연습문제
풀이 및 해답

01 총조사와 표본조사의 차이와 장단점을 예를 들어 설명하라.

[예시] 인구총조사와 같은 총조사는 모집단을 구성하는 모든 개체들을 조사하여 데이터를 얻는 방법이다. 전체를 모두 조사하는 것이기 때문에 통계적 추론을 적용할 필요가 없지만 비용과 시간이 많이 소요된다. 치료제 개발 등에서 사용되는 표본조사는 모집단 중 일부(표본)를 추출하여 모집단의 특성을 추론하는 것이다. 전수조사에 비해 비용과 시간이 절감되지만 표본이 모집단을 잘 대표할 수 없을 때는 문제가 생길 수 있다.

02 표본조사의 예를 설계하여 제시하여보고, 제시한 표본조사에서 모집단, 원소, 모수, 표본을 정의하여 보라. 또한, 본인이 설계한 표본조사에 가장 적합한 표본조사법은 무엇이며 그 이유는 무엇인지 설명하여 보라.

[예시] **특정 후보자의 지지율 조사**

모집단: 투표권이 있는 모든 사람

모수: 지지율

표본: 10000명의 유권자

표본추출법: 출구조사를 실시할 때, 5~7명당 1명씩 계통추출

모집단의 원소 수가 매우 크기 때문에 단순임의추출법보다는 계통추출이 적합하다고 생각한다.

03 다음 자료를 측정하는 척도가 무엇인지 각각 판단하여 보라.

(1) 올해 8월 서울지역의 강수량

(2) 어느 도시 30대 남자의 학력

(3) 장마철 교실의 습도

(4) 대입 정시 확대에 대한 찬성 반대 의견

(5) 사람의 혈액형

(6) 근무회사의 복지정책에 대한 만족도

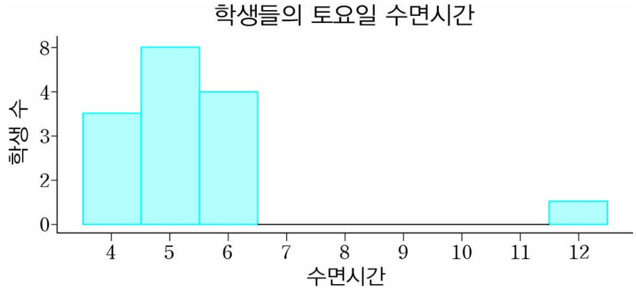
✎ **해답**

(1) 비율척도 (2) 순서척도 (3) 비율척도 (4) 명목척도 (5) 명목척도 (6) 구간척도

04 A고등학교 3학년 학생 20명의 토요일 수면시간을 측정한 다음 표본에 대하여 히스토그램을 작성하여 보라. (단위: 시간)

5	5	4	6	6	6	5	4	5	12
4	5	5	6	4	6	5	5	4	6

✎해답



05 04의 자료에 대하여 표본평균, 표본중위수, 표본최빈값, 표본분산을 구하여 보라.

✎해답 표본평균 = 5.4, 표본중위수 = 5, 표본최빈값 = 5, 표본분산 = 2.9895

06 B고등학교 3학년 학생 10명의 토요일 수면시간을 측정한 다음 표본에 대하여 표본평균과 표본분산을 구하고 04의 A고등학교 학생의 수면시간과 비교하여 보라. (단위: 시간)

6	6	6	5	5	6	7	6	6	7

✎해답 표본평균=6, 표본분산=0.4444

[예시] 토요일 수면시간에 대한 표본평균은 B고등학교가 더 높지만 분산은 A고등학교가 더 크다. 즉, B고등학교 학생들의 수면시간은 평균과 차이가 크지 않지만 A고등학교 학생들의 수면시간은 평균을 많이 벗어난 값들이 존재한다.

07 다음은 기초통계학 수강자 60명의 기말시험 성적이다. 아래 물음에 답하라.

74	87	87	82	99	93	86	96	78	91
82	69	89	90	72	92	89	86	92	69
76	74	75	67	96	84	77	87	79	80
87	80	92	85	90	79	83	79	69	81
80	90	80	75	89	86	71	74	88	86
86	76	85	74	81	85	74	85	85	88

(1) 가장 낮은 점수 (2) 점수의 평균값
(3) 점수의 중앙값 (4) 15등 학생의 점수
(5) 80점 이상 학생 비율 (6) 가장 높은 점수
(7) 표와 같이 학점을 준다고 할 때, 학점에 대한 도수분포표를 작성하라.

점수	학점
90~100	A
80~89	B
70~79	C
60~69	D
0~59	F

✎해답 (1) 67 (2) 82.683 (3) 84.5 (4) 88 (5) 0.6667 (6) 99

(7)

계급	학점	도수
0~59	F	0
60~69	D	4
70~79	C	16
80~89	B	29
90~100	A	11

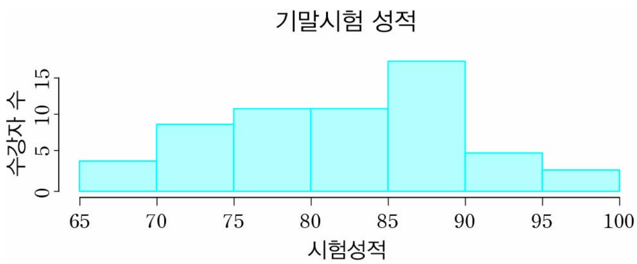
08 07의 자료에 대한 줄기-잎 그림을 작성하라.

✎해답

```
6 | 7999
7 | 1244444
7 | 556678999
8 | 0000112234
8 | 5555566666777788999
9 | 00012223
9 | 669
```

09 07의 자료를 히스토그램으로 표현하고, 분포의 형태에 대하여 설명하여 보라.

✎해답



[예시] 기초통계학 수강자들의 기말고사 성적이 평균 주변에 분포되어 있다.

10 이를 이용하여 위 자료가 따르는 분포를 추측하여 보라.

[예시] 정확한 대칭은 아니지만 좌우가 비슷한 종모양의 형태를 보이기 때문에 평균이 84.5이고 표준편차가 7.501인 정규분포를 따른다.

11 다음 자료는 학생 30명의 혈액형을 조사한 것이다.

O	A	B	A	A	A	B	O	O	AB
B	A	O	A	A	AB	AB	B	B	A
B	O	O	AB	A	A	A	B	B	AB

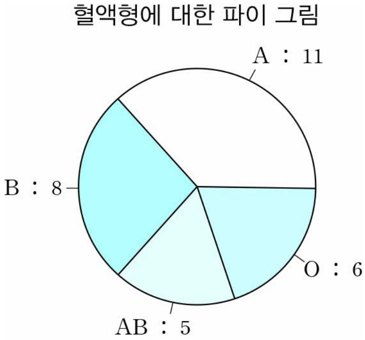
(1) 위의 자료를 이용하여 도수분포표를 작성하라.

(2) 위의 자료를 이용하여 파이 그림을 작성하라.

해답 (1)

계급	도수
A	11
B	8
AB	5
O	6

(2)



3장 연습문제

풀이 | 및 | 해답

01 주머니에 1부터 5까지 숫자가 적힌 5장의 카드가 있다. 이 상자에서 2장의 카드를 동시에 뽑는 실험을 하고자 한다.

(1) 표본공간을 정의하라.

(2) 뽑힌 2개의 카드의 합이 짝수일 사건을 모두 나열하라.

해답

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

(2) $A=$ 뽑힌 2개의 카드의 합이 짝수

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

02 01에서 처음 한 장의 카드를 꺼내어 기록하고 다시 상자에 넣어 잘 섞은 후 다시 한 장을 꺼내는 실험을 한다. 다음 각 경우에 해당하는 확률변수의 표본공간을 정의하라.

(1) $X \equiv$ 두 카드의 숫자의 합이 짝수

(2) $Y \equiv$ 첫 번째와 두 번째 숫자의 차이

해답

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

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

03 정상제품 80개와 불량품 20개가 들어있는 상자에서 2개의 제품을 연속하여 꺼내는 실험을 할 때, 다음의 확률을 구하라. 단, 한번 꺼낸 물품은 다시 상자에 넣지 않는다.

(1) 처음 꺼낸 물품이 불량일 확률

(2) 적어도 한 개의 물품이 불량일 확률

(3) 처음 꺼낸 물품이 정상제품일 때, 두 번째 꺼낸 물품도 정상제품일 확률

✎해답 (1) $\frac{20}{100}$

(2) P(적어도 한 개의 물품이 불량)$= 1 - P$(2개의 물품이 정상)

$$= 1 - \left(\frac{80}{100} \times \frac{79}{99}\right) = 0.3616$$

(3) P(두 번째 정상 첫 번째 정상)$= \dfrac{P(\text{첫 번째 정상} \cap \text{두 번째 정상})}{P(\text{첫 번째 정상})}$

$$= \frac{\frac{80}{100} \times \frac{79}{99}}{\left(\frac{80}{100} \times \frac{79}{99}\right) + \left(\frac{80}{100} \times \frac{20}{99}\right)} = 0.798$$

04 다음의 실험에서 정의되는 표본공간을 정의하라.

(1) 100원짜리 동전을 세 번 던져 관측된 앞면의 수

(2) 운전면허시험 응시자 20명 가운데 합격자의 수

(3) 두통약을 복용하고 두통이 사라질 때까지 걸리는 시간

✎해답 (1) $\Omega = \{0, 1, 2, 3\}$

(2) $\Omega = \{0, 1, 2, \cdots, 20\}$

(3) $\Omega = \{x \mid x \geq 0\}$

05 1에서 12까지의 정수로 구성된 집합 $S = \{1, 2, \cdots, 12\}$가 있다. 이때, S의 부분집합 A, B, C를 각각 다음과 같이 정의하자.

$$A = \{2, 4, 6, 8, 10, 12\}$$
$$B = \{1, 3, 5, 7, 9, 11\}$$
$$C = \{1, 2, 3, 4, 5, 6\}$$

(1) $A \cap B$, $A \cup B$, $A^C \cap C$를 구하라.

(2) 집합 A, B, C 중 서로 배반인 집합은?

(3) 집합 A, B, C는 서로 독립인가?

✎**해답** (1) $A \cap B = \{\ \}$ 또는 $A \cap B = \phi$

$A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

$A^C \cap C = \{1, 3, 5\}$

(2) A와 B

(3) $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{2}$

$P(A \cap B) = 0$, $P(A \cap C) = \frac{1}{4}$, $P(B \cap C) = \frac{1}{4}$, $P(A \cap B \cap C) = 0$

따라서 A와 C, B와 C 만 서로 독립이다.

06 다음을 증명하라.

$$\begin{aligned} P(E \cup F \cup G) = {} & P(E) + P(F) + P(G) - P(E \cap F) \\ & - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G) \end{aligned}$$

✎**해답**

$$\begin{aligned} P(E \cup F \cup G) &= P((E \cup F) \cup G) = P(E \cup F) + P(G) - P((E \cap F) \cup G) \\ &= P(E) + P(F) - P(E \cap F) + P(G) - P((E \cap G) \cup (F \cap G)) \\ &= P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) \\ &\qquad - P(F \cap G) + P(E \cap F \cap G) \end{aligned}$$

07 네 개의 동전을 던지는 실험에서 사건 A, B, C를 각각 다음과 같이 정의하자.

A: 첫 번째 동전이 앞면이 나오는 사건

B: 네 번째 동전이 앞면이 나오는 사건

C: 동전 네 개 모두 앞면이 나오는 사건

(1) 표본공간을 정의하라.

(2) $P(A)$, $P(B)$, $P(C)$, $P(A \cap C)$를 구하라.

✎**해답**

(1) $\Omega = \{HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTHT, THHT, HTTH,$
$THTH, TTHH, HTTT, THTT, TTHT, TTTH, TTTT\}$

(2) $A = \{HHHH, HHHT, HHTH, HTHH, HHTT, HTHT, HTTH, HTTT\}$

$B = \{HHHH, HHTH, HTHH, THHH, HTTH, THTH, TTHH, TTTH\}$

$C = \{HHHH\}$

$A \cap C = \{HHHH\}$

$P(A) = \frac{8}{16} = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{16}$, $P(A \cap C) = \frac{1}{16}$

08 사건 A, B가 서로 독립이고 $P(A)=0.3, P(B)=0.4$이다. 이때, $P(A|B)$, $P(B|A)$, $P(A\cap B)$, $P(A\cup B)$를 각각 구하라.

✎**해답** A와 B가 서로 독립이므로 $P(A\cap B) = P(A)\bullet P(B)$이고 $P(A|B) = P(A)$, $P(B|A) = P(B)$가 성립한다.

$\therefore\ P(A|B) = 0.3,\ P(B|A) = 0.4,\ P(A\cap B) = 0.12$

$P(A\cup B) = P(A) + P(B) - P(A\cap B) = 0.58$

09 A형 간염 항체에 대한 검사에서, A형 간염에 걸린 사람들 중에서 90%가 양성반응을 보이고 A형 간염에 걸리지 않은 사람들 중에서 10%가 양성반응을 보인다. 전체 사람 중에서 A형 간염에 걸린 사람이 1%일 때, A형 간염 항체 반응 검사에서 양성반응을 보인 사람이 실제로 A형 간염에 걸렸을 확률을 구하라.

✎**해답**

$\oplus$: 양성반응 , A^+: A형 간염에 걸린 사람 , A^-: A형 간염에 걸리지 않은 사람

$P(\oplus|A^+)=0.9$, $P(\oplus|A^-)=0.1$, $P(A^+)=0.01$, $P(A^-)=0.99$

$$\begin{aligned}P(A|\oplus) = \frac{P(A^+\cap\oplus)}{P(\oplus)} &= \frac{P(A^+\cap\oplus)}{P(A^+\cap\oplus)+P(A^-\cap\oplus)}\\ &= \frac{P(A^+)P(\oplus|A^+)}{P(A^+)P(\oplus|A^+)+P(A^-)P(\oplus|A^-)}\\ &= \frac{0.01\times0.9}{(0.01\times0.9)+(0.99\times0.1)} = 0.083\end{aligned}$$

10 정형외과에 두 명의 물리치료사 A와 B가 일하는데, A가 전체 환자의 60%를, B가 전체 환자의 40%를 담당한다고 한다. 이 정형외과를 이용하는 환자들에게 설문조사를 한 결과 A에게서 치료를 받은 환자의 10%, B에게서 치료를 받은 환자의 5%가 불만을 나타냈다고 한다. 임의로 한 환자 한명에게 만족도를 물어 봤을 때, 이 사람이 불만족스러웠다고 대답할 확률은?

✎**해답** D: 불만족스러웠다고 대답한 환자

$P(A\cap D) = P(A)P(D|A) = 0.6\times0.01 = 0.06$

$P(B\cap D) = P(B)P(D|B) = 0.4\times0.05 = 0.02$

$P(D) = P(A\cap D)+P(B\cap D) = 0.08$

11 10의 상황에서, 한 명의 환자가 불만족하다고 했을 때, 이 환자가 A에게서 치료를 받았을 확률은?

✎해답 $P(A|D) = \dfrac{P(A \cap D)}{P(D)} = \dfrac{0.06}{0.08} = 0.75$

12 휴대전화를 판매하는 가게에서 세 회사(A, B, C)에서 생산되는 제품을 판매한다. 각 회사의 시장점유율은 40%, 30%, 20%이며, 1년 이내에 휴대전화기가 고장이 나는 비율은 각각 3%, 10%, 5%이다.

(1) 이 가게에서 A 회사 제품을 사고 1년 이내에 고장이 날 확률은?

(2) 이 가게에서 휴대전화를 사고 1년 이내에 고장이 날 확률을 구하라.

(3) 만약 이 가게에서 휴대전화를 구매하고 1년 이내에 고장이 났을 경우, 그 제품이 A회사 제품일 확률을 구하라.

✎해답 $P(A) = 0.4$, $P(B) = 0.3$, $P(C) = 0.2$

$P(T|A) = 0.03$, $P(T|B) = 0.1$, $P(T|C) = 0.05$

(1) $P(T \cap A) = P(A)P(T|A) = 0.4 \times 0.03 = 0.012$

(2) $P(T \cap B) = P(B)P(T|B) = 0.3 \times 0.1 = 0.03$

$P(T \cap C) = P(C)P(T|C) = 0.2 \times 0.05 = 0.01$

$P(T) = P(T \cap A) + P(T \cap B) + P(T \cap C) = 0.052$

(3) $P(A|T) = \dfrac{P(T \cap A)}{P(T)} = \dfrac{0.012}{0.052} = 0.231$

4장 연습문제

풀이 | 및 | 해답

01 확률변수 X의 확률분포가 다음과 같을 때, $E(X)$, $Var(X)$ 를 구하라. 또한 누적분포함수 $F(x)$를 구하고, 이를 그래프로 표현하라.

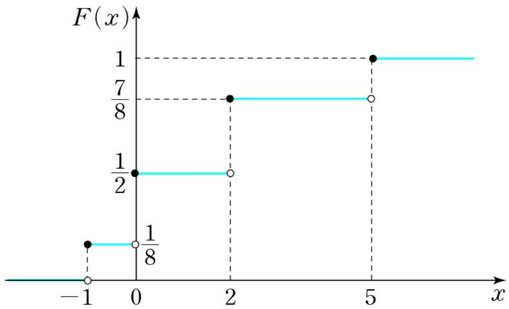
x	-1	0	2	5	합
$P(X=x)$	1/8	3/8	3/8	1/8	1

✎해답 $E(X) = ((-1) \times \frac{1}{8}) + (2 \times \frac{3}{8}) + (5 \times \frac{1}{8}) = 1.25$

$E(X^2) = ((-1)^2 \times \frac{1}{8}) + (2^2 \times \frac{3}{8}) + (5^2 \times \frac{1}{8}) = 4.75$

$Var(X) = E(X^2) - [E(X)]^2 = 3.1875$

$$F(x)=\begin{cases} 0, & x<-1 \\ \frac{1}{8}, & -1\le x<0 \\ \frac{1}{2}, & 0\le x<2 \\ \frac{7}{8}, & 2\le x<5 \\ 1, & 5\le x \end{cases}$$



02 어떤 상자에 검은색 공 8개, 흰색 공 2개가 들어있다. 이때 확률변수 X를 '흰 공을 찾을 때까지 공을 뽑은 횟수'라고 하자.

(1) 확률변수 X의 확률질량함수 $P(X=x)$를 구하라.

(2) 확률변수 X의 누적분포함수 $F(x)$를 구하고, 이를 그래프로 표현하라.

✎ **해답** (1) $P(X=1)=\frac{2}{10}$

$$P(X=2)=\frac{8}{10}\times\frac{2}{10}$$

$$P(X=3)=\frac{8}{10}\times\frac{8}{10}\times\frac{2}{10}=\left(\frac{8}{10}\right)^2\times\frac{2}{10}$$

$$\vdots$$

$$P(X=x)=\left(\frac{8}{10}\right)^{x-1}\times\frac{2}{10},\quad x=1,2,\cdots$$

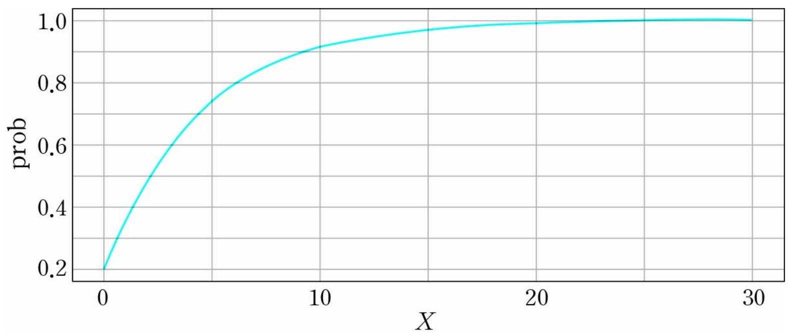
(2) $F(1)=P(X\le 1)=\frac{2}{10}$

$$F(2)=P(X\le 2)=\frac{2}{10}+\left(\frac{8}{10}\times\frac{2}{10}\right)=\frac{36}{100}$$

$$F(3)=P(X\le 3)=\frac{2}{10}+\left(\frac{8}{10}\times\frac{2}{10}\right)+\left(\frac{8}{10}\times\frac{8}{10}\times\frac{2}{10}\right)=\frac{488}{1000}$$

$$\vdots$$

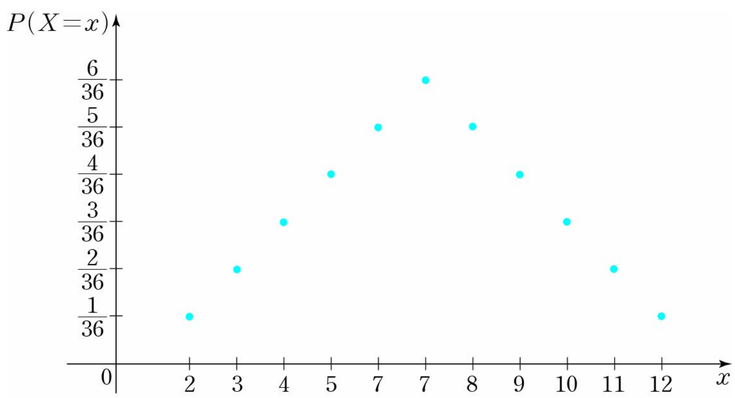
$$F(x)=P(X\le x)=1-\left(\frac{8}{10}\right)^x$$



03 주사위를 두 번 던져서 나타나는 눈금의 합을 확률변수 X라 할 때, X의 확률 분포를 구하고, 이를 그래프로 표현하라. 또한 평균과 분산을 구하라.

✎해답

x	2	3	4	5	6	7	8	9	10	11	12
$P(X=x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



$$E(X)=\left(2\times\frac{1}{36}\right)+\left(3\times\frac{2}{36}\right)+\left(4\times\frac{3}{36}\right)+\left(5\times\frac{4}{36}\right)+\left(6\times\frac{5}{36}\right)+\left(7\times\frac{6}{36}\right)$$
$$+\left(8\times\frac{5}{36}\right)+\left(9\times\frac{4}{36}\right)+\left(10\times\frac{3}{36}\right)+\left(11\times\frac{2}{36}\right)+\left(12\times\frac{1}{36}\right)=7$$

$$E(X^2)=\left(2^2\times\frac{1}{36}\right)+\left(3^2\times\frac{2}{36}\right)+\left(4^2\times\frac{3}{36}\right)+\left(5^2\times\frac{4}{36}\right)+\left(6^2\times\frac{5}{36}\right)+\left(7^2\times\frac{6}{36}\right)$$
$$+\left(8^2\times\frac{5}{36}\right)+\left(9^2\times\frac{4}{36}\right)+\left(10^2\times\frac{3}{36}\right)+\left(11^2\times\frac{2}{36}\right)+\left(12^2\times\frac{1}{36}\right)=54.8333$$

$$Var(X)=E(X^2)-[E(X)]=5.8333$$

04 확률변수 X의 확률밀도함수 $f(x)$가 다음과 같을 때 물음에 답하라.

$$f(x) = \begin{cases} k(1-x)^2, & 0 < x < 1 \\ 0 & , \text{ 그 외에서} \end{cases}$$

(1) $f(x)$가 연속형 확률변수 X의 확률밀도함수가 되도록 상수 k를 결정하라.

(2) 누적분포함수 $F(x)$를 구하라.

(3) 확률변수 X의 평균과 분산을 구하라.

✎**해답**

(1) $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 k(1-x)^2\,dx = k\left[x - x^2 + \frac{1}{3}x^3\right]_0^1 = \frac{1}{3}k = 1$이므로 $k=3$이다.

(2) $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_0^x 3(1-x)^2\,dx = x^3 - 3x^2 + 3x\ ,\quad 0 < x < 1$

(3) $E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_0^1 3x(1-x)^2\,dx = 3\left[\frac{1}{2}x^2 - \frac{2}{3}x^3 + \frac{1}{4}x^4\right]_0^1 = \frac{1}{4}$

$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 3x^2(1-x)^2\,dx = 3\left[\frac{1}{3}x^2 - \frac{1}{2}x^4 + \frac{1}{5}x^5\right]_0^1 = \frac{1}{10}$

$Var(X) = E(X^2) - [E(X)]^2 = \frac{3}{80}$

05 확률변수 X의 확률함수 $f(x)$가 다음과 같을 때 물음에 답하라.

$$f(x) = \begin{cases} kx^2, & 0 < x < 2 \\ 0 & , \text{ 그 외에서} \end{cases}$$

(1) $f(x)$가 연속형 확률변수 X의 확률밀도함수가 되도록 상수 k를 결정하라.

(2) 확률 $P(X \geq 1/2)$을 구하라.

(3) $E[2(X+1)^2]$를 구하라.

(4) 누적분포함수 $F(x)$를 구하라.

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

(2) $P(X \geq \frac{1}{2}) = \int_{\frac{1}{2}}^2 \frac{3}{8}x^2\,dx = \left[\frac{1}{8}x^3\right]_{\frac{1}{2}}^2 = \frac{63}{64}$

(3) $E[2(X+1)^2] = \int_0^2 2(x+1)^2 \cdot \frac{3}{8}x^2\,dx = \frac{3}{4}\left[\frac{1}{5}x^5 + \frac{1}{2}x^4 + \frac{1}{3}x^3\right]_0^2 = \frac{256}{15}$

(4) $F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_0^x \frac{3}{8}x^2\,dx = \frac{1}{8}x^3\ ,\quad 0 < x < 2$

06 확률변수 X와 Y의 결합확률분포가 다음과 같을 때, 다음 물음에 답하라.

$X \backslash Y$	0	1	2
0	0.1	0.0	0.1
1	0.1	0.2	0.2
2	0.0	0.1	0.2

(1) $P(X=1,\ Y=1)$ (2) $P(X \le 1,\ Y \le 1)$ (3) $P(X=x)$

(4) $P(Y=y)$ (5) $P(X \le 1)$ (6) $P(X+Y \le 3)$

✎해답

(1) $P(X=1,\ Y=1) = 0.2$

(2) $P(X \le 1,\ Y \le 1) = P(X=0,\ Y=0) + P(X=0,\ Y=1)$

$+ P(X=1,\ Y=0) + P(X=1,\ Y=1) = 0.4$

(3)

X	0	1	2
$P(X=x)$	0.2	0.5	0.3

(4)

Y	0	1	2
$P(Y=y)$	0.2	0.3	0.5

(5) $P(X \le 1) = P(X=0) + P(X=1) = 0.7$

(6)

$X+Y$	0	1	2	3	4
확률	0.1	0.1	0.3	0.3	0.2

07 확률변수 X와 Y의 결합확률분포가 06과 같을 때, 다음 물음에 답하라.

(1) X와 Y의 평균, 표준편차를 구하라.

(2) X와 Y의 공분산, 상관계수를 구하라.

(3) X와 Y의 두 확률변수는 독립인가?

✎해답

(1) $E(X) = \sum_{x=0}^{2} xP(X=x) = 1.1$, $E(Y) = \sum_{x=0}^{2} yP(Y=y) = 1.3$

$E(X^2) = \sum_{x=0}^{2} x^2 P(X=x) = 1.7$, $E(Y^2) = \sum_{x=0}^{2} y^2 P(Y=y) = 2.3$

$Var(X) = E(X^2) - [E(X)]^2 = 0.49$, $Var(Y) = E(Y^2) - [E(Y)]^2 = 0.61$

$\sigma_x = 0.7$, $\sigma_y = 0.78$

(2)

XY	0	1	2	4
확률	0.3	0.2	0.3	0.2

$E(XY) = (1\times 0.2) + (2\times 0.3) + (4\times 0.2) = 1.6$

$\therefore Cov(X, Y) = E(XY) - E(X)E(Y) = 0.17$

$$\rho(X, Y) = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_y} = 0.31$$

(3) $P(X=0, Y=0) = 0.1 \neq P(X=0)P(Y=0) = 0.4$이므로 두 확률변수 X, Y는 서로 독립이 아니다.

08 두 확률변수 X와 Y의 결합확률분포가 다음과 같을 때, 다음 물음에 답하라.

$$f(x, y) = \begin{cases} \frac{6}{5}(x^2 + y), & 0 < x < 1, 0 < y < 1 \\ 0 \quad , & \text{그 외에서} \end{cases}$$

(1) 확률변수 X의 주변확률밀도함수를 구하라.

(2) 확률 $P(0 < X < 0.5, 0 < Y < 0.5)$을 구하라.

(3) 두 확률변수의 상관계수를 구하라.

(4) 두 확률변수는 독립인가? 그 이유는?

✎ **해답**

(1) $$f(x) = \int_y f(x, y)\,dy = \int_0^1 \frac{6}{5}(x^2 + y)\,dy$$

$$= \frac{6}{5}\left[x^2 y + \frac{1}{2}y^2\right]_0^1 = \frac{6}{5}x^2 + \frac{3}{5}, \quad 0 < x < 1$$

(2) $$P(0 < X < 0.5, 0 < Y < 0.5) = \int_0^{0.5}\int_0^{0.5} \frac{6}{5}(x^2 + y)\,dx\,dy$$

$$= \frac{6}{5}\left[\frac{1}{24}y + \frac{1}{4}y^2\right]_0^{0.5}$$

(3) $$f(y) = \int_x f(x,y)\,dx = \int_0^1 \frac{6}{5}(x^2 + y)\,dx = \frac{6}{5}\left[\frac{1}{3}x^3 + xy\right]_0^1 = \frac{6}{5}y + \frac{2}{5}, \quad 0 < y < 1$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_0^1 x\left(\frac{6}{5}x^2 + \frac{3}{5}\right)dx = \frac{6}{5}\left[\frac{1}{4}x^4 + \frac{1}{4}x^2\right]_0^1 = \frac{3}{5}$$

$$E(Y) = \int_{-\infty}^{\infty} yf(y)\,dy = \int_0^1 y\left(\frac{6}{5}y + \frac{2}{5}\right)dy = \frac{6}{5}\left[\frac{1}{3}y^3 + \frac{1}{6}y\right]_0^1 = \frac{3}{5}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2\left(\frac{6}{5}x^2 + \frac{3}{5}\right)dx = \frac{6}{5}\left[\frac{1}{5}x^5 + \frac{1}{6}x^3\right]_0^1 = \frac{11}{25}$$

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 f(y)\,dy = \int_0^1 y^2\left(\frac{6}{5}y + \frac{2}{5}\right)dy = \frac{6}{5}\left[\frac{1}{4}y^4 + \frac{1}{9}y^3\right]_0^1 = \frac{13}{30}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{2}{25}, \quad Var(Y) = E(Y^2) - [E(Y)]^2 = \frac{11}{150}$$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,f(x,y)\,dx\,dy = \int_0^1\int_0^1 xy\frac{6}{5}(x^2+y)\,dx\,dy = \frac{7}{20}$$

$$Cov(X, Y) = E(XY) - E(X)E(Y) = -0.01$$

$$\rho(X, Y) = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_y} = -0.1306$$

(4) $f(x)f(y) = \left(\frac{6}{5}x^2 + \frac{3}{5}\right)\left(\frac{6}{5}y + \frac{2}{5}\right) \neq f(x,y) = \frac{6}{5}(x^2+y)$이므로 두 확률변수 X, Y는 서로 독립이 아니다.

09 두 확률변수 X, Y의 결합확률분포가 다음과 같을 때, X, Y의 주변 확률분포함수를 구하라.

$$f_{X,Y}(x, y) = \frac{1}{2}, \quad 0 < y < x < 2$$

✎해답 $f(x) = \int_y f(x,y)\,dy = \int_0^x \frac{1}{2}\,dy = \frac{1}{2}x, \quad 0 < x < 2$

$$f(y) = \int_x f(x,y)\,dx = \int_y^2 \frac{1}{2}\,dx = 1 - \frac{1}{2}y, \quad 0 < y < 2$$

10. X와 Y가 이산형 확률변수일 때 다음이 성립함을 보여라.

(1) $E(3X) = 3E(X)$

(2) $E(2X-3) = 2E(X) - 3$

(3) $E(3X - 2Y) = 3E(X) - 2E(Y)$

(4) $Var(X - 2Y) = Var(X) + 4Var(Y) - 4Cov(X, Y)$

✎해답

(1) $E(3X) = \sum_x 3x\,P(X=x) = 3\sum_x x\,P(X=x) = 3E(X)$

(2) $E(2X-3) = \sum_x (2x-3)P(X=x)$

$$= 2\sum_x x\,P(X=x) - 3\sum_x P(X=x) = 2E(X) - 3$$

(3) $E(3X - 2Y) = \sum_x (3x - 2y)P(X=x, Y=y)$

$$= 3\sum_x x \sum_y P(X=x,\ Y=y) - 2\sum_y y \sum_x P(X=x,\ Y=y)$$

$$= 3\sum_x xP(X=x) - 2\sum_y yP(Y=y) = 3E(X) - 2E(Y)$$

(4) $Var(X-2Y) = E[(X-2Y)^2] - [E(X-2Y)]^2$

$$= E(X^2 - 4XY + 4Y^2) - [E(X) - 2E(Y)]^2 \ \ (\because\ 10-(3))$$

$$= E(X^2) - 4E(XY) + 4E(Y^2) - [E(X)]^2 + 4E(X)E(Y) - 4[E(Y)]^2$$

$$Var(X) + 4Var(Y) - 4Cov(X, Y)$$

11. 빨간색 공(R) 3개, 파란색 공(B) 4개, 흰색 공(W)이 5개 들어 있는 항아리에서 임의로 두 개의 공을 선택한다. 다음 물음에 답하라.

(1) 확률변수 X를 뽑힌 공 가운데 빨간색 공의 수, 확률변수 Y를 뽑힌 공 가운데 흰색 공의 수라고 할 때, 확률변수 X와 Y의 결합확률분포를 구하라.

(2) (1)의 결과를 이용하여, 확률변수 X와 확률변수 Y의 주변확률분포를 구하라.

(3) $P(X+Y=2)$를 구하라.

(4) $E(X)$, $E(Y)$를 구하라.

(5) 확률변수 X와 Y가 독립인지 확인하라.

✎ **해답**

(1)

X \ Y	0	1	2
0	$\frac{12}{132}$	$\frac{40}{132}$	$\frac{20}{132}$
1	$\frac{24}{132}$	$\frac{30}{132}$	0
2	$\frac{6}{132}$	0	0

(2)

X	0	1	2
$P(X=x)$	$\frac{72}{132}$	$\frac{54}{132}$	$\frac{6}{132}$

(3)

Y	0	1	2
$P(Y=y)$	$\frac{42}{132}$	$\frac{70}{132}$	$\frac{20}{132}$

(4) $E(X) = (1\times\frac{54}{132}) + (2\times\frac{6}{132}) = \frac{1}{2}$, $\quad E(Y) = (1\times\frac{70}{132}) + (2\times\frac{20}{132}) = \frac{5}{6}$

(5) $P(X=0)P(Y=0)=\dfrac{21}{121} \neq P(X=0,\ Y=0)=\dfrac{12}{132}$ 이므로 두 확률변수 X, Y는 서로 독립이 아니다.

12. 확률변수 X와 Y의 결합분포가 다음과 같다.

X \ Y	-1	0	1
0	1/3	0	1/3
1	0	1/3	0

(1) $Corr(X,\ Y)$를 구하라.

(2) 두 확률변수 X와 Y는 서로 독립인가?

✎**해답**

(1)

X	0	1
$P(X=x)$	$\frac{2}{3}$	$\frac{1}{3}$

Y	0	1	2
$P(Y=y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

XY	-1	0	2
확률	0	1	0

$E(X)=\dfrac{1}{3},\quad E(Y)=0,\quad E(X^2)=\dfrac{1}{3},\quad E(Y^2)=\dfrac{2}{3},\quad E(XY)=0$

$Var(X)=\dfrac{2}{9},\quad Var(X)=\dfrac{2}{3},\quad Cov(X,\ Y)=E(XY)-E(X)E(Y)=-\dfrac{2}{9}$

$\rho(X,\ Y)=Corr(X,Y)=\dfrac{Cov(X,\ Y)}{\sigma_x\sigma_y}=-0.57735$

(2) $P(X=0)P(Y=0)=\dfrac{2}{9} \neq P(X=0,\ Y=0)=0$이므로 두 확률변수 X, Y는 서로 독립이 아니다.

01 이항분포 $B(n,\ p)$를 따르는 확률변수 X에 대하여 n과 p가 다음과 같을 때, 관측값 x에 대한 확률값을 계산하라. 또한 필요한 경우 누적 이항 확률분포표를 사용하라.

(1) $n=10,\ p=0.6,\ x=2$ (2) $n=5,\ p=0.2,\ x=3$

(3) $n=20,\ p=0.5,\ x=10$ (4) $n=100,\ p=0.8$

✎ **해답** (1) $P(X=2)=\binom{10}{2}(0.6)^2(0.4)^8=0.011$

(2) $P(X=3)=\binom{5}{3}(0.2)^3(0.8)^2=0.0512$

(3) $P(X=10)=\binom{20}{10}(0.5)^{20}=0.176$

(4) $P(X=x)=\binom{100}{x}(0.8)^x(0.2)^{100-x},\quad x=0,\ 1,\ \cdots,\ 100$

02 이항분포 $B(n,p)$에서 n과 p가 다음과 같을 때, 평균과 분산을 구하시오.

(1) $n=1000,\ p=0.3$ (2) $n=500,\ p=0.01$

(3) $n=300,\ p=0.5$ (4) $n=200,\ p=0.8$

✎ **해답** (1) $E(X)=np=300,\quad Var(X)=np(1-p)=210$

(2) $E(X)=np=5,\quad Var(X)=np(1-p)=4.95$

(3) $E(X)=np=150,\quad Var(X)=np(1-p)=75$

(4) $E(X)=np=160,\quad Var(X)=np(1-p)=32$

03 두루미가 낳은 알이 정상적으로 부화되는 확률은 0.2이다. 두루미가 낳은 알의 부화는 서로 독립이고 정상적으로 부화하는 것을 '성공'이라 하자. 20개의 두루미 알에 대하여 다음에 답하라.

(1) 정상적으로 부화되는 알의 수 X의 분포를 구하라.

(2) X의 평균과 분산을 구하라.

(3) 20개의 알 가운데 10개 이상이 정상적으로 부화될 확률은?

✎ **해답** (1) $X\sim B(20,0.2)$

$P(X=x)=\binom{20}{x}(0.2)^x(0.8)^{20-x},\quad x=0,\ 1,\ 2,\ \cdots,\ 20$

(2) $E(X)=np=4,\quad Var(X)=np(1-p)=3.2$

(3) $P(X \geq 10) = 1 - P(X \leq 9) = 1 - \sum_{x=0}^{9} \binom{20}{x} (0.2)^x (0.8)^{20-x} \approx 0.00259$

04 위궤양의 치료에 80%의 완치율을 보이는 약이 있다. 위궤양을 앓고 있는 20명의 환자에게 이 약을 투여하려고 한다.

(1) 20명 중 완치된 사람 수를 X라고 할 때, X의 분포는?

(2) 한 명도 완치되지 않을 확률과 모두 완치될 확률을 구하라.

✎해답 (1) $X \sim B(20, 0.8)$

$$P(X = x) = \binom{20}{x} (0.8)^x (0.2)^{20-x}, \quad x = 0, 1, 2, \cdots, 20$$

(2) $P(X = 0) = \binom{20}{0} (0.2)^{20} = (0.2)^{20}$

$$P(X = 20) = \binom{20}{20} (0.8)^{20} = (0.8)^{20}$$

05 경부고속도로에서 발생하는 교통사고의 수는 하루 평균 20건이라고 한다.

(1) 확률변수 X를 경부고속도로에서 발생하는 교통사고의 수라고 정의할 때, X의 확률함수를 제시하라.

(2) (1)에서 제시한 확률함수를 이용하여 $P(X=0)$, $P(X=3)$, $P(X=20)$을 구하라.

✎해답

(1) $X \sim Poisson(20)$

$$P(X = x) = \frac{e^{-20} 20^x}{x!}, \quad x = 0, 1, 2, \cdots$$

(2) $P(X = 0) = \frac{e^{-20} 20^0}{0!} = e^{-20}$

$$P(X = 3) = \frac{e^{-20} 20^3}{3!}$$

$$P(X = 20) = \frac{e^{-20} 20^{20}}{20!}$$

06 어떤 가전제품의 각 부품이 정상적으로 작동할 확률이 $p\,(0 \leq p \leq 1)$이고, 각 부품은 서로 독립적으로 작동된다. 가전제품의 부품 중에서 적어도 50% 이상이 정상적으로 작동되면 가전제품이 사용가능하다고 할 때, 다섯 개 부품으로 구성된 제품이 세 개의 부품으로 구성된 제품보다 사용가능할 확률이 더 큰 p 값의 범위를 구하라.

✎해답

$$\sum_{x=3}^{5}\binom{5}{x}p^{x}(1-p)^{5-x} > \sum_{x=2}^{3}\binom{3}{x}p^{x}(1-p)^{3-x}$$

$$10p-20p^2+10p^3+5p^2-5p^3+p^3-3+3p-p > 0$$

$$2p^3-5p^2+4p-1 > 0$$

$$\therefore\ \frac{1}{2} < p < 1$$

07 서울 시청의 인터넷 홈페이지에는 1시간에 평균 120명이 접속한다.

(1) 10분 동안에 접속한 사람이 한 명도 없을 확률을 구하라.

(2) 30분 동안에 60명 이상이 홈페이지에 접속할 확률을 구하라.

✎해답

(1) 시간당 평균 120명이므로 10분당 평균 20명이다.

$X \sim Poisson(20)$

$$P(X=0) = \frac{e^{-20}\,20^0}{0!} = e^{-20}$$

(2) 시간당 평균 120명이므로 30분당 평균 60명이다.

$Y \sim Poisson(60)$

$$P(X \ge 60) = 1-P(X \le 59) = 1-\sum_{x=0}^{59}\frac{e^{-60}\,60^x}{x!} \approx 0.5172$$

08 전단지를 제작하는 회사의 불량률이 0.01이라고 한다. 500장의 전단지를 제작하는 경우 제작된 전단지의 불량품 수를 X라 할 때, 다음의 물음에 답하라.

(1) $P(X \le 2)$의 확률을 이항분포를 이용하여 계산하라.

(2) $\Pr(X \le 2)$의 확률을 포아송 근사를 이용하여 계산하라.

(3) (1)과 (2)를 비교하여, 포아송 근사의 정확성을 확인하라.

✎해답

(1) $X \sim B(500, 0.01)$

$$P(X=x) = \binom{50}{x}(0.01)^x(0.99)^{50-x}, \quad x=0,1,2,\cdots,50$$

$$P(X \le 2) = P(X=0)+P(X=1)+P(X=2) = 0.1234$$

(2) $X \sim Poisson(5)$

$$P(X=x) = \frac{e^{-5}\,5^x}{x!}, \quad x=0,1,2,\cdots$$

$$P(X \le 2) = P(X=0) + P(X=1) + P(X=2) = 0.1247$$

(3) (1)의 정확한 확률값과 (2)의 근사함 확률값이 거의 같음을 확인할 수 있다.

09 $n = 100$일 때 다음의 p 값에 대한 이항분포의 평균과 표준편차를 계산하여 아래 표를 완성하고 표준편차를 가장 크게 하는 p를 결정하라.

p	0.1	0.3	0.5	0.7	0.8	0.9
평균						
표준편차						

즉, 표준편차를 가장 크게 하는 p값은 0.5이다.

10 A출판사에서 출판하고 있는 책의 초안에는 페이지 당 평균 3개의 오타가 있다고 한다.

(1) 어느 페이지에 오타가 하나도 없을 확률을 구하라.

(2) 한 페이지 당 오타가 3개 미만일 확률을 구하라.

(3) 한 페이지 당 오타가 10개 이상 있을 확률을 구하라.

✎**해답** $X \sim Poisson(3)$

$$P(X=x) = \frac{e^{-3}3^x}{x!}, \quad x = 0, 1, 2, \cdots$$

(1) $P(X=0) = \dfrac{e^{-3}3^0}{0!} = e^{-3} = 0.0498$

(2) $P(X<3) = P(X \le 2) = P(X=0) + P(X=1) + P(X=2) = 0.4232$

(3) $P(X \ge 10) = 0.0011$

11 사격선수가 표적을 맞출 확률이 95%라고 한다.

(1) 이 선수가 사격을 시작하여 네 번째 사격 이내에 처음 표적을 맞출 확률은?

(2) 이 선수가 표적을 처음으로 맞출 때까지 필요한 사격횟수의 평균은?

✎**해답** $X \sim Geometric(0.95)$

$$P(X=x) = (0.95)(0.05)^{x-1}, \quad x = 1, 2, \cdots$$

(1) $P(X \le 4) = 0.9999937 \approx 1$

(2) $E(X) = \dfrac{1}{p} = \dfrac{1}{0.95} = 1.0526$

12 이항분포와 포아송 분포가 실제 적용되는 사례 또는 사용할 수 있는 사례를 본문의 설명에서 예로 들은 것과 다르게 제시하여 보라.

[예시] 이항분포: 주사위를 5회 던졌을 때 눈금의 개수가 1이 나오는 횟수
인구의 5%가 쌍커풀이 있다고 했을 때, 100명 중 쌍커풀을 가진 사람의 수
포아송분포: 어떤 지역의 하루 교통사고 건수
1시간 동안 백화점을 방문한 고객 수

6장 연습문제

풀이 | 및 | 해답

01 확률변수 Z가 표준정규분포를 따른다고 할 때, 다음을 만족하는 z값을 구하라.

(1) $P(Z < z) = 0.65$

(2) $P(-z < Z < z) = 0.90$

(3) $P(Z > z) = 0.001$

(4) $P(Z < -z) = 0.001$

해답 (1) $z = 0.385$

(2) $P(-z < Z < z) = 2P(0 < Z < z) = 0.90 \quad \therefore z = 1.645$

(3) $P(Z > z) = 1 - P(Z < z) = 0.001$

$P(Z < z) = 0.999 \quad \therefore z = 3.09$

(4) $z = 3.09$

02 어느 대학의 학생식당에서는 아침식사로 크림스프와 감자샐러드를 제공한다. 이 학생식당에서 제공하는 크림스프의 열량은 평균이 164$kcal$이고 표준편차는 5$kcal$인 정규분포를 따르고, 감자샐러드의 열량은 평균이 183$kcal$이고 표준편차는 4$kcal$인 정규분포를 따른다고 가정하자. 두 분포는 서로 독립이다.

(1) 이 학생식당에서 아침식사로 1인당 섭취하는 열량의 평균과 표준편차를 구하라.

(2) 이 학생식당에서 1회 제공하는 아침식사의 열량이 360$kcal$ 이상일 확률을 구하라.

해답 $X \sim N(164, 5^2), \quad Y \sim N(183, 4^2)$

(1) $X + Y \sim N(347, 6.4^2)$

(2) $P(X + Y \geq 360) = P(Z \geq \dfrac{360 - 347}{6.4}) = P(Z \geq 2.03) = 0.021$

03 생수회사에서 300ml 생산할 때, 양이 적은 경우 발생할 수 있는 고객 불만을 고려하여 310ml를 생수양의 기준으로 하여 생산한다고 한다. 이 회사에서 생산하는 생수가 평균이 310ml이고 표준편차가 3ml인 정규분포를 따른다고 가정하고 다음 물음에 답하라.

(1) 이 회사에서 생산된 생수가 300ml 이하일 확률을 구하라.

(2) 이 회사에서 생산된 생수가 300ml 이하일 확률이 0.03 이하가 되기 위한 최소한의 기준

을 결정하라.

✎해답 $X \sim N(310,\ 3^2)$

(1) $P(X \leq 300) = P(Z \leq \dfrac{300-310}{3}) = P(Z \leq -3.33) = 0.0004$

(2) $P(Z \leq \dfrac{300-a}{3}) \leq 0.03$

이 회사에서 생산된 생수가 $300ml$ 이하일 확률이 0.03 이하가 되기 위해서는 최소한 $305.6424\,ml$ 의 생수를 기준으로 생산해야 한다.

04 기초통계학 과목의 성적은 평균이 75, 그리고 표준편차가 5인 정규분포를 따른다고 한다. 다음 물음에 답하여라.

(1) 60점 이하가 F학점을 받는다고 할 때, F학점을 받는 학생의 비율은 얼마인가?

(2) 만일 50명이 수강한다면, F학점을 받는 학생은 몇 명인가?

✎해답 $X \sim N(75,\ 5^2)$

(1) $P(X \leq 60) = P(Z \leq \dfrac{60-75}{5}) = P(Z \leq -3) = 0.0013$

(2) $50 \times 0.0013 = 0.065$

약 1명 미만의 학생이 F학점을 받는다.

05 어느 공장에서 생산되는 제품의 무게를 X라 할 때, $X \sim N(20,\ 2^2)$을 따른다고 한다. 다음 물음에 답하라.

(1) 이 공장의 품질관리 부서에서는 제품의 무게가 10 이하이거나 26 이상이면 불량품으로 판정한다. 생산되는 제품 중에서 불량품의 비율은 어느 정도 되겠는가?

(2) 제품 500개를 한 상자에 넣는다고 할 때, 한 상자를 검사하면 몇 개의 불량품이 나타나겠는가?

✎해답 (1) $P(X \leq 10$ 또는 $X \geq 26) = P(Z \leq \dfrac{10-20}{2}$ 또는 $Z \geq \dfrac{26-20}{2})$

$= P(Z \leq -5$ 또는 $Z \geq 3) = 0.0014$

(2) $500 \times 0.0014 = 0.7$

약 1개 미만의 불량품이 나타난다.

06 확률변수 W가 χ^2_r 분포를 따를 때, 다음 확률을 계산하라.

(1) $r = 1,\ P(W \leq 0.5)$

(2) $r = 5,\ P(2.180 \leq W \leq 7.344)$

(3) $r = 20,\ P(30.144 \leq W \leq 36.191)$

(4) $r = 80$, $P(W \le 46.520) + \Pr(W \ge 124.839)$

✎해답

(1) 0.5205　　(2) 0.6274　　(3) 0.0529　　(4) 0.002

07 확률변수 W가 χ^2_{15} 분포를 따를 때, 다음을 만족하는 w를 구하라.

(1) $P(W > w) = 0.5$　　(2) $P(W > w) = 0.1$

(3) $P(W > w) = 0.05$　　(4) $P(W > w) = 0.01$

✎해답 (1) $w = 14.33886$　(2) $w = 22.30713$　(3) $w = 24.99579$

(4) $w = 30.57791$

08 확률변수 T가 t_n 분포를 따를 때, 다음 확률을 계산하라.

(1) $n = 1$, $P(T \le 31)$　　(2) $n = 12$, $P(T \ge 0.695)$

(3) $n = 40$, $P(T \le 1.96)$　　(4) $n = 160$, $P(1 \le T \le 1.845)$

✎해답 (1) 0.9897　(2) 0.2501　(3) 0.9715　(4) 0.126

09 확률변수 X가 $X \sim \exp(\lambda)$일 때,

(1) 평균과 분산은?　　(2) X가 2 이상일 확률은?

(3) X가 5 이하일 확률은?

✎해답

(1) $E(X) = \dfrac{1}{\lambda}$, $VAR(X) = \dfrac{1}{\lambda^2}$

(2) 지수분포의 누적분포함수를 이용하면 다음과 같이 구할 수 있다.

$$P(X \ge 2) = 1 - P(X \le 2) = 1 - (1 - e^{-2\lambda}) = e^{-2\lambda}$$

(3) $P(X \le 5) = 1 - e^{-5\lambda}$

10 어떤 지역에서 교통사고가 평균 10분에 2건 발생한다고 한다.

(1) 교통사고가 발생한 이후, 다음 사고가 발생할 때까지 걸리는 평균 시간은?

(2) 교통사고가 발생한 후, 10분 동안 사고가 없을 확률은?

✎해답

X를 교통사고가 발생한 후 다음사고가 발생할 때까지 걸리는 시간(단위시간:1분)이라고 하면, X는 λ가 0.2인 지수분포를 따른다.

(1) $E(X) = \dfrac{1}{\lambda} = 5$

따라서 교통사고가 발생한 이후, 다음 사고가 발생하는데 걸리는 시간은 5분이다.

(2) $P(X \geq 10) = e^{-2}$

11 확률변수 X가 $B(100,\ 0.4)$를 따를 때, 다음 확률을 계산하라.

(1) $P(X \leq 40)$ (2) $P(X \leq 80)$ (3) $P(20 \leq X \leq 50)$

✎**해답**

$n = 100$이 충분이 크고 $p = 0.4$이므로 이항분포의 정규근사와 연속석 수정을 이용하여 근사확률을 구한다. $np = 40,\ np(1-p) = 24$이므로, X는 근사적으로 $N(40,\ 4.9^2)$를 따른다.

(1) $P(X \leq 40) = P(Z \leq \dfrac{40.5-40}{4.9}) = P(Z \leq 0.1) = 0.54$

(2) $P(X \leq 80) = P(Z \leq \dfrac{80.5-40}{4.9}) = P(Z \leq 8.27) = 1$

(3) $P(20 \leq X \leq 50) = P(\dfrac{19.5-40}{4.9} \leq Z \leq \dfrac{50.5-40}{4.9})$

$= P(-4.18 \leq Z \leq 2.14) = 0.984$

7장 연습문제 풀이 및 해답

01 정규분포 $N(\mu,\ \sigma^2)$를 따르는 모집단으로부터 크기 25인 확률표본을 추출하였다. 다음 확률을 구하라.

(1) $\sigma = 5$일 때 $P(-2 \leq \overline{X} - \mu \leq 2)$

(2) $\sigma = 10$일 때 $P(-2 \leq \overline{X} - \mu \leq 2)$

✎**해답** (1) $P(-2 \leq \overline{X} - \mu \leq 2) = P\left(-2 \leq \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq 2\right)$

$= P(-2 \leq Z \leq 2) = 0.9772 - 0.0228 = 0.9544$

(2) $P(-2 \leq \overline{X} - \mu \leq 2) = P\left(-1 \leq \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq 1\right)$

$= P(-1 \leq Z \leq 1) = 0.8413 - 0.1587 = 0.6826$

02 정규분포 $N(30,\ 4)$를 따르는 모집단으로부터 크기 36인 확률표본을 추출하였다. 이때의 표본평균이 $\overline{X}$라고 할 때, 다음에 답하라.

(1) $P(\overline{X} \geq 30)$의 값은?

(2) $P(\overline{X} \geq a) = 0.05$를 만족하는 a의 값은?

(3) $P(-b \leq \overline{X} - 30 \leq b) = 0.95$를 만족하는 b의 값은?

✎해답 (1) $P(\overline{X} \geq 30) = P\left(\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \geq \dfrac{30-30}{2/\sqrt{36}}\right)$

$= P(Z \geq 0) = 0.5$

(2) $P(\overline{X} \geq a) = P\left(Z \geq \dfrac{a-30}{0.33}\right) = 0.05$, $\dfrac{a-30}{0.33} = 1.645$

$a = 30 + 0.548 = 30.548$

(3) $P(-b \leq \overline{X} - 30 \leq b) = 0.95 = P\left(\dfrac{-b}{0.33} \leq \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \dfrac{b}{0.33}\right)$

$= P\left(|Z| \leq \dfrac{b}{0.33}\right) = 0.95$

$\dfrac{b}{0.33} = 1.96$, $a = 0.6468$

03 정규분포 $N(0, \sigma^2)$를 따르는 모집단으로부터 크기가 36인 표본을 추출하였다. 다음 물음에 답하라.

(1) $P(-1.96 \leq \overline{X} \leq 1.96) = 0.95$를 만족하는 모집단의 분산은?

(2) (1)을 만족하는 표준편차로부터 $P(-a \leq \overline{X} \leq a) = 0.90$을 만족하는 a의 값은?

✎해답 (1) $P(-1.96 \leq \overline{X} \leq 1.96) = P\left(\dfrac{-1.96-0}{\sigma/6} \leq \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \dfrac{1.96-0}{\sigma/6}\right)$

$= P\left(|Z| \leq \dfrac{1.96}{\sigma/6}\right) = 0.95$

$\dfrac{1.96}{\sigma/6} = 1.96$, $\sigma = 6$, $\sigma^2 = 36$

(2) $P(-a \leq \overline{X} \leq a) = P\left(\dfrac{-a}{1} \leq \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \dfrac{a}{1}\right) = P(|Z| \leq a) = 0.90$

$a = 1.645$

04 비행기 승객 1인당 평균 짐의 무게가 20kg이고, 표준편차는 5kg이라고 한다. 어느 날 임의로 추출한 36명의 승객들의 평균 짐의 무게가 23kg 이상일 확률은 얼마인가?

✎해답 $P(\overline{X} \geq 23) = P\left(Z \geq \dfrac{23-20}{5/6}\right) = P(Z \geq 3.6) = 0.000612$

05 다음 각 경우에 대하여 모평균 μ의 95% 신뢰구간을 구하라. 단, $\sigma = 1$이다.

(1) $n = 36$, $\overline{X} = 15$　　(2) $n = 64$, $\overline{X} = 20$

(3) (1)과 (2)에서 구한 모평균 μ의 95% 신뢰구간의 의미를 설명하라.

✎해답 (1) $0.95 = P(-z_{0.05/2} \le Z \le z_{0.05/2})$

$$= P\left(-z_{0.05/2} \le \frac{15-\mu}{1/\sqrt{36}} \le z_{0.05/2}\right)$$

$$= P\left(15 - 1.96 \times \frac{1}{\sqrt{36}} \le \mu \le 15 + 1.96 \times \frac{1}{\sqrt{36}}\right)$$

$\therefore$ 95% 신뢰구간은 $(14.67, 15.33)$이다.

(2) 95% 신뢰구간: $(20 \pm 1.96\frac{1}{8}) = (19.76, 20.25)$

(3) 동일한 모집단에서 동일한 크기로 표본을 100개 추출했을 때, 얻어진 100개의 신뢰구간 중에서 약 95개가 μ의 참값을 포함하고 있다고 예상되는 공식에 따라 구한 구간을 의미한다.

06 다음 각 경우에 대하여 모비율 p의 95% 근사 신뢰구간을 구하라.

(1) $n = 25$, $\hat{p} = 0.6$

(2) $n = 36$, $\hat{p} = 0.5$

(3) $n = 36$, $\hat{p} = 0.1$

(4) $n = 49$, $\hat{p} = 0.8$

✎해답 (1) $\left(0.6 - 1.96\sqrt{\frac{0.6 \times 0.4}{25}},\ 0.6 + 1.96\sqrt{\frac{0.6 \times 0.4}{25}}\right) = (0.41, 0.79)$

(2) $\left(0.5 - 1.96\sqrt{\frac{0.5 \times 0.5}{36}},\ 0.5 + 1.96\sqrt{\frac{0.5 \times 0.5}{36}}\right) = (0.34, 0.66)$

(3) $\left(0.1 - 1.96\sqrt{\frac{0.1 \times 0.9}{36}},\ 0.1 + 1.96\sqrt{\frac{0.1 \times 0.9}{36}}\right) = (0.002, 0.198)$

(4) $\left(0.8 - 1.96\sqrt{\frac{0.8 \times 0.2}{49}},\ 0.8 + 1.96\sqrt{\frac{0.8 \times 0.2}{49}}\right) = (0.69, 0.91)$

07 기초통계학 수강생 36명의 기말고사 점수를 조사해보니 평균이 70점이고 표준편차가 5점이었다. 기말고사 점수가 정규분포를 따른다고 할 때, σ^2에 대한 99% 신뢰구간을 구하라.

✎해답 $\left(\frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}},\ \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right) = \left(\frac{35 \times 25}{60.27},\ \frac{35 \times 25}{17.19}\right) = (14.52, 50.90)$

08 국회의원 후보자가 투표일에 앞서 자신의 지지율을 알아보기 위하여 선거구민 300명을 대상으로 조사한 결과 90명이 지지한다고 응답하였다. 이 후보자에 대한 지지율의 95% 신뢰구간을 구하라.

✎**해답** 비율의 추정치인 표본비율의 값은 $\hat{p} = \dfrac{90}{300} = 0.3$이다. 표본 수가 크므로 이 후보의 지지율에 대한 95% 신뢰구간은 정규 근사를 통해 임계치 $z_{0.025} = 1.96$ 을 사용하면 다음과 같다.

$$\left(0.3 - 1.96\sqrt{\frac{0.3\times 0.7}{300}},\ 0.3 + 1.96\sqrt{\frac{0.3\times 0.7}{300}}\right) = (0.25,\ 0.35)$$

09 모평균에 대한 95% 신뢰구간을 구할 때, 추정오차한계가 2를 넘지 않도록 다음 각 경우에 대한 표본크기를 결정하라.

(1) $\sigma = 5$인 경우

(2) σ가 알려져 있지 않고 사전조사에 의해 구한 표본표준편차가 7인 경우

✎**해답** (1) $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} \le d$, 즉 $n \ge \left(z_{\alpha/2}\dfrac{\sigma}{d}\right)^2 = (1.96\dfrac{5}{2})^2 = 24.01 \quad \therefore n = 25$

(2) $n \ge \left(z_{\alpha/2}\dfrac{S}{d}\right)^2 = (1.96\dfrac{7}{2})^2 = 47.06 \quad \therefore n = 48$

10 어떤 고교의 흡연율을 추정하고자 한다. 이때 신뢰수준 95%하에서 오차한계가 0.05 이하가 되기 위한 최소 표본의 크기를 결정하라. 단, 근처 고교의 예로 볼 때 흡연율은 0.3 정도가 될 것으로 예상된다.

✎**해답** $n \ge p^*(1-p^*)\left(\dfrac{z_{\alpha/2}}{d}\right)^2 = 0.3\times 0.7\left(\dfrac{1.96}{0.05}\right)^2 = 322.69$

즉, 필요한 표본의 크기는 약 323명이다.

11 성인 남자의 고혈압 환자 비율에 대한 조사를 하고자 한다. 신뢰수준 95%하에서 오차한계는 0.02를 넘지 않기 위해서, 다음 각 경우에 대해 얼마나 많은 표본을 추출해야 하는지를 결정하라.

(1) 실제 고혈압 보유 비율을 알 수 없는 경우

(2) 예년의 조사에 따라 고혈압 환자 비율이 0.07이라고 판단되는 경우

(3) (1)과 (2)의 결과에 따라 어떤 결론을 내릴 수 있는가?

✎**해답** (1) $n \ge \dfrac{1}{4}\left(\dfrac{z_{\alpha/2}}{d}\right)^2 = \dfrac{1}{4}\left(\dfrac{1.96}{0.02}\right)^2 = 2401$

(2) $n \ge p^*(1-p^*)\left(\dfrac{z_{\alpha/2}}{d}\right)^2 = 0.07\times 0.93\left(\dfrac{1.96}{0.02}\right)^2 = 625.22$

즉, 필요한 표본의 크기는 약 626명이다.

(3) 따라서, 아무 정보가 없을 때 더 많은 표본이 필요한 것을 알 수 있다.

01 정규 모집단으로부터 추출된 36개의 표본으로부터 표본평균 70, 표본표준편차 3.5를 얻었다. 다음 물음에 답하라,

(1) 모평균 μ가 75보다 작은지를 검정하고자 한다. 가설을 설정하라.

(2) 영가설하에서의 검정통계량 T_0를 계산하라.

(3) 유의수준 0.05에서 검정하라.

(4) 유의수준 0.01에서 검정하라.

✎해답 (1) $H_0 : \mu = 75$, $H_1 : \mu < 75$

(2) $T_0 = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{70 - 75}{3.5 / \sqrt{36}} = -8.571$

(3) $C_{0.05} = \{ T_0 < z_{0.05} = -1.645 \}$

$T_0 \in C_{0.05}$가 되어 H_0를 기각한다. 즉, $\alpha = 0.05$일 때 모평균이 75보다 작다고 할 수 있다.

(4) $C_{0.01} = \{ T_0 < z_{0.01} = -2.33 \}$

$T_0 \in C_{0.01}$가 되어 H_0를 기각한다. 즉, $\alpha = 0.01$일 때 모평균이 75보다 작다고 할 수 있다.

02 다음 가설과 정보를 사용하여 유의수준은 5%에서 가설검정을 실시하라.

(1) $H_0 : \mu = 50 \quad H_1 : \mu > 50$, 표본 36개, 표본평균 52, $\sigma^2 = 1$

(2) $H_0 : \mu = 30 \quad H_1 : \mu \neq 30$, 표본 100개, 표본평균 40, $\sigma^2 = 25$

(3) $H_0 : \mu = 10 \quad H_1 : \mu < 10$, 표본 64개, 표본평균 8, $S^2 = 10$

✎해답 (1) $T_0 = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{52 - 50}{1 / \sqrt{36}} = 12$

$C_{0.05} = \{ T_0 > z_{0.05} = 1.645 \}$

$T_0 \in C_{0.05}$가 되어 H_0를 기각한다. 즉, $\alpha = 0.05$일 때 모평균이 50보다 크다고 할 수 있다.

(2) $T_0 = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{40 - 30}{5 / \sqrt{100}} = 20$

$C_{0.05} = \{ |T_0| > z_{0.025} = 1.96 \}$

$T_0 \not\in C_{0.05}$가 되어 H_0 를 기각한다. 즉, $\alpha = 0.05$ 일 때 모평균이 30 이 아니라고 할 수 있다.

(3) $T_0 = \dfrac{\overline{x} - \mu_0}{S/\sqrt{n}} = \dfrac{8-10}{\sqrt{10}/\sqrt{64}} = -5.06$

$n = 64$ 이므로 정규분포를 이용

$C_{0.05} = \{\ T_0 < z_{0.05} = -1.645\ \}$

$T_0 \in C_{0.05}$ 가 되어 H_0 를 기각한다. 즉, $\alpha = 0.05$ 일 때 모평균이 10 보다 작다고 할 수 있다.

03 베르누이 분포로부터 200개의 표본을 추출하여, 성공 71, 실패 129를 얻었다.

(1) 성공의 확률이 30% 보다 큰가를 알아보고자 한다. 가설을 설정하라.

(2) 영가설하에서의 검정통계량 값을 계산하라.

(3) 유의수준 5%에서 검정하고 결론을 설명하라.

✎해답 (1) $H_0 : p = 0.3$, $H_1 : p > 0.3$

(2) $T_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{0.355 - 0.3}{\sqrt{\dfrac{0.3(1-0.3)}{200}}} = 1.697$

(3) $C_{0.05} = \{\ T_0 > z_{0.05} = 1.645\} \Rightarrow T_0 > z_{0.05}$이면 H_0 를 기각

$T_0 \not\in C_{0.05}$가 되어 영가설을 채택한다. 즉, 유의수준 $\alpha = 0.05$ 일 때 모비율이 0.3이다.

04 정규 모집단으로부터 추출된 36개의 표본으로부터 표본분산의 값이 18.4이다. 모분산 σ^2이 16보다 크다고 할 수 있는지 유의수준 0.05에서 검정하라.

✎해답 $H_0 : \sigma^2 = 16$ $H_1 : \sigma > 16$

영가설하에서의 검정통계량 T_0 의 값은 다음과 같다.

$$T_0 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(36-1)18.4}{16} = 40.25$$

이 경우는 단측 검정이고 유의수준 $\alpha = 0.05$ 일 때 자유도가 35 인 카이제곱분포의 임계치는 다음과 같으므로

$$\chi^2_{35,0.05} = 49.802$$

유의수준 $\alpha = 0.05$ 인 기각역은 다음과 같다.

$$C_{0.05} = \{\ T_0 > 49.802\ \} \Rightarrow T_0 > z_{0.05}\text{이면 } H_0 \text{ 를 기각}$$

즉, $T_0 \not\in C_{0.05}$ 가 되어 영가설을 채택한다. 즉, 유의수준 $\alpha = 0.05$ 일 때 모분산이 16 이

라고 할 수 있다.

05 다음은 정규모집단으로부터 추출된 자료이다. $H_0 : \sigma^2 = 10$, $H_1 : \sigma^2 \neq 10$을 유의수준 5% 하에서 검정하라.

13	18	10	13	15	18	28	12	15	19	18	19	22	30	27
14	21	22	14	18	12	19	21	12	32	21	19	14	15	19

✎**해답** $H_0 : \sigma^2 = 10 \quad H_1 : \sigma^2 \neq 10$

$S^2 = 30.3$

영가설하에서의 검정통계량 T_0 의 값은 다음과 같다.

$$T_0 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(30-1)30.3}{10} = 87.87$$

이 경우는 양측 검정이고 유의수준 $\alpha = 0.05$일 때 자유도가 9인 카이제곱분포의 임계치는 다음과 같으므로

$$\chi^2_{29,\,0.975} = 16.047,\ \chi^2_{29,0.025} = 45.722$$

유의수준 $\alpha = 0.05$인 기각역은 다음과 같다.

$$C_{0.05} = \{T_0 < 16.047 \text{ 또는 } T_0 > 45.722\}$$

즉, $T_0 \in C_{0.05}$가 되어 영가설을 기각한다. 즉, 유의수준 $\alpha = 0.05$일 때 모분산은 10이 아니다.

06 신용카드 결제율에 따른 고객신뢰도를 조사하고자 600명을 조사하였는데, 이 중 100명은 결제일에 정상적으로 결제를 하지 못하는 것으로 나타났다. 불량 신용률이 20% 이하인지를 알고 싶을 때, 유의수준 0.05하에서 검정하라.

✎**해답** $H_0 : p = 0.2$, $\quad H_1 : p < 0.2$

$\hat{p} = \frac{100}{600} = 0.167$

또한, 영가설하에서의 검정통계량 T_0 의 값은 다음과 같다.

$$T_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.167 - 0.2}{\sqrt{\frac{0.2(1-0.2)}{600}}} = -2.021$$

이 경우는 좌측 단측 검정이고 유의수준 $\alpha = 0.05$ 일 때 기각역은 다음과 같으므로

$$C_{0.05} = \{T_0 < z_{0.05} = -1.645\}$$

$T_0 \in C_{0.05}$가 되어 영가설을 기각한다. 즉, 유의수준 $\alpha = 0.05$일 때 모비율이 0.2보다 작다고 할 수 있다.

07 지난해에 조사한 어느 대학교 남학생의 흡연율이 전체의 30%에 달하는 것으로 나타났다. 금연교육을 실시한 후 올해 전체 남학생 가운데 1000명의 흡연율을 다시 조사하였는데, 250명이 흡연을 한다고 응답하였다. 흡연율이 감소하였다고 할 수 있는지 유의수준 0.05하에서 검정하라.

✎**해답** $H_0 : p = 0.3\,,\quad H_1 : p < 0.3$

$\hat{p} = \dfrac{250}{1000} = 0.25$

또한, 영가설하에서의 검정통계량 T_0 의 값은 다음과 같다.

$$T_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.25 - 0.3}{\sqrt{\dfrac{0.3(1-0.3)}{1000}}} = -3.45$$

이 경우는 좌측 단측 검정이고 유의수준 $\alpha = 0.05$ 일 때 기각역은 다음과 같으므로

$$C_{0.05} = \{\ T_0 < z_{0.05} = -1.645\ \}$$

$T_0 \in C_{0.05}$가 되어 영가설을 기각한다. 즉, 유의수준 $\alpha = 0.05$ 일 때 흡연율이 감소하였다고 할 수 있다.

08 다음 결과를 이용하여 $H_0 : \mu_X - \mu_Y = 3\,,\ H_1 : \mu_X - \mu_Y \neq 3$을 검정하려 한다.

표본	n	표본평균	표본분산
X	30	74.8	4.6
Y	31	68.9	5.6

(1) 분산의 동일성을 유의수준 5%하에서 검정하라.

(2) (1)의 결과에 따라, 주어진 가설을 유의수준 5%하에서 검정하고 그 결론을 도출하라.

✎**해답**

(1) 영가설 $H_0 : \sigma_X^2 = \sigma_Y^2$과 대립가설 $H_1 : \sigma_X^2 \neq \sigma_Y^2$에 대하여 H_0 하에서의 검정통계량의 관측값은 다음과 같다.

$$T_0 = \frac{s_X^2}{s_Y^2} = \frac{4.6}{5.6} = 0.821 \sim F_{29,30}$$

부록의 F-분포표로부터 $F_{29,30,0.025} = 2.08$, $F_{29,30,0.975} = 0.48$이다. 따라서 영가설을 채택하게 되고, 두 모분산은 동일하다고 결론을 내릴 수 있다.

(2) $H_0 : \mu_X - \mu_Y = 3 \quad H_1 : \mu_X - \mu_Y \neq 3$

공통분산 σ^2의 추정량 s_p^2 을 구하면 다음과 같다.

$$s_p^2 = \frac{(m-1)s_X^2 + (n-1)s_Y^2}{m+n-2} = \frac{(30-1)\times 4.6 + (31-1)\times 5.6}{30+31-2}$$

$$T_0 = \frac{(\overline{x}-\overline{y}) - (\mu_X - \mu_Y)}{s_p\sqrt{m^{-1}+n^{-1}}} = \frac{5.9-3}{\sqrt{5.108\left(\frac{1}{30}+\frac{1}{31}\right)}} = 5.010$$

유의수준 $\alpha = 0.05$하에서 기각역은 $C_{0.05} = \{\,|T_0| > z_{0.025} = 1.96\,\}$이므로 유의수준 5%에서 H_0 를 기각한다. 즉, 유의수준 $\alpha = 0.05$에서 두 모집단의 평균차이는 3이 아니라고 할 수 있다.

09 다음은 어느 대학에 수시모집(A)과 정시모집(B)으로 입학한 1학년생 가운데 일부를 뽑아 적성 적합도 테스트에 대한 자료이다.

집단	n	표본평균	표본표준편차
A	80	74점	3.2점
B	40	78점	3.8점

위의 결과로부터 적성도의 평균점수에 있어서 수시모집 입학생이 정시모집 입학생보다 낮다고 할 수 있는가를 유의수준 $\alpha = 0.05$로 검정하라. 단, 모분산은 동일하다고 가정한다.

✎ **해답** $H_0 : \mu_A = \mu_B,\ H_1 : \mu_A \neq \mu_B$

공통분산 σ^2 의 추정량 s_p^2 을 구하면 다음과 같다.

$$s_p^2 = \frac{(m-1)s_X^2 + (n-1)s_Y^2}{m+n-2} = \frac{(80-1)\times 3.2^2 + (40-1)\times 3.8^2}{80+40-2}$$

영가설하에서 검정통계량의 관측값은 다음과 같고

$$T_0 = \frac{\overline{x}-\overline{y}}{s_p\sqrt{\frac{1}{m}+\frac{1}{n}}} = \frac{74-78}{\sqrt{11.628\left(\frac{1}{80}+\frac{1}{40}\right)}} = -6.057$$

$n_A + n_B = 120$ 이므로 정규분포를 이용하며, 유의수준 $\alpha = 0.05$하에서 기각역은 $C_{0.05} = \{\,|T_0| > z_{0.025} = 1.96\,\}$이므로 유의수준 5%에서 H_0 를 기각한다. 즉, 유의수준 $\alpha = 0.05$에서 두 모집단의 적성도 평균점수는 같지 않다고 할 수 있다.

10. 환경영향 평가를 위해 A지역에서 임의 추출한 250세대 중 50세대와 B지역에서 임의 추출한 200세대 중 30세대가 수질 오염으로 고통을 겪고 있다고 답하였다. 이 사실로부터 두 지역 A와 B의 수질 오염 피해 정도에 차이가 있다고 할 수 있는가를 유의수준 5%로 검정하라.

✎해답 $H_0 : p_1 = p_2 \quad H_1 : p_1 \neq p_2$

$$\hat{p_1} = \frac{50}{250} = 0.2,\ \hat{p_2} = \frac{30}{200} = 0.15$$

$$\hat{p} = \frac{50+30}{250+200} = \frac{80}{450} = 0.178$$

$$T_0 = \frac{0.2-0.15}{\sqrt{0.178(1-0.178)(250^{-1}+200^{-1})}} = 1.378$$

유의수준 $\alpha = 0.05$하에서 기각역은 $C_{0.05} = \{|T_0| > z_{0.025} = 1.96\}$이므로 유의수준 5%에서 H_0를 채택한다. 즉, 유의수준 $\alpha = 0.05$ 에서 두 모집단의 수질 오염 피해 정도는 같다고 할 수 있다.

11. 도지사 선거운동 기간 중에 어떤 입후보자가 지역에 따른 자신의 지지율을 파악하기 위하여 A지역에서 100명, B지역에서 150명을 조사한 결과, 각각 48명과 78명이 본인을 지지하는 것으로 조사되었다. 두 지역의 지지율이 같다고 할 수 있는지 유의수준 5%에서 검정하라.

✎해답 $H_0 : p_1 = p_2 \quad H_1 : p_1 \neq p_2$

$$\hat{p_1} = \frac{48}{100} = 0.48,\ \hat{p_2} = \frac{78}{150} = 0.52$$

$$\hat{p} = \frac{48+78}{100+150} = \frac{126}{250} = 0.504$$

$$T_0 = \frac{0.48-0.52}{\sqrt{0.504(1-0.504)(100^{-1}+150^{-1})}} = -0.62$$

유의수준 $\alpha = 0.05$하에서 기각역은 $C_{0.05} = \{|T_0| > z_{0.025} = 1.96\}$이므로 유의수준 5%에서 H_0를 채택한다. 즉, 유의수준 $\alpha = 0.05$에서 두 모집단의 지지율은 같다고 할 수 있다.

12 현재 알려져 있는 주장에 의하면 긴박한 상황에서 성인 남성의 반응시간은 0.85초라고 한다. 이 주장이 맞는지를 검정하고자 36명의 성인 남성을 대상으로 조사한 결과 표본평균은 $\bar{X} = 0.83$초가 나왔다. 모표준편차가 $\sigma = 0.05$초로 알려져 있을 경우 기존의 주장은 타당한가? 유의수준 5%로 검정하라.

✎해답 $H_0 : \mu = 0.85 \quad H_1 : \mu \neq 0.85$

$$T_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{0.83-0.85}{0.05/\sqrt{36}} = -2.4$$

$$C_{0.05} = \{|T_0| > z_{0.025} = 1.96\}$$

$T_0 \not\in C_{0.05}$가 되어 H_0를 기각한다. 즉, $\alpha = 0.05$일 때 모평균이 0.85가 아니라고 할 수

있다.

13. 36개의 대응표본으로부터 계산된 차이에 대한 표본평균과 분산은 각각 $\overline{d} = 2.1$, $s_D^2 = 4.5$ 이었다. 가설 $H_0 : \mu_d = 0$, $H_1 : \mu_d > 0$을 유의수준 5%하에서 검정하라.

✎해답 $T = \dfrac{2.1 - 0}{\sqrt{4.5}\ /\ \sqrt{36}} = 5.94 \sim t_{36-1}$

$n = 36$이므로 정규분포를 이용 $C_{0.05} = \{\, T_0 > z_{0.05} = 1.645 \,\}$

$T_0 \in C_{0.05}$가 되어 H_0를 기각한다. 즉, $\alpha = 0.05$일 때 대응표본으로부터 계산된 차이 μ_d는 0보다 크다고 할 수 있다.

9장 연습문제

풀이 | 및 | 해답

01 김장철을 맞이하여 세 지역(서울, 춘천, 청주)에서의 판매되는 국산 고춧가루 값이 같은지를 알아보기 위하여 각 지역에서 고춧가루 판매점 네 곳을 임의 추출하여 kg당 가격을 조사한 결과가 다음과 같다. (단위: 천원)

서울	춘천	청주
36	35	29
33	29	32
32	34	30
37	34	33

(1) 이 자료에 대한 분산분석 모형을 설정하고 분산분석표를 작성하라.

(2) 세 지역의 고춧가루 가격이 같은지 유의수준 5%에서 검정하라.

✎해답

(1)

원 인	제곱합	자유도	평균제곱	T_0
처 리	25	2	12.5	2.3
오 차	49	9	5.4	
총 합	74	11		

(2) H_0: 각 지역에서 고춧가루 값에 차이가 없다. 즉, $H_0 : \mu_1 = \mu_2 = \mu_3$

H_1: 각 지역에서 고춧가루 값에 차이가 있다.

F-분포표로부터 $F_{2,9,0.05} = 4.26$이고 영가설하에서 검정통계량의 값은 2.3이므로 유의수준 $\alpha = 0.05$하에서 영가설을 채택하고, 각 지역에서 고춧가루 값에 차이가 없다고 결

론을 내린다.

02 다음 일원배치 분산분석표에 대한 물음에 답하라.

원인	제곱합	자유도	평균제곱	F
A 오차	a 43.69	2 b	c d	e
총계	58.47	23		

(1) 주어진 분산분석표의 a ~ e 값을 구하라.

(2) 유의수준 5%하에서 요인 A의 수준 간에는 차이가 있다고 할 수 있는지 검정하라.

✎해답

(1)

원인	제곱합	자유도	평균제곱	F
A 오차	a=14.78 43.69	2 b=21	c=7.39 d=2.08	e=3.55
총계	58.47	23		

(2) H_0: A의 수준간에는 차이가 없다. 즉, $H_0 : \mu_1 = \mu_2 = \mu_3$

H_1: A의 수준간에는 차이가 있다.

F-분포표로부터 $F_{2,21,0.05} = 3.47$이고 영가설하에서 검정통계량의 값은 3.55이므로 유의수준 $\alpha = 0.05$하에서 영가설을 기각하고, A의 수준간에는 차이가 있다고 결론을 내린다.

03 아동의 순발력을 증진시키는 데 효과적인 방법을 찾기 위해 28명의 초등학교 1학년 학생을 세 그룹으로 나누어 서로 다른 운동 방법에 임의로 배치하고, 4주간 지정된 운동방법으로 훈련시킨 후, 운동 전 반응시간과 운동 후 반응시간을 측정하였다. 결과는 다음과 같다. (단위: 초)

	방법 1	방법 2	방법 3
표본 수	10	9	9
평균반응시간	−1.34	0.32	3.69

원인	제곱합	자유도	평균제곱	F
운동방법 오차	64.31 a	b c	d e	f
총계	402.33			

(1) 주어진 분산분석표의 a~f 값을 구하라.

(2) 운동방법간 차이가 있다고 할 수 있는가? 유의수준 5%에서 검정하라.

✎해답

(1)

원인	제곱합	자유도	평균제곱	F
운동방법 오차	64.31 a=338.02	b=2 c=25	d=32.155 e=13.521	f=2.38
총계	402.33	27		

(2) H_0: 운동방법간에는 차이가 없다. 즉, $H_0 : \mu_1 = \mu_2 = \mu_3$

H_1: 운동방법간에는 차이가 있다.

F-분포표로부터 $F_{2,25,0.05} = 3.39$이고 영가설하에서 검정통계량의 값은 2.38이므로 유의수준 $\alpha = 0.05$하에서 영가설을 기각하고, 운동방법간에는 차이가 있다고 결론을 내린다.

04 다음 일원배치 분산분석표를 이용하여 물음에 답하라.

원인	제곱합	자유도	평균제곱	F
A 오차	170.45 134.47	2 9	85.23 14.94	5.70
총계	304.92	11		

(1) 실험에서 관측된 총 관측값의 수는 몇 개인가?

(2) (1)에서의 관측값을 Y_{ij}라고 할 때, 적절한 분산분석모형을 설정하라.

(3) 위의 분산분석표에서 검정하고자 하는 가설을 제시하라.

(4) 오차항의 분산 σ^2의 추정값은 얼마인가?

(5) 결정계수 R^2은 얼마인가?

✎해답

(1) 12개

(2) $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$,

$$\epsilon_{ij} \sim iid\ N(0, \sigma^2),\ i = 1, 2, 3,\ \ j = 1, 2, 3, 4$$

(3) H_0: A의 수준간에는 차이가 없다. 즉, $H_0 : \mu_1 = \mu_2 = \mu_3$

H_1: A의 수준간에는 차이가 있다.

(4) 14.94 (5) $\dfrac{134.47}{304.92} = 0.44$

05 다음은 4가지 온도가 특정 제품의 강도에 미치는 영향을 조사하기 위해 시행된 실험의 결과이다. 다음 물음에 답하라.

반복 \ 온도	10	20	30	40	평균
1 2 3	13 8 9	15 11 13	8 12 7	11 15 10	11.75 11.5 9.75
평균	10	13	9	12	$\overline{Y}=11$

(1) 다음의 분산분석표를 완성하라.

요인	제곱합	자유도	평균제곱합
온도 오차	(　　　) (　　　)	(　　　) (　　　)	(　　　) (　　　)
총계	(　　　)	(　　　)	

(2) 유의수준 5%하에서 온도간의 차이가 있는지를 검정하라.

(3) 만일 차이가 있다면, 어떤 온도가 가장 적합한지 선택하라.

✎해답

(1)

요인	제곱합	자유도	평균제곱합	F
온도 오차	30 50	3 8	10 6.25	1.6
총계	80	11		

(2) H_0: 온도간에는 차이가 없다. 즉, $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: 온도간에는 차이가 있다.

F-분포표로부터 $F_{3,8,0.05} = 4.07$이고 영가설하에서 검정통계량의 값은 1.6이므로 준 $\alpha = 0.05$하에서 영가설을 채택하고, 온도간에는 차이가 없다고 결론을 내린다.

(3) 온도간에는 차이가 없다.

06 일주일 간 주문량(단위: 만 원)이 요일(월~금) 별로 차이가 있는지를 알아보고자 6개월 동안 임의로 선정된 요일 별로 다음의 데이터를 얻었다. 요일 별로 차이가 있는지를 유의수준 5%에서 검정하라.

반복 \ 요일	월	화	수	목	금
1	143	162	160	138	110
2	128	136	132	168	130
3	110	144	180	120	135
4		158	160		
5			138		
표본 크기	3	4	5	3	3
표본 총계	381	600	770	426	375
표본 평균	127	150	154	142	125

✎ **해답** H_0: 일주일 간 주문량은 요일(월-금) 별로 차이가 없다.

H_1: 일주일 간 주문량은 요일(월-금) 별로 차이가 있다.

요인	제곱합	자유도	평균제곱합	F
처리	2517	4	3.17	2.05
오차	4000	13	8.81	
총계	6517	17		

F-분포표로부터 $F_{4,13,0.05} = 3.18$이고 영가설하에서 검정통계량의 값은 2.05이므로 유의수준 $\alpha = 0.05$하에서 영가설을 채택하고, 일주일 간 주문량(단위 : 만원)이 요일(월-금)별로 차이가 없다고 결론을 내린다.

07 [R] 운동 후의 심장 박동수(단위: 번/분)가 연령대별로 차이가 있는지를 알아보기 위해 각 연령대별로 10명을 임의로 선정하여 각 개인의 운동 전 심장 박동수와 운동 후의 심장 박동수를 측정하여 그 차이를 기록하였다. 유의수준은 5%에서 차이가 있는지 검정하라.

ID	나이			
	10~19	20~29	30~39	40~49
1	29	24	37	28
2	33	27	25	29
3	26	33	22	34
4	27	31	33	36
5	39	21	28	21
6	35	28	26	20
7	33	24	30	25
8	29	34	34	24
9	36	21	27	33
10	22	32	33	32

✎해답

```
###7###[R]
y1<-c(29,33,26,27,39,35,33,29,36,22)
y2<-c(24,27,33,31,21,28,24,34,21,32)
y3<-c(37,25,22,33,28,26,30,34,27,33)
y4<-c(28,29,34,36,21,20,25,24,33,32)

y<-c(y1,y2,y3,y4)
y

n<-rep(10,4)
n
group<-rep(1:4,n)
group

group_df<-data.frame(y,group)
group_df

sapply(group_df, class)
# transform from 'integer' to 'factor'
group_df <- transform(group_df, group = factor(group))
sapply(group_df, class)

# one-wayANOVA
summary(aov(y ~ group, data = group_df))

> summary(aov(y ~ group, data = group_df))
            Df Sum Sq Mean Sq F value Pr(>F)
group        3   67.5   22.49   0.866  0.468
Residuals   36  935.5   25.99
```

분산분석결과 유의수준 0.05 하에서 p-value가 유의수준보다 크므로 영가설을 채택하고, 대립가설을 기각한다. 운동 후의 심장 박동수가 연령대별로 차이가 없다.

08 세 개의 서로 다른 처리(A, B, C)와 네 개의 블록으로부터 얻은 자료가 다음과 같다.

처리 \ 블록	1	2	3	4
A	3	5	8	6
B	8	6	9	5
C	7	4	7	8

(1) 분산분석표를 작성하라.

(2) 각 처리의 평균이 동일한지 유의수준 0.05에서 검정하라.

✎해답

(1)

요인	제곱합	자유도	평균제곱합	F
처리	4.67	2	2.34	0.13
블럭	14	3	4.67	0.26
오차	108.75	6	18.13	
총계	127.42	11		

(2) $H_0 : \mu_A = \mu_B = \mu_C$

H_1 : H_0가 아니다.

F-분포표로부터 $F_{2,6,0.05} = 5.14$, $F_{3,6,0.05} = 4.76$이고 영가설하에서 검정통계량의 값은 $0.13, 0.26$ 이므로 유의수준 $\alpha = 0.05$ 하에서 영가설을 채택하고, 처리집단의 평균들이 모두 동일하다고 볼 수 있다.

09 [R] 서로 다른 밀 종자 5개(A, B, C, D, E)의 수확량을 비교하기 위하여 다음과 같은 실험을 행하였다. 수확량은 토질, 기후 등의 영향을 받게 되므로 지역을 블록으로 지정하여 실험 대상이 되는 모든 종자를 동일한 지역에서 실험할 수 있도록 하였으며, 지역은 모두 6곳을 택하였다. 정해진 지역에서 5종류의 종자는 1평방미터를 한 구역으로 하여 랜덤하게 배치하였다. 실험의 결과는 다음과 같다. 유의수준은 5%에서 수확량의 차이가 있는지 검정하라.

	1	2	3	4
A	35.3	31.0	32.7	36.8
B	30.7	32.2	31.4	31.7
C	38.2	33.4	33.6	37.1
D	34.9	36.1	35.2	38.3
E	32.4	28.9	29.2	30.7

✎ **해답**

```
###9####[R]
A<-c(35.3,31.0,32.7,36.8,37.2,33.1)
B<-c(30.7,32.2,31.4,31.7,35.0,32.7)
C<-c(38.2,33.4,33.6,37.1,37.3,38.2)
D<-c(34.9,36.1,35.2,38.3,40.2,36.0)
E<-c(32.4,28.9,29.2,30.7,33.9,32.1)

y<-c(A,B,C,D,E)
y

n<-rep(6,5)
n
group<-rep(1:5,n)
group

group_df<-data.frame(y,group)
group_df

sapply(group_df, class)
# transform from 'integer' to 'factor'
group_df <- transform(group_df, group = factor(group))
sapply(group_df, class)

# one-wayANOVA
summary(aov(y ~ group, data = group_df))


> summary(aov(y ~ group, data = group_df))
            Df Sum Sq Mean Sq F value   Pr(>F)
group        4  142.7   35.67   8.375 0.000198 ***
Residuals   25  106.5    4.26
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

분산분석결과 유의수준 0.05 하에서 p-value가 유의수준보다 작으므로 영가설을 기각하고, 대립가설을 채택한다. 밀 종자에 따라 수확량의 차이가 있다.

01 독립변수 X와 종속변수 Y를 관측한 자료가 다음과 같다.

X	-3	-1	0	1	3
Y	1	2	4	4	5

(1) 위의 값에 대해 산점도를 그려보라.

(2) 단순 선형 회귀모형을 설정하고 최소제곱법을 통해 회귀계수의 추정치를 구하라.

(3) (1)에서 그린 산점도 위에 추정된 회귀선을 그려보라.

(4) 추정된 식에 따라 $X=2$인 경우, Y의 추정값은?

(5) 분산의 추정치인 $\widehat{\sigma^2}$과 결정계수 R^2을 구하라.

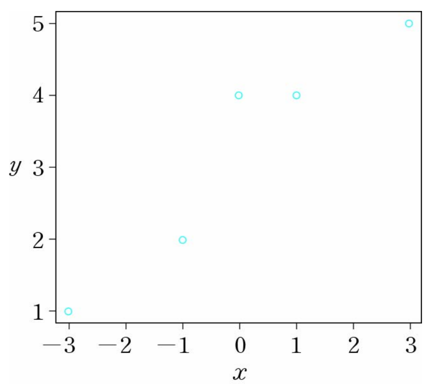
(6) 표본 상관계수 r을 구하고, 결정계수 R^2과 표본 상관계수 r의 관계를 설명하라.

(7) $H_0 : \beta_1 = 0$, $H_1 : \beta_1 \neq 0$에 대해 유의수준 5%하에서 검정하라.

(8) 분산분석표를 작성하고, 회귀모형의 유의성을 유의수준 5%에서 검정하라.

✎해답

(1)



(2) $Y=\beta_0+\beta_1 X+\epsilon$

$\overline{X} = 0$, $\overline{Y}=3.2$,

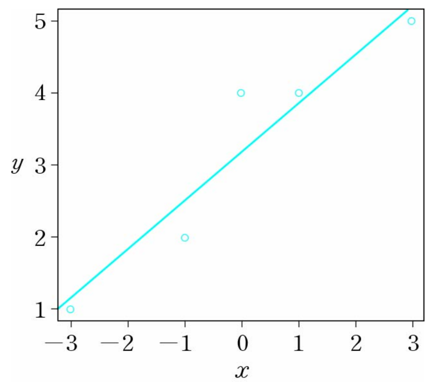
$S_{XX}=\sum_{i=1}^{5}(X_i-\overline{X})^2 = 20$, $\quad S_{YY}=\sum_{i=1}^{5}(Y_i-\overline{Y})^2 = 10.8$,

$S_{XY}=\sum_{i=1}^{5}(X_i-\overline{X})(Y_i-\overline{Y}) = 14$ 이므로

$$\widehat{\beta_1}=\frac{S_{XY}}{S_{XX}}=0.7, \quad \widehat{\beta_0}=\overline{Y}-\widehat{\beta_1}\,\overline{X}=3.2$$

따라서 추정된 단순 선형회귀 직선은 $\widehat{Y} = 3.2 + 0.7X$ 이다.

(3)



(4) $3.2 + 0.7 \times 2 = 4.6$

(5) $\widehat{\sigma^2} = \dfrac{1}{n-2}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 = \dfrac{1}{3} \times 1 = 0.33$

$SST = S_{YY} = 10.8$, $SSR = \dfrac{[S_{XY}]^2}{S_{XX}} = \dfrac{14^2}{20} = 9.8$

$R^2 = \dfrac{SSR}{SST} = 0.91$

(6) $Var(X) = 5$, $Var(Y) = 2.7$, $Cov(X, Y) = 7.2$

$r = \dfrac{Cov(x,y)}{\sqrt{Var(x) \cdot Var(y)}} = \dfrac{3.5}{\sqrt{5 \times 2.7}} = 0.95$

표본 상관계수 r 을 제곱한 것과 결정계수 R^2 는 동일하다.

(7) $T_0 = \dfrac{\widehat{\beta}_1}{\sqrt{MSE/S_{XX}}} = \dfrac{0.7}{\sqrt{0.33/14}} = 4.56$

검정통계량 $T = 4.56$ 이고 $T \sim t_{(3)}$ 이다. 자유도 3인 t분포에서 $P(|t| > 3.182) = 0.05$ 이며, $|T| = 4.56 > 3.182$ 이므로 귀무가설을 기각한다. 즉, β_1 이 0이라고 볼 수 없다.

(8)

원 인	제곱합	자유도	평균제곱	T_0
회 귀	9.8	1	9.8	$T_0 = \dfrac{9.8}{0.33} = 29.70$
잔 차	1	3	0.33	
총 합	10.8	4		

$T_0 = 29.70 > F_{1,3,0.05} = 10.13$ 이므로 영가설 $H_0 : \beta_0 = \beta_1 = 0$ 을 기각하고 추정된 회귀직선이 유의하다고 결론을 내릴 수 있다.

02 잔차를 e_i라고 할 때, 단순 선형 회귀모형에서 다음이 성립함을 증명하라.

(1) $\sum_{i=1}^{n} e_i = 0$ (2) $\sum_{i=1}^{n} X_i e_i = 0$ (3) $\sum_{i=1}^{n} Y_i e_i = 0$

✎ 해답

(1) $e_i \sim N(0,\, \sigma^2)$, $E(\sum_{i=1}^{n} e_i) = \sum_{i=1}^{n} E(e_i) = n \cdot 0 = 0$

(2) $e_i \sim N(0,\, \sigma^2)$, $E(\sum_{i=1}^{n} X_i e_i) = \sum_{i=1}^{n} X_i E(e_i) = \sum_{i=1}^{n} X_i \cdot 0 = 0$

(3) $e_i \sim N(0,\, \sigma^2)$, $E(\sum_{i=1}^{n} Y_i e_i) = \sum_{i=1}^{n} Y_i E(e_i) = \sum_{i=1}^{n} Y_i \cdot 0 = 0$

03 다음은 단순 선형 회귀모형 $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $i = 1, 2, \cdots, 30$의 표본으로부터 계산한 중간계산 결과이다. 다음 물음에 답하라.

$$\sum_{i=1}^{30} X_i = 3025.2\,, \quad \sum_{i=1}^{30} X_i^2 = 305991\,, \quad \sum_{i=1}^{30} Y_i = 8469.7\,,$$

$$\sum_{i=1}^{30} Y_i^2 = 2401495\,, \quad \sum_{i=1}^{30} X_i Y_i = 856777$$

(1) 최소제곱법에 의한 회귀계수의 추정치를 구하라.

(2) 분산 σ^2의 점추정값을 계산하고, 결정계수 R^2을 구하라.

(3) 회귀모형의 유의성을 유의수준 5%로 검정하라.

✎ 해답

(1) $S_{XX} = \sum_{i=1}^{30}(X_i - \overline{X})^2 = \sum_{i=1}^{30} X_i^2 - 30(\overline{X})^2 = 305991 - 30 \times 10168.71 = 929.83$,

$S_{YY} = \sum_{i=1}^{30}(Y_i - \overline{Y})^2 = \sum_{i=1}^{30} Y_i^2 - 30(\overline{Y})^2 = 2401495 - 30 \times 79706.46 = 10301.06$,

$S_{XY} = \sum_{i=1}^{30}(X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^{30} X_i Y_i - 30\overline{X}\,\overline{Y} = 856777 - 30 \times 28469.48 = 2692.45$

이므로

$$\widehat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{2692.45}{929.83} = 2.9,\quad \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X} = -9.67$$

이다. 그러므로 추정된 단순 선형 회귀모형은 다음과 같다.

$$\therefore\ \widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X = -9.67 + 2.9X$$

(2) $\widehat{\sigma^2} = MSE = \dfrac{SSE}{n-2} = 89.45$

$SST = S_{YY} = 10301.06$, $SSR = \dfrac{[S_{XY}]^2}{S_{XX}} = \dfrac{2692.45^2}{929.83} = 7796.35$

$R^2 = \dfrac{SSR}{SST} = 0.76$

(3)

	자유도	제곱합	평균제곱합	F값
회귀	1	7796.35	7796.35	87.15
잔차	28	2504.71	89.45	
총합	29	10301.06		

$F = 87.15 > F_{1,28,0.05} = 4.2$이므로 영가설 $H_0 : \beta_0 = \beta_1 = 0$을 기각하고 추정된 회귀직선이 유의하다고 결론을 내릴 수 있다.

04 다음은 30개의 표본으로부터 구한 분산분석표이다. 빈칸을 채우고 물음에 답하라.

	자유도	제곱합	평균제곱합	F값
회귀	4	824	( )	( )
잔차	( )	( )	( )	
총합	( )	1254		

(1) 회귀모형을 세우고, 회귀모형의 유의성 여부에 대해 영가설과 대립가설을 설정하라.
(2) 결정계수와 수정결정계수를 구하여 비교하라.
(3) F분포표를 참고하여 F값에 대해 유의수준 5%하에서 가설검정하라.

해답

	자유도	제곱합	평균제곱합	F값
회귀	4	824	206	11.98
잔차	25	430	17.2	
총합	29	1254		

(1) $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_4 X_4 + \epsilon_i,\ \epsilon_i \sim iid\ N(0,\ \sigma^2),\ i = 1,\ 2,\ \cdots,\ 30$

$H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0,\ H_1 : \text{not } H_0$

(2) $R^2 = \dfrac{SSR}{SST} = \dfrac{824}{1254} = 0.66,\ R_a^2 = 1 - (n-1)\dfrac{MSE}{SST} = 1 - 29\dfrac{17.2}{1254} = 0.60$

수정된 결정계수는 결정계수보다 작다.

(3) $F = 11.98 > F_{4,\,25,\,0.05} = 2.76$이므로 영가설 H_0을 기각하고 추정된 회귀직선이 유의하다고 결론을 내릴 수 있다.

05 다음 자료는 6개의 도매점에서 조사된 특정 제품의 가격에 따른 수요량이다. 물음에 답하라.

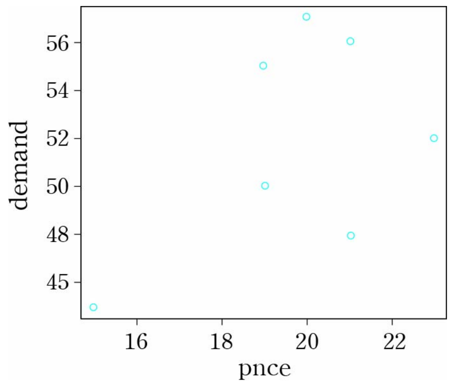
가격	21	23	19	21	15	19	20
수요량	56	52	55	48	45	50	57

(1) 수요량을 반응변수로 하는 단순회귀모형이 적합한지 산점도를 통해 확인하라.

(2) 회귀계수의 추정치를 구하라.

(3) 가격이 24일 때 예상되는 평균 수요량의 예측값을 구하라.

해답 (1)



(2) $Y = \beta_0 + \beta_1 X + \epsilon$

$\overline{X} = 19.71$, $\overline{Y} = 51.86$

$S_{XX} = \sum_{i=1}^{7}(X_i - \overline{X})^2 = 37.43$, $S_{YY} = \sum_{i=1}^{7}(Y_i - \overline{Y})^2 = 118.86$,

$S_{XY} = \sum_{i=1}^{7}(X_i - \overline{X})(Y_i - \overline{Y}) = 33.71$이므로

$\widehat{\beta_1} = \frac{S_{XY}}{S_{XX}} = 0.9$, $\widehat{\beta_0} = \overline{Y} - \widehat{\beta_1}\,\overline{X} = 34.1$

따라서 추정된 단순 선형회귀 직선은 $\widehat{Y} = 34.1 + 0.9X$이다.

(3) $34.1 + 0.9 \times 24 = 55.7$

06 [R] 다음과 같은 국고채와 콜금리의 데이터에 대해 단순 선형 회귀분석을 실시하고자 한다. 여기서 독립변수는 국고채이고, 종속변수는 콜금리이다. 다음 물음에 답하여라. (단위: %)

국고채	3.96	3.36	3.34	3.77	3.26	3.82	3.57	4.62	3.85	3.05
콜금리	2.75	3.25	3.26	3.01	3.31	3.24	2.99	2.01	1.97	2.27
국고채	4.22	3.35	3.61	4.16	3.80	3.56	4.16	3.81	4.12	3.83
콜금리	2.00	3.25	2.28	1.93	3.02	2.28	1.93	2.75	2.02	1.96

(1) 산점도를 그려보라.

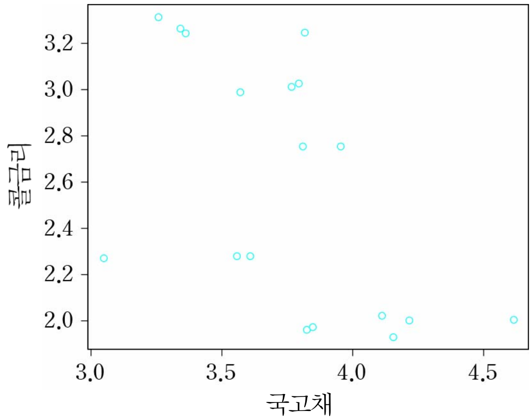
(2) 단순 선형 회귀모형을 설정하고 회귀계수의 추정치를 구하라.

(3) 산점도 위에 추정된 회귀선을 그려보라.

(4) 결정계수와 표본 상관계수를 구하라.

(5) 분산분석표를 작성하고, 회귀모형의 유의성을 유의수준 5%하에서 검정하라.

✎해답 (1)

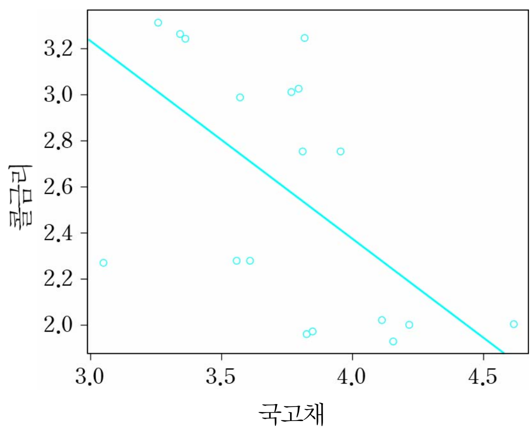


(2)

$$Y = \beta_0 + \beta_1 X + \epsilon$$

```
> summary(lm(콜금리~국고채))

Call:
lm(formula = 콜금리 ~ 국고채)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.9141 -0.3321  0.1956  0.3265  0.7166 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   5.8014     1.0219   5.677  2.2e-05 ***
국고채       -0.8581     0.2704  -3.174  0.00526 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4539 on 18 degrees of freedom
Multiple R-squared:  0.3588,    Adjusted R-squared:  0.3232 
F-statistic: 10.07 on 1 and 18 DF,  p-value: 0.005256
```

(3)



(4) Multiple R-squared: 0.3588, Adjusted R-squared: 0.3232

(5) 분산분석표를 작성하고, 회귀모형의 유의성을 유의수준 5%하에서 검정하라.

```
> summary(aov(콜금리~국고채))
            Df Sum Sq Mean Sq F value  Pr(>F)
국고채       1  2.076  2.0757   10.07 0.00526 **
Residuals   18  3.709  0.2061
```

$H_0 : \beta_0 = \beta_1 = 0,\ H_1 :$ not H_0

유의수준 0.05하에서 p-value가 유의수준보다 작기 때문에 귀무가설을 기각한다.

즉, 유의한 설명변수가 하나 이상 있다.

07 [R] 다음 자료에 대한 물음에 답하라.

Y	X_1	X_2	X_3	X_4
73	40	52	47	35
52	26	44	30	27
55	33	43	34	22
70	48	56	51	34
43	21	27	26	35
57	31	48	40	44

(1) 회귀모형을 세우고, 회귀모형의 유의성 여부에 대해 영가설과 대립가설을 설정하라.

(2) 분산분석표를 작성하라.

(3) 결정계수와 수정결정계수를 구하고 비교하라.

(4) 회귀계수가 의미가 있는지를 유의수준 5%하에서 검정하라.

✎해답

(1) $Y = \beta_0 + \beta_1 X + \beta_2 X + \beta_3 X + \beta_4 X + \epsilon,\ \epsilon_i \sim iid\ N(0, \sigma^2),\ i = 1, 2, \cdots, 4$

$H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0,\ H_1 :$ not H_0

(2)

```
> summary(aov(x7 ~ group, data = group_df))
            Df Sum Sq Mean Sq F value Pr(>F)
group        1   19.2    19.2   0.183  0.673
Residuals   22 2305.3   104.8
```

(3)

< summary (reg7)

```
Call:
lm(formula = y7 ~ x7_1 + x7_2 + x7_3 + x7_4)

Residuals:
      1       2       3       4       5       6 
 0.2043  0.8499 -1.0640  0.4429  0.2991 -0.7322 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  38.1753     6.9032   5.530    0.114
x7_1         -2.3506     0.6611  -3.555    0.175
x7_2         -0.2537     0.2109  -1.203    0.442
x7_3          3.8618     0.8136   4.746    0.132
x7_4         -1.1334     0.2803  -4.044    0.154

Residual standard error: 1.649 on 1 degrees of freedom
Multiple R-squared:  0.9957,    Adjusted R-squared:  0.9787 
F-statistic: 58.55 on 4 and 1 DF,  p-value: 0.09767
```

Multiple R-squared: 0.9957, Adjusted R-squared: 0.9787

수정된 결정계수가 결정계수보다 작다.

(4) 유의수준 0.05하에서 모든 X 변수의 p-value가 유의수준보다 크기 때문에 추정된 계수는 모두 0이라고 볼 수 있다.

11장 연습문제 풀이 I 및 I 해답

01 210명을 대상으로 어떤 안건에 대하여 조사한 자료가 다음과 같다. 세가지 의견에 대한 비율이 동일한지 유의수준 5%하에서 검정하라.

의견	찬성	반대	중립
관측빈도	85	55	70

✎해답 $H_0 : p_1 = \frac{1}{3},\ p_2 = \frac{1}{3},\ p_3 = \frac{1}{3}$

$H_1 : H_0$ 은 사실이 아님.

따라서 영가설하에서의 검정통계량의 값을 계산하면 다음과 같고

$$X_0^2 = \sum_{i=1}^{3} \frac{(O_i - E_{i0})^2}{E_{i0}} = \frac{(85-70)^2}{70} + \frac{(55-70)^2}{70} + \frac{(70-70)^2}{70} = 6.43$$

유의수준이 0.05인 기각역은 $C_{0.05} = \left\{X_0^2 \geq \chi^2_{2,\,0.05} = 5.991\right\}$이므로 영가설을 기각하고, 세 가지 의견에 대한 비율이 동일하다고 볼 수 없다고 결론을 내린다.

02 네 가지 상품의 선호도에 대하여 400명에게 조사한 자료가 다음과 같다. 네 가지 상품에 대한 소비자의 선호도가 같다는 것에 대해 유의수준 5%하에서 가설 검정하라.

상품	A	B	C	D
빈도	94	74	123	109

✎해답 H_0 : $p_1 = \dfrac{1}{4}$, $p_2 = \dfrac{1}{4}$, $p_3 = \dfrac{1}{4}$, $p_4 = \dfrac{1}{4}$

H_1 : H_0 은 사실이 아님.

따라서 영가설하에서의 검정통계량의 값을 계산하면 다음과 같고

$$X_0^2 = \sum_{i=1}^{4} \frac{(O_i - E_{i0})^2}{E_{i0}}$$

$$= \frac{(94-100)^2}{100} + \frac{(74-100)^2}{100} + \frac{(123-100)^2}{100} + \frac{(109-100)^2}{100} = 13.22$$

유의수준이 0.05인 기각역은 $C_{0.05} = \left\{X_0^2 \geq \chi^2_{3,\,0.05} = 7.815\right\}$이므로 영가설을 기각하고, 네 가지 상품에 대한 소비자의 선호도가 동일하다고 볼 수 없다고 결론을 내린다.

03 다음은 주사위를 30번 던져서 나온 결과이다. 물음에 답하여라.

눈	1	2	3	4	5	6
관측빈도	5	3	7	5	6	4

(1) 주사위 값이 공정한가에 대한 가설을 설정하라.

(2) 다음 표를 완성하라.

값	O_i	E_i	$(O_i - E_i)^2$	$\dfrac{(O_i - E_i)^2}{E_i}$
1				
2				
3				
4				
5				
6				
합계				

(3) 유의수준 5%에서 (1)의 가설을 검정하라.

✎해답 (1) H_0 : $p_1 = \dfrac{1}{6}$, $p_2 = \dfrac{1}{6}$, $p_3 = \dfrac{1}{6}$, $p_4 = \dfrac{1}{6}$, $p_5 = \dfrac{1}{6}$, $p_6 = \dfrac{1}{6}$

H_1 : H_0은 사실이 아님.

(2)

값	O_i	E_i	$(O_i - E_i)^2$	$\dfrac{(O_i - E_i)^2}{E_i}$
1	5	5	0	0
2	3	5	4	0.8
3	7	5	4	0.8
4	5	5	0	0
5	6	5	1	0.2
6	4	5	1	0.2
합계	30	30	10	2

(3) 따라서 영가설하에서의 검정통계량의 값을 계산하면 다음과 같고

$$X_0^2 = \sum_{i=1}^{6} \frac{(O_i - E_{i0})^2}{E_{i0}}$$
$$= \frac{(5-5)^2}{5} + \frac{(3-5)^2}{5} + \frac{(7-5)^2}{5} + \frac{(5-5)^2}{5} + \frac{(6-5)^2}{5} + \frac{(4-5)^2}{5} = 2$$

유의수준이 0.05인 기각역은 $C_{0.05} = \left\{ X_0^2 \le \chi^2_{5,\,0.05} = 11.070 \right\}$이므로 영가설을 채택한다. 즉, 주사위의 값은 $\dfrac{1}{6}$으로 공정하다.

04 다음은 고혈압환자 가운데 합병증이 발병한 환자를 조사한 결과이다. 물음에 답하라.

나이 \ 성별	남성	여성
~39	15	6
40~49	24	12
50~59	35	27
60~69	33	44
70~	12	13
합계	119	102

(1) 고혈압의 합병증 발병에 대해 성별과 나이가 서로 독립인지 알고 싶다. 적절한 가설을 설정하라.

(2) 다음 표를 완성하라.

성별	연령	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
남자	~39				
남자	40~49				
남자	50~59				
남자	60~69				
남자	70~				
여자	~39				
여자	40~49				
여자	50~59				
여자	60~69				
여자	70~				
합계					

(3) 유의수준 5%에서 (1)의 가설을 검정하라.

✎ 해답

(1) H_0 : 두 범주형 변수 성별과 나이는 서로 독립이다.

H_1 : 두 범주형 변수 성별과 나이는 서로 독립이 아니다.

(2)

성별	연령	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
남자	~39	15	11.31	13.63	1.21
남자	40~49	24	19.38	21.30	1.10
남자	50~59	35	33.38	2.61	0.08
남자	60~69	33	41.46	71.60	1.73
남자	70~	12	13.46	2.14	0.16
여자	~39	6	9.69	13.63	1.41
여자	40~49	12	16.62	21.30	1.28
여자	50~59	27	28.62	2.61	0.09
여자	60~69	44	35.54	71.60	2.01
여자	70~	13	11.54	2.14	0.19
합계		221	221	222.56	9.25

(3) 따라서 영가설하에서의 검정통계량의 값을 계산하면 다음과 같고

$$X_0^2 = \sum_{i=1}^{2}\sum_{j=1}^{5} \frac{(O_{ij} - \widehat{E_{ij}})^2}{\widehat{E_{ij}}} = 9.25$$

유의수준이 0.05인 기각역은 $C_{0.05} = \left\{ X_0^2 \le \chi^2_{4,\,0.05} = 11.143 \right\}$이므로 영가설을 채택하고, 성별과 나이는 서로 독립이라는 결론을 내린다.

05 어느 대학교 교학팀에서 학생들의 취업률과 성적 사이의 연관성을 파악하기 위하여 작년에 졸업한 학생 가운데 300명을 임의로 선정하여 취업유무와 평균학점을 조사하여 다음과 같은 자료를 얻었다. 평균학점과 취업여부 사이에 어떤 관계가 있는지를 유의수준 5%에서 검정하라.

	4.00 이상	3.00~3.99	2.00~2.99
취업 ○	58	51	55
취업 ×	17	36	64

✎ 해답

	4.00 이상	3.00-3.99	2.00-2.99	2.00 이하	합계
취업 ○	58	51	55	10	174
취업 ×	17	36	64	9	126
합계	75	87	119	19	300

H_0 : 취업률과 성적은 서로 관련이 없다.

H_1 : 취업률과 성적은 서로 관련이 있다.

지지정당 / 의견	4.00 이상	3.00-3.99	2.00-2.99	2.00 이하
취업 ○	$43.5 = (\widehat{E_{11}})$	$50.46 = (\widehat{E_{21}})$	$69.02 = (\widehat{E_{31}})$	$11.02 = (\widehat{E_{41}})$
취업 ×	$31.5 = (\widehat{E_{12}})$	$36.54 = (\widehat{E_{22}})$	$49.98 = (\widehat{E_{32}})$	$7.98 = (\widehat{E_{42}})$

여기서 각 셀(cell)의 값은 다음과 같이 계산되었다.

$$\widehat{E_{11}} = \frac{174}{300} \times \frac{75}{300} \times 300 = 43.5, \quad \widehat{E_{12}} = \frac{126}{300} \times \frac{75}{300} \times 300 = 31.5$$

$$\widehat{E_{21}} = \frac{174}{300} \times \frac{87}{300} \times 300 = 50.46, \quad \widehat{E_{22}} = \frac{126}{300} \times \frac{87}{300} \times 300 = 36.54$$

$$\widehat{E_{31}} = \frac{174}{300} \times \frac{119}{300} \times 300 = 69.02, \quad \widehat{E_{32}} = \frac{126}{300} \times \frac{119}{300} \times 300 = 49.98$$

$$\widehat{E_{41}} = \frac{174}{300} \times \frac{19}{300} \times 300 = 11.02, \quad \widehat{E_{42}} = \frac{126}{300} \times \frac{19}{300} \times 300 = 7.98$$

따라서 영가설하에서의 검정통계량의 값을 계산하면 다음과 같고

$$\begin{aligned} X_0^2 &= \sum_{i=1}^{4} \sum_{j=1}^{2} \frac{(O_{ij} - \widehat{E_{ij}})^2}{\widehat{E_{ij}}} \\ &= \frac{(58-43.5)^2}{43.5} + \frac{(51-50.46)^2}{50.46} + \cdots + \frac{(9-7.98)^2}{7.98} \\ &= 18.53 \end{aligned}$$

유의수준이 0.05인 기각역은 $C_{0.05} = \left\{ X_0^2 \geq \chi^2_{3,\,0.05} = 7.815 \right\}$이므로 영가설을 기각하고, 취업률과 성적은 서로 관련이 있다고 결론을 내릴 수 있다.

06 다음은 30세 이상의 성인 남자 가운데 흡연자 200명, 비흡연자 200명을 대상으로 연령대별로 흡연 여부를 조사한 결과이다. 각 연령대별로 흡연 비율이 차이가 있는가를 유의수준 5% 하에서 검정하라.

연령	흡연자	비흡연자	합계
30~39	85	75	160
40~49	45	60	105
50~59	50	40	90
60 이상	20	25	45
합계	200	200	400

✎해답 H_0 : 각 연령대와 흡연비율은 서로 독립이다(즉, 아무 연관성이 없다).

H_1 : 각 연령대와 흡연비율은 서로 독립이 아니다(즉, 각 연령에 따라 흡연비율은 차이가 있다).

영가설하에서 기대빈도의 추정치를 구하면 다음과 같다.

흡연여부 / 의연령	흡연자	비흡연자
30-39	$80 = (\widehat{E_{11}})$	$80 = (\widehat{E_{21}})$
40-49	$52.5 = (\widehat{E_{12}})$	$52.5 = (\widehat{E_{22}})$
50-59	$45 = (\widehat{E_{13}})$	$45 = (\widehat{E_{23}})$
60이상	$22.5 = (\widehat{E_{14}})$	$22.5 = (\widehat{E_{24}})$

영가설하에서의 검정통계량의 값을 계산하면 다음과 같고

$$X_0^2 = \sum_{i=1}^{4}\sum_{j=1}^{4}\frac{(O_{ij}-\widehat{E_{ij}})^2}{\widehat{E_{ij}}} = 4.435$$

유의수준이 0.05인 기각역은 $C_{0.05} = \left\{X_0^2 \le \chi^2_{9,\,0.05} = 16.919\right\}$이므로 영가설을 채택한다. 연령에 따라 흡연비율은 차이가 있다고 결론을 내릴 수 있다.