문제 **1.** 로또 **6/45**의 **1**등**, 2**등**, 3**등 당첨 확률은 얼마인가**?** 로또 **1**등 당첨자가 평균적인 확률에 비해 이상하게 많다고 할 수 있는지**?**

풀이:

우선 로또 숫자 조합을 계산해보면, 1부터 45까지 숫자 중 6개를 순서에 상관없이 선택한다. 따라서 가능한 총 조합의 수는 아래와 같다.

$$\frac{45!}{6!(45-6)!} = 8{,}145{,}060$$

1) 1등 당첨 확률 : 6개를 정확히 맞춰야 하기 때문에 경우의 수는 1개이다.
   따라서 확률은
   $P(1\text{등}) = \frac{1}{8{,}145{,}060}$

2) 2등 당첨 확률 : 6개 중 5개를 맞추고, 남은 한 개가 보너스 숫자와 같아야 한다.
   따라서 6개 중 5개를 선택할 경우의 수는 6개이고, 나머지 숫자는 반드시 보너스 숫자여야 하므로 최종적으로 경우의 수는 6이다.
   따라서 확률은
   $P(2\text{등}) = \frac{6}{8{,}145{,}060}$

3) 3등 당첨 확률 : 6개 중에 5개의 숫자만 일치하면 된다. 따라서 앞 서 본 것과 같이 6개 중 5개를 선택할 경우의 수는 6이고, 여기에 나머지 숫자는 당첨 숫자 6개와 보너스 숫자 1개를 제외한 38개의 경우의 수가 존재할 수 있다. 따라서
   $6 \times 38 = 228$
   따라서 확률은
   $P(3\text{등}) = \frac{228}{8{,}145{,}060}$

로또 1등 당첨자가 평균적인 확률에 비해 이상하게 많다고 할 수 있는지?
동행복권 로또히스토리 사이트를 접속해보자.
(https://dhlottery.co.kr/gameInfo.do?method=statHistory)

로또 6/45의 히스토리를 안내해 드립니다.

제 1회부터 제 1161회까지 누적 히스토리 입니다.

현재까지 총 판매금액	80,226,476,220,000 원
현재까지의 총 당첨금액	40,113,238,110,000 원
현재까지 누적 1등 당첨자수	9,413 명
현재까지 누적 2등 당첨자수	57,019 명
현재까지 누적 3등 당첨자수	2,152,917 명

1회부터 1161회까지의 로또 관련 데이터가 나와있다.

1등 당첨자의 수가 9,413명이라고 나와 있다.
우리가 앞서 1등 당첨자의 확률은 평균적으로 8,145,060장 판매 될 때마다 1장이 나온다고 보았다.
이 확률을 가지고 계산해보면, 1161회까지 9,413명의 1등 당첨자가 나왔을 때, 기대 판매금액은

$$\text{기대판매금액} = 8,145,060\text{장} \times 1161\text{회} \times 1000\text{원} = 76,669,449,780,000\text{원}$$

그런데 실제 판매금액은 80,226,476,220,00원이므로 평균 확률보다 더 적게 1등 당첨이 됐었음을 알 수 있다.

다만 1등 당첨자와 2, 3등 당첨자와의 확률 비는 9,413 : 57,019 : 2,152,917 = 1 : 6.057 : 228.717

이 비율은 위에서 구한 경우의 수와 유사함을 알 수 있다.

문제 **2. Bayes Theorem**

Bayes Theorem example

Suppose a drug test is 99% sensitive and 99% specific. That is, the test will produce 99% true positive results for drug users and 99% true negative results for non-drug users. Suppose that 0.5% of people are users of the drug. If a randomly selected individual tests positive, what is the probability he or she is a user?

위의 내용을 정리해보면 다음과 같다.

$P(true\ positive \mid drug\ user) = 0.99$
$P(false\ negative \mid drug\ user) = 0.01$

약 투여자 중 99%는 정상 양성이 나오지만, 1%는 오검출되어 위음성이 나온다.

$P(true\ negative \mid non - drug\ user) = 0.99$
$P(flalse\ positive \mid non - drug\ user) = 0.01$

약 비사용자 중 99%는 정상 음성이 나오지만, 1%는 오검출되어 위양성이 나온다.

$P(drug\ user) = 0.005$
$P(non - drug\ user) = 0.995$

$P(drug\ user \mid positive) = ?$

$= \frac{P(user \cap positive)}{P(positive)} = \frac{P(true\ positive|drug\ user) \cdot P(drug\ user)}{P(true\ positive \mid drug\ user) \cdot P(drug\ user) + P(false\ positive \mid non\text{-}drug\ user) \cdot P(non\text{-}drug\ user)}$
$= \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.01 \cdot 0.995}$
$= \frac{0.00495}{0.0149}$
$= 0.332214765100671$ **(**약 **33%)**

문제 **3. Bayes Inference**

A, B, C 중 한 사람이 근무 중 작업을 한다.

30일 기준으로 A의 작업 일수는 평균 18일, B의 작업 일수는 9일, C의 작업 일수는 3일이고,

A, B, C가 작업 시 실수할 확률은 각각 0.3%, 0.7%, 1%로 알려져 있다.

1) 전 날 작업한 결과물 중 하나를 샘플로 검사하였더니 실수가 발견되었다면 전 날 A, B, C 중 누가 작업했을 확률이 가장 높은가?

$P(A\text{가 실수} \mid A\text{가 근무}) = 0.003$
$P(B\text{가 실수} \mid B\text{가 근무}) = 0.007$
$P(C\text{가 실수} \mid C\text{가 근무}) = 0.01$

$P(A\text{가 근무할 확률}) = 18/30$
$P(B\text{가 근무할 확률}) = 9/30$
$P(C\text{가 근무할 확률}) = 3/30$

$$P(A\text{가 근무} \mid \text{실수가 발견}) = \frac{P(A\text{가 근무} \cap \text{실수가 발견})}{P(\text{실수가 발견})}$$

$$= \frac{P(\text{실수가 발견} \mid A\text{가 근무}) \cdot P(A\text{가 근무})}{P(\text{실수가 발견} \mid A\text{가 근무}) \cdot P(A\text{가 근무}) + P(\text{실수가 발견} \mid B\text{가 근무}) \cdot P(B\text{가 근무}) + P(\text{실수가 발견} \mid C\text{가 근무}) \cdot P(C\text{가 근무})}$$
$$= \frac{0.003 \times 18/30}{0.003 \times 18/30 + 0.007 \times 9/30 + 0.01 \times 3/30}$$
$$= \frac{0.0018}{0.0018 + 0.0021 + 0.001} = 0.367346$$

$$P(B\text{가 근무} \mid \text{실수가 발견}) = \frac{P(B\text{가 근무} \cap \text{실수가 발견})}{P(\text{실수가 발견})}$$

$$= \frac{P(\text{실수가 발견} \mid B\text{가 근무}) \cdot P(B\text{가 근무})}{P(\text{실수가 발견} \mid A\text{가 근무}) \cdot P(A\text{가 근무}) + P(\text{실수가 발견} \mid B\text{가 근무}) \cdot P(B\text{가 근무}) + P(\text{실수가 발견} \mid C\text{가 근무}) \cdot P(C\text{가 근무})}$$
$$= \frac{0.007 \times 9/30}{0.003 \times 18/30 + 0.007 \times 9/30 + 0.01 \times 3/30}$$
$$= \frac{0.0021}{0.0018 + 0.0021 + 0.001} = 0.428571$$

$$P(C\text{가 근무} \mid \text{실수가 발견}) = \frac{P(C\text{가 근무} \cap \text{실수가 발견})}{P(\text{실수가 발견})}$$

$$= \frac{P(\text{실수가 발견} \mid C\text{가 근무}) \cdot P(C\text{가 근무})}{P(\text{실수가 발견} \mid A\text{가 근무}) \cdot P(A\text{가 근무}) + P(\text{실수가 발견} \mid B\text{가 근무}) \cdot P(B\text{가 근무}) + P(\text{실수가 발견} \mid C\text{가 근무}) \cdot P(C\text{가 근무})}$$
$$= \frac{0.01 \times 3/30}{0.003 \times 18/30 + 0.007 \times 9/30 + 0.01 \times 3/30}$$

$$= \frac{0.001}{0.0018 + 0.0021 + 0.001} = 0.204081$$

따라서 B의 실수일 확률이 약 **42.85%**로 가장 높다.

2) 샘플을 하나 더 열어봤더니 또 실수가 있었다면 전 날 작업한 사람의 확률은 어떻게 달라지는가

A의 경우를 생각해보자.

$P(A\text{가 }1\text{차 실수} \mid 2\text{차 실수}) =$

$$\frac{P(2\text{차 실수} \mid A\text{가 }1\text{차 실수}) \times P(A\text{가 }1\text{차 실수})}{P(2\text{차 실수} \mid A\text{가 }1\text{차 실수}) \times P(A\text{가 }1\text{차 실수}) + P(2\text{차 실수} \mid B\text{가 }1\text{차 실수}) \times P(B\text{가 }1\text{차 실수}) + P(2\text{차 실수} \mid C\text{가 }1\text{차 실수}) \times P(C\text{가 }1\text{치 실수})}$$

여기서 $P(2\text{차 실수} \mid A\text{가 }1\text{차 실수})$는 사실 상 $P(2\text{차 실수} \mid A)$와 같다고 볼 수 있다.
왜냐하면 이미 A가 1차 때 실수 했다는 건, 이 날 작업을 A가 하고 있었다는 것을 전제하기 때문.

따라서 $P(2\text{차 실수} \mid A\text{가 }1\text{차 실수}) = 0.003$ 즉, A가 실수할 확률과 동일하다.
그리고 $P(A\text{가 }1\text{차 실수}) = P(A\text{가 근무} \mid \text{실수가 발견}) = 0.367346$

이와 같이 정리하면,

$P(2\text{차 실수} \mid B\text{가 }1\text{차 실수}) = 0.007$
$P(2\text{차 실수} \mid B\text{가 }1\text{차 실수}) = 0.01$

$P(B\text{가 }1\text{차 실수}) = 0.428571$
$P(C\text{가 }1\text{치 실수}) = 0.204081$

이를 대입하여 정리하면,

$$P(A\text{가 }1\text{차 실수} \mid 2\text{차 실수}) = \frac{0.003\times0.367346}{0.003\times0.367346+0.007\times0.428571+0.01\times0.204081} = \frac{0.001102}{0.006141} = 0.1794$$
$$P(B\text{가 }1\text{차 실수} \mid 2\text{차 실수}) = \frac{0.007\times0.428571}{0.003\times0.367346+0.007\times0.428571+0.01\times0.204081} = \frac{0.002999}{0.006141} = 0.4883$$
$$P(B\text{가 }1\text{차 실수} \mid 2\text{차 실수}) = \frac{0.01\times0.204081}{0.003\times0.367346+0.007\times0.428571+0.01\times0.204081} = \frac{0.002040}{0.006141} = 0.3321$$

연속 2개의 샘플을 확인했을 때, 2개 모두 실수인 경우에도 전 날 작업자가 **B**일 확률이 약 **49%**로 가장 높다.