

LGE - SNU AI Scientist 고급과정

확률통계 및 통계 방법론

홍현경(hyungyeong81@snu.ac.kr)

2026년 01월

Contents

1 확률과 확률분포	2
1.1 확률	2
1.2 확률변수와 확률분포	6
2 다양한 확률분포	8
2.1 확률분포의 예	8
2.2 다변량 확률변수와 확률분포	11
3 표본분포	14
3.1 표본분포의 성질	14
4 데이터 기반 추론 1	17
4.1 통계적 추정	17
4.2 통계적 추정 방법과 알고리즘	20
5 데이터 기반 의사결정	22
5.1 유의성 검정의 개념	22

1 확률과 확률분포

1.1 확률

예제 1. 한 모바일 앱의 하루 사용자 행동을 분석하는 상황을 고려하자. 임의로 선택한 한 사용자의 하루 동안의 행동에 대해 다음 두 사건을 정의하자.

- 사건 A: 해당 사용자가 앱 알림을 한 번 이상 클릭한 사건
- 사건 B: 해당 사용자가 앱 내에서 결제를 한 사건

분석 결과 다음과 같은 확률이 주어졌다고 하자.

$$P(A) = \frac{3}{5}, \quad P(B) = \frac{1}{2}, \quad P(A \cup B) = \frac{7}{10}$$

다음 물음에 답하여라.

- 해당 사용자가 알림을 클릭하고, 동시에 결제까지 했을 확률을 구하여라.
- 해당 사용자가 알림을 클릭한 사건과 결제를 한 사건이 서로 독립인지 판단하여라.

예제 2. 서버 3대(서버 1, 서버 2, 서버 3)로 구성된 클라우드 서버 시스템이 병렬로 연결되어 있다. 각 서버는 서로 독립적으로 작동하며, 하나의 서버가 고장날 확률은 0.02이다. 이 시스템은 세 대의 서버 중 적어도 두 대의 서버가 정상적으로 작동할 때 정상적으로 작동한다고 한다. 다음의 물음에 답하여라.

- (a) 시스템이 정상적으로 작동할 확률을 구하여라.
- (b) 시스템이 정상적으로 작동한다는 것이 관측되었을 때, 서버 1이 정상적으로 작동하고 있을 확률을 구하여라.

예제 3. 한 공정 라인에서는 동일한 양을 측정하는 두 개의 센서가 설치되어 있다. 두 센서는 정상 작동 시 측정값이 0부터 4 사이의 값 중에서 서로 독립적으로 균등하게 발생한다고 알려져 있다. 첫 번째 센서의 측정값을 확률변수 X , 두 번째 센서의 측정값을 확률변수 Y 라 하자. 품질 관리 기준에 따라, 두 측정값의 차이가 2보다 크면 측정 시스템에 이상이 있는 것으로 판단한다. 시스템에 이상이 있다고 판단될 확률을 구하여라.

예제 4. 한 프로야구 구장에서 자동 투구 판정 시스템(Automated Ball-Strike System: ABS)을 운영하고 있다. 이 시스템은 스트라이크 존에 들어오는 투구를 스트라이크(strike), 그 외의 투구를 볼(ball)로 자동으로 판별한다. 이 시스템에 대해 다음과 같은 정보가 주어졌다고 가정하자.

- 전체 투구 중 40%는 실제로 스트라이크이다.
- 실제 스트라이크 투구의 90%를 ABS가 스트라이크라고 올바르게 판정한다.
- 실제 볼 투구의 20%를 ABS가 스트라이크라고 잘못 판정한다.

이 때, 다음의 물음에 답하여라.

- (a) ABS가 투구를 스트라이크라고 판단할 확률을 구하여라.
- (b) ABS가 스트라이크로 판단한 투구가 실제로 스트라이크일 확률을 구하여라.

1.2 확률변수와 확률분포

예제 5. 1부터 5까지의 자연수로 이루어진 집합 $\{1, 2, 3, 4, 5\}$ 에서 서로 다른 두 원소를 임의로 선택한다. 다음 물음에 답하여라.

- (a) 선택된 두 수 중 큰 값을 확률변수 X 라고 할 때, $E(X)$ 를 구하여라.
- (b) 선택된 두 수 중 작은 값을 확률변수 Y 라고 할 때, $P(2 \leq Y \leq 4)$ 을 구하여라.

예제 6. 연속확률변수 X 의 확률밀도함수가 다음과 같을 때 물음에 답하여라.

$$f_X(x) = \frac{c(1-x^3)}{3} \quad (0 < x < 1)$$

- (a) 상수 c 의 값을 구하여라.
- (b) $P(\frac{1}{2} < X \leq 1)$ 을 구하여라.
- (c) $E(X)$, $\text{Var}(X)$ 을 각각 구하여라.

2 다양한 확률분포

2.1 확률분포의 예

예제 7. 한 음악 스트리밍 서비스에서는 사용자에게 노래를 추천하고, 사용자의 반응을 바탕으로 추천 시스템을 개선한다. 노래가 추천된 후 일정 시간 이내에 사용자가 아무런 긍정적인 행동(저장, 재생 지속 등)을 보이지 않으면, 해당 추천은 '관심 없음'으로 기록된다.

로그 분석 결과, 임의의 사용자에게 관심 없음 반응이 발생할 확률은 0.3이며, 개별 사용자들의 반응은 서로 독립이라고 가정한다. 임의로 선택한 사용자 4명에게 동일한 노래를 추천할 때, 다음 물음에 답하여라.

- (a) 한 사용자의 관심 없음 반응 여부를 나타내는 확률변수 X 를 다음과 같이 정의할 때, X 의 확률분포를 명시하여라.

$$X = \begin{cases} 1, & \text{관심 없음 반응이 발생한 경우} \\ 0, & \text{관심 없음 반응이 발생하지 않은 경우} \end{cases}$$

- (b) 4명의 사용자 중 각 사용자의 관심 없음 반응 여부를 나타내는 확률변수를 X_1, X_2, X_3, X_4 라고 하자. 관심 없음 반응이 나타난 사용자의 수를 확률변수 Y 라 할 때

- X_1, X_2, X_3, X_4 와 Y 사이의 관계식을 바탕으로 Y 의 확률 분포를 명시하고,
- $P(Y \geq 2)$ 를 계산하여라.

예제 8. 확률변수 X 의 확률밀도함수가 다음과 같이 주어져 있다.

$$f_X(x) = \begin{cases} cx^2(1-x), & 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

다음 물음에 답하여라.

- (a) $f_X(x)$ 가 베타분포의 확률밀도함수 형태임을 이용하여 X 의 분포를 $\text{Beta}(\alpha, \beta)$ 로 가정할 때, (α, β) 를 구하여라.

(참고: 베타분포의 확률밀도함수)

$$X \sim \text{Beta}(\alpha, \beta) \iff f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1$$

- (b) (a)에서 구한 분포의 형태를 이용하여 상수 c 의 값을 구하여라.

(참고: 감마함수 계산공식)

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \quad \Gamma(n) = (n-1)! \quad (n \text{은 자연수})$$

예제 9. 확률변수 X 가 평균이 10, 분산이 4인 정규분포를 따른다고 하자. 즉,

$$X \sim N(10, 4).$$

다음과 같이 정의된 확률변수 Y 를 고려하자.

$$Y = 3X - 5.$$

(a) 확률변수 Y 의 분포를 구하여라.

(b) 표준정규분포표를 이용하여 $P(19 \leq Y \leq 31)$ 을 계산하여라.

z	0.00	0.01	0.02	0.03
0.8	0.7881	0.7910	0.7939	0.7967
0.9	0.8159	0.8186	0.8212	0.8238
1.0	0.8413	0.8438	0.8461	0.8485

$P(Z \leq z)$ 표준정규분포표 일부

2.2 다변량 확률변수와 확률분포

예제 10. 한 온라인 플랫폼에서 사용자의 유료 구독 여부를 X , 모바일 앱 사용 여부를 Y 라고 하자.

$$X = \begin{cases} 1, & \text{유료 구독자} \\ 0, & \text{비구독자} \end{cases}, \quad Y = \begin{cases} 1, & \text{모바일 앱 사용} \\ 0, & \text{사용하지 않음} \end{cases}$$

조사를 통해 얻은 (X, Y) 의 결합확률질량함수는 다음과 같다.

	$Y = 0$	$Y = 1$
$X = 0$	0.30	0.10
$X = 1$	0.20	0.40

- 모바일 앱을 사용하는 사용자가 유료 구독자가 아닐 확률을 구하여라.
- $E(X)$, $E(Y)$, $\text{Cov}(X, Y)$ 를 구하여라.
- $U = 2X + 1$, $V = 3Y - 2$ 라 할 때, $\text{Cov}(U, V)$ 를 구하여라.

예제 11. 두 개의 센서 A, B가 동일한 물체의 온도($^{\circ}\text{C}$)를 측정한다고 하자. 센서 A의 측정값을 확률변수 X_1 , 센서 B의 측정값을 확률변수 X_2 라고 하자. $X_1 \sim N(25, 4^2)$, $X_2 \sim N(25, 3^2)$ 이고 X_1 과 X_2 는 서로 독립이라고 할 때, 두 센서의 측정값 차이가 4°C 보다 클 확률을 다음 표를 이용하여 구하여라. (단, $Z \sim N(0, 1)$)

z	$P(0 \leq Z \leq z)$
0.2	0.079
0.4	0.155
0.6	0.226
0.8	0.288
1.0	0.341
1.2	0.385

예제 12. KBO리그에서 투구 구속은 측정 방식에 따라 차이가 발생할 수 있다. 구속을 측정하기 위해서는 보통 **카메라 기반의 투구 추적 시스템(PTS)**과 **레이더 기반의 트랙맨**이 사용되며, 트랙맨은 도플러 레이더를 이용해 공의 궤적을 비행 전 구간에서 걸쳐 추적하여 투수 손을 떠난 지점에 가까운 구속을 측정한다. 반면, PTS는 여러 대의 카메라로 공의 이동 구간 일부를 추적하여 구속을 산출하므로, 트랙맨에 비해 상대적으로 낮은 구속이 기록되는 경우가 있다. 이러한 측정 결과는 경기 중 문자 중계에 표출되는 공식 구속 기록으로 사용된다.

한 투구의 실제 구속을 $v(\text{km/h})$ 라고 하자. 여기서 v 는 오차가 없는 값이라고 가정한다. 문자 중계에 표출되는 구속은 측정 과정에서 발생하는 차이를 포함하여 다음과 같이 나타난다고 가정한다.

$$\text{표출 구속} = v + X$$

여기서 X 는 실제 구속과 표출 구속 사이의 **속력 차이(편차)**를 나타내는 **확률변수**로, $X < 0$ 은 표출 구속이 실제 구속보다 느리게 기록된 경우를 의미한다.

구장 및 방송 환경의 차이로 인해, 문자 중계에 사용되는 구속 측정 방식은 다음 두 유형으로 구분된다고 가정한다. 한 경기에서는 하나의 측정 방식이 일관되게 사용되며, 경기 단위로 볼 때 **트랙맨 기준 방식이 사용될 확률은 0.7, 카메라 기준 방식이 사용될 확률은 0.3**이다.

- (유형 A: 카메라 기반 측정) 카메라 영상을 이용한 추정 방식으로, 표출 구속이 실제 구속보다 더 낮게 기록되는 경향이 있다. 이 경우 측정 오차 X 는

$$X \sim N(-4, 1^2).$$

- (유형 B: 트랙맨 기준 측정) 보정 절차가 적용된 측정 방식으로, 트랙맨을 이용한 측정 방식으로, 표출 구속과 실제 구속의 차이가 상대적으로 작다. 이 경우 측정 오차 X 는

$$X \sim N(-1, 1^2).$$

이 때, 문자 중계 구속이 실제보다 3km/h 이상 느리게 표출될 확률, 즉 $P(X \leq -3)$ 을 구하여라. (단, $\alpha = P(Z \leq z_\alpha)$, $Z \sim N(0, 1)$ 에 대해 $z_{0.999} = 3$, $z_{0.977} = 2$, $z_{0.933} = 1.5$, $z_{0.841} = 1$, $z_{0.691} = 0.5$ 임을 이용하여라.)

3 표본분포

3.1 표본분포의 성질

예제 13. 어떤 공장에서 생산되는 부품의 길이는 평균 50mm, 표준편차 0.05mm의 분포를 따르고, 각 부품의 생산은 서로 독립적으로 이루어진다고 한다. 생산된 제품 중 무작위로 100개의 부품을 추출했을 때, 추출된 부품의 평균 길이가 $[50\text{mm}-0.01\text{mm}, 50\text{mm}+0.01\text{mm}]$ 범위에 들어오면 해당 공정 상태를 ‘우수’로 평가한다. 공정 상태가 ‘우수’로 평가될 확률을 아래의 표를 참고하여 구하여라. (단, Z 는 표준정규분포를 따르는 확률변수이다.)

z	$P(0 \leq Z \leq z)$
1.0	0.341
1.5	0.433
2.0	0.477
2.5	0.494
3.0	0.499

예제 14. 한 카페의 에스프레소 한 샷 추출량(ml)을 확률변수 X 라 하자. 추출량은 평균 30ml, 표준편차 3ml인 분포를 따른다고 가정하며, 각 샷의 추출은 서로 독립적으로 이루어진다. 무작위로 100샷을 추출하여 추출량을 측정했을 때, 추출량의 표본평균 \bar{X} 가 $[30\text{ml} - 0.3\text{ml}, 30\text{ml} + 0.3\text{ml}]$ 범위에 들어오면 해당 날의 추출 상태를 정상으로 평가한다. 추출 상태가 정상으로 평가될 확률을 아래의 표준정규분포표를 참고하여 구하여라.

z	$P(0 \leq Z \leq z)$
1.0	0.341
1.5	0.433
2.0	0.477
2.5	0.494
3.0	0.499

예제 15. 음악 스트리밍 서비스에서 임의로 한 명의 사용자를 선택했을 때, 그 사용자가 시스템이 추천한 곡을 끝까지 재생하면 1, 중간에 종료하면 0으로 기록한다고 하자. 이를 확률변수 X 로 두면, 확률변수 X 는 다음과 같은 분포를 가진다.

$$P(X = 1) = 0.7, \quad P(X = 0) = 0.3$$

서로 다른 $n = 100$ 명의 사용자를 무작위로 선택하여 얻은 결과를 X_1, \dots, X_{100} 이라 하자. 다음 물음에 답하여라.

- (a) 새로운 확률변수 Y 를 추천한 곡을 끝까지 재생한 사용자의 수로 정의하려고 한다. 이 때, Y 를 X_1, \dots, X_{100} 을 이용하여 나타내고, 이를 바탕으로 Y 의 분포를 명시하여라.
- (b) 추천한 곡을 끝까지 재생한 사용자의 수가 75명 이상일 확률을 근사적으로 구하여라. (단, $Z \sim N(0, 1)$ 에서, $P(Z \leq 1.091) = 0.862$ 임을 이용하여라.)

4 데이터 기반 추론 1

4.1 통계적 추정

예제 16. 확률통계 및 통계방법론 과목을 수강한 학생들의 시험 성적을 분석하고자 한다. 이 시험은 20점 만점이며, 수강생 34명의 성적을 하나의 모집단으로 간주한다. 이 모집단의 모평균 성적은 17점이라고 알려져 있다. 이 모집단에서 비복원추출을 이용하여 크기가 5인 랜덤 표본을 선택할 때, 모평균에 대한 두 가지 추정량인 표본평균 \bar{X} 와 표본중앙값 X_{med} 의 확률분포가 다음과 같이 주어져 있다. 아래의 물음에 답하시오.

\bar{X}	15	16	17	18	19
$P(\bar{X})$	0.10	0.20	0.40	0.20	0.10

X_{med}	15	16	17	18	19
$P(X_{\text{med}})$	0.25	0.15	0.20	0.15	0.25

- 표본평균과 표본중앙값의 기댓값을 구하고, 두 추정량이 비편향추정량(편향 = 0)인지 여부를 확인하여라.
- 표본평균과 표본중앙값의 분산을 구하여라.
- 앞의 결과를 바탕으로, 표본평균과 표본중앙값 중 어떤 추정량이 더 좋은 추정량인지에 대해 논하여라.

예제 17. 어떤 전자부품의 수명(시간)은 정규분포를 따르며, 모표준편차는 $\sigma = 100$ 시간으로 알려져 있다고 하자. 이 전자부품의 평균 수명을 μ (시간)라 하자. 무작위로 $n = 30$ 개의 전자부품을 선택하여 수명을 측정한 결과, 수명에 대한 표본평균이 2000시간으로 나타났다. 다음 물음에 답하여라. (단, 표준정규분포를 따르는 확률변수 Z 에 대하여 $z_{1-\alpha}$ 는 $P(Z \geq z_{1-\alpha}) = \alpha$ 를 만족한다고 할 때, $z_{0.995} = 2.58$, $z_{0.99} = 2.33$, $z_{0.975} = 1.96$, $z_{0.95} = 1.65$ 임을 이용하시오.)

- (a) 모평균 μ 에 대한 95% 신뢰구간을 구하여라.
- (b) 모평균 μ 에 대한 99% 신뢰구간의 길이가 (a)에서 구한 95% 신뢰구간의 길이보다 작아지기 위해 필요한 표본의 크기 n 의 최소값을 구하여라.

예제 18. 한 식품회사는 새로 출시한 과자의 평균 중량(g) μ 가 규격을 만족하는지 확인하려고 한다. 생산 공정이 안정적이라 과자 중량은 정규분포를 따른다고 가정하며, 공정 데이터로부터 모표준편차는 $\sigma = 3(\text{g})$ 으로 알려져 있다. 회사는 μ 에 대한 95% 신뢰구간을 구성할 때, 신뢰구간의 오차한계를 0.8g 이하로 하고자 한다. 이를 위해서는 표본이 최소 몇 개 있어야 하는지 구하여라. (단, 표준정규분포를 따르는 확률변수 Z 에 대하여 $z_{1-\alpha}$ 는 $P(Z \geq z_{1-\alpha}) = \alpha$ 를 만족한다고 할 때, $z_{0.995} = 2.58$, $z_{0.99} = 2.33$, $z_{0.975} = 1.96$, $z_{0.95} = 1.65$ 임을 이용하시오.)

4.2 통계적 추정 방법과 알고리즘

예제 19. 어떤 온라인 쇼핑 플랫폼에서는 하나의 제품에 대해 사용자가 남기는 리뷰 별점 유형을 세 가지로 분류한다. 각 반응은 다음과 같이 숫자 1, 2, 3으로 기록된다.

- 1: 낮은 수준의 만족도
- 2: 보통 수준의 만족도
- 3: 높은 수준의 만족도

한 사용자가 남기는 리뷰 별점을 확률변수 X 로 나타낼 때, 이 반응은 다음과 같은 분포를 따른다고 알려져 있다.

$$P(X = 1) = \theta^2, \quad P(X = 2) = 2\theta(1 - \theta), \quad P(X = 3) = (1 - \theta)^2, \quad (0 < \theta < 1).$$

이제 서로 독립인 사용자의 리뷰 별점 X_1, \dots, X_n 을 관측하였다. 관측 결과, 리뷰 별점이 1 또는 2인 경우의 총 횟수가 m 번이었고, 반응 값이 3인 경우는 $n - m$ 번 관측되었다고 하자. 이때 모수 θ 의 최대가능도추정량 $\hat{\theta}_{MLE}$ 을 m 과 n 을 이용하여 구하여라(2차 미분은 보이지 않아도 무방하다).

Hint (미분 공식 및 근의 공식)

- (로그 미분 공식)
양의 함수 $f(\theta) > 0$ 에 대하여

$$\frac{d}{d\theta} \log f(\theta) = \frac{f'(\theta)}{f(\theta)}.$$

- (유리함수 미분 공식)
미분 가능한 함수 $u(\theta), v(\theta)$ 에 대하여

$$\frac{d}{d\theta} \left(\frac{u(\theta)}{v(\theta)} \right) = \frac{u'(\theta)v(\theta) - u(\theta)v'(\theta)}{v(\theta)^2}, \quad v(\theta) \neq 0.$$

- (근의공식)
이차방정식 $a\theta^2 + b\theta + c = 0$ ($a \neq 0$)의 해는 $\theta = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ 이다.

예제 20. 엘리베이터 두 대가 설치된 어떤 건물에서 연구자는 엘리베이터를 이용하는 사람들의 버튼 사용 습관을 관찰한다. 무작위로 서로 다른 n 명을 표본으로 선택하여, 각 사람이 엘리베이터를 기다릴 때 버튼을 양쪽 모두 누르는지 여부를 기록하였다. i 번째 사람의 행동을 다음과 같이 정의한다.

- $X_i = 1$: 버튼을 양쪽 모두 누름
- $X_i = 0$: 버튼을 한쪽만 누르거나 누르지 않음

각 사람은 서로 독립적으로 표본추출되었으며, 한 사람이 두 버튼을 모두 누르는 행동은 다른 사람의 행동에 직접적인 영향을 주지 않는다고 가정한다. 한 사람이 버튼을 양쪽 모두 누를 확률을 p 라고 하면,

$$P(X_i = 1) = p, \quad P(X_i = 0) = 1 - p, \quad (0 < p < 1),$$

즉, $X_i \sim \text{Bernoulli}(p)$ 이다. n 명의 관측 결과가 x_1, \dots, x_n 으로 주어졌을 때, 모수 p 의 최대가능도추정량 \hat{p}_{MLE} 을 구하여라.

Hint (미분 공식)

- (로그 미분 공식)
양의 함수 $f(\theta) > 0$ 에 대하여

$$\frac{d}{d\theta} \log f(\theta) = \frac{f'(\theta)}{f(\theta)}.$$

- (유리함수 미분 공식)
미분 가능한 함수 $u(\theta), v(\theta)$ 에 대하여

$$\frac{d}{d\theta} \left(\frac{u(\theta)}{v(\theta)} \right) = \frac{u'(\theta)v(\theta) - u(\theta)v'(\theta)}{v(\theta)^2}, \quad v(\theta) \neq 0.$$

5 데이터 기반 의사결정

5.1 유의성 검정의 개념

예제 21. 한 온라인 구독 서비스에서는 하루 동안 접속한 사용자들 중 서비스 이용을 중단하는 사용자 수를 기록한다. 각 사용자는 서로 독립적으로 행동하며, 정상적인 서비스 상태에서는 한 사용자가 서비스 이용을 중단할 확률이 $p = 0.03$ 으로 알려져 있다.

하루에 총 200명의 사용자를 무작위로 관측하여, i 번째 사용자가 서비스를 중단하면 $X_i = 1$, 그렇지 않으면 $X_i = 0$ 이라 하자. 각 X_i ($i = 1, \dots, 200$)는 서로 독립이라고 가정한다. 이제 다음 가설을 검정하고자 한다.

$$H_0 : p = 0.03 \quad \text{vs.} \quad H_1 : p > 0.03.$$

기각역을 자연수 M 에 대하여

$$X_1 + \dots + X_{200} \geq M$$

의 형태로 정하고, 제1종 오류를 범할 확률을 0.05 이하로 하고자 한다. 다음 이항분포 누적확률표를 이용하여 자연수 M 의 최소값을 구하여라.

k	9	10	11	12	13
$P(S \leq k), S \sim \text{Bin}(200, 0.03)$	0.919	0.960	0.982	0.992	0.997

예제 22. 연속형 확률변수 X 를 한 병에 담긴 소주의 실제 알코올 도수라고 하자. 해당 소주가 빨간뚜껑 소주인지, 혹은 초록뚜껑 소주인지를 판별하고자 한다. 제조 과정상의 변동과 측정 오차로 인해 알코올 도수는 정규분포를 따른다고 가정한다. 귀무가설과 대립가설을 다음과 같이 설정하였다.

$$H_0 : X \sim N(16, 4), \quad H_1 : X \sim N(20, 4).$$

측정된 도수가 18도를 초과하면 H_0 를 기각한다고 하자. 주어진 표준정규분포표를 이용하여 다음을 구하여라.

z	$P(0 \leq Z \leq z)$
1.0	0.341
1.5	0.433
2.0	0.477
2.5	0.494
3.0	0.499

(a) 제1종 오류를 범할 확률을 구하여라.

(b) 제2종 오류를 범할 확률을 구하여라.

예제 23. 한 커피 프랜차이즈의 라지 사이즈 커피에 대한 실제 평균 용량(ml)을 μ 라고 하자. 실제 커피의 평균 용량이 표기된 용량인 355ml와 일치하는지 확인하기 위해 아래의 가설을 검정하고자 한다.

$$H_0 : \mu = 355 \quad \text{vs} \quad H_1 : \mu \neq 355$$

해당 프랜차이즈의 라지 사이즈 커피 $n = 36$ 잔을 랜덤추출하여 용량을 측정한 결과, 표본평균과 표본표준편차가 다음과 같이 나타났다.

$$\bar{x} = 351, \quad s = 12$$

유의수준 5%에서 가설을 검정하고자 할 때, 아래의 물음에 답하시오.

- (a) 검정통계량과 임계치(critical value)를 이용하여 (즉, 기각역을 설정하여) 가설을 검정하여라.
- (b) 유의확률(p-value)를 이용하여 가설을 검정하여라.

(단, $P(T \geq t_{1-\alpha}) = \alpha$, $T \sim t(35)$ 에 대해 $t_{0.975} = 2.03$, $t_{0.95} = 1.69$ 이며, $P(T \leq 2) = 0.974$ 임을 이용하여라.)

예제 24. 한 F1 팀이 새 피트 전략 A 를 도입할지 판단하기 위해, 기존 전략 B 대비 랩타임 (초) 개선량을 시뮬레이션으로 평가한다고 하자. 독립적인 시뮬레이션을 $n = 100$ 번 수행하여, 각 반복에서 얻는 개선량을 X_1, \dots, X_n 이라 하자. X_i 는 서로 독립이고 정규분포를 따른다고 가정하며, 랩타임 개선량의 모분산은 알려져 있지 않다고 하자. 시뮬레이션 결과, 표본평균과 표본표준편차는 각각 $\bar{x} = 0.06$, $S = 0.3$ 로 관측되었다.

새 피트 전략이 통계적으로 유의하다고 판단되기 위해서는 평균 랩타임 개선량이 양수여야 한다. 따라서 다음 가설을 검정한다.

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu > 0$$

유의수준 5%에서 주어진 가설을 검정하고자 할 때, 아래의 물음에 답하시오.

(a) 검정통계량과 임계치(critical value)를 이용하여 (즉, 기각역을 설정하여) 가설을 검정하여라.

(b) 유의확률(p-value)를 이용하여 가설을 검정하여라.

(단, $P(T \geq t_{1-\alpha}) = \alpha$, $T \sim t(99)$ 에 대해 $t_{0.975} = 1.98$, $t_{0.95} = 1.66$ 임을 이용하여라.)