SALUTON DENOVE!

확률 및 통계학

동국대학교

2025년 2학기

문 동 휘

월 12:00 ~ 13:30

수 11:00 ~ 12:30

순위상관계수

- 두 변량 X와 Y가 순위로 측정되거나 측정값들이 순위가 아니라면 순위대로 정렬한 후 큰 순서에 따라 순위를 매긴다.

- 그 후 순위의 짝인(x_i, y_i)를 만든다. 이때 i 번째 개체의 순위 차를 d_i 라 하자.

- 순위상관계수 r_s 는 다음과 같이 계산한다.
- 순위가 같다면 그 산술평균이 순위이다. 예) 4, 5위 순위 같으면 둘 다 4.5위

10

107

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i}{n(n^2-1)}$$

F15 = 1 - ((6*F12)/(F13*(F14-1)))

	A	B	C	D	E	F
1	영어시험 순위	영어시험 수정순위 (x)	수학시험 순위	수학시험 수정순위 (y)	d = x - y	d^2
2	1	1	8	8	-7	49
3	5	5	2또는3	2.5	2.5	6.25
4	7또는8	7.5	4	4	3.5	12.25
5	6	6	1	1	5	25
6	2	2	2또는3	2.5	-0.5	0.25
7	3	3	5	5	-2	4
8	4	4	6	6	-2	4
9	7 또는 8	7.5	9	9	-1.5	2.25
10	9	9	7	7	2	4
11	10	10	10	10	0	0
12		순위가 같으면,			d^2 합	107
13		두 수의 평균으로 매김			n	10
14					n^2	100
15					순위상관계수	0.351515

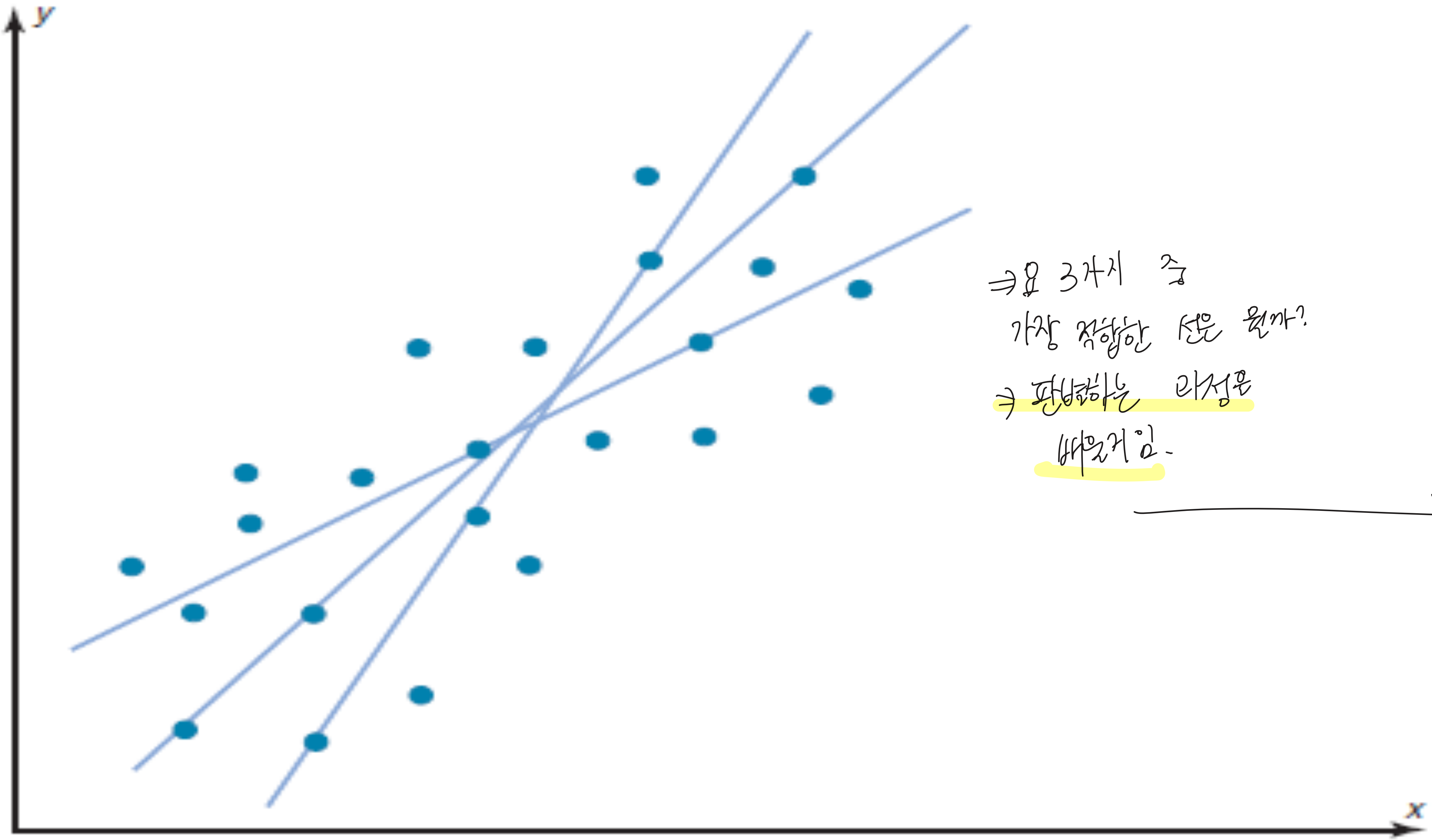
	A	B	C	D	E	F	G	H	I	J
1					문제					
2	경제시험성적	경제시험순위	통계시험성적	통계시험순위	d=x-y	d^2	순위상관계수			
3	10	10	85	3	7	49	$n=10$			
4	53	6	63	5	1	1	$n^2=100$			
5	77	3	57	6	-3	9	d^2 합=140			
6	90	1	88	2	-1	1	$\frac{140}{99\times10}$ = 순위상관계수			
7	12	9	15	10	-1	1				
8	43	7	32	8	-1	1				
9	55	5	44	7	-2	4				
10	82	2	31	9	-7	49				
11	75	4	99	1	3	9				
12	31	8	67	4	4	16				

				문제			
경제시험성적	경제시험순위	통계시험성적	통계시험순위	d = x - y	d^2		
10	10	85	3	7	49	n	10
53	6	63	5	1	1	n^2	100
77	3	57	6	-3	9	d^2 합	140
90	1	88	2	-1	1	순위상관계수	0.151515
12	9	15	10	-1	1		
43	7	32	8	-1	1		
55	5	44	7	-2	4		
82	2	31	9	-7	49		
75	4	99	1	3	9		
31	8	67	4	4	16		

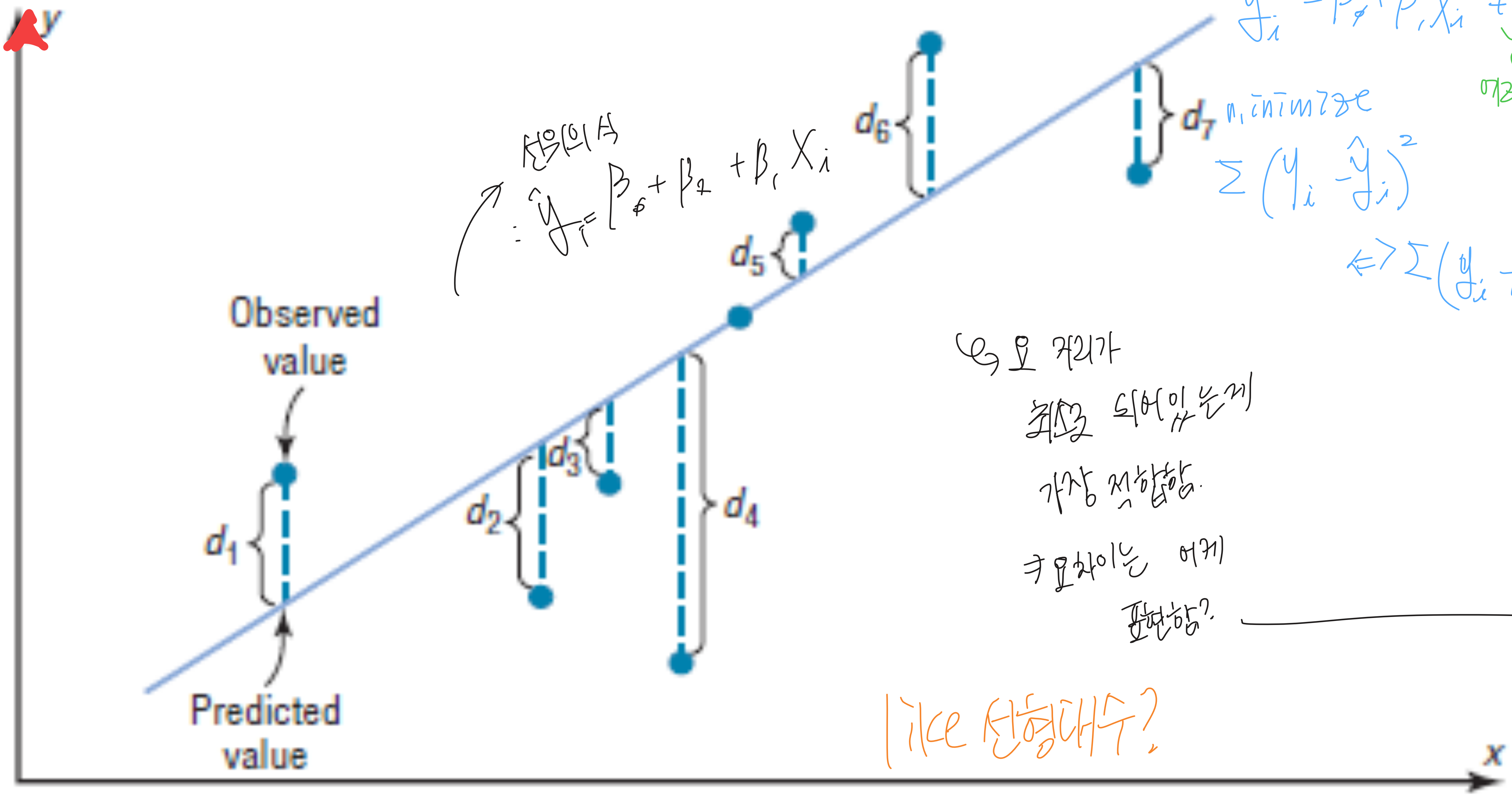
↳ 경제와 통계는 관련 없구나 라는걸 알려줌.

약간 //

회귀분석기초



⇒ 요 3가지 중
가장 적합한 선은 뭘까?
⇒ 판별하는 과정을
배울거임.



선형회귀식 : $\hat{y}_i = \beta_0 + \beta_2 + \beta_1 X_i$

$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

에러

minimize $\sum (y_i - \hat{y}_i)^2$

$\Leftrightarrow \sum (y_i - \beta_0 - \beta_1 y_i)$

↳ 요 거리가 최소화 되어있는게 가장 적합함.

→ 요차이는 어케 표현함?

like 선형대수?

The difference between the actual value y and the predicted value y' (that is, the vertical distance) is called a **residual** or a predicted error. Residuals are used to determine the line that best describes the relationship between the two variables.

The method used for making the residuals as small as possible is called the *method of least squares*. As a result of this method, the regression line is also called the *least squares regression line*.

The normal equations (1.9) can be derived by calculus. For given sample observations (X_i, Y_i), the quantity Q in (1.8) is a function of β_0 and β_1. The values of β_0 and β_1 that minimize Q can be derived by differentiating (1.8) with respect to β_0 and β_1. We obtain:

$$\frac{\partial Q}{\partial \beta_0} = -2\sum (Y_i - \beta_0 - \beta_1 X_i)$$

$$\frac{\partial Q}{\partial \beta_1} = -2\sum X_i(Y_i - \beta_0 - \beta_1 X_i)$$

편미분한거임.

β_0 와 β_1 에 대해 미분함

We then set these partial derivatives equal to zero, using b_0 and b_1 to denote the particular values of β_0 and β_1 that minimize Q:

$\Sigma(y_i - \hat{y}_i)$ 을 최소로 만드는 β_0 와 β_1 ?

$$-2\sum (Y_i - b_0 - b_1 X_i) = 0$$

$$-2\sum X_i(Y_i - b_0 - b_1 X_i) = 0$$

Simplifying, we obtain:

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = 0$$

$$\sum_{i=1}^{n}X_i(Y_i - b_0 - b_1 X_i) = 0$$

Expanding, we have:

$$\sum Y_i - nb_0 - b_1 \sum X_i = 0$$

$$\sum X_i Y_i - b_0 \sum X_i - b_1 \sum X_i^2 = 0$$

from which the normal equations (1.9) are obtained by rearranging terms.

Using the analytical approach, it can be shown for regression model (1.1) that the values b_0 and b_1 that minimize Q for any particular set of sample data are given by the following simultaneous equations:

$$\sum Y_i = nb_0 + b_1 \sum X_i \tag{1.9a}$$

$$\sum X_i Y_i = b_0 \sum X_i + b_1 \sum X_i^2 \tag{1.9b}$$

Equations (1.9a) and (1.9b) are called *normal equations; b_0* and b_1 are called *point estimators* of β_0 and β_1, respectively.

The normal equations (1.9) can be solved simultaneously for b_0 and b_1:

$$b_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} \tag{1.10a}$$

$$b_0 = \frac{1}{n}\left(\sum Y_i - b_1 \sum X_i\right) = \bar{Y} - b_1 \bar{X} \tag{1.10b}$$

where $\bar{X}$ and $\bar{Y}$ are the means of the X_i and the Y_i observations, respectively. Computer calculations generally are based on many digits to obtain accurate values for b_0 and b_1.

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \tag{1.10a}$$

$$b_0 = \frac{1}{n}\left(\sum Y_i - b_1 \sum X_i\right) = \underbrace{\bar{Y} - b_1 \bar{X}}_{y절편} \tag{1.10b}$$

Formulas for the Regression Line $y' = a + bx$

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

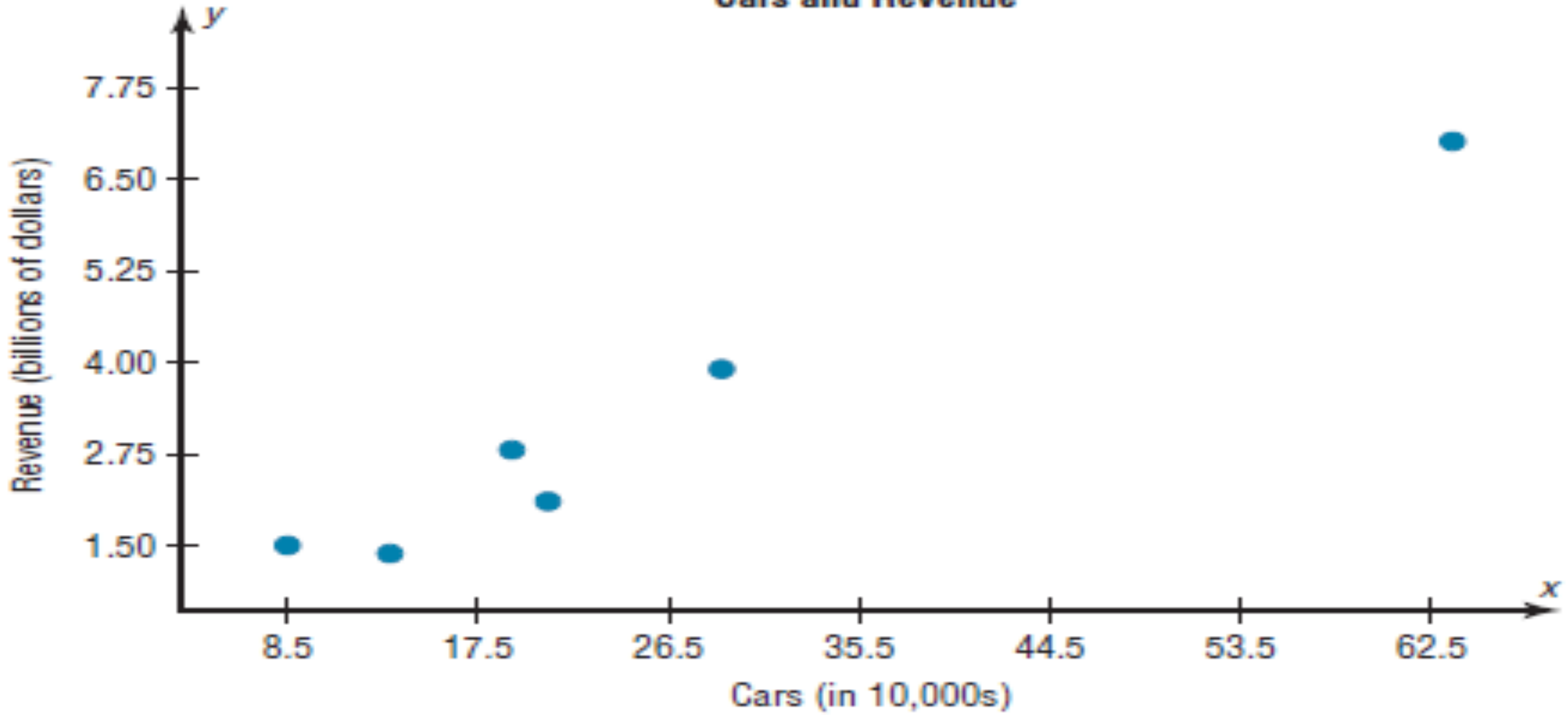
where a is the y' intercept and b is the slope of the line.

- a = b0 = y절편
- b = b1 = 기울기

회귀분석 예제 1

Company	Cars (in ten thousands)	Revenue (in billions)
A	63.0	$7.0
B	29.0	3.9
C	20.8	2.1
D	19.1	2.8
E	13.4	1.4
F	8.5	1.5

Cars and Revenue



cars (x)	revenue (y)	xy	x^2
63	7	441	3969
29	3.9	113.1	841
20.8	2.1	43.68	432.64
19.1	2.8	53.48	364.81
13.4	1.4	18.76	179.56
8.5	1.5	12.75	72.25
153.8	18.7	682.77	5859.26

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b_0 = \frac{1}{n}\left(\sum Y_i - b_1 \sum X_i\right) = \bar{Y} - b_1 \bar{X}$$

n	6
b의 분자	1220.56
b의 분모	11501.12
b (기울기)	**0.106125**
Y 평균	3.116667
X평균	25.63333
a(y절편)	**0.396321**

The values needed for the equation are $n = 6$, $\Sigma x = 153.8$, $\Sigma y = 18.7$, $\Sigma xy = 682.77$, and $\Sigma x^2 = 5859.26$. Substituting in the formulas, you get

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{(18.7)(5859.26) - (153.8)(682.77)}{(6)(5859.26) - (153.8)^2} = 0.396$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{6(682.77) - (153.8)(18.7)}{(6)(5859.26) - (153.8)^2} = 0.106$$

Hence, the equation of the regression line $y' = a + bx$ is

$$y' = 0.396 + 0.106x$$

To graph the line, select any two points for x and find the corresponding values for y. Use any x values between 10 and 60. For example, let $x = 15$. Substitute in the equation and find the corresponding y' value.

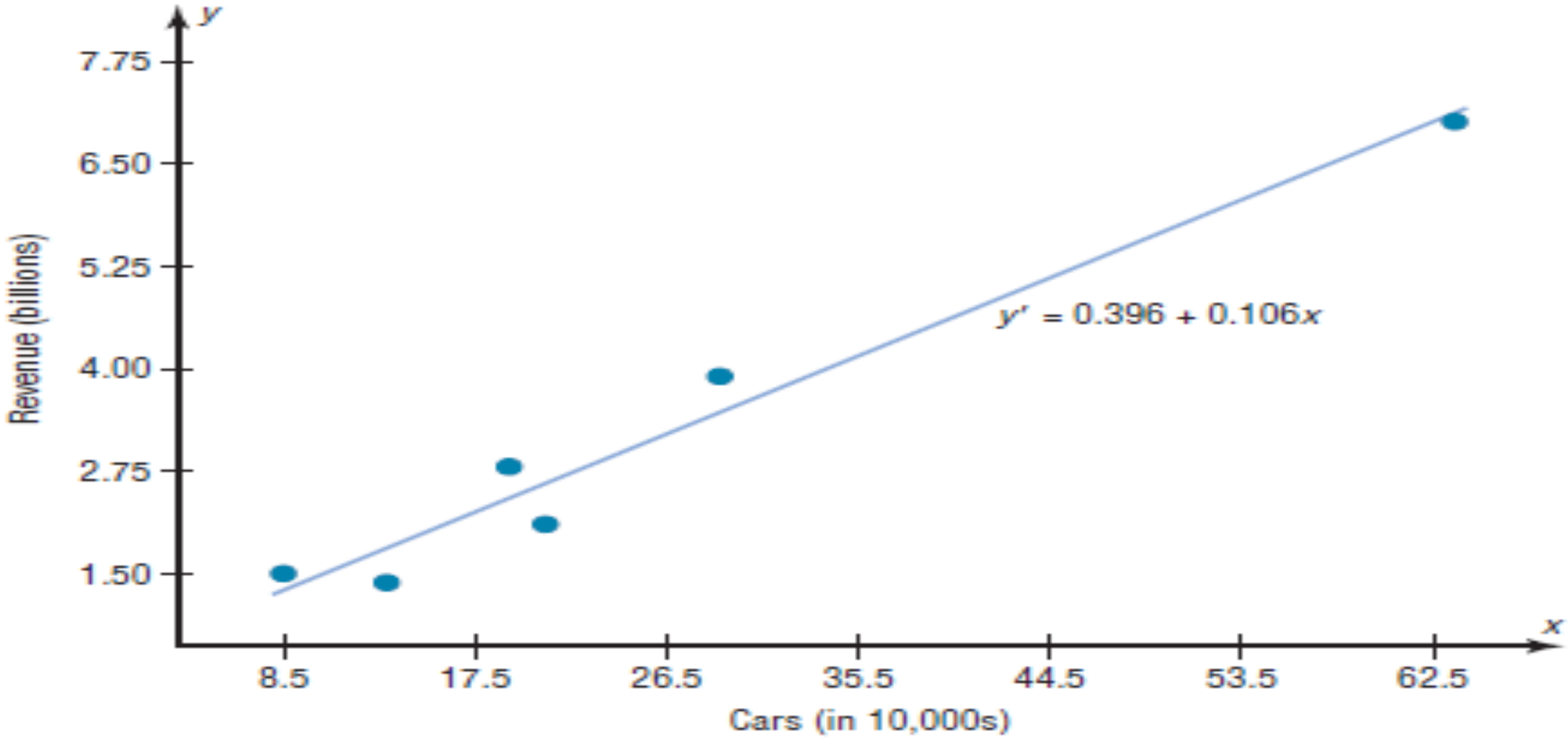
$$\begin{aligned} y' &= 0.396 + 0.106x \\ &= 0.396 + 0.106(15) \\ &= 1.986 \end{aligned}$$

Let $x = 40$; then

$$\begin{aligned} y' &= 0.396 + 0.106x \\ &= 0.396 + 0.106(40) \\ &= 4.636 \end{aligned}$$

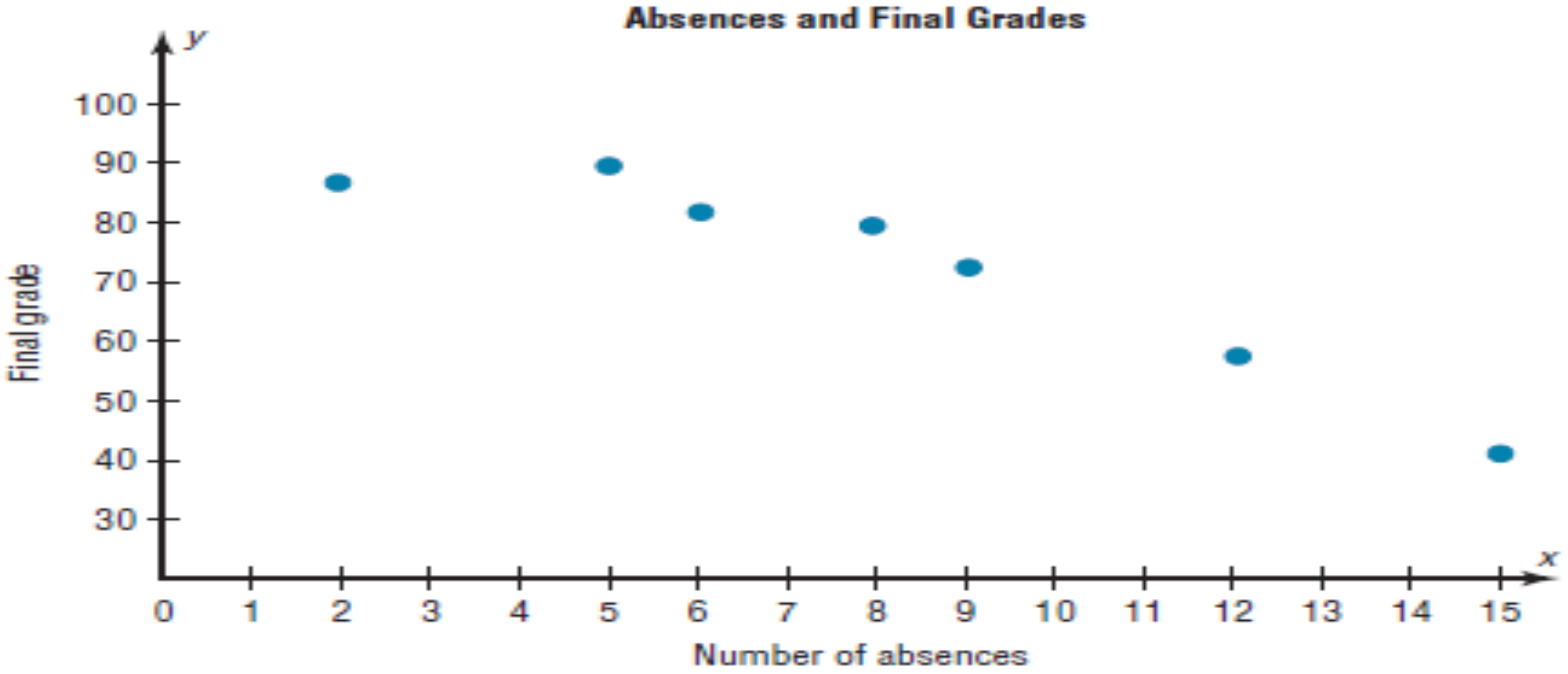
Then plot the two points (15, 1.986) and (40, 4.636) and draw a line connecting the two points. See Figure 10–14.

FIGURE 10–14 Regression Line for Example 10–9



회귀분석 예제 2

Student	Number of absences x	Final grade y (%)
A	6	82
B	2	86
C	15	43
D	9	74
E	12	58
F	5	90
G	8	78



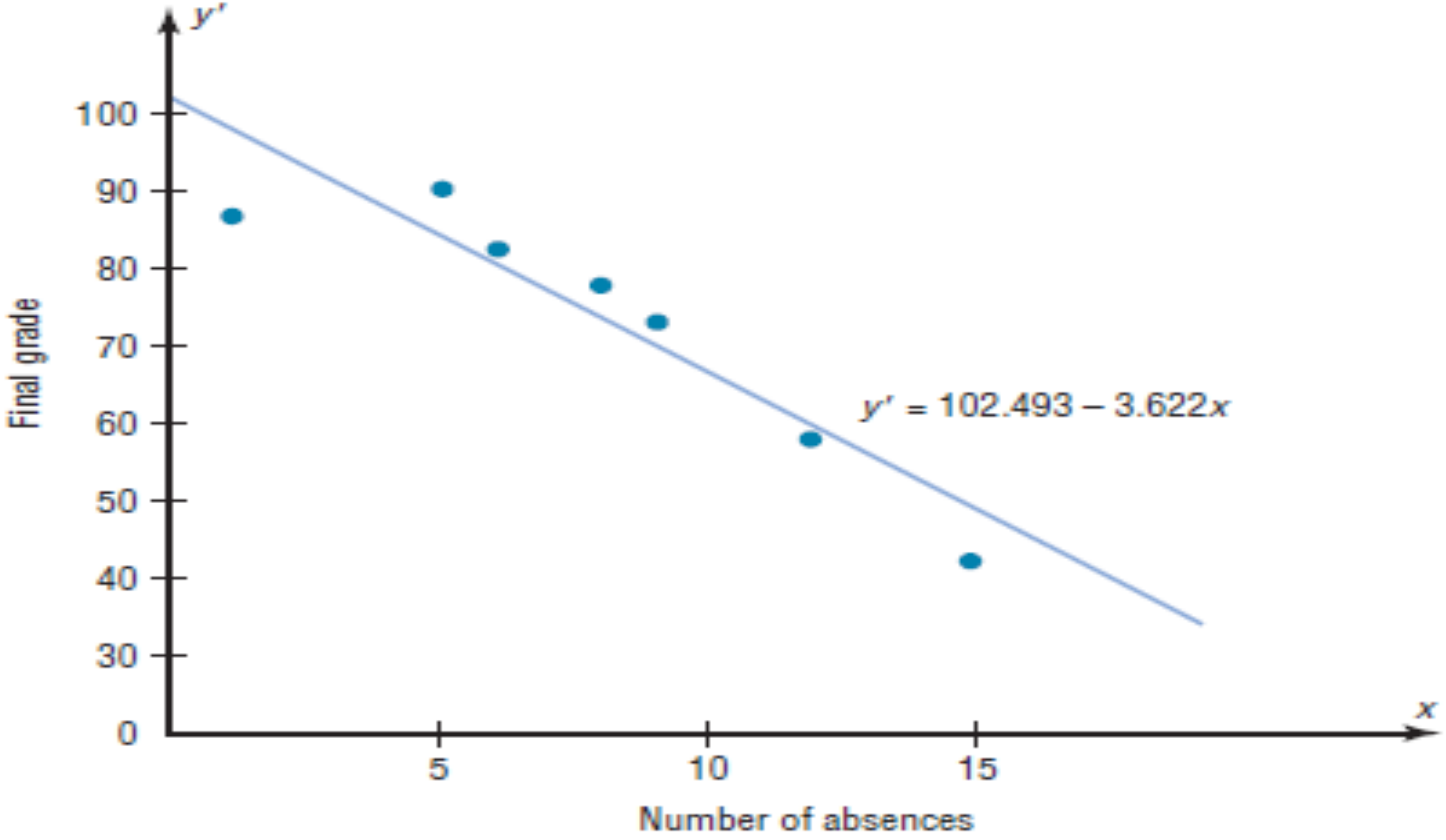
The values needed for the equation are $n = 7$, $\Sigma x = 57$, $\Sigma y = 511$, $\Sigma xy = 3745$, and $\Sigma x^2 = 579$. Substituting in the formulas, you get

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{(511)(579) - (57)(3745)}{(7)(579) - (57)^2} = 102.493$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{(7)(3745) - (57)(511)}{(7)(579) - (57)^2} = -3.622$$

Hence, the equation of the regression line $y' = a + bx$ is

$$y' = 102.493 - 3.622x$$



y'
100
90
80
70
60
50
40
30
0
Final grade
y' = 102.493 – 3.622x
x
5
10
15
Number of absences

회귀분석 문제들

$n=10$
$\sum x=100$
$\sum y=220$
$\sum xy=3680$
$\sum x^2=1140$

13.19 The following information is obtained from a sample data set.

$$n = 10, \quad \sum x = 100, \quad \sum y = 220, \quad \sum xy = 3680, \quad \sum x^2 = 1140$$

$\bar{x} = \frac{100}{10} = 10$

$\bar{y} = \frac{220}{10} = 22$

Find the estimated regression line.

$\sum xy - \frac{\sum x \sum y}{n} = 3680 - \frac{(100)(220)}{10} = 3680 - 1480$

$\sum x^2 - \frac{(\sum x)^2}{n} = 1140 - \frac{10000}{10} = 140$

$b = \frac{1480}{140} = 10.5714$

$a = \bar{y} - b\bar{x} = 22 - (10.5714)(10)$

$a = 22 - 105.714 = -83.714$

$\hat{y} = -83.714 + 10.5714$

$7.8371 = 10.57$

$a = -83.71$

$\hat{y} = -83.71 + 10.57x$

13.27 A diabetic is interested in determining how the amount of aerobic exercise impacts his blood sugar. When his blood sugar reaches 170 mg/dL, he goes out for a run at a pace of 10 minutes per mile. On different days, he runs different distances and measures his blood sugar after completing his run. Note: The preferred blood sugar level is in the range of 80 to 120 mg/dL. Levels that are too low or too high are extremely dangerous. The data generated are given in the following table.

Distance (miles)	2	2	2.5	2.5	3	3	3.5	3.5	4	4	4.5	4.5
Blood sugar (mg/dL)	136	146	131	125	120	116	104	95	85	94	83	75

b. Find the predictive regression equation of blood sugar level on the distance run.

e. Calculate the predicted blood sugar level count after a run of 3.1 miles (5 kilometers).

f. Estimate the blood sugar level after a 10-mile run. Comment on this finding.

13.19 $\hat{y} = -83.7140 + 10.5714x$

13.27 b. $\hat{y} = 191.6238 - 25.3714x$ e. 112.9724

f. -62.0905

Predictions Using the Estimated Regression Line

Assumptions for Valid Predictions in Regression

이런게 있다

1. The sample is a random sample.
2. For any specific value of the independent variable *x*, the value of the dependent variable *y* must be normally distributed about the regression line. See Figure 10–16(a).
3. The standard deviation of each of the dependent variables must be the same for each value of the independent variable. See Figure 10–16(b).

$X : \alpha$ — α, $1-\alpha$

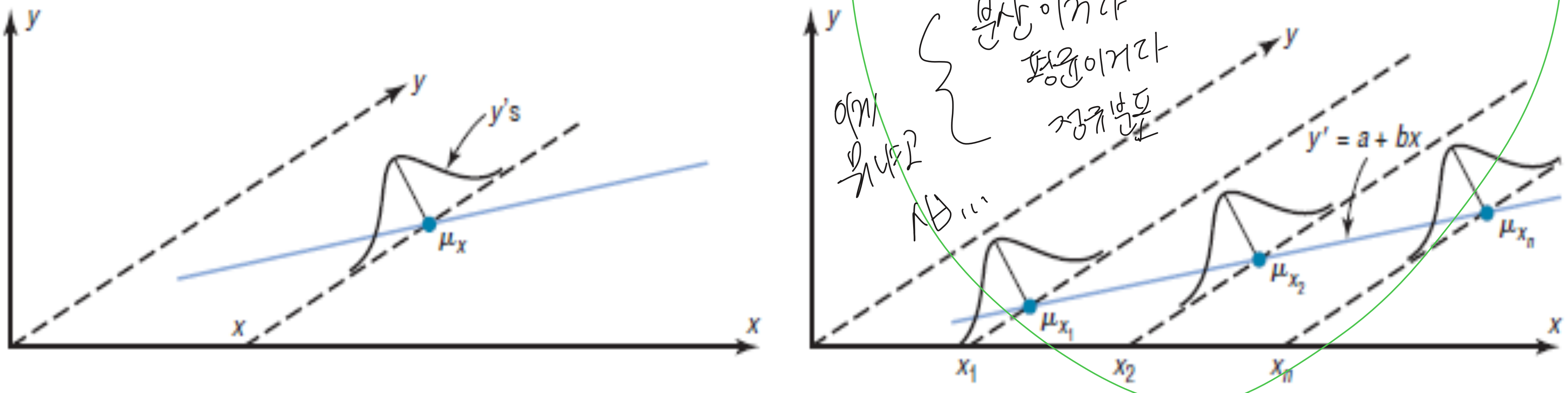
정규분포 — 평균 — 분산

$$y_i \sim N(\underbrace{\beta_0 + \beta_1 X_i}_{\text{평균}}, \underbrace{\sigma^2}_{\text{분산}})$$

등분산이 들어야함

분산이거나 평균이거나 정규분포 이게 무너지면 새...

FIGURE 10–16 Assumptions for Predictions



(a) Dependent variable *y* normally distributed

(b) $\sigma_1 = \sigma_2 = \cdots = \sigma_n$

Company	Cars (in ten thousands)	Revenue (in billions)
A	63.0	$7.0
B	29.0	3.9
C	20.8	2.1
D	19.1	2.8
E	13.4	1.4
F	8.5	1.5

$$y' = 0.396 + 0.106x$$

EXAMPLE 10–11 Car Rental Companies

Use the equation of the regression line to predict the income of a car rental agency that has 200,000 automobiles.

SOLUTION

Since the x values are in 10,000s, divide 200,000 by 10,000 to get 20, and then substitute 20 for x in the equation.

$$\begin{aligned} y' &= 0.396 + 0.106x \\ &= 0.396 + 0.106(20) \\ &= 2.516 \end{aligned}$$

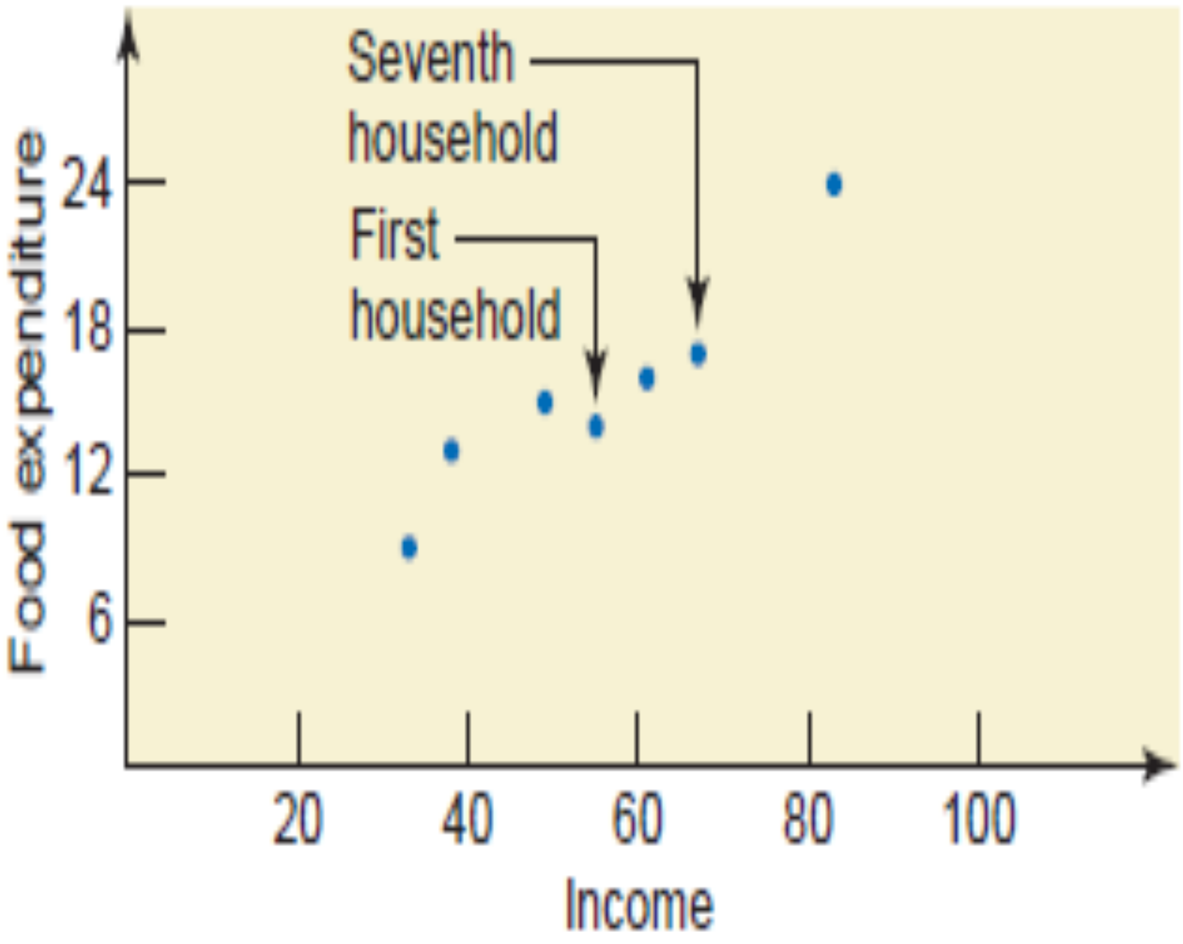
Hence, when a rental agency has 200,000 automobiles, its revenue will be approximately $2.516 billion.

Table 13.1 **Incomes and Food Expenditures of Seven Households**

Income	Food Expenditure
55	14
83	24
38	13
61	16
33	9
49	15
67	17

뭐 정확하지 않대

전 ppt는 뭐가 있어서 그나마 같은데



Interpretation of a

Consider a household with zero income. Using the estimated regression line obtained in Example 13–1, we get the predicted value of y for $x = 0$ as

$$\hat{y} = 1.5050 + .2525(0) = \$1.5050 \text{ hundred} = \$150.50$$

Thus, we can state that a household with no income is expected to spend \$150.50 per month on food. Alternatively, we can also state that the point estimate of the average monthly food expenditure for all households with zero income is \$150.50. Note that here we have used $\hat{y}$ as a point estimate of $\mu_{y|x}$. Thus, $a = 150.50$ gives the predicted or mean value of y for $x = 0$ based on the regression model estimated for the sample data.

However, we should be very careful when making this interpretation of a. In our sample of seven households, the incomes vary from a minimum of \$3300 to a maximum of \$8300. (Note that in Table 13.1, the minimum value of x is 33 and the maximum value is 83.) Hence, our regression line is valid only for the values of x between 33 and 83. If we predict y for a value of x outside this range, the prediction usually will not hold true. Thus, since $x = 0$ is outside the range of household incomes that we have in the sample data, the prediction that a household with zero income spends \$150.50 per month on food does not carry much credibility. The same is true if we try to predict y for an income greater than \$8300, which is the maximum value of x in Table 13.1.

Interpretation of b

The value of b in a regression model gives the change in y (dependent variable) due to a change of one unit in x (independent variable). For example, by using the regression equation obtained in Example 13–1, we see:

해석 ☆

$$\text{When } x = 50, \quad \hat{y} = 1.5050 + .2525(50) = 14.1300$$
$$\text{When } x = 51, \quad \hat{y} = 1.5050 + .2525(51) = 14.3825$$

Hence, when x increased by one unit, from 50 to 51, $\hat{y}$ increased by $14.3825 - 14.1300 = .2525$, which is the value of b. Because our unit of measurement is hundreds of dollars, we can state that, on average, a \$100 increase in income will result in a \$25.25 increase in food expenditure. We can also state that, on average, a \$1 increase in income of a household will increase the food expenditure by \$.2525. Note the phrase "on average" in these statements. The regression line is seen as a measure of the mean value of y for a given value of x. If one household's income is increased by \$100, that household's food expenditure may or may not increase by \$25.25. However, if the incomes of all households are increased by \$100 each, the average increase in their food expenditures will be very close to \$25.25.

아니시발 2-이든하거니 한다며기

회기분석 종합문제

In a study on speed and braking distance, researchers looked for a method to estimate how fast a person was traveling before an accident by measuring the length of the skid marks. An area that was focused on in the study was the distance required to completely stop a vehicle at various speeds. Use the following table to answer the questions.

MPH	Braking distance (feet)
20	20
30	45
40	81
50	133
60	205
80	411

Assume MPH is going to be used to predict stopping distance.

1. Find the linear regression equation.
2. What does the slope tell you about MPH and the braking distance? How about the y′ intercept?
3. Find the braking distance when MPH = 45.
4. Find the braking distance when MPH = 100.
5. Comment on predicting beyond the given data values.

1. The linear regression equation is
$y' = -151.900 + 6.451x$

2. The slope says that for each additional mile per hour a car is traveling, we expect the stopping distance to increase by 6.45 feet, on average. The y intercept is the braking distance we would expect for a car traveling 0 mph—this is meaningless in this context, but is an important part of the model.

3. $y' = -151.900 + 6.451(45) = 138.4$ The braking distance for a car traveling 45 mph is approximately 138 feet.

4. $y' = -151.900 + 6.451(100) = 493.2$ The braking distance for a car traveling 100 mph is approximately 493 feet.

5. It is not appropriate to make predictions of braking distance for speeds outside of the given data values (for example, the 100 mph in the previous problem) because we know nothing about the relationship between the two variables outside of the range of the data.

10/2

확률기초

Probability as a general concept can be defined as the chance of an event occurring. Many people are familiar with probability from observing or playing games of chance, such as card games, slot machines, or lotteries. In addition to being used in games of

Processes such as flipping a coin, rolling a die, or drawing a card from a deck are called *probability experiments.*

A **probability experiment** is a chance process that leads to well-defined results called outcomes.

An **outcome** is the result of a single trial of a probability experiment.

A **sample space** is the set of all possible outcomes of a probability experiment.

Experiment	Sample space
Toss one coin	Head, tail
Roll a die	1, 2, 3, 4, 5, 6
Answer a true/false question	True, false
Toss two coins	Head-head, tail-tail, head-tail, tail-head

EXAMPLE 4–1 Rolling Dice

Find the sample space for rolling two dice.

SOLUTION

Since each die can land in six different ways, and two dice are rolled, the sample space can be presented by a rectangular array, as shown in Figure 4–1. The sample space is the list of pairs of numbers in the chart.

FIGURE 4–1
Sample Space for Rolling Two Dice (Example 4–1)

	Die 2					
Die 1	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

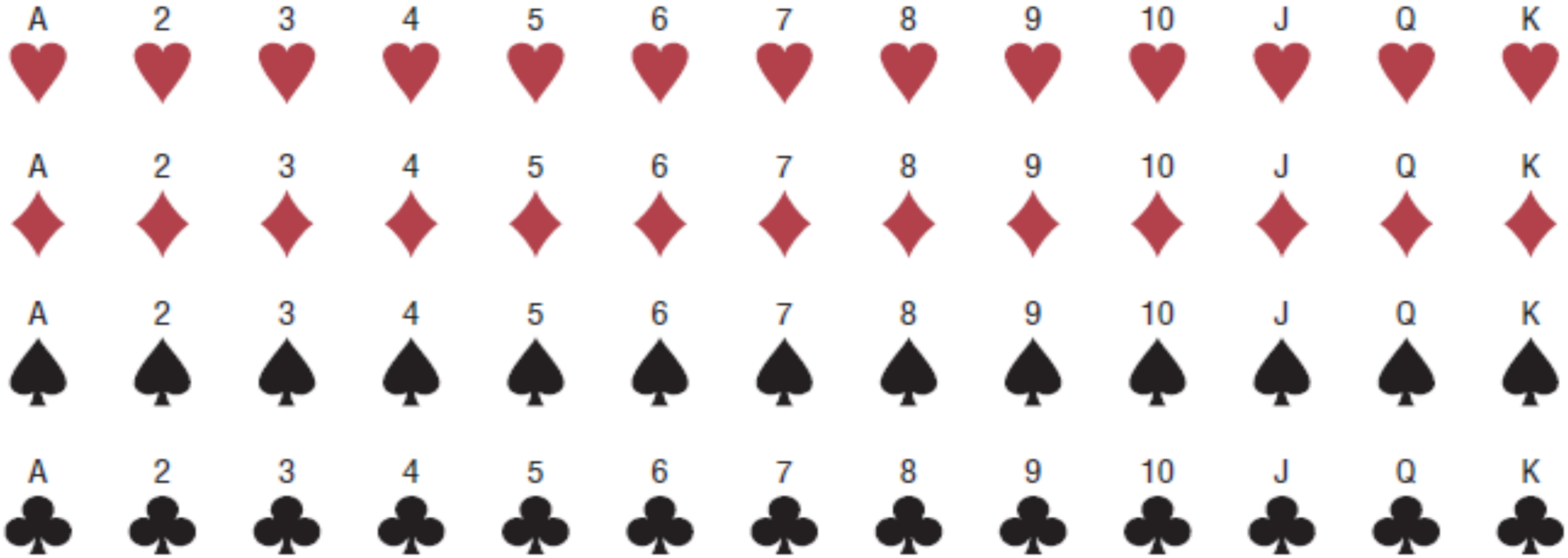
EXAMPLE 4–2 Drawing Cards

Find the sample space for drawing one card from an ordinary deck of cards.

SOLUTION

Since there are 4 suits (hearts, clubs, diamonds, and spades) and 13 cards for each suit (ace through king), there are 52 outcomes in the sample space. See Figure 4–2.

FIGURE 4–2 Sample Space for Drawing a Card (Example 4–2)



EXAMPLE 4–3 Gender of Children

Find the sample space for the gender of the children if a family has three children. Use B for boy and G for girl.

SOLUTION

There are two genders, boy and girl, and each child could be either gender. Hence, there are eight possibilities, as shown here.

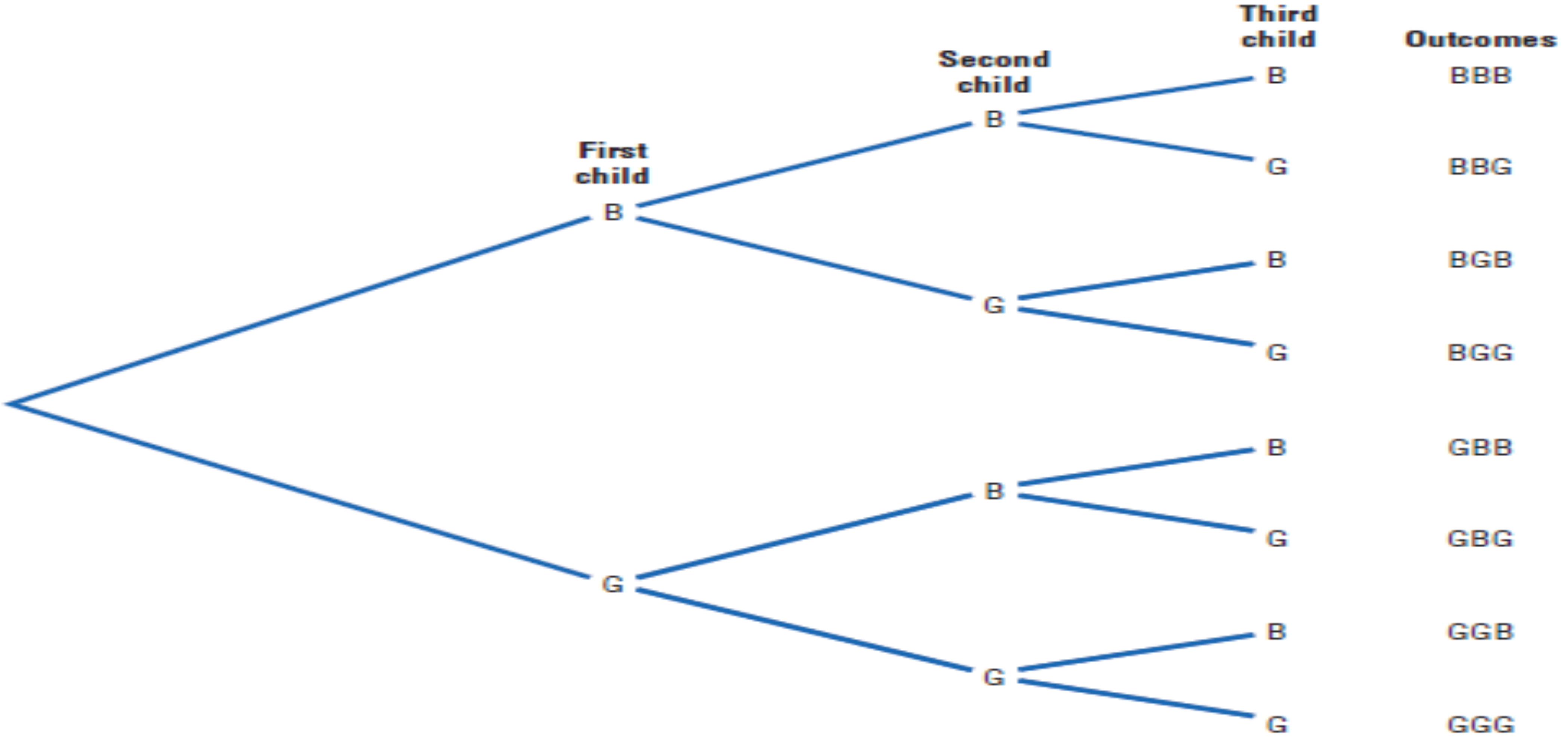
BBB BBG BGB GBB GGG GGB GBG BGG

EXAMPLE 4–4 Gender of Children

Use a tree diagram to find the sample space for the gender of three children in a family, as in Example 4–3.

Since there are two possibilities (boy or girl) for the first child, draw two branches from a starting point and label one B and the other G. Then if the first child is a boy, there are two possibilities for the second child (boy or girl), so draw two branches from B and label one B and the other G. Do the same if the first child is a girl. Follow the same procedure for the third child. The completed tree diagram is shown in Figure 4–3. To find the outcomes for the sample space, trace through all the possible branches, beginning at the starting point for each one.

FIGURE 4–3 Tree Diagram for Example 4–4



An **event** consists of a set of outcomes of a probability experiment.

samplespace : 총 집합

An event can be one outcome or more than one outcome. For example, if a die is rolled and a 6 shows, this result is called an *outcome*, since it is a result of a single trial. An event with one outcome is called a **simple event.** The event of getting an odd number when a die is rolled is called a **compound event,** since it consists of three outcomes or three simple events. In general, a compound event consists of two or more outcomes or simple events.

벤다이어그램, 트리로 구성된거 그려봄

In a group of people, some are in favor of genetic engineering and others are against it. Two persons are selected at random from this group and asked whether they are in favor of or against genetic engineering. How many distinct outcomes are possible? Draw a Venn diagram and a tree diagram for this experiment. List all the outcomes included in each of the following events and state whether they are simple or compound events.

(a) Both persons are in favor of genetic engineering.

(b) At most one person is against genetic engineering.

(c) Exactly one person is in favor of genetic engineering.

$$F = \text{a person is in favor of genetic engineering}$$
$$A = \text{a person is against genetic engineering}$$

This experiment has the following four outcomes:

$$FF = \text{both persons are in favor of genetic engineering}$$
$$FA = \text{the first person is in favor and the second is against}$$
$$AF = \text{the first person is against and the second is in favor}$$
$$AA = \text{both persons are against genetic engineering}$$

The Venn and tree diagrams in Figure 4.5 show these four outcomes.

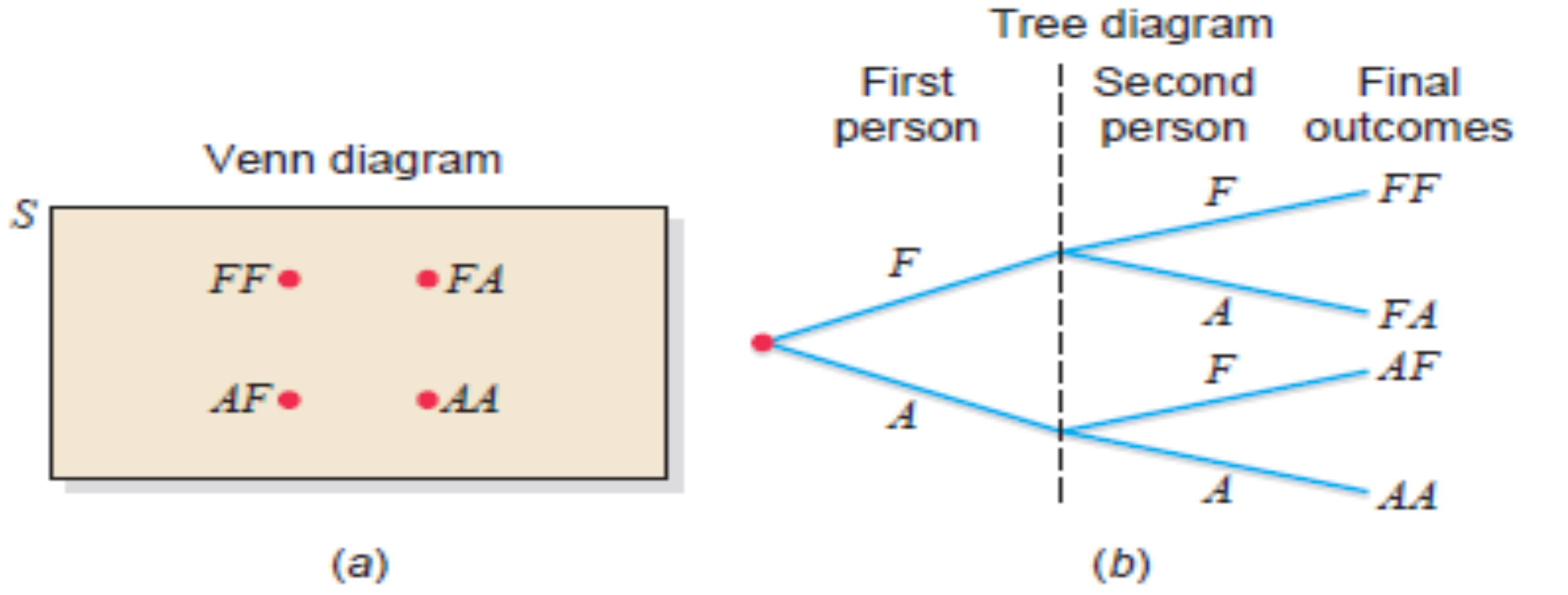


Figure 4.5 Venn and tree diagrams.

(a) The event "both persons are in favor of genetic engineering" will occur if FF is obtained. Thus,

$$\text{Both persons are in favor of genetic engineering} = \{FF\}$$

Because this event includes only one of the final four outcomes, it is a **simple event**.

(b) The event "at most one person is against genetic engineering" will occur if either none or one of the persons selected is against genetic engineering. Consequently,

$$\text{At most one person is against genetic engineering} = \{FF, FA, AF\}$$

Because this event includes more than one outcome, it is a **compound event**.

(c) The event "exactly one person is in favor of genetic engineering" will occur if one of the two persons selected is in favor and the other is against genetic engineering. Hence, it includes the following two outcomes:

$$\text{Exactly one person is in favor of genetic engineering} = \{FA, AF\}$$

Because this event includes more than one outcome, it is a **compound** event. ■

4.5 In a group of adults, some have iPods, and others do not. If two adults are randomly selected from this group, how many total outcomes are possible? Draw a tree diagram for this experiment.

Simple cump um?

경우의수 1개냐 or 여러개냐.

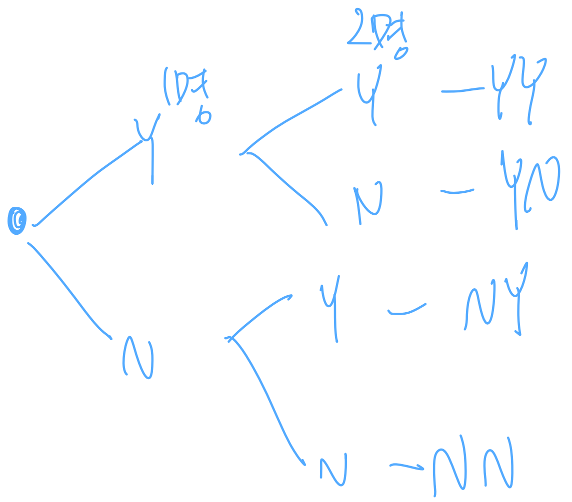
4.11 Refer to Exercise 4.5. List all the outcomes included in each of the following events. Indicate which are simple and which are compound events.

a. One person has an iPod and the other does not.

b. At least one person has an iPod.

c. Not more than one person has an iPod.

d. The first person has an iPod and the second does not.



4.5 four possible outcomes; $S = \{NN, NI, IN, II\}$

4.11 a. {NI and IN}; a compound event

b. {II, NI, and IN}; a compound event

c. {NN, IN, and NI}; a compound event

d. {IN}; a simple event

There are three basic interpretations of probability:

1. Classical probability
2. Empirical or relative frequency probability
3. Subjective probability

Classical probability uses sample spaces to determine the numerical probability that an event will happen. You do not actually have to perform the experiment to determine that probability. Classical probability is so named because it was the first type of probability studied formally by mathematicians in the 17th and 18th centuries.

Classical probability assumes that all outcomes in the sample space are equally likely to occur. For example, when a single die is rolled, each outcome has the same probability of occurring. Since there are six outcomes, each outcome has a probability of $\frac{1}{6}$.

Equally likely events are events that have the same probability of occurring.

Formula for Classical Probability

The probability of any event E is

$$\frac{\text{Number of outcomes in } E}{\text{Total number of outcomes in the sample space}}$$

This probability is denoted by

$$P(E) = \frac{n(E)}{n(S)}$$

where $n(E)$ is the number of outcomes in E and $n(S)$ is the number of outcomes in the sample space S.

EXAMPLE 4–5 Drawing Cards

Find the probability of getting a red face card (jack, queen, or king) when randomly drawing a card from an ordinary deck.

SOLUTION

There are 52 cards in an ordinary deck of cards, and there are 6 red face cards (jack, queen, and king of hearts and jack, queen, and king of diamonds). Hence, the probability of getting a red face card is

$$\frac{6}{52} = \frac{3}{26} \approx 0.115$$

EXAMPLE 4–6 Gender of Children

If a family has three children, find the probability that exactly two of the three children are girls.

SOLUTION

The sample space for the gender of the children for a family with three children has eight outcomes, that is, BBB, BBG, BGB, GBB, GGG, GGB, GBG, and BGG. (See Examples 4–3 and 4–4.) Since there are three ways to have two girls, namely, GGB, GBG, and BGG, $P(\text{two girls}) = \frac{3}{8}$.

EXAMPLE 4–7 Drawing Cards

A card is drawn from an ordinary deck. Find the probability of getting

a. A king

b. The 4 of spades

c. A face card (jack, queen, or king)

d. A red card

e. A club

a. Refer to the sample space in Figure 4–2. There are 4 kings in the event E and a total of 52 cards; hence,

$$P(\text{king}) = \frac{4}{52} = \frac{1}{13} \approx 0.077$$

b. Since there is only one 4 of spades, the probability is

$$P(4\spadesuit) = \frac{1}{52} \approx 0.019$$

c. There are 3 face cards for each suit, and there are 4 suits (hearts, clubs, diamonds, and spades). So there are 12 face cards; hence,

$$P(\text{face card}) = \frac{12}{52} = \frac{3}{13} \approx 0.231$$

d. There are 26 red cards: 13 diamonds and 13 hearts. Hence,

$$P(\text{red card}) = \frac{26}{52} = \frac{1}{2} = 0.5$$

e. There are 13 clubs, so the probability of selecting a club is

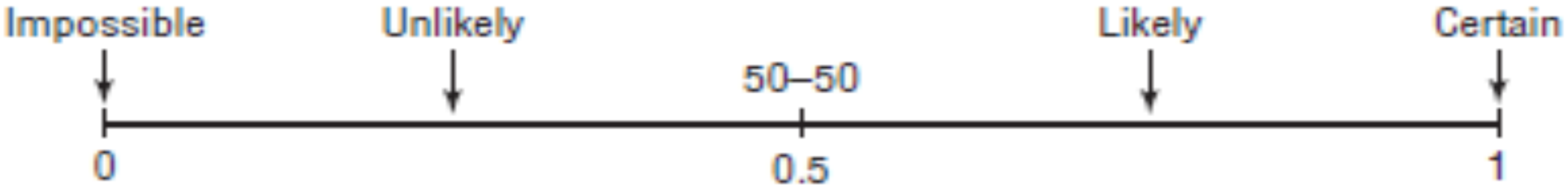
$$P(\clubsuit) = \frac{13}{52} = \frac{1}{4} = 0.25$$

통계학 성립조건

Probability Rules

1. The probability of any event E is a number (either a fraction or decimal) between and including 0 and 1. This is denoted by $0 \le P(E) \le 1$.
2. The sum of the probabilities of all the outcomes in a sample space is 1.
3. If an event E cannot occur (i.e., the event contains no members in the sample space), its probability is 0.
4. If an event E is certain, then the probability of E is 1.

Rule 1 states that probability values range from 0 to 1. When the probability of an event is close to 0, its occurrence is highly unlikely. When the probability of an event is near 0.5, there is about a 50-50 chance that the event will occur; and when the probability of an event is close to 1, the event is highly likely to occur. See Figure 4–4.



Rule 2 can be illustrated by the example of rolling a single die. Each outcome in the sample space has a probability of $\frac{1}{6}$, and the sum of the probabilities of all the outcomes is 1, as shown.

Outcome	1		2		3		4		5		6
Probability	$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$		$\frac{1}{6}$
Sum	$\frac{1}{6}$	$+$	$\frac{1}{6}$	$+$	$\frac{1}{6}$	$+$	$\frac{1}{6}$	$+$	$\frac{1}{6}$	$+$	$\frac{1}{6} = \frac{6}{6} = 1$

EXAMPLE 4–8 Rolling a Die

When a single die is rolled, find the probability of getting a 9.

SOLUTION

Since the sample space is 1, 2, 3, 4, 5, and 6, it is impossible to get a 9. Hence, the probability is $P(9) = \frac{0}{6} = 0$.

EXAMPLE 4–9 Rolling a Die

When a single die is rolled, what is the probability of getting a number less than 7?

SOLUTION

Since all outcomes—1, 2, 3, 4, 5, and 6—are less than 7, the probability is

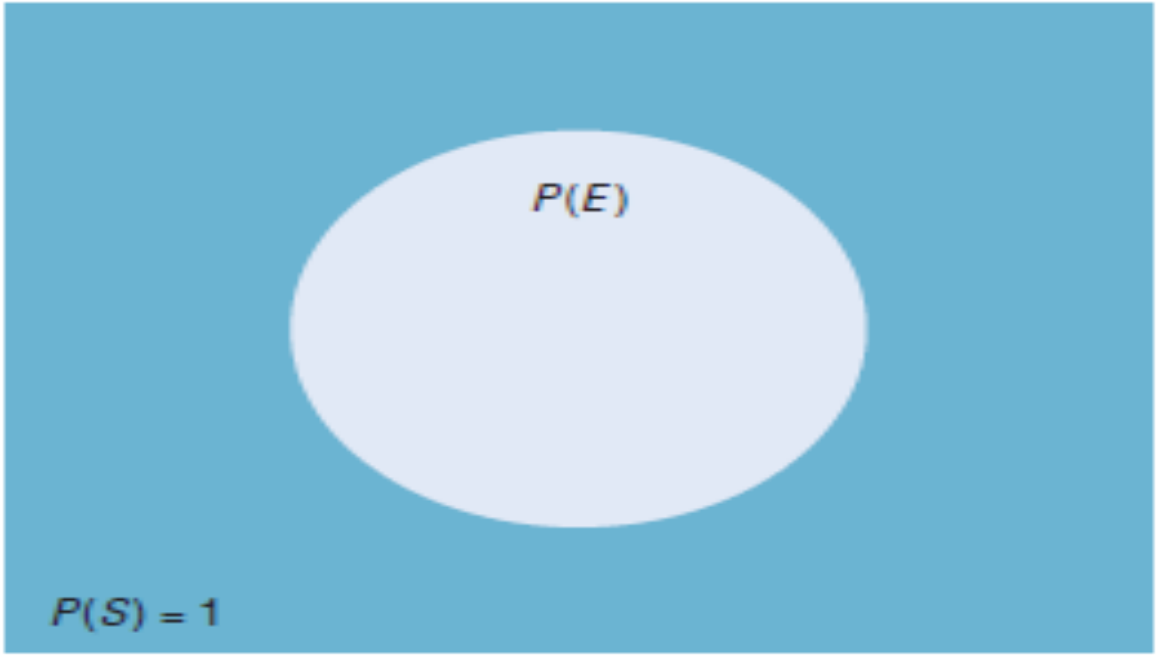
$$P(\text{number less than 7}) = \frac{6}{6} = 1$$

The event of getting a number less than 7 is certain.

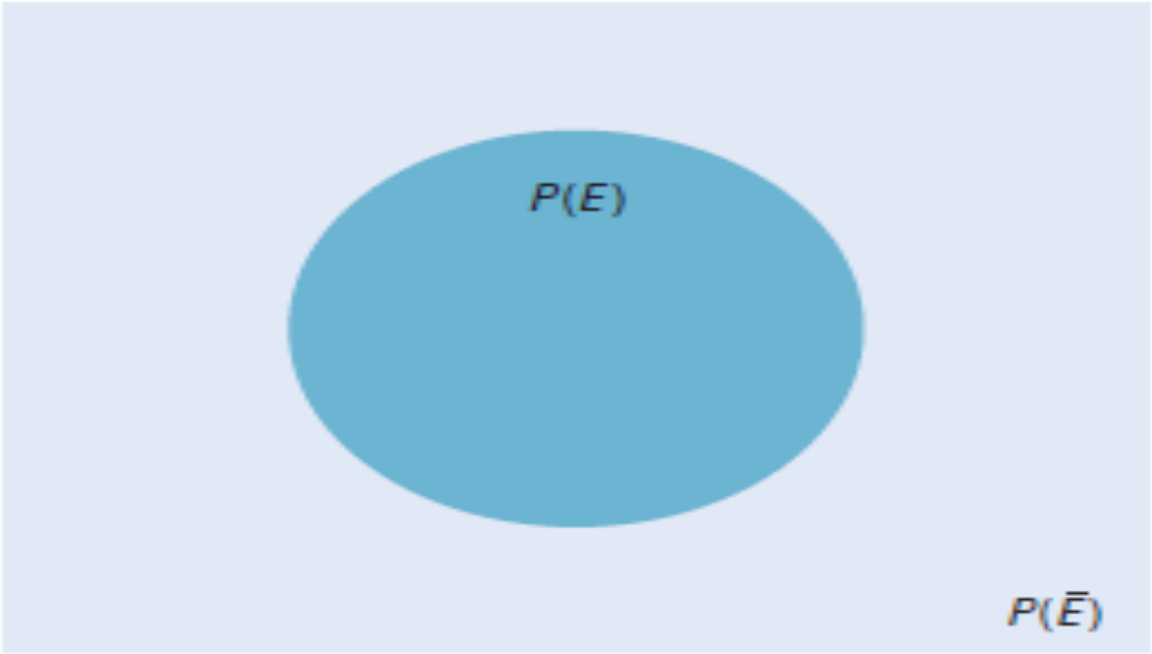
The **complement of an event** E is the set of outcomes in the sample space that are not included in the outcomes of event E. The complement of E is denoted by $\overline{E}$ (read "E bar").

Rule for Complementary Events

$$P(\overline{E}) = 1 - P(E) \qquad \text{or} \qquad P(E) = 1 - P(\overline{E}) \qquad \text{or} \qquad P(E) + P(\overline{E}) = 1$$



(a) Simple probability



(b) $P(\bar{E}) = 1 - P(E)$

EXAMPLE 4–10 Finding Complements

Find the complement of each event:

a. Selecting a month that has 31 days

b. Selecting a day of the week that begins with the letter T

c. Rolling two dice and getting a sum that is an odd number

d. Selecting a letter of the alphabet that is used as a vowel or consonant

SOLUTION

a. Select a month that has fewer than 31 days, i.e., February, April, June, September, and November.

b. Select a day of the week that does not begin with a T, i.e., Sunday, Monday, Wednesday, Friday, and Saturday.

c. Rolling two dice and getting a sum that is an even number, i.e., a sum of 2, 4, 6, 8, 10, or 12.

d. Since y is the only letter that is used as a vowel or a consonant, the complement is selecting a letter other than y.

EXAMPLE 4–11 Favorite Ice Cream Flavors

In a study, it was found that 23% of the people surveyed said that vanilla was their favorite flavor of ice cream. If a person is selected at random, find the probability that the person's favorite flavor of ice cream is not vanilla.

Source: Rasmussen Report.

SOLUTION

$$
\begin{aligned}
P(\text{not vanilla}) &= 1 - P(\text{vanilla}) \\
&= 1 - 0.23 = 0.77 = 77\%
\end{aligned}
$$

Empirical Probability

The difference between classical and **empirical probability** is that classical probability assumes that certain outcomes are equally likely (such as the outcomes when a die is rolled), while empirical probability relies on actual experience to determine the likelihood of outcomes. In empirical probability, one might actually roll a given die 6000 times, observe the various frequencies, and use these frequencies to determine the probability of an outcome.

Formula for Empirical Probability

Given a frequency distribution, the probability of an event being in a given class is

$$P(E) = \frac{\text{frequency for the class}}{\text{total frequencies in the distribution}} = \frac{f}{n}$$

This probability is called *empirical probability* and is based on observation.

Suppose, for example, that a researcher for the American Automobile Association (AAA) asked 50 people who plan to travel over the Thanksgiving holiday how they will get to their destination. The results can be categorized in a frequency distribution as shown.

Method	Frequency
Drive	41
Fly	6
Train or bus	3
	50

EXAMPLE 4–12 Travel Survey

In the travel survey just described, find the probability that a person will travel by airplane over the Thanksgiving holiday.

SOLUTION

$$P(E) = \frac{f}{n} = \frac{6}{50} = \frac{3}{25}$$

EXAMPLE 4–13 Distribution of Blood Types

In a sample of 50 people, 21 had type O blood, 22 had type A blood, 5 had type B blood, and 2 had type AB blood. Set up a frequency distribution and find the following probabilities.

a. A person has type O blood.

b. A person has type A or type B blood.

c. A person has neither type A nor type O blood.

d. A person does not have type AB blood.

Type	Frequency
A	22
B	5
AB	2
O	21
	Total 50

a. $P(\text{O}) = \dfrac{f}{n} = \dfrac{21}{50}$

b. $P(\text{A or B}) = \dfrac{22}{50} + \dfrac{5}{50} = \dfrac{27}{50}$

(Add the frequencies of the two classes.)

c. $P(\text{neither A nor O}) = \dfrac{5}{50} + \dfrac{2}{50} = \dfrac{7}{50}$

(Neither A nor O means that a person has either type B or type AB blood.)

d. $P(\text{not AB}) = 1 - P(\text{AB}) = 1 - \dfrac{2}{50} = \dfrac{48}{50} = \dfrac{24}{25}$

(Find the probability of not AB by subtracting the probability of type AB from 1.)

Hospital records indicated that knee replacement patients stayed in the hospital for the number of days shown in the distribution.

Number of days stayed	Frequency
3	15
4	32
5	56
6	19
7	5
	127

Find these probabilities.

a. A patient stayed exactly 5 days.

b. A patient stayed fewer than 6 days.

c. A patient stayed at most 4 days.

d. A patient stayed at least 5 days.

a. $P(5) = \dfrac{56}{127}$

b. $P(\text{fewer than 6 days}) = \dfrac{15}{127} + \dfrac{32}{127} + \dfrac{56}{127} = \dfrac{103}{127}$

(Fewer than 6 days means 3, 4, or 5 days.)

c. $P(\text{at most 4 days}) = \dfrac{15}{127} + \dfrac{32}{127} = \dfrac{47}{127}$

(At most 4 days means 3 or 4 days.)

d. $P(\text{at least 5 days}) = \dfrac{56}{127} + \dfrac{19}{127} + \dfrac{5}{127} = \dfrac{80}{127}$

(At least 5 days means 5, 6, or 7 days.)

Subjective Probability

The third type of probability is called *subjective probability.* **Subjective probability** uses a probability value based on an educated guess or estimate, employing opinions and inexact information.

In subjective probability, a person or group makes an educated guess at the chance that an event will occur. This guess is based on the person's experience and evaluation of a solution. For example, a sportswriter may say that there is a 70% probability that the Pirates will win the pennant next year. A physician might say that, on the basis of her diagnosis, there is a 30% chance the patient will need an operation. A seismologist might say there is an 80% probability that an earthquake will occur in a certain area. These are only a few examples of how subjective probability is used in everyday life.

All three types of probability (classical, empirical, and subjective) are used to solve a variety of problems in business, engineering, and other fields.

4.19 Which of the following values cannot be probabilities of events and why?

1/5	.97	−.55	1.56	5/3	0.0	−2/7	1.0

4.23 The president of a company has a hunch that there is a .80 probability that the company will be successful in marketing a new brand of ice cream. Is this a case of classical, relative frequency, or subjective probability? Explain why.

4.25 A hat contains 40 marbles. Of them, 18 are red and 22 are green. If one marble is randomly selected out of this hat, what is the probability that this marble is

a. red? b. green?

4.19 $-.55$, 1.56, $5/3$, $-2/7$

4.23 subjective probability

4.25 a. .450 b. .550

확률의 덧셈법칙

Two events are **mutually exclusive events** or **disjoint events** if they cannot occur at the same time (i.e., they have no outcomes in common).

1. The person is a Democrat.
2. The person is an Independent.

EXAMPLE 4–16 Drawing a Card

Determine which events are mutually exclusive and which events are not when a single card is drawn from a deck.

a. Getting a king; getting a diamond

b. Getting a 4; getting a king

c. Getting a face card; getting a club

d. Getting a face card; getting a 10

SOLUTION

a. These events are not mutually exclusive since the king of diamonds represents both events.

b. These events are mutually exclusive since you cannot draw one card that is both a 4 and a king.

c. These events are not mutually exclusive since a jack, queen, or king can be clubs.

d. These events are mutually exclusive since a 10 and a face card cannot be drawn at the same time when one card is drawn.

V → [illegible]

Determine whether the two events are mutually exclusive. Explain your answer.

a. Randomly selecting a female student
Randomly selecting a student who is a junior

b. Randomly selecting a person with type A blood
Randomly selecting a person with type O blood

c. Rolling a die and getting an odd number
Rolling a die and getting a number less than 3

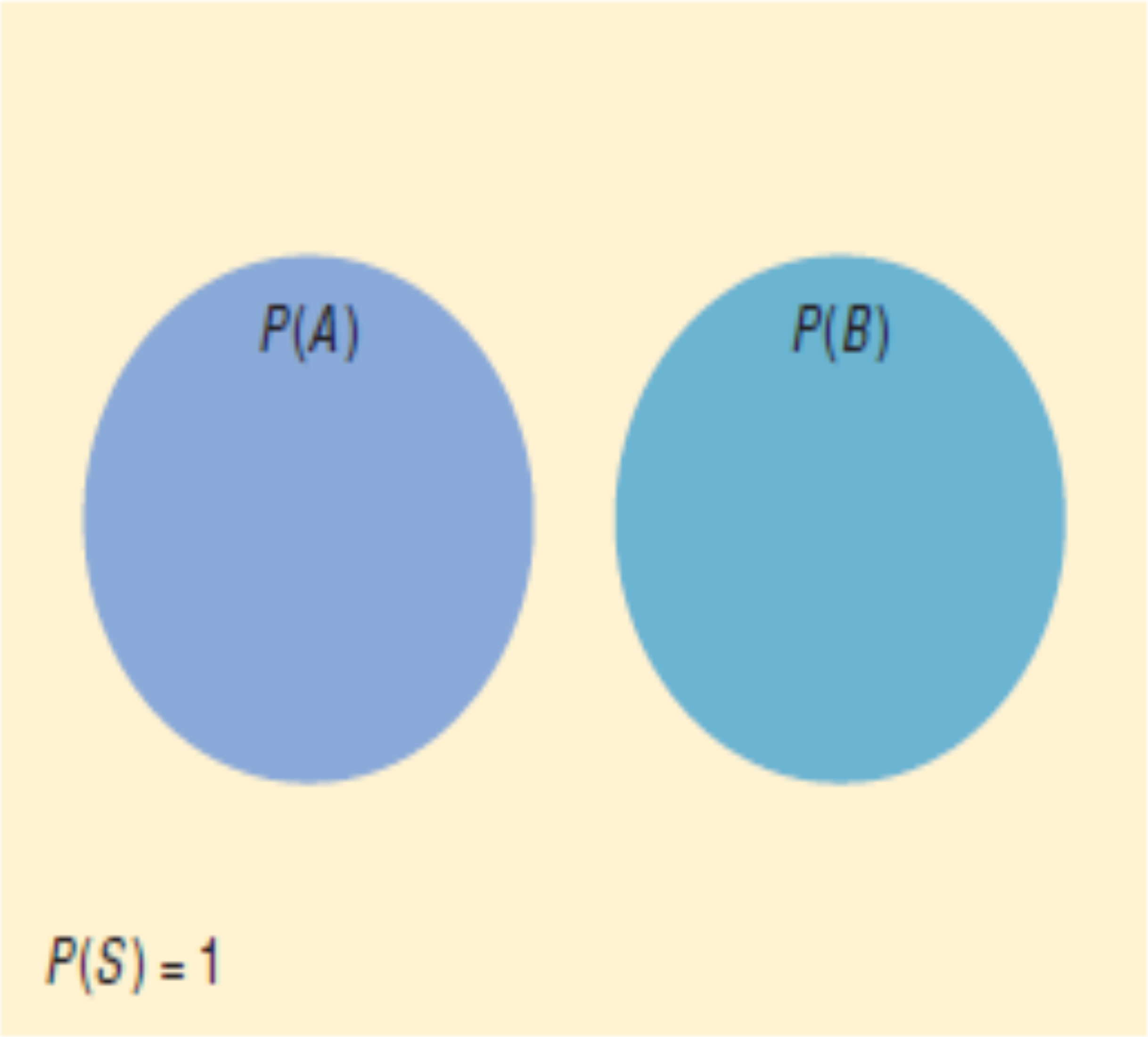
d. Randomly selecting a person who is under 21 years of age
Randomly selecting a person who is over 30 years of age

a. These events are not mutually exclusive since a student can be both female and a junior.

b. These events are mutually exclusive since a person cannot have type A blood and type O blood at the same time.

c. These events are not mutually exclusive since the number 1 is both an odd number and a number less than 3.

d. These events are mutually exclusive since a person cannot be both under 21 and over 30 years of age at the same time.



Mutually exclusive events
$P(A \text{ or } B) = P(A) + P(B)$

Addition Rule 1

When two events A and B are mutually exclusive, the probability that A or B will occur is

$$P(A \text{ or } B) = P(A) + P(B)$$

EXAMPLE 4–17 Coffee Shop Selection

A city has 9 coffee shops: 3 Starbucks, 2 Caribou Coffees, and 4 Crazy Mocho Coffees. If a person selects one shop at random to buy a cup of coffee, find the probability that it is either a Starbucks or Crazy Mocho Coffees.

SOLUTION

Since there are 3 Starbucks and 4 Crazy Mochos, and a total of 9 coffee shops, $P(\text{Starbucks or Crazy Mocho}) = P(\text{Starbucks}) + P(\text{Crazy Mocho}) = \frac{3}{9} + \frac{4}{9} = \frac{7}{9} \approx 0.778$. The events are mutually exclusive.

EXAMPLE 4–18 Research and Development Employees

The corporate research and development centers for three local companies have the following numbers of employees:

U.S. Steel	110
Alcoa	750
Bayer Material Science	250

If a research employee is selected at random, find the probability that the employee is employed by U.S. Steel or Alcoa.

Source: *Pittsburgh Tribune Review.*

SOLUTION

$$P(\text{U.S. Steel or Alcoa}) = P(\text{U.S. Steel}) + P(\text{Alcoa})$$

$$= \frac{110}{1110} + \frac{750}{1110} = \frac{860}{1110} = \frac{86}{111}$$

EXAMPLE 4–19 Favorite Ice Cream

In a survey, 8% of the respondents said that their favorite ice cream flavor is cookies and cream, and 6% like mint chocolate chip. If a person is selected at random, find the probability that her or his favorite ice cream flavor is either cookies and cream or mint chocolate chip.

Source: Rasmussen Report.

SOLUTION

$$\begin{aligned} &P(\text{cookies and cream or mint chocolate chip}) \\ &\quad = P(\text{cookies and cream}) + P(\text{mint chocolate chip}) \\ &\quad = 0.08 + 0.06 = 0.14 = 14\% \end{aligned}$$

These events are mutually exclusive.

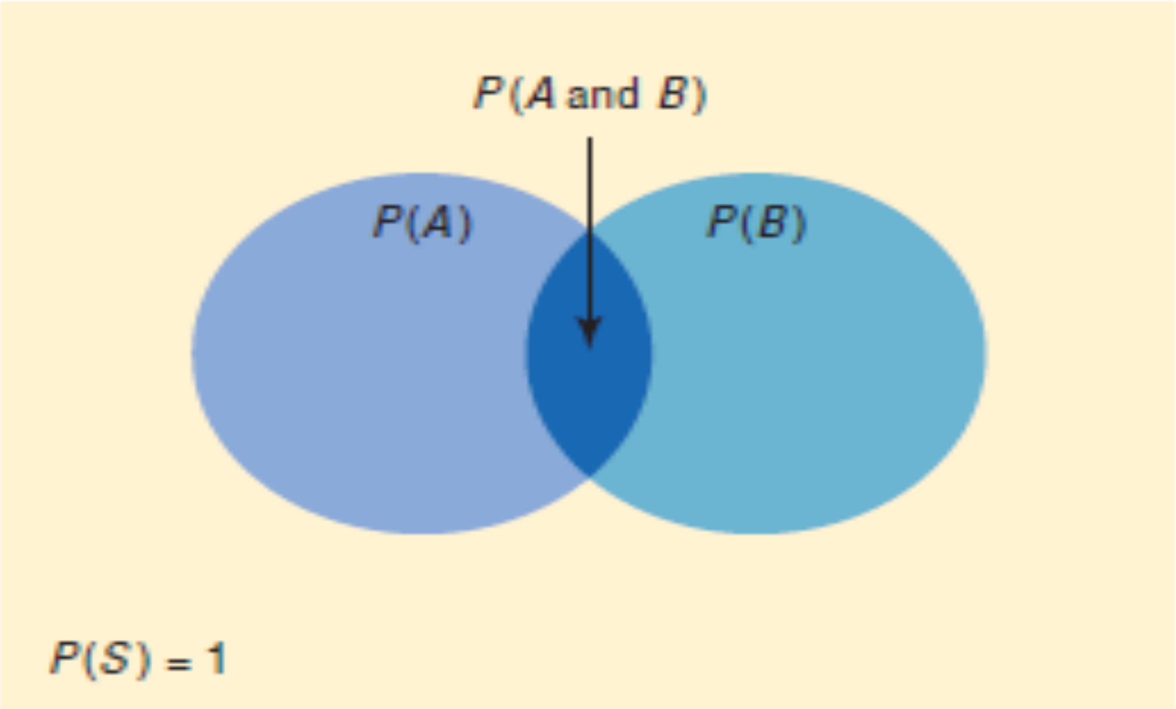
The probability rules can be extended to three or more events. For three mutually exclusive events A, B, and C,

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$$

Addition Rule 2

If A and B are *not* mutually exclusive, then

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$



Nonmutually exclusive events
$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

EXAMPLE 4–20 Drawing a Card

A single card is drawn at random from an ordinary deck of cards. Find the probability that it is either an ace or a black card.

SOLUTION

Since there are 4 aces and 26 black cards (13 spades and 13 clubs), 2 of the aces are black cards, namely, the ace of spades and the ace of clubs. Hence, the probabilities of the two outcomes must be subtracted since they have been counted twice.

$$
\begin{aligned}
P(\text{ace or black card}) &= P(\text{ace}) + P(\text{black card}) - P(\text{black aces}) \\
&= \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52} = \frac{7}{13} \approx 0.538
\end{aligned}
$$

EXAMPLE 4–21 Selecting a Medical Staff Person

In a hospital unit there are 8 nurses and 5 physicians; 7 nurses and 3 physicians are females. If a staff person is selected, find the probability that the subject is a nurse or a male.

SOLUTION

The sample space is shown here.

Staff	Females	Males	Total
Nurses	7	1	8
Physicians	3	2	5
Total	10	3	13

The probability is

$$P(\text{nurse or male}) = P(\text{nurse}) + P(\text{male}) - P(\text{male nurse})$$

$$= \frac{8}{13} + \frac{3}{13} - \frac{1}{13} = \frac{10}{13} \approx 0.769$$

EXAMPLE 4–22 Driving While Intoxicated

On New Year's Eve, the probability of a person driving while intoxicated is 0.32, the probability of a person having a driving accident is 0.09, and the probability of a person having a driving accident while intoxicated is 0.06. What is the probability of a person driving while intoxicated or having a driving accident?

SOLUTION

$$\begin{aligned} P(\text{intoxicated or accident}) &= P(\text{intoxicated}) + P(\text{accident}) \\ &\quad - P(\text{intoxicated and accident}) \\ &= 0.32 + 0.09 - 0.06 = 0.35 \end{aligned}$$

For three events that are *not* mutually exclusive,

$$\begin{aligned} P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) - P(A \text{ and } B) - P(A \text{ and } C) \\ - P(B \text{ and } C) + P(A \text{ and } B \text{ and } C) \end{aligned}$$

Assume that following an injury you received from playing your favorite sport, you obtain and read information on new pain medications. In that information you read of a study that was conducted to test the side effects of two new pain medications. Use the following table to answer the questions and decide which, if any, of the two new pain medications you will use.

	Number of side effects in 12-week clinical trial		
Side effect	Placebo $n = 192$	Drug A $n = 186$	Drug B $n = 188$
Upper respiratory congestion	10	32	19
Sinus headache	11	25	32
Stomach ache	2	46	12
Neurological headache	34	55	72
Cough	22	18	31
Lower respiratory congestion	2	5	1

1. How many subjects were in the study?
2. How long was the study?

6. What is the probability that a randomly selected person was receiving a placebo?
7. What is the probability that a person was receiving a placebo or drug A? Are these mutually exclusive events? What is the complement to this event?
8. What is the probability that a randomly selected person was receiving a placebo or experienced a neurological headache?
9. What is the probability that a randomly selected person was not receiving a placebo or experienced a sinus headache?

1. There were $192 + 186 + 188 = 566$ subjects in the study.

2. The study lasted for 12 weeks.

6. The probability that a randomly selected person was receiving a placebo is $192/566 = 0.339$ (about 34%).

7. The probability that a randomly selected person was receiving a placebo or drug A is $(192 + 186)/566 = 378/566 = 0.668$ (about 67%). These are mutually exclusive events. The complement is that a randomly selected person was receiving drug B.

8. The probability that a randomly selected person was receiving a placebo or experienced a neurological headache is $(192 + 55 + 72)/566 = 319/566 = 0.564$ (about 56%).

9. The probability that a randomly selected person was not receiving a placebo or experienced a sinus headache is $(186 + 188)/566 + 11/566 = 385/566 = 0.680$ (about 68%).

- 과제 1: 09/21/2025 (일요일) 이클래스에 올라갈 예정임

- 제출기한: 10/02/2025 (목요일) 까지 제출하기

- 제출형식: 손으로 쓰거나 프린트해서 하드카피 수업시간에 제출

Ĝis revido!