

## 11장 추정하기 (1부)

- 주의사항 :
- 11장과 12장 내용을 1, 2부로 나누어 다룬다.
  - 1부 주요 내용 : 평균의 표본분포와 신뢰구간
  - 2부 주요 내용 : 비율의 표본분포와 신뢰구간

### 1. 주요 내용

- ① 절추정 : 표본을 이용해서 모집단에 대한 정보를 추정하기

$$\left. \begin{array}{l} \text{표본 평균} \\ \text{표본 분산} \end{array} \right\} \text{계산} \xrightarrow{\text{절추정}} \text{추정} \left\{ \begin{array}{l} \text{모집단 평균} \\ \text{모집단 분산} \end{array} \right\}$$

- ② 비율의 표본분포 (2부에서 다룬다)

- ③ 평균의 표본분포

- ④ 중심극한 정리 (Central Limit Theorem)

### 2. 주요 예제

모집단 = "달콤한 풍선껌" 회사에서 생산하는 "한 번에 씹는 한 끗" 풍선껌 전체

표본 = 원활되지 않은 표본이 구성되어 있다고 가정

예습문제 = "달콤한 풍선껌" 모집단 전체에 대한 정보 주하기

2부에서  
다룬다

- ① 양이 지속되는 시간의 평균과 분산
- ② 타 회사 제품과 "달콤한 회사" 제품을 선호하는 사람들의 비율

### 3. 모집단의 평균 및 분산 추정하기

p. 483

① 절추점 = 표본을 통해 구한 정보(평균, 분산 등)를 이용하여 모집단에 대한 정보를 구하는 방법

⇒ 절추점을 이용하여 구한 모집단에 대한 정보(평균, 분산 등)가 정확하다고 장담할 수는 없다. 하지만 그렇게 하는 것이 최선이다.

\* 주의: 편향되지 않은 표본이 구성되도록 치선을 다해야 한다.

#### ② 모집단 평균값의 절추점

p. 484~485

i) 표본의 평균값  $\bar{x}$  를 구한다:

$$\bar{x} = \frac{\sum x}{n}$$

표본에 속한 모든 데이터 더하기  
표본의 크기

ii) 모집단 평균값의 추정치를 설정한다:

$$\hat{\mu} = \bar{x}$$

즉, 표본의 평균을 모집단 평균의 추정치로 사용.

\* 주의: 모집단의 진짜 평균은 서로 표시된다.

하지만 서는 일반적으로 거의 알 수가 없으며 평균값의 추정치인  $\hat{\mu}$  를 서 대신에 사용한다.

⇒ 보다 정확한  $\hat{\mu}$  를 얻기 위해 편향되지 않은면서 "적합한" 크기의 표본을 구성하는 일이 매우 중요하다.

### ③ 모집단 분산의 절추정

P. 488  
~ 490

평균값의 절추정과는 달리 모집단 분산의 절추정은

표본의 분산과 다르게 계산한다.

$$S^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

모집단 분산의  
절추정 값으로  
사용하는 경우
표본의 크기
n 대신에  
n-1 사용!

표본의  
평균값

↑ 이유: 모집단의 분산은 표본의 분산에  
비해 좀 더 큰 값을 가짐.

⇒ 모집단 분산의 절추정:  $\hat{\sigma}^2 = S^2$

예제: “달콤한 풍선껌” 회사의 제품의 표본에  
속한 개체의 끊임없는 차수(개)의 차이가 같다.  
(현제 10)

61.9	62.6	63.3	64.8	65.1
66.4	67.1	67.2	68.7	69.9

$$\Rightarrow \hat{\mu} = \bar{x} = \frac{61.9 + 62.6 + \dots + 69.9}{10}$$

$$= 65.7$$

$$\hat{\sigma}^2 = S^2 = \frac{(61.9 - 65.7)^2 + \dots + (69.9 - 65.7)^2}{9}$$

$$= 6.92$$

## 가 제

## 1) 과제 소개

11장 ~ 12장 내용을 이해하기 위해 다음의 모의실험을 각자  
집에서 실행할 것.

직접 모의실험을 해 본 경우 강의 내용을 브라 청기 이해 가능.

## 2) 모의 실험 내용

- (i) "한 개의 주사위를 6번 던지기"를 30회 반복하기
- (ii) 주사위를 6번 휘젓을 때마다 숫자 1이 나온 횟수를  
아래 표에 작성하기

1회	2회	3회	...	29회	30회	평균

각각의 번에는 주사위를  
6번 던졌을 때마다  
나온 횟수 입력

숫자 1이 나온  
횟수의 평균 입력

## 3) 질문

- (i) 본인이 주한 평균값이 적절하다고 생각되는가?
- (ii) 질문 (i)에 대한 본인의 답에 대한 근거를 설명하라.

(주제내용)

#### 4) 힌트

(i) 주제 주사를 6번 했을 때 숫자 1이 나오는 횟수  $X$ 는 이항분포를 따른다.

$$X \sim B(6, \frac{1}{6})$$

주사의 6번  
인자기

주사의 6번  
인자기  
인자를 때 숫자 1이  
나올 확률

(ii) "주사의 6번 인자기"를 30회 반복하여 숫자 1이 나온 횟수의 평균  $\bar{X}$ 는 아래 공식을 만족 시킨다.

$$\bar{X} = \frac{x_1 + \dots + x_{30}}{30}$$

단,  $x_i$ 는  $X$ 의 독립간측이다.

↑ "주사의 6번 인자기" 반복하는

행위는 서로 독립이며

동일한 확률분포를 따른다.

(iii) 이제  $\bar{X}$ 의 기대치, 즉 평균과 표준을 구할 수 있다.

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{x_1 + \dots + x_{30}}{30}\right) \\ &= \frac{1}{30} \cdot E(x_1 + \dots + x_{30}) \\ &= \frac{1}{30} \cdot 30 \cdot E(X) \\ &= E(X) \end{aligned}$$

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{x_1 + \dots + x_{30}}{30}\right) \\ &= \frac{1}{30^2} \text{Var}(x_1 + \dots + x_{30}) \\ &= \frac{1}{30^2} \cdot 30 \cdot \text{Var}(X) \\ &= \frac{\text{Var}(X)}{30} \end{aligned}$$

(추가내용)

(iv) 또한  $\bar{X}$ 는 정규분포를 따른다. 즉,

$$\bar{X} \sim N(1, \frac{1}{6^2})$$

중심극한정리에 의해!

이유:  $X \sim B(6, \frac{1}{6})$

$$\Rightarrow E(X) = 6 \cdot \frac{1}{6} = 1$$

$$Var(X) = 6 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{6}$$

(v) 결론

각자의 실현결과는 아래 구간에 속할 것이다.

①  $[1 - \frac{1}{6}, 1 + \frac{1}{6}]$  구간에 속할 확률 68.2%

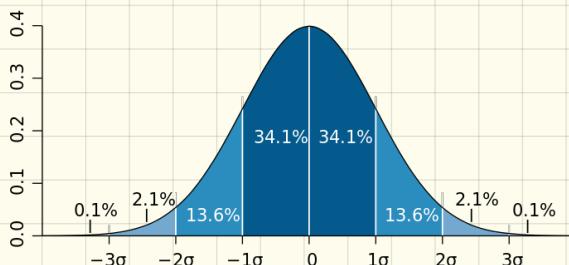
②  $[1 - \frac{1}{3}, 1 + \frac{1}{3}]$  구간에 속할 확률 95.4%

③  $[1 - \frac{1}{2}, 1 + \frac{1}{2}]$  구간에 속할 확률 99.7%

5) 추가 질문

(v)의 결론을 어떻게 구했을까?

힌트: 아래 그림 참조



## 4. 표본분포

① 통계량: 표본으로부터 얻은 표본평균, 표본총산, 표본비율과 같은 통계적인 양

② 표본분포: 조사단에서 일정한 크기의 표본을 반복적으로 선정하여 얻은 통계량의 확률분포

### 예) 평균의 표본분포

- 비율의 표본분포 (그림에서 다룬다)
- 분산의 표본분포 (여기서는 다루지 않음)

## 5. 평균의 표본분포 예제

p.510

### 예제

- 풍선껌을 봉지에 담아 판매
- 한 봉지에 담긴 풍선껌 개수의 평균은 10, 분산은 1인.
- 한 상자에 30봉지씩 넣어 판매함.

질문: ① 한 상자에 들어 있는 풍선껌 한 봉지에 담긴

풍선껌 개수의 평균은 얼마인가?

② 한 상자에 들어 있는 풍선껌 한 봉지에 담긴

풍선껌 개수의 평균이 8.5개 이하일 확률은?

예제의 질문에 대하기 위해서는 아래의 과정을 진행해야 한다.

- ① 30개의 물자를 무작위로 선택하여 하나의 물자에 들어 있는 풍선개의 개수의 평균을 구해보아야 한다.
- ② 과정①을 여러 번 반복하여 풍선개 개수의 평균의 확률을 알아낸다.

일반적으로 30회 이상.

**주의사항**은 30개의 물자의 선택은 매번 독립적으로 이루어져야 한다. 즉, 한 물자의 선택은 무작위적으로 이루어지며 한 번 선택된 물자를 다시 원래대로 들여 넣는 **복원 추출** 방식을 따른다.

예제 답안:

$X$ 를 한 물자에 들어 있는 풍선개 개수의 확률로 나타낸다고 하자. 그러면 다음 사실이 성립한다.

$$\begin{aligned} \textcircled{1} \quad E(X) &= 10 \\ \textcircled{2} \quad \text{Var}(X) &= 1 \end{aligned}$$

\Leftrightarrow \left\{ \begin{array}{l} \text{하나의 물자에 들어 있는 풍선개의 개수가} \\ \text{대한 기대치 (평균값)와 분산} \end{array} \right.

$\textcircled{3}$  물자 30개 각각의 선택은 **독립관측**이다.

따라서 30물자에 들어 있는 풍선개 개수의 평균값의 분포를  $\bar{X}$ 라 하면 아래 공식이 성립한다.

$$\textcircled{1} \quad \bar{X} = \frac{X_1 + \dots + X_{30}}{30}$$

$\textcircled{2} \quad X_i$ 는  $X$ 의 독립관측

$X_1 + X_2$ 의 확률변수이므로  $\bar{X}$ 의 평균값과 표준을 다음과 같이 구할 수 있다.

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{X_1 + \dots + X_{30}}{30}\right) \\ &= E\left(\frac{1}{30}X_1 + \dots + \frac{1}{30}X_{30}\right) \\ &= \frac{1}{30}E(X_1) + \dots + \frac{1}{30}E(X_{30}) \\ &= \frac{1}{30}(E(X_1) + \dots + E(X_{30})) \\ &= \frac{1}{30} \cdot 30 \cdot E(X) \\ &= E(X) = 10 \end{aligned}$$

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{X_1 + \dots + X_{30}}{30}\right) \\ &= \text{Var}\left(\frac{1}{30}X_1 + \dots + \frac{1}{30}X_{30}\right) \\ &= \frac{1}{30^2} \text{Var}(X_1) + \dots + \frac{1}{30^2} \text{Var}(X_{30}) \\ &= \frac{1}{30^2} \cdot 30 \cdot \text{Var}(X) \\ &= \frac{\text{Var}(X)}{30} = \frac{1}{30} \end{aligned}$$

생각적으로 이해 가능.  
분산의 경우: 한 항에 포함된 평균값의 개수가 늘어날 수록 표표가 비례하여 증식됨

여기  $P(\bar{X} < 8.5)$ 를 구하고자 한다.  
그런데  $\bar{X}$ 는 어떤 확률률을 따르는가?

$\Rightarrow$  정규분포를 따른다.

즉,  $\bar{X} \sim N(10, \frac{1}{30})$  이 성립한다.

여기 정규분포정리

에서 설명!

$$\begin{aligned} \text{따라서 } P(\bar{X} < 8.5) &= P\left(Z < \frac{8.5 - 10}{\sqrt{0.0333}}\right) \\ &= P(Z < -1.82) \\ &= 0 \end{aligned}$$

학점 허이블에서 확인 불가.  
사실상 0인줄 의미함.

(주의: 연속성분정 필요없음)

↑ 평균값들은 어떤 값으로 인의도 가질 수 있기 때문.

## 6. 평균의 표본분포 (일반화)

p.512 ~ 519  $E(X) = \mu$ ,  $\text{Var}(X) = \sigma^2$  을 따르는  $X$ 의 분포가 주어짐.

$n$  개의  $X$ 를 독립관측으로任意로 선택했을 때  
 $X$ 의 평균값을  $\bar{X}$ 라고 하자. 즉,

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

( $X_i$ 는  $X$ 의 독립관측)

그러면  $\bar{X}$ 의 평균값과 표준은 다음과 같다.

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{X_1 + \dots + X_n}{n}\right) \\ &= \frac{1}{n} \cdot E(X_1 + \dots + X_n) = \frac{1}{n} \cdot n \cdot E(X) \\ &= E(X) = \mu \end{aligned}$$

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) \\ &= \frac{1}{n^2} \cdot \text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2} \cdot n \cdot \text{Var}(X) \\ &= \frac{1}{n} \cdot \text{Var}(X) = \frac{\sigma^2}{n} \end{aligned}$$

~한번抽取로  $n$ 이 커질수록 표준오차는 줄어든다.

p.520 ~ 521

또한  $n \geq 30$  일때  $\bar{X}$ 는 중심극한정리에 의해 정규분포를 따른다. 즉,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$(\text{단}, E(X) = \mu, \text{Var}(X) = \sigma^2)$$

주의 :  $X$ 가 반드시 정규분포를 따를 필요는 없다.

## 7. 중심극한정리

- ① 서로 독립이며 동일한 분포를 따르는  $n$  개의 확률변수  $X_1, X_2, \dots, X_n$ 에 대해  $n$ 이 충분히 크면, 예를 들어  $n \geq 30$ , 아래와 같이 정의된

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

$\bar{X}$ 의 분포가 개략적으로 정규분포를 따른다.

또한,  $E(\bar{X}) = \mu$ 이고  $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$ 이고  $X_i$ 가  $X$ 와 동일한 확률분포를 가지면 다음과 성립한다.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (\text{만}, n \geq 30)$$

### ② 확률 찾기

$\bar{X}$ 가 정규분포를 따르므로 확률테이블을 이용하여 확률값을 계산할 수 있다.

$$(i) P(\bar{X} > a) = P\left(z > \frac{a - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$(ii) P(a < \bar{X} < b) = P\left(\frac{a - \mu}{\frac{\sigma}{\sqrt{n}}} < z < \frac{b - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$(iii) P(a < \bar{X}) = P\left(z < \frac{a - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

(주의: 연속성로정 필요 없음)

↑ 평균값들은 어떤가 그대로 인의로 칠 수 있기 때문.

## d. 중성주한 정리 활용

### 1) 이항분포

전제:

$$X \sim B(n, p)$$

$$\Rightarrow \mu = n \cdot p, \quad \sigma^2 = n \cdot p \cdot q \quad (q = 1 - p)$$

(n ≥ 30)

이항분포를 따르는  $X$ 를 득점 만족으로 개별 반복할 때  
중성주한 정리에 의해  $X$ 의 평균값  $\bar{X}$ 는 정규분포를  
따른다. 즉,

$$\bar{X} \sim N(n \cdot p, \frac{n \cdot p \cdot q}{n})$$

예제) 동전을 10번 던졌을 때 숫자가 나오는 횟수는  
이항분포를 따른다. 즉,  $X \sim B(10, \frac{1}{2})$ .

이제, "동전 10번 던지기"를 30번 반복했을 때

숫자가 나오는 횟수의 평균은 다음의 봉포를 따른다.

$$\bar{X} \sim N(5, \frac{1}{12})$$

예제) "동전 10번 던지기"를 30번 반복했을 때,

숫자가 평균 6번 이상 나올 확률은?

정답:  $\bar{X} \sim N(5, \frac{1}{12})$  이므로

$$\begin{aligned}
 P(\bar{X} > 6) &= P(Z > \frac{6-5}{\sqrt{\frac{1}{12}}}) \\
 &= P(Z > \sqrt{12}) = P(Z > 3.46) \\
 &= 1 - P(Z \leq 3.46) \\
 &= 1 - 0.898730 \\
 &\approx 0.000270
 \end{aligned}$$

↑ 매우 낮음.

주의: 책 설명과 같은 다른  
(책 설명을 일반화 시킨 내용인)

주의:  
 1)  $X$  자체는  
 정규분포를 따르지  
 않습니다.

## \* 주의사항 (책에 있는 내용임)

"동전 10번 헌지기"를 30회 반복해서 얻은 통계량과 동전을 300번 던져서 얻는 통계량은 어떻게 다를까?  
즉, 아래 두 질문의 결과를 비교해 보자.

- ① "동전 10번 헌지기"를 30회 반복했을 때, 숫자가 평균 6번 이상 나올 확률은?
- ② 동전을 300번 던져서 숫자가 180번 이상 나올 확률은?

$$\Rightarrow \textcircled{1} \text{의 답은 } 0.00027 \quad (\text{앞서 구한값})$$

\textcircled{2}의 답은 아래와 같이 구할 수 있다.

$X$ 가 동전을 300번 던져서 숫자가 나오는 횟수를 가지면

$$X \sim N(300, \frac{1}{2}) \text{이다.}$$

그리고  $300 \times \frac{1}{2} = 150 > 5$  이므로  $X$ 는 정규 분포를 따른다. 즉,

$$X \sim N(150, 75).$$

따라서

$$\begin{aligned} P(X > 180) &= P(Z > \frac{180.5 - 150}{\sqrt{75}}) \\ &= P(Z > 3.52) \\ &= 1 - P(Z < 3.52) \\ &= 1 - 0.999784 \\ &\approx 0.000216 \end{aligned}$$

↑  
 $0.00027$  과 비교하여  
오차는  $0.000054$ , 즉,  
 $0.0054\%$  일. 확률로는  
어떤 의미도 없는 오차임.

## 2) 평균 분포 (평균 분포)

P.522

$$\text{전제: } X \sim P_6(\lambda)$$

$$\Rightarrow \mu = \lambda, \sigma^2 = \lambda$$

평균 분포를 따르는  $X$ 를 높임 간격으로  $n$ 번 반복할 때  
증식 주변 정의에 의해  $X$ 의 평균값  $\bar{X}$ 는 정규분포를  
따른다. 즉,

$$\bar{X} \sim N(\lambda, \frac{\lambda}{n})$$

예제) 영학관의 평균 기계의 주당 평균 고장 횟수는  
평균 분포를 따르며  $X \sim P_6(3.4)$  이다.

이제, 30주 동안 주당 고장 횟수의 평균은 다음과  
정규분포를 따른다:

$$\bar{X} \sim N(3.4, \underbrace{0.11}_{\frac{3.4}{30}})$$

예제) 임의로 지정된 30주 동안 주당 고장 횟수의 평균이  
2회 이하일 확률은?

정답:  $\bar{X} \sim N(3.4, 0.11)$  이므로

$$P(\bar{X} < 2) = P(Z < \frac{2 - 3.4}{\sqrt{0.113}})$$

$$= P(Z < \frac{-1.4}{0.337})$$

$$= P(Z < -4.15)$$

$$= 0$$

↑ 거의 0임.

주의:  
 1)  $X$  자체는  
 정규분포를 따르지  
 않는다.