전산통계 과제#2

컴퓨터소프트웨어 학부 2018008559 신상윤

코드

결과

q1	q2
0.022750	1

(가) X ~ N(170, 5^2)에서 P(X >= 180) = 1 - P(X < 180)을 구했다. **0.02275** (나) 표본평균의 분포는 \overline{X} ~ N(170, $5^2/100$)이므로 P(\overline{X} <= 175)를 구했다. **1**

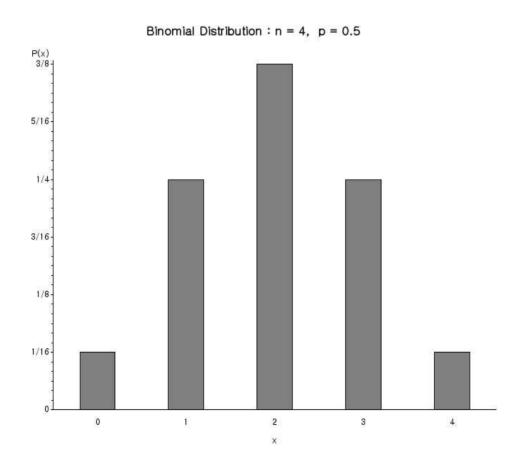
(다) 표준오차는 $\frac{\sigma}{\sqrt{n}}$ 이므로 크기가 25일때는 5/5 = 1, 크기가 10일때는 $\frac{5}{\sqrt{10}}$ 이다.

3-4

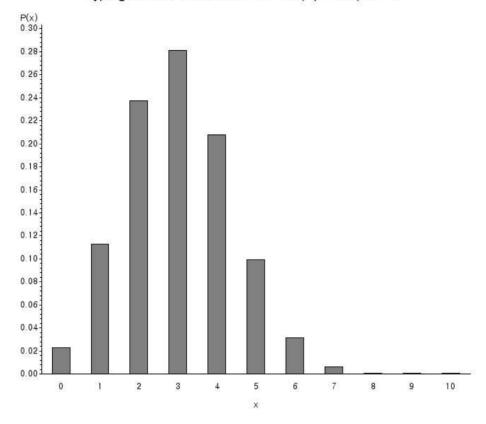
```
data binomial;
    do x=0 to 4;
    p_x = pdf('binomial',x,0.5,4);
    output;
    end;
run;
```

```
goptions device = win ctext = black graphrc;
title1 color = black height = 1.5 cell
"Binomial Distribution : n = 4, p = 0.5";
pattern1 color=gray value=S;
pattern2 color=gray value=S;
pattern3 color=gray value=S;
pattern4 color=gray value=S;
pattern5 color=gray value=S;
axis1 width=1 color=black style=1 label=("x");
axis2 width=1 color=black style=1 label=("P(x)")
       order=0 to 0.4 by 0.0625;
proc gchart data=binomial;
      format p_x fract.;
      vbar x / caxes=black ctext=black
       maxis = axis1 raxis=axis2 width=10 space=15
      frame type=sum sumvar=p_x discrete
       noframe
       patternid=midpoint;
run;
quit;
data hyper;
      do x=0 to 10;
      p_x = pdf(hyper',x,100,30,10);
      output;
      end;
run;
goptions device = win ctext = black graphrc;
title1 color = black height = 1.5 cell
"Hypergeometric Distribution: N = 100, p = 0.3, n = 10";
pattern1 color=gray value=S;
pattern2 color=gray value=S;
pattern3 color=gray value=S;
pattern4 color=gray value=S;
```

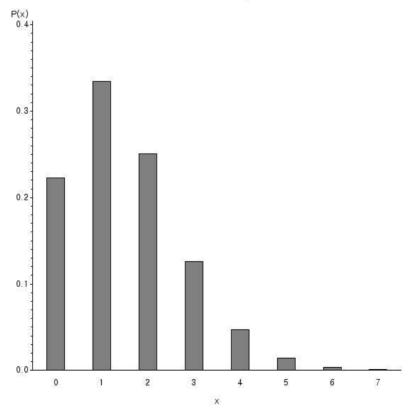
```
pattern5 color=gray value=S;
pattern6 color=gray value=S;
pattern7 color=gray value=S;
pattern8 color=gray value=S;
pattern9 color=gray value=S;
pattern10 color=gray value=S;
axis1 width=1 color=black style=1 label=("x");
axis2 width=1 color=black style=1 label=("P(x)")
       order=0 to 0.3 by 0.02;
proc gchart data=hyper;
      vbar x / caxes=black ctext=black
       maxis = axis1 raxis=axis2 width=5
      frame type=sum sumvar=p_x discrete space=6
       noframe
       patternid=midpoint;
run;
quit;
data poisson;
      do x=0 to 7;
      p_x = pdf(poisson',x,1.5);
      output;
      end:
run:
goptions device = win ctext = black graphrc;
title1 color = black height = 1.5 cell
"Poisson Distribution : \mu = 1.5";
pattern1 color=gray value=S;
pattern2 color=gray value=S;
pattern3 color=gray value=S;
pattern4 color=gray value=S;
pattern5 color=gray value=S;
pattern6 color=gray value=S;
pattern7 color=gray value=S;
```



Hypergeometric Distribution : N = 100, p = 0.3, n = 10



Poisson Distribution : $\mu = 1.5$

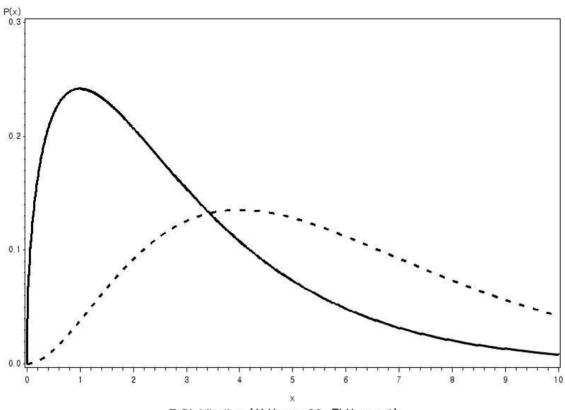


```
코드
```

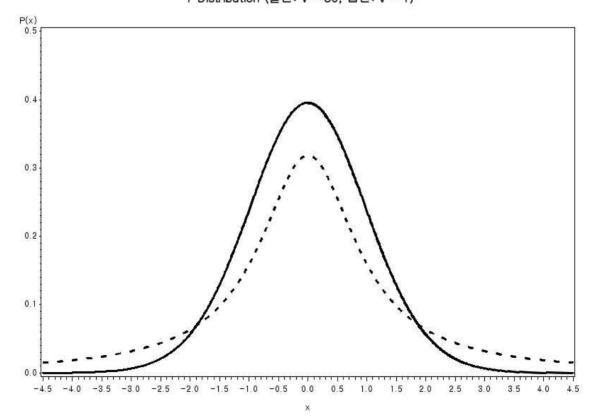
```
data T;
       do x = -4.5 to 4.5 by 0.01;
              p_x = pdf('t',x,30);
              p_y = pdf(t',x,1);
              output;
       end;
run;
goptions device=win ctext=black graphrc interpol=join;
title1 color=black height=1.5 cell
       "T Distribution (실선: v = 30, 점선: v = 1)";
symbol1 c=default I=join L=1 CI=black W=2 V=none;
symbol2 c=default I=join L=34 CI=black W=2 V=none;
Axis1 width=1 color =black style=1 label=("x")
       order=-4.5 to 4.5 by 0.5;
Axis2 width=1 color=black style=1 label=("P(x)")
       order=0 to 0.5 by 0.1;
proc gplot data = T;
       plot (p_x p_y) * x / overlay haxis=axis1 vaxis=axis2 frame;
run;
quit;
data CHI;
       do x = 0 to 10 by 0.01;
              p_x = pdf('chisquared',x,3);
              p_y = pdf('chisquared',x,6);
              output;
       end;
run;
goptions device=win ctext=black graphrc interpol=join;
title1 color=black height=1.5 cell
```

```
"X^2 Distribution (실선: v = 3, 점선: v = 6)";
symbol1 c=default I=join L=1 CI=black W=2 V=none;
symbol2 c=default I=join L=34 CI=black W=2 V=none;
Axis1 width=1 color =black style=1 label=("x")
       order=0 to 10 by 1;
Axis2 width=1 color=black style=1 label=("P(x)")
       order=0 to 0.3 by 0.1;
proc gplot data = CHI;
       plot (p_x p_y) * x / overlay haxis=axis1 vaxis=axis2 frame;
run;
quit;
data F;
       do x = 0 to 2.5 by 0.01;
             p x = pdf(f',x,10,60);
              output;
      end;
run;
goptions device=win ctext=black graphrc interpol=join;
title1 color=black height=1.5 cell
       "F Distribution":
symbol1 c=default I=join L=1 CI=black W=2 V=none;
Axis1 width=3 color=black style=1 label=none value=none
       order=0 to 2.5 by 2.5;
Axis2 width=1 color=black style=1 label=none value=none
      order=0 to 1 by 1;
proc gplot data = F;
       plot p_x * x / haxis=axis1 vaxis=axis2 noframe;
run;
quit;
```

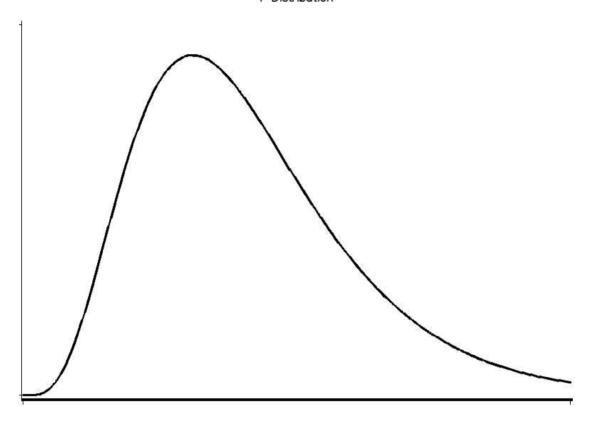
X^2 Distribution (실선: v = 3, 점선: v = 6)



T Distribution (실선: v = 30, 점선: v = 1)



F Distribution



3-7

코드

결과

q1	q 2	q 3
0,81481	0,92049	0,7353

```
X : 3명의 아이를 낳고 난 후 딸아이의 수라 하자. X ~ B(3, 0.43) 이다.
아들의 수는 3 - X이다.
```

- (가) 전체에서 딸아이가 0명일 확률을 빼면 된다. 1 P(X=0) = 0.81481
- (나) 전체에서 딸아이가 3명일 확률을 빼면 된다. 나머지 경우는 모두 아들이 1명 이상이기 때문이다. 1 P(X=3) = 0.92049
- (다) 아들, 딸이 모두 한 명 이상인 경우는 아들만 있을 때 (= 1 (가)), 딸만 있을 때 (= 1 (나))를 제외하고 나머지 모든 경우이다. 따라서 1 (1 (가)) (1 (나)) = (가) + (나) 1 = 0.7353

3-8

```
data exercise8;
      sum1 = 0; /* 앞/뒤 동전 */
      sum2 = 0; /* 앞만 동전 */
      do i = 1 to 10000000;
            x = rand('bern',1/2); /*0이면 앞/뒤 동전, 1이면 앞만 동전*/
            y = 2;
             if(x=0) then do;
                   y = rand('bern',1/2); /*0이면 앞 1이면 뒤*/
             end;
             if(y=0) then do;
                   sum1 = sum1 + 1;
             end;
             if(y=2) then do;
                   sum2 = sum2 + 1;
             end;
      end;
      p = sum1 / (sum1 + sum2);
run;
proc print data = exercise8 (drop = i x y);id;run;
```

sum1	sum2	р
2498582	5002851	0,33308

```
앞/뒤 동전을 뽑고 그 동전에서 앞이 나올 확률 = 1/4
앞만 동전을 뽑고 앞이 나올 확률 = 1/2
앞이 나왔을 때 앞/뒤 동전이었을 확률 = 1/4 / 1/4 + 1/2 = 1/3
랜덤으로 베르누이 시행을 해봤을 때에도 p = 0.33308 = 약 1/3이 나왔다.
```

3-9

```
data exercise9;
      sum1 = 0; /* W 첫번째 주머니 */
      sum2 = 0; /* W 두번째 주머니 */
      do i = 1 to 10000000;
            x = rand('bern',1/2); /*0이면 첫번째 주머니*/
            if(x=0) then do:
                   y = rand('bern',2/9); /*1이면 흰공*/
                   if(y=1) then do;
                         sum1 = sum1 + 1;
                   end;
            end;
            if(x=1) then do;
                   y = rand('bern',5/11); /*1이면 흰공*/
                   if(y=1) then do;
                         sum2 = sum2 + 1;
                   end;
            end;
      end;
      p = sum1 / (sum1 + sum2);
run;
```

proc print data = exercise9 (drop = i x y);id;run;

결과

sum1	sum2	q
1111796	2272633	0,32850

앞면, 흰공 일 확률 : 1/2 * 2/9 뒷면, 흰공 일 확률 : 1/2 * 5/11 흰공 일 때 첫 번째 주머니(앞면)일 확률 : 1/2 * 2/9 / (1/2 * 2/9 + 1/2 * 5/11) = 22/67 = 0.3283582 베르누이 시행으로 얻은 결과 = 0.32850 로 비슷함을 알 수 있다.

3-12

```
P(A) = 0.4 , P(A ∩ B) = 0.2 , P(B) = 0.5

P(A ∪ B) = P(A) + P(B) - P(A ∩ B) = 0.7 (A 또는 B 적어도 하나 구독 확률)

1 - P(A ∪ B) = 0.3 (A 와 B 어느 것도 구독하지 않을 확률)
```

3-14

```
data exercise14;
    sum1 = 0; /* A 공장 불량품 */
    sum2 = 0; /* B 공장 불량품 */
    do i = 1 to 10000000;
        x = rand('bern',0.4); /* 1이면 A 공장*/
        if(x=1) then do;
        y = rand('bern',0.07); /*1이면 불량*/
        if(y=1) then do;
        sum1 = sum1 + 1;
        end;
```

```
end;
if(x=0) then do;
y = rand('bern',0.11); /*1이면 불량*/
if(y=1) then do;
sum2 = sum2 + 1;
end;
end;
end;
p = sum1 / (sum1 + sum2);
run;
proc print data = exercise14 (drop = i x y);id;run;
```

sum1	sum2	р
279646	660205	0,29754

```
A에서 불량제품이 나올 확률: 0.4 * 0.07
B에서 불량제품이 나올 확률: 0.6 * 0.11
불량제품이 A에서 나올 확률: 0.4 * 0.07 / 0.4 * 0.07 + 0.6 * 0.11 = 14/47 = 0.29787234 직접 구한 결과인 0.29754와 비슷하다.
```

3-16

```
data exercise16_1;

x = 0;

y = 0;

do i = 1 to 10000000;

z = rand('integer',1,6);

x = 0;

y = 0;
```

```
if(z=2 or z=4 or z=6) then do;
                    x = x+1;
              end;
              if(w=2 or w=4 or w=6) then do;
                    x = x+1;
              end:
              if(z=3 or z=6) then do;
                    y = y+1;
              end;
              if(w=3 or w=6) then do;
                    y = y+1;
              end;
              output;
      end;
run;
proc freq data = exercise16_1;
      table y * x / nocol norow;
run;
data exercise16_2;
      x0 = 0;
      x1 = 0;
      x2 = 0;
      y0 = 0;
      y1 = 0;
      y2 = 0;
      do i = 1 to 10000000;
             z = rand('integer', 1, 6);
             w = rand('integer', 1, 6);
             x = 0;
             y = 0;
              if(z=2 or z=4 or z=6) then do;x = x+1;end;
              if(w=2 or w=4 or w=6) then do;x = x+1;end;
```

$$if(z=3 \text{ or } z=6) \text{ then do; } y = y+1; end; \\ if(w=3 \text{ or } w=6) \text{ then do; } y = y+1; end; \\ if(x = 0) \text{ then do; } x0 = x0 + 1; end; \\ if(x = 1) \text{ then do; } x1 = x1 + 1; end; \\ if(x = 2) \text{ then do; } x2 = x2 + 1; end; \\ if(y = 0) \text{ then do; } y0 = y0 + 1; end; \\ if(y = 1) \text{ then do; } y1 = y1 + 1; end; \\ if(y = 2) \text{ then do; } y2 = y2 + 1; end; \\ end; \\ x = x0 + x1 + x2; \\ y = y0 + y1 + y2; \\ x0 = x0 / x; x1 = x1 / x; x2 = x2 / x; \\ y0 = y0 / y; y1 = y1 / y; y2 = y2 / y; \\ \end{cases}$$

run;

proc print data = exercise16_2 (drop = x y z w i);id;run;

결과

(가)

경우의 수는 총 6 * 6 = 36가지이고, 각 (X, Y)에 대해 결합확률 분포를 구하면

$$(0, 0) = (1,1),(1,5),(5,1),(5,5) = 47|X| = 4/36 = 1/9$$

$$(0, 1) = (1,3),(3,1),(3,5),(5,3) = 1/9$$

$$(0, 2) = (3,3) = 1/36$$

$$(1, 0) = (1,2),(1,4),(2,1),(2,5),(4,1),(4,5),(5,2),(5,4) = 2/9$$

$$(1, 1) = (1,6),(2,3),(3,2),(3,4),(4,3),(5,6),(6,1),(6,5) = 2/9$$

$$(1, 2) = (3,6),(6,3) = 1/18$$

$$(2, 0) = (2,2),(2,4),(4,2),(4,4) = 1/9$$

$$(2, 1) = (2,6),(4,6),(6,2),(6,4) = 1/9$$

$$(2, 2) = (6,6) = 1/36$$

결합분포표로 나타내면

YX	0	1	2
0	1/9	2/9	1/9
1	1/9	2/9	1/9
2	1/36	1/18	1/36

이고, 직접 구한 결과와 비교해보면

FREQ 프로시저					
빈도			테이블 y	* X	
백분율			>	(
	У	0	1	2	합계
	0	1109821 11,10	2222860 22,23	1110482 11,10	4443163 44.43
	1	1111820 11,12	2223526 22,24	1110504 11,11	4445850 44.46
	2	277543 2,78	556560 5,57	276884 2,77	1110987 11,11
	합계	2499184 24,99	5002946 50,03	2497870 24,98	1E7 100,00

비슷함을 알 수 있다.

(나)

X의 주변확률분포를 구하면

$$P(X = 0) = (4 + 4 + 1) / 36 = 1/4$$

$$P(X = 1) = (8 + 8 + 2) / 36 = 1/2$$

$$P(X = 2) = (4 + 4 + 1) / 36 = 1/4$$

Y의 주변확률분포를 구하면

$$P(Y = 0) = (4 + 8 + 4) / 36 = 4/9$$

$$P(Y = 1) = (4 + 8 + 4) / 36 = 4/9$$

$$P(Y = 2) = (1 + 2 + 1) / 36 = 1/9$$

×0	x1	x2	у0	у1	у2
0,25002	0,50010	0,24989	0,44439	0.44445	0,11116

직접 구한 결과와 비슷하다.

3-18

코드

```
data exercise18;
```

```
AO = 0; /* A 공장 양품 */
```

$$BO = 0$$
; $BX = 0$; $CO = 0$; $CX = 0$;

do
$$i = 1$$
 to 10000000;

if(
$$x=1$$
) then do;

if(y=1) then do;
$$AX = AX + 1$$
; end;

else do;
$$AO = AO + 1$$
; end;

end;

else do;

y = rand('bern',3/8); /* 1이면 B공장*/

```
if(y=1) then do;
                           z = rand(bern', 0.02);
                           if(z=1) then do; BX = BX + 1; end;
                           else do; BO = BO + 1; end;
                    end;
                    else do; /* C 공장 */
                           z = rand(bern', 0.03);
                           if(z=1) then do; CX = CX + 1; end;
                           else do; CO = CO + 1; end;
                    end;
             end:
      end;
      q1 = AX / (AX + BX + CX);
      q2 = AO / (AO + BO + CO);
run;
proc print data = exercise18 (drop = AO AX BO BX CO CX x y z
i);id;run;
```

q1	q2
0,087205	0,20265

(가) 0.2*0.01/(0.2*0.01+0.3*0.02+0.5*0.03) = 2/23 = 0.08695 (나) 0.2*0.99/(0.2*0.99+0.3*0.98+0.5*0.97) = 198/977 = 0.202661 (가)의 경우 표본이 작아서 오차가 조금 있는 것 같다.

3-20

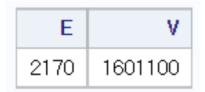
```
data exercise20; E = 5000 * 0.12 + 3000 * 0.18 + 2000 * 0.33 + 1000 * 0.37;
```

$$V = (5000 - E)*(5000 - E)*0.12+(3000 - E)*(3000 - E)*0.18+ \\ (2000 - E)*(2000 - E)*0.33+(1000 - E)*(1000 - E)*0.37;$$

run;

proc print data = exercise20;id;run;

결과



$$E(X) = 2170, V(X) = 1601100$$

3-21

코드

data exercise21;

$$E = 0*0.05+1*0.3+2*0.3+3*0.2+4*0.1+5*0.05;$$

$$V = (0-E)*(0-E)*0.05 + (1-E)*(1-E)*0.3 + (2-E)*(2-E)*0.3 + (3-E)*(3-E)*0.2 + (4-E)*(4-E)*0.1 + (5-E)*(5-E)*0.05;$$

run;

proc print data = exercise21;id;run;

결과

(가)
$$E(X) = 2.15$$
, $V(X) = 1.5275$

$$V(2X-1) = 4*V(X) = 6.11$$

3-22

(가)
$$\int_0^1 ax^2 = 1$$
이므로 a = 3

$$(\Box)P(X\leq \frac{1}{2})=\int_{0}^{\frac{1}{2}}3x^{2}=\frac{1}{8}$$

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

$$P(\frac{1}{4} \le X \le \frac{3}{4}) = \int_{\frac{1}{4}}^{\frac{3}{4}} 3x^2 = \frac{27}{64} - \frac{1}{64} = \frac{13}{32}$$

3-25

코드

data exercise25;

$$p = 1 - cdf(binomial', 2, 0.1, 15);$$

run;

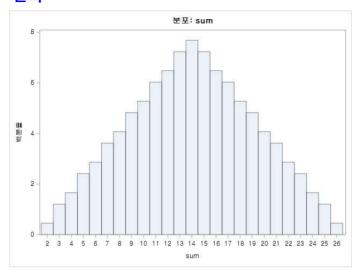
proc print data = exercise25;id;run;

결과



 $X: 15면에서 오자가 발생한 면 개수라 하면 <math>X \sim B(15, 0.1)$ 이다. $P(X \ge 3) = 1 - P(X < 3) = 1 - P(X \le 2) = 0.18406$

```
data exercise28;
   do i = 1 to 51;
      do j = i+1 to 52;
              x = i;
              y = j;
         do while (x > 13);
            x = x - 13;
         end;
         do while (y > 13);
            y = y - 13;
         end;
         sum = x + y;
         output;
      end;
   end;
run;
proc univariate data=exercise28 noprint;
   histogram sum / midpoints = 2 to 26 by 1;
run;
proc freq data = exercise28;
      table sum;
run;
```



(가) 이산확률 분포를 따른다. 이산 확률 변수는 freq로 그린 빈도표에서 확인할 수 있다.

(L†)
$$P(X < 16) = 810/1326$$

 $P(X \ge 17) = 1 - 896/1326 = 430/1326$
 $E(X) = \frac{810}{1326} * 1000 + \frac{430}{1326} * (-1000)$
=286.576168929

FREQ 프로시저

sum	빈도	백분율	누적 빈도	누적 백분율
2	6	0,45	6	0,45
3	16	1,21	22	1,66
4	22	1,66	44	3,32
5	32	2,41	76	5,73
6	38	2,87	114	8,60
7	4 8	3,62	162	12,22
8	54	4,07	216	16,29
9	64	4,83	280	21,12
10	70	5,28	350	26,40
11	80	6,03	430	32,43
12	86	6,49	516	38,91
13	96	7,24	612	46,15
14	102	7,69	714	53,85
15	96	7,24	810	61,09
16	86	6,49	896	67,57
17	80	6,03	976	73,60
18	70	5,28	1046	78,88
19	64	4,83	1110	83,71
20	54	4.07	1164	87,78
21	48	3,62	1212	91,40
22	38	2,87	1250	94,27
23	32	2,41	1282	96,68
24	22	1,66	1304	98,34
25	16	1,21	1320	99,55
26	6	0,45	1326	100,00

3-30

코드

data exercise30;

$$p = 1 - cdf('normal',300,280,sqrt(250));$$

$$p = 100 * p;$$

run;

proc print data = exercise30;id;run;

p 10,2952

(가) 빵 ~ N(200, 15^2), 우유 ~ N(80, 5^2), 빵+우유 ~ N(200+80, 15^2+5^2) 따라서 1인당 섭취하는 열량 ~ N(280, $5\sqrt{10}$)이다. (나) P(X \geq 300) = 1 - P(X < 300) = 0.102952 따라서 1년 중 약 10.3% 정도는 목표를 달성할 수 있다. 이는 일수로 약 38일이다.

3-31

코드

data exercise31; p = 1 - cdf('normal',700,670,sqrt(2420));run; proc print data = exercise31;id;run;

결과



하루 방출되는 황산의 양 ~ N(134, 22^2) 5일 방출되는 황산의 양 ~ N(134*5, 22^2*5) = N(670, 2420) P(X \geq 700) = 1 - P(X < 700) = 0.27098

3-33

코드

data exercise33;
 E = 600 * 3 / 8;
 S = 600 * 3 * 5 / (8 * 8);
 S = sqrt(S);
 q2 = cdf('binomial',220,3/8,600);
 q3 = 1 - cdf('binomial',250,3/8,600);
run;
proc print data = exercise33;id;run;

결과

	Е	S	q 2	q3
:	225	11,8585	0,35330	0,016221

(가) 3개 동전을 동시에 던지는 시행에서 앞면이 2개, 뒷면이 1개 나올 확률은 3/8이다. 즉, X는 성공확률이 3/8인 이항분포를 따른다.

$$X \sim B(600, 3/8)$$

$$E(X) = np = 600 * 3/8 = 225$$

$$V(X) = npq = 600 * 3/8 * 5/8$$

$$\sigma = \sqrt{V(X)} = 11.8585$$

$$(\Box t) P(X \ge 250) = 1 - P(X < 250) = 0.016221$$

3-34

코드

결과

q1	q2	q3
0,94999	0,97499	0,89000

(7t)
$$P(\chi^2 \le 15.507) = 0.94999$$

(Lt) $P(\chi^2 \ge 2.18) = 1 - P(\chi^2 < 2.18) = 0.97499$
(Lt) $P(1.647 \le \chi^2 \le 13.362) = P(\chi^2 \le 13.362) - P(\chi^2 < 1.647) = 0.89$

3-36

```
data exercise36;
    q1 = cdf('t',1.397,8);
    q2 = cdf('t',1.18,8) - cdf('t',-1.397,8);
    q3 = tinv(1 - 0.025,8);
run;
proc print data = exercise36;id;run;
```

q1	q 2	q3
0,90003	0,76408	2,30600

- (71) P(T \leq 1.397) = 0.90003
- (Lt) $P(-1.397 \le T \le 1.18) = P(T \le 1.18) P(T < -1.397) = 0.76408$
- (\Box) P(T \geq t) = 0.025, P(T < t) = 0.975, T < t = P⁻¹(0.975) = 2.306