **2.3** | 0.131 | 0.074 | 0.052 | 0.041 | 0.035 | 0.031 | 0.027 | 0.025 | 0.023 | 0.022 | 0.021 | 0.020 | 0.019 | 0.018 | 0.018 | 0.018 | 0.017 | 0.017 |
| **2.4** | 0.126 | 0.069 | 0.048 | 0.037 | 0.031 | 0.027 | 0.024 | 0.022 | 0.020 | 0.019 | 0.018 | 0.017 | 0.016 | 0.015 | 0.015 | 0.014 | 0.014 | 0.014 |
| **2.5** | 0.121 | 0.065 | 0.044 | 0.033 | 0.027 | 0.023 | 0.020 | 0.018 | 0.017 | 0.016 | 0.015 | 0.014 | 0.013 | 0.012 | 0.012 | 0.012 | 0.011 | 0.011 |
| **2.6** | 0.117 | 0.061 | 0.040 | 0.030 | 0.024 | 0.020 | 0.018 | 0.016 | 0.014 | 0.013 | 0.012 | 0.012 | 0.011 | 0.010 | 0.010 | 0.010 | 0.009 | 0.009 |
| **2.7** | 0.113 | 0.057 | 0.037 | 0.027 | 0.021 | 0.018 | 0.015 | 0.014 | 0.012 | 0.011 | 0.010 | 0.010 | 0.009 | 0.008 | 0.008 | 0.008 | 0.008 | 0.007 |
| **2.8** | 0.109 | 0.054 | 0.034 | 0.024 | 0.019 | 0.016 | 0.013 | 0.012 | 0.010 | 0.009 | 0.009 | 0.008 | 0.008 | 0.007 | 0.007 | 0.006 | 0.006 | 0.006 |
| **2.9** | 0.106 | 0.051 | 0.031 | 0.022 | 0.017 | 0.014 | 0.011 | 0.010 | 0.009 | 0.008 | 0.007 | 0.007 | 0.006 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| **3.0** | 0.102 | 0.048 | 0.029 | 0.020 | 0.015 | 0.012 | 0.010 | 0.009 | 0.007 | 0.007 | 0.006 | 0.006 | 0.005 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |
| **3.1** | 0.099 | 0.045 | 0.027 | 0.018 | 0.013 | 0.011 | 0.009 | 0.007 | 0.006 | 0.006 | 0.005 | 0.005 | 0.004 | 0.004 | 0.004 | 0.003 | 0.003 | 0.003 |
| **3.2** | 0.096 | 0.043 | 0.025 | 0.016 | 0.012 | 0.009 | 0.008 | 0.006 | 0.005 | 0.005 | 0.004 | 0.004 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.002 |
| **3.3** | 0.094 | 0.040 | 0.023 | 0.015 | 0.011 | 0.008 | 0.007 | 0.005 | 0.005 | 0.004 | 0.004 | 0.003 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| **3.4** | 0.091 | 0.038 | 0.021 | 0.014 | 0.010 | 0.007 | 0.006 | 0.005 | 0.004 | 0.003 | 0.003 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| **3.5** | 0.089 | 0.036 | 0.020 | 0.012 | 0.009 | 0.006 | 0.005 | 0.004 | 0.003 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 |
| **3.6** | 0.086 | 0.035 | 0.018 | 0.011 | 0.008 | 0.006 | 0.004 | 0.004 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| **3.7** | 0.084 | 0.033 | 0.017 | 0.010 | 0.007 | 0.005 | 0.004 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| **3.8** | 0.082 | 0.031 | 0.016 | 0.010 | 0.006 | 0.004 | 0.003 | 0.003 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| **3.9** | 0.080 | 0.030 | 0.015 | 0.009 | 0.006 | 0.004 | 0.003 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| **4.0** | 0.078 | 0.029 | 0.014 | 0.008 | 0.005 | 0.004 | 0.003 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.000 | 0.000 |