**SAS 데이터 분석 입문 3장**

**2019020650 김형욱**

**\* 3장 예제문제**

**<예 3-1>**

**data** discrete;

a\_1 = pdf('binomial', **1**, **0.05**, **100**);

a\_2 = cdf('binomial', **10**, **0.05**, **100**);

b\_1 = pdf('hyper', **0**, **1000**, **50**, **10**);

b\_2 = cdf('hyper', **5**, **1000**, **50**, **10**) - cdf('hyper', **2**, **1000**, **50**, **10**);

c\_1 = pdf('poisson', **1**, **5**);

c\_2 = cdf('poisson', **10**, **5**);

**run**;

**proc** **print** data=discrete;

**run**;

**data** prob1;

a\_2 = probbnml(**0.05**, **100**, **10**);

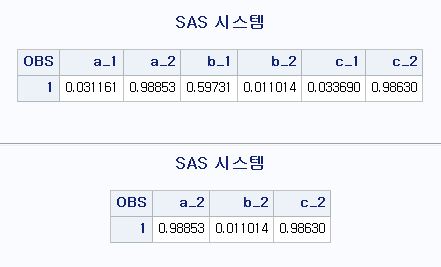
b\_2 = probhypr(**1000**, **50**, **10**, **5**) - probhypr(**1000**, **50**, **10**, **2**);

c\_2 = poisson(**5**, **10**);

**run**;

**proc** **print** data=prob1;

**run**;



**<예 3-2>**

**data** continue;

a\_1 = cdf('normal', -**1.6449**, **0**, **1**);

a\_2 = probit(**0.05**);

b\_1 = cdf('t', **1.6449**, **50**) - cdf('t', -**1.6449**, **50**);

b\_2 = tinv(**0.05**, **50**);

c\_1 = **1** - cdf('chisquared', **1**, **5**);

c\_2 = cinv(**0.95**, **5**);

d\_1 = cdf('f', **1**, **3**, **10**);

d\_2 = finv(**0.05**, **3**, **10**);

**run**;

**proc** **print** data=continue;

**run**;

**data** prob2;

a\_1 = probnorm(-**1.6449**);

b\_1 = probt(**1.6449**, **50**) - probt(-**1.6449**, **50**);

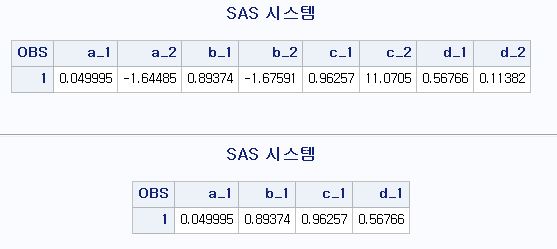
c\_1 = **1** - probchi(**1**, **5**);

d\_1 = probf(**1**, **3**, **10**);

**run**;

**proc** **print** data=prob2;

**run**;



**\* 3장 연습문제**

**<연습문제 3-2>**

/\* 복원추출인 경우 : 이항분포 사용 \*/

**data** sasadv.ex3\_2\_1;

a\_1 = pdf('binomial', **10**, **0.95**, **10**);

**run**;

**proc** **print** data=sasadv.ex3\_2\_1;

**run**;

/\* 비복원추출인 경우 : 초기하분포 사용 \*/

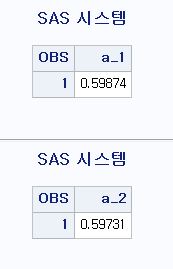
**data** sasadv.ex3\_2\_2;

a\_2 = pdf('hyper', **10**, **1000**, **950**, **10**);

**run**;

**proc** **print** data=sasadv.ex3\_2\_2;

**run**;



해석 : 복원추출을 한 경우에 여자가 한 명도 추출되지 않을 확률은 0.59874이고, 비복원추출을 한 경우에 여자가 한 명도 추출되지 않을 확률은 0.59731이다.

**<연습문제 3-3>**

/\* (가) \*/

**data** sasadv.ex3\_3\_1;

a\_1 = **1** - cdf('normal', **180**, **170**, **5**);

**run**;

**proc** **print** data=sasadv.ex3\_3\_1;

**run**;



해석 : 성인 남자 중에서 한 사람을 추출하였을 때 그 사람의 키가 180cm 이상일 확률은 0.02275이다.

/\* (나) \*/

**data** sasadv.ex3\_3\_2;

a\_2 = cdf('normal', **175**, **170**, **0.5**);

**run**;

**proc** **print** data=sasadv.ex3\_3\_2;

**run**;



해석 : 성인 남자 중에서 100명을 추출하였을 때 표본평균이 175cm 이하일 확률은 1이다.

/\* (다) \*/

해석 : 크기가 25인 랜덤표본을 추출하여 표본평균을 구하였을 떄와 크기가 10인 랜덤표본을 추출하여 표본평균을 구하였을 때 각각의 평균의 표준오차는 다음과 같다.

DRW0000119c15b6 DRW0000119c15bb

**<연습문제 3-7>**

**(가)**

DRW00002b402216

**(나)**

DRW00002b40221b

**(다)**

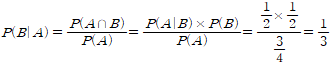
아들과 딸이 적어도 한 명 이상 나올 수 있는 경우의 수는 (아들, 딸, 아들), (아들, 딸, 딸), (딸, 아들, 아들), (딸, 아들, 딸)이다. 그러므로 아들과 딸이 적어도 한 명 이상인 확률은 4/8=0.5이다.

**<연습문제 3-8>**

동전의 앞면이 나올 사건을 A, 앞면과 뒷면이 나올 확률이 같은 동전일 사건을 B라고 하자.

DRW00001d4c34a3

주머니에서 동전 1개를 꺼내어 던졌는데 앞면이 나왔을 때, 이 동전이 앞면과 뒷면이 나올 확률이 같은 동전이었을 확률은 다음과 같다.



**<연습문제 3-30>**

**(가)**

DRW00002b402225

따라서 이 가정의 아침식사에서 일인당 섭취하는 열량의 평균은 280(칼로리)이고, 표준편차는 DRW00002b40222a이다.

**(나)**

DRW00002b40222f

그러므로 이 가정이 매일 아침식사에서 300칼로리 이상을 섭취하려고 목표한다면 1년 중 약 10.3%에 해당하는 날에 목표가 달성될 것이다.

**<연습문제 3-33>**

**(가)**

DRW00002b402234

DRW00002b402239

그러므로 X의 평균은 225이고, 표준편차는 11.8585이다.

**(나)**

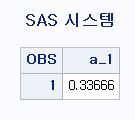
**data** sasadv.ex3\_33\_1;

a\_1 = **1** - cdf('normal', **0.4216**, **0**, **1**);

**run**;

**proc** **print** data=sasadv.ex3\_33\_1;

**run**;



그러므로 P(X<=220)=0.33666이다.

(n=600으로 크기 때문에 중심극한정리를 사용하여 이항분포의 정규근사를 사용함)

**(다)**

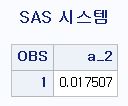
**data** sasadv.ex3\_33\_2;

a\_2 = **1** - cdf('normal', **2.1082**, **0**, **1**);

**run**;

**proc** **print** data=sasadv.ex3\_33\_2;

**run**;



그러므로 P(X>=250)=0.017507이다.

(n=600으로 크기 때문에 중심극한정리를 사용하여 이항분포의 정규근사를 사용함)