

출석수업 과제물(평가결과물) 표지(온라인제출용)

교과목명 : 확률의 개념과 응용

학 번 : 202135-367895

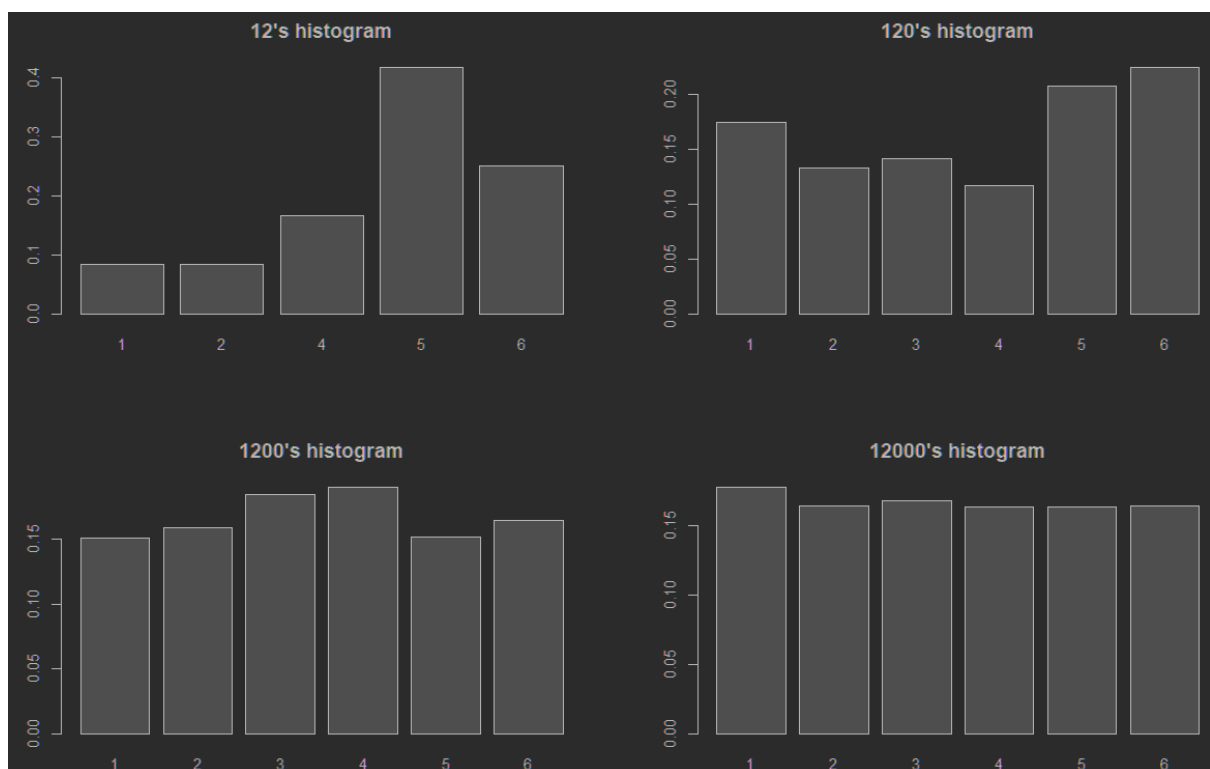
성 명 : 김태정

강 의 실 : 지역대학 호

연 락 처 : 010-4172-4516

1 번문제

```
par(mfrow = c(2, 2))
set.seed(367895)
for (num in c(12, 120, 1200, 12000)) {
  barplot(table(sample(1:6, num, replace = T))) / num, main = paste0(num, 'W's
  histogram'))
}
```



상대 도수의 개념은 (계급의 도수)/(총 도수)로 계산할 수 있다. 이 개념은 총 도수에 대한 각 변량의 도수 비를 의미한다. 다른 말로는 상대 빈도라고 하며 비의 개념이기에 값이 백분율로 나오게 된다. 해당 시험결과 횃수가 적을 경우 각각의 도수의 상대 도수가 들쭉날쭉 하지만 시행이 늘어나면서 각각이 독립시행이기에 대수의 법칙을 통해서 모든 눈금의 상대 도수가 비슷한 값에 수렴함을 알 수 있다. 그래서 이 실험은 **대수의 법칙**에 대해서 알 수 있는 실험과 동시에 **상대 도수적 확률**의 정의를 실험한 결과라는 것을 확인할 수 있다.

2 번 문제

x = 학번의 마지막 첫번째 자리 = 5

y = 학번의 마지막 두번째 자리 = 9

$P(D) = 0.01 = 1\% =$ 질병에 걸린 사건의 확률

$P(D^c) = 0.99 = 99\% =$ 질병에 걸리지 않은 사건의 확률

$P(T^+|D) = 0.95 = 95\% =$ 질병에 걸린 사람 중 양성 반응을 보인 사건의 확률

$P(T^-|D^c) = 0.99 = 99\% =$ 질병에 걸리지 않은 사람 중 음성을 보인 사건의 확률

여기에 구하려는 양성반응일 때 질병이 걸렸을 조건부 확률은 $P(D|T^+)$ 라고 표현 가능하다. 이 조건부 확률은 베이즈 정리를 이용하면 다시 아래처럼 표현할 수 있다.

$$P(D|T^+) = \frac{P(T^+|D)P(D)}{P(T^+|D)P(D) + P(T^+|D^c)P(D^c)}$$

T^+ 는 T^- 의 여사건이다. 따라서 아래와 같은 식이 성립한다.

$$P(T^+|D^c) = 1 - P(T^-|D^c) = 0.05 = 5\%$$

따라서 모든 변수를 베이즈 정리에 대입할 수 있다.

$$P(D|T^+) = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} = 0.16101694\cdots$$

테스트결과가 양성반응일 때 이 사람이 질병에 걸렸을 확률은 약 16%인걸 알 수 있다.

3번문제

확률이 실생활에 적용되는 예시는 나이팅게일 크림전쟁 사망통계와 연 스노의 콜레라 지도처럼 아주 유명한 일화나 포커와 블랙잭의 게임이론등 무궁무진하게 많다. 확률에 대한 일화중에 실생활에 접할 수 있는 스포츠는 다양한 관심 주제 중 하나일 것이다. 이번에 정한 주제는 야구의 세이버메트릭스이다.

세이버메트릭스는 사회과학의 게임 이론과 통계학적 방법론을 적용시켜서 선수의 실질적 가치 및 야구의 본질적 가치를 찾아내는 방법론이다. 이 이론을 최초로 생각한 사람은 F.C 레인으로 5 센트, 10 센트, 25 센트 50 센트를 같은 선상에 비교하는게 옳은 것인가라는 의문으로 시작하게 되었고 62 경기의 실황 중계를 바탕으로 하여 각 타율의 가치를 매겨서 단타는 0.475 점, 2 루타는 0.786 점, 3 루타는 1.15 점, 홈런은 1.55 점의 가치를 가진다는 것을 밝혀냈다. 이때 사용한 방법은 현대의 기댓값을 구하는 방법으로 구해낸 것인데 이로 인하여 그 당시에 널리 퍼져있는 절대 지표인 타율은 사실 별 가치가 없는 지표이며, 출루율과 장타율이 중요하다는 생각이졌다. 현대에서는 이 출루율과 장타율을 합쳐서 OPS 라고 부른다. 이론에만 그쳤던 세이버메트릭스를 실제로 도입한 팀은 1944 년 블루클린 다저스에서 통계학자까지 고용하였다. 그 이후 1950 년부터 야구분석학이 각광받기 시작했고 1954 년에는 득점기대치 이론이 나오면서 본격적으로 세이버메트릭스의 시대가 열렸다. 그 후 언쇼 쿡이 확률야구라는 논문까지 등장하며 현대까지 세이버메트릭스는 야구에 널리 사용되고 있다. 모든 게임에서 확률이 중요하게 현대 사용되고 있지만 이렇게 광범위하게 사용된 것은 야구가 최초라고 할 수 있을 것이다.

이 세이버메트릭스는 다른 스포츠 분야에도 영향을 미쳤다. 축구에서도 2003 년 옵타라는 사이트를 필두로 해서 전문적이고 세분화된 축구통계지표를 제공하고 있다. 과거에 없었던 CC(chance Created)나 CCC(Cleat-cut Chances)같은 지표들이 등장하며 과거에 골과 어시스트로 점철되었던 기록에 새로운 길을 제시하고 있다. 다른 스포츠 분야 역시 확률분석시도는 계속 되고 있다. 야구가 발전된 시기에 비하면 다른 스포츠가 발전이 느린편인데 이는 던제게임인 특성과 비교적 정적인 상황이 많은 야구와 다른 스포츠와의 차이에서 비롯되지만 빅데이터와 확률이 적용되면서 다른 분야에까지 장벽이 무너지고 있다.