<3-3>

```
LIBNAME MYLIB '/folders/myfolders/mylib';
DATA MYLIB.EX3_3;
        A_1 = 1 - CDF('NORMAL', 180, 170, 5);
        A_2 = CDF('NORMAL', 175, 170, 0.5);
RUN;
PROC PRINT DATA=MYLIB.EX3_3;
RUN;
```

(가) 0.02275

(나) 1

(다) $\sqrt{\frac{25}{25}} = 1,\ \sqrt{\frac{25}{10}} = \frac{5}{\sqrt{10}} = 1.5811$

<3-5>

```
DATA MYLIB.NORMAL;
        DO x=-4.5 TO 4.5 BY 0.01;
                p_x = PDF('NORMAL',x,0,1);
                p_y = PDF('NORMAL',x,0,2);
                OUTPUT;
        END;
RUN;
PROC PLOT data=MYLIB.NORMAL;
  PLOT p_x*x='*'
    p_y*x='o' / OVERLAY BOX;
  TITLE1 'Plot of normal(0,1)';
  TITLE2 'Plot of normal(0,4)';
RUN;
```

<3-7>

```
DATA MYLIB.CHILDREN;
        A_1 = 1-PMF('BINOMIAL',0,0.43,3);
        A_2 = 1-PMF('BINOMIAL',0,0.57,3);
        A_3 = A_2 - PMF('BINOMIAL',3,0.57,3);
RUN;
PROC PRINT DATA=MYLIB.CHILDREN;
RUN;
```

(가) 0.81481

(나) 0.92049

(다) 0.7353

<3-8>

$P(fair\ coin \mid head) = \frac{1}{3}$

<3-12>

$P(B|A) = 0.5$
$P(A \cap B) = 0.2$
$P(A \cup B) = 0.7$
$P((A \cup B)^c) = 0.3$

<3-14>

$P(Defective) = 0.4 * 0.07 + 0.6 * 0.11 = 0.094$
$P(A \cap D) = 0.4 * 0.07 = 0.028$
$P(A|D) = \frac{0.028}{0.094} = 0.297823$

<3-16>

(가)

	$Y=0$	$Y=1$	$Y=2$	
$X=0$	1/9	2/9	1/9	4/9
$X=1$	1/9	2/9	1/9	4/9
$X=2$	1/36	1/18	1/36	1/9
	1/4	1/2	1/4	1

(나)

$X=0$	$X=1$	$X=2$	
1/4	1/2	1/4	1

$Y=0$	$Y=1$	$Y=2$	
4/9	4/9	1/9	1

<3-18>

$P(Defective) = 0.2 * 0.01 + 0.3 * 0.06 + 0.5 * 0.03 = 0.23$
$P(A \cap D) = 0.2 * 0.01 = 0.02$

(가) $P(A|D) = \frac{0.02}{0.23} = 0.086957$

(나) $P(A|D^c) = \frac{0.2*0.99}{0.77} = 0.257143$

<3-22>

(가) $\int_0^1 ax^2dx = \frac{a}{3}[x^3]_0^1 = \frac{a}{3} = 1 \quad \therefore a = 3$

(나) $P\left(X \le \frac{1}{2}\right) = \int_0^{\frac{1}{2}} 3x^2dx = \frac{1}{8}$

$P\left(X \le \frac{1}{3}\right) = \int_0^{\frac{1}{2}} 3x^2dx = \frac{1}{27}$

$P\left(\frac{1}{4} \le X \le \frac{3}{4}\right) = \int_0^{\frac{1}{2}} 3x^2dx = \frac{13}{32}$

<3-26>

(가) 0,1,2,3,…

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

<3-28>

(가)

X	n	$P(X)$
2	1	1/169
3	2	2/169
4	3	3/169
…	…	…
13	12	12/169
14	13	13/169
15	12	12/169
…	…	…
25	2	2/169
26	1	1/169

(나) $P(X < 16) * 1000 - P(X \ge 17) * 1000$

$= \frac{1 + \cdots + 15}{169} * 1000 - \frac{1 + \cdots + 10}{169} * 1000 = 739.645$

<3-31>

```
DATA MYLIB.SULFURIC;
     SULFURIC = 1-CDF('NORMAL',700,670,22/2.236);
RUN;
PROC PRINT DATA=MYLIB.SULFURIC;
RUN;
It is .001147675
```

<3-33>

$X \sim Bin(600, \frac{3}{8})$

```
DATA MYLIB.COIN;
        A_1 = CDF('BINOMINAL',220,3/8,600);
        A_2 = 1 - CDF('BINOMIAL',250,3/8,600);
RUN;
```

```
PROC PRINT DATA=MYLIB.COIN;
RUN;
```

(가) $E[X] = 225, Var(X) = 140.625$

(나) $P(X \leq 220) = 0.35330$

(다) $P(X \geq 250) = 0.016221$

<3-34>

```
DATA MYLIB.EX3_34;
        A_1 = CDF('CHISQUARED',15.507,8);
        A_2 = 1 - CDF('CHISQUARED',2.180,8);
        A_3 = CDF('CHISQUARED',13.362,8) -
              CDF('CHISQUARED',1.647,8);
RUN;
PROC PRINT DATA=MYLIB.EX3_34;
RUN;
```

(가) 0.94999

(나) 0.97499

(다) 0.89000

<3-35>

$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$, $\frac{s^2}{30} \sim \chi^2(10)$

$P(s^2 \geq 480) = P\left(\frac{s^2}{30} \geq 16\right) = P(\chi^2 \geq 16)$

```
DATA MYLIB.BULB;
        A_1 = 1 - CDF('CHISQUARED',16,10);
RUN;
PROC PRINT DATA=MYLIB.BULB;
RUN;
It is 0.0996324
```

<3-36>

```
DATA MYLIB.TDIST;
A_1 = CDF('T',1.397,8);
A_2 = CDF('T',1.18,8) - CDF('T',-1.397,8);
A_3 = TINV(1-0.025,8);
RUN;
PROC PRINT DATA=MYLIB.TDIST;
RUN;
```

(가) 0.90003

(나) 0.76408

(다) 2.30600

<4-2>

(가) 다리길이의 차이가 정규성을 따르는지 확인하고자 하는 것으로 모평균이 0인지 아닌지 확인한다.

(나) $H_0: \mu = 0,\ H_1: \mu \neq 0$

(다) 제1종 오류: 다리길이 차이의 모평균이 0인데 그렇지 않다는 가설을 채택한 경우

제2종 오류: 다리길이 차이가 0이 아닌데 0이라는 가설을 채택한 경우

(라) 유의수준을 5%로 한다는 것은 실제 다리길이의 모평균이 0임에도 불구하고 0이 아니라는 대립가설을 채택하는 제1종 오류의 최대 허용오차를 5% 이하로 설정한다는 의미이다. 기존의 연구결과가 사실임에도 연구자의 대립가설을 채택하여 그 주장을 받아들이는 것이 반대의 경우보다 더 심각한 것으로 여기기 때문이다.

(마) 다리길이 차이의 평균은 3.3, 중위수는 4.5, 최빈값은 5이며, 표준편차는 3.06, 분산은 9.34이다. 1사분위수는 2, 3사분위수는 5이다.

(바) 아래에 극단치가 있으며 평균이 중간값 보다 낮다.

(사) 양측검정의 $p\text{-}value$가 0.077으로 유의수준 0.05보다 낮기 때문에 기각역에 존재한다. 따라서 H_0을 기각하고 H_1을 채택한다. 즉 '모평균은 0이 아니다.' 라고 말할 수 있다.

<4-5>

(가) $\left(171 - Z_{0.025}\frac{6}{\sqrt{100}}, 171 + Z_{0.025}\frac{6}{\sqrt{100}}\right) = (169.9, 172.2)$

(나) $(170.4, 171.6)$

(다) 같은 신뢰수준에서 정밀도를 높이기 위해서 즉 신뢰구간을 줄이기 위해서는 표본의 수가 커져야 한다. 표본의 수가 100에서 400으로 늘렸기 때문에 동일한 신뢰수준에서 더 좁은 신뢰구간을 갖게 되었다.

<4-9>

$$\left(0.67 - Z_{0.005}\sqrt{0.67 * \frac{0.33}{100}}, 0.67 + Z_{0.005}\sqrt{0.67 * \frac{0.33}{100}}\right)$$
$$= (0.5489, 0.7911)$$

<4-12>

```
DATA MYLIB.EX4_12;
        A_1 = (71-1)*(4**2)/30;
        A_2 = CINV(0.05,70);
RUN;
PROC PRINT DATA=MYLIB.EX4_12;
RUN;
A_1 = 37.333
A_2 = 51.739
```

5% 유의수준에서 검정통계량이 51.739보다 작으므로 기각역에 존재하여 귀무가설이 기각된다. 즉 모분산은 30보다 작다.

<4-13>

$$\left(\frac{2^2 Z_{0.025}}{0.1}\right)^2 = 6146.33 < n$$
$$\therefore n = 6147$$

<4-15>

$$\frac{113.5 - 110}{10/4} = 1.4 \ngtr Z_{0.05} = 1.6449$$

따라서 검정통계량이 기각역에 존재하지 않으므로 귀무가설을 기각할 수 없다. 즉 대학생 평균 IQ는 110이다.

<4-19>

$$\frac{18-20}{3/\sqrt{8}} = -1.8856 \nless -t_{0.05}(7) = -1.8946$$

검정통계량이 기각역에 존재하지 않으므로 귀무가설을 기각할 수 없다. 즉 유의수준 5% 하에서 1l당 평균 20㎞를 주행할 수 있다는 회사의 주장이 옳다.

<4-20>

$$\left|\frac{0.8-0.9}{\sqrt{\frac{0.9*0.1}{200}}}\right| = 4.714 > Z_{0.025} = 1.96$$

검정통계량이 기각역에 존재하므로 귀무가설을 기각한다. 즉 5% 유의수준 하에서 평균 출근율과 같다고 할 수 없다.

<4-22>

$$\left|(\frac{52}{240}-\frac{1}{6})/\sqrt{\frac{\frac{1}{6}\left(1-\frac{1}{6}\right)}{240}}\right| = 2.0784 > Z_{0.05} = 1.6448$$

검정통계량이 기각역에 존재하므로 귀무가설을 기각한다. 즉 10% 유의수준 하에서 6이 나올 확률이 1/6이 아니므로 바른 주사위가 아니다.

<4-24>

$$\frac{\frac{54}{120}-0.4}{\sqrt{\frac{\frac{54}{120}*(1-\frac{54}{120})}{120}}} = 1.1 \ngeq Z_{0.05} = 1.6448$$

기각치에 존재하지 않으므로 귀무가설을 채택한다. 즉 5% 유의수준 하에서 강남고등학교 3학년생의 영어 실력은 서울시 전체에 비교하여 열등하지 않다.

<4-25>

$$\frac{0.55-0.5}{\sqrt{\frac{0.5*0.5}{200}}} = 1.4142 \ngeq Z_{0.05} = 1.6448$$

검정통계량이 기각치에 존재하지 않으므로 귀무가설을 채택한다. 즉 5% 유의수준 하에서 요즘 남아의 출생률이 여아보다 높다는 산부인과 의사의 주장은 틀렸다. 따라서 남아의 출생률이 여아보다 높다고 볼 수 없다.