생생한 사례로 배우는 확률과 통계

추가문제 이용 안내

1 ~ 2장

1. 어느 농촌지역의 가옥내 배수설비로부터 악취발생 여부를 알기 위하여 이 지역에 거주하는 258명을 표본조사한 결과 다음 결과를 얻었다. 이 자료에 대한 도수표, 꺾은선 그래프와 원그래프를 그려라. 단위는 명이다.

자주 발생	가끔 발생	무발생	무응답
33	70	154	1

풀이

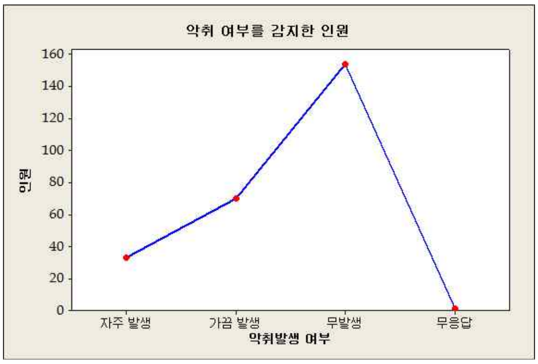
전체 표본조사 인원은 258이다. 따라서 다음 도수표를 얻는다.

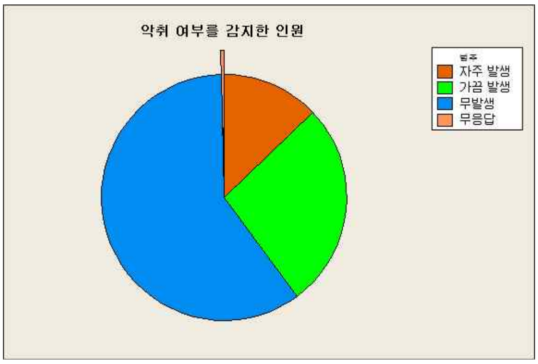
종류	건수	상대도수
자주 발생	33	0.128
가끔 발생	70	0.271
무발생	154	0.597
무응답	1	0.004

각 악취발생 여부별 중심각은 다음과 같다.

자주 발생: 46.0° 가끔 발생: 97.7° 무발생: 214.9° 무응답: 1.4°

따라서 악취발생 여부별 건수에 대한 막대그래프와 원그래프는 다음과 같다.

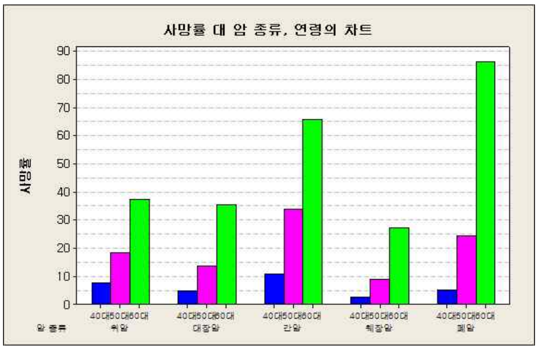


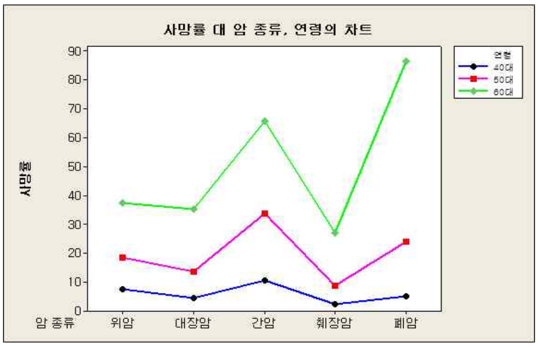


2. 다음은 연령별 암 종류에 의한 사망률을 나타낸다. 암 종류에 대한 연령별 사망률을 비교하는 막대그래프와 꺾은선 그래프를 그려라. 단위는 %이다.

종류 / 연령	위암	대장암	간암	췌장암	폐암
40대	7.7	4.7	10.8	2.4	5.2
50대	18.5	13.7	33.7	8.7	24.2
60대	37.4	35.4	65.7	27.1	86.4

풀이





3. 다음은 2010년 구미시의 석유류 소비량을 나타낸다. 단위는 kl이다.

월	1	2	3	4	5	6
휘발유	8,398	7,833	8,367	8,432	8,861	8,330
등유	3,342	2,467	2,219	1,659	1,251	1,122
경유	11,660	10,759	12,367	12,710	13,235	12,600
LPG	5,195	4,996	4,929	4,720	4,931	4,582
월	7	8	9	10	11	12
휘발유	8,882	9,232	8,948	8,928	8,655	8,949
등유	1,038	974	1,123	1,880	2,627	3,836
경유	12,960	12,805	11,869	13,945	13,549	13,861
LPG	4,651	4,824	4,821	4,681	5,349	5,455

출처: 구미시 과학경제과

(a) 각 석유류의 연평균을 구하라.

(b) 각 석유류의 월별 사용량을 비교하는 막대그래프와 꺾은선 그래프를 그려라.

풀이

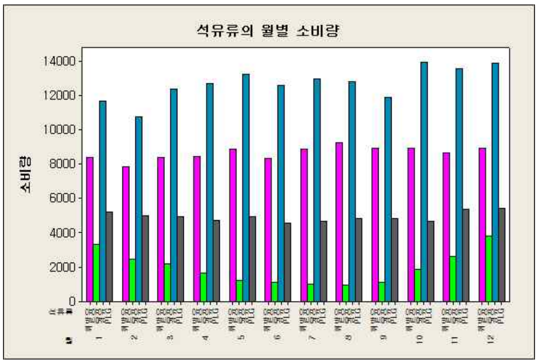
(a) 휘발유의 평균 소비량: $\overline{x}_1 = \frac{1}{12}\sum_{i=1}^{10} x_i = \frac{103815}{12} = 8651$

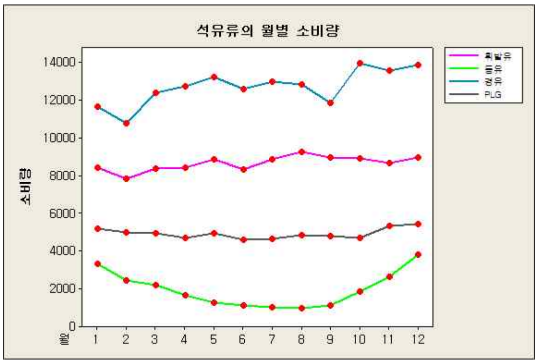
등유의 평균 소비량: $\overline{x}_2 = \frac{1}{12}\sum_{i=1}^{10} y_i = \frac{23538}{12} = 1962$

경유의 평균 소비량: $\overline{x}_3 = \frac{1}{12}\sum_{i=1}^{10} z_i = \frac{152320}{12} = 12693$

LPG의 평균 소비량: $\overline{x}_4 = \frac{1}{12}\sum_{i=1}^{10} w_i = \frac{59134}{12} = 4927.8$

(b)





4. 다음 표는 통계청의 『인구주택총조사보고서』에서 어느 도시의 15세 이상 배우자가 있는 남자의 자료이다. 이때 85세 이상은 85세 이상 89세 이하로 생각한다.

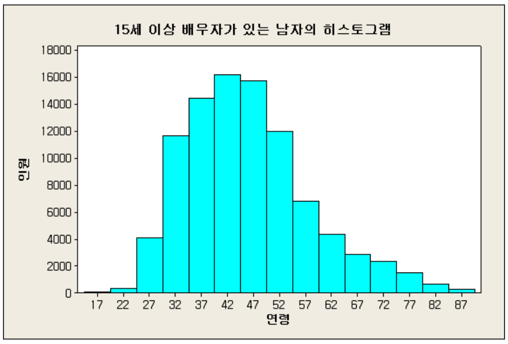
나이	15~19세	20~24세	25~29세	30~34세	35~39세	40~44세	45~49세	50~54세
인원	55	315	4,080	11,623	14,450	16,218	15,726	11,990
나이	55~59세	60~64세	65~69세	70~74세	75~79세	80~84세	85세 이상	
인원	6,809	4,330	2,871	2,352	1,484	619	247	

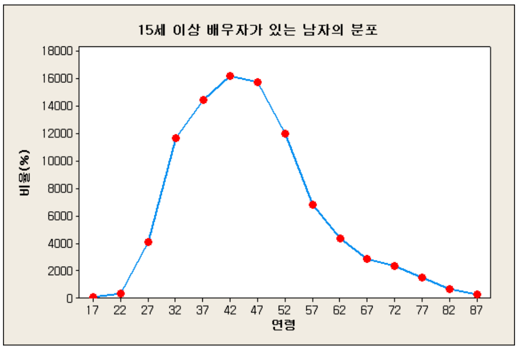
(a) 이 자료에 대한 도수분포표, 히스토그램과 도수다각형을 그려라.
(b) 누적도수히스토그램, 누적상대도수히스토그램과 누적상대도수다각형을 그려라.
(c) 이 자료에 대한 평균과 표준편차를 구하라.

풀이

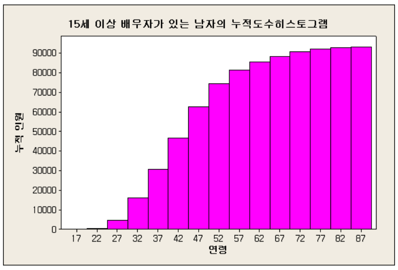
(a)

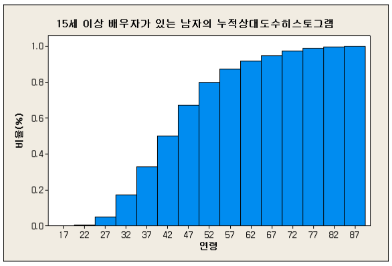
계급간격	도수	상대도수	누적도수	누적상대도수	계급값
14.5 - 19.5	55	0.0006	55	0.0006	17
19.5 - 24.5	315	0.0034	370	0.0040	22
24.5 - 29.5	4,080	0.0438	4450	0.0478	27
29.5 - 34.5	11,623	0.1248	16073	0.1726	32
34.5 - 39.5	14,450	0.1551	30523	0.3277	37
39.5 - 44.5	16,218	0.1741	46741	0.5018	42
44.5 - 49.5	15,726	0.1688	62467	0.6706	47
49.5 - 54.5	11,990	0.1287	74457	0.7993	52
54.5 - 59.5	6,809	0.0731	81266	0.8724	57
59.5 - 64.5	4,330	0.0465	85596	0.9189	62
64.5 - 69.5	2,871	0.0308	88467	0.9497	67
69.5 - 74.5	2,352	0.0252	90819	0.9749	72
74.5 - 79.5	1,484	0.0159	92303	0.9908	77
79.5 - 84.5	619	0.0066	92922	0.9974	82
84.5 -	247	0.0026	93169	1.0000	87
합 계	93169	1.00			

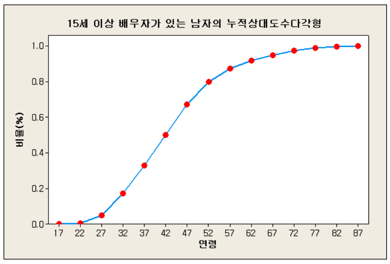




(b)







(c) 각 계급의 계급값 x_i와 상대도수 f_i/n에 대하여 평균은 다음과 같다.

$$\overline{x} = 17 \times 0.0006 + 22 \times 0.0034 + \cdots + 87 \times 0.0026 = 45.86$$

분산은 다음과 같다.

$$s^2 = (17 - 45.86)^2 \times 0.0006 + (22 - 45.86)^2 \times 0.0034 + \cdots + (87 - 45.86)^2 \times 0.0026 = 142.536$$

따라서 표준편차는 $s = \sqrt{142.536} = 11.9388$이다.

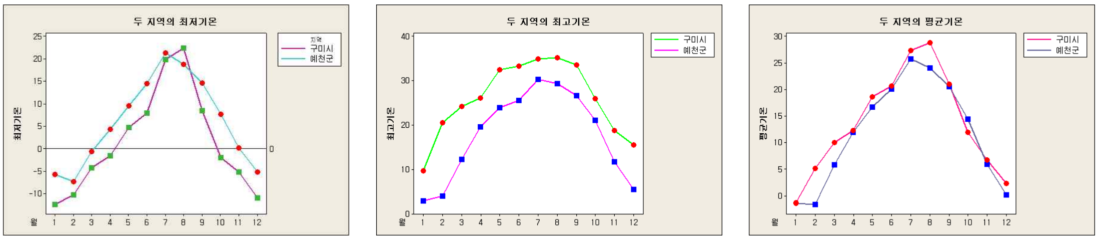
5. 다음 자료는 구미시와 예천군의 월별 기상개황에 대한 자료이다.(구미시는 2010년 대구기상대, 예천군은 2009년 안동기상대에서 측정한 기상연보이다.)

월	최저기온(℃)		최고기온(℃)		강수량(mm)		평균풍속(m/s)		상대습도(%)	
	구미시	예천군	구미시	예천군	구미시	예천군	구미시	예천군	구미시	예천군
1	-12.4	-5.7	9.8	3.0	8.8	33.1	2.2	3.3	56.8	65.0
2	-10.2	-7.3	20.5	4.1	60.6	4.3	1.7	4.3	60.6	50.0
3	-4.2	-0.6	24.2	12.3	76.3	65.0	2.0	3.0	61.2	63.0
4	-1.6	4.4	26.1	19.6	55.0	27.2	2.0	2.9	52.3	58.0
5	4.7	9.6	32.5	23.9	89.0	62.1	1.8	2.7	55.8	60.0
6	8.0	14.5	33.3	25.6	12.3	168.0	1.4	2.0	59.0	73.0
7	19.8	21.3	34.9	30.3	173.3	312.2	1.4	1.7	72.1	79.0
8	22.4	18.7	35.1	29.3	418.3	257.5	1.3	1.7	76.1	78.0
9	8.5	14.6	33.5	26.6	188.6	80.6	1.2	1.7	75.9	77.0
10	-2.0	7.7	26.0	21.1	28.2	30.1	1.3	2.3	71.7	72.0
11	-5.2	0.2	18.8	11.8	11.4	10.7	1.7	2.8	53.1	67.0
12	-10.9	-5.2	15.6	5.6	31.7	12.3	2.0	3.5	55.3	63.0

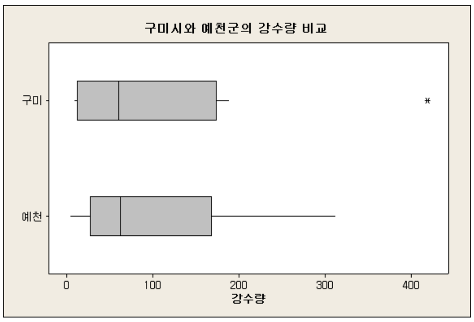
(a) 월별 구미시와 예천군의 최저기온, 최고기온 그리고 월별 평균기온을 비교하는 시계열 그림을 그려라.
(b) 구미시와 예천군의 강수량을 비교하는 상자그림을 그려라.
(c) 구미시와 예천군의 평균풍속에 대한 표준점수를 구하라.
(d) 구미시와 예천군의 강수량과 평균풍속의 변동계수를 구하라.
(e) 구미시와 예천군의 평균풍속에 대한 전체 자료의 도수히스토그램을 그리고 평균, 최빈값과 중위수를 구하라. 단, 히스토그램의 계급값은 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5이다.
(f) 구미시와 예천군의 상대습도에 대한 전체 자료의 평균과 사분위수를 구하라.

풀이

(a)



(b)



(c) 두 도시의 월별 평균풍속에 대한 평균과 분산 그리고 표준편차를 구하면 각각 다음과 같다.

	평균	분산	표준편차
구미시	1.67	0.115	0.339
예천군	2.66	0.666	0.816

따라서 두 도시의 월별 평균풍속에 대한 표준점수는 다음과 같다.

월	1	2	3	4	5	6
구미시	1.56342	0.08850	0.97345	0.97345	0.38348	-0.79646
예천군	0.78431	2.00980	0.41667	0.29412	0.04902	-0.80882
월	7	8	9	10	11	12
구미시	-0.79646	-1.09145	-1.38643	-1.09145	0.08850	0.97345
예천군	-1.17647	-1.17647	-1.17647	-0.44118	0.17157	1.02941

(d) 두 도시의 월별 강수량에 대한 평균과 분산 그리고 표준편차를 구하면 각각 다음과 같다.

	평균	분산	표준편차
구미시	102.0	14776.9	121.6
예천군	95.5	10907.1	104.4

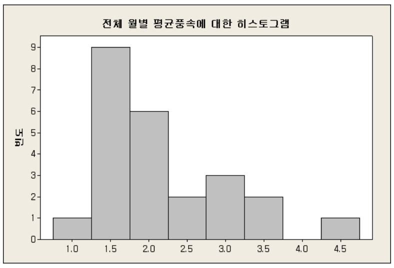
• 구미시의 강수량과 평균풍속에 대한 변동계수:

강수량의 변동계수: $\dfrac{121.6}{102} = 1.192$,　평균풍속의 변동계수: $\dfrac{0.339}{1.67} = 0.203$

• 예천군의 강수량과 평균풍속에 대한 변동계수:

강수량의 변동계수: $\dfrac{104.4}{95.5} = 1.093$,　평균풍속의 변동계수: $\dfrac{0.816}{2.66} = 0.307$

(e) 평균은 $\bar{x} = \dfrac{1}{24}\Sigma x_i = 2.163$이다. 최빈값은 1.7이다. 전체 자료의 수가 24이므로 중위수는 $Q_2 = \dfrac{x_{(12)} + x_{(13)}}{2} = \dfrac{2.0 + 2.0}{2} = 2.0$이다. 또한 도수히스토그램은 오른쪽 그림과 같다.



(f) 전체 자료의 수가 24이고 $24 \times 0.25 = 6$, $24 \times 0.5 = 12$, $24 \times 0.75 = 18$이므로 사분위수는 다음과 같다.

$$Q_1 = \frac{x_{(6)} + x_{(7)}}{2} = \frac{56.8 + 58.0}{2} = 57.4$$

$$Q_2 = \frac{x_{(12)} + x_{(13)}}{2} = \frac{63.0 + 63.0}{2} = 63.0$$

$$Q_3 = \frac{x_{(18)} + x_{(19)}}{2} = \frac{72.1 + 73.0}{2} = 72.55$$

평균은 $\bar{x} = \dfrac{1}{24}\Sigma x_i = 64.79$이다.

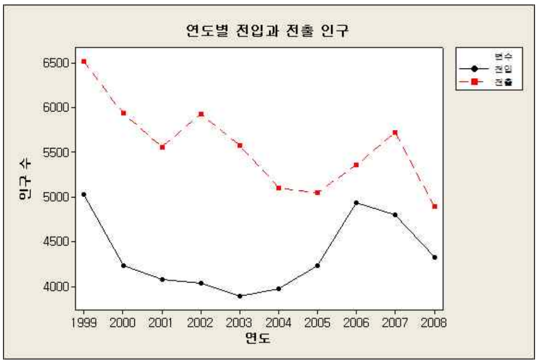
6. 어느 농촌 지역의 인구증가요인을 분석한 결과 다음과 같은 결과를 얻었다.

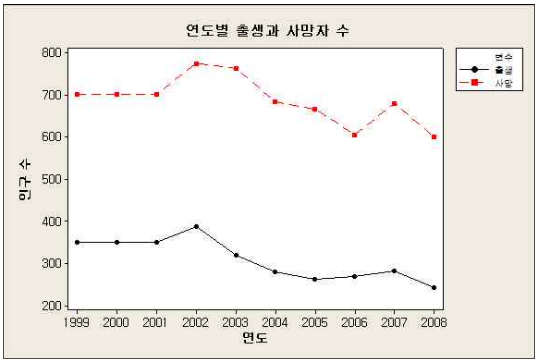
연도		1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
사회적 증감	전입	5,024	4,233	4,073	4,039	3,892	3,971	4,236	4,933	4,798	4,321
	전출	6,513	5,935	5,558	5,929	5,572	5,102	5,046	5,361	5,718	4,890
자연적 증감	출생	350	350	350	388	318	280	261	269	282	243
	사망	700	700	700	774	763	684	666	605	680	599

(a) 사회적 증감과 자연적 증감에 대한 시계열 그림을 그려라.
(b) 이 농촌에서 외지로 전출한 인구와 외지에서 이 지역으로 전입한 인구의 차를 순이동이라 한다. 순이동에 대한 시계열 그림을 그려라.
(c) 연도에 따른 출생과 사망을 비교하는 막대그래프를 그려라.

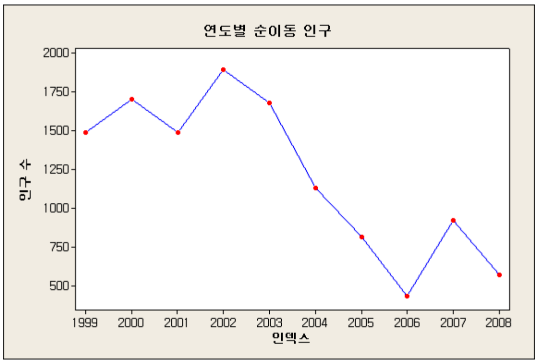
풀이

(a)





(b)



(c)

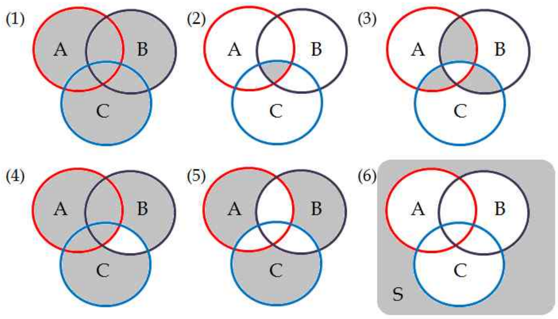
3장

1. 어느 건설회사는 서로 다른 세 도시 A, B, C에서 교량공사를 하고 있다. 이들 세 도시에서 계약일자에 맞춰 공사가 완공되는 사건을 각각 A, B, C로 나타낸다. 사건 A, B, C와 벤다이어그램을 다음 사건을 표현하라.

(a) 적어도 한 교량이 계약일자 안에 완공된다.
(b) 모든 교량이 계약일자 안에 완공된다.
(c) 어느 두 도시만 교량이 계약일자 안에 완공된다.
(d) A 도시 또는 다른 두 도시 중에서 어느 한 도시만 교량이 계약일자 안에 완공된다.
(e) 꼭 한 도시만 교량이 계약일자 안에 완공된다.
(f) 세 도시 모두 계약일자 안에 완공되지 않는다.

풀이

(a) $A \cup B \cup C$
(b) $A \cap B \cap C$
(c) $[(A \cap B) \cup (B \cap C) \cup (A \cap C)] \cap (A \cap B \cap C)^c$
(d) $[(A \cup B \cup C) \cap (B \cap C)^c] \cup (A \cap B \cap C)$
(e) $(A \cup B \cup C) \cap [(A \cap B) \cup (B \cap C) \cup (A \cap C)]^c$
(f) $(A \cup B \cup C)^c$



2. 여러 도시에 공급되는 상수도를 조사하면 두 종류의 불순물이 공통적으로 발견된다. 상수도의 20%는 아무런 불순물도 포함되지 않았고, A형의 불순물이 발견된 상수도는 45% 그리고 B형의 불순물이 발견된 상수도는 55%이었다. 한 도시를 임의로 선택하여 불순물을 조사할 때, 오로지 한 종류의 불순물만 발견될 확률을 구하라.

풀이

상수도에서 A형의 불순물이 발견되는 사건을 A, B형의 불순물이 발견되는 사건을 B라 하면, $P(A) = 0.45$, $P(B) = 0.55$ 그리고 $P(A^c \cap B^c) = 0.2$이다. 그러므로 두 사건 A와 B에 대해 다음 확률을 얻는다.

$$P(A \cup B) = 1 - P(A^c \cap B^c) = 1 - 0.2 = 0.8$$
$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.45 + 0.55 - 0.8 = 0.2$$

그리고 오로지 한 종류의 불순물만 발견되는 사건은 다음과 같이 서로 배반인 두 사건의 합사건으로 표현된다.

$$(A - B) \cup (B - A) = [A - (A \cap B)] \cup [B - (A \cap B)]$$

따라서 구하고자 하는 확률은 다음과 같다.

$$P[(A - B) \cup (B - A)] = P[A - (A \cap B)] + P[B - (A \cap B)] = P(A) + P(B) - 2P(A \cap B)$$
$$= 0.45 + 0.55 - 2 \times 0.2 = 0.6$$

3. $P(A)=P(B)=P(C)=\frac{1}{3}$, $P(A\cap B)=P(A\cap C)=P(B\cap C)=\frac{1}{8}$, $P(A\cap B\cap C)=\frac{1}{16}$일 때, $P(A\cup B\cup C)$를 구하라.

풀이

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

$$=3\times\frac{1}{3}-3\times\frac{1}{8}+\frac{1}{16}=\frac{11}{16}$$

4. 건설업자는 신도시 건설 프로젝트에 필요한 덤프트럭을 매입하려고 한다. 이 업자가 이전에 덤프트럭을 구입한 경험에 따르면, 각 덤프트럭이 적어도 6개월 이상 사용할 수 있을 확률이 50%이다. 이 사람이 새로운 신도시 건설을 위해 덤프트럭 4대를 구입했을 때, 다음을 구하라.

(a) 6개월 후에도 사용할 수 있는 덤프트럭 수에 대한 표본공간을 구하라.

(b) 단 한 대의 덤프트럭만이 6개월 후에도 사용할 수 있을 확률을 구하라.

풀이

(a) 6개월이 지난 시점에서 불도저를 사용할 수 있으면 O 그렇지 않으면 N이라 하면, 나타날 수 있는 모든 경우는 다음과 같다.

$$S=\begin{Bmatrix} OOOO, OOON, OONO, ONOO, OONN, ONON, ONNO, ONNN \\ NOOO, NOON, NONO, NNOO, NONN, NNON, NNNO, NNNN \end{Bmatrix}$$

따라서 6개월이 지난 시점에서 사용 가능한 불도저의 수는 0, 1, 2, 3, 4이므로 표본공간은 $S=\{0,1,2,3,4\}$이다.

(b) 단 한 대의 불도저만이 6개월 후에도 사용할 수 있는 사건은 다음과 같다.

$$A=\{ONNN, NONN, NNON, NNNO\}$$

따라서 구하고자 하는 확률은 $P(A)=\frac{4}{16}=\frac{1}{4}$이다.

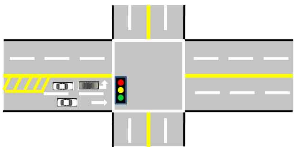
5. 어느 회사는 국가에서 실시하는 두 프로젝트를 동시에 지원한다. 이때 두 프로젝트 중에서 어느 하나가 선정될 확률은 0.8이고 두 프로젝트 모두 선정될 확률은 0.3이라 한다. 이때 두 프로젝트 중에서 꼭 하나만 선정될 확률을 구하라.

풀이

두 프로젝트가 선정되는 사건을 각각 A, B라 하면, $P(A\cup B)=0.8$, $P(A\cap B)=0.3$이다. 그러면 두 프로젝트 중에서 꼭 하나만 선정되는 사건은 $(A\cup B)-(A\cap B)$이므로 구하고자 하는 확률은 다음과 같다.

$$P[(A\cup B)-(A\cap B)]=P(A\cup B)-P(A\cap B)=0.8-0.3=0.5$$

6. 그림과 같은 교차로에서 좌회전 차로를 디자인할 때, 원활한 교통을 위하여 좌회전 차로의 길이를 결정하는 것은 매우 중요하다.



이 목적 아래서 일주일 동안 교통량이 많은 시간대에 이 교차로에서 좌회전하기 위하여 기다리는 차량의 수를 관찰하여 다음을 얻었다.

4 5 7 9 8 7 8 7 7 2 3 9 5 8 7 9 7 6 5 3 4 7 5 7 7
5 8 2 6 9 0 8 4 6 2 8 5 2 6 8 9 5 1 5 9 7 6 8 4 5

(a) 차량 수에 대한 분포표를 작성하라.
(b) 임의의 시각에 좌회전을 위해 기다리는 차량이 5대 이상일 확률을 상대도수에 의해 구하라.

풀이

(a) 차량 수에 대한 분포표를 작성하면 다음과 같다.

차량 수	관찰 횟수	상대도수	차량 수	관찰 횟수	상대도수
0	1	0.02	5	9	0.18
1	1	0.02	6	5	0.10
2	4	0.08	7	10	0.20
3	2	0.04	8	8	0.16
4	4	0.08	9	6	0.12

(b) 좌회전을 위해 기다리는 차량이 5대 이상인 사건을 A라 하면, 구하고자 하는 확률은 다음과 같다.

$$P(A) = 0.18 + 0.10 + 0.20 + 0.16 + 0.12 = 0.76$$

7. 다음 표는 2006년 7월 통계청에서 공시한 의료산업에 종사하는 전문직 근로자 자료를 재구성한 것이다.

직 종	종사자 수		전체 비율(%)	
	여성	남성	여성	남성
의사	15,744	66,254	9.14	38.47
치과의사	4,738	16,606	2.75	9.64
한의사	1,910	13,496	1.11	7.84
약사	34,128	19,364	19.81	11.24
계	56,520	115,720	32.81	67.19
	172,240		100.0	

의료산업에 종사하는 근로자 중에서 임의로 한 사람을 선정했을 때, 이 사람에 대해 다음 확률을 구하라.

(a) 이 사람이 여성일 확률
(b) 이 사람이 남성인 치과의사일 확률
(c) 약사일 확률
(d) 여성일 때, 이 여성이 약사일 확률

풀이

(a) 의료산업에 종사하는 여성의 비율이 32.81%이므로 임의로 선정된 의료산업에 종사하는 사람이 여성일 확률은 0.3281이다.

(b) 남성인 치과의사의 비율이 9.64%이므로 남성인 치과의사가 선정될 확률은 0.0964이다.

(c) 남녀 약사의 비율이 각각 11.24%, 19.81%이므로 임의로 선정한 사람이 약사일 확률은 0.3105이다.

(d) 여성인 약사의 비율이 0.3281이고 여성인 약사의 비율이 0.1981이므로 구하고자 하는 확률은 $\frac{0.1981}{0.3281} = 0.6038$이다.

8. 앞면이 나올 가능성이 $\frac{2}{3}$인 찌그러진 동전을 독립적으로 두 번 던질 때, 다음을 구하라.

(a) 앞면이 한 번도 나오지 않을 확률
(b) 앞면이 한 번 나올 확률
(c) 앞면이 두 번 나올 확률

풀이

(a) 처음에 앞면이 나오는 사건을 A, 두 번째 앞면이 나오는 사건을 B라 하면, $P(A) = \frac{2}{3}$, $P(B) = \frac{2}{3}$이다. 그러므로 두 번 모두 뒷면이 나올 확률은 다음과 같다.

$$P(A^c \cap B^c) = P(A^c)P(B^c) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$$

(b) 앞면이 한 번 나오면 뒷면도 한 번 나오므로 확률은 다음과 같다.

$$P(A \cap B^c) + P(A^c \cap B) = P(A)P(B^c) + P(A^c)P(B) = \frac{2}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3} = \frac{4}{9}$$

(c) 앞면이 두 번 모두 나올 확률은 $P(A \cap B) = P(A)P(B) = \frac{2}{3} \times \frac{2}{3} = \frac{4}{9}$이다.

9. 혈압과 심장박동 사이의 관계를 연구하고 있는 의사가 자신의 환자를 대상으로 혈압 상태(정상, 고혈압, 저혈압)와 심박 상태(정상, 비정상)를 조사하여 다음의 결과를 얻었다.

- 14%는 고혈압이고, 22%는 저혈압이다.
- 15%는 심박이 비정상이다.
- 심박이 비정상인 환자 중에서 $\frac{1}{3}$이 고혈압이다.
- 정상 혈압을 가진 환자 중에서 $\frac{1}{8}$이 비정상적인 심박을 갖는다.

임의로 한 환자를 선정할 때 심장 박동이 정상이고 저혈압인 환자가 선정될 확률을 구하라.

풀이

조사 결과를 표로 작성하면 다음과 같다.

	고혈압	저혈압	정상 혈압	합　계
정상 심박	0.09	0.20	0.56	0.85
이상 심박	0.05	0.02	0.08	0.15
합　계	0.14	0.22	0.64	1.00

따라서 심장 박동이 정상이고 저혈압인 환자는 20%이다.

10. 흰색 바둑돌 4개와 검은색 바둑돌이 6개가 들어있는 주머니에서 비복원추출에 의해 차례로 바둑돌 세 개를 꺼낸다.
(a) 3개 모두 흰색일 확률을 구하라.
(b) 바둑돌이 차례로 흰색, 검은색 그리고 흰색일 확률을 구하라.

풀이

(a) 3개 모두 흰색일 확률은 $\frac{4}{10} \times \frac{3}{9} \times \frac{2}{8} = \frac{1}{30}$이다.

(b) 차례로 흰색, 검은색 그리고 흰색일 확률은 $\frac{4}{10} \times \frac{6}{9} \times \frac{3}{8} = \frac{1}{10}$이다.

11. 고층건물의 붕괴는 불충분한 지지대 또는 과도한 침하에 기인한다. 불충분한 지지대와 과도한 침하에 의하여 붕괴되는 사건을 각각 A와 B라 하면 $P(A)=0.002$, $P(B)=0.007$ 그리고 $P(A|B)=0.12$이다.

(a) 고층건물이 붕괴할 확률을 구하라.

(b) 과도한 침하에 의해서만 붕괴할 확률을 구하라.

(c) 어느 한 원인에 의해서만 붕괴할 확률을 구하라.

풀이

(a) $P(A\cap B)=P(B)P(A|B)=0.007\times 0.12=0.00084$이므로 구하고자 하는 확률은 다음과 같다.

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=0.002+0.007-0.00084=0.00816$$

(b) 과도한 침하에 의해서만 붕괴되는 사건은 $A^c\cap B$이므로 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned}P(A^c\cap B)&=P(B)P(A^c|B)=P(B)[1-P(A|B)]\\&=0.007\times(1-0.12)=0.00616\end{aligned}$$

(c) 어느 한 원인에 의해서만 붕괴하는 사건은 $(A^c\cap B)\cup(A\cap B^c)=(A\cup B)-(A\cap B)$이므로 구하고자 하는 확률은 다음과 같다.

$$P[(A^c\cap B)\cup(A\cap B^c)]=P(A\cup B)-P(A\cap B)=0.00816-0.00084=0.00732$$

12. 위성 시스템은 컴퓨터 1에 의하여 조정되며, 이 시스템은 두 개의 독립적인 백업용 컴퓨터 2와 3으로 구성된다. 정상적으로 컴퓨터 1은 시스템을 조정하지만 이 컴퓨터가 고장 나면 자동적으로 컴퓨터 2가 작동하고, 컴퓨터 2가 고장 나면 컴퓨터 3이 작동한다. 그리고 세 컴퓨터가 모두 고장 나면, 위성 시스템은 멈춘다고 한다. 그리고 각 컴퓨터들의 멈출 확률은 0.01이고, 이 컴퓨터들이 멈추는 것은 역시 독립적이다. 이때 각 컴퓨터들이 작동할 확률을 구하라. 그리고 위성 시스템이 멈출 확률을 구하라.

풀이

컴퓨터 1, 2, 3이 작동하는 사건을 각각 A, B 그리고 C라 하면, 컴퓨터 1이 멈출 확률이 0.01이므로 컴퓨터 1이 작동할 확률은 $P(A)=0.99$이다. 컴퓨터 2는 컴퓨터 1이 멈춘 조건 아래서 작동하므로 컴퓨터 2가 작동할 확률은 다음과 같다.

$$P(B)=P(A^c)P(B|A^c)=0.01\times 0.99=0.0099$$

이고, 또한 컴퓨터 3은 컴퓨터 1과 컴퓨터 2가 멈춘 조건 아래서 작동하므로

$$P(C)=P(A^c)P(B^c|A^c)P(C|A^c\cap B^c)=0.01\times 0.01\times 0.99=0.000099$$

이다. 그리고 각 컴퓨터들이 멈출 사건은 서로 독립이므로 위성 시스템이 멈출 확률은 다음과 같다.

$$P(\text{위성 시스템이 멈춤})=P(A^c)P(B^c)P(C^c)=0.01\times 0.01\times 0.01=10^{-6}$$

13. 어떤 교차로에서 우회전하는 차량에 비하여 직진하는 차량이 두 배가량 많다. 그리고 좌회전하는 차량은 우회전하는 차량의 $\frac{1}{2}$이다.

(a) 교차로에 접근하는 어떤 차량이 직진, 좌회전 그리고 우회전할 확률을 각각 구하라.

(b) 교차로에 접근하는 어떤 차량이 회전한다고 할 때, 이 차량이 좌회전할 확률을 구하라.

풀이

(a) 교차로에 접근하는 차량이 직진하는 사건을 A, 우회전하는 사건을 B 그리고 좌회전하는 사건을 C라 하면 $P(A) = 2P(B)$, $P(C) = \frac{1}{2}P(B)$이다. 따라서 다음을 얻는다.

$$P(A) + P(B) + P(C) = 2P(B) + P(B) + \frac{1}{2}P(B) = \frac{7}{2}P(B) = 1$$

그러므로 구하고자 하는 확률은 각각 다음과 같다.

$$P(A) = \frac{4}{7},\ P(B) = \frac{2}{7},\ P(C) = \frac{1}{7}$$

(b) 교차로에 접근하는 차량이 회전할 확률은 $P(B \cup C) = \frac{3}{7}$이므로 구하고자 하는 확률은 다음과 같다.

$$P(C|B \cup C) = \frac{P(C)}{P(B \cup C)} = \frac{1/7}{3/7} = \frac{1}{3}$$

14. 근로자가 어떤 기계의 사용 설명서에 따라 사용할 때 오작동이 발생할 확률은 1%이고 그렇지 않을 때 오작동이 발생할 확률은 4%이다. 시간이 부족한 관계로 근로자가 설명서를 80% 밖에 숙지하지 못한 상태로 기계를 사용할 때, 이 기계에 오작동이 발생할 확률을 구하라.

풀이

근로자가 사용 설명서를 숙지할 사건을 A, 기계에 오작동이 발생하는 사건을 B라 하면, $P(A) = 0.8$, $P(A^c) = 0.2$이고 $P(B|A) = 0.01$, $P(B|A^c) = 0.04$이다. 따라서 구하고자 하는 확률은 다음과 같다.

$$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c) = (0.8)(0.01) + (0.2)(0.04) = 0.016$$

15. 어느 공장에서 지난해 발생한 사고를 분석하니 다음과 같이 근로환경과 근로자의 부주의 그리고 기계의 결함에 의해 발생했다. 그리고 근로자의 근무시간인 1교대(09:00-17:00), 2교대(a7:00-01:00), 3교대(01:00-09:00)도 한 요인이다.

	근로환경	근로자의 부주의	기계의 결함
1교대	3%	23%	2%
2교대	5%	28%	1%
3교대	6%	29%	3%

지난해 발생한 사고 중에서 임의로 한 사례를 선정하였다.
(a) 이 사고가 각 근무 시간대에 의해 발생할 확률을 구하라.
(b) 근로자의 부주의에 의해 사고가 발생할 확률을 구하라.
(c) 근로자의 부주의에 의해 사고가 발생했을 때, 이 사고가 1교대 시간대에서 발생했을 확률을 구하라.

풀이

(a) 1교대, 2교대, 3교대에 의한 원인을 각각 A_1, A_2, A_3 그리고 근로환경과 근로자의 부주의 그리고 기계의 결함에 의한 원인을 각각 B, C, D라 하면, 각 근무 시간에 대한 확률은 다음과 같다.

$$P(A_1) = P(A_1 \cap B) + P(A_1 \cap C) + P(A_1 \cap D) = 0.03 + 0.23 + 0.02 = 0.28$$

$$P(A_2) = P(A_2 \cap B) + P(A_2 \cap C) + P(A_2 \cap D) = 0.05 + 0.28 + 0.01 = 0.34$$

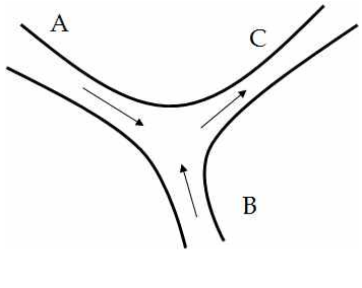
$$P(A_3) = P(A_3 \cap B) + P(A_3 \cap C) + P(A_3 \cap D) = 0.06 + 0.29 + 0.03 = 0.38$$

(b) 근로자의 부주의에 의해 발생한 사고는 각 시간대로 분할되므로 구하고자 하는 확률은 다음과 같다.

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) = 0.23 + 0.28 + 0.29 = 0.8$$

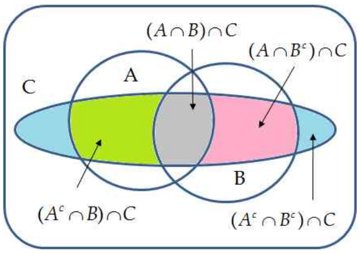
(c) 베이즈 정리에 의해 구하고자 하는 확률은 $P(A_1|B) = \dfrac{P(A_1 \cap B)}{P(B)} = \dfrac{0.23}{0.8} = 0.2875$이다.

16. 그림과 같이 두 고속도로 A와 B가 고속도로 C로 합류된다. A와 B는 동일한 교통량을 가지나 혼잡한 시간대에 고속도로 A와 B가 혼잡할 확률은 각각 0.2와 0.4이다. 그리고 고속도로 A가 혼잡할 때, 고속도로 B가 혼잡해질 확률은 0.7이다. 그리고 고속도로 B가 혼잡할 때, 고속도로 A가 혼잡해질 확률은 0.4이다. 또한 A와 B가 모두 혼잡하지 않을 때, 고속도로 C가 혼잡해질 확률은 0.3이다. 이때 고속도로 C가 혼잡해질 확률을 구하라.



풀이

고속도로 A, B 그리고 C가 혼잡해지는 사건을 각각 A, B, C라 하면 $P(A)=0.2$, $P(B)=0.4$ 그리고 $P(B|A)=0.7$, $P(A|B)=0.4$, $P(C|A^c \cap B)=0.3$이다. 그리고 사건 C는 그림과 같이 서로 배반인 $(A^c \cap B) \cap C$, $(A \cap B) \cap C$, $(A \cap B^c) \cap C$, $(A^c \cap B^c) \cap C$로 분할된다.



또한 다음 확률을 얻는다.

$$P(A^c \cap B) = P(B)P(A^c|B) = 0.4 \times (1-0.4) = 0.24$$

$$P(A \cap B) = P(B)P(A|B) = 0.4 \times 0.4 = 0.16$$

$$P(A \cap B^c) = P(A)P(B^c|A) = 0.2 \times (1-0.7) = 0.06$$

$$P(A^c \cap B^c) = 1 - 0.24 - 0.16 - 0.06 = 0.54$$

그러므로 고속도로 C가 혼잡해질 확률은 다음과 같다.

$$\begin{aligned} P(C) &= P[(A^c \cap B) \cap C] + P[(A \cap B) \cap C] + P[(A \cap B^c) \cap C] + P[(A^c \cap B^c) \cap C] \\ &= P(C|A^c \cap B)P(A^c \cap B) + P(C|A \cap B)P(A \cap B) \\ &\quad + P(C|A \cap B^c)P(A \cap B^c) + P(C|A^c \cap B^c)P(A^c \cap B^c) \\ &= 1 \times 0.24 + 1 \times 0.16 + 1 \times 0.06 + 0.3 \times 0.54 = 0.622 \end{aligned}$$

17. 어떤 보험회사의 생명보험 가입자 가운데 10%는 흡연가이고, 나머지는 비흡연가이다. 그리고 비흡연가가 올해 안에 사망할 확률은 1%이고, 흡연가가 사망할 확률은 5%라고 한다.

(a) 보험 가입자가 올해 안에 사망할 확률을 구하라.

(b) 보험 가입자가 사망했다고 할 때, 이 가입자가 흡연가일 확률을 구하라.

풀이

(a) 생명보험 가입자 가운데 흡연자가 선정될 사건을 A, 보험 가입자가 사망할 사건을 B라 하면, 문제 조건에 의하여 다음 확률을 얻는다.

$$P(A)=0.1, \quad P(A^c)=0.9, \quad P(B|A)=0.05, \quad P(B|A^c)=0.01$$

따라서 보험 가입자가 올해 안에 사망할 확률은 다음과 같다.

$$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c) = (0.1)(0.05) + (0.9)(0.01) = 0.014$$

(b) 보험 가입자가 사망했다고 할 때, 이 가입자가 흡연가일 확률은 다음과 같다.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{(0.1)(0.05)}{0.014} = 0.357$$

18. 의학 보고서에 따르면 전체 국민의 7%가 폐질환을 앓고 있으며, 그들 중 85%가 흡연가라고 한다. 그리고 폐질환을 갖지 않은 사람 중에 25%가 흡연가라 한다.
(a) 임의로 선정한 사람이 흡연가일 확률을 구하라.
(b) 임의로 선정한 사람이 흡연가라 할 때, 이 사람이 폐질환을 앓고 있을 확률을 구하라.

풀이

(a) 폐질환을 앓고 있는 사람을 A라 하면 $P(A)=0.07$이므로 여사건의 확률은 $P(A^c)=0.93$이다. 이제 임의로 선정한 사람이 흡연가일 사건을 B라 하면 $P(B|A)=0.85$, $P(B|A^c)=0.25$이다. 따라서 임의로 선정한 사람이 흡연가일 확률은 다음과 같다.

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

(b) 흡연가가 선정되었을 때, 이 사람이 폐질환을 앓고 있을 확률은 다음과 같다.

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

19. 250명의 성인을 임의로 선정하여 흡연과 고혈압의 관계를 연구하여 다음 결과를 얻었다.

	비흡연자(N)	적당한 흡연자(M)	심한 흡연자(H)
고혈압 환자(A)	27	39	43
저혈압 환자(B)	10	16	22
정상인(C)	53	28	12

(a) 임의로 한 사람을 선정했을 때, 이 사람이 심한 흡연자일 확률을 구하라.
(b) 임의로 한 사람을 선정했을 때, 이 사람이 고혈압 환자일 확률을 구하라.
(c) 선정한 사람이 심한 흡연자일 때, 이 사람이 고혈압 환자일 확률을 구하라.
(d) 선정한 사람이 고혈압 환자일 때, 이 사람이 심한 흡연자일 확률을 구하라.

풀이

(a) 임의로 선정한 사람이 비흡연자, 적당한 흡연자, 심한 흡연자인 사건을 각각 N, M, H 그리고 고혈압 환자, 저혈압 환자, 정상인인 사건을 각각 A, B, C라 하자. 그러면 이 사람이 심한 흡연자일 확률은 다음과 같다.

$$P(H)=P(H\cap A)+P(H\cap B)+P(H\cap C)=\frac{43}{250}+\frac{22}{250}+\frac{12}{250}=\frac{77}{250}$$

(b) 임의로 선정한 사람이 고혈압 환자일 확률은 다음과 같다.

$$P(A)=P(A\cap N)+P(A\cap M)+P(A\cap H)=\frac{27}{250}+\frac{39}{250}+\frac{43}{250}=\frac{109}{250}$$

(c) 선정한 사람이 심한 흡연자일 때, 이 사람이 고혈압 환자일 확률은 다음과 같다.

$$P(A|H) = \frac{P(H \cap A)}{P(H)} = \frac{43/250}{77/250} = \frac{43}{77}$$

(d) 선정한 사람이 고혈압 환자일 때, 이 사람이 심한 흡연자일 확률은 다음과 같다.

$$P(H|A) = \frac{P(H \cap A)}{P(A)} = \frac{43/250}{109/250} = \frac{43}{109}$$

20. 지난 5년 동안 어떤 단체에 가입한 사람을 대상으로 건강에 대한 연구가 이루어져 왔다. 이 연구의 초기에 흡연의 정도에 따라 담배를 많이 피우는 사람과 적게 피우는 사람 그리고 전혀 담배를 피우지 않는 사람의 비율이 각각 20%, 30% 그리고 50%이었다. 연구가 끝난 5년 동안에, 담배를 적게 피우는 사람은 전혀 피우지 않는 사람의 두 배가 사망하였고 많이 피우는 사람에 비하여 $\frac{1}{2}$만이 사망하였다는 결과를 얻었다. 이 연구의 대상인 회원을 임의로 선정하였을 때, 이 회원이 연구 기간 안에 사망하였다. 이 회원이 담배를 많이 피우는 사람이었을 확률을 구하라.

풀이

H, L 그리고 N을 각각 담배를 많이 피우는 사람과 적게 피우는 사람 그리고 전혀 담배를 피우지 않는 사람이 선정될 사건이라고 하자. 그리고 이 기간에 사망했을 사건을 D라고 하면 다음을 얻는다.

$$P(H) = 0.2,\ P(L) = 0.3,\ P(N) = 0.5,\quad P(D|L) = 2P(D|N),\quad P(D|L) = \frac{P(D|H)}{2}$$

이다. 따라서 구하고자 하는 확률은 베이즈 정리에 의하여 다음과 같다.

$$\begin{aligned} P(H|D) &= \frac{P(D|H)\,P(H)}{P(D|H)\,P(H) + P(D|L)\,P(L) + P(D|N)\,P(N)} \\ &= \frac{2P(D|L)(0.2)}{2P(D|L)(0.2) + P(D|L)(0.3) + (1/2)P(D|L)(0.5)} \\ &= \frac{0.4}{0.4 + 0.3 + 0.25} = 0.4211 \end{aligned}$$

4장

1. 자동화 조립 라인에서 생산되는 진공관의 길이를 주기적으로 측정한다. 무작위로 선정된 진공관 4개에 포함된 불량품의 개수는 다음과 같은 확률질량함수를 갖는다.

$$f(x) = \begin{cases} \dfrac{4!}{x!\,(4-x)!}(0.98)^x\,(0.02)^{4-x}, & x = 0,\,1,\,2,\,3,\,4 \\ 0 & ,\ \text{다른 곳에서} \end{cases}$$

(a) 선정된 4개의 진공관에 포함된 불량품 수에 대한 확률표를 만들어라.
(b) 불량품 수에 대한 평균과 분산을 구하라.
(c) 불량품이의 개수가 2개 이상일 확률을 구하라.

풀이

(a) 선정된 4개의 진공관에 포함된 불량품 수에 대한 확률표는 다음과 같다.

x	0	1	2	3	4
$f(x)$	$\frac{1}{6250000}$	$\frac{196}{6250000}$	$\frac{14406}{6250000}$	$\frac{470596}{6250000}$	$\frac{5764801}{6250000}$

(b) $x\,f(x)$와 $x^2 f(x)$을 구하면 다음 표와 같다.

x	0	1	2	3	4
x^2	0	1	4	9	16
$x\,f(x)$	0	$\frac{196}{6250000}$	$\frac{28812}{6250000}$	$\frac{1411788}{6250000}$	$\frac{23059204}{6250000}$
$x^2 f(x)$	0	$\frac{196}{6250000}$	$\frac{57624}{6250000}$	$\frac{4235364}{6250000}$	$\frac{92236816}{6250000}$

따라서 X의 평균과 2차 적률은 각각 다음과 같다.

$$\mu = E(X) = \frac{196}{6250000} + \frac{28812}{6250000} + \frac{1411788}{6250000} + \frac{23059204}{6250000} = \frac{98}{25}$$

$$E(X^2) = \frac{196}{6250000} + \frac{57624}{6250000} + \frac{4235364}{6250000} + \frac{92236816}{6250000} = \frac{9653}{625}$$

따라서 분산은 $\sigma^2 = E(X^2) - E(X)^2 = \frac{9653}{625} - \left(\frac{98}{25}\right)^2 = \frac{49}{625}$이다.

(c) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(X=0) + P(X=1)] = 1 - \left(\frac{1}{6250000} + \frac{196}{6250000}\right) = \frac{6249803}{6250000}$

2. 상점에서 판매되고 있는 핸드폰 25개 중에 불량품이 3개 포함되어 있다. 이 상점에서 구입한 핸드폰 2개에 포함된 불량품의 수를 확률변수 X라 할 때, 확률질량함수와 평균 그리고 분산을 구하라.

풀이

핸드폰 25개 중에서 임의로 2개 선정하는 방법의 수는 $\binom{25}{2}=300$이고 불량품이 하나도 포함되지 않는 경우의 수는 $\binom{3}{0}\binom{22}{2}=1\times 231=231$이다. 그리고 불량품이 하나 포함되는 경우의 수는 $\binom{3}{1}\binom{22}{1}=3\times 22=66$이고, 두 개 모두 불량품인 경우의 수는 $\binom{3}{2}\binom{22}{0}=3\times 1=3$이다. 따라서 다음 확률을 얻는다.

X	0	1	2
$P(X=x)$	$\frac{231}{300}=\frac{77}{100}$	$\frac{66}{300}=\frac{22}{100}$	$\frac{3}{300}=\frac{1}{100}$

따라서 평균과 X^2의 기댓값은 다음과 같다.

$$\mu=E(X)=0\times\frac{77}{100}+1\times\frac{22}{100}+2\times\frac{1}{100}=\frac{24}{100}=\frac{6}{25}$$

$$E(X^2)=0^2\times\frac{77}{100}+1^2\times\frac{22}{100}+2^2\times\frac{1}{100}=\frac{26}{100}=\frac{13}{50}$$

그러므로 분산은 다음과 같다.

$$\sigma^2=E(X^2)-\mu^2=\frac{13}{50}-\left(\frac{6}{25}\right)^2=\frac{253}{1250}$$

3. 어느 지역의 특정 시간대에 톨게이트를 지나는 자동차 수의 분포는 다음과 같다.

X	1	2	3	4	5	6	7	8	9	10
$P(X=x)$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{5}$		$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{20}$	$\frac{1}{20}$

(a) $P(X=6)$을 구하라.
(b) X의 평균과 분산을 구하라.
(c) 이 시간대에 7대 이상의 자동차가 지나갈 확률을 구하라.
(d) $Y=2X+1$의 평균과 분산을 구하라.

풀이

(a) $\sum_{x=1}^{10}P(X=x)=1$이고 $P(X=x)$를 모두 더하면 $\frac{4}{5}$이므로 $P(X=6)=\frac{1}{5}$이다.

(b) X의 평균과 2차 적률은 각각 다음과 같다.

$$\mu = E(x) = 1 \times \frac{1}{20} + 2 \times \frac{1}{20} + \cdots + 9 \times \frac{1}{20} + 10 \times \frac{1}{20} = \frac{11}{2}$$

$$E(X^2) = 1^2 \times \frac{1}{20} + 2^2 \times \frac{1}{20} + \cdots + 9^2 \times \frac{1}{20} + 10^2 \times \frac{1}{20} = \frac{353}{10}$$

따라서 분산은 $\sigma^2 = E(X^2) - E(X)^2 = \frac{101}{20}$이다.

(c) $P(X \geq 7) = \frac{1}{10} + \frac{1}{10} + \frac{1}{20} + \frac{1}{20} = \frac{3}{10}$

(d) $Y = 2X + 1$의 평균과 분산은 각각 다음과 같다.

$$E(Y) = E(2X+1) = 2E(X) + 1 = 12$$

$$Var(Y) = Var(2X+1) = 4\,Var(X) = \frac{706}{5}$$

4. 룰렛 게임은 빨간색 18개 숫자, 검은색 18개 숫자 그리고 0, 00으로 구성된다. 어떤 게이머가 빨간색 숫자에 1,000원 걸고, 빨간색이 나오면 1,000원을 벌고 그렇지 않으면 1,000원을 잃는다. 이 게이머의 기대금액을 구하라.

풀이

전체 38개의 숫자가 있으며 빨간색 숫자가 나올 확률은 $\frac{9}{19}$이고 다른 색의 숫자가 나올 확률은 $\frac{10}{19}$이다. 빨간색 숫자가 나오면 1,000원을 벌고 그렇지 않으면 1,000원을 잃으므로 이 게이머의 기대 금액은 다음과 같다.

$$E(X) = 1000 \times \frac{9}{19} + (-1000) \times \frac{10}{19} = -\frac{1000}{19} = -52.6(\text{원})$$

5. 어떤 기계의 수명은 확률밀도함수 $f(x)$가 $0 \leq x \leq 25$에서 $(x+5)^{-2}$에 비례한다. 단, 단위는 년이다.
(a) 비례상수를 구하라.
(b) 이 기계를 10년 이상 사용할 확률을 구하라.
(c) 이 기계를 5년 이상 15년 이하로 사용할 확률을 구하라.

풀이

(a) $0 \leq x \leq 25$에서 확률밀도함수 $f(x)$가 $(x+5)^{-2}$에 비례하므로 비례상수를 k라 하면, 다음이 성립한다.

$$\int_0^{25} \frac{k}{(x+5)^2}\,dx = \left[-\frac{k}{x+5} \right]_0^{25} = \frac{k}{6} = 1$$

따라서 비례상수는 $k = 6$이다.

(b) 10년 이상 사용할 확률은 다음과 같다.

$$P(X \geq 10) = \int_{10}^{25} \frac{6}{(x+5)^2}\, dx = \left[-\frac{6}{x+5} \right]_{10}^{25} = \frac{1}{5}$$

(c) 5년 이상 15년 이하로 사용할 확률은 다음과 같다.

$$P(5 \leq X \leq 15) = \int_{5}^{15} \frac{6}{(x+5)^2}\, dx = \left[-\frac{6}{x+5} \right]_{5}^{15} = \frac{3}{10}$$

6. 추적 장치의 위상오차는 다음 확률밀도함수를 갖는다.

$$f(x) = \begin{cases} \cos x\,, & 0 < x < \frac{\pi}{2} \\ 0 & , \text{ 다른 곳에서} \end{cases}$$

(a) 위상오차가 0과 $\frac{\pi}{6}$ 사이일 확률을 구하라.

(b) 위상오차가 $\frac{\pi}{3}$ 이상일 확률을 구하라.

(c) 평균 위상오차를 구하라.

풀이

(a) 위상오차가 0과 $\frac{\pi}{6}$ 사이일 확률은 다음과 같다.

$$P\left(0 < X < \frac{\pi}{4}\right) = \int_0^{\pi/4} \cos x\, dx = \left[\sin x\right]_0^{\pi/4} = \sin\frac{\pi}{4} = \frac{1}{\sqrt{2}}$$

(b) 위상오차가 $\frac{\pi}{3}$ 이상일 확률은 다음과 같다.

$$P\left(X \geq \frac{\pi}{3}\right) = \int_{\pi/3}^{\pi/2} \cos x\, dx = \left[\sin x\right]_{\pi/3}^{\pi/2} = \sin\frac{\pi}{2} - \sin\frac{\pi}{3} = 1 - \frac{\sqrt{3}}{2}$$

(c) 위상오차의 평균은 다음과 같다.

$$\mu = \int_0^{\pi/2} x \cos x\, dx = \left[\cos x + x \sin x\right]_0^{\pi/2} = \frac{\pi}{2} \sin\frac{\pi}{2} - 1 = \frac{1}{2}(\pi - 2)$$

7. 어떤 화학적 합성과정에서 발생하는 반응온도의 확률밀도함수는 -5°C에서 5°C에서 일정하게 나타난다.

(a) 확률밀도함수 $f(x)$를 구하라.

(b) $-5 < a < a+3 < 5$를 만족하는 실수 a에 대해 $P(a < X < a+3)$을 구하라.

풀이

(a) 반응온도를 X라 하면 확률밀도함수가 -5와 5 사이에서 일정하므로 다음과 같다.

$$f(x) = \begin{cases} \dfrac{1}{10}, & -5 < x < 5 \\ 0, & \text{다른 곳에서} \end{cases}$$

(b) 구하고자 하는 확률은 다음과 같다.

$$P(a < X < a+3) = \int_a^{a+3} \frac{1}{10}\,dx = \left[\frac{1}{10}x\right]_a^{a+3} = \frac{3}{10}.$$

8. 수문학에서 나타나는 확률변수에 대한 통계적 모형은 다음 분포함수로 나타난다.

$$F(x) = \begin{cases} 0, & x < 0 \\ \dfrac{x}{4}\left(1 + \ln\dfrac{4}{x}\right), & 0 \le x < 4 \\ 1, & x \ge 4 \end{cases}$$

(a) 확률밀도함수 $f(x)$를 구하라.

(b) $P(X \le 2)$를 구하라.

(c) $P(3 < X \le 4)$를 구하라.

풀이

(a) 확률변수 X의 확률밀도함수 $f(x)$는 다음과 같다.

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\left[\frac{x}{4}\left(1 + \ln\frac{4}{x}\right)\right] = \frac{1}{4}\ln\frac{4}{x},\ 0 < x < 4$$

(b) $P(X \le 2) = F(2) = \frac{1}{2}(1 + \ln 2) \approx 0.8466$

(c) $P(2 < X \le 4) = F(4) - F(2) = 1 - \frac{1}{2}(1 + \ln 2) = \frac{1}{2}(1 - \ln 2) \approx 0.1534$

9. 다음 확률밀도함수를 갖는 확률변수 X를 생각하자.

$$f(x)=\begin{cases}\frac{1}{4}, & 1 \le x \le 5 \\ 0, & \text{다른 곳에서}\end{cases}$$

(a) 분포함수 $F(x)$를 구하라.
(b) 확률 $P(2 \le X \le 4)$를 구하라.

풀이

(a) $x<1$이면 $F(x)=P(X \le x)=0$이고 $1 \le x < 5$이면 $F(x)$는 다음과 같다.

$$F(x)=\int_1^x \frac{1}{4}dx=\left[\frac{1}{4}u\right]_1^x=\frac{1}{4}(x-1)$$

그리고 $x \ge 5$이면 $F(x)$는 다음과 같다.

$$F(x)=\int_1^x \frac{1}{4}dx=\int_1^5 \frac{1}{4}dx+\int_5^x 0\,dx=\left[\frac{1}{4}u\right]_1^5=1$$

따라서 구하고자 하는 분포함수는 다음과 같다.

$$F(x)=\begin{cases}0 & , x<1 \\ \frac{1}{4}(x-1), & 1 \le x < 5 \\ 1 & , x \ge 5\end{cases}$$

(b) 분포함수를 이용하여 확률을 구하면 다음과 같다.

$$P(2 \le X \le 4)=F(4)-F(2)=\frac{1}{4}(3-1)=\frac{1}{2}$$

10. 생태학자들은 반경 10m인 원형인 영역에 선을 그어 표시하고자 한다. 그러나 실제 반경은 다음 확률밀도함수를 갖는 확률변수 X로 나타난다.

$$f(x)=\begin{cases}\frac{3}{10}(-x^2+20x-98), & 9<x<11 \\ 0 & , \text{다른 곳에서}\end{cases}$$

(a) 확률변수 X의 분포함수 $F(x)$를 구하라.
(b) $P(9.5<X<10.5)$를 구하라.

풀이

(a) $x<9$이면 $F(x)=0$이고, $x\geq 11$이면 $F(x)=1$이다. 그리고 $9\leq x<11$이면 $F(x)$는 다음과 같다.

$$F(x)=\int_9^x \frac{3}{10}(-u^2+20u-98)\,du=\left[-\frac{u^3}{10}+3u^2-\frac{147}{5}u\right]_9^x$$

$$=-\frac{x^3}{10}+3x^2-\frac{147}{5}x+\frac{189}{2}$$

(b) $P(9.5<X<10.5)=F(10.5)-F(9.5)=\frac{63}{80}-\frac{17}{80}=\frac{23}{40}$

11. 핸드폰에 사용되는 주요 부품의 수명은 다음 확률밀도함수를 갖는다. 단위는 시간이다.

$$f(x)=\begin{cases}\frac{1}{3000}e^{-x/3000}, & x>0\\ 0, & \text{다른 곳에서}\end{cases}$$

(a) X의 분포함수 $F(x)$를 구하라.

(b) 이 핸드폰을 2000시간 이상 사용할 확률을 구하라.

풀이

(a) $x<0$이면 분포함수는 $F(x)=0$이고 $x\geq 0$이면 다음과 같다.

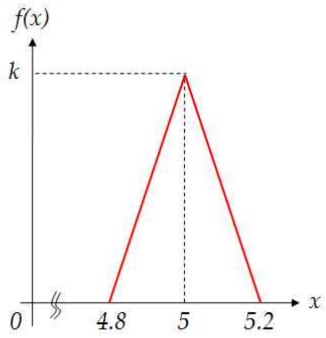
$$F(x)=\int_0^x \frac{1}{3000}e^{-u/3000}\,du=\left[-e^{-u/3000}\right]_0^x=1-e^{-x/3000}$$

그러므로 구하고자 하는 분포함수는 다음과 같다.

$$F(x)=\begin{cases}0, & x<0\\ 1-e^{-x/3000}, & x\geq 0\end{cases}$$

(b) $P(X\geq 2000)=1-F(2000)=1-\left(1-e^{-2000/3000}\right)=e^{-2/3}$

12. 테스터기를 이용하여 통신신호의 전압 X를 측정하며, 측정에는 항상 오차가 발생한다. 통신신호의 정격전압은 5볼트이며 X의 확률밀도함수는 다음 그림과 같다.



(a) 상수 k를 구하라.
(b) 확률밀도함수 $f(x)$를 구하라.
(c) 확률 $P(-4.9 \le X \le 5.1)$를 구하라.
(d) 확률 $P(X \ge 5.15)$를 구하라.

풀이

(a) 밑면의 길이가 0.4, 높이 k인 이등변삼각형의 넓이가 1이어야 하므로 다음이 성립한다.

$$\frac{1}{2}(0.4 \times k) = 1;\ k = 5$$

(b) $x < 4.8$ 또는 $x > 5.2$이면 $f(x) = 0$이다. $4.8 \le x < 5$이면 $(4.8, 0)$과 $(5, 5)$를 지나는 직선이므로 $f(x) = 25x - 120$이고 $5 \le x < 5.2$이면 $(5.2, 0)$과 $(5, 5)$를 지나는 직선이므로 $f(x) = -25x + 130$이다. 그러므로 구하고자 하는 확률밀도함수는 다음과 같다.

$$f(x) = \begin{cases} 25x - 120, & 4.8 \le x < 5 \\ -25x + 130, & 5 \le x \le 5.2 \\ 0, & \text{다른 곳에서} \end{cases}$$

(c) 확률밀도함수가 $x = 5$를 중심으로 대칭이므로 구하고자 하는 확률은 다음과 같다.

$$P(-4.9 \le X \le 5.1) = 2P(-4.9 \le X \le 5) = 2\int_{-4.9}^{5} (25x - 120)\,dx$$

$$= 2\left[\frac{25}{2}x^2 - 120x\right]_{-4.9}^{5} = 0.75$$

(d) 구하고자 하는 확률은 다음과 같다.

$$P(X \ge 5.15) = \int_{5.15}^{5.2} (-25x + 130)\,dx = \left[-\frac{25}{2}x^2 + 130x\right]_{5.15}^{5.2}$$

$$= 338 - 337.96875 = 0.03125$$

13. 미로 상자 안에 있는 쥐가 음식이 있는 곳까지 가는데 걸리는 시간은 다음 확률밀도를 갖는 확률변수 X로 나타난다. 단, 단위는 초이다.

$$f(x) = \begin{cases} \dfrac{15}{x^2}, & x > 15 \\ 0, & \text{다른 곳에서} \end{cases}$$

(a) 확률변수 X의 분포함수 $F(x)$를 구하라.
(b) $P(20 < X < 30)$을 구하라.
(c) 기댓값이 존재하지 않음을 보여라.

풀이

(a) $x < 15$이면 $F(x) = 0$이고, $x \geq 15$이면 $F(x)$는 다음과 같다.

$$F(x) = \int_{15}^{x} \frac{15}{u^2}\, du = \left[-\frac{15}{u} \right]_{15}^{x} = 1 - \frac{15}{x}$$

(b) $P(20 < X < 30) = F(30) - F(20) = \left(1 - \dfrac{15}{30}\right) - \left(1 - \dfrac{15}{20}\right) = \dfrac{1}{4}$

(c) p-급수 판정법에 의해 다음 적분 결과는 수렴하지 않으므로 X의 기댓값은 존재하지 않는다.

$$\int_{15}^{\infty} x f(x)\, dx = \int_{15}^{\infty} x \left(\frac{15}{x^2}\right) dx = 15 \int_{15}^{\infty} \frac{1}{x}\, dx = \infty$$

14. 장거리 전화통화 시간 X는 다음과 같은 확률밀도함수를 갖는다고 한다.

$$f(x) = \begin{cases} \dfrac{1}{10} e^{-x/10}, & x > 0 \\ 0, & \text{다른 곳에서} \end{cases}$$

(a) X의 기대값과 분산을 구하라.
(b) $P(\mu - \sigma \leq X \leq \mu + \sigma)$와 $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$를 구하라.

풀이

(a) X의 평균과 2차 적률을 구하면 각각 다음과 같다.

$$\mu = E(X) = \frac{1}{10} \int_{0}^{\infty} x e^{-x/10}\, dx = \left[-(x+10) e^{-x/10} \right]_{0}^{\infty} = 10$$

$$E(X^2) = \frac{1}{10} \int_{0}^{\infty} x^2 e^{-x/10}\, dx = \left[-(x^2 + 20x + 200) e^{-x/10} \right]_{0}^{\infty} = 200$$

따라서 구하고자 하는 분산은 $\sigma^2 = E(X^2) - [E(X)]^2 = 200 - 10^2 = 100$이다.

(b) $\sigma^2 = 100$이므로 표준편차는 $\sigma = 10$이고, 따라서 $\mu - \sigma = 0$, $\mu + \sigma = 20$, $\mu - 2\sigma = -10$, $\mu + 2\sigma = 30$이다. 그러므로 구하고자 하는 확률은 각각 다음과 같다.

$$P(\mu - \sigma \le X \le \mu + \sigma) = P(0 \le X \le 20) = \int_0^{20} \frac{1}{10} e^{-x/10} dx = \left[-e^{-x/10}\right]_0^{20} = 1 - e^{-2}$$

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = P(0 \le X \le 30) = \int_0^{30} \frac{1}{10} e^{-x/10} dx = \left[-e^{-x/10}\right]_0^{30} = 1 - e^{-3}$$

15. 고속도로의 특정 지점에 설치된 과속 카메라에 단속되는 승용차 사이의 시간은 다음 분포함수를 갖는다. 단, 단위는 분이다.

$$F(x) = \begin{cases} 0 & , \ x < 0 \\ 1 - e^{-x/10} & , \ x \ge 0 \end{cases}$$

(a) 확률밀도함수 $f(x)$를 구하라.
(b) 평균과 분산을 구하라.
(c) 단속 차량 사이의 시간이 15분 이상일 확률을 구하라.

풀이

(a) $x > 0$이면 $f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}(1 - e^{-x/10}) = 10e^{-x/10}$이므로 확률밀도함수 $f(x)$는 다음과 같다.

$$f(x) = \begin{cases} \frac{1}{10} e^{-x/10} & , \ x > 0 \\ 0 & , \ \text{다른 곳에서} \end{cases}$$

(b) 평균과 2차 적률은 각각 다음과 같다.

$$\mu = E(X) = \int_0^{\infty} \frac{1}{10} x e^{-x/10} dx = \left[-e^{-x/10}(x + 10)\right]_0^{\infty} = 10$$

$$E(X^2) = \int_0^{\infty} \frac{1}{10} x^2 e^{-x/10} dx = \left[-e^{-x/10}(x^2 + 20x + 200)\right]_0^{\infty} = 200$$

따라서 분산은 $\sigma^2 = E(X^2) - E(X)^2 = 200 - 10^2 = 100$이다.

(c) $P(X > 15) = \int_{15}^{\infty} \frac{1}{10} e^{-x/10} dx = \left[-e^{-x/10}\right]_{15}^{\infty} = e^{-3/2}$

16. 나사의 피치 지름의 코드측정은 다음 확률밀도함수를 갖는다.

$$f(x)=\begin{cases} \dfrac{k}{\pi(1+x^2)}, & 0 \le x \le 1 \\ 0, & \text{다른 곳에서} \end{cases}$$

(a) 상수 k를 구하라.

(b) X의 평균과 분산을 구하라.

풀이

(a) $\displaystyle\int_0^1 \frac{k}{\pi(1+x^2)}\,dx = \left[\frac{k}{\pi}\tan^{-1}x\right]_0^1 = \frac{k}{4} = 1$이므로 $k=4$이다.

(b) X의 평균과 2차 적률은 각각 다음과 같다.

$$\mu = E(X) = \int_0^1 \frac{4}{\pi}\frac{x}{1+x^2}\,dx = \left[\frac{2}{\pi}\ln(1+x^2)\right]_0^1 = \frac{2\ln 2}{\pi}$$

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

따라서 분산은 $\sigma^2 = E(X^2)-E(X)^2 = \dfrac{4}{\pi}-1-\dfrac{4(\ln 2)^2}{\pi^2}$이다.

17. 연속확률변수 X의 확률밀도함수가 $f(x)=\dfrac{2}{3}(x+1)$, $0<x<1$이다.

(a) X의 분포함수 $F(x)$를 구하라.

(b) $P(X\le q_1)=\dfrac{1}{4}$, $P(X\le q_2)=\dfrac{1}{2}$, $P(X\le q_3)=\dfrac{3}{4}$을 만족하는 q_1, q_2, q_3을 구하라. 이때 q_1, q_2, q_3을 X의 사분위수라 한다.

풀이

(a) $x<0$이면 분포함수는 $F(x)=0$이고 $0\le x<1$이면 다음과 같다.

$$F(x) = \int_0^x \frac{2}{3}(u+1)\,du = \left[\frac{1}{3}(u^2+2u)\right]_0^x = \frac{1}{3}(x^2+2x)$$

그리고 $x\ge 1$이면 $F(x)=1$이다. 그러므로 구하고자 하는 분포함수는 다음과 같다.

$$F(x)=\begin{cases} 0, & x<0 \\ \dfrac{1}{3}(x^2+2x), & 0\le x<1 \\ 1, & x\ge 1 \end{cases}$$

(b) $P(X \le q_1) = F(q_1) = \frac{1}{3}(q_1^2 + 2q_1) = \frac{1}{4}$; $4q_1^2 + 8q_1 - 3 = 0$; $q_1 = \frac{1}{2}(-2 + \sqrt{7})$

$P(X \le q_2) = F(q_2) = \frac{1}{3}(q_2^2 + 2q_2) = \frac{1}{2}$; $2q_2^2 + 4q_2 - 3 = 0$; $q_2 = \frac{1}{2}(-2 + \sqrt{10})$

$P(X \le q_3) = F(q_3) = \frac{1}{3}(q_3^2 + 2q_3) = \frac{3}{4}$; $4q_1^2 + 8q_1 - 9 = 0$; $q_3 = \frac{1}{2}(-2 + \sqrt{13})$

18. 임의의 확률변수 X에 대하여 표준화 확률변수의 평균과 분산은 각각 $E(Z) = 0$, $Var(Z) = 1$인 것을 보여라.

풀이

임의의 확률변수 X에 대하여 표준화 확률변수는 $Z = \frac{X-\mu}{\sigma}$이므로 Z의 평균과 분산은 각각 다음과 같다.

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}(E(X) - \mu) = 0$$

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(X-\mu) = \frac{1}{\sigma^2}Var(X) = \frac{1}{\sigma^2} \times \sigma^2 = 1$$

19. 확률변수 X의 평균과 분산은 각각 $\mu = 5$, $\sigma^2 = 4$이다. 다음 확률의 상한 또는 하한을 구하라.

(a) $P(|X-5| \ge 3)$ (b) $P(|X-5| < 4)$ (c) $P(0 < X < 10)$

(d) $P(|X-5| \ge \alpha) \le 0.0016$를 만족하는 상수 k를 구하라.

풀이

(a) $\mu = 5$, $\sigma = 2$이므로 체비쇼프 부등식에 의하여 다음을 얻는다.

$$P(|X-5| \ge 3) = P\left(|X-\mu| \ge \frac{3}{2}\sigma\right) \le \frac{1}{(3/2)^2} = \frac{4}{9}$$

(b) $P(|X-5| < 4) = P(|X-\mu| < 2\sigma) \ge 1 - \frac{1}{2^2} = \frac{3}{4}$

© $P(0 < X < 10) = P(-5 < X-5 < 5) = P(|X-5| < 5) = P\left(|X-5| < \frac{5}{2}\sigma\right)$

$$\ge 1 - \frac{1}{(5/2)^2} = \frac{21}{25}$$

(d) $\alpha = k\sigma$라 하면 $P(|X-5| \ge \alpha) \le 0.0016 = \frac{1}{25^2}$이므로 $k = 25$이고 $\sigma = 2$이다. 그러므로 구하고자 하는 상수는 $\alpha = k\sigma = 50$이다.

20. 용접기의 수명 X는 $x > 0$에서 확률밀도함수 $f(x) = \frac{1}{2}e^{-x/2}$를 갖는다. 용접기 수명의 왜도와 첨도를 구하라.

풀이

먼저 평균과 분산을 구하기 위해 X의 기댓값과 2차 적률을 구한다.

$$\mu = E(X) = \int_0^\infty \frac{1}{2} x e^{-x/2} dx = \left[-(x+2)e^{-x/2}\right]_0^\infty = 2$$

$$E(X^2) = \int_0^\infty \frac{1}{2} x^2 e^{-x/2} dx = \left[-(x^2+4x+8)e^{-x/2}\right]_0^\infty = 8$$

따라서 X의 평균과 분산은 각각 $\mu = 2$, $\sigma^2 = 8 - 2^2 = 4$이고 표준편차는 $\sigma = 2$이다. 또한 3차, 4차 중심적률은 각각 다음과 같다.

$$E[(X-\mu)^3] = \frac{1}{2}\int_0^\infty (x-2)^3 e^{-x/2} dx = \left[-(x^3+12x+16)e^{-x/2}\right]_0^\infty = 16$$

$$E[(X-\mu)^4] = \frac{1}{2}\int_0^\infty (x-2)^4 e^{-x/2} dx = \left[-(x^4+24x^2+64x+144)e^{-x/2}\right]_0^\infty = 144$$

따라서 왜도 s와 첨도 k는 각각 다음과 같다.

$$s = \frac{E[(X-\mu)^3]}{\sigma^3} = \frac{16}{8} = 2, \qquad k = \frac{E[(X-\mu)^4]}{\sigma^4} = \frac{144}{16} = 9$$

5장

1. 확률변수 X와 Y의 주변확률질량함수와 조건부 확률질량함수가 각각 다음과 같다.

$$f_X(x) = \begin{cases} 0.4, & x=0 \\ 0.6, & x=1 \\ 0 & , \text{다른 곳에서} \end{cases}$$

$$f_{Y|X}(y|0) = \begin{cases} 0.6, & y=0 \\ 0.4, & y=1 \\ 0 & , \text{다른 곳에서} \end{cases}$$

$$f_{Y|X}(y|1) = \begin{cases} 0.4, & y=0 \\ 0.6, & y=1 \\ 0 & , \text{다른 곳에서} \end{cases}$$

(a) X와 Y의 결합확률표를 작성하라.

(b) X와 Y의 독립성을 조사하라.

풀이

(a) X와 Y의 결합확률을 구하면 다음과 같다.

$$P(X=0,\ Y=0) = f_X(0)f_{Y|X}(0|0) = 0.4 \times 0.6 = 0.24$$

$$P(X=0,\ Y=1) = f_X(0)f_{Y|X}(1|0) = 0.4 \times 0.4 = 0.16$$

$$P(X=1,\ Y=0) = f_X(1)f_{Y|X}(0|1) = 0.6 \times 0.4 = 0.24$$

$$P(X=1,\ Y=1) = f_X(1)f_{Y|X}(1|1) = 0.6 \times 0.6 = 0.36$$

따라서 X와 Y의 결합확률표는 다음과 같다.

X \ Y	0	1	$f_X(x)$
0	0.24	0.16	0.4
1	0.24	0.36	0.6
$f_Y(y)$	0.48	0.52	0.1

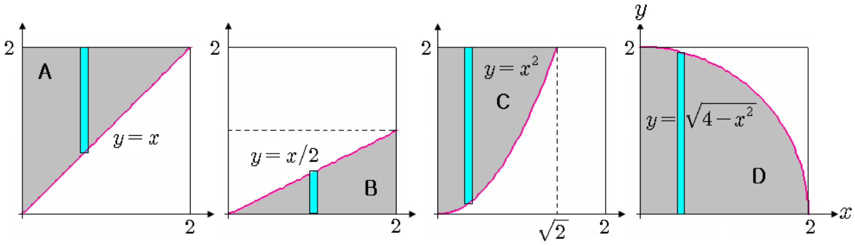
$x=0, 1,\ y=0, 1$에 대해 $f(0,0) = 0.24 \neq f_X(0)f_Y(0) = 0.4 \times 0.48$이므로 X와 Y의 독립이 아니다.

2. 연속확률변수 X, Y의 결합밀도함수가 $f(x, y) = \dfrac{x+y}{8}$, $0 < x < 2$, $0 < y < 2$일 때 다음을 구하라.

(a) $P(X \le Y)$ (b) $P(X \ge 2Y)$ (c) $P(Y \ge X^2)$ (d) $P(X^2 + Y^2 \le 4)$

풀이

각 경우의 (x, y) 영역은 그림과 같다.



따라서 구하고자 하는 확률은 각각 다음과 같다.

(a) $P(X \le Y) = \displaystyle\int_0^2 \int_x^2 \frac{x+y}{8}\, dy\, dx = \frac{1}{2}$

(b) $P(X \ge 2Y) = \displaystyle\int_0^2 \int_0^{x/2} \frac{x+y}{8}\, dy\, dx = \frac{5}{24}$

(c) $P(Y \ge X^2) = \displaystyle\int_0^{\sqrt{2}} \int_{x^2}^2 \frac{x+y}{8}\, dy\, dx = \frac{5+8\sqrt{2}}{40}$

(d) $P(X^2 + Y^2 \le 4) = \displaystyle\int_0^2 \int_0^{\sqrt{4-x^2}} \frac{x+y}{8}\, dy\, dx = \frac{2}{3}$

3. 이산확률변수 X와 Y의 결합확률질량함수가 다음과 같다.

$$f(x, y) = \frac{3^x 2^y}{x!\, y!} e^{-5}, \quad x = 0, 1, 2, \cdots, \quad y = 0, 1, 2, \cdots$$

(a) X와 Y의 주변확률질량함수를 구하라.
(b) X와 Y의 독립성을 조사하라.
(c) 확률 $P(X \ge 1, Y \ge 1)$을 구하라.

풀이

(a) 메클로린 급수에 의하여 $e^m = \displaystyle\sum_{x=0}^{\infty} \frac{m^x}{x!}$이므로 $e^2 = \displaystyle\sum_{y=0}^{\infty} \frac{2^y}{y!}$이다. 따라서 X와 Y의 주변확률질량함수는 각각 다음과 같다.

$$f_X(x) = \sum_{y=0}^{\infty} \frac{3^x 2^y}{x!\, y!} e^{-5} = \frac{3^x}{x!} e^{-5} \sum_{y=0}^{\infty} \frac{2^y}{y!} = \frac{3^x}{x!} e^{-5} \times e^2 = \frac{3^x}{x!} e^{-3}, \quad x = 0, 1, 2, \cdots,$$

$$f_Y(y) = \sum_{x=0}^{\infty} \frac{3^x 2^y}{x!\, y!} e^{-5} = \frac{2^y}{y!} e^{-5} \sum_{x=0}^{\infty} \frac{3^x}{x!} = \frac{2^y}{y!} e^{-5} \times e^3 = \frac{2^y}{y!} e^{-2}, \quad y = 0, 1, 2, \cdots$$

(b) $x=0,\ 1,\ 2,\ \cdots,\ y=0,\ 1,\ 2,\ \cdots$에 대해 다음이 성립하므로 X와 Y는 독립이다.

$$f(x,y)=\frac{3^x 2^y}{x!\,y!}e^{-5}=f_X(x)\cdot f_Y(y)=\left(\frac{3^x}{x!}e^{-3}\right)\left(\frac{2^y}{y!}e^{-2}\right)$$

(c) X와 Y는 독립이므로 $P(X\ge 1,\ Y\ge 1)=P(X\ge 1)P(Y\ge 1)$이고 다음을 얻는다.

$$P(X\ge 1)=1-P(X=0)=1-\frac{3^0}{0!}e^{-3}=1-e^{-3}$$
$$P(Y\ge 1)=1-P(Y=0)=1-\frac{2^0}{0!}e^{-2}=1-e^{-2}$$

따라서 구하고자 하는 확률은 $P(X\ge 1,\ Y\ge 1)=(1-e^{-3})(1-e^{-2})$이다.

4. 이산확률변수 X와 Y의 결합질량함수가 다음과 같다.

$$f(x,y)=\frac{x+y}{110},\quad x=1,\ 2,\ 3,\ 4,\ \ y=1,\ 2,\ 3,\ 4,\ 5$$

(a) X와 Y의 주변질량함수를 구하라.
(b) $E(X),\ E(Y)$를 구하라.
(c) 확률 $P(X<Y)$를 구하라.
(d) 확률 $P(3\le X+Y\le 5)$를 구하라.

풀이

(a) X와 Y의 주변질량함수는 각각 다음과 같다.

$$f_X(x)=\sum_{y=1}^{5}\frac{x+y}{110}=\frac{x+3}{22},\quad x=1,\ 2,\ 3,\ 4$$

$$f_Y(y)=\sum_{x=1}^{4}\frac{x+y}{110}=\frac{2y+5}{55},\quad y=1,\ 2,\ 3,\ 4,\ 5$$

(b) X와 Y의 기댓값은 각각 다음과 같다.

$$E(X)=\sum_{x=1}^{4}\frac{x(x+3)}{22}=\frac{4}{22}+\frac{10}{22}+\frac{18}{22}+\frac{28}{22}=\frac{30}{11}$$

$$E(Y)=\sum_{y=1}^{5}\frac{y(2y+5)}{55}=\frac{7}{55}+\frac{18}{55}+\frac{33}{55}+\frac{52}{55}+\frac{75}{55}=\frac{37}{11}$$

(c) $P(X<Y)=\left(\sum_{y=2}^{5}f(1,y)\right)+\left(\sum_{y=3}^{5}f(2,y)\right)+f(3,4)+f(3,5)+f(4,5)=\frac{6}{11}$

(d) $P(3 \le X+Y \le 5) = f(1,2)+f(2,1)+f(1,3)+f(2,2)+f(3,1)+P(X+Y=5) = \dfrac{3}{11}$

5. 확률변수 X와 Y의 결합확률밀도함수가 $f(x,y) = 60xy^2$, $0<x<1$, $0<y<1$, $x+y<1$이다.

(a) 확률 $P(0<X<0.5,\ 0<Y<0.5)$를 구하라.

(b) 확률 $P(0<X<0.75,\ 0<Y<0.5)$를 구하라.

(c) 확률 $P(X+Y<0.5)$를 구하라.

풀이

(a) 구하고자 하는 확률은 다음과 같다(그림 (a) 참조).

$$P(0<X<0.5,\ 0<Y<0.5) = \int_0^{0.5}\int_0^{0.5} 60xy^2\,dy\,dx = \int_0^{0.5}\left[20xy^3\right]_{y=0}^{0.5} dx$$

$$= \int_0^{0.5} \frac{5}{2}x\,dx = \left[\frac{5}{4}x^2\right]_0^{0.5} = \frac{5}{16}$$

(b) 구하고자 하는 확률은 다음과 같다(그림 (b) 참조).

$$P(0<X<0.75,\ 0<Y<0.5) = \int_0^{0.5}\int_0^{0.5} 60xy^2\,dy\,dx + \int_{0.5}^{0.75}\int_0^{1-x} 60xy^2\,dy\,dx$$

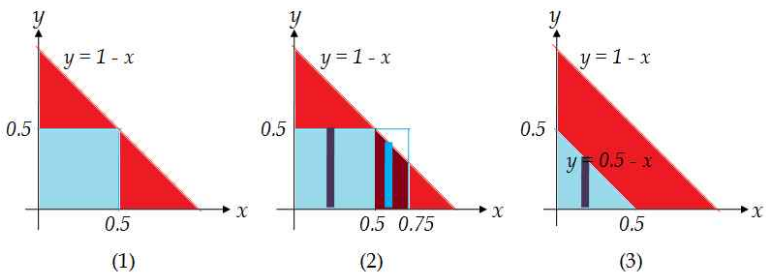
$$= \frac{5}{16} + \int_{0.5}^{0.75}\int_0^{1-x}\left[20xy^3\right]_{y=0}^{y=1-x} dy\,dx$$

$$= \frac{5}{16} + \int_{0.5}^{0.75} 20x(1-x)^3\,dx = \frac{5}{16} + \frac{11}{64} = \frac{31}{64}$$

(c) 구하고자 하는 확률은 다음과 같다(그림 (c) 참조).

$$P(X+Y<0.5) = \int_0^{0.5}\int_0^{0.5-x} 60xy^2\,dy\,dx = \int_0^{0.5}\left[20xyy^3\right]_{y=0}^{0.5-x} dx$$

$$= \int_0^{0.5}\frac{5}{2}x(1-2x)^3\,dx = \left[\frac{1}{4}(5-20x+30x^2-16x^3)\right]_0^{0.5} = \frac{1}{32}$$



6. 두 확률변수 X와 Y의 평균은 각각 -2, 2이고 분산은 각각 1, 3이고 $E(XY)=3$이다.
(a) $E(X+Y)$를 구하라.
(b) X와 Y의 공분산 $Cov(X, Y)$를 구하라.
(c) $Var(X+Y)$를 구하라.

풀이

(a) 두 확률변수 X와 Y에 $E(X+Y)=E(X)+E(Y)=-2+2=0$이다.

(b) X와 Y의 공분산 $Cov(X, Y)=E(XY)-E(X)E(Y)=3-(-2)\times 2=7$이다.

(c) 두 확률변수 X와 Y에 대하여 $Var(X+Y)=Var(X)+Var(Y)+2Cov(X, Y)$이므로
$Var(X+Y)=11$이다.

7. 매일 인접한 두 도시 A와 B의 교통사고 발생 시간을 조사한 결과, 두 도시의 교통사고 발생 시간을 각각 X와 Y라 할 때, 다음 결합밀도함수를 갖는다고 한다.

$$f(x, y)=\begin{cases} ke^{-(x+2y)}, & x>0,\ y>0 \\ 0, & \text{다른 곳에서} \end{cases}$$

(a) 상수 k를 구하라.
(b) X와 Y의 주변확률밀도함수를 구하라.
(c) X와 Y의 결합분포함수를 구하라.
(d) X와 Y의 독립성을 조사하라.
(e) 확률 $P(X\le 1,\ Y\le 1)$을 구하라.

풀이

(a) $f(x,y)$가 결합밀도함수이므로 다음이 성립해야 한다.

$$\int_0^\infty \int_0^\infty f(x,y)\,dy\,dx=\int_0^\infty \int_0^\infty ke^{-(x+2y)}\,dy\,dx=\int_0^\infty \left[-\frac{k}{2}e^{-x}e^{-2y}\right]_{y=0}^{y=\infty} dx$$

$$=\int_0^\infty \frac{k}{2}e^{-x}\,dx=\left[-\frac{k}{2}e^{-x}\right]_{x=0}^{x=\infty}=\frac{k}{2}=1$$

그러므로 구하고자 하는 상수는 $k=2$이다.

(b) X와 Y의 주변확률밀도함수는 각각 다음과 같다.

$$f_X(x)=\int_0^\infty f(x,y)\,dy=\int_0^\infty 2e^{-(x+2y)}\,dy=\left[-e^{-x}e^{-2y}\right]_{y=0}^{y=\infty}=e^{-x},\ x>0$$

$$f_Y(y)=\int_0^\infty f(x,y)\,dx=\int_0^\infty 2e^{-(x+2y)}\,dx=\left[-2e^{-2y}e^{-x}\right]_{x=0}^{x=\infty}=2e^{-2y},\ y>0$$

(c) $x>0,\ y>0$에서 X와 Y의 결합분포함수는 다음과 같다.

$$F(x,y)=\int_0^x\int_0^y f(u,v)\,dv\,du=\int_0^x\int_0^y 2e^{-(u+2v)}\,dv\,du$$
$$=\int_0^x\left[-2e^{-u}e^{-2v}\right]_{v=0}^{v=y}du$$
$$=\int_0^x e^{-u}(1-e^{-2y})\,du=\left[-(1-e^{-2y})e^{-u}\right]_{u=0}^{u=x}$$
$$=(1-e^{-x})(1-e^{-2y})$$

(d) $x>0,\ y>0$에서 $f(x,y)=2e^{-(x+2y)}=f_X(x)f_Y(y)=e^{-x}\times 2e^{-2y}$이므로 X와 Y의 독립이다.

(e) 결합분포함수를 이용하면 구하고자 하는 확률은 다음과 같다.

$$P(X\le 1,\ Y\le 1)=F(1,1)=(1-e^{-1})(1-e^{-2})$$

8. 이산확률변수 X와 Y의 결합확률질량함수가 다음과 같다.

$$f(x,y)=k,\ x=1,2,\cdots,10,\ y=11-x,12-x,\cdots,10$$

(a) 상수 k를 구하라.
(b) X와 Y의 주변확률질량함수를 구하고, 항등분포인지 조사하라.
(c) X와 Y의 독립성을 조사하라.
(d) $X=5$일 때, Y의 조건부 확률질량함수를 구하라.

풀이

(a) $x=1$이면 $y=10$, $x=2$이면 $y=9,10$이고 반복적으로 $x=10$이면 $y=1,2,\cdots,10$이다. 따라서 쌍 (x,y)가 취할 수 있는 값들은 모두 55개이고, $k=\frac{1}{55}$이다.

(b) X와 Y의 결합확률은 아래 표와 같다. 따라서 X와 Y의 주변확률질량함수는 각각 다음과 같다.

$$f_X(x)=\frac{x}{55},\quad x=1,2,\cdots,10,\ f_Y(y)=\frac{y}{55},\quad y=1,2,\cdots,10$$

그러므로 $x=1,2,\cdots,10$에 대하여 $f_X(x)=f_Y(x)=\frac{x}{55}$이고, X와 Y는 항등분포이다,

Y \ X	1	2	3	4	5	6	7	8	9	10	$f_X(x)$
1	0	0	0	0	0	0	0	0	0	$\frac{1}{55}$	$\frac{1}{55}$
2	0	0	0	0	0	0	0	0	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{2}{55}$
3	0	0	0	0	0	0	0	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{3}{55}$
4	0	0	0	0	0	0	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{4}{55}$
5	0	0	0	0	0	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{5}{55}$
6	0	0	0	0	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{6}{55}$
7	0	0	0	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{7}{55}$
8	0	0	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{8}{55}$
9	0	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{9}{55}$
10	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{1}{55}$	$\frac{10}{55}$
$f_Y(y)$	$\frac{1}{55}$	$\frac{2}{55}$	$\frac{3}{55}$	$\frac{4}{55}$	$\frac{5}{55}$	$\frac{6}{55}$	$\frac{7}{55}$	$\frac{8}{55}$	$\frac{9}{55}$	$\frac{10}{55}$	1

(c) $f(0,0)=0 \neq f_X(0)\,f_Y(0)=\left(\frac{1}{55}\right)^2$ 이므로 X와 Y는 독립이 아니다.

(d) $P(X=5)=\frac{5}{55}$ 이므로 $y=6,7,8,9,10$에 대하여 Y의 조건부 확률은 다음과 같다.

$$P(Y=y\,|\,X=5)=\frac{P(X=5,\ Y=y)}{P(X=5)}=\frac{1/55}{5/55}=\frac{1}{5}$$

따라서 구하고자 하는 조건부 확률질량함수는 $f_{Y|X}(y|5)=\frac{1}{5}$, $y=6,7,8,9,10$이다.

9. 확률변수 X와 Y의 결합확률밀도함수가 $f(x,y)=e^{-x}$, $0<y<x<\infty$이다.
(a) X와 Y의 주변확률밀도함수를 구하라.
(b) X와 Y의 독립성을 조사하라.
(c) 확률 $P(0<X<2,\ 0<Y<2)$를 구하라.
(d) 확률 $P(2Y\le X)$를 구하라.
(e) 확률 $P(X+Y\le 2)$을 구하라.

풀이

(a) X와 Y의 상태공간은 그림 (a)와 같다. 그러므로 X와 Y의 주변확률밀도함수는 각각 다음과 같다.

$$f_X(x) = \int_0^x e^{-x}\, dy = \left[y e^{-x} \right]_{y=0}^{y=x} = x e^{-x}, \ \ x > 0$$

$$f_Y(y) = \int_y^\infty e^{-x}\, dx = \left[-e^{-x} \right]_{x=y}^{x=\infty} = e^{-y}, \ \ y > 0$$

(b) $x > 0$, $y > 0$에서 $f(x, y) = e^{-x} \neq f_X(x) f_Y(y) = x e^{-(x+y)}$이므로 X와 Y는 독립이 아니다.

(c) $\{(x, y) | 0 < x < 2, 0 < y < 2\}$는 그림 (b)와 같다. 그러므로 구하고자 하는 확률은 다음과 같다.

$$P(0 < X < 2, 0 < Y < 2) = \int_0^2 \int_0^x e^{-x}\, dy\, dx = \int_0^2 x e^{-x}\, dx$$

$$= \left[-(x+1) e^{-x} \right]_0^2 = 1 - 3e^{-2}$$

(d) $\{(x, y) | 2y < x, y > 0\}$은 그림 (c)와 같다. 그러므로 구하고자 하는 확률은 다음과 같다.

$$P(2Y \le X) = \int_0^\infty \int_0^{x/2} e^{-x}\, dy\, dx = \int_0^\infty \frac{1}{2} x e^{-x}\, dx$$

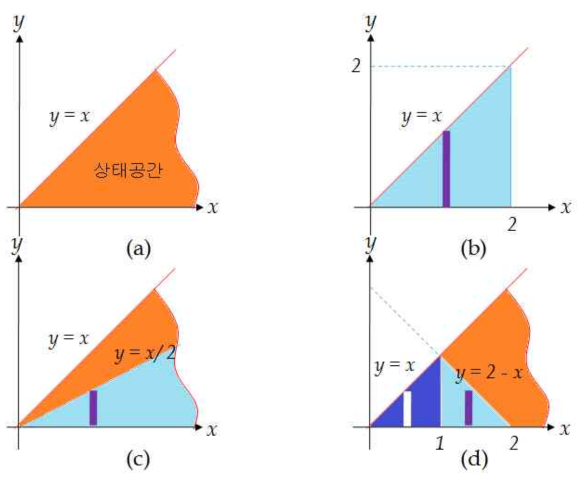
$$= \left[-\frac{1}{2}(x+1) e^{-x} \right]_0^\infty = \frac{1}{2}$$

(e) 상태공간에서 $x + y \le 2$인 영역은 그림 (d)와 같이 두 영역 $\{(x, y) | y < x, 0 < x < 1\}$와 $\{(x, y) | y < 2 - x, 1 < x < 2\}$로 분할된다. 그러므로 구하고자 하는 확률은 다음과 같다.

$$P(X + Y \le 2) = \int_0^1 \int_0^x e^{-x}\, dy\, dx + \int_1^2 \int_0^{2-x} e^{-x}\, dy\, dx$$

$$= \int_0^1 x e^{-x}\, dx + \int_1^2 \int_0^{2-x} (2 - x) e^{-x}\, dx$$

$$= \left[-(x+1) e^{-x} \right]_0^1 + \left[(x-1) e^{-x} \right]_1^2 = 1 - 2e^{-1} + e^{-2}$$



10. 영역 $A = \{(x, y) | x \geq 0,\ y \geq 0,\ x^2 + y^2 \leq 1\}$에서 확률변수 X와 Y의 결합확률밀도함수가 $f(x, y) = kxy$이다.

(a) 상수 k를 구하라.

(b) X와 Y의 주변확률밀도함수를 구하라.

(c) $X = x$일 때, Y의 조건부 확률밀도함수를 구하라.

(d) $X = \frac{1}{2}$일 때, Y의 조건부 확률밀도함수를 구하라.

풀이

(a) 결합확률밀도함수는 주어진 영역에서 $\int\int_A f(x, y)\,dy\,dx = 1$을 만족해야 한다. (그림 (a) 참조).

$$\int_0^1 \int_0^{\sqrt{1-x^2}} kxy\,dy\,dx = \int_0^1 \left[kxy^2\right]_{y=0}^{y=\sqrt{1-x^2}} dx = k\int_0^1 x(1-x^2)\,dx = \left[\frac{k}{8}(2x^2 - x^4)\right]_0^1 = \frac{k}{8} = 1$$

따라서 구하고자 하는 상수는 $k = 8$이다.

(b) X의 주변확률밀도함수는 그림 (b)와 같이 $0 < x < 1$에서 변수 y에 대해 적분한다.

$$f_X(x) = \int_0^{\sqrt{1-x^2}} 8xy\,dy = \left[4xy^2\right]_{y=0}^{y=\sqrt{1-x^2}} = 4x(1-x^2)$$

Y의 주변확률밀도함수는 그림 (c)과 같이 $0 < y < 1$에서 변수 x에 대해 적분한다.

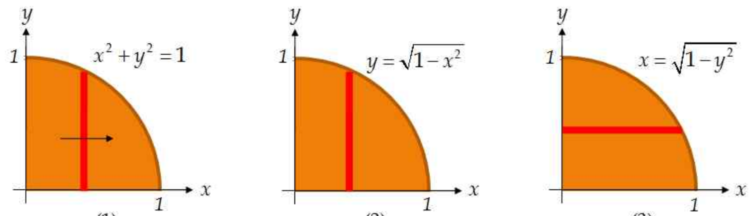
$$f_Y(y) = \int_0^{\sqrt{1-y^2}} 8xy\,dy = \left[4x^2y\right]_{x=0}^{x=\sqrt{1-y^2}} = 4y(1-y^2)$$

(c) $f_X(x) = 4x(1-x^2)$이므로 $X = x$일 때, Y의 조건부 확률밀도함수는 다음과 같다.

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{8xy}{4x(1-x^2)} = \frac{2y}{1-x^2}, \quad 0 < y < \sqrt{1-x^2}$$

(d) (c)에서 $X = \frac{1}{2}$이므로 구하고자 하는 조건부 확률밀도함수는 다음과 같다.

$$f_{Y|X}(y|1/2) = \frac{2y}{1-(1/2)^2} = \frac{8}{3}y, \quad 0 < y < \frac{\sqrt{3}}{2}$$



(1) (2) (3)

11. 0에서 9까지의 정수 중에서 비복원추출에 의하여 100단위의 숫자를 만들 때, 가장 작은 자릿수를 X, 가징 큰 자릿수를 Y라고 한다.

(a) X와 Y의 결합확률질량함수를 구하라.

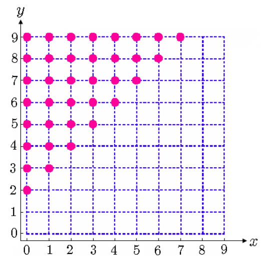
(b) $Z=Y-X$의 확률질량함수를 구하라.

(c) $P(Z \le 5)$를 구하라.

풀이

(a) 비복원추출에 의하여 세 자릿수를 추출하므로 세 자릿수는 서로 달라야 한다. 따라서 최소 자릿수 x와 최대 자릿수 y가 취할 수 있는 경우는 다음 그림과 같이 $0 \le x \le 7$, $x+2 \le y$, $y \le 9$이어야 한다. 이때 0에서 9 사이의 정수 중에서 3개의 숫자를 뽑는 방법의 수는 $\binom{10}{3}=120$이다.

한편 예를 들어, 최소 자릿수 3과 최대 자릿수 6인 경우, 즉 $x=3$, $y=6$이 되는 경우에 숫자의 조합은 $\{3,4,6\}$와 $\{3,5,6\}$뿐이며, 따라서 $6-3-1=2$가지가 있다. 또한 $x=3$, $y=7$이 되는 경우에 숫자의 조합은 $\{3,4,7\}$, $\{3,5,7\}$ 그리고 $\{3,6,7\}$뿐으로 $7-3-1=3$가지이다. 일반적으로 최소 자릿수 x와 최대 자릿수 y인 숫자 집합의 개수는 $y-x-1$개 있다. 따라서 최소 자릿수 x와 최대 자릿수 y일 확률 즉, X와 Y의 결합확률질량함수는 다음과 같다.



$$f(x,y)=P(X=x,\ Y=y)=\frac{y-x-1}{120},\quad x=0,\ 1,\ \cdots,\ 7,\ x+2 \le y,\ y \le 9$$

(b) 확률변수 $Z=Y-X$가 취할 수 있는 값은 위의 그림에서 보듯이 명백히 $2,\ 3,\ \cdots,\ 9$이다. 또한 $Z=z$인 경우의 수, 예를 들어 $Z=2$인 경우는 $(0,2)$, $(1,3)$, $(2,4)$, $(3,5)$, $(4,6)$, $(5,7)$, $(6,8)$, $(7,9)$ 등이며, $10-2=8$가지이다. 이와 같이 일반적으로 $Z=z$인 경우의 수는 $10-z$가지 있다. 한편 각각의 경우에 대하여 z는 최소 자릿수와 최대 자릿수의 쌍 즉, $(x,\ y)$로부터 나와야 하며, 이와 같은 최소 자릿수와 최대 자릿수의 쌍에 대한 확률은 $\frac{y-x-1}{120}=\frac{z-1}{120}$이다. 따라서 $Z=z$일 확률 즉, Z의 확률질량함수는 다음과 같다.

$$f_Z(z)=P(Z=z)=\frac{(10-z)(z-1)}{120}\quad,\qquad z=2,\ 3,\ \cdots,\ 9$$

(c) 구하고자 하는 확률은 $P(Z \le 5)=\sum_{z=2}^{5}\frac{(10-z)(z-1)}{120}=\frac{1}{2}$이다.

6장

1. 어떤 보안회사는 99.5%의 신뢰성을 갖는 보안시스템을 개발하고자 한다. 5명의 해커를 상대로 새로 개발한 보안시스템을 시험하였다.
(a) 이 보안시스템을 풀어낸 해커의 수에 대한 확률질량함수를 구하라.
(b) 이 보안시스템을 풀어낸 해커가 4명 이상일 확률을 구하라.
(c) 적어도 한 명이 보안시스템을 풀어낼 확률을 구하라.

풀이

(a) 보안시스템을 풀어낸 해커의 수를 X라 하면, 해커가 보안시스템을 풀어낼 가능성은 0.005이므로 확률질량함수는 다음과 같다.

$$f(x) = \binom{5}{x}(0.005)^x(0.995)^{5-x}, \quad x = 0, 1, 2, 3, 4, 5$$

(b) 구하고자 하는 확률은 다음과 같다.

$$P(X \geq 4) = f(4) + f(5) = \binom{5}{4}(0.005)^4(0.995) + \binom{5}{5}(0.005)^5$$

$$= 3.10938 \times 10^{-9} + 3.125 \times 10^{-12} = 3.1125 \times 10^{-9}$$

(c) 구하고자 하는 확률은 다음과 같다.

$$P(X \geq 1) = 1 - f(0) = 1 - \binom{5}{0}(0.995)^5 = 1 - 0.9752 = 0.0248$$

2. 한 인간의 신체적 특성은 한 쌍의 유전자에 의하여 결정된다. 이때 우성인자를 a 그리고 열성인자를 b이라 하면, (a, a)를 순수우성, (b, b)를 순수열성 그리고 (a, b) 또는 (b, a)를 혼성이라고 한다. 그리고 자녀는 부모로부터 각각 어느 한 유전자를 물려받게 되며, 이 유전자는 두 종류의 유전자 중에서 동등한 기회로 대물림한다.
(a) 혼성 유전자를 가지고 있는 두 부모의 자녀가 순수열성일 확률을 구하라
(b) 혼성 유전자를 가진 부모에게 3명의 자녀가 있다. 자녀 중에서 어느 1명만 순수열성일 확률을 구하라.
(c) 평균 순수열성인 자녀수를 구하라.

풀이

(a) 자녀가 순수열성일 가능성은 서로 독립적인 두 부모로부터 각각 열성인자를 물려받는 경우뿐이고, 각각의 확률은 $\frac{1}{2}$이므로 $p = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 0.25$이다.

(b) 3명의 자녀 중에서 순수열성인 자녀의 수를 X라 하면, $X \sim B(3, 0.25)$이다. 따라서 정확히 한 명만 순수열성일 확률은 다음과 같다.

$$P(X = 1) = P(X \leq 1) - P(X = 0) = 0.8438 - 0.4219 = 0.4219$$

(c) 구하고자 하는 평균은 $\mu = 3 \times 0.25 = 0.75$이다.

3. 건설업자는 신도시 건설을 위해 중고 불도저를 구입하려고 한다. 이 업자가 이전에 불도저를 구입한 경험에 의하면 각 불도저를 적어도 1년 이상 사용할 수 있을 확률이 35%이다. 이 사람이 새로운 신도시 건설을 위해 불도저 5대를 구입했다.

(a) 1년 후에도 사용할 수 있는 불도저의 수에 대한 확률질량함수를 구하라.

(b) 두 대 이상 1년 후에도 사용할 수 있을 확률을 구하라.

풀이

(a) 1년이 지난 시점에서 사용 가능한 불도저의 수를 X라 하면, 각 불도저가 적어도 1년 이상 사용할 수 있을 확률이 35%이므로 $X \sim B(5, 0.35)$이고 확률질량함수는 다음과 같다.

$$f(x) = \binom{5}{x}(0.35)^x (0.65)^{5-x}, \quad x = 0, 1, 2, 3, 4, 5$$

(b) 두 대 이상 1년 후에도 사용할 수 있을 확률은 다음과 같다.

$$P(X \ge 2) = 1 - P(X \le 1) = 1 - 0.4284 = 0.5716$$

4. 스마트 폰 생산라인에서 제조한 폰 10개를 대리점에 보냈으며, 이 중에는 약간의 흠집이 있는 폰이 2개 포함되어 있다. 10개의 폰 중에서 5개를 매장에 진열했을 때, 흠집이 있는 폰의 개수를 X라 한다.

(a) X의 확률질량함수를 구하라.

(b) 매장에 진열된 폰 중에서 흠집이 있는 폰이 한 개일 확률을 구하라.

(c) 매장에 진열된 폰 중에서 흠집이 있는 폰의 평균과 분산을 구하라.

풀이

(a) $N = 10$, $M = 2$, $n = 5$이므로 $X \sim H(10, 2, 5)$이므로 구하고자 하는 확률질량함수는 다음과 같다.

$$f(x) = \frac{\binom{2}{x}\binom{8}{5-x}}{\binom{10}{5}}, \quad x = 0, 1, 2$$

(b) 구하고자 하는 확률은 다음과 같다.

$$P(X = 1) = f(1) = \frac{\binom{2}{1}\binom{8}{4}}{\binom{10}{5}} = \frac{2 \times 70}{252} = \frac{5}{9}$$

(c) 구하고자 하는 평균과 분산은 각각 다음과 같다.

$$\text{평균} : \mu = 5 \times \frac{2}{10} = 1$$

$$\text{분산} : \sigma^2 = 5 \times \frac{2}{10} \times \left(1 - \frac{2}{10}\right)\left(\frac{10-5}{10-1}\right) = \frac{4}{9}$$

5. 어떤 공정라인에서 생산되는 제품은 10개 당 평균 1개의 불량품이 발생한다.
(a) 불량품이 발견될 때까지 생산된 제품의 개수에 대한 확률질량함수를 구하라.
(b) 다섯 번째에서 처음으로 불량품이 발생할 확률을 구하라.
(c) 다섯 번째에서 두 번째 불량품이 나올 확률을 구하라.

풀이

(a) 10개 당 평균 1개의 불량품이 생산되므로 불량품이 발견될 때까지 생산된 제품의 수 X는 모수 $p=0.1$인 기하분포에 따른다. 따라서 구하고자 하는 확률질량함수는 다음과 같다.

$$f(x)=(0.1)(0.9)^{x-1},\ x=1,2,3,\cdots$$

(b) 구하고자 하는 확률은 $P(X=5)=f(5)=(0.1)(0.9)^4=0.0656$이다.

(c) 두 번째 불량품이 나올 때까지 생산한 제품의 수를 Y라 하면, 확률질량함수는 다음과 같다.

$$f(y)=\binom{y-1}{1}(0.1)^2(0.9)^{y-2},\ y=2,3,\cdots$$

따라서 다섯 번째에서 두 번째 불량품이 나올 확률은 다음과 같다.

$$P(Y=5)=f(5)=\binom{4}{1}(0.1)^2(0.9)^3=0.0292$$

6. 지질학자들은 지르콘의 표면에 있는 우라늄의 분열 흔적의 횟수를 가지고 지르콘의 연대를 측정한다. 특정한 지르콘은 1㎠ 당 평균 4개의 흔적을 가지고 있다. 이 지르콘의 2㎠에 많아야 3개의 흔적을 가질 확률을 구하라. 이 분열 흔적의 수는 푸아송분포에 따른다고 한다.

풀이

1㎠ 당 평균 4개의 흔적을 가지고 있으므로 2㎠에 평균 8개의 흔적을 가지고 있으며, 따라서 분열 흔적의 횟수는 $X\sim P(8)$이다. 그러므로 2㎠에 많아야 3개의 흔적을 가질 확률은 $P(X\le 3)=0.042$이다.

7. 새 모델의 자동차를 개발한 어느 자동차 제조회사는 고속주행에서 파국의 원인이 되는 제동장치의 결함에 관심을 갖는다. 이 결함을 가지고 생산된 자동차의 수는 연간 평균 8대인 푸아송분포에 따른다.
(a) 어느 해 1년 동안 생산된 자동차 중에서 이 결함을 가진 자동차가 2대 이하로 생산될 확률을 구하라.
(b) 이 기간 동안 생산된 자동차 중에서 이 결함을 가진 자동차가 적어도 14대 이상 생산될 확률을 구하라.

풀이

(a) 푸아송 누적확률표를 이용하면 구하고자 하는 확률은 $P(X\le 2)=0.014$이다.

(b) 구하고자 하는 확률은 다음과 같다.

$$P(X\ge 14)=1-P(X\le 13)=1-0.966=0.034$$

8. 양수 m과 μ 그리고 $p+q=1$인 두 양수 p와 q에 대하여 다음 함수를 생각하자.

$$f(x)=\begin{cases} p e^{-m x} \dfrac{m^x}{x!}+q e^{-\mu x} \dfrac{\mu^x}{x!}, & x=0,1,2, \cdots \\ 0, & \text{다른 곳에서} \end{cases}$$

(a) 함수 $f(x)$가 확률질량함수임을 보여라.
(b) $p=2q$일 때 X의 확률질량함수를 구하라.
(c) $p=2q$일 때 X의 평균과 분산을 구하라.

풀이

(a) $\sum_{x=0}^{\infty} \frac{m^x}{x!} e^{-m x}=1$, $\sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu x}=1$이므로 다음을 얻는다.

$$\sum_{x=0}^{\infty} f(x)=\sum_{x=0}^{\infty}\left(p e^{-m x} \frac{m^x}{x!}+q e^{-\mu x} \frac{\mu^x}{x!}\right)=p \sum_{x=0}^{\infty} e^{-m x} \frac{m^x}{x!}+q \sum_{x=0}^{\infty} e^{-\mu x} \frac{\mu^x}{x!}$$
$$=p+q=1$$

그리고 모든 $x=0,1,2, \cdots$에 대하여 $f(x)>0$이므로 $f(x)$는 확률질량함수이다.

(b) $p=2q$, $p+q=1$이므로 $p=\frac{2}{3}$, $q=\frac{1}{3}$이므로 X의 확률질량함수는 다음과 같다.

$$f(x)=\begin{cases} \dfrac{2}{3} e^{-m x} \dfrac{m^x}{x!}+\dfrac{1}{3} e^{-\mu x} \dfrac{\mu^x}{x!}, & x=0,1,2, \cdots \\ 0, & \text{다른 곳에서} \end{cases}$$

(c) X의 평균은 다음과 같다.

$$E(X)=\sum_{x=0}^{\infty} x f(x)=\sum_{x=0}^{\infty}\left(\frac{2}{3} x e^{-m x} \frac{m^x}{x!}+\frac{1}{3} x e^{-\mu x} \frac{\mu^x}{x!}\right)$$
$$=\frac{2}{3} \sum_{x=0}^{\infty} x e^{-m x} \frac{m^x}{x!}+\frac{1}{3} \sum_{x=0}^{\infty} x e^{-\mu x} \frac{\mu^x}{x!}$$
$$=\frac{2}{3} m+\frac{1}{3} \mu=\frac{2 m+\mu}{3}$$

그리고 $X(X-1)$의 기댓값은 다음과 같다.

$$E[X(X-1)]=\sum_{x=1}^{\infty} x(x-1) f(x)=\sum_{x=1}^{\infty}\left(\frac{2}{3} x(x-1) e^{-m x} \frac{m^x}{x!}+\frac{1}{3} x(x-1) e^{-\mu x} \frac{\mu^x}{x!}\right)$$
$$=\frac{2}{3} \sum_{x=1}^{\infty} x(x-1) e^{-m x} \frac{m^x}{x!}+\frac{1}{3} \sum_{x=1}^{\infty} x(x-1) e^{-\mu x} \frac{\mu^x}{x!}$$
$$=\frac{2}{3} m^2+\frac{1}{3} \mu^2=\frac{2 m^2+\mu^2}{3}$$

따라서 X^2의 기댓값은 다음과 같다.

$$E(X^2) = E[X(X-1)] + E(X) = \frac{2m^2+\mu^2}{3} + \frac{2m+\mu}{3}$$

그러므로 X의 분산은 다음과 같다.

$$\sigma^2 = E(X^2) - E(X)^2 = \left(\frac{2m^2+\mu^2}{3} + \frac{2m+\mu}{3}\right) - \left(\frac{2m+\mu}{3}\right)^2$$

$$= \frac{1}{9}\left[2(m-\mu)^2 + 3(2m+\mu)\right]$$

9. 어느 주유소는 보통 가솔린, 고급 가솔린과 디젤 연료를 하루에 각각 50%, 20%, 30%씩 판매한다. 어느 날 20명의 고객이 주유소를 찾아왔을 때, 고객들은 독립적으로 연료를 선택한다.

(a) 고객들이 선정한 연료에 대한 결합확률질량함수를 구하라.

(b) 보통 가솔린, 고급 가솔린과 디젤 연료를 선택한 고객이 각각 15명, 3명, 2명일 확률을 구하라.

(c) 고급 가솔린을 선택한 사람이 7명 이상일 확률을 구하라.

(d) 고급 가솔린을 선택한 고객의 평균을 구하라.

풀이

(a) 보통 가솔린, 고급 가솔린과 디젤 연료를 선택한 고객 수를 각각 X, Y, Z라 하면, 구하고자 하는 결합확률질량함수는 다음과 같다.

$$f(x, y, z) = \binom{20}{x, y, z}(0.5)^x (0.2)^y (0.3)^z, \quad x, y, z = 0, 1, \cdots, 20, \quad x+y+z = 20$$

(b) 구하고자 하는 확률은 다음과 같다.

$$P(X=15, Y=3, Z=2) = f(15, 3, 2) = \binom{20}{15, 3, 2}(0.5)^{15}(0.2)^3(0.3)^2$$

$$= \frac{20!}{15!\,3!\,2!}(0.5)^{15}(0.2)^3(0.3)^2 = 0.0034$$

(c) $Y \sim B(20, 0.3)$이므로 구하고자 하는 확률은 다음과 같다.

$$P(Y \ge 7) = 1 - P(Y \le 6) = 1 - 0.9133 = 0.0867$$

(d) 구하고자 하는 평균은 $\mu_Y = 20 \times 0.2 = 4$이다.

10. 어느 건전지 제조회사에서 만들어진 1.5볼트 건전지는 실제로 1.45볼트에서 1.55볼트 사이에서 균등분포를 이룬다고 한다.
(a) 생산된 건전지 중에서 임의로 하나를 선정했을 때, 기대되는 전압과 표준편차를 구하라.
(b) 건전지 전압이 1.49볼트보다 크고 1.51볼트보다 작을 확률을 구하라.
(c) 20개의 건전지가 들어 있는 상자 안에 전압이 1.49볼트보다 크고 1.51볼트보다 작은 건전지 수의 평균과 분산을 구하라.
(d) (c)에서 1.49볼트보다 크고 1.51볼트보다 작은 전압을 가진 건전지가 4개 이상 들어 있을 확률을 구하라.

풀이

(a) 건전지의 전압을 X라 하면, $X \sim U(1.45, 1.55)$이므로 X의 평균과 분산은 각각 다음과 같다.

$$\text{평균} : \mu = \frac{1.45 + 1.55}{2} = 1.5$$

$$\text{분산} : \sigma^2 = \frac{(1.55 - 1.45)^2}{12} = 0.00083$$

(b) X의 확률밀도함수는 $f(x) = \dfrac{1}{1.55 - 1.45} = 10$, $1.45 \le x \le 1.55$이다. 따라서 구하고자 하는 확률은 (높이)×(밑변)이므로 $P(X \le 1.5) = 10 \times (1.51 - 1.49) = 0.2$이다.

(c) 한 건전지의 전압이 1.49볼트보다 크고 1.51볼트보다 작을 확률이 0.2이므로, 20개 안에 이러한 건전지의 수를 Y라 하면, $Y \sim B(20, 0.2)$이다. 따라서 건전지 수의 평균과 분산은 각각 다음과 같다.

$$\text{평균} : \mu = 20 \times 0.2 = 4$$

$$\text{분산} : \sigma^2 = 20 \times 0.2 \times 0.8 = 3.2$$

(d) 구하고자 하는 확률은 $P(Y \ge 4) = 1 - P(Y \le 3) = 1 - 0.4114 = 0.5886$이다.

11. 어떤 질병에 감염되어 증세가 나타날 때까지 걸리는 시간은 평균 15일이고, 감염기간은 지수분포를 이룬다고 한다.
(a) 이 질병에 감염된 환자가 20일 안에 증세를 보일 확률을 구하라.
(b) 적어도 25일 동안 이 질병에 대한 증세가 나타나지 않을 확률을 구하라.
(c) 적어도 20일 동안 이 질병에 대한 증세가 나타나지 않았다고 할 때, 이 환자가 25일 이후에 증세를 보일 확률을 구하여라.

풀이

(a) 질병의 증세가 나타날 때까지 걸리는 시간을 확률변수 X라고 하면, 평균 15인 지수분포를 이루므로 구하고자 하는 확률은 다음과 같다.

$$P(X < 20) = F(20) = 1 - e^{-20/15} = 0.7364$$

(b) 25일 안에 증세가 나타나지 않을 확률은 $P(X>25)=S(30)=e^{-25/15}=0.1889$이다.

(c) 구하고자 하는 확률은 비기억성 성질에 의하여 다음과 같다.

$$P(X>25 \mid X>20)=P(X>5)=S(5)=e^{-5/15}=0.7165$$

12. 어느 대학교의 교내에서 발생하는 교통사고는 연평균 73건이고[조선일보 2015.09.25] 푸아송분포에 따라 발생한다. 교통사고 발생 이후로 다음 사고가 발생할 때까지 걸리는 시간을 X라 한다. 단, 단위는 일이다.
(a) X의 확률밀도함수를 구하라.
(b) 사고 발생 이후로 다음 사고가 일주일 이내에 발생할 확률을 구하라.
(c) 30일 이상 사고가 발생하지 않았다고 할 때, 앞으로 30일 동안 사고가 발생하지 않을 확률을 구하라.
(d) 30일 동안 사고가 발생하지 않을 확률을 구하라.

풀이

(a) 연간 73건의 사고가 발생하므로 하루에 $\lambda=\dfrac{73}{365}$의 비율로 사고가 발생한다. 따라서 X의 확률밀도함수는 다음과 같다.

$$f(x)=\frac{73}{365}\,e^{-73x/365},\quad x>0$$

(b) 일주일 이내에 사고가 발생할 확률은 $P(X<7)=F(7)=1-e^{-73\cdot 7/365}=0.7734$이다.

(c) 비기억성 성질에 의하여 구하고자 하는 확률은 다음과 같다.

$$P(X>60 \mid X>30)=P(X>30)=S(30)=e^{-73\cdot 30/365}=0.0025$$

(d) 사고가 연평균 73건 발생하므로 30일 동안 평균 $\mu=\dfrac{73\times 30}{365}=\dfrac{2190}{365}=6$건 발생한다. 따라서 30일 동안 발생한 건수를 Y라 하면, $Y\sim P(6)$이고, 구하고자 하는 확률은 $P(X=0)=0.002$이다.

13. 자동차 차체 생산라인의 판넬생산공정에 도착하는 판넬이 1시간당 평균 $\mu=1.6$의 비율인 포아송분포에 따라 도착한다.
(a) 판넬생산공정에 도착하는 두 판넬 사이의 평균 시간을 구하라.
(b) 판넬생산공정에 도착하는 두 판넬 사이의 시간이 1.5시간 이상일 확률을 구하라.
(c) 5시간 동안에 도착한 판넬의 수에 대한 분포를 구하라.
(d) 5시간 동안에 적어도 3개의 판넬이 도착할 확률을 구하라.

풀이

(a) 판넬생산공정에 도착하는 두 판넬 사이의 대기시간을 T라 하면, $T\sim \mathrm{Exp}(1.6)$이므로 T의 평균시간은 $E(T)=\dfrac{1}{1.6}=0.625$이다.

(b) $T \sim \text{Exp}(1.6)$이므로 생존함수는 $S(x) = e^{-1.6x}$, $x > 0$이고 따라서 구하고자 하는 확률은 $P(X \geq 1.5) = S(1.5) = e^{-1.6 \cdot 1.5} = 0.0907$이다.

(c) 1시간당 평균이 1.6이므로 5시간 당 평균 8인 푸아송분포에 따른다. 따라서 5시간 동안에 도착한 판넬의 수 X는 모수 8인 푸아송분포에 따른다.

(d) 구하고자 하는 확률은 $P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.014 = 0.986$이다.

14. $X \sim U(0,1)$일 때, X의 분포함수를 이용하여 $Y = -\lambda \ln(1-X)$은 모수 $\frac{1}{\lambda}$인 지수분포에 따르는 것을 보여라.

풀이

$X \sim U(0,1)$이므로 X의 분포함수는 $F_X(x) = x$, $0 < x < 1$이므로 $0 < y = -\lambda \ln(1-x) < \infty$이다. 이제 확률변수 Y의 분포함수를 다음과 같이 구한다.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P[-\lambda \ln(1-X) \leq y] = P\left[\ln(1-X) \geq -\frac{y}{\lambda}\right] \\ &= P\left(1-X \geq e^{-y/\lambda}\right) = P\left(X \leq 1 - e^{-y/\lambda}\right) = F_X\left(1 - e^{-y/\lambda}\right) \\ &= 1 - e^{-y/\lambda} \end{aligned}$$

즉, Y의 분포함수는 $F_Y(y) = 1 - e^{-y/\lambda}$, $y > 0$이고 확률밀도함수는 $f_Y(y) = \frac{1}{\lambda} e^{-y/\lambda}$, $y > 0$이다. 따라서 Y는 모수 $\frac{1}{\lambda}$인 지수분포에 따른다.

15. 지수분포의 변형인 다음과 확률밀도함수를 가지는 확률분포를 모수 λ와 θ인 **라플라스분포**Laplace distribution라 한다.

$$f(x) = \frac{1}{2} \lambda e^{-\lambda |x - \theta|}, \quad -\infty < x < \infty$$

(a) X의 분포함수를 구하라.
(b) $\lambda = 3$, $\theta = 1$일 때, $P(X \leq 0)$을 구하라.
(c) (b)의 조건에서 $P(X \leq 2)$를 구하라.
(d) (b)의 조건에서 $P(0 \leq X \leq 2)$를 구하라.

풀이

(a) $x < \theta$에 대하여 $|x - \theta| = -(x - \theta)$이므로 다음을 얻는다.

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \frac{1}{2} \lambda \int_{-\infty}^{x} e^{\lambda(t - \theta)}\, dt$$

$$= = \frac{1}{2} \lambda \, \frac{1}{\lambda} e^{\lambda(t - \theta)} \Big|_{-\infty}^{x} = \frac{1}{2} e^{\lambda(x - \theta)} = \frac{1}{2} e^{-\lambda(\theta - x)}$$

$x > \theta$에 대하여 $|x-\theta| = x-\theta$이므로 다음을 얻는다.

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \frac{1}{2}\lambda\int_{-\infty}^{\theta} e^{\lambda(t-\theta)}\,dt + \frac{1}{2}\lambda\int_{\theta}^{x} e^{-\lambda(t-\theta)}\,dt$$

$$= \frac{1}{2} + \frac{1}{2}\lambda\ -\frac{1}{\lambda}e^{-\lambda(t-\theta)}\Big|_{\theta}^{x} = \frac{1}{2} + \frac{1}{2}\left(1-e^{-\lambda(x-\theta)}\right) = 1-\frac{1}{2}e^{-\lambda(x-\theta)}$$

따라서 구하고자 하는 분포함수는 다음과 같다.

$$F(x) = \begin{cases} \frac{1}{2}e^{-\lambda(\theta-x)} & ,\ x < \theta \\ 1-\frac{1}{2}e^{-\lambda(x-\theta)} & ,\ x \geq \theta \end{cases}$$

(b) $\lambda = 3$, $\theta = 1$이면 분포함수는 다음과 같다.

$$F(x) = \begin{cases} \frac{1}{2}e^{-3(1-x)} & ,\ x < 1 \\ 1-\frac{1}{2}e^{-3(x-1)} & ,\ x \geq 1 \end{cases}$$

그러므로 구하고자 하는 확률은 $P(X \leq 0) = F(0) = \frac{1}{2}e^{-3} = 0.0249$이다.

(c) 구하고자 하는 확률은 $P(X \leq 2) = F(2) = 1-\frac{1}{2}e^{-3} = 0.9751$이다.

(d) 구하고자 하는 확률은 다음과 같다.

$$P(0 \leq X \leq 2) = F(2) - F(0) = 0.9751 - 0.0249 = 0.9502$$

16. 우리나라는 규모 2.0이상인 지진이 연평균 44.5인 푸아송분포에 따라 발생한다.
(a) 1월 1일 관측한 이후 3번째 지진이 발생할 때까지 걸리는 시간의 확률밀도함수를 구하라.
(b) 3번째 지진이 1월 달 안에 발생할 확률을 구하라.

풀이

(a) 지진이 관측된 이후로 다음 지진이 관측될 때까지 걸리는 시간 T는 모수 $\lambda = 44.5$인 지수분포를 이루므로 3번째 지진이 발생할 때까지 걸리는 시간 X는 $\alpha = 3$과 $\beta = \frac{1}{44.5}$인 감마분포를 이룬다. 따라서 구하고자 하는 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\Gamma(3)(1/44.5)^3}\,x^{3-1}\exp\left(-\frac{x}{1/44.5}\right) = \frac{704969}{16}\,x^2\,e^{-(44.5)x},\ x > 0$$

(b) 1월 달은 31일이고 $\frac{31}{365}$년이므로 3번째 지진이 1월 달 안에 발생할 확률은 다음과 같다.

$$P\left(0<X<\frac{31}{365}\right)=\int_0^{31/365}\frac{704969}{16}\,x^2\,e^{-(44.5)x}\,dx=-\frac{1}{8}\left[(7921x^2+356x+8)e^{89x/2}\right]_0^{31/365}$$

$$=-\frac{12706021}{1065800}e^{-2759/730}+1=0.7278$$

17. 어떤 기계의 부속품의 수명은 $\alpha=2$, $\beta=1.5$인 와이블 분포에 따른다고 한다. 단, 단위는 년이다.
(a) 이 기계를 처음 사용한 후 6개월 안에 고장 날 확률을 구하라.
(b) 1년 이상 고장 나지 않고 사용할 확률을 구하라.
(c) 1년 이상 고장 나지 않고 사용했다 했을 때, 앞으로 6개월 이상 더 사용할 확률을 구하라.
(d) 고장률을 구하라.

풀이

(a) 기계 부속품의 수명을 X라 하면, 모수 $\alpha=2$, $\beta=1.5$인 와이블 분포에 따르므로 분포함수는 $F(x)=1-e^{-(1.5x)^2}$이고, 6개월은 0.5년이므로 6개월 안에 고장 날 확률은 다음과 같다.

$$P(X<0.5)=F(0.5)=1-e^{-((1.5)\cdot(0.5))^2}=1-0.5698=0.4302$$

(b) 1년 이상 고장 나지 않고 사용할 확률은 다음과 같다.

$$P(X\geq 1)=1-F(1)=e^{-((1.5)\cdot 1)^2}=0.1054$$

(c) 구하고자 하는 확률은 다음과 같다.

$$P(X\geq 1.5\,|\,X\geq 1)=\frac{P(X\geq 1.5)}{P(X\geq 1)}=\frac{1-F(1.5)}{1-F(1)}=\frac{e^{-((1.5)\cdot(1.5))^2}}{e^{-((1.5)\cdot 1)^2}}=0.06$$

(d) $\alpha=2$, $\beta=1.5$이므로 고장률은 다음과 같다.

$$h(x)=\frac{1.5}{2^{1.5}}\,x^{1.5-1}=\frac{3\sqrt{2}}{8}\sqrt{x},\ x>0$$

18. 어떤 발전기의 수명은 고장률이 시간 당 $h(x)=10^{-3}x$인 와이블 분포를 이룬다.
(a) 이 발전기의 수명에 대한 확률밀도함수를 구하라.
(b) 발전기의 평균 수명을 구하라.
(c) 이 발전기를 30시간 이상 사용할 확률을 구하라.

풀이

(a) $X\sim \mathrm{Wei}(\alpha,\beta)$의 고장률은 $h(x)=\frac{\beta}{\alpha^\beta}x^{\beta-1}=\frac{1}{10^3}x$, $x>0$이므로 $x^{\beta-1}=x$이고 $\frac{\beta}{\alpha^\beta}=\frac{1}{10^3}$이다.

따라서 $\beta=2$이고 $\dfrac{2}{\alpha^2}=\dfrac{1}{10^3}$으로부터 $\alpha=20\sqrt{5}$이므로 구하고자 하는 확률밀도함수는 다음과 같다.

$$f(x)=\frac{1}{1000}\,x\,e^{-x^2/2000},\ \ x>0$$

(b) 발전기의 평균 수명은 다음과 같다.

$$\mu=20\sqrt{5}\,\Gamma\left(1+\frac{1}{2}\right)=20\sqrt{5}\times\frac{1}{2}\times\Gamma\left(\frac{1}{2}\right)=10\sqrt{5\pi}=79.2665(\text{시간})$$

(c) $X\sim \mathrm{Wei}(\alpha,\beta)$의 생존함수는 $S(x)=e^{-(x/\alpha)^\beta},\ x>0$이므로 구하고자 하는 확률은 다음과 같다.

$$P(X\geq 30)=S(30)=e^{-30^2/2000}=e^{-9/20}=0.6376$$

7장

1. X는 평균 4이고 표준편차 3인 정규분포에 따른다.
(a) 확률 $P(X<7)$을 구하라.
(b) $P(4 \le X \le x_0) = 0.4750$인 x_0을 구하라.
(c) $P(1 < X < x_0) = 0.756$인 x_0을 구하라.

풀이

X를 표준화하면, $Z=(X-4)/3$이다.

(a) $P(X<7) = P(Z<1) = 0.8413$이다.

(b) $P(X \le 4) = 0.5$, $P(4 \le X \le x_0) = 0.4750$이므로 $P(X \le x_0) = 0.9750$이다. 한편 표준정규분포표에서 $\Phi(1.96) = 0.9750$이므로 $x_0 = 4 + 3 \times 1.96 = 9.88$이다.

(c) $1 < X < x_0 \Leftrightarrow \dfrac{1-4}{3} < Z < \dfrac{x_0 - 4}{3}$이므로

$$\begin{aligned} P(1 < X < x_0) &= P\left(-1 < Z < \frac{x_0-4}{3}\right) = \Phi\left(\frac{x_0-4}{3}\right) - \Phi(-1) \\ &= \Phi\left(Z < \frac{x_0-4}{3}\right) - (1-\Phi(1)) \\ &= \Phi\left(Z < \frac{x_0-4}{3}\right) - (1-0.8413) = 0.756 \end{aligned}$$

이다. 즉, $z_0 = (x_0 - 4)/3$에 대하여 $\Phi(z_0) = 0.9147$이고, 표준정규분포표에서 $z_0 = 1.37$을 얻는다. 그러므로 $x_0 = 4 + 3 \times 1.37 = 8.11$이다.

2. $X_1, X_2, \cdots, X_{25} \sim$ i.i.d. $N(4, 9)$일 때, 표본평균을 $\overline{X}$라 한다.
(a) 확률 $P(\overline{X} < 5.5)$를 구하라.
(b) $P(\overline{X} \le x_0) = 0.9750$인 x_0을 구하라.
(c) $P(2 < \overline{X} < x_0) = 0.7320$인 x_0을 구하라.

풀이

$X_1, X_2, \cdots, X_{25} \sim$ i.i.d. $N(4, 9)$이므로 표본평균은 $\overline{X} \sim N\left(4, \dfrac{9}{25}\right)$이다.

(a) $P(\overline{X} < 5.5) = P\left(\dfrac{\overline{X}-4}{3/5} < \dfrac{5.5-4}{3/5}\right) = \Phi(2.5) = 0.9938$이다.

(b) $P(\overline{X} < x_0) = P\left(\dfrac{\overline{X}-4}{3/5} < \dfrac{x_0-4}{3/5}\right) = 0.975$이다. 그리고 표준정규분포표에서 $\Phi(1.96) = 0.9750$이므로 $x_0 = 4 + 1.96 \times 0.6 = 5.176$이다.

© $P(2<\overline{X}<x_0)=P\left(\frac{2-4}{3/5}<\frac{\overline{X}-4}{3/5}<\frac{x_0-4}{3/5}\right)$

$$=P\left(-3.33<Z<\frac{x_0-4}{0.6}\right)=\Phi\left(\frac{x_0-4}{0.6}\right)-\Phi(-3.33)$$

$$=\Phi\left(\frac{x_0-4}{0.6}\right)-(1-\Phi(3.33))$$

$$=\Phi\left(\frac{x_0-4}{0.6}\right)-(1-0.9996)=0.7320$$

즉 $\Phi\left(\frac{x_0-4}{0.6}\right)=0.7324$, 표준정규분포표에서 $\Phi(0.62)=0.7324$이므로 $\frac{x_0-4}{0.6}=0.62$ 즉, $x_0=4.372$이다.

3. $T\sim t(r)$에 대하여 $\sigma^2=1.25$이라 한다. 이때 자유도 r과 $P(|T|\le 2.228)$을 구하라.

풀이

$T\sim t(r)$이므로 $\sigma^2=\frac{r}{r-2}=1.25$; $r=1.25\times(r-2)=1.25r-2.5$; $r=10$이다. 또한 $t_\alpha(10)=2.228$를 만족하는 $\alpha=0.025$이다. 즉, 다음을 얻는다.

$$P(T>2.228)=0.025,\ P(T<-2.228)=0.025$$

그러므로 구하고자 하는 확률은 다음과 같다.

$$P(|T|\le 2.228)=1-2\times 0.025=0.95$$

4. 독립인 두 확률변수가 $X\sim\chi^2(6)$, $Y\sim\chi^2(3)$일 때, $Z=X+Y$의 99% 백분위수를 구하라.

풀이

$X\sim\chi^2(6)$, $Y\sim\chi^2(3)$이고 독립이므로 $Z\sim\chi^2(9)$이다. 그러므로 카이제곱분포표로부터 99% 백분위수는 $\chi^2_{0.01}(9)=21.67$이다.

5. 건축물에 사용되는 철근은 평균 5m와 표준편차 0.1m인 정규분포에 따라 절단된다.

(a) 절단된 길이가 오차한계인 4.85m와 5.15m를 벗어날 확률을 구하라.

(b) 절단된 길이가 오차한계 km 안에 놓일 확률이 95%가 되는 k를 구하라.

풀이

(a) 절단된 철근의 길이는 $X\sim N(5,0.1^2)$이고 절단된 길이와 평균의 차에 대한 절댓값이 0.15를 초과할 확률은 다음과 같다.

$$P(|X-\mu|\ge 0.15)=P\left(\frac{|X-\mu|}{0.1}\ge\frac{0.15}{0.1}\right)=P(|Z|\ge 1.5)=2P(Z\ge 1.5)$$

$$=2(1-\Phi(1.5))=2(1-0.9332)=0.1336$$

(b) 오차한계 km 안에 놓일 확률이 95%이므로 다음이 성립한다.

$$P(|X-\mu| \le k) = P\left(\frac{|X-\mu|}{0.1} \le \frac{k}{0.1}\right) = P\left(|Z| \le \frac{k}{0.1}\right) = P(|Z| \le z_{0.05}) = 0.95$$

이때 $z_{0.05} = 1.96$이므로 $\frac{k}{0.1} = 1.96$, 즉 $k = 0.1 \times 1.96 = 0.196(\mathrm{m})$이다.

6. 평균 μ와 표준편차 σ인 정규분포에 따르는 확률변수 X가 다음을 만족한다. 이때 평균 μ와 표준편차 σ를 구하라.

$$P(X < 2.5) = 0.9788,\ P(X > 4.2) = 0.002$$

풀이

2.5와 4.2를 표준화하면 $z_r = \frac{2.5-\mu}{\sigma}$, $z_l = \frac{4.2-\mu}{\sigma}$이다. 따라서 다음을 얻는다.

$$P(X < 2.5) = \Phi(z_r) = 0.9788, \quad P(X > 4.2) = P(Z > z_l) = 1 - \Phi(z_l) = 0.002$$

즉, $\Phi(z_r) = 0.9788$, $\Phi(z_l) = 0.9980$이고 표준정규분포표에서 $z_r = 2.03$, $z_l = 2.88$을 얻는다. 따라서 다음 연립방정식을 얻는다.

$$\frac{2.5-\mu}{\sigma} = 2.03, \ \frac{4.2-\mu}{\sigma} = 2.88; \ \begin{cases} \mu + (2.03)\sigma = 2.5, \\ \mu + (2.88)\sigma = 4.2 \end{cases}; \quad \mu = -1.56,\ \sigma = 2$$

7. 장마철에 어느 도시에서 배출되는 물의 양은 시간당 평균 26.5만m^3, 표준편차 5만m^3인 정규분포에 따른다. 배수시설은 시간당 최대 32만m^3을 배출하도록 고안되어있다.

(a) 시간당 배출량이 23.5만m^3과 30.8만m^3 사이일 확률을 구하라.

(b) 장마철에 이 도시가 침수로 피해를 입을 확률을 구하라.

풀이

(a) 배출되는 물의 양을 X라 하면 $X \sim N(26.5, 25)$이고, 23.5와 30.8을 표준화하면 다음과 같다.

$$z_l = \frac{23.5-26.5}{5} = -0.6, \quad z_r = \frac{30.8-26.5}{5} = 0.86$$

따라서 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned} P(23.5 < X < 30.8) &= P(-0.6 < Z < 0.86) = \Phi(0.86) - \Phi(-0.6) \\ &= \Phi(0.86) - (1 - \Phi(0.6)) = 0.8051 - (1 - 0.7257) = 0.5308 \end{aligned}$$

(b) 최대 배출량은 32이므로 침수로 피해를 입을 확률은 다음과 같다.

$$P(X>32)=P\left(\frac{X-26.5}{5}>\frac{32-26.5}{5}\right)=1-\Phi(1.1)=1-0.8643=0.1357$$

8. 어떤 금속의 인장강도는 $1\,\text{cm}^2$당 평균 10톤, 분산 0.16톤인 정규분포에 따른다.
(a) 인장강도가 9톤과 11.2톤 사이일 확률을 구하라.
(b) 인장강도가 10.86톤을 초과할 확률을 구하라.
(c) 이 금속 10개의 평균 강도가 10.3톤을 초과할 확률을 구하라.

풀이

(a) 인장강도를 X라 하면 $X \sim N(10,\,0.16)$이고, 9와 11.2을 표준화하면 다음과 같다.

$$z_l=\frac{9-10}{0.4}=-2.5,\quad z_r=\frac{11.2-10}{0.4}=3.0$$

따라서 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned}P(9<X<11.2)&=P(-2.5<Z<3.0)=\Phi(3.0)-\Phi(-2.5)\\&=\Phi(3)-(1-\Phi(2.5))=0.9987-(1-0.9938)=0.9925\end{aligned}$$

(b) 10.86을 표준화하면, $z_l=\dfrac{10.86-10}{0.4}=2.15$이므로 구하고자 하는 확률은 다음과 같다.

$$P(X>10.86)=P(Z>2.15)=1-\Phi(2.15)=1-0.9842=0.0158$$

(c) 이 금속 10개의 평균 강도는 $\overline{X} \sim N(10,\,0.016)$이고 10.3을 표준화하면, $z_l=\dfrac{10.3-10}{\sqrt{0.016}}=2.37$이므로 구하고자 하는 확률은 다음과 같다.

$$P(\overline{X}>10.3)=P(Z>2.27)=1-\Phi(2.37)=1-0.9911=0.0089$$

9. X는 모수 $n=50$, $p=0.3$인 이항분포에 따른다.
(a) 정규근사에 의하여 $P(13\le X\le 17)$을 구하라.
(b) 연속성을 수정하여 $P(13\le X\le 17)$을 구하라.
(c) 연속성을 수정하여 $P(13<X<17)$을 구하라.

풀이

$X \sim B(50,\,0.3)$이므로 $X \approx N(15,\,10.5)$이다.

(a) $P(13\le X\le 17)=P\left(\dfrac{13-15}{\sqrt{10.5}}\le Z\le \dfrac{17-15}{\sqrt{10.5}}\right)=P(-0.62\le Z\le 0.62)$

$$=2\Phi(0.62)-1=2\times 0.7324-1=0.4648$$

(b) $P(13 \le X \le 17) = P\left(\frac{12.5-15}{\sqrt{10.5}} \le Z \le \frac{17.5-15}{\sqrt{10.5}}\right) = P(-0.77 \le Z \le 0.77)$

$$= 2\Phi(0.77) - 1 = 2 \times 0.7794 - 1 = 0.5584$$

(c) $P(13 < X < 17) = P(14 \le X \le 16) = P\left(\frac{13.5-15}{\sqrt{10.5}} \le Z \le \frac{16.5-15}{\sqrt{10.5}}\right)$

$$= P(-0.46 \le Z \le 0.46) = 2\Phi(0.46) - 1$$

$$= 2 \times 0.6772 - 1 = 0.3544$$

10. $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$이고, X_1과 X_2는 독립이다.
(a) $Y = pX_1 + (1-p)X_2$의 확률분포를 구하라.
(b) Y의 최소분산과 이 경우에 p를 구하라.

풀이

(a) Y의 평균과 분산은 각각 다음과 같은 정규분포를 이룬다.

평균 : $\mu_Y = E(pX_1 + (1-p)X_2) = pE(X_1) + (1-p)E(X_2) = p\mu_1 + (1-p)\mu_2$
분산 : $\sigma_Y^2 = Var(pX_1 + (1-p)X_2) = p^2\, Var(X_1) + (1-p)^2\, Var(X_2) = p^2\sigma_1^2 + (1-p)^2\sigma_2^2$

(b) $\sigma_Y^2 = p^2\sigma_1^2 + (1-p)^2\sigma_2^2 = (\sigma_1^2 + \sigma_2^2)\left(p - \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^2 + \sigma_2^2\left(\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)$이므로 $p = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$일 때,
Y의 분산 $\sigma_Y^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$이 최소가 된다.

11. 탑의 높이(h)를 측량하기 위하여 두 개의 기구가 사용되며, 이 기구 중에서 정확성이 떨어지는 기구에 의한 오차는 평균 0과 표준편차 $0.0056h$인 정규분포에 따른다. 또한 정확성이 좀 더 좋은 기구에 의하여 측정한 오차는 평균 0과 표준편차 $0.0044h$인 정규분포에 따른다고 한다. 두 개의 기구가 독립이라 할 때, 두 측정값의 평균오차가 탑의 높이의 $0.005h$ 편차 안에 들어올 확률을 구하라.

풀이

정확성이 떨어지는 기구에 의한 오차와 정확성이 좀 더 좋은 기구에 의하여 측정한 오차를 각각 X_1, X_2라고 하자. 그러면 X_1과 X_2는 독립이고 각각 다음 정규분포를 이룬다.

$$X_1 \sim N(0, (0.0056h)^2), \quad X_2 \sim N(0, (0.0044h)^2)$$

따라서 다음을 얻는다.

$$Y = \frac{X_1 + X_2}{2} \sim N\left(0, \frac{(0.0056h)^2 + (0.0044h)^2}{4}\right) = N(0, (0.00356h)^2)$$

그러면 구하고자 하는 확률은 다음과 같다.

$$P(|Y| \le 0.005h) = P(-0.005h \le Y \le 0.005h) = P\left(\frac{-0.005h - 0}{0.00356h} \le Z \le \frac{0.005h - 0}{0.00356h}\right)$$
$$= P(-1.4045 \le Z \le 1.4045) = 2\Phi(1.4) - 1 = 0.84$$

12. 두 회사 A와 B에서 제조된 동일한 장치의 수명은 각각 $X \sim N(70, 16)$과 $Y \sim N(72, 4)$인 정규분포에 따른다고 한다. 구매자는 이 장치를 적어도 75개월 이상 사용하고자 한다. 어느 회사 제품을 선정하는 것이 좋은지 결정하라.

풀이

A 회사에서 생산된 장치의 수명은 $X \sim N(70, 16)$이므로 75를 표준화하면 다음과 같다.

$$z_l = \frac{75 - 70}{4} = 1.25$$

B 회사에서 생산된 장치의 수명은 $Y \sim N(72, 4)$이므로 75를 표준화하면 다음과 같다.

$$z_l = \frac{75 - 72}{2} = 1.5$$

따라서 두 회사의 제품을 75개월 이상 사용할 확률은 각각 다음과 같다.

회사 A : $P(X \ge 75) = P(Z \ge 1.25) = 1 - \Phi(1.25) = 1 - 0.8944 = 0.1056$
회사 B : $P(Y \ge 75) = P(Z \ge 1.50) = 1 - \Phi(1.50) = 1 - 0.9394 = 0.0606$

그러므로 A 회사의 재품을 선정하는 것이 좋다.

13. 특정 시간대에 두 지역의 교차로에서 한 달 동안 차량의 흐름을 조사한 결과 다음 결과를 얻었다. 하루 동안 두 교차로를 지나는 차량은 정규분포를 이룬다고 한다.

차량 수(일)	평균	표준편차
A 지역	257.5	5.5
B 지역	245.3	6.3

(a) A 지역과 B 지역을 지나는 차량의 차에 대한 확률분포를 구하라.
(b) A 지역을 지나는 차량이 B 지역보다 25대 이상 많을 확률을 구하라.
(c) B 지역을 지나는 차량이 A 지역보다 많을 확률을 구하라.

풀이

(a) A 지역과 B 지역을 지나는 차량 수를 각각 X, Y라 하면 $X \sim N(257.5, 5.5^2)$, $Y \sim N(245.3, 6.3^2)$이므로 $X - Y \sim N(12.2, 69.94)$이다.

(b) 구하고자 하는 확률은 다음과 같다.

$$P(X-Y \geq 25) = P\left(Z \geq \frac{25-12.2}{\sqrt{69.94}}\right) = 1-\Phi(1.53) = 1-0.9730 = 0.0270$$

(c) B 지역을 지나는 차량이 A 지역보다 많을 확률은 다음과 같다.

$$P(Y > X) = P(X-Y < 0) = P\left(Z < \frac{0-12.2}{\sqrt{69.94}}\right) = \Phi(-1.46)$$
$$= 1-\Phi(1.46) = 1-0.9279 = 0.0721$$

14. 시중에서 판매되고 있는 두 회사의 땅콩 잼에 포함된 카페인의 양은 각각 다음과 같은 정규분포를 이룬다고 한다.

카페인 양(mg)	평균	표준편차
A 회사	78	3.25
B 회사	75	3.60

(a) A 회사와 B 회사에서 생산된 땅콩 잼에 포함된 카페인 양의 차에 대한 확률분포를 구하라.
(b) A 회사의 땅콩 잼에 포함된 카페인 양이 B 회사보다 8mg 이상 많을 확률을 구하라.
(c) B 회사의 땅콩 잼에 포함된 카페인 양이 A 회사보다 많을 확률을 구하라.

풀이

(a) A 회사와 B 회사에서 생산된 땅콩 잼에 포함된 카페인 양을 각각 X, Y라 하면 $X \sim N(78, 3.25^2)$, $Y \sim N(75, 3.6^2)$이므로 $X-Y \sim N(3, 23.5225)$이다.

(b) 구하고자 하는 확률은 다음과 같다.

$$P(X-Y \geq 8) = P\left(Z \geq \frac{8-3}{\sqrt{23.5225}}\right) = 1-\Phi(1.65) = 1-0.9505 = 0.0495$$

(c) B 회사의 땅콩 잼에 포함된 카페인 양이 A 회사보다 많을 확률은 다음과 같다.

$$P(Y > X) = P(X-Y < 0) = P\left(Z < \frac{0-3}{\sqrt{23.5225}}\right) = \Phi(-0.62)$$
$$= 1-\Phi(0.62) = 1-0.7324 = 0.2676$$

8장

1. 모집단분포가 이산균등분포 $X \sim DU(4)$인 모집단으로부터 크기 2인 표본을 복원추출에 의해 임의로 추출한다.
(a) 표본으로 나올 수 있는 모든 경우를 구하라.
(b) (a)에서 구한 각 표본의 평균을 구하라.
(c) 표본평균 $\overline{X}$의 확률분포를 구하라.
(d) 표본평균 $\overline{X}$의 평균과 분산을 구하라.

풀이

(a) 복원추출에 의하여 두 개를 꺼내므로 표본으로 나올 수 있는 모든 경우는 다음과 같다.

$$\{1,1\}\ \{1,2\}\ \{1,3\}\ \{1,4\}\ \{2,1\}\ \{2,2\}\ \{2,3\}\ \{2,4\}$$

$$\{3,1\}\ \{3,2\}\ \{3,3\}\ \{3,4\}\ \{4,1\}\ \{4,2\}\ \{4,3\}\ \{4,4\}$$

(b) (a)에서 구한 각 표본의 평균을 구하면, 1, 1.5, 2, 2.5, 3, 3.5, 4이다.

(c) $i, j = 1, 2, 3, 4$에 대하여 복원추출에 의해 표본 $\{i, j\}$가 나올 확률은 $\frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$이므로 각 표본과 표본평균의 확률분포는 다음과 같다.

$\overline{x}$	1	1.5	2	2.5	3	3.5	4
$p_{\overline{X}}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

(d) 표본평균 $\overline{X}$의 평균은 $\mu_{\overline{X}} = \sum \overline{x}\, p_{\overline{x}} = 2.5$이고 분산은 $\sigma_{\overline{X}}^2 = E(\overline{X}^2) - E(\overline{X})^2 = 0.625$이다.

2. 모집단분포가 이산균등분포 $X \sim DU(4)$인 모집단으로부터 크기 2인 표본을 비복원추출에 의해 임의로 추출한다.
(a) 표본으로 나올 수 있는 모든 경우를 구하라.
(b) (a)에서 구한 각 표본의 평균을 구하라.
(c) 표본평균 $\overline{X}$의 확률분포를 구하라.
(d) 표본평균 $\overline{X}$의 평균과 분산을 구하라.

풀이

(a) 비복원추출에 의하여 두 개를 꺼내므로 표본으로 나올 수 있는 모든 경우는 다음과 같다.

$$\{1,2\}\ \{1,3\}\ \{1,4\}\ \{2,1\}\ \{2,3\}\ \{2,4\}$$

$$\{3,1\}\ \{3,2\}\ \{3,4\}\ \{4,1\}\ \{4,2\}\ \{4,3\}$$

(b) (a)에서 구한 각 표본의 평균을 구하면, 1.5, 2, 2.5, 3, 3.5이다.

(c) $i, j = 1, 2, 3, 4$에 대하여 비복원추출에 의해 표본 $\{i, j\}$가 나올 확률은 $\frac{1}{4} \cdot \frac{1}{3} = \frac{1}{12}$이므로 각 표본과 표본평균의 확률분포는 다음과 같다.

$\overline{x}$	1.5	2	2.5	3	3.5
$p_{\overline{X}}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(d) 표본평균 $\overline{X}$의 평균은 $\mu_{\overline{X}} = \sum \overline{x}\, p_{\overline{x}} = 2.5$이고 분산은 $\sigma^2_{\overline{X}} = E(\overline{X}^2) - E(\overline{X})^2 = 0.4167$이다.

3. 모평균 $\mu = 8$, 모분산 $\sigma^2 = 4$인 정규모집단으로부터 크기 25인 표본을 임의로 추출할 때, $P(\overline{X} \geq x_0) = 0.05$를 만족하는 x_0을 구하라.

풀이

$\mu = 8$, $\sigma^2 = 4$, $n = 25$이므로 표본평균은 $\overline{X} \sim N(8, 0.16)$이다. 따라서 x_0을 표준화하면 $\frac{x_0 - 8}{0.4} = z_{0.05} = 1.645$이고 구하고자 하는 x_0은 $x_0 = 8 + 0.4 \times 1.645 = 8.658$이다.

4. 우리나라 20세 이상 성인 남자의 혈중 콜레스테롤 수치는 평균 $\mu = 198$, 분산 $\sigma^2 = 36$인 정규분포에 따른다고 가정하자. 단, 단위는 mg/dl이다.

(a) 임의로 1명을 선정하였을 때, 이 사람의 콜레스테롤 수치가 195와 204 사이일 확률을 구하라.

(b) 16명을 임의로 선정하여 표본을 만들 때, 표본평균 $\overline{X}$의 표본분포를 구하라.

(c) 표본평균이 $\mu - 2$와 $\mu + 2$ 사이일 확률을 구하라.

풀이

(a) 임의로 선정한 사람의 콜레스테롤 수치를 X라 하면 $X \sim N(198, 36)$이다. 따라서 구하고자 하는 확률은 다음과 같다.

$$P(195 < X < 204) = P\left(\frac{195 - 198}{6} < Z < \frac{204 - 198}{6}\right) = P(-0.5 < Z < 1)$$
$$= \Phi(1) - [1 - \Phi(0.5)] = 0.8413 - (1 - 0.6915) = 0.5328$$

(b) $\mu = 198$, $\sigma^2 = 36$, $n = 16$이므로 $\overline{X} \sim N(198, 2.25)$이다.

(c) 구하고자 하는 확률은 다음과 같다.

$$P(\mu - 2 < \overline{X} < \mu + 2) = P(-2 < \overline{X} - \mu < 2) = P\left(-\frac{2}{1.5} < Z < \frac{2}{1.5}\right)$$
$$= 2\Phi(1.33) - 1 = 2 \times 0.9082 - 1 = 0.8164$$

5. 모비율이 $p=0.009$인 모집단에서 다음과 같은 크기를 가진 표본을 임의로 선정한다. 이때 표본비율이 $p-0.005$와 $p+0.005$ 사이에 있을 근사확률을 구하라.

(a) $n=25$ (b) $n=100$ (c) $n=400$

풀이

표본비율을 $\hat{p}$라 하면, $\hat{p} \approx N\left(0.009, \dfrac{0.008919}{n}\right)$이다.

(a) $n=25$이므로 $\hat{p} \approx N(0.009, 0.01889^2)$이고, 따라서 구하고자 하는 근사확률은 다음과 같다.

$$\begin{aligned} P(p-0.005<\hat{p}<p+0.005) &= P\left(-\frac{0.005}{0.01889}<\frac{\hat{p}-p}{0.01889}<\frac{0.005}{0.01889}\right) \\ &= P(-0.26<Z<0.26) \approx 2\Phi(0.26)-1 \\ &= 2\times 0.6026-1=0.2052 \end{aligned}$$

(b) $n=100$이므로 $\hat{p} \approx N(0.009, 0.00944^2)$이고, 따라서 구하고자 하는 근사확률은 다음과 같다.

$$\begin{aligned} P(p-0.005<\hat{p}<p+0.005) &= P\left(-\frac{0.005}{0.00944}<\frac{\hat{p}-p}{0.00944}<\frac{0.005}{0.01889}\right) \\ &= P(-0.53<Z<0.53) \approx 2\Phi(0.53)-1 \\ &= 2\times 0.7019-1=0.4038 \end{aligned}$$

(c) $n=400$이므로 $\hat{p} \approx N(0.009, 0.00472^2)$이고, 따라서 구하고자 하는 근사확률은 다음과 같다.

$$\begin{aligned} P(p-0.005<\hat{p}<p+0.005) &= P\left(-\frac{0.005}{0.00472}<\frac{\hat{p}-p}{0.01889}<\frac{0.005}{0.00472}\right) \\ &= P(-1.06<Z<1.06) \approx 2\Phi(1.06)-1 \\ &= 2\times 0.8554-1=0.7108 \end{aligned}$$

6. 우리나라 성인이 하루 동안 마시는 물의 양은 평균 1.65리터, 표준편차 0.07리터인 정규분포에 따른다고 한다. 성인 16명을 임의로 선정하여 하루 동안 마시는 물의 양을 측정하여 다음을 얻었다.

1.74	1.63	1.60	1.74	1.62	1.64	1.61	1.65
1.59	1.34	1.67	1.53	1.66	1.67	1.65	1.72

(a) 통계량 $V=(n-1)\dfrac{S^2}{\sigma^2}$의 분포를 구하라.

(b) 관찰된 표본분산의 값 s_0^2을 구하라.

(c) 이 표본을 이용하여 통계량의 관찰값 $v_0=(n-1)\dfrac{s_0^2}{\sigma^2}$을 구하라.

(d) 표본분산 S^2이 (b)에서 구한 s_0^2보다 클 근사확률을 구하라.

(e) $P(S^2 > v_1)=0.05$인 v_1을 구하라.

풀이

(a) $n=16$, $\sigma^2=0.0049$이므로 V의 확률분포는 $V=\dfrac{15S^2}{0.0049}\sim\chi^2(15)$이다.

(b) 표본평균과 표본분산은 각각 다음과 같다.

$$\bar{x}=\frac{1.74+1.63+\cdots+1.72}{16}=1.62875$$

$$s_0^2=\frac{1}{15}\sum_{i=1}^{16}(x_i-1.62875)^2=\frac{0.0134375}{15}=0.00896$$

(c) 통계량의 관찰값은 다음과 같다.

$$v_0=\frac{15\times 0.00896}{0.0049}=27.43$$

(d) 카이제곱분포표로부터 구하고자 하는 확률은 다음과 같다.

$$P(S^2>0.00896)=P\left(\frac{15S^2}{0.0049}>27.43\right)=P(V>27.43)\approx 0.025$$

(e) $P(S^2>v_1)=P\left(\dfrac{15S^2}{0.0049}>\dfrac{15v_1}{0.0049}\right)=0.05$이므로 구하고자 하는 임계값은 다음과 같다.

$$\chi^2_{0.05}(15)=\frac{15v_1}{0.0049}=25;\ v_1=\frac{25\times 0.0049}{15}=0.00817$$

7. 두 모집단은 평균이 각각 $\mu_1 = 70$, $\mu_2 = 68$이고 모분산은 미지이지만 동일한 정규분포를 이룬다고 한다. 이때 두 모집단으로부터 표본을 추출하여 다음 결과를 얻었다.

표본 A	$n = 12$, $\overline{x} = 72$, $s_1 = 2$
표본 B	$m = 15$, $\overline{y} = 67$, $s_2 = 4$

(a) 합동표본분산 s_p^2을 구하라.

(b) 두 표본평균의 차 $\overline{X} - \overline{Y}$에 대한 확률분포를 구하라.

(c) $P(\overline{X} - \overline{Y} > t_0) = 0.05$를 만족하는 t_0을 구하라.

풀이

(a) $n = 12$, $m = 15$, $s_1 = 2$, $s_2 = 4$이므로 구하고자 하는 합동표본분산은 다음과 같다.

$$s_p^2 = \frac{1}{12 + 15 - 2}(11 \times 4 + 14 \times 16) = \frac{268}{25} = 10.72$$

(b) $s_p = \sqrt{10.72} = 3.2741$, $\mu_1 = 70$, $\mu_2 = 68$ 그리고 $n = 12$, $m = 15$이므로 다음을 얻는다.

$$\mu_1 - \mu_2 = 2, \quad s_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}} = 3.2741 \times \sqrt{\frac{1}{12} + \frac{1}{15}} = 1.268$$

따라서 $\overline{X} - \overline{Y}$에 관한 확률분포는 $\dfrac{(\overline{X} - \overline{Y}) - 2}{1.268} \sim t(25)$이다.

(c) $\dfrac{(\overline{X} - \overline{Y}) - 2}{1.268} \sim t(25)$이므로 다음이 성립한다.

$$P(\overline{X} - \overline{Y} > t_0) = P\left(\frac{(\overline{X} - \overline{Y}) - 2}{1.268} > \frac{t_0 - 2}{1.268}\right) = P(T > t_{0.05}(25)) = P(T > 1.708) = 0.05$$

따라서 구하고자 하는 t_0은 다음과 같다.

$$\frac{t_0 - 2}{1.268} = 1.708 \ ; \ t_0 = 2 + 1.268 \times 1.708 = 4.1657$$

8. 두 종류의 약품 A와 B의 치료율이 동일하게 90%인지 알아보기 위해 임의로 표본을 선정하여 다음 결과를 얻었다.

	표본의 크기	치료율
약품 A	250	91%
약품 B	250	88%

(a) 약품 A와 B의 표본비율을 각각 $\hat{p}_1$, $\hat{p}_2$라 할 때 표본비율의 차 $\hat{p}_1-\hat{p}_2$에 대한 확률분포를 구하라.

(b) $\hat{p}_1-\hat{p}_2$가 3% 이상일 확률을 구하라.

(c) $\hat{p}_1-\hat{p}_2$가 p_0보다 클 확률이 0.05인 p_0을 구하라.

풀이

(a) 약품 A과 B의 치료율을 각각 p_1, p_2라 하면, $p_1=p_2=0.9$이므로 $p_1-p_2=0$이고, $\sqrt{\frac{0.9\times0.1}{250}+\frac{0.9\times0.1}{250}}=0.0268$이므로 $\hat{p}_1-\hat{p}_2$의 확률분포는 $\hat{p}_1-\hat{p}_2 \approx N(0,\, 0.0268^2)$이다.

(b) 구하고자 하는 확률은 다음과 같다.

$$P(\hat{p}_1-\hat{p}_2 \geq 0.03)=P\left(\frac{\hat{p}_1-\hat{p}_2}{0.0268} \geq \frac{0.03}{0.0268}\right)=P(Z\geq 1.12)$$
$$=1-\Phi(1.12)=1-0.8686=0.1314$$

(c) 관찰된 표본비율 $\hat{p}_1=\frac{72}{1500}=0.048$, $\hat{p}_2=\frac{68}{2000}=0.034$이므로 $\hat{p}_1-\hat{p}_2=0.014$이다. 따라서 $\hat{p}_1-\hat{p}_2$이 0.014보다 클 확률은 다음과 같다.

$$P(\hat{p}_1-\hat{p}_2>0.014)=P\left(\frac{\hat{p}_1-\hat{p}_2-0.02}{0.0068}>\frac{0.014-0.02}{0.0068}\right)=P(Z>-0.88)$$
$$=\Phi(0.88)=0.8106$$

(d) $z_{0.05}=1.645$이므로 다음이 성립한다.

$$P(\hat{p}_1-\hat{p}_2>p_0)=P\left(Z>\frac{p_0}{0.0268}\right)=P(Z>z_{0.05})=P(Z>1.645)=0.05$$

따라서 구하고자 하는 p_0은 다음과 같다.

$$\frac{p_0}{0.0268}=1.645; \quad p_0=1.645\times0.0268=0.0441$$

9장

1. 모분산이 4인 정규모집단에서 크기 3인 확률표본을 선정하여 모평균을 추정하기 위해 다음과 같이 점추정량을 설정하였다.

$$\hat{\mu}_1 = \frac{1}{3}(X_1 + X_2 + X_3), \quad \hat{\mu}_2 = \frac{1}{4}(X_1 + X_2 + 2X_3), \quad \hat{\mu}_3 = \frac{1}{3}(3X_1 + X_2 + X_3)$$

(a) 각 추정량의 편의를 구하라.
(b) 불편추정량과 편의추정량을 구하라.
(c) 불편추정량의 분산을 구하고, 최소분산불편추정량을 구하라.

풀이

(a) 각 추정량의 평균을 구하면 다음과 같다.

$$E(\hat{\mu}_1) = \frac{1}{3}E(X_1 + X_2 + X_3) = \frac{1}{3}(\mu + \mu + \mu) = \mu$$
$$E(\hat{\mu}_2) = \frac{1}{4}E(X_1 + X_2 + 2X_3) = \frac{1}{4}(\mu + \mu + 2\mu) = \mu$$
$$E(\hat{\mu}_3) = \frac{1}{3}E(3X_1 + X_2 + X_3) = \frac{1}{3}(3\mu + \mu + \mu) = \frac{5}{3}\mu$$

그러므로 각 추정량의 편의는 다음과 같다.

$$b_1 = E(\hat{\mu}_1) - \mu = 0, \; b_2 = E(\hat{\mu}_2) - \mu = 0, \; b_3 = E(\hat{\mu}_3) - \mu = \frac{2}{3}\mu$$

(b) 불편추정량 : $\hat{\mu}_1$, $\hat{\mu}_2$, 편의추정량 : $\hat{\mu}_3$

(c) $Var(\hat{\mu}_1) = \frac{1}{9}Var(X_1 + X_2 + X_3) = \frac{1}{9}(4 + 4 + 4) = \frac{4}{3}$

$$Var(\hat{\mu}_2) = \frac{1}{16}Var(X_1 + X_2 + 2X_3) = \frac{1}{16}(4 + 4 + 4 \cdot 4) = \frac{3}{2}$$

따라서 $Var(\hat{\mu}_1) < Var(\hat{\mu}_2)$이다. 그러므로 최소분산불편추정량은 $\hat{\mu}_1$ 이다.

2. Θ_1과 Θ_2는 모수 θ에 대한 서로 다른 불편추정량이고 $Var(\Theta_1) = 2\,Var(\Theta_2)$이다. 이때 $a\Theta_1 + b\Theta_2$가 모수 θ에 대한 최소분산불편추정량이 되기 위한 a와 b를 구하라. 단, $a+b=1$이다.

풀이

$a\Theta_1 + b\Theta_2$의 분산은 다음과 같다.

$$Var(a\Theta_1 + b\Theta_2) = a^2\,Var(\Theta_1) + b^2\,Var(\Theta_2) = 2a^2\,Var(\Theta_2) + b^2\,Var(\Theta_2)$$
$$= (2a^2 + b^2)\,Var(\Theta_2)$$

따라서 $2a^2 + b^2$이 최소인 경우에 $Var(a\Theta_1 + b\Theta_2)$가 최소이다. 특히 $a+b=1$이므로 다음을 얻는다.

$$2a^2 + b^2 = 2a^2 + (1-a)^2 = 3\left(a - \frac{1}{3}\right)^2 + \frac{2}{3}$$

그러므로 $a = \frac{1}{3}$, $b = \frac{2}{3}$일 때, $Var(a\Theta_1 + b\Theta_2)$은 최소분산을 갖는다.

3. $X \sim B(5, p)$에 대한 점추정량 $\hat{p} = \frac{X}{6}$에 대하여 이 추정량의 편의와 분산을 구하라.

풀이

$E(X) = 5p$, $Var(X) = 5p(1-p)$이므로 $E(\hat{p}) = E\left(\frac{X}{6}\right) = \frac{5p}{6}$이다. 그러므로 $\hat{p}$의 편의와 분산은 각각 다음과 같다.

$$\text{bias} = E(\hat{p}) - p = \frac{5p}{6} - p = -\frac{p}{6}$$
$$Var(\hat{p}) = Var\left(\frac{X}{6}\right) = \frac{Var(X)}{36} = \frac{5p(1-p)}{36}$$

4. 모수 λ인 푸아송분포를 이루는 모집단으로부터 크기 12인 표본을 선정하여 다음을 얻었다.

8	5	0	10	0	3	4	11	2	8	6	7

(a) 모수 λ에 대한 불편추정량을 하나 제시하라.
(b) (a)에서 구한 불편추정량에 대한 λ의 추정값을 구하라.

풀이

(a) 모집단 분포는 $X \sim P(\lambda)$이므로 모평균 μ는 모수 λ와 일치한다. 그리고 표본평균 $\overline{x}$는 모평균 μ에 대한 불편추정량의 하나이므로 $\hat{\lambda} = \frac{1}{12}(X_1 + X_2 + \cdots + X_{12})$이다.

(b) 표본의 측정값이 $x_1 = 8$, $x_2 = 5$, $\cdots$, $x_{12} = 7$이므로 $\hat{\lambda} = \frac{1}{12}(x_1 + x_2 + \cdots + x_{12}) = 5.33$이다.

5. $\{X_1, X_2, \cdots, X_{10}\}$은 모수 p인 모집단으로부터 추출된 크기 10인 확률표본이고, 모비율을 추정하기 위한 불편추정량으로 표본비율 $\hat{p}=\dfrac{\overline{X}}{10}$를 사용한다. 관찰된 표본이 다음과 같을 때, 모수 p에 대한 점추정값을 구하라.

1	8	2	5	7	6	2	9	4	7

풀이

표본평균은 $\overline{x}=\dfrac{1}{10}\sum x_i = 5.1$이므로 모수 p에 대한 추정값 $\hat{p}=\dfrac{5.1}{10}=0.51$이다.

6. $\{X_1, X_2, \cdots, X_{10}\}$은 구간 $(0, b)$에서 균등분포를 이루는 모집단으로부터 추출된 확률표본이고, 관찰값이 다음 표와 같다고 한다. 이때 모평균을 추정하기 위하여 표본평균을 사용한다.

10	7	11	12	8	8	9	10	9	13

(a) 모평균 μ에 대한 추정값을 구하라.
(b) 모수 b에 대한 추정값을 구하라.

풀이

(a) 표본평균이 모평균에 대한 불편추정량이므로, 모평균에 대한 불편추정값은 다음과 같다.

$$\hat{\mu}=\frac{1}{10}(x_1+x_2+\cdots+x_{10})=9.7$$

(b) 모평균의 추정값이 9.7이고, 모평균은 $\mu=\dfrac{b+0}{2}$이므로 b에 대한 추정값은 $\hat{\mu}=\dfrac{\hat{b}}{2}=9.7$로부터 $\hat{b}=19.4$이다.

7. 모분산 4인 정규모집단의 모평균에 대한 95% 신뢰도를 갖는 구간을 추정하고자 한다. 표본의 크기가 다음과 같을 때 오차한계를 구하라.
(a) $n=50$ (b) $n=100$ (c) $n=200$ (d) $n=500$

풀이

(a) $e=1.96\times\dfrac{2}{\sqrt{50}}=0.5544$

(b) $e=1.96\times\dfrac{2}{\sqrt{100}}=0.3920$

(c) $e=1.96\times\dfrac{2}{\sqrt{200}}=0.2772$

(d) $e=1.96\times\dfrac{2}{\sqrt{500}}=0.1753$

8. 다음과 같은 크기의 표본을 선정하여 모비율에 대한 95% 신뢰도를 갖는 구간을 추정하고자 한다. 표본비율이 $\hat{p}=0.75$일 때 오차한계를 구하라.

(a) $n=50$　　(b) $n=100$　　(c) $n=200$　　(d) $n=500$

풀이

(a) $e=1.96\times\sqrt{\dfrac{(0.75)(0.25)}{50}}=0.1200$

(b) $e=1.96\times\sqrt{\dfrac{(0.75)(0.25)}{100}}=0.0849$

(c) $e=1.96\times\sqrt{\dfrac{(0.75)(0.25)}{200}}=0.0600$

(d) $e=1.96\times\sqrt{\dfrac{(0.75)(0.25)}{500}}=0.0380$

9. 모분산이 2인 정규모집단에서 크기 20인 표본을 조사한 결과 $\sum_{i=1}^{20}x_i=48.6$이었다.

(a) 표본평균에 의한 모평균 μ의 점추정값을 구하라.

(b) 모평균에 대한 95% 신뢰구간을 구하라.

풀이

(a) 표본평균에 의한 모평균의 점추정값은 $\hat{\mu}=\dfrac{48.6}{20}=2.43$이다.

(b) $\sigma^2=2$, $n=20$이므로 95% 오차한계는 다음과 같다.

$$e=(1.96)\mathrm{S.E.}(\overline{X})=(1.96)\sqrt{\frac{2}{20}}=0.6198$$

따라서 모평균에 대한 95% 신뢰구간은 $(2.43-0.6198,\ 2.43+0.6198)=(1.8102,\ 3.0498)$이다.

10. 정규모집단으로부터 추출한 크기 100인 표본에 대하여 $\overline{x}=24.5$, $s^2=5.05$를 얻었다.

(a) 모평균에 대한 점추정값을 구하라.

(b) 95% 오차 한계를 구하라.

(c) 95% 신뢰구간을 구하라.

풀이

(a) 표본평균이 $\overline{x}=24.5$이므로 모평균에 대한 점추정값은 $\hat{\mu}=24.5$이다.

(b) 모분산을 모르지만 표본의 크기가 충분히 크므로 z-추정을 한다. 이때 95% 오차 한계는 $(1.96)\times\dfrac{5.05}{\sqrt{100}}=0.9898$이다.

(c) 신뢰구간의 하한과 상한은 각각 다음과 같다.

$$l = \overline{x} - \text{S.E.}(\overline{X}) = 24.5 - 0.9898 = 23.5102$$

$$u = \overline{x} + \text{S.E.}(\overline{X}) = 24.5 + 0.9898 = 25.4898$$

따라서 95% 신뢰구간은 $(23.5102,\ 25.4898)$이다.

11. 어느 회사에서 제조되는 저항기는 표준편차가 2Ω인 정규분포에 따른다고 한다. 이 회사에서 생산되는 저항기의 평균 저항을 알기 위하여 40개를 임의로 선정하여 측정한 결과 다음과 같았다. 이 회사에서 제조되는 저항기의 평균 저항에 대한 90% 신뢰구간을 구하라.

38.7	40.6	40.5	40.3	40.9	38.2	38.7	40.6	40.2	40.7
39.2	40.8	41.3	39.4	40.2	41.6	41.9	42.0	40.0	39.1
38.2	38.4	39.0	41.7	41.9	39.4	39.4	40.4	38.6	41.6
41.3	40.2	41.0	39.4	40.7	38.4	38.2	40.7	39.0	40.5

풀이

표본으로부터 얻은 표본평균은 $\overline{x} = 40.075$이고 $\sigma = 2$, $z_{0.05} = 1.645$, $n = 40$이므로 90% 신뢰구간은 다음과 같다.

$$\left(\overline{x} - z_{0.05}\frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{0.05}\frac{\sigma}{\sqrt{n}}\right) = \left(40.075 - (1.645)\cdot\frac{2}{\sqrt{40}},\ 40.075 + (1.645)\cdot\frac{2}{\sqrt{40}}\right)$$
$$= (39.555,\ 40.595)$$

12. 다음 자료는 어느 직장에 근무하는 직원 20명에 대한 혈중 콜레스테롤 수치를 조사한 자료이다. 이 직장에 근무하는 직원들의 콜레스테롤 평균 수치에 대한 95% 신뢰구간을 구하라. 단, 콜레스테롤 수치는 정규분포 $N(\mu,\ 144)$에 따른다고 한다.

193.27	193.88	253.26	237.15	188.83	200.56	274.31	230.36	212.08	222.19
198.48	202.50	215.35	218.95	233.16	222.23	218.53	204.64	206.72	199.37

풀이

표본으로부터 얻은 표본평균은 $\overline{x} = 216.29$이다. 한편 $\sigma = 12$, $n = 20$, $z_{0.025} = 1.96$이므로 95% 신뢰구간은 다음과 같다.

$$\left(\overline{x} - z_{0.025}\frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{0.025}\frac{\sigma}{\sqrt{n}}\right) = \left(216.29 - (1.96)\cdot\frac{12}{\sqrt{20}},\ 216.29 + (1.96)\cdot\frac{12}{\sqrt{20}}\right)$$
$$= (211.03,\ 221.55)$$

13. 모분산이 각각 $\sigma_1^2 = 9$, $\sigma_2^2 = 4$이고 독립인 두 정규모집단으로부터 각각 크기 25와 36인 표본을 추출하여 표본평균 $\overline{x} = 35$, $\overline{y} = 32.5$를 얻었다. 두 모평균 차에 대한 99% 신뢰구간을 구하라.

풀이

$\overline{x} = 35$, $\overline{y} = 32.5$이므로 $\mu_1 - \mu_2$에 대한 점추정은 $\overline{x} - \overline{y} = 35 - 32.5 = 2.5$이다. $\sigma_1^2 = 9$, $\sigma_2^2 = 4$이고 $n = 25$, $m = 36$이므로 표준오차는 $\text{S.E.}(\overline{X} - \overline{Y}) = \sqrt{\frac{9}{25} + \frac{4}{36}} = \sqrt{0.4711} = 0.6864$이므로 95% 오차한계는
$e = 2.58 \times \text{S.E.}(\overline{X} - \overline{Y}) = 2.58 \times 0.6864 = 1.771$이다. 그러므로 구하고자 하는 신뢰구간은
$(2.5 - 1.771,\ 2.5 + 1.771) = (0.729,\ 4.271)$이다.

14. 우리나라 80세 이상인 남자와 여자의 혈당치는 각각 표준편차가 13mg/dℓ와 11mg/dℓ인 정규분포에 따른다고 한다. 전국 80세 이상인 남자 15명과 여자 20명을 대상으로 조사하여 각각 평균 혈당치가 115.9 mg/dℓ와 104mg/dℓ를 얻었다. 전국 80세 이상 남자와 여자의 평균 혈당치의 차에 대한 95% 신뢰구간을 구하라.

풀이

80세 이상인 남자와 여자의 표본평균을 각각 $\overline{x}$, $\overline{y}$라 하면 $\mu_M - \mu_F$에 대한 점추정값은 $\overline{x} - \overline{y} = 11.9$이다. $\sigma_M^2 = 169$, $\sigma_F^2 = 121$이고 $n_M = 15$, $n_F = 20$이므로 표준오차는 다음과 같다.

$$\text{S.E.}(\overline{X} - \overline{Y}) = \sqrt{\frac{169}{15} + \frac{121}{20}} = \sqrt{17.3167} = 4.1613$$

그리고 95% 오차한계는 $e = 1.96 \times \text{S.E.}(\overline{X} - \overline{Y}) = 1.96 \times 4.1613 = 8.156$이므로 구하고자 하는 신뢰구간은 $(11.9 - 8.156,\ 11.9 + 8.156) = (3.744,\ 20.056)$이다.

15. 한국영양학회지에 발표된 '고등학생의 식습관과 건강인지에 관한 연구'에 따르면, 서울지역 고등학교 남학생 260명과 여학생 250명을 표본조사하여 다음 결과를 얻었다.

- 주 1회 이상 아침식사를 결식하는 비율이 남학생 41.1%, 여학생 44.1%이다.
- 자신이 건강하다고 생각하는 비율은 남학생 68.9%, 여학생 55.6%이다.
- 평균 키는 남학생이 174.1cm이고 여학생이 161.6cm이다.
- 평균 몸무게은 남학생이 65.9kg이고 여학생이 52.5kg이다.

(a) 서울지역 고등학교 남학생의 평균 키와 여학생의 평균 키의 차에 대한 95% 신뢰구간을 구하라. 단, 남학생과 여학생의 키에 대한 표준편차는 각각 5cm와 3cm이고 정규분포에 따른다고 가정한다.
(b) 서울지역 고등학교 남학생의 평균 몸무게와 여학생의 평균 몸무게의 차에 대한 99% 신뢰구간을 구하라. 단, 남학생과 여학생의 키에 대한 표준편차는 각각 4.5kg과 2.5kg이고 정규분포에 따른다고 가정한다.
(c) 서울지역 고등학생이 주 1회 이상 아침식사를 결식하는 여학생의 비율과 남학생의 비율의 차에 대한 90% 신뢰구간을 구하라.
(d) 서울지역 고등학생이 건강하다고 생각하는 남학생의 비율과 여학생의 비율의 차에 대한 95% 신뢰구간을 구하라.

풀이

(a) 남학생의 평균 키를 μ_1, 여학생의 평균 키를 μ_2라 하면, $\sigma_1^2=25$, $\sigma_2^2=9$, $n=260$, $m=250$, $\overline{x}=174.1$, $\overline{y}=161.6$이므로 $\mu_1-\mu_2$의 점추정은 $\overline{x}-\overline{y}=174.1-161.6=12.5$이다. 그러므로 95% 오차한계는 다음과 같다.

$$e=(1.96)\mathrm{S.E.}(\overline{X}-\overline{Y})=(1.96)\sqrt{\frac{25}{260}+\frac{9}{250}}=0.713$$

따라서 남학생과 여학생의 비율의 차에 대한 95% 신뢰구간은 다음과 같다.

$$(12.5-0.713,\ 12.5+0.713)=(11.787,\ 13.213)$$

(b) 남학생의 평균 몸무게를 μ_1, 여학생의 평균 몸무게를 μ_2라 하면, $\sigma_1^2=20.25$, $\sigma_2^2=6.25$, $n=260$, $m=250$, $\overline{x}=65.9$, $\overline{y}=52.5$이므로 $\mu_1-\mu_2$의 점추정은 $\overline{x}-\overline{y}=65.9-52.5=13.4$이다. 그러므로 99% 오차한계는 다음과 같다.

$$e=(2.58)\mathrm{S.E.}(\overline{X}-\overline{Y})=(2.58)\sqrt{\frac{20.25}{260}+\frac{6.25}{250}}=0.828$$

따라서 남학생과 여학생의 비율의 차에 대한 99% 신뢰구간은 다음과 같다.

$$(13.4-0.8283,\ 13.4+0.828)=(12.572,\ 14.228)$$

(c) 여학생의 결식률을 p_1, 남학생의 결식률을 p_2라 하면, $n=250$, $m=260$, $\hat{p}_1=0.441$, $\hat{p}_2=0.411$이므로 p_1-p_2의 점추정은 $\hat{p}_1-\hat{p}_2=0.441-0.411=0.03$이다. 그러므로 90% 오차한계는 다음과 같다.

$$e=(1.645)\mathrm{S.E.}(\hat{p}_1-\hat{p}_2)=(1.645)\sqrt{\frac{(0.441)(0.559)}{250}+\frac{(0.411)(0.589)}{260}}=0.072$$

따라서 여학생과 남학생의 결식률의 차에 대한 90% 신뢰구간은 다음과 같다.

$$(0.03-0.072,\ 0.03+0.072)=(-0.042,\ 0.102)$$

(d) 건강하다고 생각하는 남학생의 비율을 p_1, 여학생의 비율을 p_2라 하면, $n=260$, $m=250$, $\hat{p}_1=0.689$, $\hat{p}_2=0.556$이므로 p_1-p_2의 점추정은 $\hat{p}_1-\hat{p}_2=0.689-0.556=0.133$이다. 그러므로 95% 오차한계는 다음과 같다.

$$e=(1.96)\mathrm{S.E.}(\hat{p}_1-\hat{p}_2)=(1.96)\sqrt{\frac{(0.689)(0.311)}{260}+\frac{(0.556)(0.444)}{250}}=0.0835$$

따라서 남학생과 여학생의 비율의 차에 대한 95% 신뢰구간은 다음과 같다.

$$(0.133-0.0835,\ 0.133+0.0835)=(0.0495,\ 0.2165)$$

16. 어느 회사에서 생산한 장난감 5,000개 중에 포함된 불량품의 수를 알고자 한다.
(a) 임의로 200개를 조사하여 불량품이 5개이었다. 이 장난감의 불량률에 대한 90% 신뢰구간을 구하라.
(b) (a)의 결과를 이용하여 5,000개 중 포함된 불량품 수의 최솟값과 최댓값을 구하라.

풀이

(a) 200개 포함된 불량품의 수를 X라 하면 $x=5$, $n=200$이므로 모비율 p에 대한 점추정값은 $\hat{p}=\dfrac{5}{200}=0.025$이고, 90% 오차한계는 $e=1.645\times\sqrt{\dfrac{0.025\times0.975}{200}}=0.0182$이다. 그러므로 구하고자 하는 신뢰구간은 $(0.025-0.0182,\ 0.025+0.0182)=(0.0068,\ 0.0432)$이다.

(b) 신뢰구간의 최소는 0.0068, 최대는 0.0432이므로 5,000개 중에서 최소 34개, 최대 216개다.

17. 새로 개발된 50나노 디램의 신뢰성을 조사하기 위하여 250개를 임의로 추출하여 조사한 결과 7개가 불량이었다. 이 회사에서 생산된 램의 불량률 p에 대한 99% 신뢰구간을 구하라.

풀이

250개 램 중에서 불량품의 수가 7개이므로 표본비율은 $\hat{p}=\dfrac{7}{250}=0.028$이므로 모비율 p에 대한 점추정값은 $\hat{p}=0.028$이다. 또한 99% 오차한계는 $e=2.58\times\sqrt{\dfrac{0.028\times0.972}{250}}=0.0269$이므로 99% 신뢰구간은 $(0.028-0.0269,\ 0.028+0.0269)=(0.0011,\ 0.0549)$이다.

18. 두 종류의 약품 A, B의 효능을 조사하기 위하여, 400명의 환자에게 두 약품을 각각 200명씩 투여하여 각각 165명과 150명이 효과가 있다는 사실을 얻었다. 두 약품 A와 B의 효율의 차이에 대한 95% 신뢰구간을 구하라.

풀이

A 약품과 B 약품의 효율을 각각 p_1, p_2라 하면, 두 약품의 효율에 대한 점추정값은 각각 $\hat{p}_1=\dfrac{165}{200}=0.825$, $\hat{p}_2=\dfrac{150}{200}=0.75$이다. 그러므로 $\hat{p}_1-\hat{p}_2=0.825-0.75=0.075$이고, 95% 오차한계는 다음과 같다.

$$e=1.96\times\sqrt{\frac{0.825\times0.175+0.75\times0.25}{200}}=0.0798$$

따라서 두 약품 A와 B의 효율에 대한 차의 95% 신뢰구간은 다음과 같다.

$$(0.075-0.0798,\ 0.075+0.0798)=(-0.005,\ 0.1548)$$

19. $B(400,\ p_1)$로부터 $x=120$인 성공을 얻었고, $B(200,\ p_2)$로부터 $y=60$인 성공을 얻었다. 이때 p_1-p_2에 대한 90% 신뢰구간을 구하라.

풀이

$\hat{p}_1=\dfrac{120}{400}=0.3$, $\hat{p}_2=\dfrac{60}{200}=0.3$이므로 $\hat{p}_1-\hat{p}_2=0$이고, 90% 오차한계는 다음과 같다.

$$e=1.645\sqrt{\frac{0.3\times0.7}{400}+\frac{0.3\times0.7}{200}}=0.0774$$

그러므로 90% 신뢰구간은 $(-0.0774,\ 0.0774)$이다.

20. 특수 공정에 의하여 생산되는 강철판은 모표준편차가 4mm인 정규분포에 따른다고 한다. 허용오차 0.5mm 이하에서 평균 두께에 대한 95% 신뢰구간을 얻고자 한다. 이때 표본조사해야 할 강철판의 개수를 구하라.

풀이

$\sigma=4$, $d=0.5$이므로 모평균에 대한 95% 신뢰구간을 얻기 위한 표본의 크기는 다음과 같다.

$$n\ge\left(1.96\times\frac{4}{0.5}\right)^2=245.86;\ n=246$$

21. 표준편차가 0.51Ω과 0.49Ω인 정규분포에 따르는 두 종류의 동선에 대하여 평균 저항의 차에 대한 $\pm0.15\Omega$ 오차범위에서 99% 신뢰구간을 얻고자 한다. 이때 표본조사해야 할 동선의 개수를 구하라. 단, 두 표본의 크기는 동일하다.

풀이

$\sigma_1^2=0.2601$, $\sigma_2^2=0.2401$이고 오차범위가 $d=0.1$이므로 99% 신뢰구간을 얻기 위한 표본의 크기는 다음과 같다.

$$n=m\ge\left(\frac{2.58}{0.15}\right)^2(0.2601+0.2401)=147.979;\ n=m=148$$

22. 핸드폰의 불량률 p를 오차범위 2%와 신뢰도 90%에서 구간추정하고자 한다. 이를 위하여 예비로 100개를 조사하여 $\hat{p}=0.04$를 얻었다면, 추가적으로 조사해야할 개수를 구하라.

풀이

$\hat{p}=0.04$, $\hat{q}=0.96$이고 오차범위가 $d=0.02$이므로 신뢰도 90%에서 구간추정하기 위한 표본의 크기는 다음과 같다.

$$n\ge\left(\frac{1.645}{0.02}\right)^2\times0.04\times0.96=259.778;\ n=260$$

한편 예비로 100개를 조사하였으므로 160개를 더 조사해야 한다.

10장

1. 환한 곳에서 어두운 방에 들어갈 경우에 우리의 눈이 어두운 곳에 적응하는데 걸리는 시간은 평균 7.88초라고 안과 협회에서 주장한다. 이 주장을 검정하기 위해 임의로 50명을 추출하여 실험한 결과 적응시간이 평균 7.83초였다. 이때 적응시간은 표준편차 0.15인 정규분포에 따른다고 한다.

(a) 귀무가설과 대립가설을 설정하라.
(b) 유의수준 5%에서 기각역을 구하라.
(c) 검정통계량의 관찰값을 구하라.
(d) p-값을 구하라.
(e) 검정통계량의 관찰값 또는 p-값을 이용하여 유의수준 5%에서 귀무가설을 검정하라.

풀이

(a) 귀무가설은 $H_0 : \mu_0 = 7.88$이고 대립가설은 $H_1 : \mu_0 \neq 7.88$이다.

(b) 유의수준 $\alpha = 0.05$에 대한 양측검정 기각역은 $Z \le -1.96$ 또는 $Z \ge 1.96$이다.

(c) 모표준편차가 $\sigma = 0.15$이므로 검정통계량은 $Z = \dfrac{\overline{X} - 7.88}{0.15/\sqrt{50}}$이고 $\overline{x} = 7.83$이므로 검정통계량의 관찰값은 $z_0 = \dfrac{7.83 - 7.88}{0.15/\sqrt{50}} = -2.36$이다.

(d) p-값$= 2[1 - P(Z < 2.36)] = 2(1 - 0.9909) = 0.0182$

(e) 검정통계량의 관찰값 $z_0 = -2.36$이 기각역 안에 놓이므로 귀무가설 $H_0 : \mu_0 = 7.88$을 유의수준 5%에서 귀무가설을 기각한다. 또한 p-값$= 0.0182 < \alpha = 0.05$이므로 귀무가설을 유의수준 5%에서 기각한다.

2. 어느 회사는 근로자의 혈중 콜레스테롤 평균수치가 220mg/dl 이상이라 한다. 이 주장을 확인하기 위하여 이 회사에 근무하는 근로자 20명을 무작위로 선정하여 측정한 결과 다음과 같았다. 단, 전체 근로자의 콜레스테롤 수치는 $\sigma = 15.4$인 정규분포에 따른다고 알려져 있다.

194.2	192.8	243.2	237.1	198.8	202.5	245.3	230.3	221.0	203.1
201.4	212.5	215.3	218.9	223.1	212.2	218.5	214.6	204.7	199.3

(a) 귀무가설과 대립가설을 설정하라.
(b) 유의수준 5%에서 기각역을 구하라.
(c) 검정통계량의 관찰값을 구하라.
(d) 기각역을 구하여 유의수준 5%에서 귀무가설을 검정하라.
(e) p-값을 구하라.
(f) p-값을 이용하여 유의수준 5%에서 귀무가설을 검정하라.

풀이

(a) 귀무가설은 $H_0 : \mu_0 \geq 220$이고 대립가설은 $H_1 : \mu_0 < 220$이다.

(b) 유의수준 $\alpha = 0.05$에 대한 하단측검정의 기각역은 $Z \leq -1.645$이다.

(c) $\sigma = 15.4$, $n = 20$이므로 검정통계량은 $Z = \dfrac{\overline{X} - 220}{15.4/\sqrt{20}}$이고 표본평균을 구하면 $\overline{x} = 214.44$이므로 검정통계량의 관찰값은 $z_0 = \dfrac{214.44 - 220}{15.4/\sqrt{20}} = -1.61$이다.

(d) 검정통계량의 관찰값 $z_0 = -1.61$이 기각역 $Z < -1.645$ 안에 놓이지 않으므로 귀무가설 $H_0 : \mu_0 \geq 220$을 유의수준 5%에서 기각할 수 없다.

(e) 검정통계량의 관측값이 $z_0 = -1.61$이므로 p-값은 다음과 같다.

$$p - \text{값} = P(Z < -1.61) = 1 - P(Z < 1.61) = 1 - 0.9463 = 0.0537$$

(e) $p - \text{값} = 0.0537 > \alpha = 0.05$이므로 귀무가설 $H_0 : \mu_0 \geq 220$을 유의수준 5%에서 기각할 수 없다.

3. 어느 지자체의 하루에 배출되는 폐기물량이 300톤 이상이라고 환경단체가 주장한다. 이러한 사실을 알아보기 위하여 일일 폐기물량을 임의로 50번 측정한 결과 평균 폐기물양이 290톤이었다. 유의수준 2.5%에서 환경단체의 주장을 검정하라. 단, 이 도시의 일일 폐기물양은 표준편차 35톤인 정규분포에 따르는 것으로 알려져 있다.

풀이

① 귀무가설 $H_0 : \mu \geq 300$와 대립가설 $H_0 : \mu < 300$를 설정한다.
② 유의수준 $\alpha = 0.025$에 대한 하단측검정의 기각역은 $Z \leq -1.96$이다.
③ 표준편차가 $\sigma = 35$, $n = 50$이므로 검정통계량은 $Z = \dfrac{\overline{X} - 300}{35/\sqrt{50}} = \dfrac{\overline{X} - 300}{4.9498}$이고 $\overline{x} = 290$이므로 검정통계량의 관찰값은 $z_0 = \dfrac{290 - 300}{4.9498} = -2.02$이다.
④ 검정통계량의 관찰값이 기각역 안에 들어가므로 이 도시의 일일 폐기물량은 300톤 이상이라는 근거는 타당성이 없다.

4. 어느 회사에서 생산하는 디지털 사진기용 전구의 수명이 900시간 이하라고 소비자 단체는 주장한다. 이것을 알아보기 위하여 전구 45개를 수거하여 조사한 결과 평균 909시간이었다. 유의수준 5%에서 소비자 단체의 주장을 검정하라. 단, 전구의 수명은 표준편차가 45시간인 정규분포에 따른다고 한다.

풀이

① 귀무가설 H_0 : $\mu \le 900$과 대립가설 H_1 : $\mu > 900$을 설정한다.

② 유의수준 $\alpha = 0.05$에 대한 상단측검정이므로 기각역은 $Z \ge 1.645$이다.

③ $\sigma = 45$이므로 검정통계량은 $Z = \dfrac{\overline{X} - 900}{45/\sqrt{60}} = \dfrac{\overline{X} - 900}{5.809}$이다.

④ $\overline{x} = 909$이므로 검정통계량의 관찰값은 $z_0 = \dfrac{909 - 900}{5.809} = 1.54$이다.

⑤ 검정통계량의 관찰값 $z_0 = 1.54$는 기각역 안에 놓이지 않으므로 소비자 단체의 주장은 타당성이 있다.

5. 두 그룹 A와 B의 모평균이 동일한지 알아보기 위해 각각 $n = 65$, $m = 96$인 표본조사하여 $\overline{x} = 201$와 $\overline{y} = 199$를 얻었다. 이것을 근거로 유의수준 5%에서 모평균이 동일한지 검정하라. 단, 그룹 A와 B는 각각 $\sigma_1 = 5$와 $\sigma_2 = 6$인 정규분포에 따른다고 알려져 있다.

풀이

① 그룹 A와 B의 평균을 각각 μ_1, μ_2라 하면 $\mu_1 = \mu_2$임을 보이고자 하므로 귀무가설 $H_0 : \mu_1 - \mu_2 = 0$과 대립가설 $H_1 : \mu_1 - \mu_2 \ne 0$을 설정한다.

② 유의수준 $\alpha = 0.05$에 대한 양측검정의 기각역은 $Z \le -1.96$ 또는 $Z \ge 1.96$이다.

③ 표본표준편차가 각각 $\sigma_1 = 5$, $\sigma_2 = 6$이고 $n = 65$, $m = 96$므로 검정통계량은 다음과 같다.

$$Z = \frac{(\overline{X} - \overline{Y}) - 0}{\sqrt{(5^2/65) + (6^2/96)}} = \frac{\overline{X} - \overline{Y}}{0.8716}$$

④ $\overline{x} = 201$, $\overline{y} = 199$이므로 검정통계량의 관찰값은 $z_0 = \dfrac{201 - 199}{0.8716} = 2.29$이다.

⑤ 관찰값 $z_0 = 2.29$은 기각역 안에 놓이므로 $H_0 : \mu_1 - \mu_2 = 0$을 기각한다. 즉, $\mu_1 = \mu_2$라는 결론은 신빙성이 없다.

6. 두 그룹 A와 B의 모평균 μ_1과 μ_2에 대한 주장 $H_0 : \mu_1 \le \mu_2$와 $H_1 : \mu_1 > \mu_2$을 유의수준 1%에서 검정하기 위하여 동일한 크기 20인 표본조사하여 $\overline{x}_A = 218.5$, $\overline{x}_B = 209.7$을 얻었다. 그룹 A는 $\sigma_1 = 14$이고 그룹 B는 다음 표준편차 σ_2인 정규분포에 따른다고 할 때, p-값을 구하여 μ_1과 μ_2에 대한 주장을 검정하라. 단, $P(Z < 2.326) = 0.99$이다.

(a) $\sigma_2 = 7$ (b) $\sigma_2 = 14$ (c) $\sigma_2 = 28$

풀이

(a) 다음 순서에 따라 검정한다.

① 유의수준 $\alpha = 0.01$에 대한 상단측검정의 기각역은 $Z \ge 2.326$이다.

② 모표준편차가 각각 $\sigma_1 = 14$, $\sigma_2 = 7$이고 $n = m = 20$이므로 검정통계량은 다음과 같다.

$$Z = \frac{(\overline{X} - \overline{Y}) - 0}{\sqrt{(14^2 + 7^2)/20}} = \frac{\overline{X} - \overline{Y}}{3.5}$$

③ $\bar{x}-\bar{y}=8.8$이므로 검정통계량의 관찰값은 $z_0=\dfrac{8.8}{3.5}=2.51$이다.

④ p-값$=P(Z>2.51)=0.006<0.01$이므로 $H_0:\mu_1\le\mu_2$을 기각한다.

(b) 다음 순서에 따라 검정한다.

① 유의수준 $\alpha=0.01$에 대한 상단측검정의 기각역은 $Z\ge 2.326$이다.
② 모표준편차가 각각 $\sigma_1=14$, $\sigma_2=14$이고 $n=m=20$이므로 검정통계량은 다음과 같다.

$$Z=\frac{(\overline{X}-\overline{Y})-0}{\sqrt{(14^2+14^2)/20}}=\frac{\overline{X}-\overline{Y}}{4.43}$$

③ $\bar{x}-\bar{y}=8.8$이므로 검정통계량의 관찰값은 $z_0=\dfrac{8.8}{4.43}=1.99$이다.

④ p-값$=P(Z>1.99)=0.0233>0.01$이므로 $H_0:\mu_1\le\mu_2$을 기각하지 않는다.

(c) 다음 순서에 따라 검정한다.
① 유의수준 $\alpha=0.01$에 대한 상단측검정의 기각역은 $Z\ge 2.326$이다.
② 모표준편차가 각각 $\sigma_1=14$, $\sigma_2=28$이고 $n=m=20$이므로 검정통계량은 다음과 같다.

$$Z=\frac{(\overline{X}-\overline{Y})-0}{\sqrt{(14^2+28^2)/20}}=\frac{\overline{X}-\overline{Y}}{7}$$

③ $\bar{x}-\bar{y}=8.8$이므로 검정통계량의 관찰값은 $z_0=\dfrac{8.8}{7}=1.23$이다.

④ p-값$=P(Z>1.23)=0.1093>0.01$이므로 $H_0:\mu_1\le\mu_2$을 기각하지 않는다.

7. 매뉴얼에 의하면 응급차는 출동하여 적어도 64% 이상 응급환자를 살려야한다. 구급대가 이를 지키는지 알아보기 위하여 응급차의 출동 기록을 살펴본 결과, 250회 출동 중에서 환자를 살린 횟수가 174회이었다.
(a) 응급차는 매뉴얼에 따랐는지 유의수준 5%에서 검정하라.
(b) p-값을 구하고 유의수준 5%에서 검정하라.

풀이

(a) 다음 순서에 따라 검정한다.

① 귀무가설은 $H_0:p\ge 0.64$이고 대립가설은 $H_1:p<0.64$이다.

② 유의수준 $\alpha=0.05$에 대한 하단측검정 기각역은 $Z\le -1.645$이다.

③ 검정통계량은 $Z=\dfrac{\hat{p}-0.64}{\sqrt{0.64\times 0.46/250}}=\dfrac{\hat{p}-0.64}{0.0343}$이다.

④ 표본비율이 $\hat{p}=\dfrac{174}{250}=0.696$이므로 검정통계량의 관찰값은 $z_0=\dfrac{0.64-0.696}{0.0343}=-1.63$이다.

⑤ 검정통계량의 관찰값 $z_0 = -1.63$은 기각역 안에 놓이지 않으므로 $H_0 : p \ge 0.64$를 기각한다. 즉, 매뉴얼에 따르지 않았다는 주장이 타당성을 갖는다.

(b) 검정통계량의 관찰값이 $z_0 = -1.63$이므로 p-값$= P(Z < -1.63) = 0.0516 > 0.05$이다. 그러므로 유의수준 5%에서 귀무가설을 기각할 수 없다.

8. 자동차 보험회사는 미혼인 가입자와 기혼인 가입자를 동일하게 500명씩 선정하여 조사하였다. 미혼인 가입자의 45명과 기혼인 가입자의 34명이 지난해 사고를 냈을 때, 사고율이 동일한지 유의수준 5%에서 검정하라.

풀이

① 미혼과 기혼에 의한 사고율을 각각 p_1, p_2라 하면, 귀무가설은 $H_0 : p_1 - p_2 = 0$이고 대립가설은 $H_1 : p_1 - p_2 \ne 0$이다.

② 유의수준 5%에서 양측검정이므로 기각역은 $Z \le -1.96$ 또는 $Z \ge 1.96$이다.

③ 합동표본비율은 $\hat{p} = \dfrac{45+34}{500+500} = 0.079$이므로 검정통계량은 다음과 같다.

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{0.079 \times 0.921\left(\dfrac{1}{500} + \dfrac{1}{500}\right)}} = \frac{\hat{p}_1 - \hat{p}_2}{0.017}$$

④ 미혼과 기혼의 표본비율은 각각 $\hat{p}_1 = \dfrac{45}{500} = 0.09$, $\hat{p}_2 = \dfrac{34}{500} = 0.068$이므로 검정통계량의 관찰값은 다음과 같다.

$$z_0 = \frac{0.09 - 0.068}{0.017} = 1.2896$$

⑤ 검정통계량의 관찰값 $z_0 = 1.2896$은 기각역 안에 놓이지 않으므로 귀무가설을 기각하지 않는다. 즉, 미혼과 기혼의 자동차 사고 비율은 차이가 없다는 주장이 타당하다.

9. 어린이와 노인의 신종 플루에 대한 감염률이 동일한지 알아보기 위하여 각각 360명과 325씩 선정하여 조사하였다. 이들 중에서 어린이의 101명과 노인의 69명이 신종 플루에 감염되었을 때, 감염률이 동일한지 유의수준 5%에서 검정하라.

풀이

① 어린이와 노인의 감염률을 각각 p_1, p_2라 하면, 귀무가설은 $H_0 : p_1 - p_2 = 0$이고 대립가설은 $H_1 : p_1 - p_2 \ne 0$이다.

② 유의수준 5%에서 양측검정이므로 기각역은 $Z \le -1.96$ 또는 $Z \ge 1.96$이다.

③ 합동표본비율은 $\hat{p} = \dfrac{101+69}{360+325} = 0.245$이므로 검정통계량은 다음과 같다.

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{0.245 \times 0.752\left(\dfrac{1}{360} + \dfrac{1}{325}\right)}} = \frac{\hat{p}_1 - \hat{p}_2}{0.033}$$

④ 어린이와 노인의 표본비율은 각각 $\hat{p}_1 = \dfrac{101}{360} = 0.281$, $\hat{p}_2 = \dfrac{69}{325} = 0.212$이므로 검정통계량의 관찰값은 다음과 같다.

$$z_0 = \frac{0.281 - 0.212}{0.033} = 2.09$$

⑤ 검정통계량의 관찰값 $z_0 = 2.09$는 기각역 안에 놓이므로 귀무가설을 기각한다. 즉, 어린이와 노인의 신종 플루에 대한 감염률이 동일하다는 주장은 타당하지 않다.

10. 전국적으로 발생하는 교통사고 건수가 요일별로 동일한가를 살펴보기 위하여, 지난해 1년 동안 발생한 교통사고 건수를 조사한 결과 다음 표와 같았다. 유의수준 5%에서 요일별 사고 건수가 동일한지 검정하라.

월	화	수	목	금	합계
86	56	51	55	72	320

풀이 ① 요일별 교통사고 비율이 동등한가에 대하여 검정하므로 다음과 같은 귀무가설과 대립가설을 설정한다.

$$H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = 0.2, \quad H_1 : H_0\text{이 아니다.}$$

② χ^2-통계량 $\chi^2 = \sum_{i=1}^{k} \dfrac{(o_i - e_i)^2}{e_i}$을 사용하며, 범주의 수가 5이므로 자유도 4인 χ^2-분포에 대하여 유의수준 $\alpha = 0.05$인 상단측검정에 대한 기각역은 $\chi^2 \geq 9.49$이다.

③ 검정통계량의 측정값은 다음 표와 같이 $\chi_0^2 = 13.47$이고, 이 관찰값은 기각역 안에 놓이므로 귀무가설 H_0을 기각한다. 즉, 교통사고 건수는 요일별로 동일하다고 할 수 없다.

범주	o_i	p	$e_i = np_i$	$n_i - e_i$	$(n_i - e_i)^2$	$\dfrac{(n_i - e_i)^2}{e_i}$
1	86	0.2	64	22	484	7.56
2	56	0.2	64	-8	64	1.00
3	51	0.2	64	13	169	2.64
4	55	0.2	64	-9	81	1.27
5	72	0.2	64	8	64	1.00
	$n = 320$					합 : 13.47

11장

1. 어느 회사에서 제조되는 건전지의 평균수명을 알아보기 위하여 16개를 임의로 조사하여 평균 70.1시간 표준편차 2.8시간을 얻었다. 이 회사에서 제조되는 건전지의 평균수명에 대한 90% 신뢰구간을 구하라. 단, 건전지 수명은 정규분포를 이룬다고 한다.

풀이

$\overline{x}=70.1$, $s=2.8$이고 $t_{0.05}(15)=1.753$이므로 오차한계는 $e=1.753\times\dfrac{2.8}{\sqrt{16}}=1.227$이다. 그러므로 평균수명에 대한 90% 신뢰구간은 $(70.1-1.227,\ 70.1+1.227)=(68.873,\ 71.327)$이다.

2. 건강한 성인이 평균적으로 하루에 소비하는 물의 양을 조사하여 다음을 얻었다. 성인이 평균적으로 하루에 소비하는 물의 양에 대한 95% 신뢰구간을 구하라. 단, 소비량은 정규분포에 따른다.

1.1	1.7	1.5	2.4	2.1	1.8	2.5	1.7	2.8	2.2	1.9

풀이

표본평균과 표본분산, 표본표준편차를 구하면 다음과 같다.

$$\overline{x}=\frac{1}{11}(1.1+1.7+\cdots+1.9)=1.973$$

$$s^2=\frac{1}{10}\sum_{i=1}^{11}(x_i-1.973)^2=0.238,\quad s=\sqrt{0.238}=0.488$$

그리고 $t_{0.025}(10)=2.228$이므로 오차한계는 $e=2.228\times\dfrac{0.488}{\sqrt{11}}=0.3278$이다. 그러므로 평균 소비량에 대한 95% 신뢰구간은 $(1.973-0.3278,\ 1.973+0.3278)=(1.6452,\ 2.3008)$이다.

3. 모분산이 동일하고 독립인 두 정규모집단으로부터 표본조사하여 다음을 얻었다. 모평균 차에 대한 90% 신뢰구간을 구하라.

표본 1	$n=4$, $\overline{x}=67.75$, $s_1^2=12.25$
표본 2	$m=4$, $\overline{y}=49.75$, $s_2^2=14.25$

풀이

두 표본에 대한 합동표본분산과 합동표본표준편차는 다음과 같다.

$$s_p^2=\frac{3\times12.25+3\times14.25}{4+4-2}=13.25,\quad s_p=\sqrt{13.25}=3.64$$

또한 자유도 6인 t-분포에서 $t_{0.05}(6) = 1.943$이므로 90% 신뢰구간에 대한 오차한계는 다음과 같다.

$$e = 1.943 \times 3.64 \times \sqrt{\frac{1}{4} + \frac{1}{4}} = 5.001$$

이때 $\overline{x} - \overline{y} = 18$이므로 $\mu_1 - \mu_2$에 대한 90% 신뢰구간은 다음과 같다.

$$(18 - 5.001,\ 18 + 5.001) = (12.999,\ 23.001)$$

4. 효능이 같은 서로 다른 두 약품 A와 B에 의해 치료하는데 걸리는 기간의 차이를 알기 위해, 각각 15명과 12명의 환자를 선정하여 조사하였다. 약품 A와 B에 의한 환자의 평균 치료 기간은 각각 6.4일과 4.2일, 표준편차는 1.2일과 1.5일이었다. 두 약품에 의한 평균 치료기간의 차에 대한 95% 신뢰구간을 구하라. 단, 약품 A와 B는 각각 정규분포에 따르고 모분산은 동일하다고 한다.

풀이

두 표본에 대한 합동표본분산과 합동표본표준편차는 다음과 같다.

$$s_p^2 = \frac{14 \times 1.2^2 + 11 \times 1.5^2}{15 + 12 - 2} = 1.7964,\ \ s_p = \sqrt{1.7964} = 1.34$$

또한 자유도 25인 t-분포에서 $t_{0.025}(25) = 2.06$이므로 90% 신뢰구간에 대한 오차한계는 다음과 같다.

$$e = 2.06 \times 1.34 \times \sqrt{\frac{1}{15} + \frac{1}{12}} = 1.07$$

이때 $\overline{x} - \overline{y} = 2.2$이므로 $\mu_1 - \mu_2$에 대한 90% 신뢰구간은 다음과 같다.

$$(2.2 - 1.07,\ 2.2 + 1.07) = (1.13,\ 3.27)$$

5. 어느 회사에서 제조되는 저항기는 정규분포에 따른다고 한다. 이 회사에서 생산되는 저항의 표준편차를 추정하기 위하여 30개를 임의로 선정한 결과 평균 40.1Ω, 표준편차 1.2Ω을 얻었다. 이 회사에서 생산되는 저항기의 표준편차에 대한 90% 신뢰구간을 구하라.

풀이

$s = 1.2$이고 자유도 29인 χ^2-분포에서 $\chi^2_{0.05}(29) = 42.56$, $\chi^2_{0.95}(29) = 17.71$이므로 모분산 σ^2에 대한 90% 신뢰구간의 하한과 상한은 각각 다음과 같다.

$$l = \frac{29 \times 1.2^2}{42.56} = 0.9812, \quad u = \frac{29 \times 1.2^2}{17.71} = 2.358$$

따라서 $\sqrt{0.9812} = 0.9906$, $\sqrt{2.358} = 1.536$이므로 모표준편차 σ에 대한 90% 신뢰구간은 $(0.9906,\ 1.5360)$이다.

6. 두 전자회사에서 생산되는 동일 급의 가전제품의 수명을 조사하여 다음을 얻었다. 이 자료를 이용하여 분산 비에 대한 90% 신뢰구간을 구하라. 단, 가전제품의 수명은 정규분포에 따르고 단위는 년이다.

	표본크기	표본표준편차
표본 1	11	0.52
표본 2	9	0.43

풀이

표본의 크기가 각각 11과 9이므로 분자와 분모의 자유도는 10과 8이고 $f_{0.05}(10, 8) = 3.35$, $f_{0.05}(8, 10) = 3.07$이다. 그러므로 90% 신뢰구간의 하한과 상한은 각각 다음과 같다.

$$l = \frac{0.52^2}{0.43^2} \times \frac{1}{3.35} = 0.436,\ u = \frac{0.52^2}{0.43^2} \times 3.07 = 4.488$$

따라서 σ_1^2/σ_2^2에 대한 90% 신뢰구간은 $(0.436,\ 4.488)$이다.

7. 정규모집단에 대한 $H_0 : \mu = 2.5$를 알아보기 위하여 크기 14인 표본을 조사하여 $\overline{x} = 2.26$, $s = 0.4$를 얻었다. 유의수준 5%에서 H_0을 검정하라.

풀이

다음 순서에 따라 검정한다.

① $t_{0.025}(13) = 2.16$이므로 $\alpha = 0.05$에 대한 양측검정의 기각역은 $T \le -2.16$ 또는 $T \ge 2.16$이다.

② $s = 0.4$이므로 검정통계량은 $T = \dfrac{\overline{X} - 2.5}{0.4/\sqrt{14}}$이다.

③ $\overline{x} = 2.26$이므로 검정통계량의 관찰값은 $t_0 = \dfrac{2.26 - 2.5}{0.4/\sqrt{14}} = -2.24$이다.

④ 관찰값이 기각역 안에 놓이므로 귀무가설을 기각한다.

8. 모분산이 동일한 두 정규모집단에서 임의로 표본을 선정하여 다음 결과를 얻었다. 이때 두 모평균이 동일한지 유의수준 1%에서 검정하라.

표본 A	$n = 10,\ \overline{x} = 25.5,\ s_1 = 2.1$
표본 B	$m = 8,\ \overline{y} = 22.7,\ s_2 = 3.2$

풀이

다음 순서에 따라 검정한다.

① 귀무가설 $H_0 : \mu_1 - \mu_2 = 0$과 대립가설 $H_1 : \mu_1 - \mu_2 \ne 0$을 설정한다.

② $n = 10$, $m = 8$, $\alpha = 0.01$이므로 $t_{0.005}(16) = 2.921$이고 양측검정의 기각역은 $T \le -2.921$ 또는 $T \ge 2.921$이다.

③ $s_1 = 2.1$, $s_2 = 3.2$이므로 합동표본분산은 다음과 같다.

$$s_p^2 = \frac{9\times 2.1^2 + 7\times 3.2^2}{10+8-2} = 6.9606,\ s_p = \sqrt{6.9606} = 2.6383$$

따라서 검정통계량은 다음과 같다.

$$T = \frac{(\overline{X} - \overline{Y}) - 0}{2.6383 \times \sqrt{\frac{1}{10} + \frac{1}{8}}} = \frac{\overline{X} - \overline{Y}}{1.251}$$

④ $\overline{x} - \overline{y} = 2.8$이므로 검정통계량의 관찰값은 $t_0 = \dfrac{2.8}{1.251} = 2.238$이다.

⑤ 관찰값 $t_0 = 2.238$은 기각역 안에 놓이지 않으므로 두 모평균은 동일하다는 주장을 기각할 수 없다.

9. 40대에는 남자가 여자보다 혈압이 높다고 한다. 이것을 확인하기 위해 40대 남자와 여자를 임의로 조사하여 다음을 얻었다. 남자의 혈압이 여자보다 높은지 유의수준 5%에서 검정하라. 단, 혈압은 정규분포에 따른다고 한다.

남자	$n = 6$, $\overline{x} = 141$, $s_1 = 3.6$
여자	$m = 9$, $\overline{y} = 138$, $s_2 = 2.7$

풀이

남자와 여자의 평균 혈압을 μ_1과 μ_2라 하고, 다음 순서에 따라 검정한다.

① 귀무가설 $H_0 : \mu_1 - \mu_2 \le 0$과 대립가설 $H_1 : \mu_1 - \mu_2 > 0$(주장)을 설정한다.

② $n = 6$, $m = 9$이므로 $t_{0.025}(13) = 2.16$이고, $\alpha = 0.05$에 대한 상단측검정이므로 기각역은 $T \ge 2.16$이다.

③ 합동표본분산과 합동표준편차는 각각 다음과 같다.

$$s_p^2 = \frac{5\times 3.6^2 + 8\times 2.7^2}{6+9-2} = 9.4708,\ s_p = \sqrt{9.4708} = 3.0775$$

따라서 검정통계량은 다음과 같다.

$$T = \frac{(\overline{X} - \overline{Y}) - 0}{(3.0775)\sqrt{\frac{1}{6} + \frac{1}{9}}} = \frac{\overline{X} - \overline{Y}}{1.62}$$

④ $\overline{x} - \overline{y} = 3$이므로 검정통계량의 관찰값은 $t_0 = \dfrac{3}{1.62} = 1.851$이다.

⑤ 관찰값이 기각역 안에 놓이지 않으므로 귀무가설을 기각할 수 없다. 즉, 남자가 여자보다 혈압이 높다는 주장은 타당성이 떨어진다.

10. 독립이고 모분산이 동일한 두 정규모집단에서 각각 표본을 선정하여 다음 결과를 얻었다. 이 자료를 근거로 $\mu_1 < \mu_2$인 주장을 유의수준 5%에서 검정하라.

표본 1	15, 12, 14, 11, 12
표본 2	12, 15, 16, 15, 16, 15

풀이

표본 1과 2의 모평균을 각각 μ_1, μ_2라 하고 다음 순서에 따라 검정한다.

① 귀무가설 $H_0 : \mu_1 - \mu_2 \geq 0$와 대립가설 $H_1 : \mu_1 - \mu_2 < 0$(주장)을 설정한다.

② 두 표본의 크기가 각각 5와 6이므로 $t_{0.05}(9) = 1.833$이고 $\alpha = 0.01$에 대한 하단측검정이므로 기각역은 $T \leq -1.833$이다.

③ 두 표본의 평균과 분산을 구하면 각각 다음과 같다.

$$\overline{x} = \frac{1}{5}\Sigma x_i = 12.8, \qquad s_1^2 = \frac{1}{4}\Sigma(x_i - 12.8)^2 = 2.7$$

$$\overline{y} = \frac{1}{6}\Sigma y_i = 14.833, \qquad s_2^2 = \frac{1}{5}\Sigma(y_i - 14.833)^2 = 2.167$$

그러므로 합동표본분산과 합동표준편차는 각각 다음과 같다.

$$s_p^2 = \frac{4 \times 2.7 + 5 \times 2.167}{5 + 6 - 2} = 2.4, \quad s_p = \sqrt{2.404} = 1.55$$

따라서 검정통계량은 다음과 같다.

$$T = \frac{(\overline{X} - \overline{Y}) - 0}{1.55 \times \sqrt{\frac{1}{5} + \frac{1}{6}}} = \frac{\overline{X} - \overline{Y}}{0.9386}$$

④ $\overline{x} - \overline{y} = -2.033$이므로 검정통계량의 관찰값은 $t_0 = -\dfrac{2.033}{0.9386} = -2.166$이다.

⑤ 관찰값이 기각역 안에 놓이므로 귀무가설을 기각한다. 즉, $\mu_1 < \mu_2$인 주장은 근거는 충분하다.

11. 자동차 사고가 빈번히 일어나는 교차로의 신호체계를 바꾸면 사고를 줄일 수 있다고 경찰청에서 말한다. 이것을 알아보기 위하여 시범적으로 사고가 많이 발생하는 지역을 선정하여 지난 한달 동안 발생한 사고건수와 신호체계를 바꾼 후의 사고건수를 조사한 결과 다음과 같았다. 유의수준 5%에서 신호체계를 바꾸면 사고를 줄일 수 있는지 검정하라. 단, 사고 건수는 정규분포를 이룬다고 알려져 있다.

지역	1	2	3	4	5	6	7	8
교체 전	5	10	8	9	5	7	6	8
교체 후	4	9	8	8	4	8	5	8

풀이

우선 각 지역의 신호체계를 바꾸기 전후의 사고 건수에 대한 차를 구한다.

지역	1	2	3	4	5	6	7	8
d_i	1	1	0	1	1	-1	1	0

① 신호체계를 바꾸기 전과 후의 평균 사고 건수를 각각 μ_1, μ_2라 하면, 밝히고자 하는 것은 $\mu_1 > \mu_2$이고 등호가 들어가지 않으므로 대립가설로 설정한다. 따라서 귀무가설은 $H_0 : \mu_1 \le \mu_2$이고 대립가설은 $H_1 : \mu_1 > \mu_2$(주장)이다.
② 유의수준 $\alpha = 0.05$에 대한 상단측검정이고 이때 자유도 7인 t-분포를 사용하므로 기각역은 $R : T > t_{0.05} = 1.895$이다.
③ d_i에 대한 평균과 표준편차를 구한다.

$$\overline{d} = \frac{1}{8}(1+1+0+1+1-1+1+0) = 0.5$$

그리고 $\sum d_i^2 = 6$, $\left(\sum d_i\right)^2 = 16$이므로 타수의 차에 대한 표준편차는 다음과 같다.

$$s_d = \sqrt{\frac{8\times 6 - 16}{56}} = 0.7559$$

④ 검정통계량 $T = \dfrac{\overline{d} - 0}{0.7559/\sqrt{8}} = \dfrac{\overline{d}}{0.26725}$의 관찰값은 $t_0 = \dfrac{0.5}{0.26725} = 1.871$이므로 기각역 안에 놓이지 않는다.
⑤ 따라서 귀무가설을 기각할 수 없다. 즉, 신호체계를 바꾸면 사고를 줄일 수 있다는 근거는 없다.

12. 정규모집단의 모분산을 추정하기 위하여 크기 25인 표본을 조사하여 표본분산 3.2를 얻었다. 모분산과 모표준편차에 대한 99% 신뢰구간을 구하라.

풀이

표본의 크기가 25이므로 자유도 24인 χ^2-분포에서 $\chi^2_{0.005}(24) = 45.56$, $\chi^2_{0.995}(24) = 9.89$이므로 모분산 σ^2에 대한 99% 신뢰구간의 하한과 상한은 각각 다음과 같다.

$$l = \frac{24\times 3.2}{45.56} = 1.6857, \qquad u = \frac{24\times 3.2}{9.89} = 7.7654$$

따라서 모분산에 대한 99% 신뢰구간은 $(1.6857,\ 7.7654)$이고 모표준편차에 대한 99% 신뢰구간은 $(1.298,\ 2.787)$이다.

13. 직경 20mm인 볼트를 생산하는 회사에서 볼트 직경의 표준편차는 0.7mm보다 작다고 한다. 이것을 알아보기 위하여 15개의 볼트를 임의로 조사하여 다음을 얻었다. 유의수준 5%에서 조사하라.

19.4	21.1	20.2	18.6	19.4	20.8	20.2	19.2	21.1	20.6	20.0	20.3	19.5	19.7	20.3

풀이

표본평균과 표본분산을 먼저 구하면 다음과 같다.

$$\overline{x} = \frac{1}{15}\sum x_i = 20.027,\ s^2 = \frac{1}{14}\sum (x_i - 20.027)^2 = 0.524$$

① $\sigma < 0.7$을 검정하므로 귀무가설 $H_0 : \sigma^2 \geq 0.49$와 대립가설 $H_1 : \sigma^2 < 0.49$(주장)를 설정한다.

② 크기 15이므로 $\chi^2_{0.95}(14) = 6.57$이고 $\alpha = 0.05$에 대한 하단측검정의 기각역은 $V < 6.57$이다.

③ 검정통계량은 $V = \dfrac{14S^2}{0.49}$이고, $s^2 = 0.524$이므로 관찰값은 $v_0 = \dfrac{14 \times 0.524}{0.49} = 14.9714$이다.

④ 관찰값이 기각역 안에 놓이지 않으므로 귀무가설을 기각하지 않는다. 즉, 표준편차는 0.7mm보다 작다는 근거는 미약하다.

14. 두 가지 실험방법에 의한 반응 시간의 분산을 비교하기 위하여 크기가 9와 7인 표본을 조사하여 각각 $s_1 = 3.4$, $s_2 = 6.1$을 얻었다. 두 실험방법에 대한 반응 시간의 분산의 비에 대한 95% 신뢰구간을 구하라.

풀이

크기 9와 7인 표본에 대하여 $f_{0.025}(8, 6) = 5.6$, $f_{0.025}(6, 8) = 4.65$이고, $\dfrac{s_1^2}{s_2^2} = \dfrac{3.4^2}{6.1^2} = 0.31$이므로 95% 신뢰구간의 하한은 $l = \dfrac{0.31}{5.6} = 0.055$, 상한은 $u = 0.31 \times 4.65 = 1.44$이다. 따라서 구하고자 하는 신뢰구간은 $(0.055,\ 1.44)$이다.

12장

1. 보험회사에 청구되는 보험금 신청 횟수는 푸아송과정에 따르며, 연속적인 청구 사이의 평균 시간은 이틀이라고 한다.
(a) 3일 동안 적어도 한 건의 보험금 청구가 신청될 확률을 구하라.
(b) 두 번째 보험금 신청이 4일째에 나타날 확률을 구하라.

풀이

(a) 보험금을 신청하는데 소요되는 시간이 평균 2일 이므로 하루 당 $\lambda = 1/2$이고, 따라서 보험금 신청 횟수 $N(t)$는 평균 비율 $\lambda = 1/2$인 푸아송과정에 따른다. 즉, $P[N(t) = x] = \frac{(t/2)^x}{x!} e^{-t/2}$, $t \geq 0$이다. 그러므로 다음을 얻는다.

$$P[N(3) = 0] = \frac{(3/2)^0}{0!} e^{-3/2} = e^{-1.5} = 0.2231$$

따라서 구하고자 하는 확률은 다음과 같다.

$$P[N(3) \geq 1] = 1 - P[X(3) = 0] = 1 - 0.2231 = 0.7769$$

(b) 두 번째 보험금이 신청될 때까지 경과시간 X는 모수 $\alpha = 2$와 $\beta = 2$인 감마분포에 따르므로 확률밀도함수는 $f(x) = \frac{x}{4} e^{-x/2}$, $x \geq 0$이다. 따라서 구하고자 하는 확률은 다음과 같다.

$$P(3 < X < 4) = \int_3^4 \frac{x}{4} e^{-x/2}\, dx = -\frac{1}{2}(x+2)e^{-x/2} \Big|_3^4$$

$$= \frac{5}{2} e^{-3/2} - 3e^{-2} = 0.5578 - 0.4060 = 0.1518$$

2. 보험계리인은 지난해의 자료를 분석한 결과, 보험회사에 가입한 보험가입자들로부터 청구되는 보험금 신청횟수는 푸아송과정에 따르며, 한 보험가입자로부터 보험금 신청이 있은 후 다음 번 신청까지 평균적으로 3일이 소요된다는 결론을 얻었다.
(a) 이틀 동안 보험금 청구가 신청되지 않을 확률을 구하라.
(b) 적어도 두 건의 보험금 청구가 이틀 안에 이루어질 확률을 구하라.
(c) 세 번째 보험금 신청이 4일째에 나타날 확률을 구하라.

풀이

(a) 보험금 신청 사이의 소요시간이 평균 3일 이므로 하루 당 신청비율은 $\lambda = 1/3$이고, 따라서 보험금 신청횟수 $N(t)$는 평균 비율 $\lambda = 1/3$인 푸아송과정에 따른다. 즉, $P[N(t) = n] = \frac{(t/3)^n}{n!} e^{-t/3}$, $t > 0$이다. 그러므로 구하고자 하는 확률은 다음과 같다.

$$P[N(2) = 0] = e^{-2/3} = 0.5134$$

(b) 이틀 동안에 한 건의 보험금 신청이 있을 확률은 $P[N(2)=1]=\frac{2}{3}e^{-2/3}=0.3423$이므로 구하고자 하는 확률은

$$P[N(2)\geq 2]=1-(P[X(2)=0]+P[X(2)=1])=1-(0.5134+0.3423)=0.1443$$

이다.

(c) 세 번째 보험금이 신청될 때까지 경과시간 X는 모수 $\alpha=3$과 $\beta=3$인 감마분포를 이루므로 확률밀도함수는 $f(x)=\frac{1}{54}x^2 e^{-x/3}$, $x\geq 0$이다. 따라서 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned} P(3<X<4) &= \int_3^4 \frac{1}{54}x^2 e^{-x/3}\,dx = \left[-\frac{1}{54}(3x^2+18x+54)e^{-x/3}\right]_3^4 \\ &= 0.9197-0.8494=0.0703 \end{aligned}$$

3. $\lambda=5$의 비율로 보험가입자들이 보험금 신청을 하는 보험회사에 청구된 보험금 지급요구는 10,000\$, 20,000\$, 30,000\$, 40,000\$ 등 네 종류가 있다. 한편 청구된 각 보험급부금의 비율이 $p_1=1/2$, $p_2=1/4$, $p_3=3/16$, $p_4=1/16$이라 한다.
(a) 단위 시간 동안에 40,000\$의 보험금 지급 요청이 없을 확률을 구하라.
(b) 어떤 시각에서 시작하여, 40,000\$의 보험금 지급 요청이 있기 전에 10,000\$의 요청이 적어도 두 번 있을 확률을 구하라.

풀이

(a) 시각 t까지 40,000\$의 보험금 지급 요청 횟수를 $N_4(t)$라 하면, $\{N_4(t)|t\geq 0\}$는 비율 $\lambda p_4=5(1/16)=5/16$인 푸아송과정이다. 따라서 단위 시간 동안 이 종류의 보험금 지급 요청이 없을 확률은 $P[N_4(1)=0]=e^{-5/16}=0.7316$이다.

(b) 시각 t까지 10,000\$의 보험금 지급 요청 횟수를 $N_1(t)$라 하면, $\{N_1(t)|t\geq 0\}$는 비율 $\lambda p_1=5(1/2)=5/2$인 푸아송과정이다. 한편 시각 t에서 40,000\$의 보험금 지급 요청이 있다는 조건을 부여하면, 조건부 확률은 다음과 같다.

$$P[N_1(t)\geq 2\,|\,T_4=t]=1-e^{-(5/2)t}-(5/2)t\,e^{-(5/2)t}$$

T_4의 밀도함수는 $f_{T_4}(t)=(5/16)e^{-(5/16)t}$이다. 그러므로 전확률 공식에 의하여 다음 확률을 얻는다.

$$\begin{aligned} P[N_1(t)\geq 2] &= \int_0^\infty P[N_1(t)\geq 2\,|\,T_4=t]\,f_{T_4}(t)\,dt \\ &= \frac{5}{16}\int_0^\infty \left[1-e^{-(5/2)t}-(5/2)t\,e^{-(5/2)t}\right]e^{-(5/16)t}\,dt=\frac{64}{81}=0.7901 \end{aligned}$$

4. 자동차 보험회사는 금년에 보험증권을 구입한 운전자를 정상적인 그룹과 위험성한 그룹으로 구분하고 있으며, 보험 가입자 중에서 70%가 정상적인 그룹에 속한다고 한다. 한편 이 회사에 신청되는 보험금 청구 횟수는 주당 평균 6건이고, 푸아송과정에 따른다고 한다.
(a) 1주일 동안 정상적인 운전자에 의한 보험금 청구 횟수에 대한 확률분포를 구하라.
(b) 이 기간 동안에 정상적인 운전자에 의한 청구 횟수가 한 건 이하일 확률을 구하라.

풀이

(a) 정상적인 그룹에 속하는 운전자에 의한 청구 횟수를 $N_1(1)$라 하면, 1주일 동안 청구된 횟수는 비율 $\lambda pt = (6)(0.7)(1) = 4.2$인 푸아송과정에 따른다. 따라서 $N_1(1)$은 평균 4.2인 푸아송 분포를 갖는다.

(b) 구하고자 하는 확률은 다음과 같다.

$$P[N_1(1) \le 1] = e^{-4.2} + (4.2)e^{-4.2} = 5.2e^{-4.2} = 0.078$$

5. 모수 $p = \frac{1}{4}$인 베르누이과정 $\{X_n : n = 0, 1, 2, \cdots\}$에 대하여 $N_n = X_1 + \cdots + X_n$이라 할 때, 다음이 성립한다.

$$P(N_{m+n} - N_m = k | N_0, N_1, \cdots, N_n)$$
$$= P(N_{m+n} - N_m = k) = \binom{n}{k}\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{n-k}, \quad k = 0, 1, \cdots, n$$

(a) 확률 $P(N_4 = 3, N_8 - N_4 = 2, N_{12} - N_8 = 1)$을 구하라.
(b) 기댓값 $E(N_4 N_8)$을 구하라.

풀이

(a) $N_{12} - N_8$은 $N_0, \cdots, N_8$과 독립이고 $N_8 - N_4$은 $N_0, \cdots, N_4$와 독립이므로 다음을 얻는다.

$$P(N_4 = 3, N_8 - N_4 = 2, N_{12} - N_8 = 2) = P(N_4 = 3)P(N_8 - N_4 = 2)P(N_{12} - N_8 = 1)$$
$$= \binom{4}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right) \times \binom{4}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^2 \times \binom{4}{2}\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^3 = \frac{2187}{524288}$$

(b) $N_8 = N_4 + (N_8 - N_4)$이므로 $N_4 N_8 = N_4[N_4 + (N_8 - N_4)] = N_4^2 + N_4(N_8 - N_4)$이다. 그리고 N_4와 $N_8 - N_4$이 독립이므로 다음을 얻는다.

$$E(N_4 N_8) = E[N_4^2 + N_4(N_8 - N_4)] = E(N_4^2) + E(N_4)E(N_8 - N_4)$$

한편 $N_4 \sim B\left(4, \frac{1}{4}\right)$, $N_8 - N_4 \sim B\left(4, \frac{1}{4}\right)$이므로 $E(N_4) = 1$, $Var(N_4) = \frac{3}{4}$이고 $E(N_5^2) = \frac{3}{4} + 1 = \frac{7}{4}$이다. 그리고 $E(N_8 - N_5) = 1$이므로 다음을 얻는다.

$$E(N_5 N_8) = \frac{7}{4} + 1 \times 1 = \frac{11}{4}$$

6. 어느 교차로에서 연평균 5건인 푸아송분포에 따라 교통사고가 발생한다.
(a) 어느 한 해 동안 정확히 교통사고가 3번 발생할 확률을 구하라.
(b) 전반부 6개월 동안 3건의 사고가 발생할 확률을 구하라.

풀이

(a) 1년간 발생한 교통사고 건수를 $N(1)$이라 하면, $N(1) \sim P(5)$이다. 그러므로 구하고자 하는 확률은 <부록 A.2>에 의해 다음과 같다.

$$P(N(1)=3)=P(N(1) \le 3)-P(N(1) \le 2)=0.265-0.125=0.140$$

(b) 전반부 6개월 동안 발생한 교통사고 건수는 $N(0.5) \sim P(2.5)$이다. 그러므로 구하고자 하는 확률은 <부록 A.2>에 의해 다음과 같다.

$$P(N(0.5)=3)=P(N(10.5 \le 3)-P(N(0.5) \le 2)=0.758-0.544=0.214$$