Attendance Workbook

For Statistics 213

Probability and Decision Theory

Spring 2010

by

Jonathan Kuhn, Ph.D.
Associate Professor of Statistics,
Mathematics, Statistics and Physics,
Purdue University North Central

Preface

This course provides an introduction to a few finite mathematical topics, including, linear programming, probability and statistics.

This workbook is a necessary component for a student to successfully complete this course. Without the workbook, a student will not be able to participate in the course.

- This attendance workbook is *based* on the text.
- Although the material covered in each is very similar, the *presentation* of the material in the workbook is quite different from the presentation given in the text. The text consists essentially of definitions, formulas, worked out examples and exercises; this workbook, on the other hand, consists *solely* of exercises to be worked out by the student.
- The overheads presented during each lecture are based *exclusively* on the workbook. A student fills in this workbook during the lecture.
- This attendance workbook essentially mimics what goes on during the lectures.
- There are different kinds of exercises, including multiple choice, true/false, matching and fill–in–the–blank.
- Each week, I recommend you read the text, answer questions given here in attendance workbook, look over TI–84+ instructions and then do either quiz or homework assignment, in that order.

On the one hand, the workbook is, as you will see, quite a bit more elaborate than typical lecture notes, which are usually a summary of what the instructor finds important in a recommended course text. On the other hand, this workbook is not quite a text, because although it has many exercises, it does not have quite enough exercises to qualify it as a complete text. I should also point out that this workbook, unfortunately, possesses a number of typographical errors. In short, this workbook aspires to be text and, in the next few years, when enough exercises have been collected, and when most of the typographical errors have been weeded out, it will become a text.

Dr. Jonathan Kuhn,
Associate Professor of Statistics,
Purdue University North Central
November 2009.

Chapter 1

Applications of Linear Functions

We will first look at some of the definitions and formulas related to the Cartesian coordinate system and how to draw points and lines in this system. We will then look at both the graphical as well as algebraic description of linear functions. Economic and statistical examples will then be given.

1.1 The Cartesian Plane and Graphing

We will look at points on the Cartesian coordinate system in this section.

Exercise 1.1 (The Cartesian Plane and Graphing)

1. *Cartesian plane.* Consider rectangular coordinate system (also called Cartesian plane) below with four points.

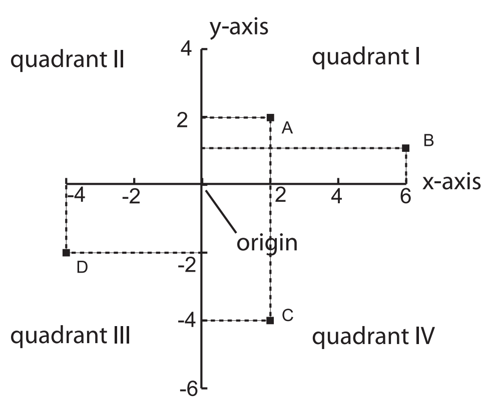

Figure 1.1 (Points on rectangular coordinate system)

   (a) Rectangular coordinate system has **2** / **3** / **4** quadrants.

   (b) The x–coordinate and y–coordinate for four points are

      i. (2,2), (6,1), (2,-4), (-4,-2)

ii. (2,2), (1,6), (2,-4), (-4,-2)
iii. (2,2), (6,1), (-4,2), (-4,-2)
iv. (2,2), (6,1), (2,-4), (-2,-4)

An alternate way of describing these four points is the following table.

x	y
-4	-2
2	-4
2	2
6	1

2. *Equations, Functions, Identities and Graphs.*
An *equation* is a statement where two mathematical expressions are equal. A *conditional equation* is only true for certain values of the variables; these values are called *solutions*. An *identity* is a equation which is true for all values of the variables. A *function*[1] is a rule f which assigns of variable x, one and only one value, $f(x)$. The x is *independent* variable; all possible values of x is *domain*. The y is *dependent* variable; all possible values of y is *range*.

(a) *What is it?*

i. $y = -2x^2 + 2x - 2$
(conditional) equation / function / solution / identity

ii. $y = 2x + 3$
(conditional) equation / function / solution / identity

iii. $(x, y) = (0, 3)$ for $y = 2x + 3$
(conditional) equation / function / solution / identity

iv. $(x, y) = (1, 5)$ for $y = 2x + 3$
(conditional) equation / function / solution / identity

v. circle, centered $(h, k) = (1, 3)$ with radius $r = 5$, $(x-1)^2+(y-3)^2 = 5^2$
(conditional) equation / function / solution / identity

vi. $(x + y)^2 = x^2 + 2xy + y^2$
(conditional) equation / function / solution / identity

vii. $5 = 5$
(conditional) equation / function / solution / identity

(b) *TI–84+: Graphing Functions.*

i. Graph $y = 2x + 3$ with domain $-10 \leq x \leq 10$
Clear previous plots and functions to prevent conflicts: press $Y =$, clear any functions along side $Y =$s, turn off all three plots on top of screen–arrow up, press enter to un-black any active plot. Enter $2x + 3$ beside $Y_1 =$; "x" is "X,Tθ,n" button.

[1]Function "$f(x)$" is often written simply as "y".

Enter domain: Press WINDOW, set Xmin to -10, Xmax to 10, Xscl to 1, Yscl to 1, Xres to 1.

Graph: Press ZOOM, ZoomFit.

ii. Evaluate $y = 2x + 3$ with domain $-10 \leq x \leq 10$ at $x = 2$
$y = 2(2) + 3 =$ **5 / 6 / 7**

Press TRACE, press 2, ENTER. $X = 2$ and $Y = 7$ appear bottom of screen.

iii. Graph $y = -2x^2 + 2x - 2$, domain $-10 \leq x \leq 10$, evaluate at $x = 2$
$y = -2(2)^2 + 2(2) - 2 =$ **−5 / −6 / −7**

First clear $y = 2x + 3$: press $Y =$ and clear it.

Enter $-2x^2 + 2x - 2$ beside $Y_1 =$; use "(-)" button for first negative, "-" button for second negative, "x^2" for square.

Graph: Press ZOOM, ZoomFit.

Press TRACE, press 2, ENTER. $X = 2$ and $Y = -6$ appear bottom of screen.

iv. Graph $y = (x - 2)^3$, domain $-10 \leq x \leq 10$, evaluate at $x = 2$
$y = (2 - 2)^3 =$ **0 / 1 / 2**

First clear $y = -2x^2 + 2x - 2$: press $Y =$ and clear it.

Enter $(x - 2)^3$ beside $Y_1 =$: type "(", then "x", then "-", then 2, then ")", then ∧, then 3.

Graph: Press ZOOM, ZoomFit.

Press TRACE, press 2, ENTER. $X = 2$ and $Y = 0$ appear bottom of screen.

(c) *Sensible domain and range: xbox price and demand.* Let p be price, in hundreds of dollars, for an xbox and $D(p) = 25\left(15 - \frac{p^2}{2}\right)$ number of units sold at Walmart in a month.

i. At $p = 0$, $D(0) = 25\left(15 - \frac{0^2}{2}\right) =$ **300 / 350 / 375**
both price, p and demand $D(p)$ are *not* negative, they are "sensible"

ii. At $p = 6$, $D(6) = 25\left(15 - \frac{6^2}{2}\right) =$ **−75 / −50 / −25**
price, p, is positive but demand $D(p)$ negative, which is *not* "sensible"

iii. Graph $D(p) = 25\left(15 - \frac{p^2}{2}\right)$ to identity sensible domain, start with domain $0 \leq p \leq 6$
sensible domain is $\mathbf{0 \leq p \leq 5.35}$ / $\mathbf{0 \leq p \leq 5.48}$ / $\mathbf{0 \leq p \leq 5.98}$

First clear previous plots.

Enter $25\left(15 - \frac{x^2}{2}\right)$ beside $Y_1 =$.

Enter domain: Press WINDOW, set Xmin to 0, Xmax to 6, Xscl to 1, Yscl to 1, Xres to 1.

Graph: Press ZOOM, ZoomFit.

Determine x-intercept: 2nd CALC, then zero, then arrow just above x-axis, ENTER, arrow to just below x-axis, ENTER, arrow back to axis ENTER, x-intercept ("zero") is a X = 5.477....

1.2 Equations of Straight Lines

We will now look at (straight) lines. After reviewing what a slope of a line is, we look at different equations of lines:

- slope-intercept $y = mx + b$

- point-slope $y - y_1 = m(x - x_1)$
- general form $Ax + By = C$

All *equations* of lines, except vertical, are also linear *functions*.

Exercise 1.2 (Equations of Straight Lines)

1. *Slope: reading ability versus brightness*

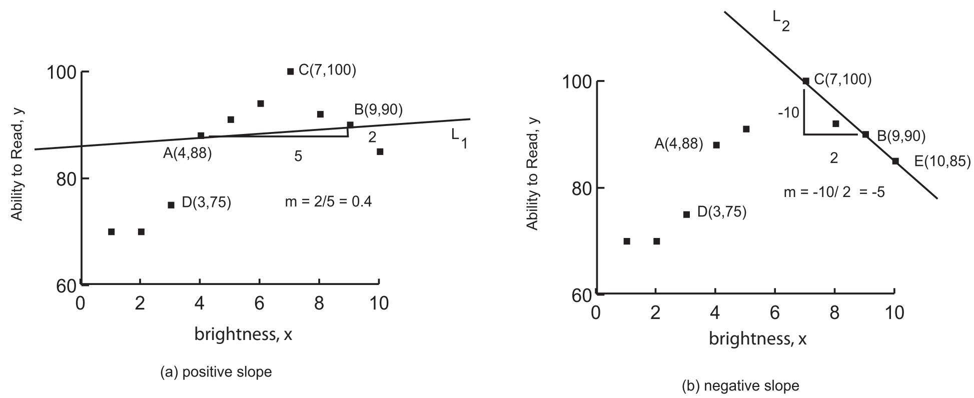

Figure 1.2 (Slope: reading ability versus brightness)

(a) Slope, m, of line L_1 in figure (a) above:

$$m = \frac{90 - 88}{9 - 4} = \frac{2}{5} =$$

(circle one) **0.1** / **0.2** / **0.4**.

(b) Slope of line L_1, $m = 0.4$, says

i. brightness increases by 0.4 units for unit increase reading ability.

ii. reading ability increases by 0.4 units for unit increase in brightness.

When $m > 0$, line **rises** / **falls**.

(c) Slope, m, of line L_2 in figure (b) above:

$$m = \frac{90 - 100}{9 - 7} = -\frac{10}{2} =$$

(circle one) **−5** / **−1** / **5**.

(d) The slope of line L_2, $m = -5$, says (circle one)

i. brightness *de*creases 5 units for unit increase in reading ability.

ii. reading ability *de*creases 5 units for each unit increase in brightness.

When $m < 0$, line **rises** / **falls**.

2. *Equation of a line*

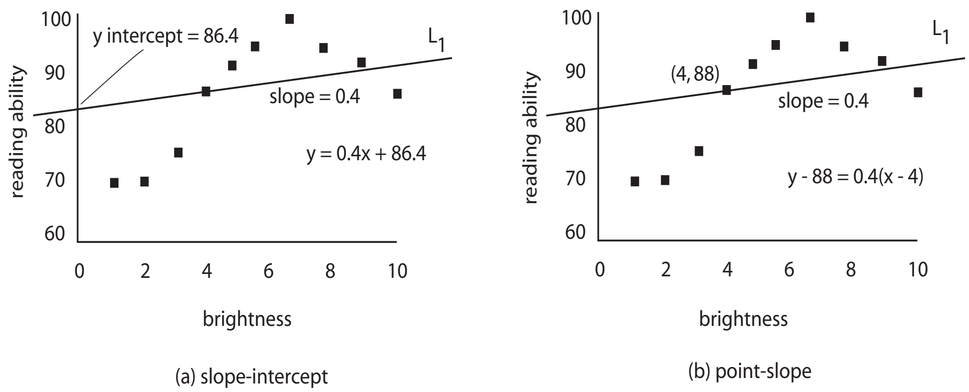

Figure 1.3 (Slope: reading ability versus brightness)

(a) *Slope–Intercept (figure (a)).*
If slope $m = 0.4$ and y–intercept $b = 86.4$, $y = mx + b =$

i. $y = 0.4x - 86.4$
ii. $y = 86.4x + 0.4$
iii. $y = 0.4x + 86.4$

(b) *Slope–Intercept.*
If slope $m = -0.2$ and y–intercept $b = 55$, $y = mx + b =$

i. $y = 55x - 0.2$
ii. $y = -0.2x + 55$
iii. $y = -0.2x - 0.2$

If $x = 5$, $y = -0.2x + 55 = -0.2(5) + 55 =$ **51** / **54** / **56**
If $x = -0.1$, $y = -0.2x + 55 = -0.2(-0.1) + 55 =$ **53.22** / **54.98** / **55.02**.
x is **dependent** / **independent** variable
and y is **dependent** / **independent** variable.

(c) *Point–Slope (figure (b)).*
Equation of line that passes point $(x_1, y_1) = (4, 88)$ with slope $m = 0.4$ is
$y - y_1 = m(x - x_1) =$

i. $(x - 4) = 0.4\,(y - 88)$
ii. $y - 88 = 0.4\,(x - 88)$
iii. $y - 88 = 0.4(x - 4)$

or $y - 88 = 0.4x - 1.6$ or $y = 0.4 + 86.4$.

(d) *Point–Slope.*
Equation of line that passes through point $(4, 70)$ with slope $m = -0.2$ is
$y - y_1 = m(x - x_1) =$ (circle none, one or more)

i. $y - 70 = -0.2(x - 4)$
ii. $y = -0.2x + 70.8$
iii. $y - 4 = -0.2(x - 70)$

3. *More equations of lines.*

(a) *Point–Slope.* Equation of line that passes through points (-2,3) and (5,8)

i. $y - 8 = \frac{5}{7}(x - 5)$
ii. $y + 8 = \frac{5}{7}(x - 5)$
iii. $y - 8 = \frac{6}{7}(x - 5)$
iv. $y - 8 = \frac{5}{7}(x + 5)$

[Hint: calculate slope first, then use point–slope.]

(b) *Equation, Function*: Equation $y = 0.4x + 86.4$ is also function

i. $f(x) = 86.4x + 0.4$
ii. $f(x) = 0.4x + 86.4$
iii. $f(x) = 0.4x - 86.4$

(c) *General Function.*
Equation $y = 0.4x + 86.4$ is also general function

i. $Ax + By = C$, where $A = -0.4$, $B = -1$ and $C = 86.4$
ii. $Ax + By = C$, where $A = 0.4$, $B = 1$ and $C = 86.4$
iii. $Ax + By = C$, where $A = -0.4$, $B = 1$ and $C = 86.4$

(d) *General Function.*
Equation $y = \frac{3}{4}x - 2$ is also general function

i. $Ax + By = C$, where $A = \frac{3}{4}$, $B = 1$ and $C = 2$
ii. $Ax + By = C$, where $A = -\frac{3}{4}$, $B = -1$ and $C = 2$
iii. $Ax + By = C$, where $A = -\frac{3}{4}$, $B = 1$ and $C = -2$

(e) *Horizontal and Vertical Lines.*
Horizontal line that passes through point (0,6) is
$\boldsymbol{x = 6}$
$\boldsymbol{y = 6}$

(f) *Horizontal and Vertical Lines.*
Vertical line that passes through point (4,0) is
$\boldsymbol{x = 4}$
$\boldsymbol{y = 4}$

(g) *Horizontal and Vertical Lines.*
Equation $y = 2$ is **horizontal** / **vertical** line.
Equation $x = 2$ is **horizontal** / **vertical** line.

(h) *Horizontal and Vertical Lines.*
Slope of a vertical line **undefined** / **zero**.
Slope of a horizontal line **undefined** / **zero**.

1.3 Linear Modeling

Many real-world situations can be modeled by linear models. In business, short-run *total costs* are given by sum of *variable costs* (which vary according to production) and *fixed costs*. Total costs can often be represented by *linear cost function*:

$$C(x) = mx + b$$

where m is *marginal cost* (or *direct cost per item*), mx is variable cost and b is *fixed cost*. Related to this, *average cost* is given by $\frac{C(x)}{x}$. Also, if P is *purchase price*, $D = \frac{N}{n}$ is *annual depreciation* where N is *net cost of item* and n is number of years of useful life of item, then, after x years, remaining *depreciated value* is

$$V(x) = P - Dx$$

Exercise 1.3 (Applications of Linear Functions)

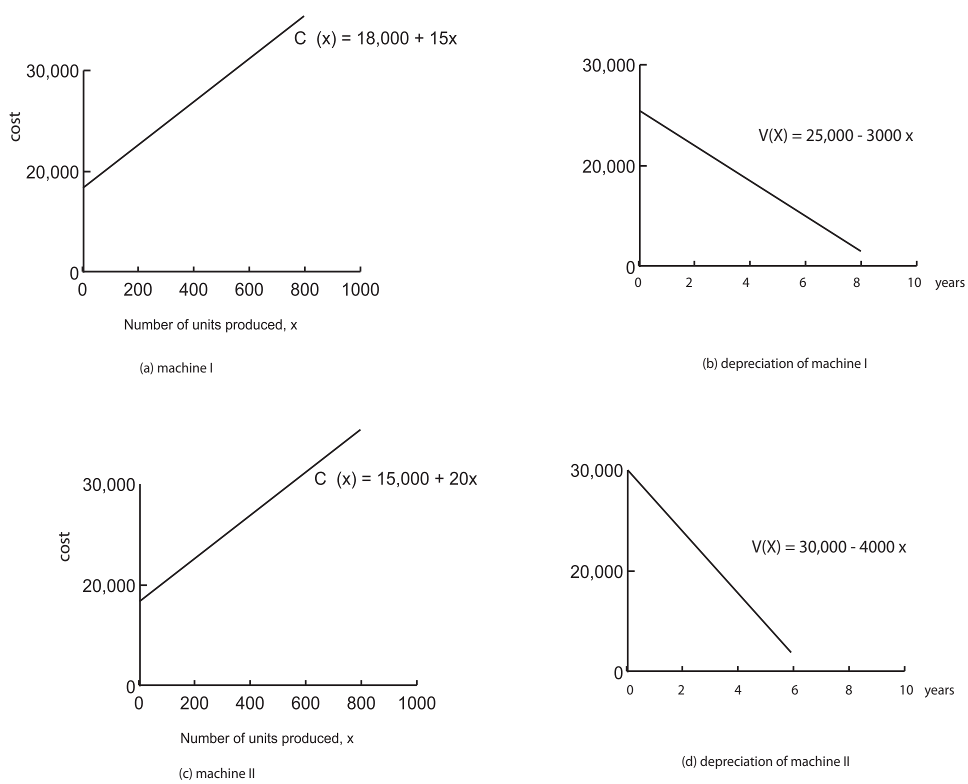

Figure 1.4 (Two cost functions and two depreciation functions)

1. *Machine costs (figure (a)).* Monthly fixed costs of using machine I are \$18,000. Marginal costs of manufacturing one widget using machine I is \$15.
   (a) Linear cost function in terms of x widgets is
      i. $C(x) = 15x + 15000$
      ii. $C(x) = 18000x + 15$
      iii. $C(x) = 15x + 18000$
   (b) Total cost of 300 widgets is
      $C(300) = 15(300) + 18000 = \mathbf{22,000} \ / \ \mathbf{22,500} \ / \ \mathbf{23,000}$
   (c) Additional cost of making 301st widget: **\$15 / \$16 / \$17**
   (d) Average cost per widget of making 300 widgets:
      $\frac{C(300)}{300} = \frac{15(300)+18000}{300} = \mathbf{50} \ / \ \mathbf{75} \ / \ \mathbf{100}$

2. *Depreciation of machine I (figure (b)).* Machine I is initial valued at \$25,000 and is depreciated by straight-line method over eight years, with a salvage of \$1,000.
   (a) How much is depreciated each year? $\frac{25000-1000}{8} = \mathbf{1,000} \ / \ \mathbf{2,000} \ / \ \mathbf{3,000}$
   (b) Linear depreciation function in terms of x years is
      i. $V(x) = 15x + 15000$
      ii. $V(x) = 18000x + 15$
      iii. $V(x) = 25,000 - 3000x$

3. *More machine costs (figure (c)).* Monthly fixed costs of using machine II are \$15,000. Marginal costs of manufacturing one widget using machine II is \$20.
   (a) Linear cost function in terms of x widgets is
      i. $C(x) = 20x + 15000$
      ii. $C(x) = 18000x + 15$
      iii. $C(x) = 15x + 18000$
   (b) Total cost of 300 widgets is
      $C(300) = 20(300) + 15000 = \mathbf{21,000} \ / \ \mathbf{22,500} \ / \ \mathbf{23,000}$
   (c) Additional cost of making 301st widget: **\$15 / \$17 / \$20**
   (d) Average cost per widget of making 300 widgets:
      $\frac{C(300)}{300} = \frac{20(300)+15000}{300} = \mathbf{60} \ / \ \mathbf{70} \ / \ \mathbf{80}$

4. *Depreciation of machine II (figure (d)).* Machine II is initial valued at \$30,000 and is depreciated by straight-line method over six years, with a salvage of \$6,000.

(a) How much is depreciated each year? $\frac{30000-6000}{6}$ = **3,000 / 4,000 / 5,000**

(b) Linear depreciation function in terms of x years is

i. $V(x) = 15x + 15000$

ii. $V(x) = 30,000 - 4000x$

iii. $V(x) = 25,000 - 3000x$

1.4 Two Lines: Relating the Geometry to the Equations

We look at solving a system of two linear equations in two unknowns. In particular, we look at economic applications where we look for an *equilibrium point* between revenue and cost, for an equilibrium quantity and price. Three possible solutions exist: one point, no point (inconsistent solution) or infinite point (dependent, identity solution) intersection. Methods of solution include methods of substitution and elimination, as well as Gauss-Jordan elimination method.

Exercise 1.4 (Two Lines: Relating the Geometry to the Equations)

1. *Cost and revenue function for machine.*

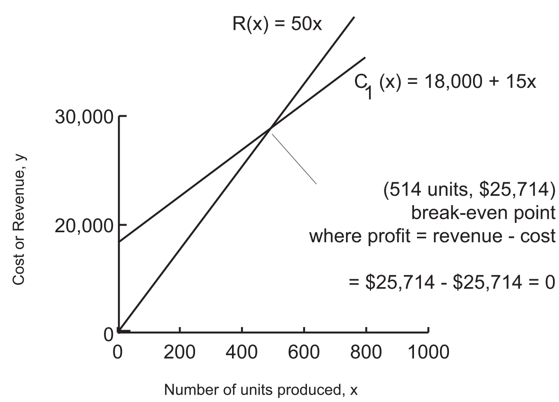

Figure 1.5 (Cost and revenue function for machine)

On one hand, monthly fixed cost of using a machine is $18,000, with variable costs of manufacturing one unit of product at $15. On other hand, each unit of product sells for $50. Determine equilibrium point, where revenue equals costs.

(a) *Using TI-84+ to geometrically find equilibrium (break-even) point.*
Quantity of units and corresponding cost/revenue where revenue equals costs is (quantity, cost/revenue)
$= (x, y) \approx$ **(514, $25,714) / (515, $25,714) / (516, $25,714)**

First clear previous plots.

Enter $18000 + 15x$ beside $Y_1 =$ and $50X$ beside $Y_2 =$.

Enter domain: Press WINDOW, set 0, 1000, 1, 0, 50000, 1, 1.

Graph: Press ZOOM, ZoomFit.

Determine intersection: 2nd CALC, intersect, ENTER to First curve? and ENTER to Second curve?, arrow close to intersection, ENTER, and intersection is X = 514.28.., Y = 25714.28...

(b) *Using algebra to find equilibrium point.*

i. Cost and revenue functions for machine are
$\boldsymbol{C(x) = 18000 + 15x, R(x) = 20x}$
$\boldsymbol{C(x) = 15000 + 20x, R(x) = 15x}$
$\boldsymbol{C(x) = 18000 + 15x, R(x) = 50x}$

ii. Break-even occurs at intersection of cost and revenue

$$C(x) = R(x)$$

or,

$$18000 + 15x = 50x,$$

so $35x = 18000$ and $x = \frac{18000}{35} \approx$ **500 / 514.3 / 525.4** units
where $C(514.3) = 18000 + 15(514.3) \approx$ **\$24, 714 / \$25, 714**
so equilibrium is **(514, \$25, 714) / (515, \$25, 714)**

2. *Intersection: supply, demand and equilibrium for a market of vacuum cleaners.*

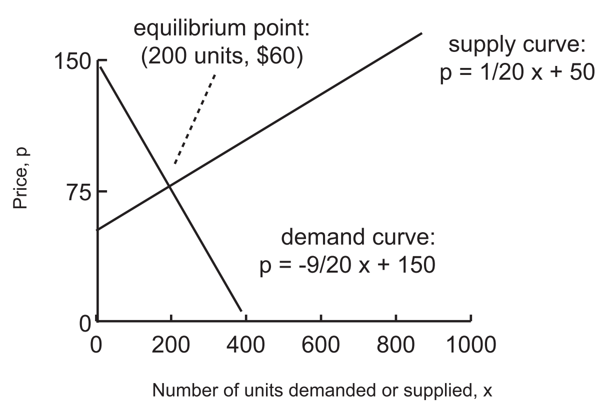

Figure 1.6 (Supply, demand and equilibrium for vacuum cleaners)

Determine equilibrium point, where supply and demand equal one another.

(a) *Using TI-84+ to geometrically find equilibrium point.*
Quantity, price where supply equals demand is
$(x, y) =$ **(\$60, 200) / (200, \$60) / (260, \$60)**

First clear previous plots.

Enter $\left(\frac{1}{20}\right)x + 50$ beside $Y_1 =$ and $-\left(\frac{9}{20}\right)x + 150$ beside $Y_2 =$.

Enter domain: Press WINDOW, set 0, 1000, 1, 0, 150, 1, 1.

Graph: Press Graph.

Determine intersection: 2nd CALC, intersect, ENTER to First curve? and ENTER to Second curve?, arrow close to intersection, ENTER, and intersection is X = 200, Y = 60.

(b) *Using algebra to find equilibrium point.*

i. Supply and demand functions for vacuum cleaners are
$\boldsymbol{p = \left(\frac{1}{20}\right) x + 30, \quad p = -\left(\frac{9}{20}\right) x + 150}$
$\boldsymbol{p = \left(\frac{1}{20}\right) x + 40, \quad p = -\left(\frac{9}{20}\right) x + 150}$
$\boldsymbol{p = \left(\frac{1}{20}\right) x + 50, \quad p = -\left(\frac{9}{20}\right) x + 150}$

ii. Break-even occurs at intersection of supply and demand

$$\left(\frac{1}{20}\right) x + 50 = -\left(\frac{9}{20}\right) x + 150,$$

so $0.5x = 100$ and $x = \frac{100}{0.5} =$ **100** / **200** / **300** units
where $p = \left(\frac{1}{20}\right)(200) + 50 =$ **\$50** / **\$60**
so equilibrium is **(200, \$50)** / **(200, \$60)**

3. *Equations and corner points of shaded regions.*

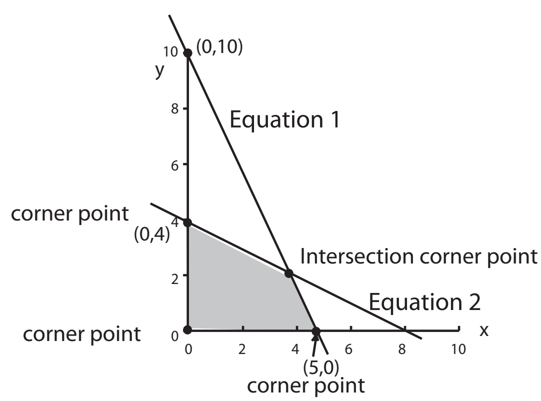

Figure 1.7 (Equations and corner points of shaded region)

(a) Equation 1 passes through y-intercept $(x, y) = (0, 10)$ and x-intercept $(x, y) = (5, 0)$ and so has slope $m = \frac{10-0}{0-5} = -2$ and so
$y - y_1 = m(x - x_1)$ or $y - 10 = -2(x - 0)$ or
$\boldsymbol{2x + y = 10}$
$\boldsymbol{-2x + y = 10}$
$\boldsymbol{2x + y = -10}$

(b) Equation 2 passes through y-intercept $(x, y) = (0, 4)$ and x-intercept $(x, y) = (8, 0)$ and so has slope $m = \frac{4-0}{0-8} = -0.5$ and so
$y - y_1 = m(x - x_1)$ or $y - 4 = -0.5(x - 0)$ or
$\boldsymbol{2x + y = 8}$
$\boldsymbol{x + 2y = 8}$
$\boldsymbol{2x + y = -8}$

(c) Corner point intersection of two equations,

$$\begin{aligned} 2x + y &= 10 \\ x + 2y &= 8 \end{aligned}$$

is, since $y = 10 - 2x$ and $2y = 8 - x$ or $y = 4 - 0.5x$, so

$$10 - 2x = 4 - 0.5x,$$

so $1.5x = 6$ and $x = \frac{6}{1.5} =$ **2 / 3 / 4**
where $y = 10 - 2x = 10 - 2(4) =$ **0 / 2**
so $(x, y) =$ **(0, 1) / (2, 2) / (4, 2)**

4. *Intersection of lines: one, none or infinity of points.*

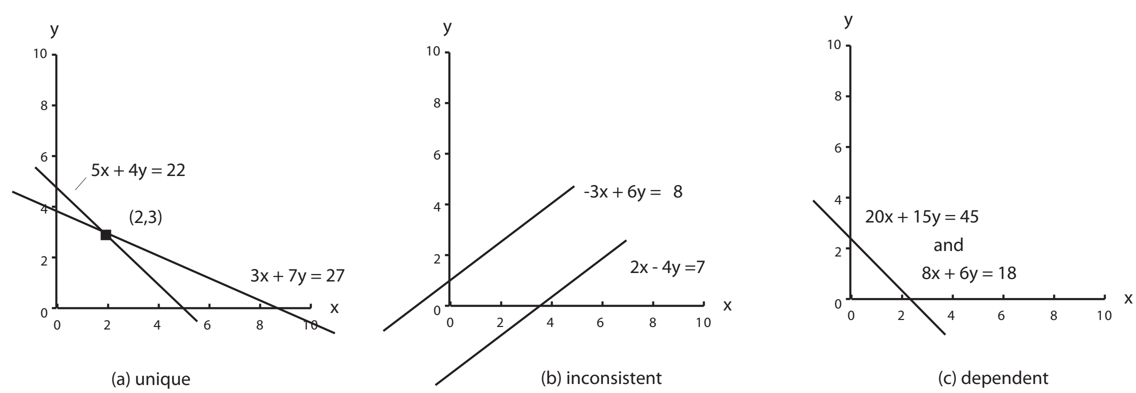

Figure 1.8 (Intersection of lines: one, none or infinity of points)

Two lines intersect

- at one point (they would *not* be parallel to one another)
- at no point (they are parallel and distinct)
- along a *line* (an infinity) of points (they are parallel and coincident)

(a) *Intersect at one point.*

$$\begin{aligned} 5x + 4y &= 22 \\ 3x + 7y &= 27 \end{aligned}$$

is, since $4y = 22 - 5x$, or $y = \frac{22}{4} - \frac{5}{4}x$ and $7y = 27 - 3x$ or $y = \frac{27}{7} - \frac{3}{7}x$, so

$$\frac{22}{4} - \frac{5}{4}x = \frac{27}{7} - \frac{3}{7}x,$$

so $\frac{23}{28}x = \frac{23}{14}$ and $x = \frac{28}{14} =$ **2 / 3 / 4**
Use calculator: $\frac{22}{4} - \frac{27}{7} = 1.64..$ then MATH ENTER ENTER for $\frac{23}{14}$; similar for $\frac{23}{28}$
where $y = \frac{22}{4} - \frac{5}{4}x = \frac{22}{4} - \frac{5}{4}(2) = \frac{27}{7} - \frac{3}{7}x = \frac{27}{7} - \frac{3}{7}(2) =$ **1 / 3**
so $(x, y) =$ **(0, 1) / (2, 2) / (2, 3)**

(b) *Intersect at no point.*

$$\begin{aligned} -3x + 6y &= 8 \\ 2x - 4y &= 7 \end{aligned}$$

is, since $6y = 8 + 3x$, or $y = \frac{8}{6} - \frac{3}{6}x$ and $4y = -7 + 2x$ or $y = -\frac{7}{4} - \frac{2}{4}x$, so

$$\frac{8}{6} - \frac{3}{6}x = -\frac{7}{4} - \frac{2}{4}x,$$

so $\frac{37}{12} = 0x$ and $x =$ **0 / 3 / 4 / huh?**
Use calculator: $\frac{8}{6} + \frac{7}{4} = 3.083...$ then MATH ENTER ENTER for $\frac{37}{12}$
so $(x, y) = \left(\mathbf{0}, \frac{\mathbf{8}}{\mathbf{6}}\right)$ / $\left(\mathbf{0}, -\frac{\mathbf{7}}{\mathbf{4}}\right)$ / **inconsistent (no solution)**

(c) *Intersect at infinity of points.*

$$\begin{aligned} 20x + 15y &= 45 \\ 8x + 6y &= 18 \end{aligned}$$

is, since $15y = 45 - 20x$, or $y = 3 - \frac{20}{15}x$ and $6y = 18 - 8x$ or $y = 3 - \frac{8}{6}x$, so

$$3 - \frac{20}{15}x = 3 - \frac{8}{6}x,$$

or $0x = 0$
so $(x, y) =$ $\mathbf{(0, 1)}$ / $\mathbf{(2, 2)}$ / **identity (infinity of points)**

1.5 Regression and Correlation

Fit a line to a scatter plot of data using *method of least squares* and also measure linearity of scatter plot by *correlation*, r.

Exercise 1.5 (Regression and correlation)

1. *Scatter Diagram: Reading Ability Versus Brightness*[2].

brightness, x	1	2	3	4	5	6	7	8	9	10
ability to read, y	70	70	75	88	91	94	100	92	90	85

[2]Notice scatter plot may be misleading because y-axis ranges 60 to 80, rather than 0 to 80.

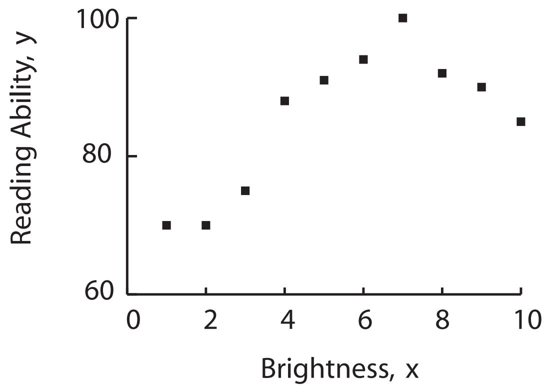

Figure 1.9 (Scatter Diagram, Reading Ability Versus Brightness)

(a) *TI-84+: Scatter diagram.*

- Remember to turn off all STAT PLOTs and Y = plots, and type (x, y) values of reading ability versus brightness data into (L_1, L_2) lists.
- To display scatter plot, press,
  - 2nd STAT PLOT ENTER
  - On ENTER
  - Type: scatter plot figure first row, far left ENTER
  - Xlist: L1 (for x values) ENTER
  - Ylist: L2 (for y values) ENTER
  - Mark: (choose any one of the three)

  then press ZOOM 9:ZoomStat. Press TRACE for values of various (x, y) points.

(b) There are (circle one) **10** / **20** / **30** data points.
One particular data point is (circle one) $\mathbf{(70, 75)}$ / $\mathbf{(75, 2)}$ / $\mathbf{(2, 70)}$.
Data point (9,90) means (circle one)

i. for brightness 9, reading ability is 90.
ii. for reading ability 9, brightness is 90.

(c) Reading ability **positively** / **not** / **negatively** *associated* to brightness.
As brightness increases, reading ability (circle one) **increases** / **decreases**.

(d) Association **linear** / **nonlinear (curved)** because straight line cannot be drawn on graph where all points of scatter fall on or near line.

(e) "Reading ability" is **response** / **explanatory** variable and "brightness" is **response** / **explanatory** variable because reading ability depends on brightness, not the reverse[3].

2. *Scatter Diagram: Grain Yield (tons) versus Distance From Water (feet).*

[3]Sometimes it is not so obvious which is response variable and which is explanatory variable. For example, it is not immediately clear which is explanatory variable and response variable for a scatter plot of husband's IQ scores and wife's IQ scores. If you were interested in knowing husband's IQ score, *given* the wife's IQ score, say, then wives's IQ score would be explanatory variable and husband's IQ score would be response variable.

dist, x	0	10	20	30	45	50	70	80	100	120	140	160	170	190
yield, y	500	590	410	470	450	480	510	450	360	400	300	410	280	350

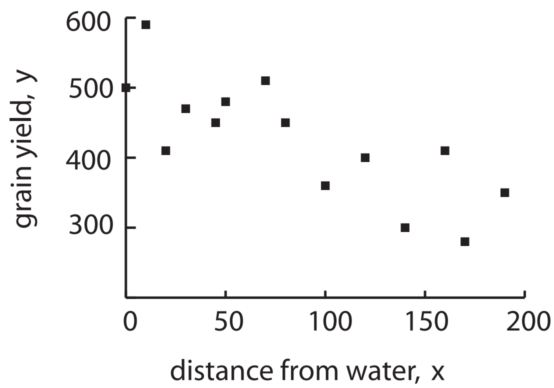

Figure 1.10 (Scatter Diagram, Grain Yield Versus Distance from Water)

Type (x, y) values of grain yield versus distance from water data into (L_3, L_4) (not (L_1, L_2) !) lists. Then 2nd STAT PLOT 1 ENTER and replace L1 by L3 and L2 by L4.

Scatter diagram has **pattern** / **no pattern (randomly scattered)** with (choose one) **positive** / **negative** association,
which is (choose one) **linear** / **nonlinear**, that is a
(choose one) **weak** / **moderate** / **strong** (non)linear relationship,
where grain yield is (choose one) **response** / **explanatory** variable.

3. *Linear Correlation Coefficient: Using Calculator.*
Linear correlation coefficient statistic, r, measures *linearity* of scatter diagram.

$r = +1$	x and y perfectly positively linear
$r \geq 0.8$ or $r \leq -0.8$	x and y strongly linear
$0.5 \leq r \leq 0.8$ or $-0.8 \leq r \leq -0.5$	x and y moderately linear
$-0.5 \leq r \leq 0.5$, $r \neq 0$	x and y weakly linear
$r = 0$	x and y uncorrelated
$r = -1$	x and y perfectly negatively linear

(a) *TI–84+: Linear Correlation Coefficient.*
You will (almost without exception) be able to use TI–84+ to determine correlation coefficient. TI–84+ does *not* automatically calculate r; its default values must be reset so that it does. This *one–time* operation involves turning "diagnostics" *on*:

- Set "Diagnostics" to On by entering CATALOG and then "arrowing" down "DiagnosticOn"; in other words,
  - 2nd CATALOG DiagnosticOn ENTER

  Calculator should return the word "Done".

Once diagnostics have been turned on, calculator can now be used to calculate r.

- First, type (x, y) values of data into (L_1, L_2) lists.
- To determine correlation coefficient of scatter plot of data:
  - STAT CALC 8: LinReg(a + bx) ENTER 2nd L_1 , 2nd L_2

Calculator should return a number of calculated quantities, including, at the bottom, correlation coefficient, r.

(b) *Reading ability versus brightness*

brightness, x	1	2	3	4	5	6	7	8	9	10
reading ability, y	70	70	75	88	91	94	100	92	90	85

In this case, $r \approx$ (circle one) **0.704 / 0.723 / 0.734**.

Type (x, y) into (L_1, L_2), then STAT CALC LinReg(a + bx) ENTER L_1 , L_2

Although same answer is calculated, do *not* use LinReg(ax + b)!

So, association between reading ability and brightness is (circle one)
positive strong linear
negative moderate linear
positive moderate linear

(c) *Grain yield versus distance from water*

dist, x	0	10	20	30	45	50	70	80	100	120	140	160	170	190
yield, y	500	590	410	470	450	480	510	450	360	400	300	410	280	350

In this case, $r \approx$ (circle one) **−0.724 / −0.785 / −0.950**.

Type (x, y) into (L_3, L_4), then STAT CALC LinReg(a + bx) ENTER L_3 , L_4

So, association between grain yield and distance from water is (circle one)
positive strong linear
negative moderate linear
positive moderate linear

4. *Linear correlation coefficient: understanding.*

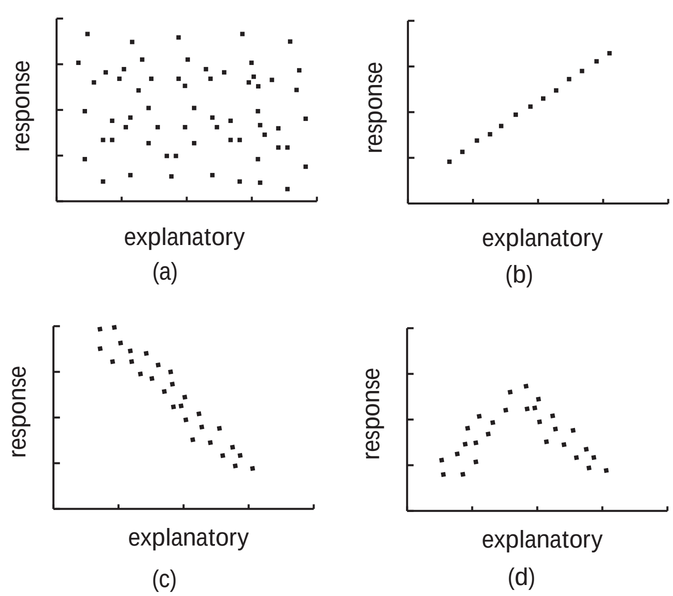

Figure 1.11 (Scatter Diagrams and Possible Correlation Coefficients)

Match correlation coefficients with scatter plots.

(a) scatter diagram (a): $\boldsymbol{r = -0.7}$ / $\boldsymbol{r = 0}$ / $\boldsymbol{r = 0.3}$

(b) scatter diagram (b): $\boldsymbol{r = -0.7}$ / $\boldsymbol{r = 0.1}$ / $\boldsymbol{r = 1}$

(c) scatter diagram (c): $\boldsymbol{r = -0.7}$ / $\boldsymbol{r = 0}$ / $\boldsymbol{r = 0.7}$

(d) scatter diagram (d): $\boldsymbol{r = -0.7}$ / $\boldsymbol{r = 0}$ / $\boldsymbol{r = 0.7}$

When $r \neq 0$, x and y are *linearly* related to one another. If $r = 0$, x and y are *nonlinearly* related to one another, which *often* means diagram (a) or sometimes means diagram (d) where positive and negative associated data points cancel one another out. Always show scatter diagram with correlation r.

5. *Linear Correlation Coefficient: Formulas.*
Computational formula:

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right)\left(\sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}\right)}}$$

Computational formula.
If $\sum_{i=1}^n x_i = -13$, $\sum_{i=1}^n y_i = 12$, $\sum_{i=1}^n x_i^2 = 160$, $\sum_{i=1}^n y_i^2 = 930$, $\sum_{i=1}^n x_i y_i = -345$, and $n = 5$, then
$SS_{xy} = \sum_{i=1}^n x_i y_i - [\sum_{i=1}^n x_i \sum_{i=1}^n y_i / n] = -345 - [(-13)(12)/5] =$ $\mathbf{-189}$ / $\mathbf{-234}$ / $\mathbf{-313.8}$
$SS_x = \sum_{i=1}^n x_i^2 - [(\sum_{i=1}^n x_i)^2/n] = 160 - [(-13)^2/5] =$ **110.2** / **126.2** / **231.3**
and $SS_y = \sum_{i=1}^n y_i^2 - [(\sum_{i=1}^n y_i)^2/n] = 930 - [(12)^2/5] =$ **640.2** / **901.2** / **960.8**
and so $r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{-313.8}{\sqrt{(126.2)(901.2)}} \approx$ $\mathbf{-0.560}$ / $\mathbf{-0.621}$ / $\mathbf{-0.93}$

6. *Least–Squares Line: Calculation, Prediction and Understanding.*

(a) *Reading ability versus brightness.*
Use TI–84+ calculator to create scatter diagram, calculate least-squares regression line and then superimpose line on scatter diagram.

brightness, x	1	2	3	4	5	6	7	8	9	10
reading ability, y	70	70	75	88	91	94	100	92	90	85

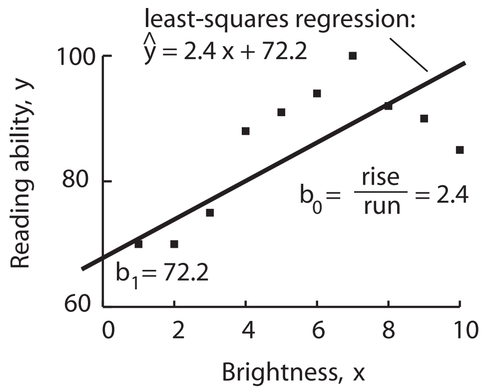

Figure 1.12 (Least–squares Line, reading ability versus brightness)

i. *TI–84+: Scatter diagram (review)*
   - Turn off all STAT PLOTs and Y = plots, then type (x, y) values into (L_1, L_2) lists.
   - Display scatter diagram by pressing
     - 2nd STAT PLOT ENTER
     - On ENTER
     - Type: scatter plot figure first row, far left ENTER
     - Xlist: L1 (for x values) ENTER
     - Ylist: L2 (for y values) ENTER
     - Mark: (choose any one of the three)

   then hit ZOOM 9:ZoomStat. TRACE key used to view various (x, y) points.

ii. *TI–84+: Calculating least-squares regression line.*

   Least–squares line is (circle one)
   $\hat{y} = 2.418x + 72.2$
   $\hat{y} = 2.418 + 72.2x$
   $\hat{y} = 2.944 + 47.04x$

   Calculate least-squares line using TI–84+:
   - First, type (x, y) values into the (L_1, L_2) lists.
   - For least–squares line, type in[4]: STAT CALC LinReg$(ax + b)$ ENTER 2nd L_1 , 2nd L_2 ENTER

iii. *TI–84+: Slope and y-intercept of least-squares regression line.*

   Least–squares regression line is $\hat{y} = 2.418x + 72.2$.

   Slope is b_1 = (circle one) **72.2 / 2.418**.
   Slope, $b_1 = 2.418$, means, on average, reading ability increases 2.418 units for an increase of *one* unit of brightness.

   The *y–intercept* is b_0 = (circle one) **72.2 / 2.418**.
   The y–intercept, $b_0 = 72.2$, means average reading ability is 72.2, if brightness is zero.

   Overlay least-squares on scatter diagram:
   - Create scatterplot.
   - Calculate least-squares line.

[4]It is also possible to type in STAT CALC LinReg$(ax+b)$ ENTER 2nd L_1 , 2nd L_2 ENTER. You will get the *same identical* answer, whether using STAT CALC 4 or STAT CALC 8. Just be careful what is the slope and what is the y–intercept when writing down the least–squares line!

- Press Y =
- VARS, down to 5:Statistics ENTER, over to EQ ENTER
- GRAPH
- Press TRACE , then press UP or DOWN to trace along observed values or least-squares.
- On least-squares regression, type x value ENTER for a y value.

iv. *Prediction.*
At brightness $x = 6.5$, predicted reading ability is
$\hat{y} \approx 2.418x + 72.2 = 2.418(6.5) + 72.2 \approx$ **83.9** / **85.5** / **87.9**..
See Figure below. Press TRACE, arrow onto regression line, type 6.5 ENTER. Read "Y = 87.9".

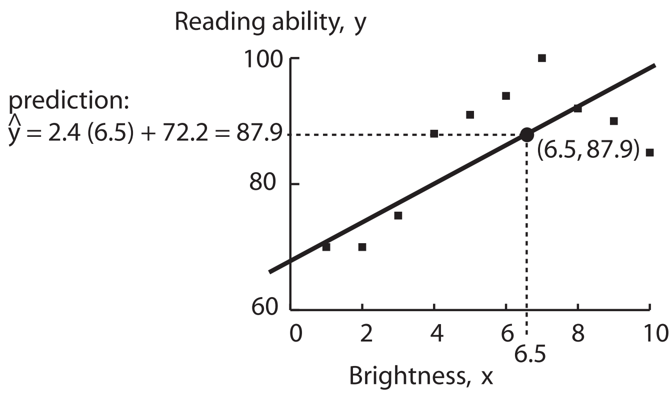

Figure 1.13 (Least–Squares Line: Prediction)

v. *More Prediction.*
At brightness $x = 5.5$, $\hat{y} \approx 2.418(5.5) + 72.2 \approx$ **83.9** / **85.5** / **87.6**.
At brightness $x = 7.5$, $\hat{y} \approx 2.418(7.5) + 72.2 \approx$ **83.9** / **89.5** / **90.4**.

(b) *Grain yield (tons) versus distance from water (feet)*

dist, x	0	10	20	30	45	50	70	80	100	120	140	160	170	190
yield, y	500	590	410	470	450	480	510	450	360	400	300	410	280	350

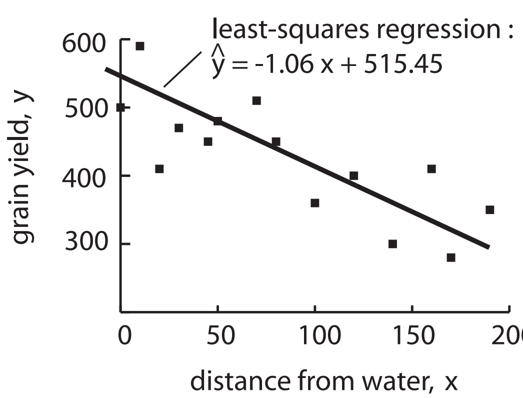

Figure 1.14 (Least–squares regression, grain yield versus distance)

i. The least–squares line is (circle one)
$\boldsymbol{\hat{y} = -1.56x + 515.45}$
$\boldsymbol{\hat{y} = -2.56x + 535.45x}$

$\hat{\boldsymbol{y}} = \mathbf{-1.06}\boldsymbol{x} + \mathbf{515.45}$.

Type data into L_3, L_4; then STAT CALC LinReg(a + bx) L_3, L_4.

ii. *Slope and y-intercept.*
Slope is $b_1 =$ (circle one) **515.45** / **−1.06**.
Slope, $b_1 = -1.06$, means, on average, grain yield decreases 1.06 tons for an increase of one foot away from water.

The *y–intercept* is $b_0 =$ (circle one) **515.45** / **−1.06**.
The y–intercept, $b_0 = 515.45$, means average grain yield is 515.45 at water's edge.

iii. *Prediction.*
At distance $x = 100$,
$\hat{y} = -1.06x + 515.45 = -1.06(100) + 515.45 =$ **400** / **407.3** / **409.5**.
At distance $x = 165$,
$\hat{y} = -1.06x + 515.45 = -1.06(165) + 515.45 =$ **340.5** / **367.0** / **404.8**

7. *Normal Equations*

x_i	y_i	x_i^2	$x_i y_i$
1	70	1	70
2	70	4	140
3	75	9	225
4	88	16	352
5	91	25	455
6	94	36	564
7	100	49	700
8	92	64	736
9	90	81	810
10	85	100	850
$\sum_{i=1}^n x_i = 55$	$\sum_{i=1}^n y_i = 855$	$\sum_{i=1}^n x_i^2 = 385$	$\sum_{i=1}^n x_i y_i = 4902$

Type (x, y) into L_1 and L_2; then type STAT CALC 2-Var Stats ENTER L_1, L_2 for $\sum_{i=1}^n x_i = 55$, $\sum_{i=1}^n y_i = 855$, $\sum_{i=1}^n x_i^2 = 385$, $\sum_{i=1}^n x_i y_i = 4902$.

(a) So *normal equations* given by

$$\begin{aligned} nb + \left(\sum_{i=1}^n x_i\right) m &= \sum_{i=1}^n y_i \\ \left(\sum_{i=1}^n x_i\right) b + \left(\sum_{i=1}^n x_i^2\right) m &= \sum_{i=1}^n x_i y_i \end{aligned}$$

in other words,

i. $55b + 55m = 855; \qquad 55b + 385m = 4902$

ii. $10b + 55m = 385$; $55b + 385m = 4902$

iii. $10b + 55m = 855$; $55b + 385m = 4902$

iv. $10b + 55m = 855$; $55b + 4902m = 4902$

(b) Solving normal equations,

$$\begin{aligned} 10b + 55m &= 855 \\ 55b + 385m &= 4902 \end{aligned}$$

is, since $10b + 55m = 855$, or $m = \frac{855}{55} - \frac{10}{55}b$
and $385m = 4902 - 55b$ or $m = \frac{4902}{385} - \frac{55}{385}b$, so

$$\frac{855}{55} - \frac{10}{55}b = \frac{4902}{385} - \frac{55}{385}b,$$

so $\frac{3}{77}b = \frac{1083}{385}$ and $b =$ **2.418 / 72.2**

Use calculator: $\frac{10}{55} - \frac{55}{385} = 0.038..$ then MATH ENTER ENTER for $\frac{3}{77}$; similar for $\frac{1083}{385}$

where $m = \frac{855}{55} - \frac{10}{55}b = \frac{855}{55} - \frac{10}{55}(72.2) = \frac{4902}{385} - \frac{55}{385}b = \frac{4902}{385} - \frac{55}{385}(72.2) \approx$ **2.418 / 72.2**

so $(b, m) \approx$ $\mathbf{(0, 1)}$ / $\mathbf{(2, 2)}$ / $\mathbf{(72.2, 2.418)}$