[연습문제 해답 이용 안내]

Chapter 01 연습문제

1.1

(a) $\bar{x} = \frac{1}{12}(1 + 3 + \cdots + 8) = 5$

(b) 12개의 자료를 크기순으로 나열하면 다음과 같다.

1, 2, 2, 3, 3, 5, 5, 5, 8, 8, 9, 9

따라서 중앙값은 $Me = \frac{1}{2}(5 + 5) = 5$ 이다.

(c) 빈도수가 3으로 가장 많은 5가 최빈값이다.

(d) $Q_1 = 2.5$, $Q_2 = 5$, $Q_3 = 8$

(e) 사분위수 범위$= 8 - 2.5 = 5.5$

(f) $s^2 = \frac{1}{12-1}\{(1-5)^2 + (3-5)^2 + \cdots + (8-5)^2\} = 8.36$

$s = \sqrt{8.36} = 2.892$

(g) $C.V = \frac{2.892}{5} \times 100 = 57.84(\%)$

(h) 왜도≒0.2165

(i) 첨도≒-1.5013

(j) $[x_{\min}, Q_1, Me, Q_3, x_{\max}] = [1, 2, 5, 8, 9]$

1.3

Frequency	Stem &	Leaf
1.00	1 .	6
2.00	2 .	28
8.00	3 .	00233678
4.00	4 .	0458
7.00	5 .	1234669
3.00	6 .	267
1.00	7 .	3
1.00	8 .	2

Stem width: 10

Each leaf: 1 case(s)

1.5

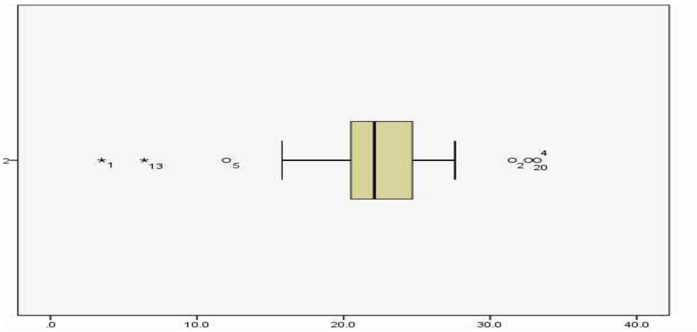
(a) 제1사분위수 $Q_1 = 20.5$, 제2사분위수 $Q_2 = 22.1$, 제3사분위수 $Q_3 = 24.7$

(b) 사분위수 범위(I.Q.R)$= Q_3 - Q_1 = 4.2$

(c) $f_l = Q_1 - 1.5 \times IQR = 20.5 - 1.5 \times 4.2 = 14.2$

$f_u = Q_3 + 1.5 \times IQR = 24.7 + 1.5 \times 4.2 = 31$

(d), (e)



1.7

(a) 평균이 1.76이므로 다음이 성립한다.

값	2.5	1.5	1.8	1.3	1.7	합
편차	2.5-1.76=0.74	1.5-1.76=-0.26	1.8-1.76=0.04	1.3-1.76=-0.46	1.7-1.76=-0.06	0

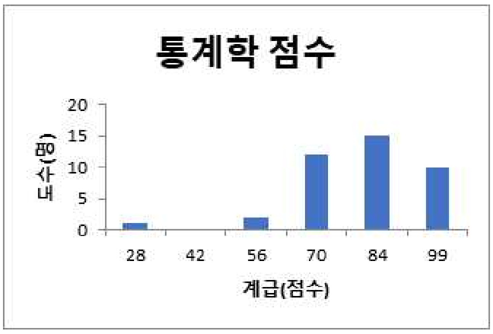
(b) 분산 : 0.208, 표준편차 : 0.456

1.9

(a) 자료의 최댓값은 99, 최솟값은 15이고 계급의 수가 6이므로 계급의 간격은 $(99-15)/6 = 14$이다. 따라서 계급의 간격을 14로 정하고, 각 계급의 도수와 상대도수를 구하여 도수분포표를 작성하면 다음과 같다.

계급(세)	계급값	도수	상대도수
15이상 ~ 29미만	22	1	1/40
29 ~43	36	0	0
43 ~ 57	50	2	2/40
57 ~ 71	64	12	12/40
71 ~85	78	15	15/40
85 ~ 100	92.5	10	10/40
합계	-	40	1

(b) 도수분포표에 따라 가로축에는 계급, 세로축에는 도수를 기입하여 히스토그램을 그리면 다음과 같다. 한편, 이 그래프에서 세로축을 상대도수로 바꿔도 그래프의 모양은 변함이 없다.



(c)

평균 : 74.525

분산 : 232.2994

1.11

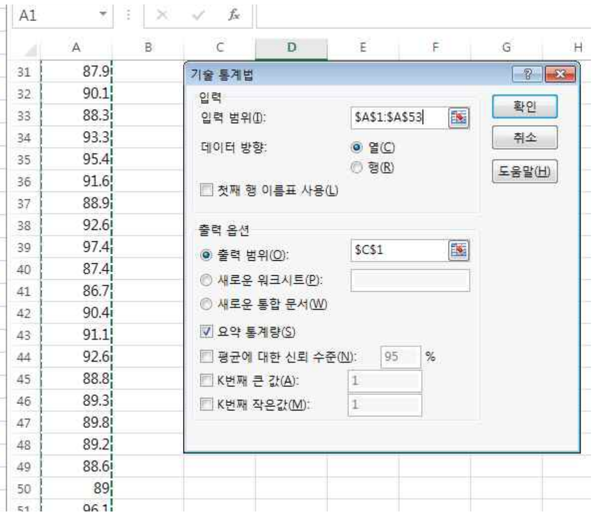
(a) $Q_1 = 42 \sim 46$, $Q_2 = 46 \sim 50$, $Q_3 = 54 \sim 58$

(b) 평균 : 50.933, 표준편차 : 7.926

1.13

1.15

다음과 같이 [데이터분석] ⇒ [기술 통계법]을 선택하고 데이터를 입력한다.



출력 결과를 보면 다음과 같다.

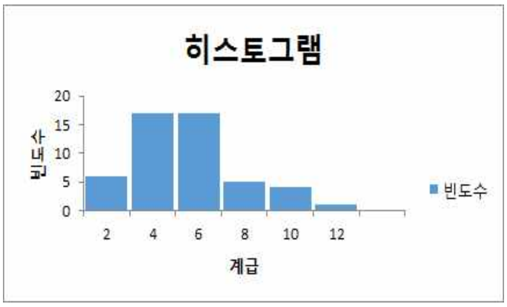
Column1	
평균	90.91887
표준오차	0.514055
중앙값	90.4
최빈값	91
표준편차	3.74238
분산	14.00541
첨도	0.318184
왜도	0.260491
범위	18.1
최솟값	82.1
최댓값	100.2
합	4818.7
관측수	53

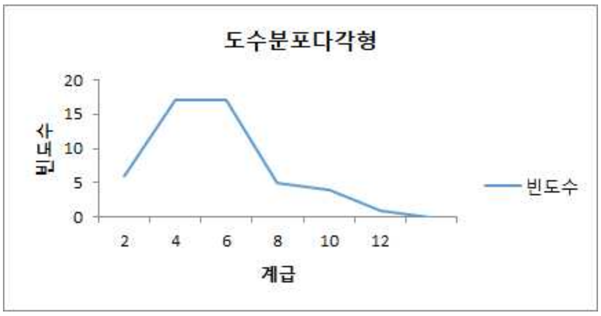
→ 왜도는 0보다 크고 첨도는 3보다 작으므로, 왼쪽으로 약간 치우치고 표준정규분포보다 정점이 낮으며 완만한 분포 형태를 이루고 있다.

1.17

(a) 자료를 워크시트에 입력하고 적당한 계급구간을 정한다.

계급	빈도수
2	6
4	17
6	17
8	5
10	4
12	1





(b) [데이터분석] ⇒ [기술 통계법]을 선택하여 출력 결과를 얻는다.

Column1	
평균	4.734
표준오차	0.324869
중앙값	4.3
최빈값	3.5
표준편차	2.297169
분산	5.276984
첨도	0.367833
왜도	0.770978
범위	10.2
최솟값	1
최댓값	11.2
합	236.7
관측수	50

Chapter 02 연습문제

2.1

(a) $\mathrm{P}(A \bigcup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \bigcap B)$

$$= \frac{1}{2} + \frac{1}{3} - \frac{1}{4} = \frac{7}{12}$$

(b) $\mathrm{P}(A \mid B) = \dfrac{\mathrm{P}(A \bigcap B)}{\mathrm{P}(B)} = \dfrac{\frac{1}{4}}{\frac{1}{3}} = \dfrac{3}{4}$

(c) $\mathrm{P}(B \mid A) = \dfrac{\mathrm{P}(A \bigcap B)}{\mathrm{P}(A)} = \dfrac{\frac{1}{4}}{\frac{1}{2}} = \dfrac{1}{2}$

(d) $\mathrm{P}(B^c \mid A^c) = \dfrac{\mathrm{P}(A^c \bigcap B^c)}{\mathrm{P}(A^c)} = \dfrac{1 - \mathrm{P}(A \bigcup B)}{1 - \mathrm{P}(A)}$

$$= \frac{1 - \frac{7}{12}}{1 - \frac{1}{2}} = \frac{\frac{5}{12}}{\frac{1}{2}} = \frac{5}{6}$$

2.3

주사위 두 개를 던질 때의 표본공간은 $\{(1,1), (1,2), \cdots, (6,6)\}$ 이다.

(a) 나타난 눈의 합이 3 이하인 경우는 $\{(1,1), (1,2), (2,1)\}$ 이므로, 구하고자 하는 확률은 $p = \dfrac{3}{36} = \dfrac{1}{12}$ 이다.

(b) 두 개 모두 같은 눈이 나오는 경우는 $\{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$ 이므로 구하고자 하는 확률은 $p = \dfrac{6}{36} = \dfrac{1}{6}$ 이다.

(c) 나타난 눈의 합이 8인 경우는 $\{(2,6), (3,5), (4,4), (5,3), (6,2)\}$ 이고, 주사위 하나의 눈이 4가 나오는 경우는 $\{(4,4)\}$ 이므로, 구하는 확률은 $p = \dfrac{1}{36}$ 이다.

(d) 첫 번째 던진 주사위의 눈의 수와 두 번째 주사위의 눈의 수의 합이 8일 사상을 A라 하고, 두 주사위의 눈의 수가 같을 사상을 B라 하면 다음과 같다.

$$A = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$$
$$B = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$$

따라서 $\mathrm{P}(A) = \dfrac{5}{36}$, $\mathrm{P}(B) = \dfrac{6}{36}$ 이다. 또한, $A \bigcap B = \{(4,4)\}$ 이므로 $\mathrm{P}(A \bigcap B) = \dfrac{1}{36}$ 이다. 그러므로 구하고자 하는 조건부 확률 $\mathrm{P}(B \mid A)$는 다음과 같다.

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)}$$

$$= \frac{\frac{1}{36}}{\frac{5}{36}} = \frac{1}{5}$$

2.5

$\mathrm{P}(\text{뒷면}) = p$, $\mathrm{P}(\text{앞면}) = 2p$ 이므로 다음이 성립한다.

$$\mathrm{P}(\text{전체}) = \mathrm{P}(\text{뒷면}) + \mathrm{P}(\text{앞면}) = 3p = 1 \;\Rightarrow\; p = \frac{1}{3}$$

따라서 주어진 조건에 의해 $\mathrm{P}(\text{앞면}) = \frac{2}{3}$,$\mathrm{P}(\text{뒷면}) = \frac{1}{3}$ 이다.

2.7

$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) \;\Rightarrow\; 0.6 = 0.4\mathrm{P}(B)$

$\Rightarrow\; \mathrm{P}(B) = 0.2$

2.9

$A \subset B$이므로 $A \cap B = A$이다.

따라서 $\mathrm{P}(B \mid A) = \dfrac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} = \dfrac{\mathrm{P}(A)}{\mathrm{P}(A)} = 1$이다.

2.11

$\mathrm{P}(A \cap B) = \mathrm{P}(B)\mathrm{P}(A \mid B) = 0.6 \times 0.2 = 0.12$ 이므로 다음과 같다.

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} = \frac{0.12}{0.4} = 0.3$$

2.13

$A = \{2, 4, 6, 8, 10\}$, $B = \{5, 10\}$, $A \cap B = \{10\}$ 이므로 다음이 성립한다.

$$\mathrm{P}(A \cap B) = 1/10,\ \mathrm{P}(A) \times \mathrm{P}(B) = \frac{5}{10} \times \frac{2}{10} = \frac{1}{10}$$

따라서 $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$이므로, A와 B는 서로 독립이다.

2.15

$\mathrm{P}(A \bigcup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A)\mathrm{P}(B)$

$\Rightarrow\ \frac{1}{3} = \frac{1}{4} + p - \frac{1}{4p}$

$\Rightarrow\ p = \frac{1}{9}$

2.17

첫 번째 뽑은 구슬이 붉은 색일 확률 : $\frac{7}{10}$

두 번째 뽑은 구슬이 붉은 색일 확률 : $\frac{6}{9}$

세 번째 뽑은 구슬이 흰색일 확률 : $\frac{3}{8}$

따라서 곱셈 공식에 의하여 구하고자 하는 확률은 $p = \frac{7}{10} \times \frac{6}{9} \times \frac{3}{8} = \frac{7}{40}$ 이다.

2.19

(a) 전체 진공관 10개 중 3개를 추출하는 경우의 수는 ${}_{10}\mathrm{C}_3$ 이고, 10개 중 4개가 불량품이므로 양품이 나올 수 있는 것은 6개 중 3개를 추출하는 경우만 가능하므로 ${}_{6}\mathrm{C}_3$ 이 된다. 따라서 불량품이 하나도 나오지 않을 확률은 $p = \frac{{}_{6}\mathrm{C}_3}{{}_{10}\mathrm{C}_3} = \frac{20}{120} = \frac{1}{6}$ 이다.

(b) 불량품 4개 중에서 2개를 추출하고 나머지 양품 6개 중 1개를 추출하는 경우의 수는 ${}_{4}\mathrm{C}_2 \times {}_{6}\mathrm{C}_1$ 이다. 그러므로 구하고자 하는 확률은 $p = \frac{{}_{4}\mathrm{C}_2 \times {}_{6}\mathrm{C}_1}{{}_{10}\mathrm{C}_3} = \frac{36}{120} = \frac{3}{10}$ 이다.

2.21

불량품이 나오는 사건을 X라 하면, 베이즈 정리에 의해 구하고자 하는 확률은 다음과 같다.

$$\mathrm{P}(A \mid X) = \frac{\mathrm{P}(A)\mathrm{P}(X \mid A)}{\mathrm{P}(A)\mathrm{P}(X \mid A) + \mathrm{P}(B)\mathrm{P}(X \mid B)}$$

$$= \frac{\frac{1}{2} \times \frac{3}{10}}{\frac{1}{2} \times \frac{3}{10} + \frac{1}{2} \times \frac{1}{10}}$$

$$= \frac{3}{4}$$

2.23

처음 꺼낸 퓨즈가 불량일 사건을 A, 두 번째 꺼낸 퓨즈가 불량일 사건을 B라고 하자. 그러면 $A \cap B$는 추출된 2개의 퓨즈가 모두 불량일 사건이다.

처음 꺼낸 퓨즈가 불량일 확률은 $\mathrm{P}(A) = \frac{5}{20} = \frac{1}{4}$ 이다. 그리고 처음에 불량 퓨즈가 나왔다면 상자 안에는 19개의 퓨즈가 들어 있고, 이 중 4개가 불량 퓨즈이다. 즉, 두 번째로 꺼낸 퓨즈가 불량일 확률은 처음에 꺼낸 퓨즈를 제외해야 하므로 $\mathrm{P}(B \mid A) = \frac{4}{19}$ 이다. 따라서 구하고자 하는 확률은 다음과 같다.

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B|A) = \frac{1}{4} \times \frac{4}{19} = \frac{1}{19}$$

2.25

폐질환을 앓고 있는 사람을 A라 하면, $\mathrm{P}(A) = 0.07$이므로 여사건의 확률의 $\mathrm{P}(A^c) = 0.93$이다. 임의로 선택한 사람이 흡연자일 사건을 B라 하면, $\mathrm{P}(B|A) = 0.85$, $\mathrm{P}(B|A^c) = 0.25$이다. 따라서 임의로 선택한 사람이 흡연자일 확률은 다음과 같다.

$$\begin{aligned}\mathrm{P}(B) &= \mathrm{P}(B \cap A) + \mathrm{P}(B \cap A^c) = \mathrm{P}(A)\mathrm{P}(B|A) + \mathrm{P}(A^c)\mathrm{P}(B|A^c) \\ &= 0.07 \times 0.85 + 0.93 \times 0.25 = 0.292\end{aligned}$$

흡연자를 선택하였을 때, 이 사람이 폐질환을 앓고 있을 확률은 다음과 같다.

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A)\mathrm{P}(B|A)}{\mathrm{P}(B)} = \frac{0.07 \times 0.85}{0.292} = \frac{0.0595}{0.292} = 0.2038$$

Chapter 03 연습문제

3.1

전체 확률의 합은 1이므로 $a+\frac{1}{8}+\frac{1}{4}+b=1$이다.

따라서 $a+b=1-\frac{3}{8}=\frac{5}{8}$이다.

3.3

(a) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ 이므로, $\int_0^2 kx^2\,dx = k\frac{1}{3}[x^3]_0^2 = 1$ 이다. 따라서 $k=\frac{3}{8}$ 이다.

(b) $$\begin{aligned} F(x) &= \int_0^x \frac{3}{8}t^2\,dt \\ &= \frac{3}{8}\left[\frac{1}{3}t^3\right]_0^x \\ &= \frac{1}{3}+4\ln 2 \end{aligned}$$

따라서 $F(x)=\begin{cases} 0 & , x<0 \\ \frac{1}{8}x^3 & , 0 \le x \le 2 \\ 1 & , 2<x \end{cases}$ 이다.

(c) $\mathrm{P}(0<X<1) = F(1)-F(0) = \frac{1}{8}-0 = \frac{1}{8}$

(d) $\mathrm{E}(X) = \int_0^2 \frac{3}{8}x^3\,dx = \frac{3}{8}\left[\frac{1}{4}x^4\right]_0^2 = \frac{3}{2}$

(e) $\mathrm{E}(X^2) = \int_0^2 \frac{3}{8}x^4\,dx = \frac{3}{8}\left[\frac{1}{5}x^5\right]_0^2 = \frac{12}{5}$

$\mathrm{Var}(X) = \frac{12}{5} - \left(\frac{3}{2}\right)^2 = \frac{3}{20}$

3.5

매 짝수 번째일 때 주사위의 눈이 6이 나올 확률은 독립적으로 각각 더해 주어야 하므로, 구하고자 하는 확률 p는 다음과 같다.

$$\begin{aligned} p &= \frac{5}{6} \times \frac{1}{6} + \left(\frac{5}{6}\right)^3 \times \frac{1}{6} + \left(\frac{5}{6}\right)^5 \times \frac{1}{6} + \cdots \\ &= \sum_{i=1}^{\infty} \frac{1}{6}\left(\frac{5}{6}\right)^{2i-1} = \frac{\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)}{1-\left(\frac{5}{6}\right)^2} = \frac{5}{11} \end{aligned}$$

3.7

(a)

X	**1**	**3**	**합**
$h(x_i)$	$\frac{1}{2}$	$\frac{1}{2}$	1

(b)

Y	**4**	**10**	**합**
$g(y_j)$	$\frac{1}{2}$	$\frac{1}{2}$	1

(c) $\mathrm{E}(X) = \sum_i x_i h(x_i) = 1 \times \frac{1}{2} + 3 \times \frac{1}{2} = 2$

$$\begin{aligned} \mathrm{Var}(X) &= \sum_i x_i^2 h(x_i) - [\mathrm{E}(X)]^2 \\ &= 1^2 \times \frac{1}{2} + 3^2 \times \frac{1}{2} - 2^2 = 1 \end{aligned}$$

(d) $\mathrm{E}(Y) = \sum_j y_j g(y_j) = 4 \times \frac{1}{2} + 10 \times \frac{1}{2} = 7$

$$\begin{aligned} \mathrm{Var}(Y) &= \sum_j y_j^2 g(y_j) - [E(Y)]^2 \\ &= 4^2 \times \frac{1}{2} + 10^2 \times \frac{1}{2} - 7^2 = 9 \end{aligned}$$

(e) $\mathrm{E}(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j)$

$$= 1 \times 4 \times \frac{1}{4} + 1 \times 10 \times \frac{1}{4} + 3 \times 4 \times \frac{1}{4} + 3 \times 10 \times \frac{1}{4} = 14$$

(f) $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = 14 - 2 \times 7 = 0$

(g) $\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{0}{\sqrt{1}\sqrt{9}} = 0$

(h) $f(x_i, y_j) = h(x_i) g(y_j)$이므로 X와 Y는 독립이다.

3.9

$P(X=100)=\frac{1}{100}=0.01$, $P(X=50)=\frac{4}{100}=0.04$,

$P(X=10)=\frac{5}{100}=0.05$, $P(X=1)=\frac{10}{100}=0.1$,

$$P(X=0)=1-P(X=100)-P(X=50)-P(X=10)-P(X=1)$$
$$=1-0.2=0.8$$

여기서, $P(X=0)$은 복권에 당첨되지 않을 확률이다.

X	**0**	**1**	**10**	**50**	**100**
$P(X=x)$	0.8	0.1	0.05	0.04	0.01

3.11

$$E(X)=\frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)=\frac{91}{6}$$

3.13

$Y=X^2$ 의 관계로부터 확률변수 Y의 값은 1, 4이며, 이에 대응하는 $P(Y=y)$는 각각 $\frac{1}{2}$, $\frac{1}{2}$ 이다. 즉, 다음이 성립한다.

$$P(Y=1)=P(X=-1)+P(X=1)=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$$
$$P(Y=4)=P(X=-2)+P(X=2)=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$$

따라서 Y의 확률질량함수는 다음과 같다.

y_j	**1**	**4**	**합**
$P(Y=y)$	$\frac{1}{2}$	$\frac{1}{2}$	1

3.15

(a) $f(x) = \int_0^5 \frac{1}{210}(2x + y)\,dy$

$= \frac{1}{210}\left[2xy + \frac{1}{2y^2}\right]_0^5$

$= \frac{1}{84}(4x + 5)$

따라서 $f(x) = \begin{cases} \frac{1}{84}(4x+5) & , \ 2 < x < 6 \\ 0 & , \text{그 외} \end{cases}$ 이다.

(b) $g(y) = \int_2^6 \frac{1}{210}(2x + y)\,dx$

$= \frac{1}{210}\left[x^2 + xy\right]_2^6$

$= \frac{1}{105}(2y + 16)$

따라서 $g(y) = \begin{cases} \frac{1}{105}(2y+16) & , \ 0 < y < 5 \\ 0 & , \text{그 외} \end{cases}$ 이다.

3.17

(a) 확률변수 X가 계획대로 완공할 수 있는 공사 건수이므로 6건의 공사 중 x건이 완공되었다면 완성이 되지 않는 공사 건수는 $(6-x)$건이다. 따라서 구하고자 하는 확률질량함수는 $f(x) = {}_6\mathrm{C}_x(0.6)^x(0.4)^{5-x}$, $x = 0, 1, 2, 3, 4, 5, 6$이다.

(b) 평균은 3.6건, 분산은 1.44건이다.

(c) 변동계수는 $CV = \frac{\sigma}{\mu} \times 100 = \frac{1.2}{3.6} \times 100 ≒ 33.33$이다.

3.19

(a) 손해를 입을 확률은 $\mathrm{P}(X < 0) = \int_{-10}^{0} 0.02\,dx = 0.2$이다.

(b) $\mathrm{P}(X \geq 40) = \int_{40}^{70}\left(-\frac{0.01}{30}x + \frac{0.07}{3}\right)dx = 0.15$

3.21

(a) $$\mathrm{E}(X)=\int_0^2 \frac{x^2}{8}dx+\int_2^8 \frac{2}{x}dx$$

$$=\left[\frac{x^3}{24}\right]_0^2+[2\ln x]_2^8$$

$$=\frac{1}{3}+4\ln 2$$

(b) $$\mathrm{P}(X\le 3)=\int_0^2 \frac{x}{8}dx+\int_2^3 \frac{2}{x^2}dx$$

$$=\frac{1}{4}+\frac{1}{3}=\frac{7}{12}$$

3.23

(a) $\mathrm{P}(X\ge 2,\ Y\ge 20)=1-\mathrm{P}(X=1,\ Y=10)=1-0.05=0.95$

(b) $\mathrm{P}(X=1,\ Y\ge 20)=0.25+0.10=0.35$

(c) $0.05=\mathrm{P}(X=1,\ Y=10)\ne \mathrm{P}(X=1)\cdot \mathrm{P}(Y=10)=0.15\times 0.2=0.03$

그러므로 X와 Y는 독립이 아니다.

3.25

X \ Y	0	1	2	3	합계
0	0.05	0.05	0.1	0	0.2
1	0.05	0.1	0.25	0.1	0.5
2	0	0.15	0.1	0.05	0.3
합계	0.1	0.3	0.45	0.15	1

E(X)	1.1	Var(X)	0.49
E(Y)	1.65	Var(Y)	0.7275

XY	0	1	2	3	4	6	합계
P	0.25	0.1	0.4	0.1	0.1	0.05	1

E(XY) 1.9

Cov(X,Y) 0.085

Corr(X,Y) 0.142365

Chapter 04 연습문제

4.1

(a) $f(x) = {}_5\mathrm{C}_x\left(\frac{3}{4}\right)^x\left(\frac{1}{4}\right)^{5-x}$, $x = 0, 1, 2, 3, 4, 5$

(b) $\mathrm{P}(X=3) = {}_5\mathrm{C}_3\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2 = 0.2637$

(c) $\mathrm{P}(X \geq 2) = 1 - \mathrm{P}(X < 2)$

$$= 1 - \mathrm{P}(X=0) - \mathrm{P}(X=1)$$

$$= 1 - {}_5\mathrm{C}_0\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^5 - {}_5\mathrm{C}_1\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^4$$

$$= 1 - 0.0156 = 0.9984$$

(d) $\mathrm{E}(X) = 5 \times \frac{3}{4} = 3.75$

$\mathrm{Var}(X) = 5 \times \frac{3}{4} \times \frac{1}{4} = 0.9375$

4.3

(a) $\mathrm{E}(X) = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}$$

$$= np \sum_{x=1}^{n} {}_{n-1}\mathrm{C}_{x-1} p^{x-1} (1-p)^{n-x}$$

$$= np \sum_{x=0}^{n-1} {}_{n-1}\mathrm{C}_x p^x (1-p)^{n-1-x}$$

$$= np \{p + (1-p)\}^{n-1}$$

$$= np$$

(b) $\mathrm{E}(X^2) = \mathrm{E}[X(X-1)] + \mathrm{E}(X)$ 로부터 다음이 성립한다.

$$\begin{aligned}\mathrm{E}[X(X-1)] &= \sum_{x=0}^{n} x(x-1)\,{}_n\mathrm{C}_x\, \mathrm{p}^{\mathrm{x}}(1-\mathrm{p})^{\mathrm{n-x}} \\ &= n(n-1)p^2 \sum_{x=2}^{n} {}_{n-2}\mathrm{C}_{\mathrm{x}-2}\, \mathrm{p}^{\mathrm{x}-2}(1-\mathrm{p})^{\mathrm{n-x}} \\ &= n(n-1)p^2 \sum_{k=0}^{n-2} {}_{n-2}\mathrm{C}_{\mathrm{k}}\, \mathrm{p}^{\mathrm{k}}(1-\mathrm{p})^{\mathrm{n-2-k}} \\ &= n(n-1)p^2\{p+(1-p)\}^{n-2} \\ &= n(n-1)p^2\end{aligned}$$

또한 $E(X) = np$ 이므로 다음이 성립한다.

$$\begin{aligned}\mathrm{E}(X^2) &= \mathrm{E}[X(X-1)] + \mathrm{E}(X) \\ &= n(n-1)p^2 + np\end{aligned}$$

따라서 다음이 성립한다.

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= np(1-p)\end{aligned}$$

(c) $M_X(t) = \sum_{x=0}^{n} e^{tx}\,{}_n\mathrm{C}_x\, p^x(1-p)^{n-x}$

$$= \sum_{x=0}^{n} {}_n\mathrm{C}_x\,(pe^t)^x(1-p)^{n-x}$$

$$= (pe^t + 1 - p)^n$$

4.5

주사위를 던져서 4 또는 6의 눈이 나올 확률은 $\frac{1}{3}$ 이다. 6번 던져서 2번 나와야 하므로, 구하고자 하는 확률은 다음과 같다.

$$p = {}_6\mathrm{C}_2\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^4 = \frac{80}{243}$$

4.7

X를 안타를 친 횟수라고 하면, 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned}\mathrm{P}(X \geq 2) &= 1 - \mathrm{P}(X=0) - \mathrm{P}(X=1) \\ &= 1 - {}_4\mathrm{C}_0 (0.3)^0 (0.7)^4 - {}_4\mathrm{C}_1 (0.3)^1 (0,7)^3 \\ &= 1 - (0.7)^4 - 4(0.3)(0.7)^3 = 0.35\end{aligned}$$

4.9

안타를 친 횟수를 X라 하면, $X \sim \mathrm{B}(20, 0.25)$이다.
따라서 $\mathrm{E}(X) = np = 20 \times 0.25 = 5$이다.

4.11

주어진 시간 또는 정해진 영역 내에서 일어나는 사건의 횟수는 푸아송분포를 따른다.

4.13

(a) $M_X(t) = \sum_{x=0}^{\infty} e^{tx} \frac{m^x e^{-m}}{x!}$

$$= e^{-m} \sum_{x=0}^{\infty} \frac{(me^t)^x}{x!}$$

$$= e^{-m} e^{me^t}$$

$$= e^{m(e^t - 1)}$$

(b) $M_X{}'(t) = me^t e^{m(e^t-1)}$

$$\mathrm{E}(X) = M_X{}'(0)$$

$$= m$$

(c) $M_X{}''(t) = (me^t)^2 e^{m(e^t-1)} + me^t e^{m(e^t-1)}$

$$\begin{aligned}\mathrm{E}(X^2) &= M_X{}''(0) \\ &= m^2 + m \\ \mathrm{Var}(X) &= M_X{}''(0) - [M_X{}'(0)]^2 \\ &= m^2 + m - m^2 \\ &= m\end{aligned}$$

4.15

사고를 낸 건수를 X라 하면, 구하고자 하는 확률은 다음과 같다.

$$\mathrm{P}(X=0)=\frac{e^{-0.3}0.3^0}{0!}=e^{-0.3}=0.741$$

4.17

$X\sim \mathrm{P}(2)$, $Y\sim \mathrm{P}(3)$ 이면 $Z=X+Y$는 $\mathrm{P}(5)$인 푸아송분포를 따른다.

4.19

(a) $M_X(t)=\sum_{x=1}^{\infty}e^{tx}q^{x-1}p$

$$=\frac{p}{q}\sum_{x=1}^{\infty}(qe^t)^x$$

$$=\frac{p}{q}\frac{qe^t}{1-qe^t}$$

$$=\frac{pe^t}{1-qe^t}$$

(b) $M_X{}'(t)=\dfrac{pe^t(1-qe^t)-pe^t(-qe^t)}{(1-qe^t)^2}$

$$=\frac{pe^t}{(1-qe^t)^2}$$

$$\mathrm{E}(X)=M_X{}'(0)$$

$$=\frac{p}{(1-q)^2}$$

$$=\frac{1}{p}$$

(c) $M_X{}''(t)=\dfrac{pe^t(1-qe^t)^2-pe^t\cdot 2(1-qe^t)(-qe^t)}{(1-qe^t)^4}$

$$=\frac{pe^t(1-qe^t)+2pe^tqe^t}{(1-qe^t)^3}$$

$$=\frac{pe^t+pe^tqe^t}{(1-qe^t)^3}$$

$$\begin{aligned}E(X^2) &= M_X''(0)\\ &= \frac{p+pq}{(1-q)^3}\\ &= \frac{1+q}{p^2}\end{aligned}$$

$$\begin{aligned}Var(X) &= M_X''(0)-[M_X'(0)]^2\\ &= \frac{1+q}{p^2}-\frac{1}{p^2}\\ &= \frac{q}{p^2}\end{aligned}$$

4.21

$$\begin{aligned}b(n,p) = {}_nC_x\, p^x(1-p)^{n-x} &= {}_nC_x\left(\frac{m}{n}\right)^x\left(1-\frac{m}{n}\right)^{n-x}\\ &= \frac{n(n-1)\cdots(n-x+1)}{x!}\left(\frac{m}{n}\right)^x\left(1-\frac{m}{n}\right)^{n-x}\\ &= \frac{m^x}{x!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\left(1-\frac{m}{n}\right)^n\left(1-\frac{m}{n}\right)^{-x}\end{aligned}$$

$e = \lim_{x\to 0}(1+x)^{\frac{1}{x}}$ 을 이용하면,

$\lim_{n\to\infty}\left(1-\frac{m}{n}\right)^n = \left\{\lim_{n\to\infty}\left(1+\frac{(-m)}{n}\right)^{\frac{n}{-m}}\right\}^{-m} = e^{-m}$ 이고,

$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right) = 1,\ \lim_{n\to\infty}\left(1-\frac{m}{n}\right)^{-x} = 1$ 이므로 다음이 성립한다.

$$\lim_{n\to\infty} b(n,p) = \frac{m^x e^{-m}}{x!}$$

4.23

(a) $f(x) = \begin{cases}\frac{1}{3}, & 0 \le x \le 3\\ 0, & x<0 \text{ 또는 } x>3\end{cases}$

(b) $F(x) = \begin{cases} 0 & , \ x < 0 \\ \frac{x}{3} & , \ 0 \le x \le 3 \\ 1 & , \ 3 < x \end{cases}$

(c) $\mathrm{E}(X) = \frac{0+3}{2} = \frac{3}{2}$, $\mathrm{Var}(X) = \frac{(3-0)^2}{12} = \frac{9}{12} = \frac{3}{4}$

4.25

(a) $\mathrm{E}(X) = \int_a^b x \frac{1}{b-a} dx$

$$= \frac{1}{b-a} \left[\frac{1}{2} x^2 \right]_a^b$$

$$= \left(\frac{1}{b-a} \right) \left(\frac{b^2 - a^2}{2} \right)$$

$$= \frac{a+b}{2}$$

(b) $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$ 에서

$\mathrm{E}(X^2) = \int_a^b x^2 \frac{1}{b-a} dx$

$$= \frac{1}{b-a} \left[\frac{1}{3} x^3 \right]_a^b$$

$$= \left(\frac{1}{b-a} \right) \left(\frac{b^3 - a^3}{3} \right)$$

$$= \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)}$$

$$= \frac{b^2 + ab + a^2}{3}$$

이고, $\mathrm{E}(X) = \frac{a+b}{2}$ 이므로 다음이 성립한다.

$$\mathrm{Var}(X) = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2} \right)^2$$

$$= \frac{(b-a)^2}{12}$$

(c) $M_X(t) = \int_a^b e^{tx} \frac{1}{b-a} dx$

$$= \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b$$

$$= \frac{e^{bt}-e^{at}}{(b-a)t}, \quad t \neq 0$$

4.27

$$\mathrm{P}(70 \le X \le 80) = \mathrm{P}\left(\frac{70-75}{7} \le \frac{X-75}{7} \le \frac{80-75}{7}\right)$$

$$= \mathrm{P}(-0.71 \le Z \le 0.71)$$

$$= 2\mathrm{P}(0 \le Z \le 10.71)$$

$$= 2 \times 0.2611 = 0.5222$$

따라서 구하고자 하는 학생 수는 $500 \times 0.5222 = 261$ (명)이다.

4.29

(a) X가 정규분포 $\mathrm{N}(10, 2^2)$이므로, 평균 $\mu = 10$, 표준편차 $\sigma = 2$이다. 따라서 다음이 성립한다.

$$\mathrm{P}(10 \le X \le 16) = \mathrm{P}(\mu \le X \le \mu + 3\sigma)$$

$$= 0.498$$

(b) X가 정규분포 $\mathrm{N}(80, 5^2)$이므로, 평균 $\mu = 80$, 표준편차 $\sigma = 5$이다. 따라서 다음이 성립한다.

$$\mathrm{P}(70 \le X \le 90) = \mathrm{P}(80-10 \le X \le 80+10)$$

$$= \mathrm{P}(80-2\times 5 \le X \le 80+2\times 5)$$

$$= \mathrm{P}(\mu - 2\sigma \le X \le \mu + 2\sigma)$$

$$= 2\mathrm{P}(\mu \le X \le \mu + 2\sigma)$$

$$= 2(0.4772) = 0.9544$$

4.31

Z는 표준정규분포를 따르며, 이때 Z의 평균과 분산은 0과 1이다.

4.33

$\mu = 400$, $\sigma = 50$이므로 다음이 성립한다.

$$\begin{aligned} \mathrm{P}(450 \le X \le 500) &= \mathrm{P}\left(\frac{450-400}{50} \le Z \le \frac{500-400}{50}\right) \\ &= \mathrm{P}(1 \le Z \le 2) \\ &= 0.9772 - 0.8413 = 0.1359 \end{aligned}$$

따라서 직업에 적합하다고 생각하는 점수는 상위 13.59%에 해당한다.

4.35

X가 정규분포 $\mathrm{N}(3, 2^2)$을 따르므로, $Z = \frac{k-3}{2}$이다.

또한 Y가 표준정규분포 $\mathrm{N}(0, 1^2)$을 따르므로, $Z = \frac{k-0}{1}$이다.

따라서 $\mathrm{P}(Z \ge \frac{k-3}{2}) = \mathrm{P}(Z \ge k)$이다.

즉, $\frac{k-3}{2} = k$, $k-3 = 2k$이므로, $k = -3$이다.

4.37

$\mathrm{E}(W) = \mathrm{E}(X) + \mathrm{E}(Y) = \mu_1 + \mu_2$

$\mathrm{Var}(W) = \mathrm{Var}(X) + \mathrm{Var}(Y) = \sigma_1^2 + \sigma_2^2$

$\Rightarrow$ $W \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

4.39

확률밀도함수 $f(x)$가 지수분포의 함수이므로, 구하고자 하는 평균수명은

$\mathrm{E}(X) = \frac{1}{\lambda} = \frac{1}{1/200} = 200$ (시간)이다.

4.41

$$\mathrm{P}(X<1)=\int_0^1 f(x)dx$$

$$=\left[-e^{-0.15x}\right]_0^1$$

$$=1-e^{-0.15}$$

$$=1-0.861=0.139$$

4.43

(a) $M_X(t)=\int_0^\infty e^{tx}\frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-\frac{x}{\beta}}dx$ 에서 $y=\frac{1-\beta t}{\beta}x$, $t<\frac{1}{\beta}$ 라 하면,

$x=\frac{\beta}{1-\beta t}y$ 이고, $dx=\frac{\beta}{1-\beta t}dy$ 이므로 다음이 성립한다.

$$M_X(t)=\frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty\left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1}e^{-y}\frac{\beta}{1-\beta t}dy$$

$$=\frac{1}{\beta^\alpha\Gamma(\alpha)}\left(\frac{\beta}{1-\beta t}\right)^\alpha\int_0^\infty y^{\alpha-1}e^{-y}dy$$

$$=\frac{1}{\beta^\alpha\Gamma(\alpha)}\left(\frac{\beta}{1-\beta t}\right)^\alpha\Gamma(\alpha)$$

$$=(1-\beta t)^{-\alpha}$$

(b) $M_X{}'(t)=-\alpha(1-\beta t)^{-\alpha-1}(-\beta)$

$\mathrm{E}(X)=M_X{}'(0)$

$=\alpha\beta$

(c) $M_X{}''(t)=\alpha\beta(-\beta)(-\alpha-1)(1-\alpha t)^{-\alpha-2}$

$\mathrm{E}(X^2)=M_X{}''(0)$

$=\alpha\beta^2(\alpha+1)$

$\mathrm{Var}(X)=M_X{}''(0)-[M_X{}'(0)]^2$

$=\alpha\beta^2(\alpha+1)-\alpha^2\beta^2$

$=\alpha\beta^2$

4.45

$X \sim \mathrm{B}(720,\ \frac{1}{6})$ 이므로, 다음 결과를 얻는다.

$$\mu = np = 720 \times \frac{1}{6} = 120$$

$$\sigma^2 = npq = 720 \times \frac{1}{6} \times \frac{5}{6} = 100$$

그러므로 X는 근사적으로 $\mathrm{N}(120, 10^2)$을 따른다.

4.47

$P(2 \le X \le 4) = P(X \le 4) - P(X \le 2)$

$= BINOMDIST(4, 10, 0.05, 1) - BINOMDIST(2, 10, 0.05, 1)$

$\fallingdotseq 0.99994 - 0.988496 = 0.01144$

E3 =BINOMDIST(4,10,0.05,1)-BINOMDIST(2,10,0.05,1)

	A	B	C	D	E	F	G	H
1	[4-47] B(10, 0.05)인 이항분포에서 P(2≤X≤4)의 확률을 엑셀을 이용하여 구하라.							
2								
3		P(2≤X≤4)= P(X≤4) - P(X≤2)=			0.01144			
4								

4.49

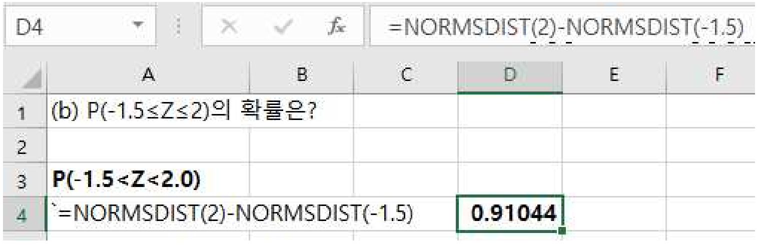
(a)

B9 =1-B7

	A	B	C
3			
4	X	180	
5	평균	172	
6	표준편차	15	
7	누적확률 P(X≤180)	0.70309857	
8			
9	P(X>180)=	0.29690143	

따라서 $\mathrm{P}(X > 180) \fallingdotseq 0.2969$이다.

(b)



D4 =NORMSDIST(2)-NORMSDIST(-1.5)

	A	B	C	D	E	F
1	(b) P(-1.5≤Z≤2)의 확률은?					
2						
3	**P(-1.5<Z<2.0)**					
4	`=NORMSDIST(2)-NORMSDIST(-1.5)			**0.91044**		

따라서 $\mathrm{P}(-1.5 \leq Z \leq 2) \fallingdotseq 0.9104$이다.

Chapter 05 연습문제

5.1

모집단은 정규분포 $N(30, 6^2)$을 따르고 표본 크기는 $n=9$이므로 $E(\overline{X}) = \mu = 30$이고,

$Var(\overline{X}) = \dfrac{\sigma^2}{n} = \dfrac{36}{9} = 4$이다. 표본평균 $\overline{X}$는 정규분포 $N(30, 2^2)$을 따른다.

따라서 구하고자 하는 확률을 구하면 다음과 같다.

$$
\begin{aligned}
P[26 \le \overline{X} \le 34] &= P\left[\frac{26-30}{2} \le Z \le \frac{34-30}{2}\right] \\
&= P[-2 \le Z \le 2] \\
&= P[Z \le 2] - P[Z \le -2] \\
&= 0.9772 - 0.0228 = 0.9544
\end{aligned}
$$

5.3

(a)

가능한 표본	$\overline{X}$
$\{1, 2, 3\}$	2
$\{1, 2, 4\}$	2.33
$\{2, 3, 4\}$	3

(b)

$\overline{X}$	$P(\overline{X} = x)$
2	1/3
2.33	1/3
3	1/3

(c) $E(\overline{X}) = \sum \overline{X} \, P(\overline{X})$

$$= 2 \times \frac{1}{3} + 2.33 \times \frac{1}{3} + 3 \times \frac{1}{3} = 2.443$$

(d) $\mathrm{Var}(\overline{X}) = \sum \overline{X}^2 \mathrm{P}(\overline{X}) - \overline{X}^2$

$$= 2^2 \times \frac{1}{3} + 2.33^2 \times \frac{1}{3} + 3^2 \times \frac{1}{3} - (2.443)^2$$

$$= 0.1747$$

(e) $\frac{N-n}{N-1} = \frac{4-3}{4-1} = \frac{2}{3}$

5.5

표본평균 $\overline{X}$의 표준오차 $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$는 표본크기 n이 커지면 작아진다.

5.7

랜덤하게 추출한 10개월 간의 평균을 $\overline{X}$라 하자. 그러면 $\overline{X}$는 $\mathrm{N}\left(5650, \frac{700^2}{10}\right)$을 따른다고 할 수 있다. 따라서 $\mathrm{P}\left(\left|\frac{\overline{X}-5650}{700/\sqrt{10}}\right| \leq z_{\alpha/2}\right) = 1-\alpha$이므로, 구하고자 하는 확률은 다음과 같다.

$$\mathrm{P}(|\overline{X}-\mu| \leq 210) = \mathrm{P}\left(\left|\frac{\overline{X}-\mu}{700/\sqrt{10}}\right| \leq 210 \cdot \frac{\sqrt{10}}{700}\right)$$

$$\fallingdotseq \mathrm{P}(|Z| \leq 0.95)$$

$$= 1 - 2\mathrm{P}(Z > 0.95) \fallingdotseq 0.6578$$

5.9

신장의 표본평균을 $\overline{X}$라고 하면, 구하고자 하는 확률은 다음과 같다.

$$\mathrm{P}(168 \leq \overline{X} \leq 172) = \mathrm{P}\left(\frac{168-170}{6/\sqrt{36}} \leq Z \leq \frac{172-170}{6/\sqrt{36}}\right)$$

$$= \mathrm{P}(-2 \leq Z \leq 2) = 0.9544$$

5.11

(a) $\overline{X_1} + \overline{X_2} \sim \mathrm{N}\left(\mu_1 + \mu_2 \,,\, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)$

(b) $\overline{X_1} - \overline{X_2} \sim \mathrm{N}\left(\mu_1 - \mu_2 \,,\, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)$

5.13

(a) $M_X(t) = \displaystyle\int_0^\infty e^{tx} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx$

$$= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty x^{\frac{n}{2}-1} e^{-\frac{x(1-2t)}{2}} dx$$

위 식에서 $y = \dfrac{x(1-2t)}{2}$ 로 두면, $dx = \dfrac{2}{1-2t} dy \quad \left(t < \dfrac{1}{2}\right)$ 이다.
따라서 다음 결과를 얻는다.

$$M_X(t) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty \left(\frac{2y}{1-2t}\right)^{\frac{n}{2}-1} e^{-y} \frac{2}{1-2t} dy$$

$$= \frac{1}{\Gamma\left(\frac{n}{2}\right)(1-2t)^{\frac{n}{2}}} \int_0^\infty y^{\frac{n}{2}-1} e^{-y} dy = (1-2t)^{-\frac{n}{2}}$$

(b) $M_X{}'(t) = -\dfrac{n}{2}(1-2t)^{-\frac{n}{2}-1}(-2)$

$$= n(1-2t)^{-\frac{n}{2}-1}$$

$$\mathrm{E}(X) = M_X{}'(0)$$

$$= n$$

(c) $M_X{}''(t) = \left(-\dfrac{n}{2}\right)\left(-\dfrac{n}{2} - 1\right)(1-2t)^{-\frac{n}{2}-2}(-2)^2$

$$\begin{aligned}\mathrm{E}(X^2) &= M_X''(0) \\ &= 4\left(\frac{n}{2}\right)\left(\frac{n+2}{2}\right) \\ &= n(n+2) \\ \mathrm{Var}(X) &= M_X''(0) - [M_X'(0)]^2 \\ &= n(n+2) - n^2 \\ &= 2n\end{aligned}$$

5.15

(a) $t-$ 분포의 확률밀도함수는 다음과 같다.

$$\begin{aligned}f(x) &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{(n+1)}{2}} \\ &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n}\,\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{(n+1)}{2}} \\ &= \frac{1}{\sqrt{n}\,B\left(\frac{n}{2},\frac{1}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{(n+1)}{2}}\end{aligned}$$

따라서 $\mathrm{E}(X)$는 다음과 같이 구한다.

$$\begin{aligned}\mathrm{E}(X) &= \int_{-\infty}^{\infty} x\frac{1}{\sqrt{n}\,B\left(\frac{n}{2},\frac{1}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{(n+1)}{2}}dx \\ &= \frac{1}{\sqrt{n}\,B\left(\frac{n}{2},\frac{1}{2}\right)}\int_{-\infty}^{\infty} x\left(1+\frac{x^2}{n}\right)^{-\frac{(n+1)}{2}}dx = 0\end{aligned}$$

(b) $$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{n}\, B\left(\frac{n}{2}, \frac{1}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}} dx$$

$$= \frac{1}{\sqrt{n}\, B\left(\frac{n}{2}, \frac{1}{2}\right)} \int_{-\infty}^{\infty} x^2 \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}} dx$$

$$= \frac{2}{\sqrt{n}\, B\left(\frac{n}{2}, \frac{1}{2}\right)} \int_{0}^{\infty} x \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}} x\, dx$$

이때 $\left(1 + \frac{x^2}{n}\right)^{-1} = t$ 로 두면 $x^2 = n(1-t)t^{-1}$, $xdx = \left(-\frac{n}{2t^2}\right)dt$ 이다.

따라서 다음이 성립한다.

$$\mathrm{E}(X^2) = \frac{2}{\sqrt{n}\, B\left(\frac{n}{2}, \frac{1}{2}\right)} \int_{1}^{0} \sqrt{n}\, (1-t)^{\frac{1}{2}} t^{-\frac{1}{2}} \cdot t^{\frac{n+1}{2}} \left(-\frac{n}{2t^2}\right) dt$$

$$= \frac{n^{\frac{3}{2}}}{\sqrt{n}\, B\left(\frac{n}{2}, \frac{1}{2}\right)} \int_{0}^{1} t^{\frac{n}{2}-1-1} (1-t)^{\frac{1}{2}+1-1} dt$$

$$= \frac{n}{B\left(\frac{n}{2}, \frac{1}{2}\right)} B\left(\frac{n}{2}-1, \frac{3}{2}\right), \quad \left(\frac{n}{2}-1 \text{ 에서 } n > 2\right)$$

$$= n \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{\Gamma\left(\frac{n}{2}-1\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}$$

$$= n \frac{\Gamma\left(\frac{n}{2}-1\right)\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\left(\frac{n}{2}-1\right)\Gamma\left(\frac{n}{2}-1\right)} = \frac{n}{n-2}$$

그러므로 $\mathrm{Var}(X)$ 는 다음과 같이 구할 수 있다.

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$$

$$= \frac{n}{n-2} - 0^2 = \frac{n}{n-2}$$

5.17
자유도가 $n_1 - 1$, $n_2 - 1$ 인 F분포이다.

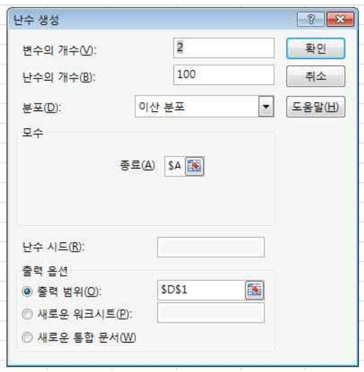
5.19
표본비율 $\hat{p} = \dfrac{X}{n} \approx \mathrm{N}\left(0.9, \dfrac{0.9 \times 0.1}{100}\right)$을 따른다. 그러므로 $Z = \dfrac{\hat{p} - 0.9}{0.03} \sim \mathrm{N}(0, 1)$을 따른다.
따라서 다음이 성립한다.

$$
\begin{aligned}
\mathrm{P}(|\hat{p} - 0.9| \le 0.05) &= \mathrm{P}\left(\frac{|\hat{p} - 0.9|}{0.03} \le \frac{5}{3}\right) \\
&= \mathrm{P}\left(|Z| \le \frac{5}{3}\right) \\
&\fallingdotseq 2 \times 0.9525 - 1 = 0.905
\end{aligned}
$$

5.21
먼저 다음과 같이 자료를 입력한다.

	A	B
1	X	P(X)
2	1	0.25
3	2	0.25
4	3	0.25
5	4	0.25

[데이터분석]⇒[난수생성]을 통해 난수를 생성한다.



이때 분포는 이산분포를 선택하고 종료(A)에 위에서 입력한 내용을 선택하여 설정해준다.
그러면 원하는 결과를 얻을 수 있다.

Chapter 06 연습문제

6.1

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = 1$$

6.3

$$\bar{x} = \frac{1}{15}(2.31 + 1.97 + \cdots + 2.25) = 2.159$$

$$2.159 - 1.96\frac{0.25}{\sqrt{15}} < \mu < 2.159 + 1.96\frac{0.25}{\sqrt{15}}$$

$$\Rightarrow\ 2.032 < \mu < 2.286$$

6.5

강판 무게에 대한 신뢰구간

$$:\ 1.0795 - 1.96\frac{0.01}{\sqrt{10}} < \mu < 1.0795 + 1.96\frac{0.01}{\sqrt{10}}$$

$$\Rightarrow\ 1.073302 < \mu < 1.085698$$

200매 강판 전부 무게에 대한 신뢰구간

$$:\ 200 \times 1.073302 < \mu < 200 \times 1.085698$$

$$\Rightarrow\ 214.6604 < \mu < 217.1396$$

6.7

$$6,200 - 1.96\frac{140}{\sqrt{49}} < \mu < 6,200 + 1.96\frac{140}{\sqrt{49}}$$

$$\Rightarrow\ 6,160.8 < p < 6,239.2$$

6.9

$$115,000 - 1.96\frac{56,000}{\sqrt{4,000}} < \mu < 115,500 + 1.96\frac{56,000}{\sqrt{4,000}}$$

$$\Rightarrow\ 113,765 < \mu < 117,235$$

6.11

$\bar{x} = \frac{1}{7}(61+57+53+58+51+56+55) = 55.857,$

$S^2 = \frac{1}{7-1}\{(61-55.867)^2+(57-55.857)^2+\cdots+(55-55.857)^2\} \fallingdotseq 10.810,$

$t_{0.05}(6) = 2.447$ 이므로 다음 결과를 얻는다.

$$55.857 - 2.447\frac{\sqrt{10.810}}{\sqrt{7}} < \mu < 55.857 + 2.447\frac{\sqrt{10.810}}{\sqrt{7}}$$

$\Rightarrow\ 53.41 < \mu < 58.897$

6.13

$$0.55 - 1.645\sqrt{\frac{(0.55)(0.45)}{100}} < p < 0.55 + 1.645\sqrt{\frac{(0.55)(0.45)}{100}}$$

$\Rightarrow\ 0.7216 < p < 0.8784$

6.15

$\hat{p} = \frac{2}{100} = 0.02$ 이므로 모불량률의 95% 신뢰구간은 다음과 같다.

$$0.02 - 1.96\sqrt{\frac{0.02\times0.98}{100}} < p < 0.02 + 1.96\sqrt{\frac{0.02\times0.98}{100}}$$

$\Rightarrow\ 0 < p < 0.04744$

6.17

$\hat{p} = \frac{5}{100} = 0.05$

$$n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{d}\right)^2$$

$$= (0.05)(0.95)\left(\frac{1.96}{0.05}\right)^2$$

$$= 73$$

따라서 필요한 표본의 최소의 크기는 73이다.

6.19

$$(418-402)-1.96\sqrt{\frac{26^2}{40}+\frac{22^2}{50}}<\mu_1-\mu_2<(418-402)+1.96\sqrt{\frac{26^2}{40}+\frac{22^2}{50}}$$

$\Rightarrow\ 10.84<\mu_1-\mu_2<21.16$

6.21

자유도가 커짐에 따라 σ^2의 추정량의 신뢰구간의 폭은 점점 좁아진다.

6.23

$$\bar{x}=\frac{1}{9}(8.6+7.9+\cdots+7.5)=7.99$$

$$S^2=\frac{1}{9-1}\{(8.6-7.99)^2+(7.9-7.99)^2+\cdots+(7.5-7.99)^2\}$$

$$=0.91875$$

$\chi^2_{0.05}(8)=15.5,\ \chi^2_{0.95}(8)=2.73$ 이므로, σ^2의 95% 신뢰구간은 다음과 같다.

$$\frac{9\times 0.91875}{15.5}<\sigma^2<\frac{7\times 0.91875}{2.73}$$

$\Rightarrow\ 0.533<\sigma^2<3.029$

6.25

95% 신뢰구간으로 오차한계가 0.01 이하이므로 $1.96\sqrt{\dfrac{\hat{p}\times(1-\hat{p})}{n}}\le 0.01$이다.

$\hat{p}=0.6$ 이므로 다음 결과를 얻는다.

$$1.96\sqrt{\frac{0.6\times 0.4}{n}}\le 0.01\ \Rightarrow\ 0.01\sqrt{n}\ge 1.96\sqrt{0.6\times 0.4}$$

$\Rightarrow\ \sqrt{n}\ge 1.96\sqrt{0.24}\ \Rightarrow\ n\ge 9220$

따라서 표본 크기 n은 9220만 명 이상이어야 한다.

6.27

자료를 다음과 같이 입력하고 신뢰수준을 구한 후 신뢰구간을 구한다.
구하고자 하는 신뢰구간은 $1.1 \pm 2.395 = (-1.295, 3.495)$이다.

교육전	교육후	차(후-전)
85	83	-2
70	74	4
73	75	2
69	68	-1
83	88	5
78	80	2
77	80	3
74	70	-4
66	71	5
79	76	-3

교육후-교육전	
평균	1.1
표준 오차	1.058825345
중앙값	2
최빈값	2
표준편차	3.348299734
분산	11.21111111
첨도	-1.528552247
왜도	-0.330330519
범위	9
최솟값	-4
최댓값	5
합	11
관측수	10
신뢰 수준(95.0%)	2.395229338

Chapter 07 연습문제

7.1

귀무가설 $H_0 : \mu_1 = \mu_2$

대립가설 $H_1 : \mu_1 > \mu_2$

7.3

① 양쪽검정이므로 귀무가설 $H_0 : \mu = 121.4$에 대한 대립가설을 $H_1 : \mu \neq 121.4$로 설정한다.

② 모표준편차를 알고 있으므로 모평균에 대한 검정통계량의 확률분포는 다음과 같다.

$$Z = \frac{\overline{X} - 121.4}{6/\sqrt{100}} \sim N(0,1)$$

③ 유의수준 $\alpha = 0.05$에 대한 양쪽검정의 임곗값은 $z_{0.025} = 1.96$이므로,

기각역은 $R : |Z| \geq 1.96$이다.

④ 표본으로부터 얻은 표본평균이 $\overline{X} = 119.4$이므로 검정통계량의 관측값은 다음과 같다.

$$|z_0| = \frac{119.4 - 121.4}{6/\sqrt{100}} \fallingdotseq -3.33$$

이 관측값은 기각역 안에 들어가므로 귀무가설을 기각한다.

따라서 유의수준 0.05에서 귀무가설인 $\mu = 121.4$는 타당성이 없다.

7.5

크기 25인 표본으로부터 $\overline{x} = 41.57,\ S = 11.05$를 얻는다.

왼쪽 한쪽검정이므로 다음이 성립한다.

(a) $T > -t_{0.1}(24) = -1.318$

(b) $T \leq -t_{0.01}(24) = -2.492$

(c) $t_0 = \dfrac{\overline{x} - \mu_0}{S/\sqrt{n}} = \dfrac{41.57 - 45}{11.05/\sqrt{24}} \fallingdotseq -1.52$

(d) $\alpha = 0.1$에서는 귀무가설을 기각하고, $\alpha = 0.01$에서는 기각하지 못한다.

(e) $p -$값$= 0.0669$

7.7

① H_0 : $\mu = 60.2$

H_1 : $\mu > 60.2$

② 검정통계량 : $z_0 = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{62.3 - 60.2}{3 / \sqrt{9}} = 2.1$

③ $\alpha = 0.05$ 이므로 $z_{0.05} = 1.645$ 이다.

④ $z_0 = 2.1 > z_{0.05} = 1.645$ 이므로 H_0 를 기각한다.

따라서 공정 변화에 따라 제품 치수가 커졌다고 할 수 있다.

7.9

① H_0 : $\mu = 1{,}600$

H_1 : $\mu \neq 1{,}600$

② 검정통계량 : $|z_0| = \left| \dfrac{\overline{x} - \mu_0}{S / \sqrt{n}} \right| = \left| \dfrac{1{,}570 - 1{,}600}{120 / \sqrt{100}} \right| = 2.50$

③ $z_{0.025} = 1.96$ 이며 $|z_0| > z_{0.025}$ 이므로 유의수준 5% 에서는 귀무가설을 기각한다.
그리고 $z_{0.005} = 2.58$ 이며 $|z_0| < z_{0.005}$ 이므로 유의수준 1% 에서는 귀무가설을 채택한다.

④ 유의수준 5% 에서는 이 회사 제품의 평균 수명에 변화가 있다고 판단하며, 유의수준 1% 에서는 평균 수명에 변화가 없다고 판단된다.

7.11

① H_0 : $p = 0.35$

H_1 : $p \neq 0.35$

② $\hat{p} = \dfrac{8}{20} = 0.4$

검정통계량 : $|z_0| = \left| \dfrac{0.4 - 0.35}{\sqrt{\dfrac{0.35 \times 0.65}{20}}} \right| = 0.469$

③ $\alpha = 0.01$ 이므로, $z_{0.005} = 2.576$ 이다.

④ $|z_0| = 0.469 < z_{0.005} = 2.576$ 이므로 H_0 를 기각하지 못한다.

따라서 불량률이 변했다고 말할 수 없다.

7.13

① H_0 : $\mu_{\mathrm{S}} = \mu_{\mathrm{Q}}$

H_1 : $\mu_{\mathrm{S}} \neq \mu_{\mathrm{Q}}$

② $\overline{x}_{\mathrm{S}} = 24.33$, $\overline{x}_{\mathrm{Q}} = 15.83$

검정통계량 :

$$|z_0| = \left| \frac{\overline{x}_{\mathrm{S}} - \overline{x}_{\mathrm{Q}}}{\sqrt{\frac{\sigma_{\mathrm{S}}^2}{n_{\mathrm{S}}} + \frac{\sigma_{\mathrm{Q}}^2}{n_{\mathrm{Q}}}}} \right| = \frac{24.33 - 15.83}{\sqrt{\frac{3^2}{6} + \frac{5^2}{6}}} \fallingdotseq 3.57$$

③ $\alpha = 0.05$ 이므로 $z_{0.025} = 1.96$ 이다.

④ $|z_0| = 3.57 > z_{0.025} = 1.96$ 이므로 H_0 를 기각한다. 따라서 두 회사의 제품의 평균에 차이가 있다고 할 수 있다.

7.15

① H_0 : $\mu_A = \mu_B$

H_1 : $\mu_A \neq \mu_B$

② $\overline{x}_A = 57300$, $\overline{x}_Q = 56100$

검정통계량 : $$|z_0| = \left| \frac{\overline{x}_A - \overline{x}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \right| = \frac{57300 - 56100}{\sqrt{\frac{3550^2}{36} + \frac{3800^2}{36}}} \fallingdotseq 1.3846$$

③ $\alpha = 0.05$ 이므로 $z_{0.025} = 1.96$ 이다.

④ $|z_0| = 1.3846 < z_{0.025} = 1.96$ 이므로 H_0 를 기각하지 못한다.

따라서 두 회사 타이어의 평균 수명에는 차이가 있다고 말할 수 없다.

7.17

① H_0 : $p_1 = p_2$

H_1 : $p_1 < p_2$

② $\hat{p}_1 = \frac{35}{1,000} = 0.035$, $\hat{p}_2 = \frac{43}{1,000} = 0.043$

$$\hat{p} = \frac{35 + 43}{1,000 + 1,000} = 0.039$$

검정통계량 : $$z_0 = \frac{0.035 - 0.043}{\sqrt{0.039\,(1 - 0.039)\left(\frac{1}{1,000} + \frac{1}{1,000}\right)}} = -0.924$$

③ $\alpha = 0.05$ 이므로, $-z_{0.05} = -1.645$ 이다.

④ $z_0 = -0.924 > -z_{0.05} = -1.645$ 이므로 H_0 를 기각하지 못한다.

따라서 B 공정의 결점률이 A 공정보다 높다고 할 수 없다.

7.19

① $H_0 \;:\; p_1 = p_2$

$\quad H_1 \;:\; p_1 \neq p_2$

② $\hat{p}_1 = 0.55\,,\; \hat{p}_2 = 0.488$

$$\hat{p} = \frac{275+244}{1000} = 0.519$$

검정통계량 : $z_0 = \dfrac{0.55-0.488}{\sqrt{0.519\,(1-0.519)\left(\dfrac{1}{500}+\dfrac{1}{500}\right)}} \fallingdotseq 1.9620$

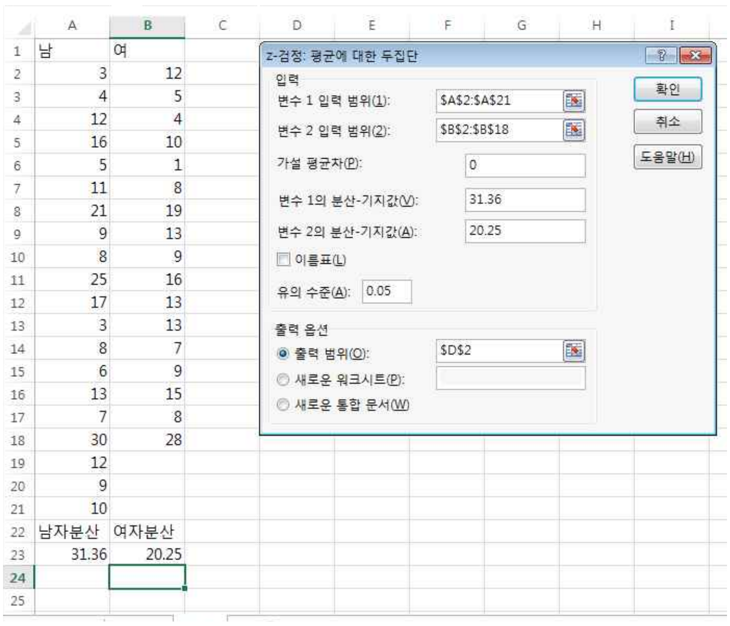
③ $\alpha = 0.05$ 이므로, $z_{0.025} = 1.96$ 이다.

④ $z_0 = 1.9620 > z_{0.025} = 1.96$ 이므로 H_0 를 기각한다.

따라서 두 도시 사이의 정당 지지율에 차이가 있다고 할 수 있다.

7.21

아래 그림의 엑셀 시트와 같이 자료를 입력하고, [데이터분석] ⇒ [Z-검정;평균에 대한 두 집단]을 선택한 후 아래 입력창의 내용을 입력한다.



	A	B
1	남	여
2	3	12
3	4	5
4	12	4
5	16	10
6	5	1
7	11	8
8	21	19
9	9	13
10	8	9
11	25	16
12	17	13
13	3	13
14	8	7
15	6	9
16	13	15
17	7	8
18	30	28
19	12	
20	9	
21	10	
22	남자분산	여자분산
23	31.36	20.25
24		
25		

그러면 다음과 같은 결과를 얻는다.

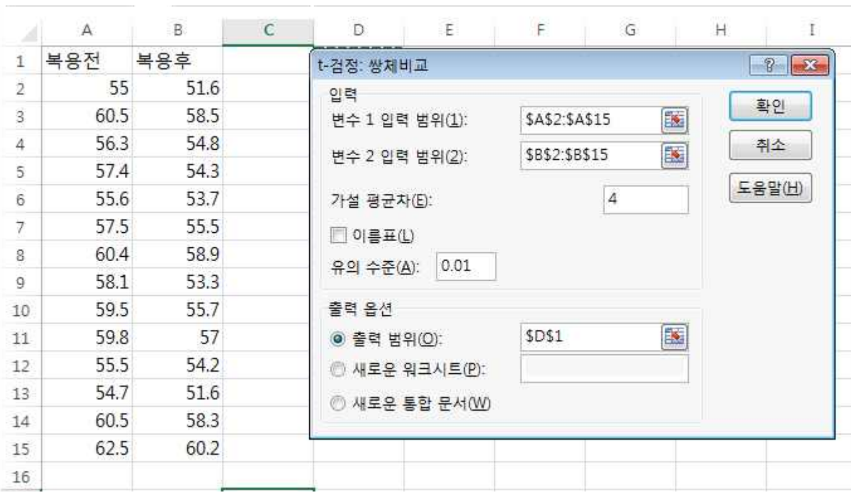
z-검정: 평균에 대한 두 집단

	변수 1	변수 2
평균	11.45	11.17647
기지의 분산	31.36	20.25
관측수	20	17
가설 평균차	0	
z 통계량	0.16467	
P(Z<=z) 단측 검정	0.434602	
z 기각치 단측 검정	1.644854	
P(Z<=z) 양측 검정	0.869204	
z 기각치 양측 검정	1.959964	

유의수준 5%에서 귀무가설을 기각하지 못함을 알 수 있다. 즉, 남녀의 입원 기간에 차이가 있다고 할 수 없다.

7.23

다음과 같이 [데이터분석] ⇒ [t-검정 : 쌍체비교]를 선택한 후 자료를 입력한다. 이때 가설 평균 차는 4, 유의수준은 0.01로 입력한다.



이때 출력결과는 다음과 같다.

t-검정: 쌍체 비교

	변수 1	변수 2
평균	58.09286	55.54286
분산	6.099176	7.298022
관측수	14	14
피어슨 상관 계수	0.929114	
가설 평균차	4	
자유도	13	
t 통계량	-5.42645	
P(T<=t) 단측 검정	5.79E-05	
t 기각치 단측 검정	2.650309	
P(T<=t) 양측 검정	0.000116	
t 기각치 양측 검정	3.012276	

따라서 p-값이 0.000으로 0.01보다 작으므로 귀무가설을 기각한다. 즉, 다이어트 식품을 한 달 간 복용하면 4kg의 체중 감량효과가 있다고 할 수 있다.

7.25

위와 같은 방법으로 95% 신뢰구간을 구하면 (118.05, 127.75)이고 이 구간은 120분을 포함하고 있다. 따라서 페인트가 마르는 데 평균 2시간이 걸린다는 주장은 타당하다.

7.27

유의확률(p-값)이 0.018로 유의수준 0.05보다 작으므로 귀무가설을 기각한다. 즉, 두 광산에서 생산한 석탄의 열량에 차이가 있다고 할 수 있다. 그러나 유의수준 0.01에서는 p-값이 더 크므로 귀무가설을 기각하지 못한다.

Chapter 08 연습문제

8.1

(a)

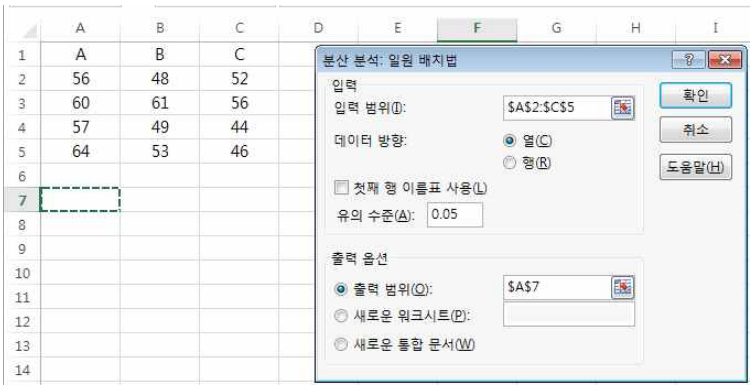
요인	제곱합	자유도	평균제곱	F
처리	320	2	160	
잔차	180	9	20	8
계	500	11		

(b) 기각역은 $F \geq \mathrm{F}_{0.05}(2,9) = 4.26$

(c) F 비가 8이므로 귀무가설을 기각한다.

8.3

자료를 입력하고 [데이터분석] ⇒ [분산분석-일원배치법]을 선택하여 다음과 같이 입력한다.



출력결과를 보면 다음과 같다.

분산 분석: 일원 배치법

요약표

인자의 수준	관측수	합	평균	분산
Column 1	4	237	59.25	12.91667
Column 2	4	211	52.75	34.91667
Column 3	4	198	49.5	30.33333

분산 분석

변동의 요인	제곱합	자유도	제곱 평균	F 비	p-값	F 기각치
처리	197.1667	2	98.58333	3.783582	0.064191	4.256495
잔차	234.5	9	26.05556			
계	431.6667	11				

여기서 p-값이 0.064로 0.05보다 크기 때문에 귀무가설을 기각할 수 없다. 따라서 벼의 종류에 따라 수확량에 차이가 있다고 볼 수 없다.

8.5

위와 같은 방법으로 입력하고 실행한 출력결과는 다음과 같다.

분산 분석: 일원 배치법

요약표

인자의 수준	관측수	합	평균	분산
Row 1	6	171	28.5	9.5
Row 2	6	97	16.16667	4.166667
Row 3	6	126	21	11.2

분산 분석

변동의 요인	제곱합	자유도	제곱 평균	F 비	p-값	F 기각치
처리	463.4444	2	231.7222	27.95576	8.72E-06	3.68232
잔차	124.3333	15	8.288889			
계	587.7778	17				

p-값이 8.7×10^{-6}으로 작기 때문에 귀무가설을 기각한다. 따라서 디자인 차이에 따라 매출고에 차이가 있다고 할 수 있다.

8.7

[데이터분석] ⇒ [분산분석 : 반복이 없는 이원배치법]을 적용하여 분석하면 다음과 같다.

분산 분석

변동의 요인	제곱합	자유도	제곱 평균	F 비	p-값	F 기각치
인자 A(행)	6	1	6	3	0.225403	18.51282
인자 B(열)	52	2	26	13	0.071429	19
잔차	4	2	2			
계	62	5				

따라서 자동차 운전을 익히는 데 소요된 시간은 성별 및 연령에 따라 모두 차이를 보이지 않는다.

8.9

자료를 다음과 같이 입력하고 [데이터분석]⇒[분산분석 : 반복이 있는 이원배치법]을 선택한 후 입력범위를 정한다.

	A	B	C	D	E
1		학생1	학생2	학생3	학생4
2	교사1	40	56	61	46
3		46	65	64	54
4		52	67	70	62
5	교사2	41	56	69	57
6		44	64	73	63
7		50	69	77	72
8	교사3	72	84	90	87
9		81	93	97	92
10		87	96	98	94

이때 출력결과는 다음과 같다.

분산 분석: 반복 있는 이원 배치법

요약표	학생1	학생2	학생3	학생4	계
교사1					
관측수	3	3	3	3	12
합	138	188	195	162	683
평균	46	62.66667	65	54	56.91667
분산	36	34.33333	21	64	89.90152
교사2					
관측수	3	3	3	3	12
합	135	189	219	192	735
평균	45	63	73	64	61.25
분산	21	43	16	57	137.4773
교사3					
관측수	3	3	3	3	12
합	240	273	285	273	1071
평균	80	91	95	91	89.25
분산	57	39	19	13	57.29545
계					
관측수	9	9	9	9	
합	513	650	699	627	
평균	57	72.22222	77.66667	69.66667	
분산	326.25	227.4444	195	308.25	

분산 분석

변동의 요인	제곱합	자유도	제곱 평균	F 비	p-값	F 기각치
인자 A(행)	7392.889	2	3696.444	105.5289	1.28E-12	3.402826
인자 B(열)	2068.75	3	689.5833	19.68676	1.17E-06	3.008787
교호작용	222	6	37	1.056305	0.415207	2.508189
잔차	840.6667	24	35.02778			
계	10524.31	35				

이 결과를 보면 교사에 따른 차이와 학생에 따른 차이는 있고, 교사와 학생 간의 상호작용(교호작용) 효과는 없다.

8.11

위와 같은 방법으로 자료를 입력하여 얻은 결과는 다음과 같다.

분산 분석

변동의 요인	제곱합	자유도	제곱 평균	F 비	P-값	F 기각치
인자 A(행)	121	1	121	8	0.01522	4.747225
인자 B(열)	42.25	1	42.25	2.793388	0.120506	4.747225
교호작용	930.25	1	930.25	61.50413	4.61E-06	4.747225
잔차	181.5	12	15.125			
계	1275	15				

이 결과로부터 일반 아동과 ADHD 아동 간에는 차이가 있고, 위약과 리탈린 간에는 차이를 보이지 않는다. 또한 상호작용(교호작용) 효과는 있다고 할 수 있다.

Chapter 09 연습문제

9.1

i 번째 기대도수 : $e_i = np_i = 500 \times 0.2 = 100$

검정통계량 : $\chi_0^2 = \dfrac{(30-100)^2}{100} + \dfrac{(70-100)^2}{100} + \cdots + \dfrac{(180-100)^2}{100} = 126$

$\alpha = 0.05$ 이므로 $\chi^2_{0.05}(4) = 9.49$ 이다.

$\chi_0^2 = 126 > \chi^2_{0.05}(4) = 9.490$이므로 H_0를 기각한다. 따라서 컬러 모니터의 크기에 따른 소비자의 성향에는 차이가 있다고 할 수 있다.

9.3

① H_0 : $p_{ij} = p_{i\cdot}p_{\cdot j}$ $(i, j = 1, 2, 3)$

H_1 : H_0가 아니다.

② 귀무가설 H_0가 참일 때의 기대도수 :

$$e_{11} = \frac{60 \times 60}{200} = 18,\ e_{12} = \frac{60 \times 70}{200} = 21,$$

$$e_{13} = \frac{60 \times 70}{200} = 21,\ e_{21} = \frac{100 \times 60}{200} = 30,$$

$$e_{22} = \frac{100 \times 70}{200} = 35,\ e_{23} = \frac{100 \times 70}{200} = 35,$$

$$e_{31} = \frac{40 \times 60}{200} = 12,\quad e_{32} = \frac{40 \times 70}{200} = 14,\quad e_{33} = \frac{40 \times 70}{200} = 14$$

검정통계량 : $\chi_0^2 = \dfrac{(30-18)^2}{18} + \dfrac{(20-21)^2}{21} + \cdots + \dfrac{(20-14)^2}{14} \fallingdotseq 22.619$

③ $\alpha = 0.05$ 이므로 $\chi^2_{0.05}((3-1)(3-1)) = \chi^2_{0.05}(4) = 9.49$ 이다.

④ $\chi_0^2 = 22.619 > \chi^2_{0.05}(4) = 9.49$이므로 H_0를 기각한다. 따라서 경제 사정과 쇼핑장소의 선택이 서로 독립이 아니므로 경제 사정에 따라 쇼핑장소의 선택이 다름을 알 수 있다.

9.5

적합도 검정을 실시한다.

① $H_0 \; : p_1 = 0.45, p_2 = 0.2, \; p_3 = 0.2, \; p_4 = 0.15$

H_1 : H_0가 아니다.

② 귀무가설 H_0가 참일 때의 기대도수 :

$e_1 = 235 \times 0.45 = 105.75, \; e_2 = 235 \times 0.2 = 47,$
$e_3 = 235 \times 0.2 = 47, \qquad e_4 = 235 \times 0.15 = 35.25$

검정통계량 : $\chi_0^2 = \dfrac{(92-105.75)^2}{105.75} + \dfrac{(69-47)^2}{47} + \dfrac{(32-47)^2}{47} + \dfrac{(42-35.25)^2}{35.25} \fallingdotseq 18.17$

③ $\alpha = 0.05$이므로 $\chi^2_{0.05}(3) = 7.815$이다.

④ $\chi_0^2 = 18.17 > \chi^2_{0.05}(3) = 7.815$이므로 H_0를 기각한다. 따라서 취업률이 예년과 다르다고 할 수 있다.

9.7

적합도 검정을 실시한다.

① 접촉 사고횟수를 확률변수 X라고 하자.

$H_0 \; : X \sim P(\lambda)$

H_1 : H_0가 아니다.

② 귀무가설 H_0가 참일 때 $f(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$이므로 모수 λ값을 추정해야 한다. 푸아송분포에서 λ는 평균이므로 이를 표본평균으로 추정하면 다음과 같다.

$$\hat{\lambda} = \bar{x} = \frac{1}{200}(0 \times 23 + 1 \times 55 + 2 \times 58 + \cdots + 7 \times 1) = 2$$

따라서 $f(x) = \dfrac{2^x e^{-2}}{x!}, \; x = 0, 1, 2, \cdots$이다.

이때 접촉 사고횟수 X에 대한 기대도수를 구하면 다음과 같다.

$$e_0 = nf(0) = 200 \times \frac{2^0 e^{-2}}{0!} \fallingdotseq 27.07$$
$$e_1 = nf(1) = 200 \times \frac{2^1 e^{-2}}{1!} \fallingdotseq 54.13$$

$$e_2 = nf(2) = 200 \times \frac{2^2 e^{-2}}{2!} \fallingdotseq 54.13$$

$$e_3 = nf(3) = 200 \times \frac{2^3 e^{-2}}{3!} \fallingdotseq 36.09$$

$$\vdots$$

$$e_7 = nf(7) = 200 \times \frac{2^7 e^{-2}}{7!} \fallingdotseq 0.69$$

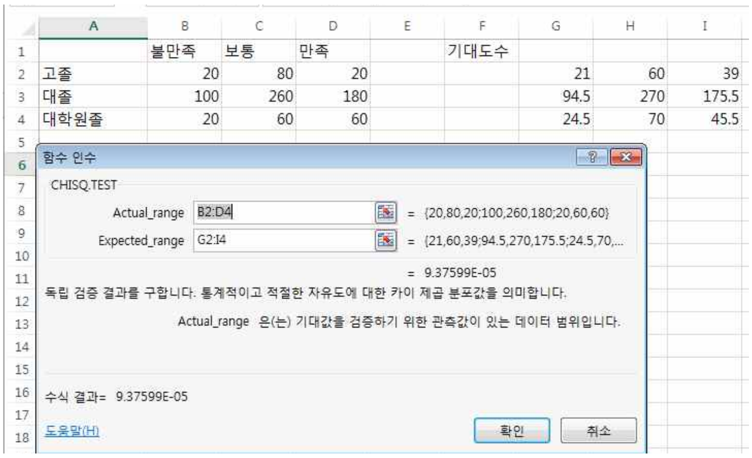
검정통계량 : $\chi_0^2 = \frac{(23-27.07)^2}{27.07} + \frac{(55-54.13)^2}{54.13} + \cdots + \frac{(1-0.69)^2}{0.69} \fallingdotseq 2.09$

③ $\alpha = 0.05$ 이므로 $\chi_{0.05}^2(7) = 14.07$이다.

④ $\chi_0^2 = 2.09 < \chi_{0.05}^2(7) = 14.07$이므로 H_0를 기각하지 못한다. 따라서 자동차 접촉 사고횟수는 푸아송분포를 따른다고 할 수 있다.

9.9

위와 같이 기대도수를 먼저 구하고 같은 방법으로 검정하면 $p-$값$=9.37\times 10^{-5}$을 얻는다. 따라서 귀무가설은 기각되고 학력에 따라 직장생활의 만족도에 차이가 있음을 알 수 있다.



Chapter 10 연습문제

10.1

(a) $\overline{x} = \frac{1}{5}(1+2+\cdots+5) = 3$

$\overline{y} = \frac{1}{5}(2+5+\cdots+8) = 5$

$$\sum_{i=1}^{5}(x_i - \overline{x})^2 = (1-3)^2 + (2-3)^2 + \cdots + (5-3)^2 = 10$$

$$\sum_{i=1}^{5}(y_i - \overline{y})^2 = (2-5)^2 + (5-5)^2 + \cdots + (8-5)^2 = 26$$

$$\sum_{i=1}^{5}(x_i - \overline{x})(y_i - \overline{y}) = (1-3)(2-5) + (2-3)(5-5) + \cdots + (5-3)(8-5) = 14$$

$$r = \frac{14}{\sqrt{10}\sqrt{26}} \fallingdotseq 0.868$$

(b) $H_0 \; : \; \rho = 0$

$H_1 \; : \; \rho > 0$

$n = 5$ 이고 $r = 0.868$ 이므로 검정통계량 :

$$t_0 = \frac{0.868\sqrt{5-2}}{\sqrt{1-(0.868)^2}} \fallingdotseq 3.028$$

$t_{0.05}(5-2) = t_{0.05}(3) = 2.353$

$t_0 = 3.028 > t_{0.05}(3) = 2.353$ 이므로 H_0 를 기각한다.

따라서 첨가물의 양과 수율 사이에 양의 상관관계가 있다고 할 수 있다.

(c) $\hat{b} = \frac{14}{10} = 1.4$

$\hat{a} = 5 - 1.4 \times 3 = 0.8 \qquad \Rightarrow \; \hat{y} = 0.8 + 1.4x$

(d) $SSE = 26 - \frac{14^2}{10} = 6.4$

$\hat{\sigma}^2 = s^2 = \frac{6.4}{5-2} \fallingdotseq 2.133$

(e) $r^2 = \frac{14^2}{10 \times 26} \fallingdotseq 0.754$

(f) $t_{0.025}(5-2) = t_{0.025}(3) = 3.182$

$$0.8 - 3.182\sqrt{2.133\left(\frac{1}{5}+\frac{3^2}{10}\right)} < a < 0.8 + 3.182\sqrt{2.133\left(\frac{1}{5}+\frac{3^2}{10}\right)}$$

$\Rightarrow$ $0.8 - 4.874 < a < 0.8 + 4.874$

$\Rightarrow$ $-4.074 < a < 5.674$

(g) $1.4 - 3.182\sqrt{\frac{2.133}{10}} < b < 1.4 + 3.182\sqrt{\frac{2.133}{10}}$

$\Rightarrow$ $1.4 - 1.470 < b < 1.4 + 1.470$

$\Rightarrow$ $-0.07 < b < 2.87$

(h) $x = 6$ 일 때, $\hat{a} + \hat{b}(6) = 0.8 + 1.4 \times 6 = 9.2$

$$9.2 - 3.182\sqrt{2.133\left(\frac{1}{5}+\frac{(6-3)^2}{10}\right)} < a + bx$$

$$< 9.2 + 3.182\sqrt{2.133\left(\frac{1}{5}+\frac{(6-3)^2}{10}\right)}$$

$\Rightarrow$ $9.2 - 4.874 < a + bx < 9.2 + 4.874$

$\Rightarrow$ $4.326 < a + bx < 14.074$

(i) 검정통계량 : $|t_0| = \left|\dfrac{0.8-0}{\sqrt{2.133\left(\frac{1}{5}+\frac{3^2}{10}\right)}}\right| \fallingdotseq 0.522$

$t_{0.025}(5-2) = t_{0.025}(3) = 3.182$

$|t_0| = 0.522 < t_{0.025}(3) = 3.182$ 이므로 H_0 를 기각하지 못한다.

(j) 검정통계량 : $t_0 = \dfrac{1.4-0}{\sqrt{\frac{2.133}{10}}} \fallingdotseq 3.031$

$t_{0.05}(5-2) = t_{0.05}(3) = 2.353$

$t_0 = 3.031 > t_{0.05}(3) = 2.353$ 이므로 H_0 를 기각한다.

10.3

$r(aX+b, cY+d)$은 $ac > 0$ 일 때 r 이고, $ac < 0$ 일 때 $-r$ 이 되므로 $r(2X+3, -7Y+4) = -r$ 이다.

10.5

$H_0 \;:\; \rho = 0$

$H_1 \;:\; \rho > 0$

$n = 101$ 으로 표본의 크기가 크고, $r = 0.61$ 이므로 다음이 성립한다.

검정통계량 : $t_0 = \dfrac{0.61\sqrt{109}}{\sqrt{1-0.61^2}} \fallingdotseq 8.04$

$t_0 = 8.04 > t_{0.05}(109) \approx Z_{0.05} = 1.645$ 이므로 H_0 를 기각한다. 따라서 반응 온도와 품질의 특성치 간에 양의 상관관계가 있다고 할 수 있다.

10.7

(a) $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \dfrac{4}{10} = 0.4$

$\overline{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i = \dfrac{6}{10} = 0.6$

$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\overline{x})^2 = 14 - 10(0.4)^2 = 12.4$

$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n(\overline{y})^2 = 18 - 10(0.6)^2 = 14.4$

$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y} = 12 - 10(0.4)(0.6) = 9.6$

$\Rightarrow\; r = \dfrac{9.6}{\sqrt{12.4}\sqrt{14.4}} \fallingdotseq 0.718$

(b)

$$\hat{b} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{9.6}{12.4} \fallingdotseq 0.774$$

$\hat{a} = \overline{y} - \hat{b}\,\overline{x}$

$= 0.6 - 0.774 \times 0.4 = 0.2904$

$\Rightarrow\; \hat{y} = 0.2904 + 0.774x$

(c) $SSE = 14.4 - \dfrac{9.6^2}{12.4} \fallingdotseq 6.97$

$\hat{\sigma}^2 = s^2 = \dfrac{6.97}{10-2} \fallingdotseq 0.87$

(d) $r^2 = \dfrac{9.6^2}{12.4 \times 14.4} = 0.516$

(e) $t_{0.025}(10-2) = t_{0.025}(8) = 2.306$

$$0.774 - 2.306\sqrt{\frac{0.87}{12.4}} < b < 0.774 + 2.306\sqrt{\frac{0.87}{12.4}}$$

$\Rightarrow\ 0.774 - 0.611 < b < 0.774 + 0.611$

$\Rightarrow\ 0.163 < b < 1.385$

(f) 검정통계량 : $|t_0| = \left| \dfrac{0.774 - 0}{\sqrt{\dfrac{0.87}{12.4}}} \right| \fallingdotseq 2.922$

$|t_0| = 2.922 > t_{0.025}(8) = 2.306$ 이므로 H_0 를 기각한다.

10.9

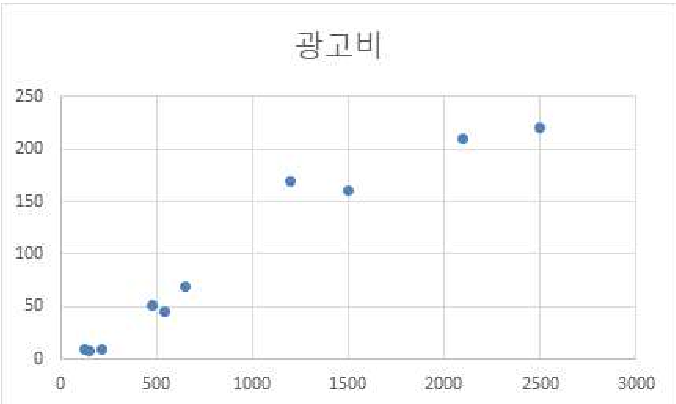
$$b = r\frac{s_y}{s_x} = 0.74 \times \frac{3}{4} = 0.555$$

10.11

Excel에 자료를 다음과 같이 입력하고 매출액과 광고비에 대한 상관도를 아래와 같이 그릴 수 있다.

그리고 엑셀에서 [데이터 분석] 탭에서 [상관분석]을 선택하여 구하면 상관계수는 0.97088로 거의 1에 가까우므로 매출액과 광고비는 강한 양의 상관관계를 갖는다고 할 수 있다.

매출액	광고비
2500	220
2100	210
1500	160
1200	170
650	70
540	46
480	51
220	10
150	8
125	9



	매출액	광고비
매출액	1	
광고비	0.97088	1

10.13

Excel에 자료를 다음과 같이 입력하고 검정하면 다음과 같다.

결석시간(X)	성적(Y)
1	90
3	93
3	81
0	92
0	95
2	85
4	77
7	83
5	82
6	79

(a) Excel에서 [데이터]탭에서 [데이터 분석] 탭을 선택하고 [상관분석]을 선택하여 표본상관계수를 구하면 다음과 같다.

	결석시간(X)	성적(Y)
결석시간(X)	1	
성적(Y)	-0.734816424	1

(b) 엑셀을 이용하여 결석시간과 성적 사이에 상관관계가 있는지 검증해 보면 다음과 같다.

표본수	10	
유의수준	0.05	
표본상관계수	-0.73482	=CORREL(B2:B11,C2:C11)
통계량	-3.06425	=F3/SQRT((1-F3^2)/(F1-2))
임곗값(좌측)	-2.306	=-TINV(F2,8)
임곗값(우측)	2.306004	=TINV(F2,8)
판정	채택	=IF(OR(F4<=F5, F4>=F6), "채택", "기각")

즉, 결석시간과 성적 사이에는 음의 상관관계가 존재함을 알 수 있다.