

Homework 3

1. 1981년에 조지프 쉘리츠 양조회사는 쇠락해가는 브랜드를 되살리기 위해 170만 달러를 들여 과감한 마케팅 행사를 열었다. NFL 슈퍼볼 하프타임에 자사의 맥주와 주요 경쟁사인 미켈롭 맥주의 맛을 겨루는 블라인드 테스트를 생방송으로 내보냈다. 미켈롭 등 경쟁 브랜드의 맥주를 즐겨마시는 소비자 100명을 대상으로 그들이 좋아하는 맥주와 쉘리츠 맥주를 두고 블라인드 테스트가 진행됐다. 소비자가 쉘리츠를 선택할 확률이 50%로 동일하고 서로 독립적이라고 가정하자. (쉘리츠가 속한 제품군에 있는 맥주들은 대부분 맛이 비슷하므로 일반인들은 구분하지 못한다는 가정)



(a) 만약 100명의 테스트 참가자 중에서 적어도 40%가 쉘리츠를 고르면 쉘리츠 사가 만족했을 거라고 가정해 보자. (미켈롭을 좋아하는 소비자의 40%가 쉘리츠를 선택했다니!!) 쉘리츠 사가 만족할 확률은?

In [3]:

```
import numpy as np
from scipy.stats import binom

try_count = 100 # 시도 : 테스트 참가자
hope_count = (40 / 100) * try_count # 기대 : 100명 중에 40%가 선택하면 숄리츠 만족. 즉, 40명이
    선택하면 숄리츠 만족
prob = 0.5 # 숄리츠를 선택할 확률 50%, 서로 독립

# 적어도 40%가 숄리츠를 고르는 경우 : 40% ~ 100%
# 확률질량함수
x = np.arange(hope_count, try_count)
result = binom.pmf(x, try_count, prob)
result.sum()
```

Out[3]:

0.982399899891119

In [4]:

```
# cdf 를 이용한 검증
1 - binom.cdf(hope_count - 1, try_count, prob) ## 왜 조금 다르지. !?!?!?! 파이썬 부동소숫점
    계산이
```

Out[4]:

0.9823998998911476

(b) 만약 테스트 참가자가 500명이라면 숄리츠 사가 만족할 확률은?

In [5]:

```
import numpy as np
from scipy.stats import binom

try_count = 500 # 시도 : 테스트 참가자
hope_count = (40 / 100) * try_count # 기대 : 100명 중에 40%가 선택하면 숄리츠 만족. 즉, 200명이
    선택하면 숄리츠 만족
prob = 0.5 # 숄리츠를 선택할 확률 50%, 서로 독립

# 적어도 40%가 숄리츠를 고르는 경우 : 40% ~ 100%
# 확률질량함수
x = np.arange(hope_count, try_count)
result = binom.pmf(x, try_count, prob)
result.sum()
```

Out[5]:

0.9999970742216363

In [6]:

```
# cdf 를 이용한 검증
1 - binom.cdf(hope_count - 1, try_count, prob) ## 마찬가지로..왜 조금 다르지. !?!?!?!? 원래 그런가 ?? 무시하기에 째째함...
```

Out[6]:

0.9999970742217401

2. TV로 방송되는 프로미식축구의 평균 작전타임 시간은 정규분포로서 평균이 84초이고 표준편차는 10초이다. 어떤 방송국이 90초의 광고시간을 갖고 있다면, 광고가 끝나기 전에 경기가 다시 시작될 확률은?

In [7]:

```
import numpy as np
from scipy.stats import norm

mu = 84 # 평균( $\mu$ )
sd = 10 # 표준편차( $\sigma$ )
x = 90 # 관심 확률변수

norm.cdf(x, mu, sd) # 이산확률변수와 연속확률변수의 누적분포함수. 표본 값을 입력하면 해당 표본 값에 대한 누적확률을 출력.

# import matplotlib.pyplot as plt
# xx = np.linspace(0, 100, 100)
# cdf = norm.cdf(xx, mu, sd)
# plt.plot(xx, cdf)
# plt.title("cdf(x, 84, 10) ")
# plt.xlabel("$x$")
# plt.ylabel("$F(x)$")
# plt.show()
```

Out[7]:

0.7257468822499265

3. 콜로라도에서 잡히는 송어의 길이는 정규분포를 따르고 평균이 12.5인치이고 표준편차는 1.2인치이다. 잡히는 송어의 80%를 집에 가져가고 나머지는 다시 방류하게 하려면, 담당 부서에서 허용하는 최소 송어의 길이는?

In [8]:

```
import numpy as np
from scipy.stats import norm

mu = 12.5 # 평균( $\mu$ )
sd = 1.2 # 표준편차( $\sigma$ )
x = 1 - (80 / 100) # 관심 확률변수, 최소 송어길이 : 방류되는 20% 지점
norm.ppf(x, mu, sd) # 누적분포함수의 역함수. 즉, cdf의 역함수로 x 분위수를 구하는 함수
```

Out[8]:

11.490054519712503

4. 개인이 지불하는 자동차 보험료의 연간 평균은 687달러이다. 이 값을 모집단 평균으로 하고 모집단의 표준편차가 230달러이다. 45개의 자동차 보험계약을 생각해보자. 45개 보험료의 평균이 모집단 평균의 ± 25 달러 이내에 있을 확률은 얼마인가?

In [9]:

```
import numpy as np
from scipy.stats import norm

mu = 687 # 모집단 평균( $\mu$ )
sd = 230 # 모집단 표준편차( $\sigma$ )
try_count = 45 # 시행수
hope_range = 25 # 기대 범위

# 일단 표준오차 구하고
se = sd / np.sqrt(try_count)

#  $\bar{X}$ 가 662 ~ 712 사이 확률 구하기
prob_max = norm.cdf(mu + hope_range, loc=mu, scale=se)
prob_min = norm.cdf(mu - hope_range, loc=mu, scale=se)
prob_max - prob_min
```

Out[9]:

0.5340916787127394