확률 및 통계학

- 8장 확률표본과 표본분포 -

손 우 영(wooyoung@pel.sejong.ac.kr)

세종대학교 프로토콜공학연구실

목 차

목 차

확률표본 (1/2)

- 모집단(Population)
  - 정의
    - 관심 있는 대상과 관련된 모든 관측 가능한 값의 집합
  - 특징
    - 모집단을 구성하는 관측값 전체의 집합을 관측하는 것은 불가능함
- 표본(Sample)
  - 정의
    - 모집단의 부분집합
  - 특징
    - 편향(Bias)되지 않고 모집단의 성질을 잘 반영하도록 추출되어야 함

확률표본 (2/2)

- 확률표본(Random Sample)
  - 정의
    - 다수의 확률변수가 상호독립적이고, 모두 동일한 확률분포를 따른다는 조건에서 추출한 표본
  - 특징
    - 확률표본을 구성하는 각 개체들은 상호 독립적이고, 동일한 모집단에서 추출되었으므로 동일한 분포를 가짐
    - 서로 독립인 n개의 확률변수 $X_1, X_2, \cdots, X_n$이 동일한 확률분포 $f(x)$를 따를 때, $X_1, X_2, \cdots, X_n$을 모집단 $f(x)$로부터 크기가 n인 확률표본이라고 함
      - 결합확률분포는 $f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$으로 표현됨

목 차

대표적 통계량 (1/7)

- 통계량(Statistic)
  - 정의
    - 확률표본을 구성하는 확률변수들의 함수
      - e.g., 표본평균, 표본분산
  - 필요성
    - 표본 데이터를 기반으로 모집단의 특성(e.g., 평균, 분산 등)을 추정할 수 있음
  - 특징
    - 통계량은 특정 확률표본의 데이터에 기반함에 따라 다른 표본을 사용하면 값이 달라짐

대표적 통계량 (2/7)

- 중심 통계량 (1/2)
  - $X_1, X_2, \cdots, X_n$을 크기가 n인 확률표본이라고 할 때,
  - 표본 평균(Sample Mean)
    - 정의
      - 표본을 구성하는 확률변수의 평균값
    - 공식
      - $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - 표본중앙값(Sample Median)
    - 정의
      - 주어진 표본 데이터를 크기 순으로 나열했을 때 중앙에 위치하는 값
    - 공식
      - $\tilde{x} = \begin{cases} x_{(n+1)/2}, & n\text{이 홀수일 때} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right), & n\text{이 짝수일 때} \end{cases}$

대표적 통계량 (3/7)

- 중심 통계량 (2/2)
  - $X_1, X_2, \cdots, X_n$을 크기가 n인 확률표본이라고 할 때,
  - 표본최빈값(Sample Mode)
    - 정의
      - 주어진 표본 데이터 집합에서 가장 많이 발생한 표본값

- 예제 8.1

관측값들이 다음과 같다고 할 때, 가장 많이 관측되는 0.43이 표본최빈값임

0.32, 0.53, 0.28, 0.37, 0.47, **0.43**, 0.36, 0.42, 0.38, **0.43**

대표적 통계량 (4/7)

- 산포 통계량 (1/3)
  - $X_1, X_2, \cdots, X_n$을 크기가 n인 확률표본이라고 할 때,
  - 표본분산(Sample Variance)
    - 정의
      - 표본 내 데이터 값들이 표본평균으로부터 얼마나 퍼져 있는지를 측정하는 통계량
    - 공식
      - $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

대표적 통계량 (5/7)

- 산포 통계량 (2/3)
  - $X_1, X_2, \cdots, X_n$을 크기가 n인 확률표본이라고 할 때,
  - 표본분산(Sample Variance)
    - 특징
      - S^2이 크기가 n인 확률표본의 분산이라면, $S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} X_i{}^2 - (\sum_{i=1}^{n} X_i)^2\right]$으로 표현할 수 있음

증명

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i{}^2 - 2\bar{X}X_i + \bar{X}^2)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i{}^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2\right]$$

$\bar{X}$를 $\sum_{i=1}^{n} X_i / n$으로 치환하고 분모, 분자에 n을 곱하면, 다음과 같은 식을 얻음

$$S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} X_i{}^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]$$

대표적 통계량 (6/7)

- 산포 통계량 (3/3)
  - $X_1, X_2, \cdots, X_n$을 크기가 n인 확률표본이라고 할 때,
  - 표본표준편차(Sample Standard Deviation)
    - 정의
      - 표본 데이터 내 개별 데이터 값들이 표본 평균으로부터 얼마나 떨어져 있는지를 나타내는 측정값으로, 표분 분산의 제곱근으로 계산됨
    - 공식
      - $S = \sqrt{S^2}$
  - 표본범위(Sample Range)
    - 정의
      - 주어진 표본 데이터 집합 내의 가장 큰 값과 가장 작은 값의 차이
    - 공식
      - $R = X_{max} - X_{min}$

대표적 통계량 (7/7)

- 예제 8.2

샌디에이고에 있는 상점 중 임의로 4곳을 선택하여 200g짜리 병커피의 가격을 비교해 본 결과 지난 달보다 12, 15, 17, 20 센트씩 인상되었다. 가격인상에 대한 이 <u>**확률표본의 분산**</u>을 구하라

- 표본평균 계산
  - $\bar{X} = \frac{12+15+17+20}{4} = 16$센트
- 표본분산 계산
  - $s^2 = \frac{1}{3}\sum_{i=1}^{4}(x_i - 16)^2 = \frac{(12-16)^2+(15-16)^2+(17-16)^2+(20-16)^2}{3} = \frac{34}{3}$

- 예제 8.3

1996년 6월 19일 무스코카 호수에서 임의로 선정된 6명의 낚시꾼들에 의해 잡힌 송어의 수는 3, 4, 5, 6, 6, 7마리였다. <u>**표본분산**</u>을 구하라

- $\sum_{i=1}^{6} {x_i}^2 = 171,\ \sum_{i=1}^{6} x_i = 31, n = 6$
- $S^2 = \frac{1}{(6)(5)}[(6)(171) - (31)^2] = \frac{13}{6}$

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- 표본평균의 분포
- 표본분산의 분포
- t 분포
- F 분포
- 분위수와 확률 그림
- 부록

표본분포

- 표본분포(Sampling Distribution)
  - 정의
    - 통계량의 확률분포
      - 모집단의 통계량은 ‘상수’로 취급되나, 표본 통계량은 ‘확률변수’로 취급됨
  - 필요성
    - 모집단 전체를 조사할 수 없을 때, 선택된 표본의 통계량으로 모집단의 특성을 추정할 수 있음
      - e.g., 표본평균의 분포, 표본분산의 분포

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- 표본평균의 분포
- 표본분산의 분포
- t 분포
- F 분포
- 분위수와 확률 그림
- 부록

표본평균의 분포 (1/9)

- 표본평균의 분포 (1/2)

정리 7.11

확률변수 $X_1, X_2, \cdots, X_n$이 서로 독립이면서 각각 평균이 $\mu_1, \mu_2, \cdots, \mu_n$이고 분산이 $\sigma^2{}_1, \sigma^2{}_2, \cdots, \sigma^2{}_n$인 정규분포를 따르면, $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$에 대해

- $\mu_Y = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$
- $\sigma^2{}_Y = a^2{}_1\sigma^2{}_1 + a^2{}_2\sigma^2{}_2 + \cdots + a^2{}_n\sigma^2{}_n$

확률표본 추출의 특징

평균이 μ, 분산이 σ^2인 정규모집단으로부터 크기가 n인 확률표본 추출 시, 확률표본 X_i $(i = 1, 2, \cdots, n)$은 모두 모집단과 동일한 정규분포를 따름

- 표본평균 ($\bar{X}$)
  - 표현
    - $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$
  - $\bar{X}$의 평균
    - $\mu_{\bar{X}} = \frac{1}{n}(\mu + \mu + \cdots + \mu) = \frac{1}{n}\times\mu\times n = \mu$

- 표본평균의 분포 (2/2)

정리 7.11

확률변수 $X_1, X_2, \cdots, X_n$이 서로 독립이면서 각각 평균이 $\mu_1, \mu_2, \cdots, \mu_n$이고 분산이 ${\sigma^2}_1, {\sigma^2}_2, \cdots, {\sigma^2}_n$인 정규분포를 따르면, $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$에 대해

- $\mu_Y = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$
- ${\sigma^2}_Y = {a^2}_1{\sigma^2}_1 + {a^2}_2{\sigma^2}_2 + \cdots + {a^2}_n{\sigma^2}_n$

확률표본 추출의 특징

평균이 μ, 분산이 σ^2인 정규모집단으로부터 크기가 n인 확률표본 추출 시, 확률표본 X_i $(i = 1, 2, \cdots, n)$은 모두 모집단과 동일한 정규분포를 따름

- 표본평균 ($\bar{X}$)
  - 표현
    - $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$
  - $\bar{X}$의 분산
    - ${\sigma^2}_{\bar{X}} = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{1}{n^2} \times \sigma^2 \times n = \frac{\sigma^2}{n}$

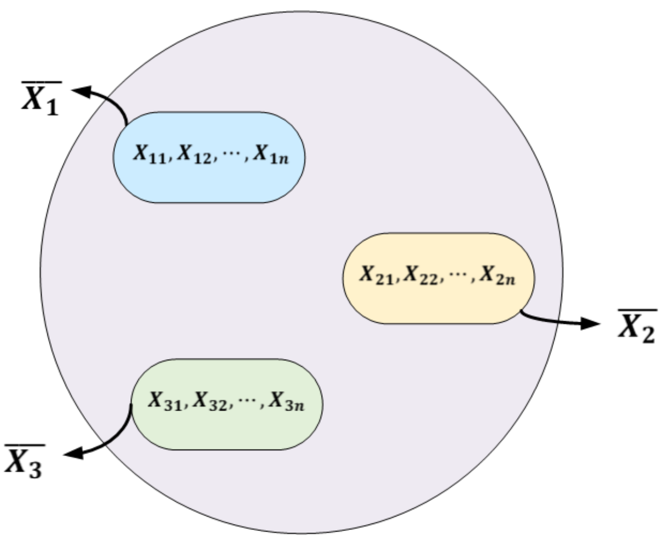
표본평균의 분포 (3/9)

- 중심극한정리(CLT, Central Limit Theorem) (1/2)
  - 정의
    - 독립적이고 동일하게 분포된 모집단에서 추출한 변수의 수가 충분히 크면, 추출한 표본의 평균 $\bar{X}$는 정규분포에 가까워짐
      - 평균이 μ, 분산이 σ^2인 모집단으로부터 크기가 n인 확률표본을 추출하면, $\bar{X}$에 대하여 $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$은 n이 충분히 클 때 표준정규분포 $n(z; 0,1)$에 근접함
  - 특징
    - 일반적으로 n이 30 이상이면 모집단의 분포에 관계 없이 $\bar{X}$에 대한 정규분포 적용 가능

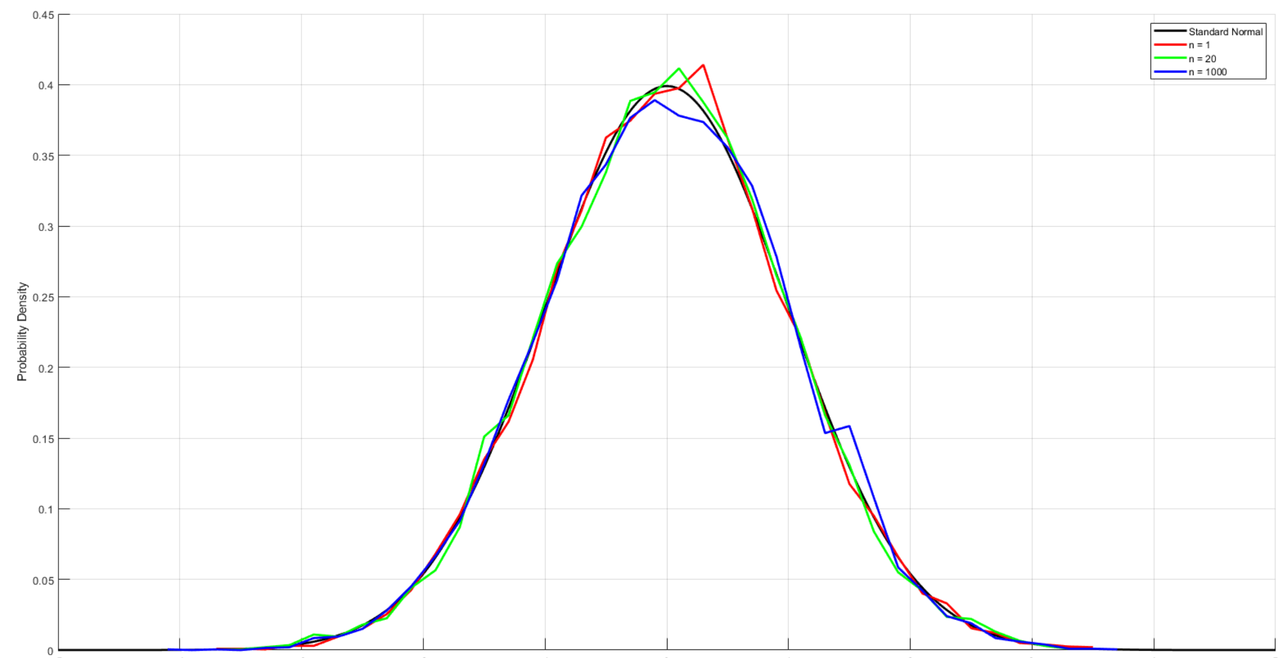
표본평균의 분포 (4/9)

- 중심극한정리(CLT, Central Limit Theorem) (2/2)
  - 일반적으로 n이 30 이상이면 모집단의 분포에 관계 없이 $\bar{X}$에 대한 정규분포 적용 가능
    - n은 표본평균의 개수를 말하는 것이 아닌, 표본평균을 도출하기 위해 추출하는 데이터의 수

모집단



<그림 8.1 동일하게 분포된 모집단에서 추출한 표본>



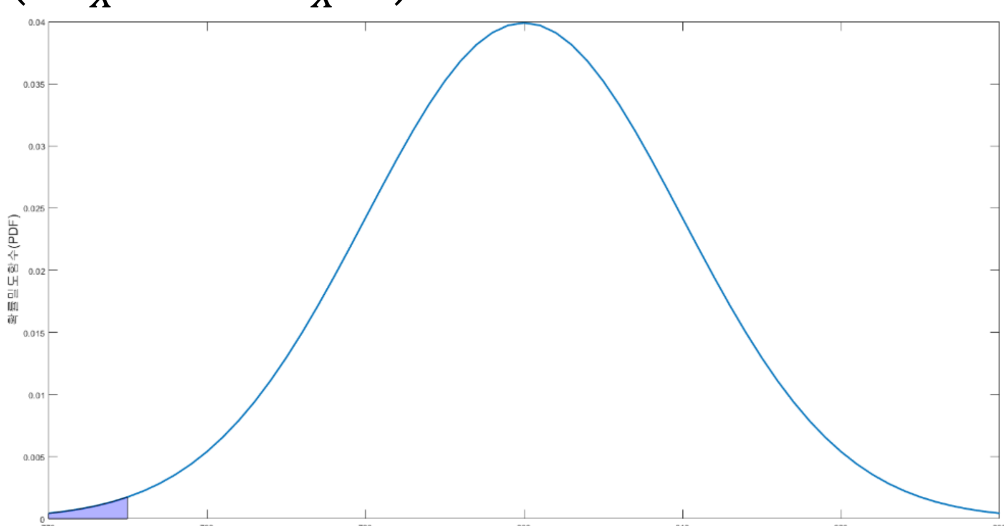
<그림 8.2 n=1, n=20, n=100일 때 표본평균의 확률분포 및 표준정규분포>

표본평균의 분포 (5/9)

- 예제 8.4

어떤 전구생산공장에서 생산되는 전구의 수명은 **평균이 800시간이고, 표준편차가 40시간인 정규분포를 따른다**고 알려져 있다. **임의로 추출된 16개의 전구의 평균 수명이 775시간 미만일 확률**을 구하라

- $\bar{X}$의 분포: $\mu_{\bar{X}} = 800, \sigma_{\bar{X}} = \frac{40}{\sqrt{16}} = 10$인 정규분포를 따름
- $P(\bar{X} < 775) = P\left(\frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{775-\mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) = P(Z < -2.5) = 0.0062$



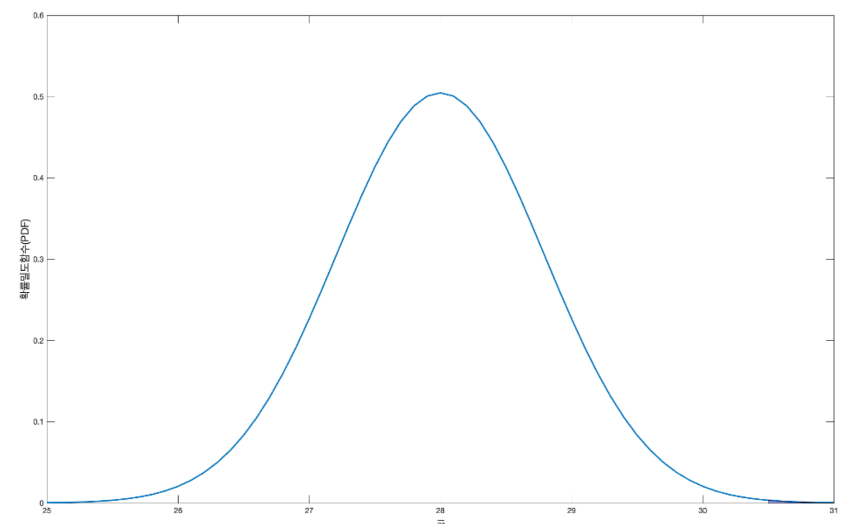
<그림 8.3 $\mu_{\bar{X}} = 800, \sigma_{\bar{X}} = 10$인 정규분포>

표본평균의 분포 (6/9)

- 예제 8.5

어느 대학의 두 캠퍼스 사이를 운행하는 셔틀버스의 운행시간은 **평균 28분, 표준편차 5분의 분포**를 따른다고 한다. 어느 **한 주 동안 40번의 운행**이 있었다고 할 때, **평균 운행시간이 30분보다 길 확률**은 얼마인가? 평균시간은 분단위로 반올림하여 측정된다고 가정한다.

- $\mu = 28, \sigma = 5, n = 40$
- $P(\bar{X} \geq 30.5) = P\left(\frac{\bar{X}-28}{\frac{5}{\sqrt{40}}} \geq \frac{30.5-28}{\frac{5}{\sqrt{40}}}\right) = P(Z \geq 3.16) = 0.008$



<그림 8.4 $\mu_{\bar{X}} = 28, \sigma_{\bar{X}} = \frac{5}{\sqrt{40}}$인 정규분포>

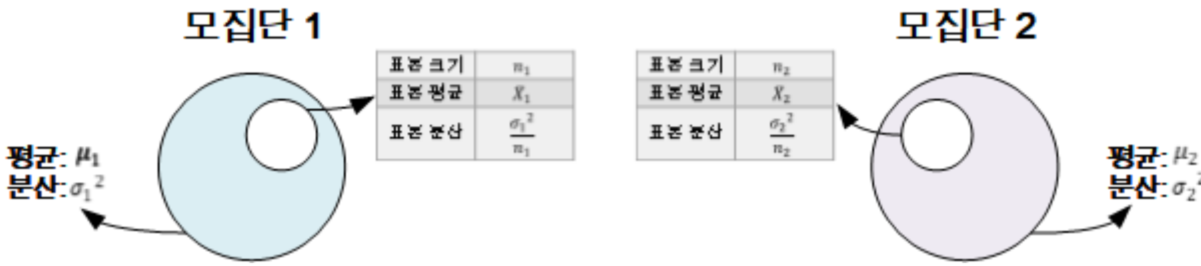
표본평균의 분포 (7/9)

- 두 표본평균 차이($\overline{X_1}$- $\overline{X_2}$)의 분포 (1/2)

정리 7.11

확률변수 $X_1, X_2, \cdots, X_n$이 서로 독립이면서 각각 평균이 $\mu_1, \mu_2, \cdots, \mu_n$이고 분산이 $\sigma^2{}_1, \sigma^2{}_2, \cdots, \sigma^2{}_n$인 정규분포를 따르면, $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$에 대해

- $\mu_Y = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$
- $\sigma^2{}_Y = a^2{}_1\sigma^2{}_1 + a^2{}_2\sigma^2{}_2 + \cdots + a^2{}_n\sigma^2{}_n$



표본 크기	n_1
표본 평균	$\overline{X}_1$
표본 분산	$\frac{\sigma_1^2}{n_1}$

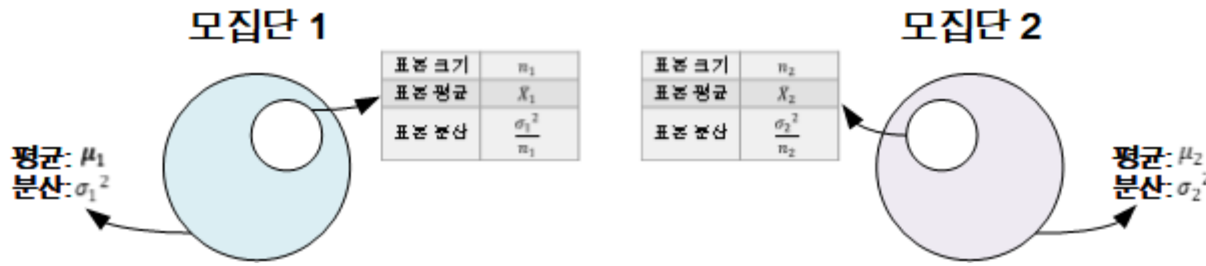
표본 크기	n_2
표본 평균	$\overline{X}_2$
표본 분산	$\frac{\sigma_2^2}{n_2}$

<그림 8.5 모집단 추정을 위한 표본 추출>

- $\overline{X_1} - \overline{X_2}$의 평균
  - $\mu_{\overline{X_1} - \overline{X_2}} = \mu_{\overline{X_1}} - \mu_{\overline{X_2}} = \mu_1 - \mu_2$
- $\overline{X_1} - \overline{X_2}$의 분산
  - $\sigma^2{}_{\overline{X_1} - \overline{X_2}} = \frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}$

표본평균의 분포 (8/9)

- 두 표본평균 차이($\overline{X_1}$- $\overline{X_2}$)의 분포 (2/2)



<그림 8.5 모집단 추정을 위한 표본 추출>

- 특징
  - $\overline{X_1}$- $\overline{X_2}$의 분포는 $N\left(\mu_1 - \mu_2, \frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}\right)$을 따르며,
    $Z = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{(\frac{{\sigma_1}^2}{n_1}) + (\frac{{\sigma_2}^2}{n_2})}}$은 근사적으로 표준정규분포를 따름

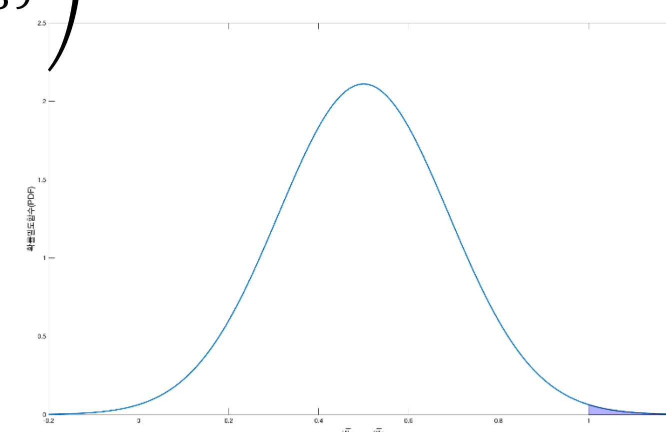
표본평균의 분포 (9/9)

- 예제 8.6

A, B 두 회사에서 텔레비전 브라운관을 생산하고 있다. **A 회사에서 생산되는 브라운관의 수명은 평균이 6.5년, 표준편차가 0.9년, B 회사에서 생산되는 브라운관의 수명은 평균 6년, 표준편차가 0.8년**이라고 알려져 있다. **A 회사에서는 36개의 확률표본을 추출하고, B 회사에서는 49개의 확률표본을 추출**했을 때, A 회사 제품의 표본 평균이 B 회사 제품의 표본평균보다 적어도 1년 이상 길 확률을 구하라

- $\mu_{\overline{X_A}} - \mu_{\overline{X_B}} = 6.5 - 6.0 = 0.5, \sigma_{\overline{X_A} - \overline{X_B}} = \sqrt{\frac{0.81}{36} + \frac{0.64}{49}} = 0.189$
- $P\left(\overline{X_A} - \overline{X_B} \geq 1.0\right) = P\left(\frac{\left(\overline{X_A} - \overline{X_B}\right) - \left(\mu_{\overline{X_A}} - \mu_{\overline{X_B}}\right)}{\sqrt{\left(\frac{\sigma_A{}^2}{n_A}\right) + \left(\frac{\sigma_B{}^2}{n_B}\right)}} \geq \frac{1-0.5}{0.189}\right) = P(Z \geq 2.65)$

$= 1 - P(Z < 2.65) = 1 - 0.9960 = 0.0040$



<그림 8.6 $\mu_{\overline{X_A}} - \mu_{\overline{X_B}} = 0.5$, $\sigma_{\overline{X_A} - \overline{X_B}} = 0.189$인 정규분포>

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- 표본평균의 분포
- 표본분산의 분포
- t 분포
- F 분포
- 분위수와 확률 그림
- 부록

표본분산의 분포 (1/4)

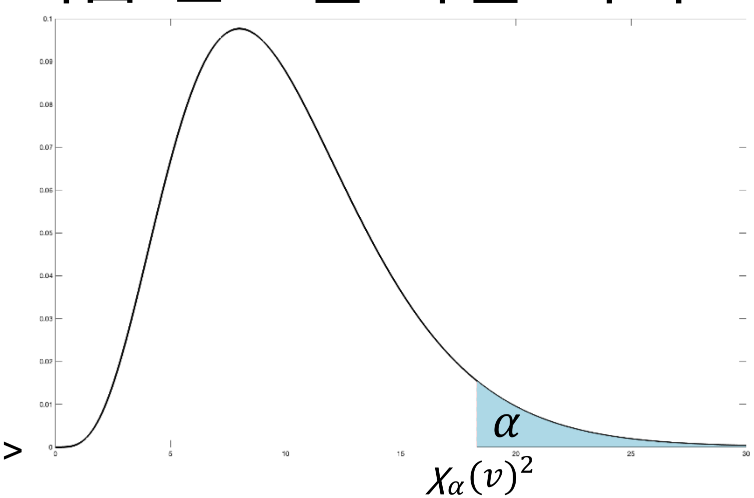
- 표본분산의 분포
  - 필요성
    - 모집단에서 표본을 추출하였을 때, 해당 표본의 분산 S^2이 모집단 σ^2에 어떻게 비례하는지 파악함으로써 표본을 통해 모집단에 대한 정보를 얻음
  - 특징
    - 분산이 σ^2인 정규모집단으로부터 크기 n인 표본을 추출하였을 때, 표본분산을 S^2이라 하면 통계량

$$X^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

은 자유도 $v = n - 1$인 카이제곱분포(Chi-square Distribution)를 따름

표본분산의 분포 (2/4)

- 카이제곱 분포(Chi-squared Distribution)
  - 정의
    - k개의 서로 독립적인 표준 정규 확률변수를 각각 제곱한 다음 합함으로써 얻어지는 분포
  - 특징
    - k의 자유도를 가짐
    - 확률변수 X가 자유도가 v인 카이제곱 분포를 따를 때의 확률은 $P(X > \chi_\alpha{}^2) = \alpha$



<그림 8.7 카이제곱분포 수표값>

표본분산의 분포 (3/4)

- 표본분산의 분포
  - 표본분산
    - 공식
      - $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

        $\rightarrow (n-1)S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2$ … ① 양변에 $(n-1)$ 곱셈 수행

        $\rightarrow \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$ … ② 양변을 σ^2 으로 나눗셈 수행
    - 특징
      - $\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$ 의 경우, 표준정규분포 제곱의 합
        - 따라서, 이는 카이제곱 분포를 따르는 것을 확인할 수 있음

표본분산의 분포 (4/4)

- 예제 8.7

어떤 자동차배터리 제조업자는 자기회사에서 제조한 배터리의 수명이 평균은 3년, **<u>표준편차는 1년</u>**이라고 주장하고 있다. 이 회사에서 제조한 5개의 배터리를 임의로 추출하여 시험한 결과 그 수명이 각각 1.9년, 2.4년, 3.0년, 3.5년, 4.2년이었다. 이 결과를 가지고 제조업자가 주장하는 표준편차가 1년이라는 것을 믿을 수 있는가를 보여라. 배터리의 수명은 정규분포를 따른다고 가정하자

- 표본 데이터를 기반으로 계산한 결과
  - 표본 평균 $\bar{x} = 3.0$년, 표본분산 $S^2 = 0.815$, 표본 표준편차 $S = 0.903$
  - 자유도: $n - 1 = 5 - 1 = 4$

- $\chi^2 = \frac{(n-1)(s^2)}{\sigma^2} = \frac{(4)(0.815)}{1} = 3.26$
- 자유도가 4일 때의 임계값인 0.484와 11.143의 범위 안에 카이제곱 값이 포함되므로 제조업자의 주장은 합당함

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- 표본평균의 분포
- 표본분산의 분포
- t 분포
- F 분포
- 분위수와 확률 그림
- 부록

t 분포 (1/6)

- t 분포 (1/3)
  - 정의
    - 표준정규분포 확률변수 Z와 자유도가 v인 카이제곱확률변수 Y에 대해 $T = \frac{Z}{\sqrt{Y/v}}$가 가지는 분포
      - $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, $Y = \frac{(n-1)S^2}{\sigma^2}$
  - 필요성
    - 모집단의 분산에 대해 알려지지 않았을 때, 표본을 통해 모평균을 추정하고자 함

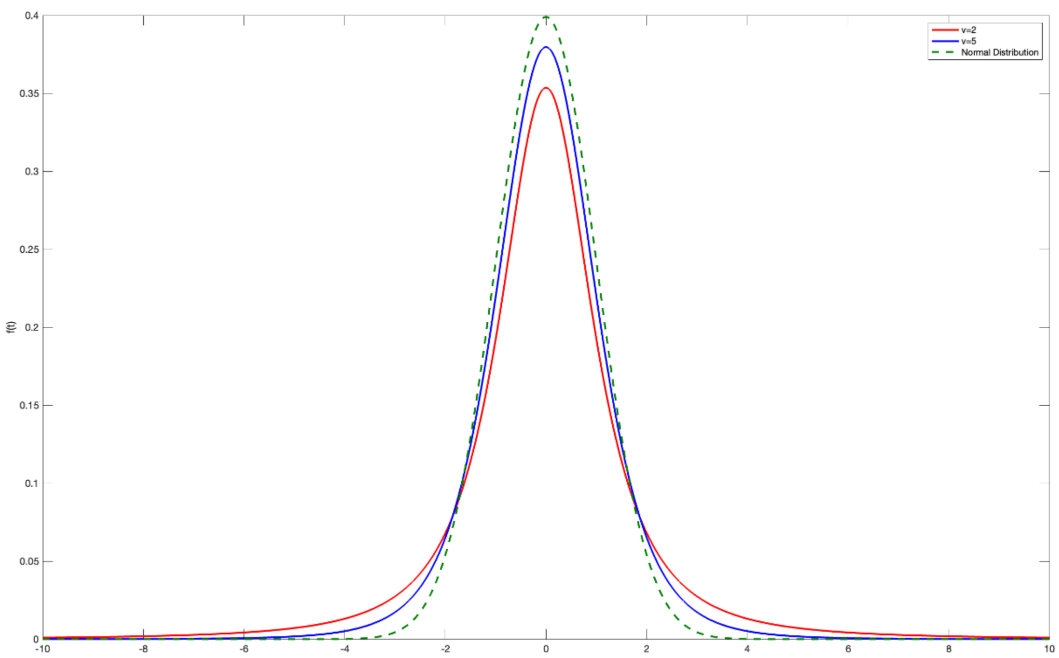
t 분포 (2/6)

- t 분포 (2/3)
  - 특징 (1/2)
    - t 분포의 확률변수인 통계량 $T = \frac{Z}{\sqrt{Y/v}}$은 자유도가 $v = n - 1$인 t분포를 따르며, 확률밀도함수 $h(t) = \frac{\Gamma\left[\frac{v+1}{2}\right]}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}}(1 + \frac{t^2}{v})^{-(v+1)/2},\ -\infty < t < \infty$을 가짐
    - $T = \frac{Z}{\sqrt{Y/v}}$에 대해 $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}},\ Y = \frac{(n-1)S^2}{\sigma^2}$이므로,

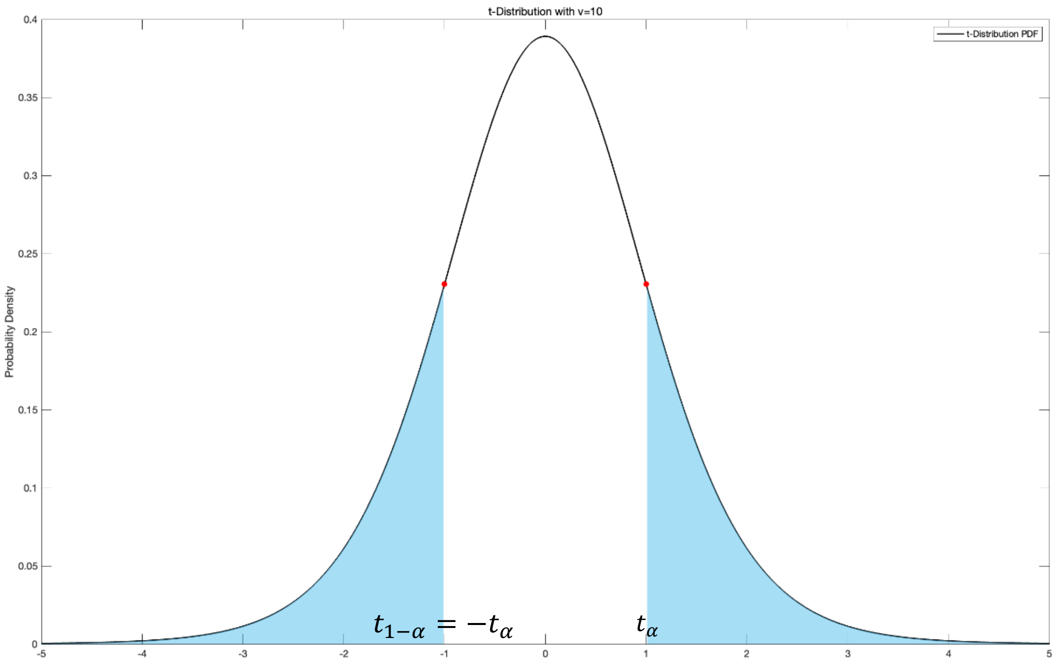
      $$T = \frac{Z}{\sqrt{Y/v}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\frac{(n-1)S^2}{\sigma^2}} = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

      - 모집단의 분산을 알 때는 표준정규분포를 사용하며, 모집단의 분산을 알 수 없을 때는 t 분포를 사용함

- t 분포 (3/3)
  - 특징 (2/2)
    - 표준정규분포와 같이 원점을 중심으로 대칭을 이룸
    - 양쪽 면적이 각각 α가 되는 점을 t_α, $t_{1-\alpha}$로 나타냄
      - $t_{1-\alpha} = -t_\alpha$ (e.g., $t_{0.95} = -t_{0.05}$)



<그림 8.8 $v = 2, 5$인 t 분포 및 표준정규분포>

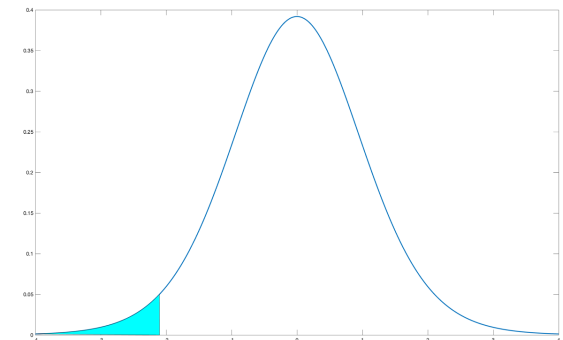


<그림 8.9 t 분포의 대칭성>

t 분포 (4/6)

- 예제 8.8

자유도 14인 t 분포에서 왼쪽 면적이 0.025가 되는 t값은 오른쪽 면적이 0.975가 되는 t값과 같게 되므로 $t_{0.975} = -t_{0.025} = -2.145$ 이다.

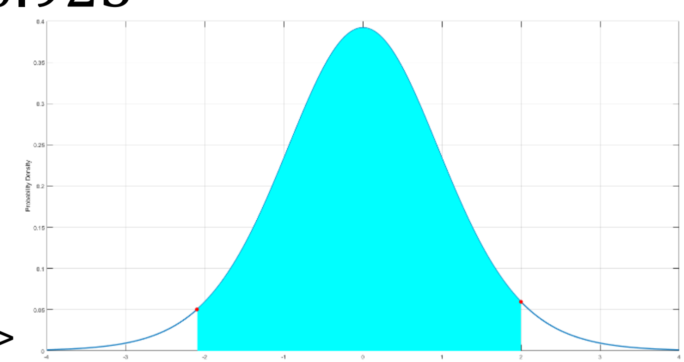


<그림 8.10 $v = 14$인 t 분포>

- 예제 8.9

$P(-t_{0.025} < T < t_{0.05})$를 구하라

- $P(-t_{0.025} < T < t_{0.05}) = 1 - 0.05 - 0.025 = 0.925$



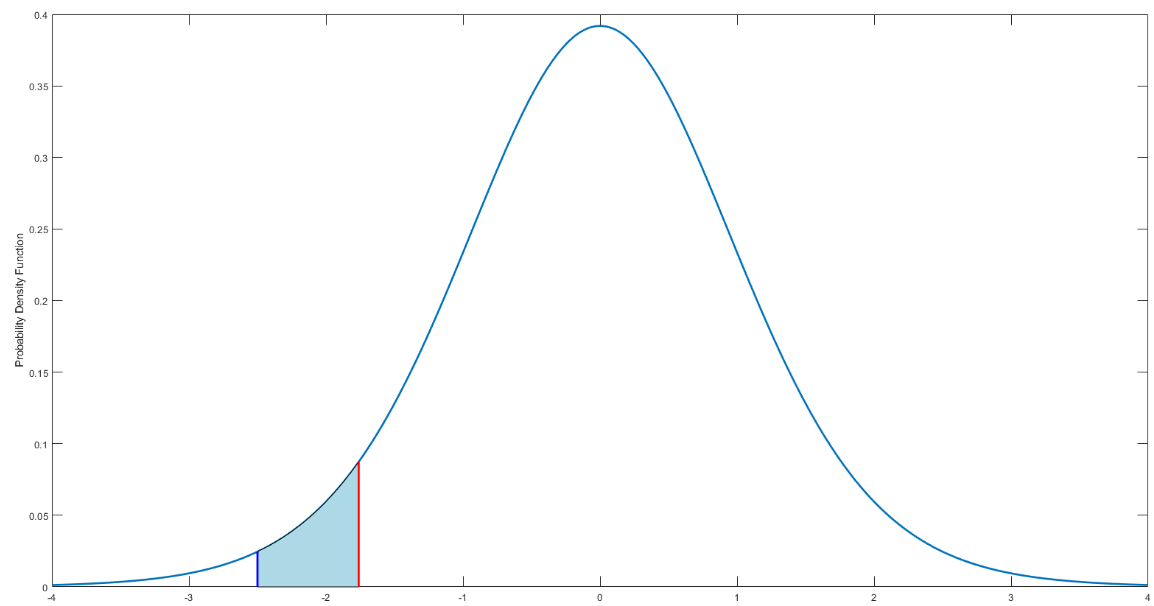
<그림 8.11 t 분포곡선에서의 면적>

t 분포 (5/6)

- 예제 8.10

정규모집단에서 **<u>15개의 표본</u>**을 추출했을 때, $P(k < T < -1.761) = 0.045$를 만족하는 k값을 구하라

- 자유도가 14인 t 분포에서 $1.761 = t_{0.05}$
- $k = -\ t_{\alpha}$ 라고 하면, $0.045 = 0.05 - \alpha$
- $\alpha = 0.005$, $k = -\ t_{0.005} = -2.977$



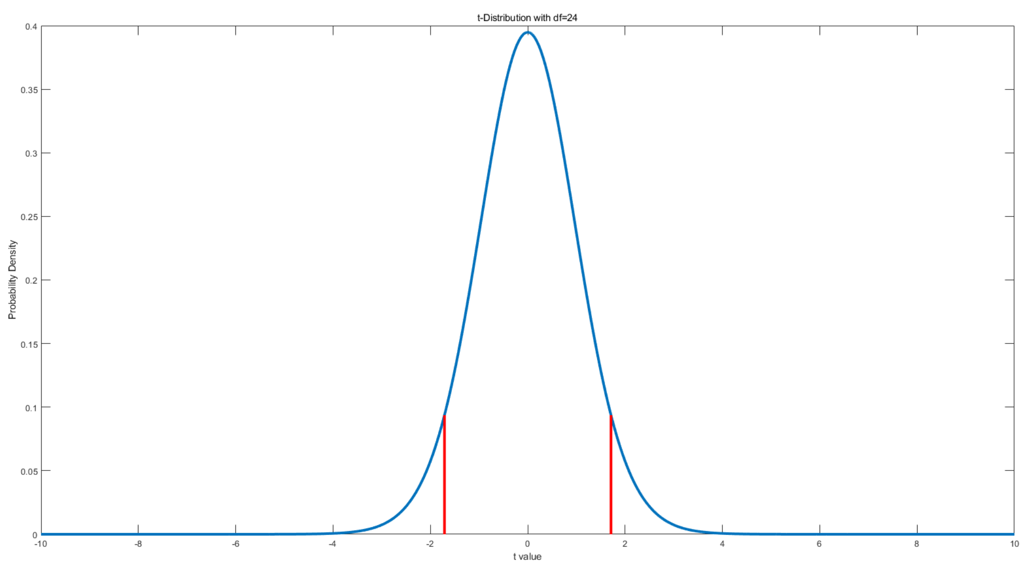
<그림 8.12 $v = 14$인 t 분포>

t 분포 (6/6)

- 예제 8.11

어떤 화공기사는 어느 배치공정의 수율이 원재료의 리터 당 500g이라고 주장하고 있다. 그는 이를 입증하기 위해 매월 **<u>25개의 배치를 추출</u>**하여 시험을 하였다. 시험 결과 계산된 t 값이 $-t_{0.05}$와 $t_{0.05}$ 사이에 있으면 그의 주장이 타당성이 있다고 하기로 한다. **<u>25개 배치의 시험결과 표본평균은 518g이고 표준편차는 40g</u>**이었다면 어떤 결론을 낼 수 있는가? 단, 모집단은 근사적으로 정규분포를 따른다고 가정하라

- $t = \frac{518-500}{40/\sqrt{25}} = 2.25$
- 자유도: $n - 1 = 25 - 1 = 24$
- 자유도가 24일 때, $t_{0.05} = 1.711$
  - t 값은 $t_{0.05}$보다 크므로 실제의 수율은 500g보다 큼



<그림 8.13 $v = 24$인 t 분포>

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- 표본평균의 분포
- 표본분산의 분포
- t 분포
- F 분포
- 분위수와 확률 그림
- 부록

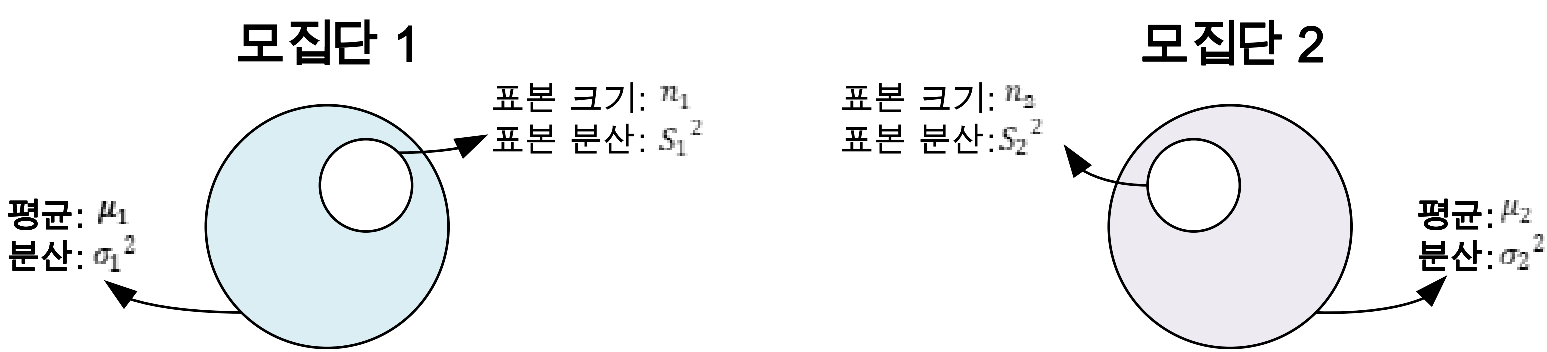
F 분포 (1/4)

- F 분포 (1/4)
  - 정의
    - 두 개의 서로 독립적인 카이제곱변수의 비율 $F = \frac{U/v_1}{V/v_2}$가 가지는 분포
      - U와 V는 서로 독립이면서 각각 자유도가 v_1, v_2인 카이제곱확률변수
  - 필요성
    - 카이제곱변수의 비율을 통해 두 집단의 분산을 비교하고자 함

- F 분포 (2/4)
  - 표현
    - $U = \frac{(n_1-1){S_1}^2}{{\sigma_1}^2}, V = \frac{(n_2-1){S_2}^2}{{\sigma_2}^2}$
    - $F = \frac{\frac{(n_1-1){S_1}^2}{{\sigma_1}^2}/(n_1-1)}{\frac{(n_2-1){S_2}^2}{{\sigma_2}^2}/(n_2-1)} = \frac{{S_1}^2/{\sigma_1}^2}{{S_2}^2/{\sigma_2}^2} = \frac{{\sigma_2}^2 \cdot {S_1}^2}{{\sigma_1}^2 \cdot {S_2}^2}$
      - $F(n_1 - 1, n_2 - 1)$을 따름

F 분포 정의

- 두 개의 서로 독립적인 카이제곱변수의 비율 $F = \frac{U/v_1}{V/v_2}$가 가지는 분포
  - U와 V는 서로 독립이면서 각각 자유도가 v_1, v_2인 카이제곱확률변수

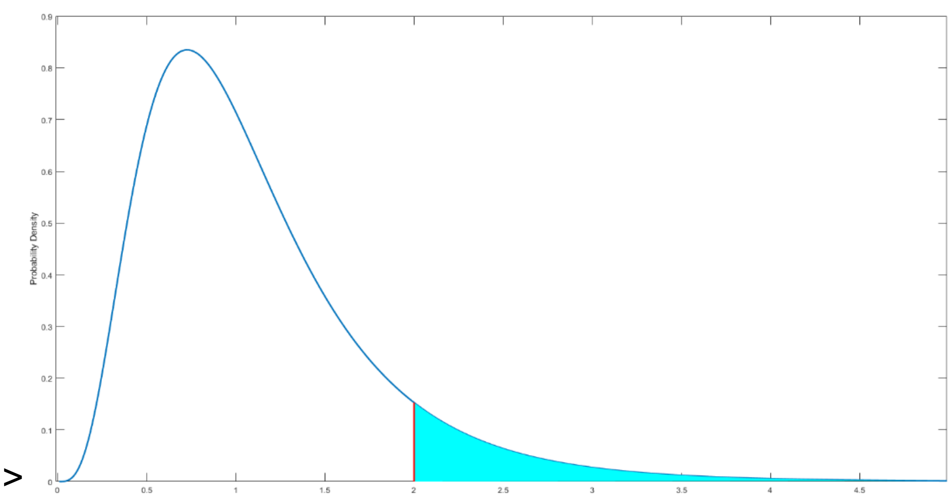


<그림 8.14 모집단 추정을 위한 표본 추출>

- F 분포 (3/4)
  - 특징 (1/2)
    - F 분포의 확률변수인 통계량 $F = \frac{U/v_1}{V/v_2}$은 확률밀도함수

$$h(f) = \begin{cases} \frac{\Gamma[(v_1+v_2)/2](v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_2/2)} \frac{f^{\left(\frac{v_1}{2}\right)-1}}{(1+v_1 f/v_2)}, f > 0 \\ 0, f \leq 0 \end{cases}$$ 을 가짐

    - f_α는 f_α보다 큰 면적이 α가 됨을 의미함



<그림 8.15 F 분포의 수표값>

F 분포 (4/4)

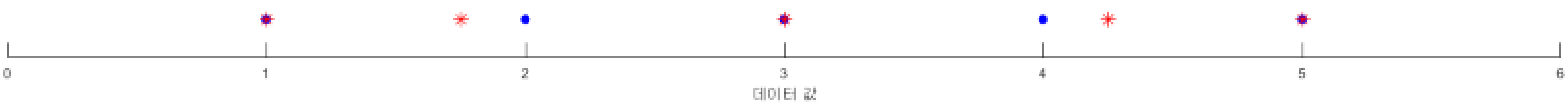
- F 분포 (4/4)
  - 특징 (2/2)
    - 자유도 v_1, v_2에서 f_α의 값을 $f_\alpha(v_1, v_2)$으로 나타내며, $f_{1-\alpha}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}$
      - F 분포는 분산 비율의 분포를 나타냄에 따라 비율의 역수는 그 분포의 역방향을 얻게됨
      - 분산 비율이 특정 임계값보다 클 확률은, 그 역수가 해당 임계값의 역수보다 작을 확률과 동일함

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- 표본평균의 분포
- 표본분산의 분포
- t 분포
- F 분포
- 분위수와 확률 그림
- 부록

분위수와 확률 그림 (1/5)

- 분위수 그림(Quantile Plot) (1/2)
  - 분위수(Quantile) $q(f)$
    - 자료값 중에서 $q(f)$ 보다 작거나 같은 자료값이 차지하는 비율이 f가 되는 값
      - e.g., 표본 중앙값: $q(0.5)$

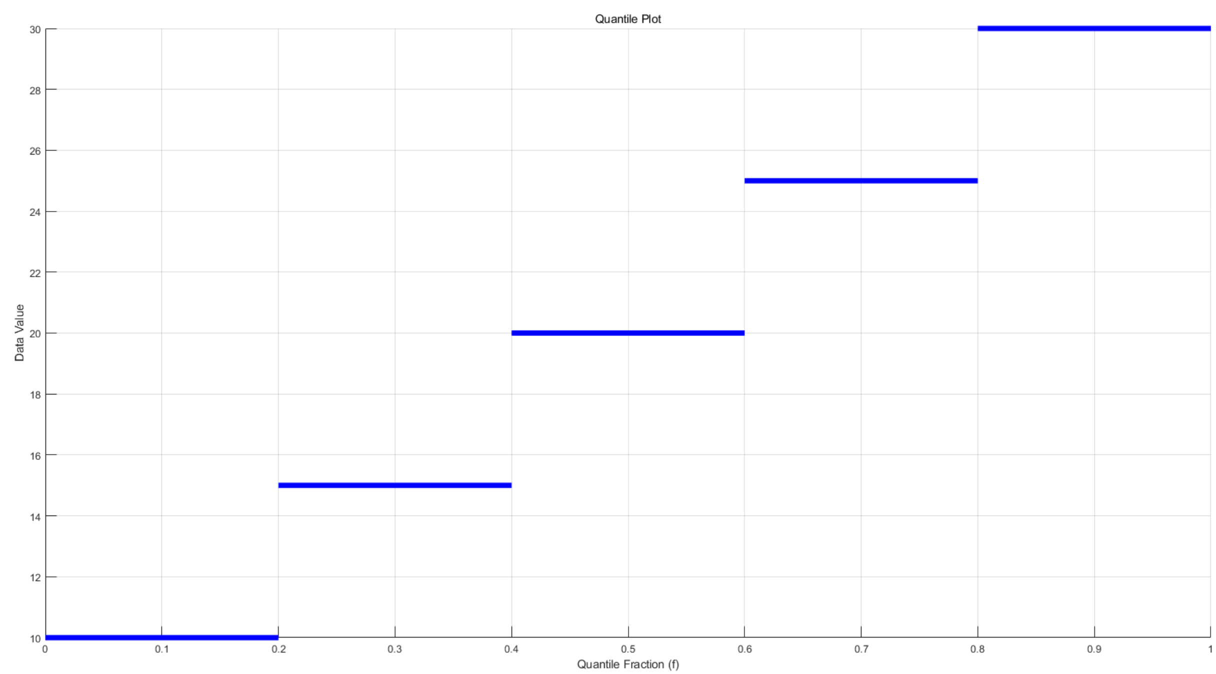


<그림 8.16 5개의 데이터를 가질 때의 분위수>

  - 정의
    - 데이터의 분포 형태 시각적으로 평가하는 도구
  - 필요성
    - 데이터 집합의 분포를 이해하고, 이론적 분포와의 비교를 통해 분포 특성을 평가할 수 있음

분위수와 확률 그림 (2/5)

- 분위수 그림(Quantile Plot) (2/2)
  - 특징
    - 수직선상에 자료값의 눈금을 매기고 전체 자료 중에서 이 값들보다 작게 되는 자료의 비율을 $f_i = \frac{i-\frac{3}{8}}{n+\frac{1}{4}}$으로 나타냄



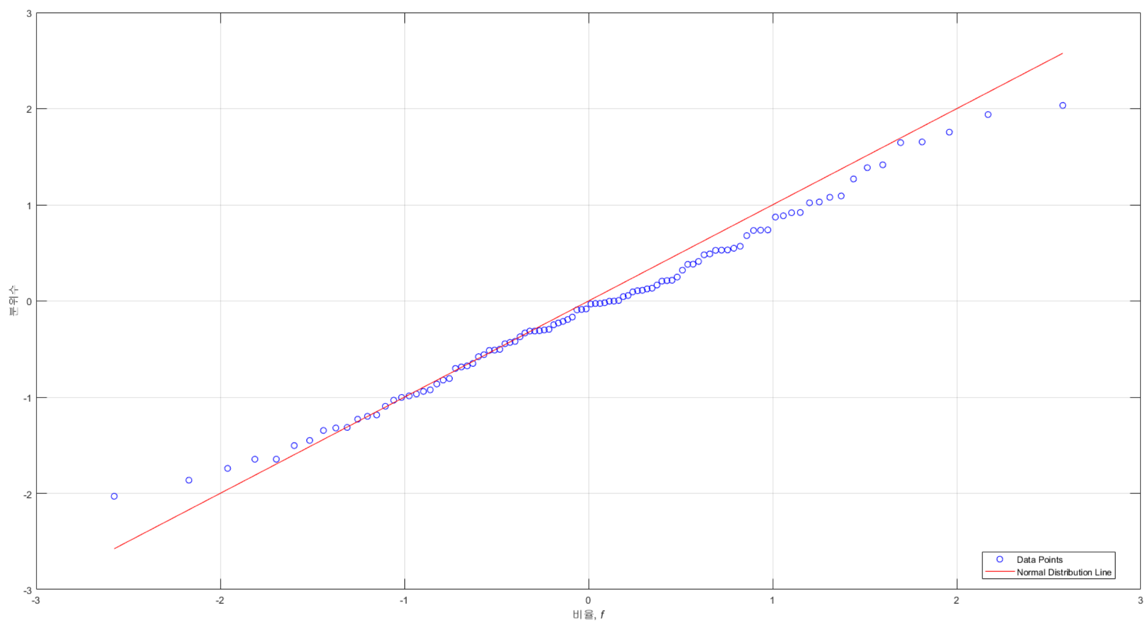
<그림 8.17 5개의 데이터(10, 15, 20, 25, 30) 를 갖는 분위수 그림>

분위수와 확률 그림 (3/5)

- 정규 분위수 그림(Normal Quantile-Quantile Plot) (1/2)
  - 정의
    - 주어진 표본이 정규분포를 따르는지 시각적으로 평가하는 도구
  - 특징
    - 정규확률변수의 분위수 표현의 근사값은 $q_{\mu,\sigma}(f) = \mu + \sigma\{4.19[f^{0.14} - (1-f)^{0.14}]\}$으로 나타낼 수 있음
      - $N(0,1)$인 확률변수에 대한 근사식은 $q_{0,1}(f) = 4.19[f^{0.14} - (1-f)^{0.14}]$으로 표현할 수 있음

분위수와 확률 그림 (4/5)

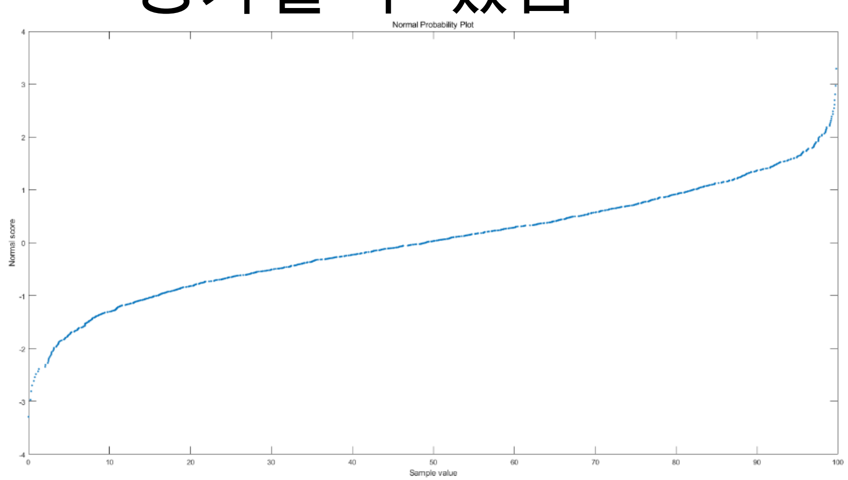
- 정규 분위수 그림(Normal Quantile-Quantile Plot) (2/2)
  - 필요성
    - 데이터 포인트가 정규 분포를 나타내는 표준 참조선 주변에 배열되어 있음을 확인함으로써 데이터가 정규 분포를 따르는지 여부 판단 가능



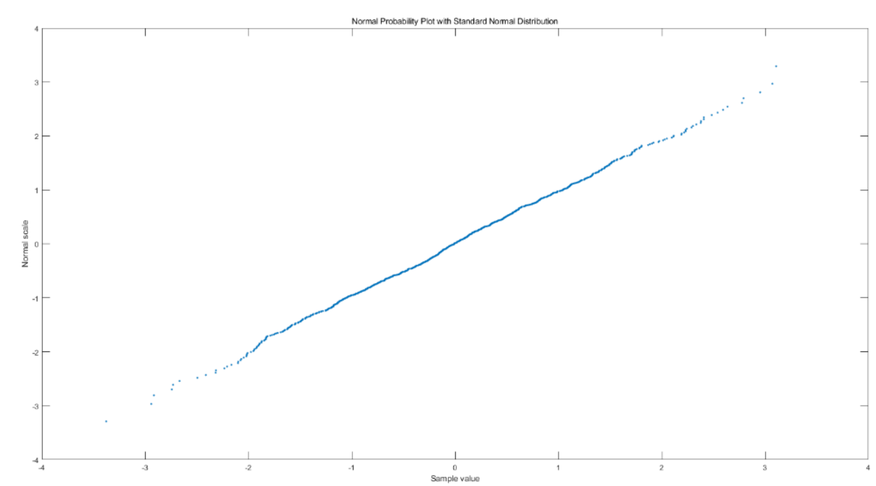
<그림 8.18 정규 분위수 그림 및 표준 참조선>

분위수와 확률 그림 (5/5)

- 정규 확률 그림(Normal Probability Plot)
  - 정의
    - 데이터가 정규 분포를 따르는지 여부를 시각적으로 평가하기 위해 데이터의 분위수를 정규 분포의 이론적 분위수와 비교하여 그래프로 나타낸 것
  - 필요성
    - 관측된 데이터의 분포가 정규분포를 따르는지 시각적으로 평가할 수 있음



<그림 8.19 정규분포를 따르지 않는 정규 확률 그림>

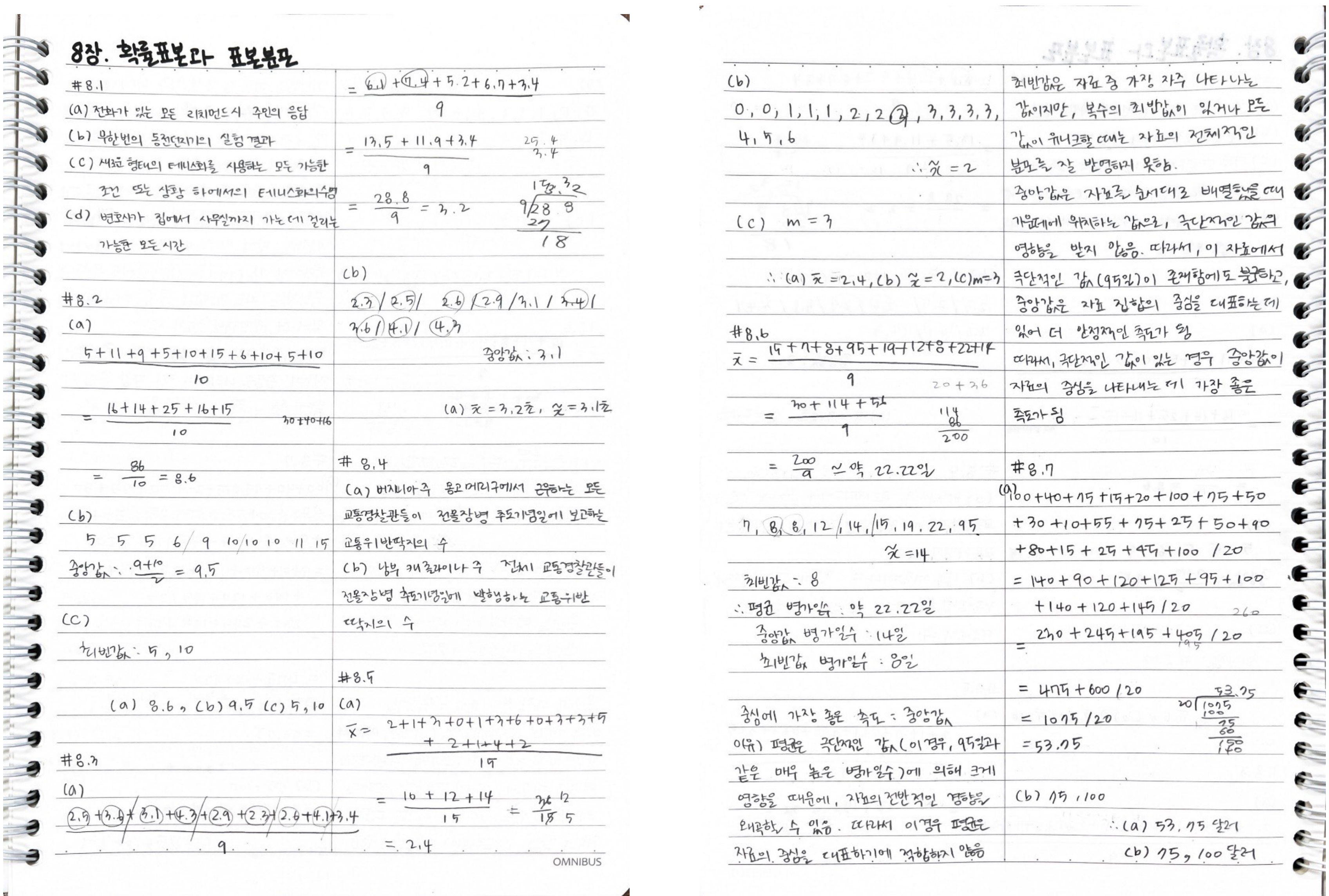


<그림 8.20 정규분포를 따르는 정규 확률 그림>

Thanks!

손 우 영 (wooyoung@pel.sejong.ac.kr)

부록 #1 – 연습문제 (8.1~8.7)



8장. 확률표본과 표본분포

#8.1

(a) 전화가 있는 모든 리치먼드시 주민의 응답

(b) 무한번의 동전던지기의 실험 결과

(c) 새로운 형태의 테니스화를 사용하는 모든 가능한 조건 또는 상황 하에서의 테니스화의 수명

(d) 변호사가 집에서 사무실까지 가는 데 걸리는 가능한 모든 시간

#8.2

(a)

$$\frac{5+11+9+5+10+15+6+10+5+10}{10}$$

$$= \frac{16+14+25+16+15}{10} \quad 30+40+16$$

$$= \frac{86}{10} = 8.6$$

(b)

5 5 5 6 / 9 10 / 10 10 11 15

중앙값 : $\frac{9+10}{2} = 9.5$

(c)

최빈값 : 5, 10

(a) 8.6, (b) 9.5 (c) 5, 10

#8.3

(a)

$$\frac{2.5+3.6+3.1+4.3+2.9+2.3+2.6+4.1+3.4}{9}$$

$$= \frac{6.1+7.4+5.2+6.7+3.4}{9}$$

$$= \frac{13.5+11.9+3.4}{9}$$

$$= \frac{28.8}{9} = 3.2$$

(b)

2.3 / 2.5 / 2.6 / 2.9 / 3.1 / 3.4 / 3.6 / 4.1 / 4.3

중앙값 : 3.1

(a) $\bar{x} = 3.2$초, $\tilde{x} = 3.1$초

#8.4

(a) 버지니아주 몽고메리구에서 근무하는 모든 교통경찰관들이 전몰장병 추도기념일에 보고하는 교통위반딱지의 수

(b) 남부 캐롤라이나 주 전체 교통경찰관들이 전몰장병 추도기념일에 발행하는 교통위반 딱지의 수

#8.5

(a)

$$\bar{x} = \frac{2+1+3+0+1+3+6+0+3+3+5+2+1+4+2}{15}$$

$$= \frac{10+12+14}{15} = \frac{36}{15} = 2.4$$

(b)

0, 0, 1, 1, 1, 2, 2, 3, 3, 3, 3, 3, 4, 5, 6

$\therefore \tilde{x} = 2$

(c) m = 3

$\therefore$ (a) $\bar{x} = 2.4$, (b) $\tilde{x} = 2$, (c) m = 3

#8.6

$$\bar{x} = \frac{15+7+8+95+19+12+8+22+14}{9}$$

$$= \frac{30+114+56}{9}$$

$$= \frac{200}{9} \simeq \text{약 } 22.22\text{일}$$

7, 8, 8, 12, 14, 15, 19, 22, 95

$\tilde{x} = 14$

최빈값 : 8

$\therefore$ 평균 병가일수 : 약 22.22일

중앙값 병가일수 : 14일

최빈값 병가일수 : 8일

중심에 가장 좋은 측도 : 중앙값

이유) 평균은 극단적인 값 (이 경우, 95일과 같은 매우 높은 병가일수)에 의해 크게 영향을 때문에, 자료의 전반적인 경향을 왜곡할 수 있음. 따라서 이 경우 평균은 자료의 중심을 대표하기에 적합하지 않음

최빈값은 자료 중 가장 자주 나타나는 값이지만, 복수의 최빈값이 있거나 모든 값이 유니크할 때는 자료의 전체적인 분포를 잘 반영하지 못함.

중앙값은 자료를 순서대로 배열했을 때 가운데에 위치하는 값으로, 극단적인 값의 영향을 받지 않음. 따라서, 이 자료에서 극단적인 값 (95일)이 존재함에도 불구하고, 중앙값은 자료 집합의 중심을 대표하는 데 있어 더 안정적인 측도가 됨

따라서, 극단적인 값이 있는 경우 중앙값이 자료의 중심을 나타내는 데 가장 좋은 측도가 됨

#8.7

(a) 100+40+75+15+20+100+75+50+30+10+55+75+25+50+90+80+15+25+45+100 / 20

= 140+90+120+125+95+100+140+120+145 / 20

= 230+245+195+405 / 20

= 475+600 / 20

= 1075 / 20

= 53.75

(b) 75, 100

$\therefore$ (a) 53.75 달러

(b) 75, 100 달러

OMNIBUS

<「이공학도를 위한 확률 및 통계학」 p.261>

부록 #2 – 연습문제 (8.8~8.12)

#8.8

(a) $\dfrac{48+47+42+42+41+34+31+30+29+29+29+26}{12}$

$= \dfrac{95+84+75+61+58+55}{12}$

$= \dfrac{179+136+133}{12}$

$= \dfrac{448}{12} =$ 약 37.33

(b)

$\tilde{x} = \dfrac{34+31}{2} = \dfrac{65}{2} = 32.5$

(c) 29

∴ (a) $\bar{x}$ = 약 37.33g, (b) $\tilde{x}$ = 32.5g

(c) 29g

#8.9

(a) $15-5=10$

(b)

$S^2 = \dfrac{1}{n(n-1)}\left[n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2\right]$

1) $n=10$,

2) $\sum_{i=1}^{n}x_i^2 = 25+121+81+25+100+225+36+100+25+100$

$= 146+106+325+136+125$

$= 838$

252+325+261

577+261 838

3) $\sum_{i=1}^{n}x_i = 86$

따라서,

$S^2 = \dfrac{1}{10\times 9}\left[10\times 838 - 86^2\right]$

$= \dfrac{1}{90}(8380-7396)$

$= \dfrac{984}{90}$

= 약 10.933…

$S = \sqrt{10.933}$

= 약 3.3065…

∴ (a) 범위: 10, (b) S: 약 3.307

#8.10

(a) $x_{max} = 4.3$

$x_{min} = 2.3$

$\rightarrow R = x_{max} - x_{min}$

$= 4.3 - 2.3$

$= 2$

∴ 2

(b)

표본분산 $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

$n=9$, $\bar{x}=3.2$

OMNIBUS

(7.2)

$S^2 = \dfrac{1}{9-1}\{(-0.7)^2+0.4^2+(-0.1)^2+(1.1)^2+(-0.3)^2+(-0.9)^2+(-0.6)^2+(0.9)^2+(0.2)^2\}$

$= \dfrac{1}{8}(0.49+0.16+0.01+1.21+0.09+0.81+0.36+0.81+0.04)$

$= \dfrac{1}{8}(3.98) = 0.4975$

∴ 0.4975

#8.11

(a)

$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

$n=15$, $\bar{x}=2.4$

$\Rightarrow S^2 = \dfrac{1}{15-1}(5.76\times 2+1.96\times 3+0.16\times 3+0.36\times 4+2.56\times 1+6.76\times 1+12.96)$

$= \dfrac{1}{14}(11.52+5.88+0.48+1.44+2.56+6.76+12.96)$

$= \dfrac{1}{14}(41.6)$

$= 2.971$

∴ 2.971

(b) $S^2 = \dfrac{1}{n(n-1)}\left[n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2\right]$

$n=15$

$\sum_{i=1}^{n}x_i^2 = (0^2\times 2+1^2\times 3+2^2\times 3+3^2\times 4+4^2\times 1+5^2\times 1+6^2\times 1)$

$= 3+12+36+16+25+36$

$= 15+52+61$

67
61

$= 128$

$\left(\sum_{i=1}^{n}x_i\right)^2$

$= (0\times 2+1\times 3+2\times 3+3\times 4+4+5+6)^2$

$= (3+6+12+15)^2$

$= 36^2$

$= 1296$

$S^2 = \dfrac{1}{15\times 14}(15\times 128 - 1296)$

$= \dfrac{1}{15\times 14}\times(1920-1296)$

$= \dfrac{1}{15\times 14}\times 624$

7)624 = 8 … 56, 64

$= 2.971\ldots$

∴ 2.971

#8.12

(a)

$\dfrac{7.3+8.6+10.4+16.1+12.2+15.1+14.5+9.3}{8}$

$= \dfrac{15.9+26.5+27.3+23.8}{8}$

$= \dfrac{93.5}{8}$

$= 11.6875$

(b)

$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

<「이공학도를 위한 확률 및 통계학」 p.261~262>

부록 #3 – 연습문제 (8.13~8.16)

$n=8$

$S^2 = \frac{1}{8-1}\{(-4.3875)^2 + (-3.0875)^2$
$+ (-1.2875)^2 + (4.4125)^2$
$+ (0.5125)^2 + (3.4125)^2$
$+ (2.8125)^2 + (-2.3875)^2\}$
$=$

#8.13

평균: $\frac{5.1+5.1+5.7+6.8+5.8+4.3+6.5+5.2+4.4+4.4}{20}$

$= \frac{10.2+12.5+10.1+11.7+8.8}{20}$

$= \frac{22.7+21.8+8.8}{20}$

$= \frac{53.3}{20} = 2.665$

$S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]$

$n=20$이므로,

$\sum_{i=1}^{n} x_i^2$

$= (3.2^2 + 1.9^2 + 2.7^2 + 2.4^2 + 2.8^2$
$+ 2.9^2 + 3.8^2 + 3.0^2 + 2.5^2 + 3.3^2$
$+ 1.8^2 + 2.5^2 + 3.7^2 + 2.8^2 + 2.0^2$
$+ 3.2^2 + 2.3^2 + 2.1^2 + 2.5^2 + 1.9^2)$
$= 10.24 + 3.61 + 7.29 + 5.76 + 7.84$
$+ 8.41 + 14.44 + 9 + 6.25 + 10.89 + 3.24$
$+ 6.25 + 13.69 + 7.84 + 4 + 10.24$
$+ 5.29 + 4.41 + 6.25 + 3.61$
$= 148.55$

$\left(\sum_{i=1}^{n} x_i\right)^2$
$= (3.2 + \cdots + 1.9)^2$
$= 53.3^2$
$= 2840.89$

$S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]$
$= \frac{1}{20 \times 19}(20 \times 148.55 - 2840.89)$
$= \frac{1}{20 \times 19}(2971 - 2840.89)$
$= \frac{130.11}{20 \times 19} = 0.342$

$\therefore S^2 = 0.342$이므로, $S = \sqrt{0.342} = 0.585$

$\therefore 0.585$

#8.14

(a)

표본분산은 데이터의 변동성을 측정하는 값으로, 데이터의 개별 값들이 얼마나 서로 다른지를 나타낸다. 표본분산 S^2은 다음과 같이 계산된다.

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

x_i는 각 표본 값이고 $\bar{x}$는 표본의 평균이다.

상수 c를 모든 표본 값에 더하거나 빼는 경우, 각 표본 값은 x_i+c 또는 x_i-c이다. 이렇게 변형된 표본의 새로운 평균을 $\bar{x}_{new}$라고 하면, $\bar{x}_{new} = \bar{x}+c$ or $\bar{x}_{new} = \bar{x}-c$.

변형된 표본에 대한 분산

$\therefore S^2_{new} = \frac{1}{n-1}\sum_{i=1}^{n}((x_i+c) - (\bar{x}+c))^2$

$S^2_{new} = \frac{1}{n-1}\sum_{i=1}^{n}((x_i-c) - (\bar{x}-c))^2$

$\Rightarrow S^2_{new} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

결과적으로, S^2_{new}는 원래 표본분산 S^2와 동일하다.

이는 모든 표본 값에 동일한 상수를 더하거나 빼도 표본의 분산이 변하지 않음을 보여줌

(b)

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

x_i는 각 표본 값이고, $\bar{x}$는 표본의 평균

새로운 데이터 세트에 대한 분산 S^2_{new}는 다음과 같이 계산됨

$S^2_{new} = \frac{1}{n-1}\sum_{i=1}^{n}(cx_i - c\bar{x})^2$
$= \frac{1}{n-1}\sum_{i=1}^{n}c^2(x_i - \bar{x})^2$
$= c^2\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$
$= c^2S^2$

#8.15

(a) (a)에서 주어진 데이터의 경우, 문제에서 처음 주어진 값의 각각 3배이다. 연습문제 8.14의 (b)에서, 표본의 각 관측값에 c를 곱하면 표본분산은 원래 값의 c^2배가 되므로, (a)에서 주어진 타율의 표본분산은

$5.2 \times c^2 = 5.2 \times 3^2$
$= 5.2 \times 9 = 46.8$

$\therefore 46.8$

(b) (b)에서 주어진 데이터의 경우, 문제에서 처음 주어진 값에 5를 더한 값이다. 연습문제 8.14의 (a)에서 상수 c가 표본의 각 값에 더해져도 표본분산은 변하지 않으므로, (b)에서 주어진 자료의 표본분산은 5.1

#8.16

크기순으로 정렬하면,

3 5 6 8 11 30 31 32 36 38 38 40 49

(a) 평균 점수

$= \frac{3+5+6+8+11+30+31+32+36+38 \times 2+40+49}{13}$

$= \frac{8+14+41+63+36+76+89}{13}$

$= \frac{22+104+112+89}{13}$

OMNIBUS

<「이공학도를 위한 확률 및 통계학」 p.262>

부록 #4 – 연습문제 (8.17~8.22)

$= \frac{126+201}{13} = \frac{327}{13} =$ 약 25.15

(b) 점수의 중앙값을 구할 때, 데이터의 개수는 홀수개이기 때문에,

중앙값 : $x_{\frac{n+1}{2}} = x_{\frac{13+1}{2}} = x_7 = 31$

#8.17

평균이 $\mu = 50$이고 표준편차가 $\sigma = 5$인 정규분포로부터 크기가 $n=16$인 표본을 추출할 때, 표본평균 $\bar{X}$의 분포도 정규분포를 따른다.

㉮ $P(\mu_{\bar{x}} - 1.9\sigma_{\bar{x}} < X < \mu_{\bar{x}} - 0.4\sigma_{\bar{x}})$

$= P(\mu - 1.9 \cdot \frac{\sigma}{\sqrt{n}} < X < \mu - 0.4 \cdot \frac{\sigma}{\sqrt{n}})$

$= P\left(\frac{(\mu - 1.9 \cdot \frac{\sigma}{\sqrt{n}} - \mu)}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{(\mu - 0.4\frac{\sigma}{\sqrt{n}}) - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$

$= P(-1.9 < Z < -0.4) \fallingdotseq 31.59\%$

∴약 31.59%

#8.18

$n = 36,\ \sigma = 2,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$2 = \frac{\sigma}{\sqrt{36}} \Rightarrow \sigma = 12$

따라서, $\sigma_{\bar{x}} = \frac{12}{\sqrt{n}} = 1.2$가 되기위해

$\sqrt{n} = 10,\ n = 100$

∴100

#8.19

$\mu = 78.3kg,\ \sigma = 5.6kg$

(a) 64에서 196으로 증가할 때,

분산 $= \frac{\sigma^2}{n}$

$n=64$일 때 : $\sigma_{\bar{x}}^2 = \frac{5.6^2}{64} = 0.49$

$n=196$일 때 : $\sigma_{\bar{x}}^2 = \frac{5.6^2}{196} = 0.16$

∴표본의 크기가 64에서 196으로 증가하여 분산은 0.49에서 0.16으로 감소한다.

(b) 784에서 49로 감소할 때

분산 $= \frac{\sigma^2}{n}$

$n = 784$일 때

: $\sigma_{\bar{x}}^2 = \frac{5.6^2}{784} = \frac{5.6 \times 5.6}{28 \times 28} = \frac{1}{25}$

$n = 49$일 때

: $\sigma_{\bar{x}}^2 = \frac{5.6^2}{49} = \frac{5.6 \times 5.6}{49}$

$= \frac{\cancel{56}8 \times \cancel{56}8}{49\cancel{00}\ 100} = \frac{64}{100}$

$= 0.64$

∴표본의 크기가 784에서 49로 감소하여 분산은 0.04에서 0.64로 증가한다.

#8.20

$n = 54$

㉮ $P(4.05 \le \bar{X} < 4.45)$

$\mu = \frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 4 + \frac{1}{3} \cdot 6$

$= \frac{1}{3} \cdot 12 = 4$

$\sigma^2 = E(X^2) - [E(X)]^2$

$E(X^2) = \frac{4+16+36}{3} = \frac{56}{3}$ 이므로,

$\sigma^2 = \frac{56}{3} - 16 = \frac{56-48}{3} = \frac{8}{3}$

$\sigma = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{\frac{8}{3}}{54}} = \sqrt{\frac{8}{27 \times 3}}$

$= \frac{2}{9}$

$P\left(\frac{4.05-4}{\frac{2}{9}} \le Z < \frac{4.45-4}{\frac{2}{9}}\right)$

$= P\left(\frac{0.05}{\frac{2}{9}} \le Z < \frac{0.45}{\frac{2}{9}}\right)$

$= P\left(\frac{0.45}{2} \le Z < \frac{4.05}{2}\right)$

$= P(0.225 \le Z < 2.025)$

≒ 약 38.96%

∴약 38.96%

#8.21

$\mu = 240,\ \sigma = 15,\ n = 40$

$\Rightarrow \sigma_{\bar{x}} = \frac{15}{\sqrt{40}} \fallingdotseq 2.37$

$\mu_{\bar{x}} + 2\sigma_{\bar{x}} = 240 + 2 \times 2.37$

$= 240 + 4.74$

$= 244.74$

$\mu_{\bar{x}} - 2\sigma_{\bar{x}} = 240 - 2 \times 2.37$

$= 240 - 4.74$

$= 235.26$

236ml은 235.26ml와 244.74ml 내에 있으므로 정상적으로 작동하고 있다고 판단하는 것은 합리적인 결정이다.

#8.22

$\mu = 174.5cm\ /\ \sigma = 6.9cm$

$n = 25\ /\ 200$회 추출

(a) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ 이므로, $n = 25$일 때

표본평균의 평균은 모집단의 평균인 174.5cm이고, 표본평균의 표준편차는 약 1.38cm 이다

∴약 1.38cm

(b)

㉮ : $P(172.45 \le X < 175.85)$

$= P\left(\frac{172.45 - 174.5}{\frac{6.9}{\sqrt{25}}} < Z < \frac{175.85 - 174.5}{\frac{6.9}{\sqrt{25}}}\right)$

$= P\left(\frac{5 \times (-2.05)}{6.9} < Z < \frac{1.35 \times 5}{6.9}\right)$

$= P(-1.48 < Z < 0.97)$

$= 0.7533$

$\therefore \cancel{200}\,2 \times \frac{75.33}{\cancel{100}} = 150.66$

∴150개

OMNIBUS

<「이공학도를 위한 확률 및 통계학」 p.271>

부록 #5 – 연습문제 (8.23~8.26)

(c)

④ $P(X<172.05)$

$= P\left(Z < \frac{172.05-174.5}{\frac{6.9}{\sqrt{25}}}\right)$

$= P\left(Z < \frac{-2.45 \times 5}{6.9}\right)$

$= P(Z < -1.775)$

$= 0.0379$

$\therefore \frac{\cancel{200}}{2} \times \frac{3.79}{\cancel{100}} = 7.58$

∴약 8개

#8.23

(a) $\mu = 4 \times 0.2 + 5 \times 0.4 + 6 \times 0.3 + 7 \times 0.1$

$= 0.8 + 2 + 1.8 + 0.7$

$= 2.8 + 2.5$

$= 5.3$

$\sigma^2 = E(X^2) - [E(X)]^2$

$E(X^2) = 16 \times 0.2 + 25 \times 0.4 + 36 \times 0.3 + 49 \times 0.1$

$= 3.2 + 10 + 10.8 + 4.9$

$= 13.2 + 15.7$

$= 28.9$

$\therefore \sigma^2 = 28.9 - (5.3)^2$

$= 28.9 - 28.09$

$= 0.81$

$\therefore \mu = 5.3,\ \sigma^2 = 0.81$

(b)

$n = 36$

$\mu_{\bar{x}} = \mu = 5.3$

$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{0.81}{36} = \frac{\cancel{81}\ 9}{2^2 \times \cancel{3^2} \times 100}$

$= \frac{9}{400}$

(c)

② $P(X < 5.5)$

$= P\left(Z < \frac{5.5-5.3}{\sqrt{\frac{9}{400}}}\right)$

$= P\left(Z < \frac{0.2}{\frac{3}{20}}\right)$

$= P\left(Z < \frac{4}{3}\right)$

$= 1 - P\left(Z \geq \frac{4}{3}\right)$

$= 0.9082$

#8.24

$\mu = 40,\ \sigma = 2,\ n = 36$

36개를 임의 추출하였을 때 저항의 합이 1458옴 이상

⑦ $P\left(X > \frac{1458}{36}\right)$

$= P\left(X > \frac{81}{2}\right)$

$= P\left(Z > \frac{40.5-40}{\frac{2}{\sqrt{36}}}\right)$

$= P\left(Z > \frac{0.5}{\frac{1}{3}}\right)$

$= P(Z > 1.5)$

$= 1 - P(Z \leq 1.5) = 1 - 0.9332$

$= 0.0668$

OMNIBUS

#8.25

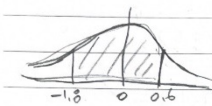
$\mu = 7,\ \sigma = 1$

(a) $n = 9$

⑦ $P(6.4 < X < 7.2)$

$= P\left(\frac{6.4-7}{\frac{1}{3}} < Z < \frac{7.2-7}{\frac{1}{3}}\right)$

$= P(-1.8 < Z < 0.6)$



0.2743
0.0359
0.3102

$P(Z < 1.8) = 0.9641$

$1 - P(Z < 1.8) = 1 - 0.9641$

$= 0.0359$

$P(Z < 0.6) = 0.7257$

$1 - P(Z < 0.6) = 0.2743$

$1 - 0.0359 - 0.2743$

$= 0.6898$

$\therefore 0.6898$

(b)

$\sigma_{\bar{x}} = \frac{1}{\sqrt{9}} = \frac{1}{3}$

$\bar{x} = \mu + Z \cdot \sigma_{\bar{x}}$

상위 15%에 해당하는 Z : 약 1.04

$\bar{x} = 7 + 1.04 + \frac{1}{3}$

$= 7.346\ldots$

3) 1.04 → .346.xx

9

14

12

20

약 7.35년

#8.26

$\mu = 3.2,\ \sigma = 1.6,\ n = 64$

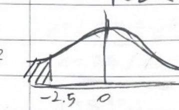
(a) $P(X < 2.7)$

$= P\left(Z < \frac{2.7-3.2}{\frac{1.6}{8}}\right)$

$= P\left(Z < \frac{-0.5 \times 8}{1.6}\right)$

$= P\left(Z < -\frac{5}{2}\right)$

$= P(Z < -2.5)$



$P(Z < 2.5) = 0.9938$

$1 - P(Z < 2.5) = 0.0062$

$\therefore 0.0062$

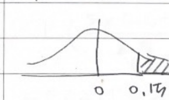
(b)

$P(X \geq 3.5)$

$= P\left(Z \geq \frac{3.5-3.2}{\frac{1.6}{8}}\right)$

$= P\left(Z \geq \frac{0.3 \times 8}{1.6}\right)$

$= P(Z \geq 0.15) = 0.5596$



$\therefore 0.5596$

(c)

$P(3.2 \leq X < 3.4)$

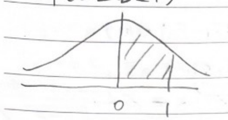
$= P\left(\frac{3.2-3.2}{\frac{1.6}{8}} \leq Z < \frac{3.4-3.2}{\frac{1.6}{8}}\right)$

$= P\left(0 \leq Z < \frac{0.2 \times 8}{1.6}\right)$

<「이공학도를 위한 확률 및 통계학」 p.271~272>

부록 #6 – 연습문제 (8.27~8.31)

$= P(0 \le Z < 1)$



$P(Z > 1) = 0.8413$

$P(Z \le 1) = 1 - 0.8413$

$= 0.1587$

$P(0 \le Z < 1) = 0.5 - 0.1587$

$= 0.3413$

$\therefore 0.3413$

\# 8.27

$\mu = 0.02g, \ \sigma = 0.10$

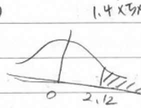
㉮ $P(X \ge 0.23)$

$= P\left(Z \ge \frac{0.23 - 0.2}{\frac{0.1}{\sqrt{50}}}\right)$

$= P\left(Z \ge \frac{0.03 \times 5\sqrt{2}}{0.1}\right)$

$= P(Z \ge 0.3 \times 5\sqrt{2})$

$= P(Z \ge 2.12)$



$P(Z < 2.12)$

$= 0.9830$

$P(Z \ge 2.12) = 1 - 0.9830$

$= 0.017$

$\therefore 0.017$

\# 8.28

① $\mu = 80, \ \sigma = 5, \ n_1 = 25$

② $\mu = 75, \ \sigma = 3, \ n_2 = 36$

㉮ $P(3.4 \le \bar{X}_1 - \bar{X}_2 < 5.9)$

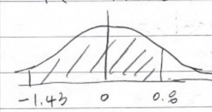
$= P\left(\frac{3.4 - (80-75)}{\sqrt{\frac{25}{25} + \frac{9}{36}}} \le Z < \frac{5.9 - (80-75)}{\sqrt{\frac{25}{25} + \frac{9}{36}}}\right)$

$= P\left(\frac{3.4 - 5}{\sqrt{\frac{5}{4}}} \le Z < \frac{0.9}{\sqrt{\frac{5}{4}}}\right)$

$= P\left(\frac{2 \cdot (-1.6)}{\sqrt{5}} \le Z < \frac{1.8}{\sqrt{5}}\right)$

$= P\left(-\frac{3.2}{\sqrt{5}} \le Z < \frac{1.8}{\sqrt{5}}\right)$

$= P(-1.43 \le Z < 0.8)$



$P(Z < 1.43) = 0.9236$

$P(Z < 0.8) = 0.7881$

$P(Z \ge 1.43) = 1 - 0.9236$

$= 0.0764$

$P(Z \ge 0.8) = 0.2119$

$1 - 0.0764 - 0.2119$

$= 0.9236 - 0.2119$

$= 0.7117$

$\therefore 0.7117$

\#8.29

티라미수: $\mu = 72, \ \sigma = 10, \ n = 64$

푸딩: $\mu = 28, \ \sigma = 5, \ n = 100$

㉮ $P(0 < \bar{X}_1 - \bar{X}_2 \le 44.2)$

$= P\left(\frac{0 - (72-28)}{\sqrt{\frac{100}{64} + \frac{25}{100}}} < Z \le \frac{44.2 - (72-28)}{\sqrt{\frac{100}{64} + \frac{25}{100}}}\right)$

OMNIBUS

$\sqrt{\frac{25}{16} + \frac{4}{16}} = \sqrt{\frac{29}{16}}$

$= P\left(-\frac{44}{\sqrt{29}} \times 4 < Z < \frac{0.2 \times 4}{\sqrt{29}}\right)$

$= P\left(-\frac{176}{\sqrt{29}} < Z < \frac{0.8}{\sqrt{29}}\right)$

$=$ 약 0.559

$\therefore 0.559$

\#8.30

$\mu = 540, \ \sigma = 50$

$n_1 = 32, \ n_2 = 50$

(a) ㉮ : $P(|\bar{X}_1 - \bar{X}_2| \ge 20)$

$P\left(\frac{\bar{X}_1 - \bar{X}_2 - (540-540)}{\sqrt{\frac{50^2}{32} + \frac{50^2}{50}}} \ge \frac{20 - (540-540)}{\sqrt{\frac{50^2}{32} + \frac{50^2}{50}}}\right)$

$= P\left(Z \ge \frac{20}{\sqrt{\frac{625}{8} + \frac{400}{8}}}\right)$

$= P\left(Z \ge \frac{20}{\sqrt{\frac{1025}{8}}}\right)$

$= P\left(Z \ge \frac{8\sqrt{2}}{\sqrt{41}}\right)$

1.414

6.4031

$= P(Z \ge 1.7666)$

$= 1 - 0.9616$

$= 0.0384$

$0.0384 \times 2 = 0.0768$

$\therefore$ 약 7.68%

(b)

㉮ $P(5 < |\bar{X}_1 - \bar{X}_2| < 10)$

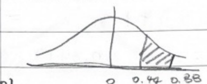
$P(5 < \bar{X}_1 - \bar{X}_2 < 10)$

$= P\left(\frac{5 - (540-540)}{\sqrt{\frac{1025}{8}}} < Z < \frac{10 - (540-540)}{\sqrt{\frac{1025}{8}}}\right)$

$= P\left(\frac{5 \times 2\sqrt{2}}{5\sqrt{41}} < Z < \frac{10 \times 2\sqrt{2}}{5\sqrt{41}}\right)$

$= P\left(\frac{2\sqrt{2}}{\sqrt{41}} < Z < \frac{4\sqrt{2}}{\sqrt{41}}\right)$

$= P(0.44 < Z < 0.88)$



$= 0.8106 - 0.6700$

$= 0.1406$

$0.1406 \times 2 = 0.2812$

$\therefore 0.2812$

\#8.31

$n_1 = n_2 = 18, \ \sigma_1 = \sigma_2 = 1.0$

(a) $P(\bar{X}_A - \bar{X}_B \ge 1.0)$

$= P\left(\frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{1}{18} + \frac{1}{18}}} \ge \frac{1.0 - 0}{\sqrt{\frac{1}{18} + \frac{1}{18}}}\right)$

$= P(Z \ge 3)$

$= 1 - 0.9987$

$= 0.0013$

평균건조시간의 차이가 1.0보다 클

확률은 0.0013

(b) $10{,}000 \times 0.0013$

$= 13$

$\therefore 13$

<「이공학도를 위한 확률 및 통계학」 p.272>

부록 #7 – 연습문제 (8.32~8.40)

#8.32

$n_A = n_B = 36$

$\bar{x}_A = 4.5,\ \bar{x}_B = 4.7$

(a) $P\left(\dfrac{\bar{X}_B - \bar{X}_A - (\mu_B - \mu_A)}{\sqrt{\frac{1}{36} + \frac{1}{36}}} \geq \dfrac{0.2 - (\mu_B - \mu_A)}{\sqrt{\frac{1}{36} + \frac{1}{36}}}\right)$

$= P\left(Z \geq \dfrac{0.2}{\sqrt{\frac{1}{18}}}\right)$

$= P(Z \geq 0.2 \times 3\sqrt{2})$

$= P(Z \geq 0.85)$

$= 1 - 0.8023$

$= 0.1977$

$\therefore 19.77\%$

(b)

(a)에서의 결과는 19.77%이므로, 두 기계의 모평균이 다르다는 것을 강력하게 뒷받침한다

#8.33

$n = 25,\ \bar{x} = 7960,\ \sigma = 100$

(a) ⑦ $P\left(Z \geq \dfrac{7960 - 7950}{\frac{100}{5}}\right)$

$= P\left(Z \geq \dfrac{10}{20}\right) = P(Z \geq 0.5)$

$= 1 - 0.6915$

$= 0.3085$

$\therefore 30.85\%$

(b)

(a)에서 확률이 30.85%로 도출되었으므로, 모표준편차 정부의 규제한 제곱

초과한다는 증거가 된다.

#8.34

$n_A = n_B = 30$

$\sigma_A = \sigma_B = 5$

(a) $P(\bar{X}_A - \bar{X}_B > 4)$

$= P\left(\dfrac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{25}{30} + \frac{25}{30}}} > \dfrac{4 - (\mu_A - \mu_B)}{\sqrt{\frac{25}{30} + \frac{25}{30}}}\right)$

$\frac{45}{30} = \frac{3}{2}$

$= P\left(Z > \dfrac{4}{\sqrt{\frac{3}{2}}}\right)$

$= P\left(Z > \dfrac{4}{1.224}\right)$

$= P(Z > 3.27)$

$= 1 - 0.9995$

$= 0.0005$

즉, $\bar{X}_A - \bar{X}_B > 4$의 확률은 0.0005으로 매우 작으므로 B의 주장은 틀렸다.

(b)

$P = 0.005$는 주어진 결과가 우연히 발생할 확률이 0.5%임을 의미한다

이러한 유의확률은 합금 A와 B의 성능이 동일하다는 가설을 지지하는 증거가 얼마나 약한지를 나타낸다

따라서, 실험결과는 합금 A의 우수성을 강력하게 뒷받침한다고 할 수 있다.

#8.35

$\mu = 800 \rightarrow 760,\ \sigma = 40,\ n = 16$

$P(\bar{X} < 775)$

$= P\left(Z < \dfrac{775 - 760}{\frac{40}{4}}\right)$

$= P(Z < 1.5)$

$= 0.9332$

$\therefore 93.32\%$

#8.36

중심극한정리는 독립적이고 동일하게 분포된 무작위 표본의 합이 표본의 크기가 충분히 크면 정규분포에 근사한다고 설명한다.

① $Y = X_1 X_2 \cdots X_n$

$\rightarrow \log(Y) = \log(X_1 X_2 \cdots X_n)$

$= \log(X_1) + \log(X_2) + \cdots + \log(X_n)$

n개의 독립적인 $\log(X_i)$ 값들의 합

② $\log(X_i)$ 값들은 독립적이며 동일한 분포를 따르므로 n이 충분히 클 때, 중심극한정리에 따라 $\log(X_1) + \log(X_2) + \cdots + \log(X_n)$의 분포는 근사적으로 정규분포를 따른다.

③ 변수 Z가 정규분포를 따를 때, e^Z는 로그정규분포를 따른다. $\log(Y)$가 근사적으로 정규분포를 따른다면, $Y = e^{\log(Y)}$는 로그정규분포를 따른다

#8.37

(a) 27.488

(b) 18.475

(c) 36.415

#8.38

(a) 16.750

(b) 30.144

(c) 26.217

#8.39

(a) $\chi_\alpha^2 = 0.297$

(b) $\chi_\alpha^2 = 32.852$

(c) $37.652 = \chi_{0.05}$

$0.05 - 0.045 = 0.005$

$\Rightarrow \alpha = 0.005$

$\chi^2_{0.005} = 46.928$

#8.40

0.01

χ_α^2

(a) $\chi_{0.01}^2 = 38.932$

(b)

0.95

0.05

χ_α^2

$\chi_{0.05}^2 = 12.592$

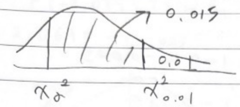
OMNIBUS

<「이공학도를 위한 확률 및 통계학」 p.272~273, p.285~286>

부록 #8 – 연습문제 (8.41~8.46)

(c)

$v=10$일때 $27.209=\chi^2_{0.01}$



$\alpha = 0.015 + 0.01$

$= 0.025$

$\therefore \chi_{0.025}^2 = 20.483$

$\therefore 20.483$

#8.41

$\sigma^2=6$, $n=25$

(a)

㉮ $P(S^2>9.1)$

$= P(\frac{S^2(n-1)}{\sigma^2} > \frac{9.1(n-1)}{\sigma^2})$

$= P(Y > \frac{9.1 \times \cancel{24}^{4}}{\cancel{6}})$

$= P(Y > 36.4)$

$36.4 = \chi^2_{0.05}$

$\therefore 0.05$

(b)

㉮ $P(3.462 < S^2 < 10.745)$

$= P(\frac{3.462(n-1)}{\sigma^2} < \frac{(n-1)S^2}{\sigma^2} < \frac{10.745(n-1)}{\sigma^2})$

$= P(\frac{3.462}{6} \times 24 < Y < \frac{10.745 \times 24}{6})$

$= P(13.848 < Y < 42.98)$

$13.848 = \chi^2_{0.95}$, $42.98 = \chi^2_{0.01}$

$= 0.95 - 0.01 = 0.94$

$\therefore 0.94$

#8.42

$\mu=74$, $\sigma^2=8$ / $n=20$, $s^2=20$

카이제곱검정 → 자유도 $n-1 =$ 19

if) 계산된 통계값이 임계값보다 크면
표본분산이 모집단 분산과 통계적으로
유의미하게 다르다고 볼 수 있음

자유도가 19인 카이제곱 분포에서의
임계값 : 30.14

카이제곱 통계값 : $\frac{(n-1)s^2}{\sigma^2} = \frac{19 \times \cancel{20}^{5}}{\cancel{8}_{2}}$

$= 47.5$

해당값은 임계값보다 크므로 과거처럼
$\sigma^2=8$이라고 여길 수 없음

#8.43

S^2 = 표본분산

$\frac{(n-1)S^2}{\sigma^2}$의 분산으로부터 S^2의 분산 도출

→ $Var((n-1)S^2) = \sigma^4 \cdot 2(n-1)$

→ $Var(S^2) = \frac{\sigma^4 2(n-1)}{(n-1)^2}$

$= \frac{\sigma^4 \cdot 2}{(n-1)}$

따라서, S^2의 분산은 n이 커질수록
감소함을 파악할 수 있다

⇒ 이는 표본 크기 n이 커질수록
표본분산 S^2이 모집단 분산 σ^2에
더 가까워진다는 것.

OMNIBUS

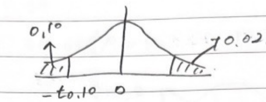
#8.44

(a) 2.145

(b) $-t_{0.10} = -1.345$

(c) t분포는 원점을 중심으로 대칭이므로,
$t_{1-\alpha} = -t_\alpha$ 이다.

따라서, $t_{0.995} = -t_{0.005} = -3.499$



따라서, $1-0.10-0.025$

$= 1-0.125$

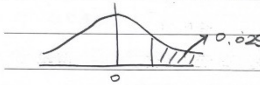
$= 0.875$

#8.45

(a) $P(T<2.365)$

$2.365 = t_{0.025}$

즉, $P(T<t_{0.025})$

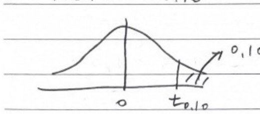


따라서, $1-0.025=0.975$

$\therefore 0.975$

(b) $P(T>1.318)$

$1.318 = t_{0.10}$



$\therefore 0.10$

(c)

$P(-1.356 < T < 2.179)$

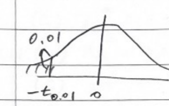
$-1.356 = -t_{0.10}$

$2.179 = t_{0.025}$

(d)

$P(T > -2.567)$

$-2.567 = -t_{0.01}$

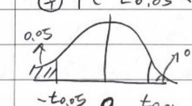


따라서, $1-0.01=0.99$

#8.46

(a)

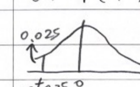
㉮ $P(-t_{0.05} < T < t_{0.01})$



$1-0.05-0.01 = 1-0.06$

$= 0.94$

(b) $P(T > -t_{0.025})$



$1-0.025=0.975$

$\therefore 0.975$

<「이공학도를 위한 확률 및 통계학」 p.286>

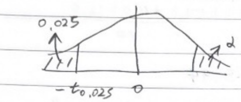
부록 #9 – 연습문제 (8.47~8.53)

#8.47

$n=24 \rightarrow v=23$

(a) $P(-2.069<T<K)=0.965$

$-2.069=-t_{0.025}$



$1-0.025-\alpha=0.965$

$0.975-\alpha=0.965$

$\alpha=0.01$

$K=t_{0.01}=2.500$

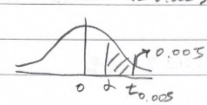
∴ 2.500

(b)

$P(K<T<2.807)=0.095$

$2.807=t_{0.005}$

$P(K<T<t_{0.005})=0.095$



$\alpha-0.005=0.095$

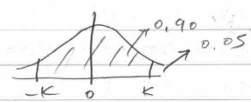
$\alpha=0.1$

$K=t_{0.1}=1.318$

∴ 1.318

(c)

$P(-K<T<K)=0.90$



$t_\alpha=t_{0.05}$

$t_{0.05}=1.714$

∴ 1.714

#8.48

$\mu=30,\ \bar{x}=27.5,\ s=5,\ n=16$

$\alpha=0.05$, 자유도 : 15

$$t=\frac{\bar{x}-\mu}{s/\sqrt{n}}=\frac{27.5-30}{\frac{5}{\sqrt{16}}}=\frac{-2.5\times4}{5}$$

$=-2$

$-t_{0.025}=-2.131$

$t_{0.025}=2.131$

$-t_{0.025}<t<t_{0.025}$를 만족하므로

평균수명 요구 만족시키고 있다고 볼 수 있음

#4.49

$\mu=20,\ \bar{x}=24,\ s=4.1,\ n=9$

$$z=\frac{\bar{x}-\mu}{s/\sqrt{n}}=\frac{24-20}{4.1/\sqrt{9}}\fallingdotseq 2.93$$

이 값은 z분포에서 상위 1% 혹은 상위 5%의 꼬리 부분에 위치한다 따라서, 이 모집단으로부터 평균이 24시간이고, 표준편차가 4.1인 크기 9인 확률분포를 추출할 확률은 매우 낮다 따라서 평균에 대해 20시간이라고 한 가정이 틀렸으며, 모집단의 실제 평균이 20시간보다 훨씬 높을 수 있다는 결론을 내릴 수 있다.

OMNIBUS

#8.50

$$\bar{x}=\frac{0.6+0.7+0.7+0.3+0.4+0.5+0.4+0.2}{8}$$

$=0.475g$

$s=0.183g$,

$$t=\frac{\bar{x}-\mu}{s/\sqrt{n}}=\frac{0.475-0.5}{0.183/\sqrt{8}}\fallingdotseq-0.39$$

t 검정통계량을 해석할 때, 일반적으로 사용되는 유의수준(α)은 0.05 → 계산된 t 검정 통계량의 절대값이 작으며, 이는 표본 데이터가 모평균 0.5g과 통계적으로 유의미한 차이가 없음을 시사함

#8.51

(a) $f_{0.05}(7,15)=2.71$

(b) $f_{0.05}(15,7)=3.51$

(c) $f_{0.01}(24,19)=2.92$

(d) $f_{0.95}(19,24)=\dfrac{1}{f_{0.05}(24,19)}=\dfrac{1}{2.11}$

(e) $f_{0.99}(28,12)=\dfrac{1}{f_{0.01}(12,28)}=\dfrac{1}{2.90}$

#8.52

모분산이 같은지 검증하기 위해 F검정을 사용할 수 있다

첫번째 데이터세트

평균 : $\dfrac{19.8+12.7+13.2+16.9+10.6+8.8+11.1+14.7+17.0+12.5}{10}$

$=25.94$

→ 표본분산 : 10.44 / 자유도 9

두번째 $s^2=1.85$ / 자유도 7

⇒ F 검증 통계량 $=\dfrac{10.44}{1.85}\fallingdotseq 5.643$

∴유의 수준에서 유의미한 차이가 있음

#8.53

광산 1 $s^2=15.750$ / $v_1=4$

광산 2 $s^2=10.920$ / $v_2=5$

F 검증 통계량 $=\dfrac{15.750}{10.920}\fallingdotseq 1.44$

유의수준에 위치

∵ 모분산이 같다고 볼 수 있음

<「이공학도를 위한 확률 및 통계학」 p.286~287>

부록 #10 – MATLAB 코드 (1/24)

- 그림 8.2 (1/2)

```
figure;
hold on;

colors = ['r', 'g', 'b'];

x_range = linspace(-5, 5, 1000);
std_normal_pdf = (1 / sqrt(2 * pi)) * exp(-x_range.^2 / 2);

plot(x_range, std_normal_pdf, 'k-', 'LineWidth', 2);

for i = 1:length(n_values)
    [counts, edges] = histcounts(normalized_sample_means(:, i), 'Normalization', 'pdf', 'BinMet
hod', 'auto');
    binCenters = (edges(1:end-1) + edges(2:end)) / 2;
    plot(binCenters, counts, colors(i), 'LineWidth', 2);
end
```

- 그림 8.2 (2/2)

```
xlim([-5, 5]);
xlabel('Z');
ylabel('Probability Density');
legend({'Standard Normal', 'n = 1', 'n = 20', 'n = 1000'}, 'Location', 'Northeast');
grid on;
hold off;
```

- 그림 8.3

```
mu = 800;
sigma = 10;
x = 770:1:830;

pdf_vals = (1 / (sigma * sqrt(2 * pi))) * exp(-((x - mu).^2 / (2 * sigma^2)));

figure;
plot(x, pdf_vals, 'LineWidth', 2);
xlabel('$\bar{x}$', 'Interpreter', 'latex', 'FontSize', 25);
ylabel('확률밀도함수(PDF)', 'Interpreter', 'latex', 'FontSize', 15);
hold on;

idx = x <= 775;
fill_x = [x(idx), fliplr(x(idx))];
fill_y = [zeros(1, sum(idx)), fliplr(pdf_vals(idx))];
fill(fill_x, fill_y, 'b', 'FaceAlpha', 0.3);

hold off;
```

부록 #13 – MATLAB 코드 (4/24)

• 그림 8.4

```
mu = 28.0;
sigma = sqrt(10) / 4;
x = 25:0.1:31;

pdf_vals = (1 / (sigma * sqrt(2 * pi))) * exp(-((x - mu).^2 / (2 * sigma^2)));

figure;
plot(x, pdf_vals, 'LineWidth', 2);
xlabel('$\bar{x}$', 'Interpreter', 'latex', 'FontSize', 25);
ylabel('확률밀도함수(PDF)', 'Interpreter', 'latex', 'FontSize', 15);
hold on;

idx = x >= 30.5;
fill_x = [x(idx), fliplr(x(idx))];
fill_y = [zeros(1, sum(idx)), fliplr(pdf_vals(idx))];
fill(fill_x, fill_y, 'b', 'FaceAlpha', 0.3);

hold off;
```

부록 #14 – MATLAB 코드 (5/24)

- 그림 8.6

```
mu = 0.5;
sigma = 0.189;
x = -0.2:0.01:1.2;

pdf_vals = (1 / (sigma * sqrt(2 * pi))) * exp(-((x - mu).^2 / (2 * sigma^2)));

figure;
plot(x, pdf_vals, 'LineWidth', 2);
xlabel('$\bar{x_1}-\bar{x_2}$', 'Interpreter', 'latex', 'FontSize', 25);
ylabel('확률밀도함수(PDF)', 'Interpreter', 'latex', 'FontSize', 15);
hold on;

idx = x >= 1.0;
fill_x = [x(idx), fliplr(x(idx))];
fill_y = [zeros(1, sum(idx)), fliplr(pdf_vals(idx))];
fill(fill_x, fill_y, 'b', 'FaceAlpha', 0.3);

hold off;
```

부록 #15 – MATLAB 코드 (6/24)

- 그림 8.7 (1/2)

```
alpha = 0.05;
v = 10;

x = 0:0.01:30;
pdf_vals = x.^(v/2-1) .* exp(-x/2) ./ (2^(v/2) * gamma(v/2));

figure;
plot(x, pdf_vals, 'k', 'LineWidth', 3);

hold on;
area_under_curve = cumsum(pdf_vals) * 0.01;
chi2_val = x(find(area_under_curve >= (1 - alpha), 1, 'first'));
```

부록 #16 – MATLAB 코드 (7/24)

- 그림 8.7 (2/2)

```
idx = x >= chi2_val;
fill_x = [x(idx), fliplr(x(idx))];
fill_y = [zeros(1, sum(idx)), fliplr(pdf_vals(idx))];
fill(fill_x, fill_y, [0.6784, 0.8471, 0.9020]);

line([chi2_val chi2_val], [0 max(pdf_vals(idx))], 'Color', 'r', 'LineStyle', '--');

hold off;
```

부록 #17 – MATLAB 코드 (8/24)

- 그림 8.8 (1/2)

```
x = -10:0.01:10;

v1 = 2;
v2 = 5;
v3 = 1e6;

gamma_v1 = gamma((v1+1)/2) / (sqrt(v1*pi)*gamma(v1/2));
gamma_v2 = gamma((v2+1)/2) / (sqrt(v2*pi)*gamma(v2/2));

pdf_v1 = gamma_v1 * (1 + x.^2/v1).^(-(v1+1)/2);
pdf_v2 = gamma_v2 * (1 + x.^2/v2).^(-(v2+1)/2);

pdf_normal = (1 / (sqrt(2*pi))) * exp(-0.5 * x.^2);
```

부록 #18 – MATLAB 코드 (9/24)

- 그림 8.8 (2/2)

```
figure;
p1 = plot(x, pdf_v1, 'Color', [1, 0, 0], 'LineWidth', 2);
hold on;
p2 = plot(x, pdf_v2, 'Color', [0, 0, 1], 'LineWidth', 2);
p3 = plot(x, pdf_normal, 'Color', [0, 0.5, 0], 'LineStyle', '--', 'LineWidth', 2);

legend([p1, p2, p3], {'v=2', 'v=5', 'Normal Distribution'}, 'Location', 'NorthEast');

xlabel('t');
ylabel('f(t)');

hold off;
```

부록 #19 – MATLAB 코드 (10/24)

- 그림 8.9 (1/2)

```
v = 10;
t_symmetry_point = 1;

x = -5:0.01:5;

pdf_vals = gamma((v+1)/2) ./ (sqrt(v*pi) .* gamma(v/2)) .* (1 + (x.^2) ./ v).^(-(v+1)/2);

figure;
plot(x, pdf_vals, 'k', 'LineWidth', 1.5);
hold on;

skyBlue = [0.3010, 0.7450, 0.9330];

x_fill_left = x(x < -t_symmetry_point);
y_fill_left = pdf_vals(x < -t_symmetry_point);
fill([x_fill_left, fliplr(x_fill_left)], [y_fill_left, zeros(1, length(y_fill_left))], skyBlue, 'EdgeColor', 'no
ne', 'FaceAlpha', 0.5);
```

부록 #20 – MATLAB 코드 (11/24)

- 그림 8.9 (2/2)

```
x_fill_right = x(x > t_symmetry_point);
y_fill_right = pdf_vals(x > t_symmetry_point);
fill([x_fill_right, fliplr(x_fill_right)], [y_fill_right, zeros(1, length(y_fill_right))], skyBlue, 'EdgeColor
', 'none', 'FaceAlpha', 0.5);

plot(-t_symmetry_point, gamma((v+1)/2) / (sqrt(v*pi) * gamma(v/2)) * (1 + (t_symmetry_point.
^2) / v)^(-(v+1)/2), 'ro', 'MarkerFaceColor', 'r');
plot(t_symmetry_point, gamma((v+1)/2) / (sqrt(v*pi) * gamma(v/2)) * (1 + (t_symmetry_point.^
2) / v)^(-(v+1)/2), 'ro', 'MarkerFaceColor', 'r');

xlabel('t');
ylabel('Probability Density');
title(sprintf('t-Distribution with v=%d', v));
legend('t-Distribution PDF', 'Location', 'NorthEast');

hold off;
```

부록 #21 – MATLAB 코드 (12/24)

- 그림 8.10 (1/2)

```
df = 14;

t_pdf = @(t, df) gamma((df+1)/2) / (sqrt(df*pi) * gamma(df/2)) * (1 + t.^2 / df).^(-(df+1)/2);

t = linspace(-4, 4, 1000);

pdf_values = t_pdf(t, df);

figure;
plot(t, pdf_values, 'LineWidth', 2);
hold on;

t_approx_left = -2.1;
t_approx_right = 2.1;
```

부록 #22 – MATLAB 코드 (13/24)

- 그림 8.10 (2/2)

```
fillAreaLeft = [t(t <= t_approx_left), t_approx_left];
pdfFillLeft = [t_pdf(t(t <= t_approx_left), df), 0];

fill(fillAreaLeft, pdfFillLeft, 'c');

plot([t_approx_left, t_approx_right], [t_pdf(t_approx_left, df);

xlabel('t value');
ylabel('Probability Density');
grid on;
hold off;
```

- 그림 8.11 (1/2)

```
df = 14;
t_pdf = @(t, df) gamma((df+1)/2) / (sqrt(df*pi) * gamma(df/2)) * (1 + t.^2 / df).^(-(df+1)/2);
t = linspace(-4, 4, 1000);
pdf_values = t_pdf(t, df);
figure;
plot(t, pdf_values, 'LineWidth', 2);
hold on;
t_approx_left = -2.1;
t_approx_right = 2.0;
```

부록 #24 – MATLAB 코드 (15/24)

- 그림 8.11 (2/2)

```
fillArea = t(t >= t_approx_left & t <= t_approx_right);
pdfFill = t_pdf(fillArea, df);
fill([fillArea(1), fillArea, fillArea(end)], [0, pdfFill, 0], 'c', 'EdgeColor', 'none');
plot([t_approx_left, t_approx_right], [t_pdf(t_approx_left, df), t_pdf(t_approx_right, df)], 'ro', 'Ma
rkerFaceColor','r');

xlabel('t value');
ylabel('Probability Density');

grid on;
hold off;
```

부록 #25 – MATLAB 코드 (16/24)

- 그림 8.12 (1/2)

```
gammaFcn = @(x) exp(gammaln(x));
pdf_t = @(x, df) gammaFcn((df+1)/2)/(sqrt(df*pi)*gammaFcn(df/2)) * (1+x.^2/df).^(-(df+1)/2);
t_range = linspace(-4, 4, 1000);
t_value = -1.761;
k_value = -2.5;

figure;
plot(t_range, pdf_t(t_range, 14), 'LineWidth', 2);
hold on;
```

- 그림 8.12 (2/2)

```
fill_idx = t_range >= k_value & t_range <= t_value;
fill_area_x = [t_range(fill_idx) t_value k_value];
fill_area_y = [pdf_t(t_range(fill_idx), 14) 0 0];
fill(fill_area_x, fill_area_y, [0.6784, 0.8471, 0.9020]);

plot([t_value t_value], [0 pdf_t(t_value, 14)], 'Color', 'red', 'LineWidth', 2);
plot([k_value k_value], [0 pdf_t(k_value, 14)], 'Color', 'blue', 'LineWidth', 2);

xlabel('t');
ylabel('Probability Density Function');
hold off;
```

- 그림 8.13

```
df = 24;
x = -10:0.01:10;
y = gamma((df+1)/2) ./ (sqrt(df*pi) .* gamma(df/2)) .* (1 + x.^2 ./ df).^(-(df+1)/2);
figure;
plot(x, y, 'LineWidth', 3)
hold on;
t0_05 = 1.711;
y_t0_05 = gamma((df+1)/2) ./ (sqrt(df*pi) .* gamma(df/2)) .* (1 + t0_05^2 ./ df).^(-(df+1)/2);
plot([-t0_05, -t0_05], [0, y_t0_05], 'r', 'LineWidth', 3);
plot([t0_05, t0_05], [0, y_t0_05], 'r', 'LineWidth', 3);
xlabel('t value');
ylabel('Probability Density');
title('t-Distribution with df=24');
hold off;
```

부록 #28 – MATLAB 코드 (19/24)

- 그림 8.15

```
df1 = 10;
df2 = 20;
f_alpha = 2.0;
pdf_f = @(x, df1, df2) sqrt(((df1 * x).^df1 .* df2^df2) ./ ((df1 * x + df2).^(df1 + df2))) ...
    ./ (x .* beta(df1/2, df2/2));
f_values = linspace(0.01, 5, 1000);
pdf_values = pdf_f(f_values, df1, df2);
figure;
plot(f_values, pdf_values, 'LineWidth', 2);
hold on;
fillIdx = f_values >= f_alpha;
h = fill([f_values(fillIdx) f_alpha], [pdf_values(fillIdx) 0], 'c');
set(h, 'EdgeColor', 'none');
line([f_alpha, f_alpha], [0, pdf_f(f_alpha, df1, df2)], 'Color', 'red', 'LineWidth', 2);
xlabel('f value');
ylabel('Probability Density');

hold off;
```

부록 #29 – MATLAB 코드 (20/24)

- 그림 8.16

```
data = [1, 2, 3, 4, 5];

q0 = quantile(data, 0);
q25 = quantile(data, 0.25);
q50 = quantile(data, 0.5);
q75 = quantile(data, 0.75);
q100 = quantile(data, 1);

figure;
plot(data, zeros(size(data)), 'bo', 'MarkerFaceColor', 'b');
hold on;
plot([q0 q25 q50 q75 q100], zeros(1,5), 'r*', 'MarkerSize', 10);

xlabel('데이터 값');
xlim([0 6]);

set(gca, 'YColor', 'none');

hold off;
```

부록 #30 – MATLAB 코드 (21/24)

- 그림 8.17

```
data = [10, 15, 20, 25, 30];
n = length(data);

figure;
hold on;

line([0, 0.2], [data(1), data(1)], 'Color', 'b', 'LineWidth', 5);
line([0.2, 0.4], [data(2), data(2)], 'Color', 'b', 'LineWidth', 5);
line([0.4, 0.6], [data(3), data(3)], 'Color', 'b', 'LineWidth', 5);
line([0.6, 0.8], [data(4), data(4)], 'Color', 'b', 'LineWidth', 5);
line([0.8, 1.0], [data(5), data(5)], 'Color', 'b', 'LineWidth', 5);

xlabel('Quantile Fraction (f)');
ylabel('Data Value');
title('Quantile Plot');
grid on;
hold off;
```

부록 #31 – MATLAB 코드 (22/24)

- 그림 8.18

```
data = randn(100, 1);
sortedData = sort(data);
n = length(data);
p = ((1:n)' - 0.5) / n;
z = sqrt(2) * erfinv(2*p - 1);
plot(z, sortedData, 'bo');
hold on;
plot(z, z, 'r-');
xlabel('비율, \it{f}');
ylabel('분위수');
legend('Data Points', 'Normal Distribution Line', 'Location', 'Best');
grid on;
```

부록 #32 – MATLAB 코드 (23/24)

- 그림 8.19

```
data = rand(1, 1000) * 100;
sorted_data = sort(data);
p = ((1:1000) - 0.5) / 1000;
normal_scores = sqrt(2) * erfinv(2*p - 1);
figure;
plot(sorted_data, normal_scores, '.');
title('Normal Probability Plot');
xlabel('Sample value');
ylabel('Normal score');
```

부록 #33 – MATLAB 코드 (24/24)

- 그림 8.20

```
data = randn(1, 1000);
sorted_data = sort(data);
p = ((1:1000) - 0.5) / 1000;
normal_scores = sqrt(2) * erfinv(2*p - 1);
figure;
plot(sorted_data, normal_scores, '.');
title('Normal Probability Plot');
xlabel('Sample value');
ylabel('Normal scale');
```