2026/10/16, 2026 확률 기초 세미나

이공학도를 위한 확률 및 통계학

-8장 확률분포와 표본분포-

서 진 우(jinwoo@pel.sejong.ac.kr)

세종대학교 프로토콜공학연구실

목 차

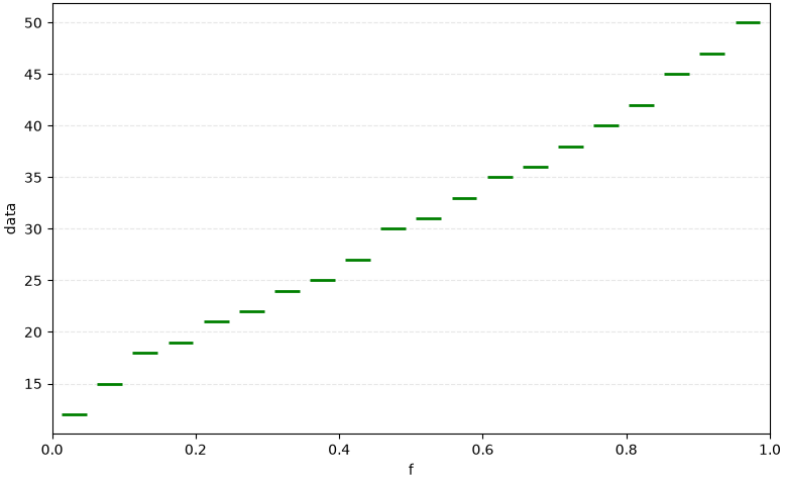
- 보충
- 확률표본
- 대표적 통계량
- 표본분포
- t 분포
- F 분포
- 분위수와 확률 그림

목 차

- 보충
- 확률표본
- 대표적 통계량
- 표본분포
- t 분포
- F 분포
- 분위수와 확률 그림

보충

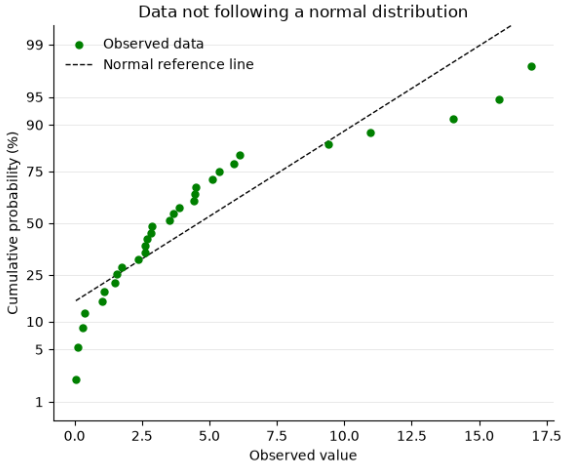
- 분위수그림(Quantile Plot)
  - 특징
    - x축은 누적비율 f, y축은 정렬된 관측값을 표시
    - 데이터를 오름차순으로 정렬 후, i번째 값에 비율 f_i를 계산하여 표시
      - $f_i = \frac{i-3/8}{n+1/4}$, n=조사한 데이터 수, $i = 1,2,\dots,n$
        - 기본형 $\frac{i}{n+1}$을 정규분포에 맞게 보정한 식
        - $\frac{3}{8}$은 정규분포에서 i번째 값의 평균 위치를 근사하는 보정값으로 수치적 절충값
        - $\frac{1}{4}$는 최솟값과 최댓값에 대응하는 누적확률이 0.5를 기준으로 대칭이 되도록 결정



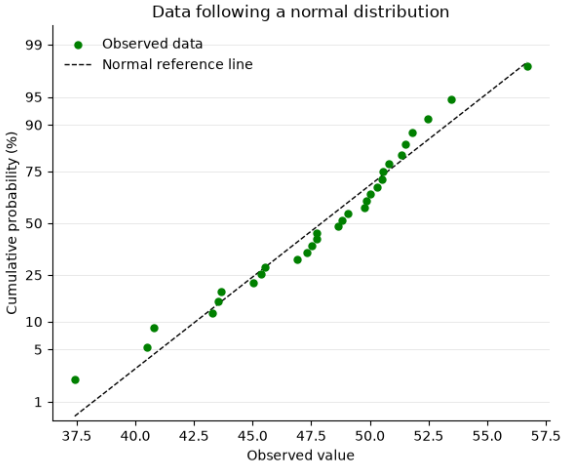
<그림 11> 분위수그림

분위수와 확률 그림

- 정규확률그림(Normal Probability Plot)
  - 특징
    - x축은 정렬된 관측값, y축은 누적확률 $f_i(\%)$
      - $f_i = \frac{i-3/8}{n+1/4}$ 으로 분위수그림과 동일하며 눈금은 $q_{0,1}(f_i)$에 배치
    - 기준선은 표본평균과 표본표준편차를 이용해 계산한 정규분포의 이론적 분위수를 나타내는 직선
      - $y = \bar{x} + s \cdot z$
        - $\bar{x}$는 표본평균. s는 표본표준편차,z는 표준정규 이론분위수



<그림 13> 정규분포를 따르지 않는 정규확률그림



<그림 14> 정규분포를 따르는 정규확률그림

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- t 분포
- F 분포
- 분위수와 확률 그림

확률표본

- 관련 용어
  - 모집단(Population)
    - 정의
      - 관심이 있는 대상과 관련된 모든 관측가능한 값의 집합
    - 특징
      - 범위가 넓어 현실적으로 전체를 전부 조사하기 어려움
        - e.g., 대한민국에 거주하는 20대 남성
  - 표본(Sample)
    - 정의
      - 모집단에서 조사 대상으로 추출한 일부 관측 값의 집합
    - 특징
      - 모집단에서 일부를 추출하여 모집단의 성질을 추측
        - e.g., 대한민국에 거주하는 20대 남성 중 무작위로 뽑은 1,000명

확률표본

- 확률표본
  - 정의
    - 모집단에 속한 표본점들이 표본으로 추출될 확률이 모두 동일하다는 조건 하에서 추출된 표본
  - 필요성
    - 표본 추출 시, 발생할 수 있는 편향(Bias)을 예방하기 위해 확률표본 선택
  - 특징
    - 추출된 표본들은 상호 독립적이고 동일한 모집단에서 추출되었으므로 동일한 분포를 가짐
    - 서로 독립인 n개의 확률변수 $X_1, X_2, \dots, X_n$이 동일한 확률분포 $f(x)$를 따를 때, $X_1, X_2, \dots, X_n$을 모집단 $f(x)$로부터의 크기 n인 확률표본이라고 부르고, 결합확률분포는 $f(x_1, x_2, \dots x_n) = f(x_1)f(x_2)\cdots f(x_n)$가 됨

*편향(Bias)
과대추정 및 과소추정을 하여 추정량이 모수의 실제 값에서 벗어나는 현상.
*표본점(Sample Point)
모집단에서 하나의 관측 값을 나타내는 개별 대상.

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- t 분포
- F 분포
- 분위수와 확률 그림

대표적 통계량

- 통계량(Statistic)
  - 정의
    - 확률표본을 구성하는 확률변수들의 함수
  - 예시
    - 특정 상표의 커피에 대한 선호도 조사
      - 모든 커피 애호가들을 대상으로 전부 조사하는 것은 불가능
      - 대표본(Large Sample)을 추출하여 특정 커피의 선호도 p에 대한 정보 $\hat{p}$를 조사
        - $\hat{p}$는 참 값인 P를 추론하는데 사용
        - $\hat{p}$는 확률표본의 측정된 값의 함수이므로 표본에 따라 그 값이 달라짐
        - $\hat{p}$는 확률변수 P에 대한 실제 값으로 표본에 따라 달라지는 값
      - 이와 같은 확률변수를 통계량이라 부름

*대표본(Large Sample)
표본의 크기(n)이 30 이상인 표본.

대표적 통계량

- 표본의 중심위치
  - 정의
    - 데이터의 분포에서 중심이 되는 위치를 나타내는 대푯값
  - 종류
    - 평균(Mean)
      - 모든 값을 더한 뒤 개수로 나눈 값
      - $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
    - 중앙값(Median)
      - 크기순으로 나열 시 정가운데에 있는 값
      - $\tilde{x} = \begin{cases} \frac{x_{n+1}}{2}, & n\text{이 홀수일 경우} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\left(\frac{n}{2}\right)+1}\right), & n\text{이 짝수일 경우} \end{cases}$
    - 최빈값(Mode)
      - 가장 많이 발생한 표본값

대표적 통계량

- 표본의 중심위치
  - 예제 8.1

관측값들이 다음과 같다고 하자.

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

이때 최빈값는 어떤 값인가?

- 풀이
  - 0.43이 가장 많이 등장

$\therefore 0.43$

대표적 통계량

- 표본의 산포(Variability) (1/5)
  - 정의
    - 데이터가 중심위치로부터 퍼짐 정도를 나타내는 척도
  - 종류 (1/5)
    - 표본분산(Sample Variance)
    - 표본표준편차(Sample Standard Deviation)
    - 표본범위(Sample Range)

대표적 통계량

- 표본의 산포(Variability) (2/5)
  - 종류 (2/5)
    - 표본분산(Sample Variance) (1/2)
      - 모집단에서 추출한 표본의 흩어진 정도를 나타내는 통계량
      - $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
      - S^2이 크기가 n인 확률표본의 분산일 경우 아래와 같이 표기 가능
        - $S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n}X_i^2 - (\sum_{i=1}^{n}X_i)^2\right]$

정의에 의하여

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

$\bar{X}$를 $\sum_{i=1}^{n}\frac{X_i}{n}$로 치환하고 분모, 분자에 n을 곱하면 다음 식을 얻을 수 있음

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

대표적 통계량

- 표본의 산포(Variability) (3/5)
  - 종류 (3/5)
    - 표본분산(Sample Variance) (2/2)
      - 예제 8.2

샌디에이고에 있는 상점 중 임의로 4곳을 선택하여 200g짜리 병커피의 가격을 비교해 본 결과 지난 달보다 12, 15, 17, 20센트씩 인상되었다. 가격인상에 대한 이 확률분포의 분산을 구하라.

- 풀이
  - 표본평균
    - $\bar{x} = \frac{x_1+x_2+x_3+x_4}{4} = \frac{12+15+17+20}{4} = 16$센트
  - $s^2 = \frac{1}{3}\sum_{i=1}^{4}(x_i - 16)^2 = \frac{(12-16)^2+(15-16)^2+(17-16)^2+(20-16)^2}{3} = \frac{(-4)^2+(-1)^2+(1)^2+(4)^2}{3} = \frac{34}{3}$

  $\therefore \frac{34}{3}$

대표적 통계량

- 표본의 산포(Variability) (4/5)
  - 종류 (4/5)
    - 표본표준편차(Sample Standard Deviation)
      - 모집단에서 추출한 표본이 표본평균을 중심으로 얼마나 퍼져 있는지를 나타내는 측정값
      - $S = \sqrt{S^2}$
    - 표본범위(Sample Range)
      - 모집단에서 최댓값과 최솟값의 차이
      - $R = X_{max} - X_{min}$

대표적 통계량

- 표본의 산포(Variability) (5/5)
  - 종류 (5/5)
    - 예제 8.3

1996년 6월 19일 무스코가 호수에서 임의로 선정된 6명의 낚시꾼들에 의해 잡힌 송어의 수는 3, 4, 5, 6, 6, 7마리였다. 이때 표본분산, 표본표준편차, 표본범위를 구하라.

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

$\therefore$ 분산은 $\frac{13}{6}$, 표본표준편차(s)는 $\sqrt{\frac{13}{6}} \approx 1.47$, 표본범위는 $7 - 3 = 4$이므로 4

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- t 분포
- F 분포
- 분위수와 확률 그림

표본분포

- 표본분포(Sampling Distribution)
  - 정의
    - 동일한 모집단에서 반복적으로 표본을 추출하여 얻은 통계량의 확률분포
  - 특징
    - 표본 자체가 아닌 표본평균, 표본비율 등 통계량의 분포
    - 표본의 크기와 추출 방법에 따라 분포의 형태가 달라짐
    - 표본을 반복 추출할수록 통계량의 분포 특성을 파악하기 쉬움
  - 필요성
    - 모집단 전체 조사가 어려울 경우, 선택된 표본의 통계량으로 모집단의 특성 추정 가능
      - e.g., 표본평균의 분포, 표본분산의 분포

표본분포

- 표본평균의 분포 (1/5)
  - 개요
    - 표본평균 $\bar{X}$의 확률분포
  - 특징
    - 평균이 μ, 분산이 σ^2인 정규모집단으로부터 크기 n의 확률표본 추출 시, 확률표본 $X_i(i = 1,2, \dots, n)$은 모집단과 동일한 정규분포를 따름
      - 분포의 평균 $\mu_{\bar{x}}$
        - $\mu_{\bar{X}} = \frac{1}{\mathrm{n}}(\mu + \mu + \cdots + \mu) = \mu$
      - 분포의 분산 $\sigma_{\bar{X}}^2$
        - $\sigma_{\bar{X}}^2 = \frac{1}{\mathrm{n}}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{\sigma^2}{n}$
    - 중심극한정리(Central Limit Theorem)에 의하여 표본의 크기가 상당히 클 경우, 표본평균의 분포는 근사적으로 평균이 μ, 분산이 $\frac{\sigma^2}{n}$인 정규분포를 따름

정리 7.11

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

표본분포

- 표본평균의 분포 (2/5)
  - 중심극한정리(Central Limit Theorem)
    - 정리

평균이 μ이고, 분산이 σ^2인 모집단으로부터 크기가 n인 확률표본을 추출했을 때, 표본의 평균에 대해서

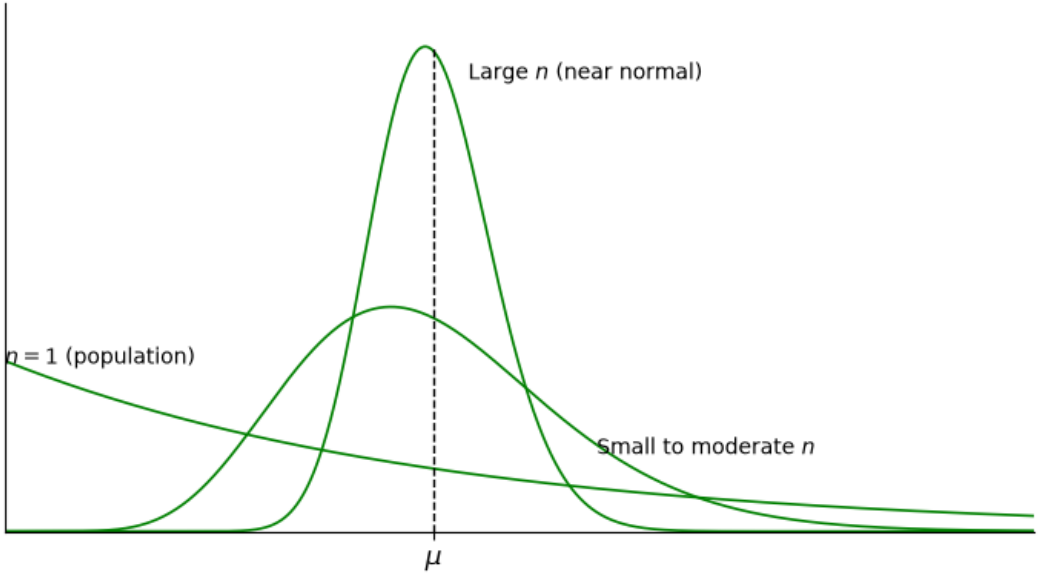
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

는 n이 상당히 클 때 표준정규분포$(z; 0,1)$에 접근

    - 특징 (1/2)
      - $n = 30$ 기준
        - n이 30이상일 경우, 모집단의 분포에 관계 없이 $\bar{X}$에 대한 정규분포의 근사
        - n이 30미만일 경우, 모집단의 분포가 정규분포에 근사할 경우에만 정규분포로의 근사

표본분포

- 표본평균의 분포 (3/5)
  - 중심극한정리(Central Limit Theorem)
    - 특징 (2/2)
      - n이 커짐에 따라 정규분포에 근사
      - $\bar{X}$ 의 평균은 표본에 관계없이 항상 μ
      - $\bar{X}$ 의 분산은 n이 커질수록 감소



<그림 1> 중심극한정리 예시

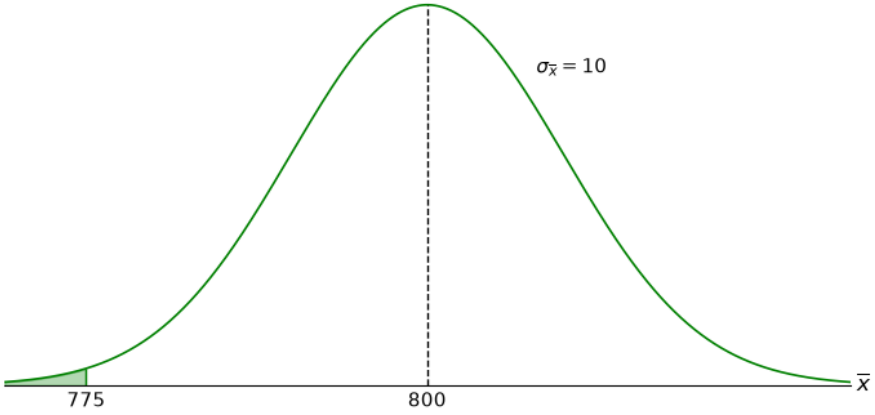
표본분포

- 표본평균의 분포 (4/5)
  - 예제 8.4

어떤 전구생산공장에서 생산되는 전구의 수명은 평균이 800시간이고, 표준편차가 40시간인 정규분포를 따른다고 알려져 있다. 임의로 추출된 16개의 전구의 평균수명이 775시간 미만일 확률을 구하라.

- 풀이
  - $\bar{X}$의 분포는 $\mu_{\bar{X}} = 800, \sigma_{\bar{X}}^2 = \frac{40}{\sqrt{16}} = 10$인 정규분포를 따름
  - 구하고자 하는 확률값은 그림 2 속 색칠된 부분
  - $\bar{x} = 775$에 해당하는 표준정규계수 z
    - $z = \frac{775-800}{10} = -2.5$
  - 구하고자 하는 확률값
    - $P(\bar{X} < 775) = P(Z < -2.5) = 0.0062$

$\therefore$ 0.0062



<그림 2> 예제 8.4 그림

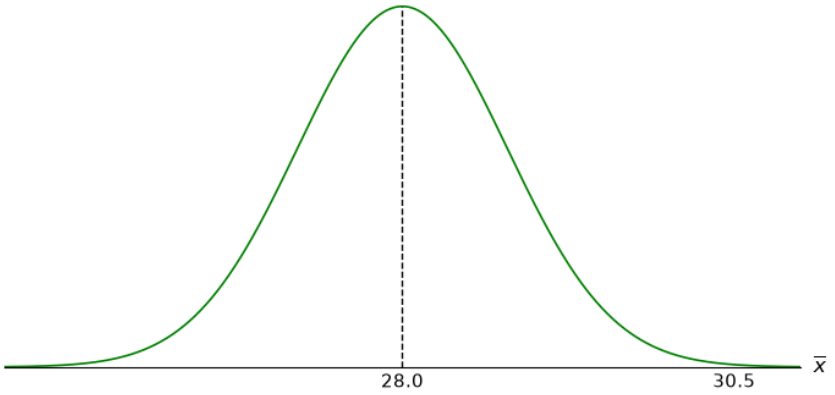
표본분포

- 표본평균의 분포 (5/5)
  - 예제 8.5

어느 대학의 두 캠퍼스 사이를 운행하는 셔틀버스이 운행시간은 평균 28분, 표준편차 5분의 분포를 따른다고 한다. 어느 한 주 동안 40번의 운행이 있었다고 할 때, 평균 운행시간이 30분보다 길 확률은 얼마인가? 평균시간은 분단위로 반올림하여 측정한다고 가정한다.

- 풀이
  - $\mu = 28, \sigma = 3, n = 3$일 때, $P(\bar{X} > 30)$을 구하면 됨
  - 시간은 분 단위로 반올림 측정되므로 $\bar{x}$가 30보다 크다는 것은 $\bar{x} > 30.5$와 동일
  - $\mathrm{P}(\bar{X} > 30) = P\left(\frac{\bar{X}-28}{5/\sqrt{40}} \geq \frac{30.5-28}{5/\sqrt{40}}\right) = P(Z \geq 3.16) = 0.0008$

∴ 0.0008



<그림 3> 예제 8.5 그림

표본분포

- 두 표본평균 차이의 분포 (1/2)
  - 정리

두 모집단이 서로 독립이고 각각의 평균이 μ_1, μ_2 분산이 σ_1^2, σ_2^2 일 때, 각 모집단으로부터 추출된 크기 n_1, n_2 인 두 표본의 표본평균 차이 $\overline{X_1} - \overline{X_2}$의 분포는 근사적으로 평균이 $\mu_{\overline{X_1}-\overline{X_2}} = \mu_1 - \mu_2$이고, 분산이 $\sigma^2_{\overline{X_1}-\overline{X_2}} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$인 정규분포를 따름

$$Z = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

는 근사적으로 표준정규분포를 따름

  - 특징
    - 두 표본의 크기 n_1과 n_2가 30이상일 경우, 두 모집단의 분포에 관계 없이 정규근사
    - n_1과 n_2가 30보다 작더라도, 두 모집단이 명백하게 비정규모집단이 아닌 한 어느 정도 정규근사
    - 두 모집단이 모두 정규모집단일 경우, 표본의 크기에 관계 없이 $\overline{X_1} - \overline{X_2}$는 정규분포를 따름

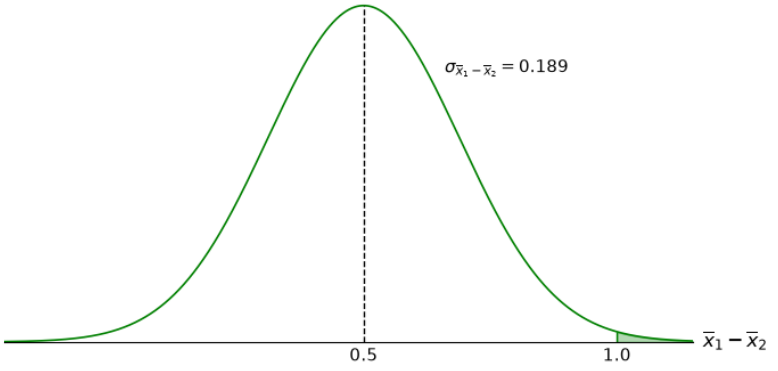
표본분포

- 두 표본평균 차이의 분포 (2/2)
  - 예제 8.6

A, B 두 회사에서 텔레비전 브라운관을 생산하고 있다. A 회사에서 생산되는 브라운관의 수명은 평균이 6.5년, 표준편차가 0.9년이고, B 회사에서 생산되는 브라운관의 수명은 평균이 6년, 표준편차가 0.8년이라고 알려져 있다. A 회사에서 36개의확률표본을 추출하고, B 회사에서는 49개의 확률표본을 추출했을 때, A 회사 제품의 표본평균이 B 회사 제품의 표본평균보다 적어도 1년 이상 길 확률을 구하라.

- 풀이
  - 두 표본평균 차이의 분포 정리에 의하면 $\overline{X_A} - \overline{X_B}$ 의 분포는 다음 그림 4와 같음
  - 근사적으로 평균이 $\mu_{\overline{X_A}-\overline{X_B}} = 6.5 - 6.0 = 0.5$이고, 표준편차가 $\sigma_{\overline{X_A}-\overline{X_B}} = \sqrt{\left(\frac{0.81}{36} + \frac{0.64}{49}\right)} = 0.189$인 정규분포를 따름
  - $\overline{x_A} - \overline{x_B} = 1.0$에 대응하는 z값
    - $z = \frac{1.0-0.5}{0.189} = 2.65$
  - $\mathrm{P}(\overline{X_A} - \overline{X_B} > 1.0) = P(Z > 2.65) = 1 - P(Z \leq 2.65)$
    $= 1 - 0.9960 = 0.0040$

  $\therefore$ 0.0040



<그림 4> 예제 8.6 그림

표본분포

- 표본분산의 분포 (1/3)
  - 필요성
    - 모집단에서 표본 추출 시, 해당 표본의 분산 S^2이 모집단 σ^2과의 관계를 파악함으로써 표본을 통해 모집단에 대한 정보 파악
  - 특징
    - 자유도 $v = n - 1$인 카이제곱분포를 따름

분산이 σ^2인 정규모집단으로부터 크기 n의 표본을 추출하였을 때, 표본분산을 S^2이라 하면 통계량

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

은 자유도 $v = n - 1$인 카이제곱분포를 따름

표본분포

- 표본분산의 분포 (2/3)
  - 증명
    - 평균이 μ이고 분산이 σ^2인 정규모집단으로부터 크기 n의 표본을 추출하였을 때, 표본분산의 통계량 S^2의 값을 얻기 위해 표본분산 s^2을 계산
      - $\sum_{i=1}^{n}(X_i-\mu)^2=\sum_{i=1}^{n}[(X_i-\bar{X})+(\bar{X}-\mu)]^2=\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{i=1}^{n}(\bar{X}-\mu)^2+2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\bar{X})=\sum_{i=1}^{n}(X_i-\bar{X})^2+n(\bar{X}-\mu)^2$
      - 양변을 σ^2으로 나누고, $\sum_{i=1}^{n}(X_i-\bar{X})^2$를 $(n-1)S^2$로 치환 시,
        - $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2=\frac{(n-1)S^2}{\sigma^2}+\frac{(X_i-\mu)^2}{\sigma^2/n}$
      - $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$는 자유도가 n인 카이제곱분포를 따르고, $\frac{(X_i-\mu)^2}{\sigma^2/n}$ 또한 자유도가 1인 카이제곱분포를 따름

표본분포

- 표본분산의 분포 (3/3)
  - 예제 8.7

어떤 자동차배터리 제조업자는 자기회사에서 제조한 배터리의 수명이 평균은 3년, 표준편차는 1년이라고 주장하고 있다. 이 회사에서 제조된 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.

- 풀이
  - 먼저 5개의 배터리의 표본분산 계산
    - $s^2 = \frac{(5)(48.26)-(15)^2}{(5)(4)} = 0.815$
  - 따라서 $\chi^2 = \frac{(4)(0.815)}{1} = 3.26$이 되고, 이는 자유도 4인 카이제곱분포를 따르는 값이 됨
  - 자유도 4인 경우 χ^2 값의 0.484와 11.143 사이에 오게 되므로, $\sigma^2 = 1$로 하여 계산된 χ^2의 값 3.26으로 미루어 볼 때 표준편차가 1년이라는 제조업자의 주장은 합당하다고 할 수 있음

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- t 분포
- F 분포
- 분위수와 확률 그림

t 분포

- t 분포 (1/7)
  - 정의
    - 표준정규분포 확률변수 Z와 자유도 v인 카이제곱확률변수 V에 대해 $T = \frac{Z}{\sqrt{V/v}}$가 가지는 분포
      - $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, $V = \frac{(n-1)S^2}{\sigma^2}$
  - 필요성
    - 모집단의 분산에 대해 알려지지 않았을 때, 표본을 통해 모평균을 추정하고자 할 때 사용
      - 모집단의 분산을 알 경우, 표준정규분포를 사용
      - 모집단의 분산을 모를 경우, t 분포 사용

t 분포

- t 분포 (2/7)
  - 정리

Z와 V가 각각 표준정규확률변수와 자유도 v인 카이제곱확률변수이고, Z와 V가 서로 독립일 경우,

$$T = \frac{Z}{\sqrt{V/v}}$$

의 확률밀도함수는

$$h(t) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2}, \qquad -\infty < t < \infty$$

가 되고, 이는 자유도 v인 t 분포를 따름

  - 따름 정리 8.1

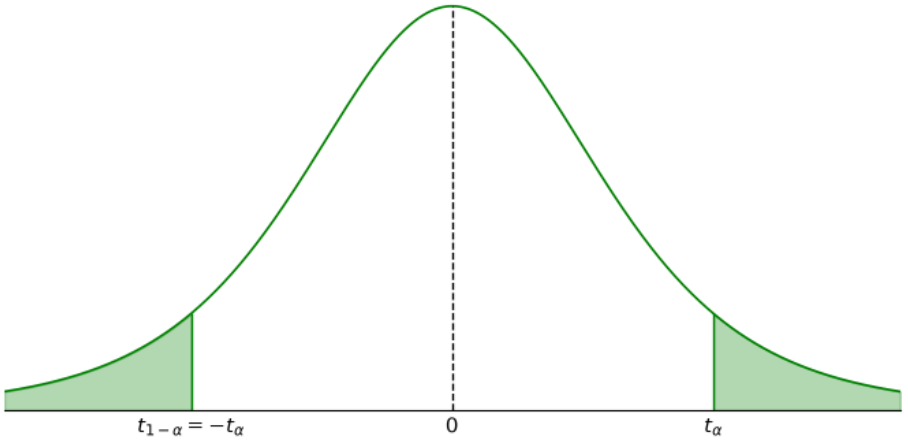
$X_1, X_2, \cdots, X_n$이 모두 평균이 μ이고 표준편차가 σ인 정규분포를 따르고 서로 독립인 확률변수일 때, 확률변수 $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$는 자유도 $v = n-1$인 t 분포를 따름.

단,

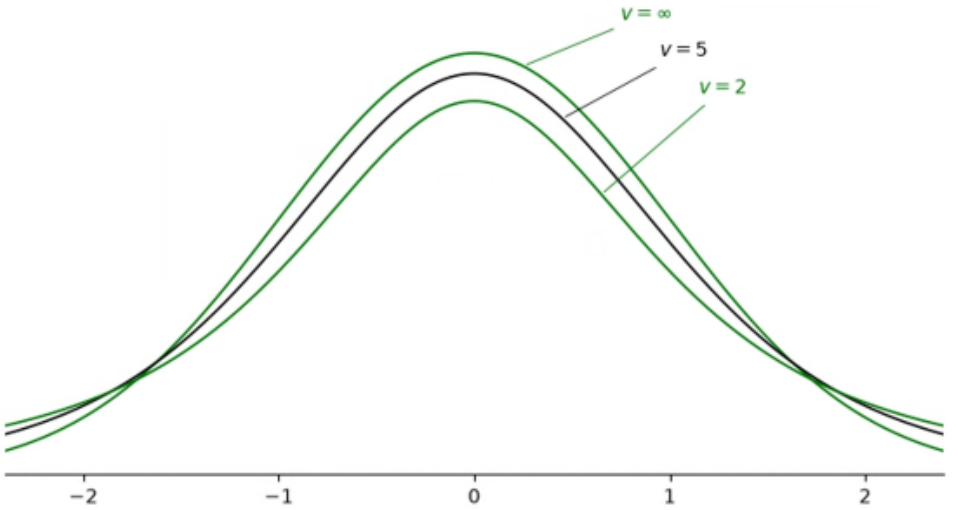
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

t 분포

- t 분포 (3/7)
  - 특징
    - 표준정규분포와 같이 원점을 중심으로 대칭
    - 양쪽 면적이 각각 가 되는 점을 t_α, $t_{1-\alpha}$ 로 나타냄
    - 자유도 v가 커질수록 표준정규분포에 접근
      - $v \leq 30$, 표준정규분포보다 낮고 완만
      - $v > 30$, 표준정규분포와 유사



<그림 5> t 분포의 대칭성



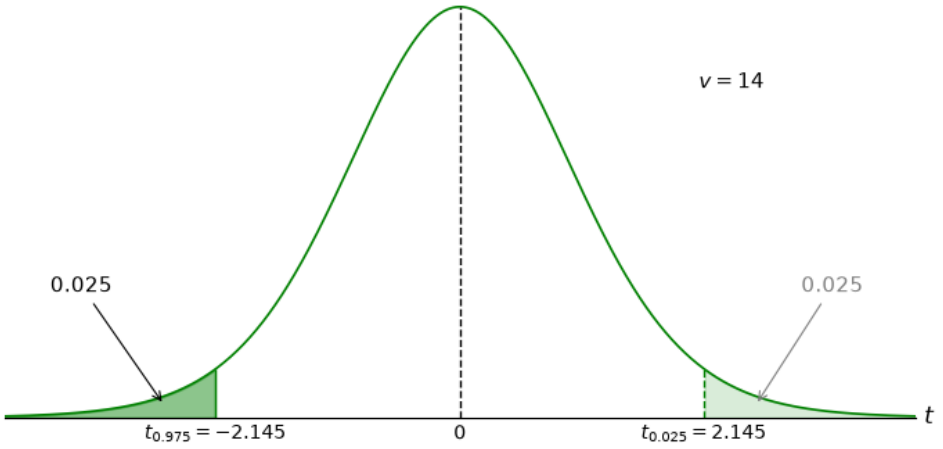
<그림 6> 자유도에 따른 t 분포

t 분포

- t 분포 (4/7)
  - 예제 8.8

자유도 14인 t 분포에서 왼쪽 면적이 0.025가 되는 t 값을 구하라.

- 풀이
  - $v = 14$를 t 분포 수식에 대입
    - $\frac{\Gamma(15/2)}{\Gamma(7)\sqrt{14\pi}} = \left(\frac{13}{2}\right)! \times \frac{1}{6! \times \sqrt{14\pi}} \times \left(1 + \frac{t^2}{14}\right)^{(-15)/2} = 0.025$
  - $t_{0.025} = 2.145$
  - 구하고자 하는 값은 왼쪽 면적이 0.025가 되는 t 값이므로 대칭 이용
    - $t_{0.975} = -t_{0.025} = -2.145$



<그림 7> 예제 8.8 그림

t 분포

- t 분포 (5/7)
  - 예제 8.9

$P(-t_{0.025} < T < t_{0.05})$를 구하라.

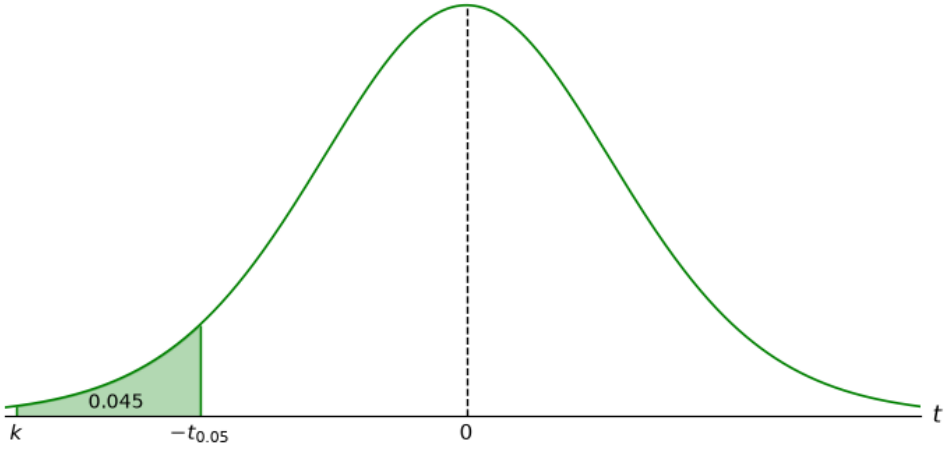
- 풀이
  - $t_{0.05}$는 t 분포곡선에서 오른쪽 면적이 0.05이고, $-t_{0.025}$는 왼쪽 면적이 0.025 이므로 나머지 면적은 $1 - 0.05 - 0.025 = 0.925$
  - 따라서 $P(-t_{0.025} < T < t_{0.05}) = 0.925$

t 분포

- t 분포 (6/7)
  - 예제 8.10

정규모집단에서 15개의 표본을 추출했을 때, $P(k < T < -1.761) = 0.045$를 만족하는 k 값을 구하라. 자유도는 14라고 가정한다.

- 풀이
  - 자유도 14인 t 분포에서 1.761은 $t_{0.05}$에 대응
  - $-t_{0.05} = -1.761$이므로, $k = -t_{\alpha}$ 라고 할 때, $0.045 = 0.05 - \alpha$ 이므로 $\alpha = 0.005$
  - 따라서, 자유도 14일 때 이를 만족하는 k 값은 $k = -t_{0.005} = -2.977$이 되고, $P(k < T < -1.761) = 0.045$이 됨



<그림 8> 예제 8.9 그림

ν값	α값				
	0.40	0.20	0.15	0.05	0.025
11	0.260	0.876	1.088	1.796	2.201
12	0.259	0.873	1.083	1.782	2.179
13	0.259	0.870	1.079	1.771	2.160
14	0.258	0.868	1.076	1.761	2.145
15	0.258	0.866	1.074	1.753	2.131

t 분포

- t 분포 (7/7)
  - 예제 8.11

어떤 화공기사는 어느 배치공정의 수율이 원재료의 리터당 500g이라고 주장하고 있다. 그는 이를 입증하기 위해 매월 25개의 배치를 추출하여 시험을 하였다. 시험결과 계산된 t 값이 $-t_{0.05}$와 $t_{0.05}$ 사이에 있으면 그의 주장이 타당성이 있다고 하기로 한다. 25개 배치의 시험결과 표본평균이 518g이고 표준편차는 40g이었다면 어떤 결론을 낼 수 있겠는가? 단, 모집단은 근사적으로 정규분포를 따른다고 가정하라.

- 풀이
  - 시험결과를 가지고 t 값 계산
    - $t = \frac{518-500}{40/\sqrt{5}} = 2.25$
  - 자유도가 24일 때 $t_{0.05} = 1.711$
  - 계산된 t 값이 $t_{0.05}$보다 크므로 실제의 수율은 500g보다 크다고 할 수 있음

ν값	α값				
	0.40	0.20	0.15	0.05	0.025
21	0.257	0.859	1.063	1.721	2.080
22	0.256	0.858	1.061	1.717	2.074
23	0.256	0.858	1.060	1.714	2.069
24	0.256	0.857	1.059	1.711	2.064
25	0.256	0.856	1.058	1.708	2.060

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- t 분포
- F 분포
- 분위수와 확률 그림

F 분포

- F 분포 (1/5)
  - 정의
    - 두 개의 서로 독립적인 카이제곱변수의 비율 $F = \frac{U/v_1}{V/v_2}$가 가지는 분포
      - U와 V는 서로 독립, 각각 자유도가 v_1, v_2인 카이제곱확률변수
  - 필요성
    - 카이제곱변수의 비율을 통해 두가지 이상의 표본집단의 분산을 비교하거나 모집단의 분산을 추정할 때 사용
      - 두 가지 이상의 표본평균들이 동일한 모평균을 가진 집단에서 추출되었는지 아닌지를 판단

F 분포

- F 분포 (2/5)
  - 정리

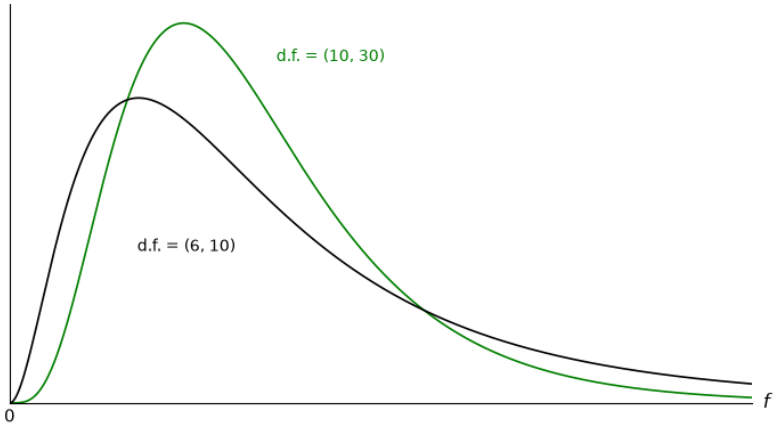
확률변수 U와 V는 각각 자유도 v_1, v_2를 가지는 카이제곱분포를 따를 때,

$$F = \frac{U/v_1}{V/v_2}$$

의 확률밀도함수는

$$h(f) = f(x) = \begin{cases} \dfrac{\Gamma[(v_1+v_2)/2](v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_2/2)} \dfrac{f^{(v_1/2)-1}}{(1+v_1 f/v_2)^{(v_1+v_2)/2}}, & f > 0 \\ 0, & f \le 0 \end{cases}$$

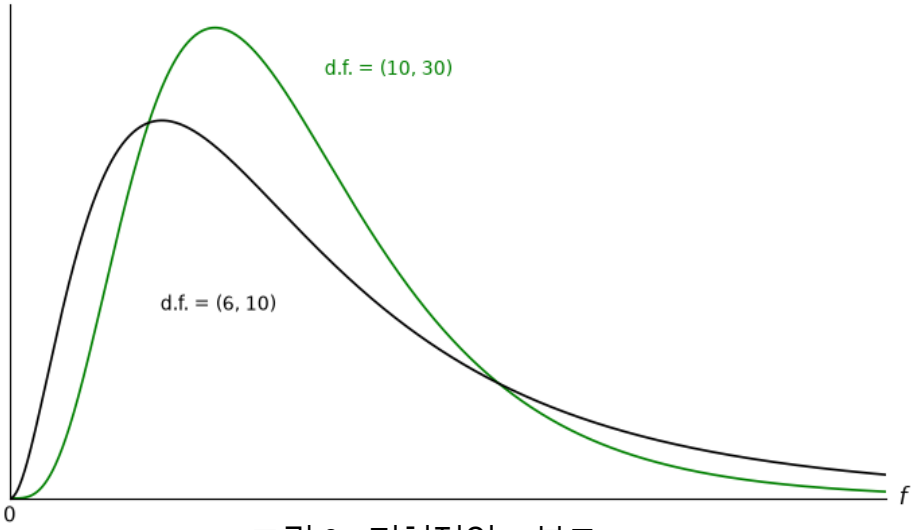
이 되고, 이는 자유도 v_1, v_2인 F 분포를 따름



<그림 9> 전형적인 F 분포

F 분포

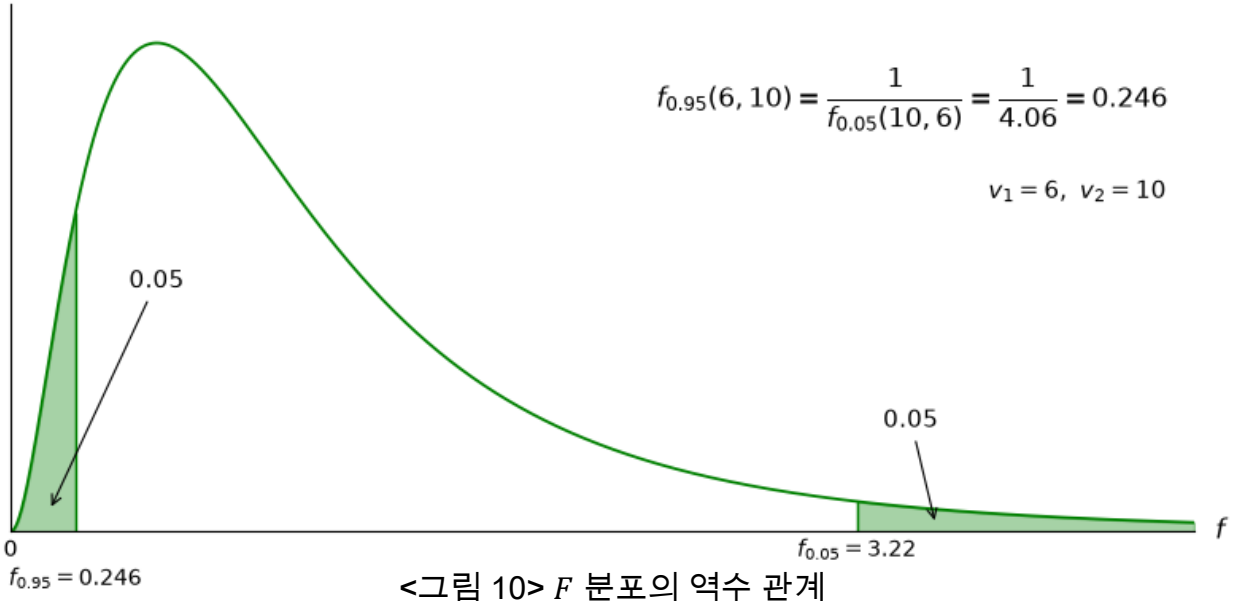
- F 분포 (3/5)
  - 특징 (1/3)
    - 두 자유도 v_1, v_2에 따라 분포곡선의 형태 결정
    - 분자의 자유도 v_1, 분모에 자유도 v_2 순서로 표기
      - v_1은 분자에 있는 카이제곱확률변수 U의 자유도
      - v_2는 분모에 있는 카이제곱확률변수 V의 자유도



<그림 9> 전형적인 F 분포

F 분포

- F 분포 (4/5)
  - 특징 (2/3)
    - 특정 면적의 값을 알고 있을 경우, 분자와 분모의 자유도를 바꾼 뒤 역수를 취해 반대쪽 면적의 값 계산 가능
      - 자유도 v_1, v_2에서 f_α 값을 $f_\alpha(v_1, v_2)$로 나타낼 때, $f_{(1-\alpha)}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}$



<그림 10> F 분포의 역수 관계

F 분포

- F 분포 (5/5)
  - 특징 (3/3)
    - 모분산이 각각 σ_1^2, σ_2^2인 두 정규모집단에서 크기 n_1, n_2인 표본 추출
      - 두 집단의 실제 분산이 같은지 판별하는 데 활용
    - 각 표본분산을 이용해 카이제곱확률변수 계산
      - 자유도: $v_1 = n_1 - 1, v_2 = n_2 - 1$
      - $\chi_1^2 = \frac{(n_1-1)S_1^2}{\sigma_1^2}, \chi_2^2 = \frac{(n_2-1)S_2^2}{\sigma_2^2}$
    - $\chi_1^2 = U,\ \chi_2^2 = V$로 치환하여 정리
      - $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$

목 차

- 확률표본
- 대표적 통계량
- 표본분포
- t 분포
- F 분포
- 분위수와 확률 그림

분위수와 확률 그림

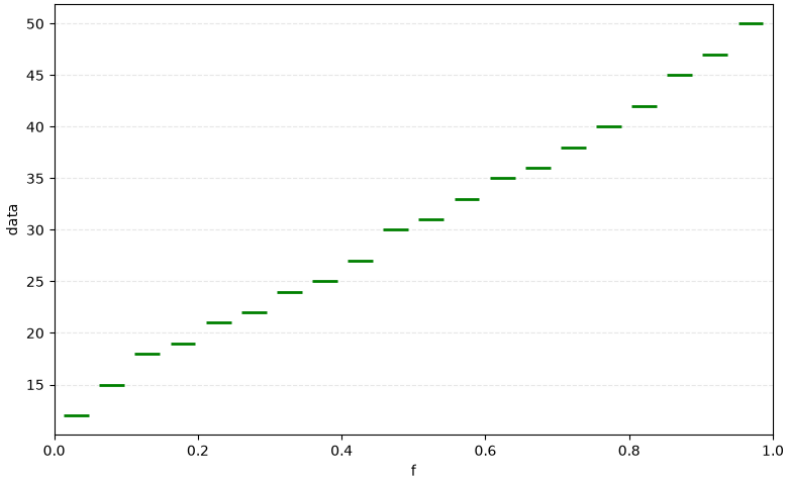
- 분위수(Quantile)
  - 정의
    - 자료값 중에서 $q(f)$보다 작거나 같은 자료값이 차지하는 비율이 f가 되는 값
  - 특징
    - 확률분포에서 확률변수의 구간을 나누는 기준이 되는 수
    - f는 0과 1사이의 값
      - $q(0.5)$ 는 자료의 중간값인 표본중앙값
      - $q(0.25)$는 하위 25% 지점의 값인 하위사분위수
  - 필요성
    - 정규분포 크기를 벗어나거나, 산포가 큰 상황에서 분위수가 대푯값으로 사용

분위수와 확률 그림

- 분위수그림(Quantile Plot) (1/3)
  - 정의
    - 데이터를 크기순으로 정렬하고, 각 데이터값을 해당 누적 비율에 대응시켜 나타낸 그림
  - 필요성
    - 데이터가 어느 구간에 집중되어 있는지 파악
    - 데이터의 치우침과 값의 변화 양상 확인
    - 데이터의 분포 특성과 집단 간 차이 비교

분위수와 확률 그림

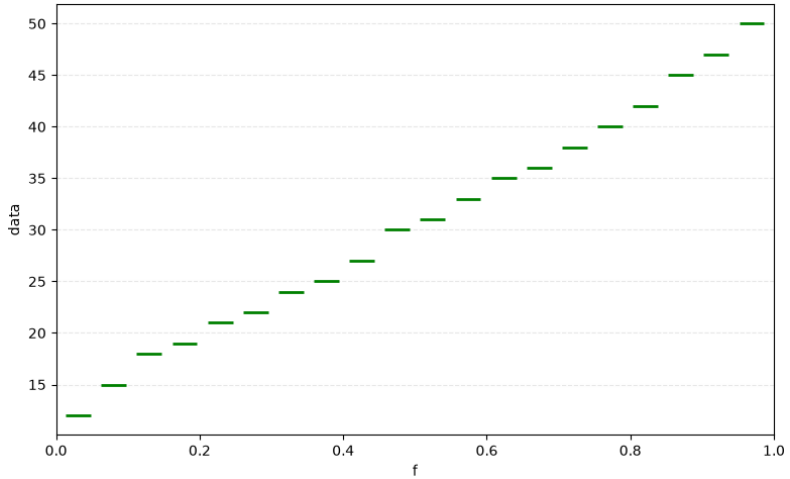
- 분위수그림(Quantile Plot) (2/3)
  - 특징
    - x축은 누적비율 f, y축은 정렬된 관측값을 표시
    - 데이터를 오름차순으로 정렬 후, i번째 값에 비율 f_i를 계산하여 표시
      - $f_i = \frac{i-3/8}{n+1/4}$, n=조사한 데이터 수, $i = 1,2,\dots,n$
        - 기본형 $\frac{i}{n+1}$을 정규분포에 맞게 보정한 식
        - $\frac{3}{8}$은 정규분포에서 i번째 값의 평균 위치를 근사하는 보정값으로 수치적 절충값
        - $\frac{1}{4}$는 양 끝 대칭 을 만족하도록 결정



<그림 11> 분위수그림

분위수와 확률 그림

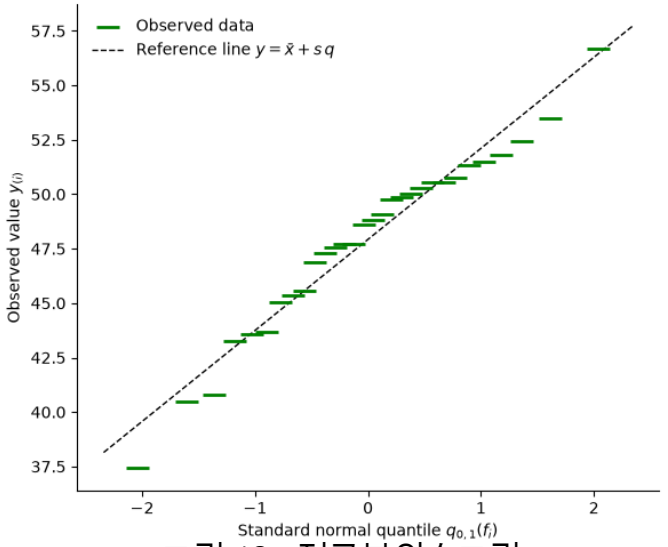
- 분위수그림(Quantile Plot) (3/3)
  - 예시
    - 학생 20명의 퀴즈 점수 분위수그림($n = 20$)
      - 퀴즈 점수
        - 12, 15, 18, 19, 21, 22, 24, 25, 27, 30, 31, 33, 35, 36, 38, 40, 42, 45, 47, 50
      - 계산 예시
        - 12점의 경우, 최소값이므로 $i = 1$이며, $f_1 = \frac{1-3/8}{20+1/4} \approx 0.031$
        - 50점의 경우, 최댓값이므로 $i = 20$이며, $f_{20} = \frac{20-3/8}{20+1/4} \approx 0.969$



<그림 11> 분위수그림

분위수와 확률 그림

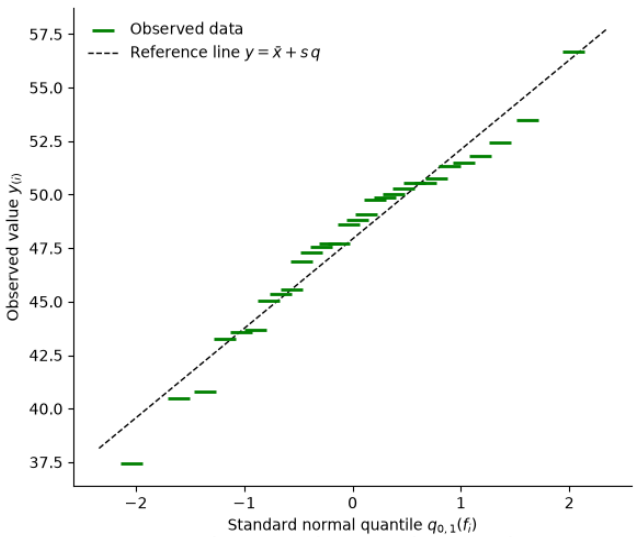
- 정규분위수그림(Normal Quantile-Quantile Plot) (1/3)
  - 정의
    - 정렬한 관측값을 같은 순위에서 정규분포가 가질 이론적 분위수와 비교하여 그린 그림
  - 필요성
    - 데이터가 정규분포를 따르는지 시각적으로 확인
    - 정규분포를 가정하는 통계분석의 적용 가능성 판단



<그림 12> 정규분위수그림

분위수와 확률 그림

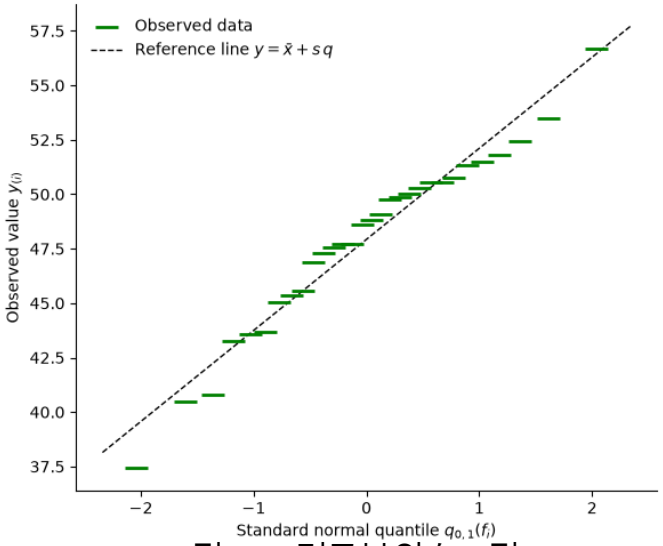
- 정규분위수그림(Normal Quantile-Quantile Plot) (2/3)
  - 특징
    - x축은 표준정규분위수 $q_{0,1}(f_i)$, y축은 정렬된 관측값 $y_{(i)}$
    - 기준선 $y = \bar{x} + s \cdot q$
      - $\bar{x}$는 관측 데이터의 표본평균, s는 관측 데이터의 표본표준편차, q는 x축 값 $q_{0,1}(f)$
    - 정규분포의 분위수 계산식은 복잡하여 근사식으로 계산
      - $q_{\mu,\sigma}(f) = \mu + \sigma\{4.91[f^{0.14} - (1-f)^{0.14}]\}$
        - $q_{\mu,\sigma}(f)$는 평균이 μ, 표준편차가 σ인 정규분포에서 누적확률이 f가 되는 지점



<그림 12> 정규분위수그림

분위수와 확률 그림

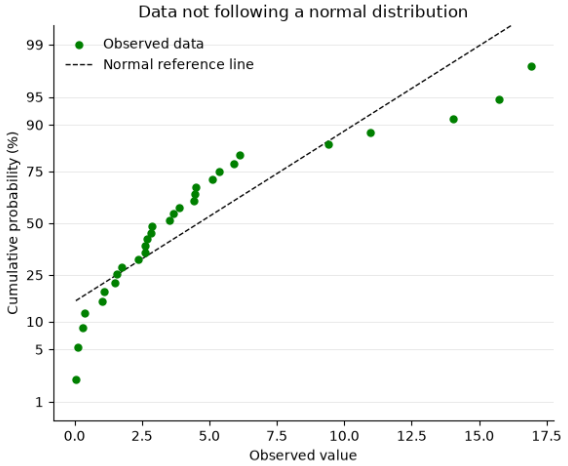
- 정규분위수그림(Normal Quantile-Quantile Plot) (3/3)
  - 예시
    - 생산 부품 30개의 무게 정규분위수그림 ($n = 30, \bar{x} = 47.95, s = 4.18$)
      - 무게를 오름차순 정렬하여 $y_{(i)}$로 표시
        - 37.4, 40.5, 40.8, 43.3, 43.6, 43.7, 45.1, 45.4, 45.6, 46.9, 47.3, 47.6, 47.7, 47.8, 48.7, 48.9, 49.1, 49.8, 49.9, 50.0, 50.3, 50.5, 50.6, 50.7, 51.4, 51.5, 51.8, 52.5, 53.5, 56.7
      - 각 f_i에 대응하는 $q_{0,1}(f_i)$를 x축으로 사용



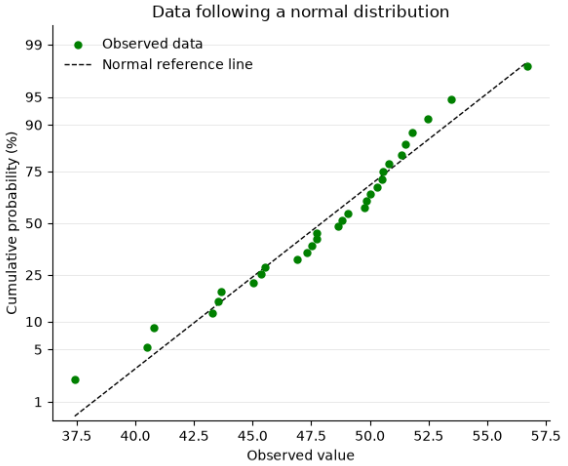
<그림 12> 정규분위수그림

분위수와 확률 그림

- 정규확률그림(Normal Probability Plot) (1/4)
  - 정의
    - 정렬한 관측값을 같은 순위의 누적확률과 비교하여 나타낸 그림
  - 필요성
    - 정규분포에서 벗어나는 데이터의 패턴 확인
    - 분포의 비대칭성이나 꼬리 부분의 특성 파악
    - 극단적인 관측값이나 이상값의 가능성 탐색



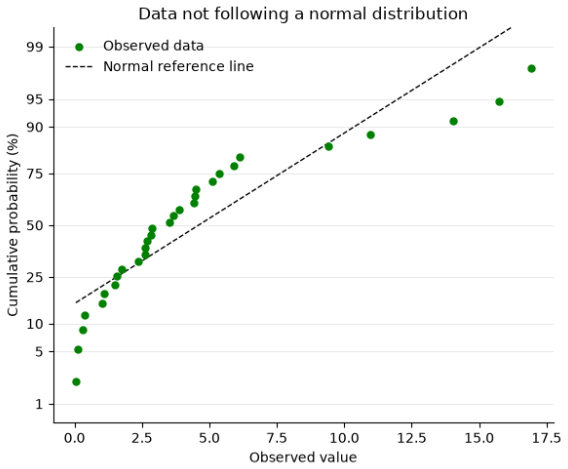
<그림 13> 정규분포를 따르지 않는 정규확률그림



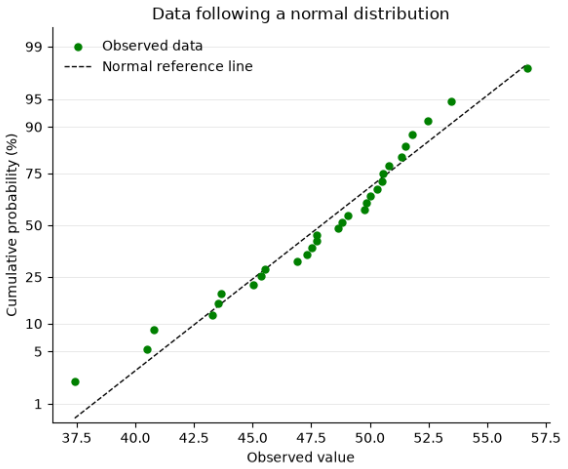
<그림 14> 정규분포를 따르는 정규확률그림

분위수와 확률 그림

- 정규확률그림(Normal Probability Plot) (2/4)
  - 특징
    - x축은 정렬된 관측값, y축은 누적확률 $f_i(\%)$
      - $f_i = \frac{i-3/8}{n+1/4}$ 으로 분위수그림과 동일하며 눈금은 $q_{0,1}(f_i)$에 배치
    - 기준선은 관측 데이터의 평균, 표준편차를 갖는 정규분포의 누적확률



<그림 13> 정규분포를 따르지 않는 정규확률그림



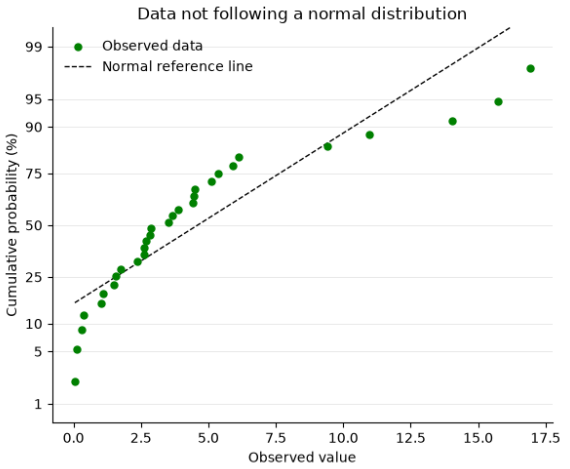
<그림 14> 정규분포를 따르는 정규확률그림

분위수와 확률 그림

- 정규확률그림(Normal Probability Plot) (3/4)
  - 예시 – 정규분포를 따르지 않는 경우
    - 고객 30명의 대기시간 정규확률그림 ($n = 30$)
      - 대기시간
        - 0.0, 0.1, 0.3, 0.4, 1.0, 1.1, 1.5, 1.6, 1.8, 2.4, 2.6, 2.7, 2.7, 2.8, 2.9, 3.6, 3.7, 3.9, 4.5, 4.5, 4.5, 5.2, 5.4, 5.9, 6.1, 9.4, 11.0, 14.1, 15.7, 16.9
      - 계산 예시
        - 0.0분의 경우 최솟값이므로,

          $i = 1$이며, $f_1 = \frac{1-\frac{3}{8}}{30+\frac{1}{4}} \approx 0.021$
        - 16.9분의 경우 최댓값이므로,

          $i = 30$이며, $f_{30} = \frac{30-\frac{3}{8}}{30+\frac{1}{4}} \approx 0.979$
      - 해석
        - 상단 점들이 기준선 오른쪽으로 벗어나므로 오른쪽 꼬리가 긴 비대칭 분포



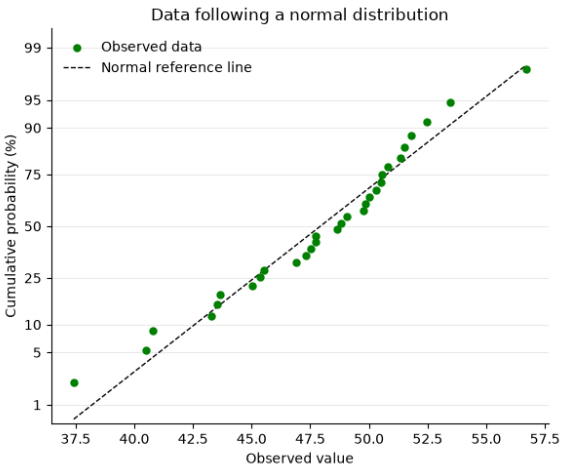
<그림 13> 정규분포를 따르지 않는 정규확률그림

분위수와 확률 그림

- 정규확률그림(Normal Probability Plot) (4/4)
  - 예시 – 정규분포를 따르는 경우
    - 생산 부품 30개의 무게 정규확률그림 ($n = 30$)
      - 무게
        - 37.4, 40.5, 40.8, 43.3, 43.6, 43.7, 45.1, 45.4, 45.6, 46.9, 47.3, 47.6, 47.7, 47.8, 48.7, 48.9, 49.1, 49.8, 49.9, 50.0, 50.3, 50.5, 50.6, 50.7, 51.4, 51.5, 51.8, 52.5, 53.5, 56.7
      - 계산 예시
        - 37.4g의 경우 최솟값이므로,

          $i = 1$이며, $f_1 = \frac{1-\frac{3}{8}}{30+\frac{1}{4}} \approx 0.021$
        - 56.7g의 경우 최댓값이므로,

          $i = 30$이며, $f_{30} = \frac{30-\frac{3}{8}}{30+\frac{1}{4}} \approx 0.979$
      - 해석
        - 대부분의 점이 기준선 근처에 위치하므로 정규분포와 부합



<그림 14> 정규분포를 따르는 정규확률그림

Thanks!

서 진 우(jinwoo@pel.sejong.ac.kr)

부록#1

- 표준정규분포표 (1/2)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

부록#1

- 표준정규분포표 (2/2)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

부록#2

- 카이제곱분포표 (1/2)

	α									
v	**0.995**	**0.99**	**0.98**	**0.975**	**0.95**	**0.90**	**0.80**	**0.75**	**0.70**	**0.50**
1	0.0^4393	0.0^3157	0.0^3628	0.0^3982	0.00393	0.0158	0.0642	0.102	0.148	0.455
2	0.0100	0.0201	0.0404	0.0506	0.103	0.211	0.446	0.575	0.713	1.386
3	0.0717	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	2.366
4	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	3.357
5	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	4.351
6	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	5.348
7	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	6.346
8	1.344	1.647	2.032	2.180	2.733	3.490	4.594	5.071	5.527	7.344
9	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	8.343
10	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	9.342
11	2.603	3.053	3.609	3.816	4.575	5.578	6.989	7.584	8.148	10.341
12	3.074	3.571	4.178	4.404	5.226	6.304	7.807	8.438	9.034	11.340
13	3.565	4.107	4.765	5.009	5.892	7.041	8.634	9.299	9.926	12.340
14	4.075	4.660	5.368	5.629	6.571	7.790	9.467	10.165	10.821	13.339
15	4.601	5.229	5.985	6.262	7.261	8.547	10.307	11.037	11.721	14.339
16	5.142	5.812	6.614	6.908	7.962	9.312	11.152	11.912	12.624	15.338
17	5.697	6.408	7.255	7.564	8.672	10.085	12.002	12.792	13.531	16.338
18	6.265	7.015	7.906	8.231	9.390	10.865	12.857	13.675	14.440	17.338
19	6.844	7.633	8.567	8.907	10.117	11.651	13.716	14.562	15.352	18.338
20	7.434	8.260	9.237	9.591	10.851	12.443	14.578	15.452	16.266	19.337
21	8.034	8.897	9.915	10.283	11.591	13.240	15.445	16.344	17.182	20.337
22	8.643	9.542	10.600	10.982	12.338	14.041	16.314	17.240	18.101	21.337
23	9.260	10.196	11.293	11.689	13.091	14.848	17.187	18.137	19.021	22.337
24	9.886	10.856	11.992	12.401	13.848	15.659	18.062	19.037	19.943	23.337
25	10.520	11.524	12.697	13.120	14.611	16.473	18.940	19.939	20.867	24.337
26	11.160	12.198	13.409	13.844	15.379	17.292	19.820	20.843	21.792	25.336
27	11.808	12.878	14.125	14.573	16.151	18.114	20.703	21.749	22.719	26.336
28	12.461	13.565	14.847	15.308	16.928	18.939	21.588	22.657	23.647	27.336
29	13.121	14.256	15.574	16.047	17.708	19.768	22.475	23.567	24.577	28.336
30	13.787	14.953	16.306	16.791	18.493	20.599	23.364	24.478	25.508	29.336
40	20.707	22.164	23.838	24.433	26.509	29.051	32.345	33.66	34.872	39.335
50	27.991	29.707	31.664	32.357	34.764	37.689	41.449	42.942	44.313	49.335
60	35.534	37.485	39.699	40.482	43.188	46.459	50.641	52.294	53.809	59.335

부록#2

- 카이제곱분포표 (2/2)

	α									
v	**0.30**	**0.25**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.001**
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819	34.527
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.124
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.698
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718	40.791
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.819
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.314
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.796
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.619
26	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.051
27	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645	55.475
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.994	56.892
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.335	58.301
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.702
40	44.165	45.616	47.269	51.805	55.758	59.342	60.436	63.691	66.766	73.403
50	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154	79.490	86.660
60	65.226	66.981	68.972	74.397	79.082	83.298	84.58	88.379	91.952	99.608

부록#3

- 그래프 코드
  - <그림 1> 중심극한정리 예시

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt

4.  # 모집단: 지수분포(평균 mu) -> 표본평균은 Gamma(n, mu/n)
5.  mu = 1.0
6.  x = np.linspace(0.001, 2.4, 1000)

7.  def xbar_pdf(n):
8.      # Gamma(shape=n, scale=mu/n) pdf, 로그 공간에서
    계산 (n 커도 overflow 없음)
9.      rate = n / mu
10.     log_pdf = n * np.log(rate) + (n - 1) * np.log(x)
    - rate * x - math.lgamma(n)
11.     return np.exp(log_pdf)

12. color = "green"
13. fig, ax = plt.subplots(figsize=(7, 4))

14. ax.plot(x, xbar_pdf(1),  color=color, lw=1.2)   # n
    = 1 (population)
15. ax.plot(x, xbar_pdf(10), color=color, lw=1.2)   #
    small to moderate n
16. ax.plot(x, xbar_pdf(50), color=color, lw=1.2)   #
    large n
```

```
17. # mu 위치 점선
18. ax.vlines(mu, 0, xbar_pdf(50).max(), colors="black",
    linestyles="--", lw=1)

19. # 라벨
20. ax.text(0.0, 0.95, r"$n = 1$ (population)",
    fontsize=10, va="bottom")
21. ax.text(1.38, 0.45, r"Small to moderate $n$",
    fontsize=10)
22. ax.text(1.08, 2.65, r"Large $n$ (near normal)",
    fontsize=10)

23. # 축 스타일
24. ax.spines["top"].set_visible(False)
25. ax.spines["right"].set_visible(False)
26. ax.set_yticks([])
27. ax.set_xticks([mu])
28. ax.set_xticklabels([r"$\mu$"], fontsize=12)
29. ax.set_xlim(0, 2.4)
30. ax.set_ylim(0, 3.1)

31. plt.tight_layout()
32. plt.show()
```

부록#3

- 그래프 코드
  - <그림 2> 예제 8.4 그림

```
1.  import numpy as np
2.  import matplotlib.pyplot as plt

3.  # 표본평균 분포: N(800, 10^2)
4.  mu, sigma = 800, 10
5.  x = np.linspace(769, 831, 1000)
6.  pdf = np.exp(-0.5 * ((x - mu) / sigma) ** 2) /
    (sigma * np.sqrt(2 * np.pi))

7.  color = "green"
8.  fig, ax = plt.subplots(figsize=(7, 3.6))

9.  ax.plot(x, pdf, color=color, lw=1.2)

10. # 775 왼쪽 꼬리 영역
11. tail = x <= 775
12. ax.fill_between(x[tail], pdf[tail], color=color,
    alpha=0.3)
13. ax.vlines(775, 0, pdf[tail][-1], colors=color, lw=1)

14. # 평균 위치 점선
15. ax.vlines(mu, 0, pdf.max(), colors="black",
    linestyles="--", lw=1)
```

```
16. # 라벨
17. ax.text(808, pdf.max() * 0.82,
    r"$\sigma_{\overline{x}} = 10$", fontsize=11)
18. ax.text(831.5, 0, r"$\overline{x}$", fontsize=12,
    va="center")

19. # 축 스타일: x축 선만 남김
20. for s in ["top", "right", "left"]:
21.     ax.spines[s].set_visible(False)
22. ax.set_yticks([])
23. ax.set_xticks([775, 800])
24. ax.tick_params(axis="x", length=0, labelsize=11)
25. ax.set_xlim(769, 831)
26. ax.set_ylim(0, pdf.max() * 1.05)

27. plt.tight_layout()
28. plt.show()
```

부록#3

- 그래프 코드
  - <그림 3> 예제 8.5 그림

```
1.  import numpy as np
2.  import matplotlib.pyplot as plt

3.  # 표본평균 분포: N(28, (5/sqrt(40))^2)
4.  mu = 28.0
5.  sigma = 5 / np.sqrt(40)          # ≈ 0.79
6.  x = np.linspace(25.0, 31.0, 1000)
7.  pdf = np.exp(-0.5 * ((x - mu) / sigma) ** 2) /
    (sigma * np.sqrt(2 * np.pi))

8.  color = "green"
9.  fig, ax = plt.subplots(figsize=(7, 3.6))

10. ax.plot(x, pdf, color=color, lw=1.2)

11. # 평균 위치 점선
12. ax.vlines(mu, 0, pdf.max(), colors="black",
    linestyles="--", lw=1)

13. # x축 끝 라벨
14. ax.text(31.1, 0, r"$\overline{x}$", fontsize=12,
    va="center")
```

```
15. # 축 스타일: x축 선만 남김
16. for s in ["top", "right", "left"]:
17.     ax.spines[s].set_visible(False)
18. ax.set_yticks([])
19. ax.set_xticks([28.0, 30.5])
20. ax.set_xticklabels(["28.0", "30.5"])
21. ax.tick_params(axis="x", length=0, labelsize=11)
22. ax.set_xlim(25.0, 31.0)
23. ax.set_ylim(0, pdf.max() * 1.05)

24. plt.tight_layout()
25. plt.show()
```

부록#3

- 그래프 코드
  - <그림 4> 예제 8.6 그림

```
1.  import numpy as np
2.  import matplotlib.pyplot as plt

3.  # 두 표본평균 차의 분포: N(0.5, 0.189^2)
4.  mu, sigma = 0.5, 0.189
5.  x = np.linspace(-0.21, 1.15, 1000)
6.  pdf = np.exp(-0.5 * ((x - mu) / sigma) ** 2) /
    (sigma * np.sqrt(2 * np.pi))

7.  color = "green"
8.  fig, ax = plt.subplots(figsize=(7, 3.6))

9.  ax.plot(x, pdf, color=color, lw=1.2)

10. # 평균 위치 점선
11. ax.vlines(mu, 0, pdf.max(), colors="black",
    linestyles="--", lw=1)

12. # 라벨
13. ax.text(0.66, pdf.max() * 0.8,
14.         r"$\sigma_{\overline{x}_1 - \overline{x}_2}
    = 0.189$", fontsize=11)
15. ax.text(1.17, 0, r"$\overline{x}_1 -
    \overline{x}_2$", fontsize=12, va="center")
```

```
16. # 축 스타일: x축 선만 남김
17. for s in ["top", "right", "left"]:
18.     ax.spines[s].set_visible(False)
19. ax.set_yticks([])
20. ax.set_xticks([0.5, 1.0])
21. ax.set_xticklabels(["0.5", "1.0"])
22. ax.tick_params(axis="x", length=0, labelsize=11)
23. ax.set_xlim(-0.21, 1.15)
24. ax.set_ylim(0, pdf.max() * 1.05)

25. # 1.0 오른쪽 꼬리 영역
26. tail = x >= 1.0
27. ax.fill_between(x[tail], pdf[tail], color=color,
    alpha=0.3)
28. ax.vlines(1.0, 0, pdf[tail][0], colors=color, lw=1)

29. plt.tight_layout()
30. plt.show()
```

부록#3

- 그래프 코드
  - <그림 5> t 분포의 대칭성

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt

4.  def t_pdf(x, v):
5.      c = math.lgamma((v + 1) / 2) - math.lgamma(v / 2)
    - 0.5 * math.log(v * math.pi)
6.      return np.exp(c - (v + 1) / 2 * np.log1p(x**2 /
    v))

7.  v = 5            # 자유도
8.  t_a = 1.75       # t_alpha 위치 (그림용)
9.  x = np.linspace(-3, 3, 1000)
10. pdf = t_pdf(x, v)

11. color = "green"
12. fig, ax = plt.subplots(figsize=(7, 3.6))
13. ax.plot(x, pdf, color=color, lw=1.2)

14. # 양쪽 꼬리 영역
15. for mask, edge in [(x <= -t_a, -t_a), (x >= t_a,
    t_a)]:
16.     ax.fill_between(x[mask], pdf[mask], color=color,
    alpha=0.3)
17.     ax.vlines(edge, 0, t_pdf(np.array(edge), v),
    colors=color, lw=1)
```

```
18. # 0 위치 점선
19. ax.vlines(0, 0, pdf.max(), colors="black",
    linestyles="--", lw=1)

20. # 라벨
21. ax.text(3.08, 0, r"$t$", fontsize=12, va="center")

22. for s in ["top", "right", "left"]:
23.     ax.spines[s].set_visible(False)
24. ax.set_yticks([])
25. ax.set_xticks([-t_a, 0, t_a])
26. ax.set_xticklabels([r"$t_{1-\alpha} = -t_{\alpha}$",
    "0", r"$t_{\alpha}$"])
27. ax.tick_params(axis="x", length=0, labelsize=11)
28. ax.set_xlim(-3, 3)
29. ax.set_ylim(0, pdf.max() * 1.05)

30. plt.tight_layout()
31. plt.show()
```

부록#3

- 그래프 코드
  - <그림 6> 자유도에 따른 t 분포

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt

4.  def t_pdf(x, v):
5.      c = math.lgamma((v + 1) / 2) - math.lgamma(v / 2)
    - 0.5 * math.log(v * math.pi)
6.      return np.exp(c - (v + 1) / 2 * np.log1p(x**2 /
    v))

7.  def norm_pdf(x):
8.      return np.exp(-0.5 * x**2) / np.sqrt(2 * np.pi)

9.  x = np.linspace(-2.4, 2.4, 1000)

10. fig, ax = plt.subplots(figsize=(7, 4))
11. ax.plot(x, norm_pdf(x), color="green", lw=1.2)   # v
    = inf (표준정규)
12. ax.plot(x, t_pdf(x, 5), color="black", lw=1.2)   # v
    = 5
13. ax.plot(x, t_pdf(x, 2), color="green", lw=1.2)   # v
    = 2
```

```
14. # 라벨: 곡선 밖에 두고 선으로 연결
15. arrow = dict(arrowstyle="-", lw=0.6)
16. ax.annotate(r"$v = \infty$", xy=(0.25,
    norm_pdf(0.25)), xytext=(0.75, 0.43),
17.             color="green", fontsize=10,
    arrowprops={**arrow, "color": "green"})
18. ax.annotate(r"$v = 5$", xy=(0.45, t_pdf(0.45, 5)),
    xytext=(0.95, 0.395),
19.             color="black", fontsize=10,
    arrowprops={**arrow, "color": "black"})
20. ax.annotate(r"$v = 2$", xy=(0.65, t_pdf(0.65, 2)),
    xytext=(1.15, 0.36),
21.             color="green", fontsize=10,
    arrowprops={**arrow, "color": "green"})

22. for s in ["top", "right", "left"]:
23.     ax.spines[s].set_visible(False)
24. ax.set_yticks([])
25. ax.set_xticks([-2, -1, 0, 1, 2])
26. ax.set_xlim(-2.4, 2.4)
27. ax.set_ylim(0, 0.46)

28. plt.tight_layout()
29. plt.show()
```

부록#3

- 그래프 코드
  - <그림 7> 예제 8.8 그림 (1/2)

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt

4.  def t_pdf(x, v):
5.      c = math.lgamma((v + 1) / 2) - math.lgamma(v / 2)
    - 0.5 * math.log(v * math.pi)
6.      return np.exp(c - (v + 1) / 2 * np.log1p(x**2 /
    v))

7.  v = 14
8.  t_a = 2.145      # t_0.025 (v=14)
9.  x = np.linspace(-4, 4, 1000)
10. pdf = t_pdf(x, v)

11. color = "green"
12. fig, ax = plt.subplots(figsize=(7, 3.6))
13. ax.plot(x, pdf, color=color, lw=1.2)

14. # 왼쪽 꼬리: 구하는 영역 (진하게)
15. left = x <= -t_a
16. ax.fill_between(x[left], pdf[left], color=color,
    alpha=0.45)
17. ax.vlines(-t_a, 0, t_pdf(np.array(-t_a), v),
    colors=color, lw=1)
```

```
18. # 오른쪽 꼬리: 대칭 참고용 (연하게)
19. right = x >= t_a
20. ax.fill_between(x[right], pdf[right], color=color,
    alpha=0.15)
21. ax.vlines(t_a, 0, t_pdf(np.array(t_a), v),
    colors=color, lw=1, linestyles="--")

22. # 0 위치 점선
23. ax.vlines(0, 0, pdf.max(), colors="black",
    linestyles="--", lw=1)

24. # 면적 라벨 + 화살표
25. ax.annotate("0.025", xy=(-2.6, 0.012), xytext=(-3.6,
    0.12),
26.             fontsize=11,
    arrowprops=dict(arrowstyle="->", lw=0.8))
27. ax.annotate("0.025", xy=(2.6, 0.012), xytext=(3.0,
    0.12),
28.             fontsize=11, color="gray",
    arrowprops=dict(arrowstyle="->", lw=0.8,
    color="gray"))
29. ax.text(2.1, pdf.max() * 0.8, r"$v = 14$",
    fontsize=11)
30. ax.text(4.08, 0, r"$t$", fontsize=12, va="center")
```

부록#3

- 그래프 코드
  - <그림 7> 예제 8.8 그림 (2/2)

```
31. for s in ["top", "right", "left"]:
32.     ax.spines[s].set_visible(False)
33. ax.set_yticks([])
34. ax.set_xticks([-t_a, 0, t_a])
35. ax.set_xticklabels([r"$t_{0.975} = -2.145$", "0",
    r"$t_{0.025} = 2.145$"])
36. ax.tick_params(axis="x", length=0, labelsize=10)
37. ax.set_xlim(-4, 4)
38. ax.set_ylim(0, pdf.max() * 1.05)

39. plt.tight_layout()
40. plt.show()
```

부록#3

- 그래프 코드
  - <그림 8> 예제 8.9 그림

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt

4.  def t_pdf(x, v):
5.      c = math.lgamma((v + 1) / 2) - math.lgamma(v / 2)
    - 0.5 * math.log(v * math.pi)
6.      return np.exp(c - (v + 1) / 2 * np.log1p(x**2 /
    v))

7.  v = 14
8.  k = -2.977       # -t_0.005 (v=14)
9.  b = -1.761       # -t_0.05  (v=14)
10. x = np.linspace(-3.05, 3.0, 1000)
11. pdf = t_pdf(x, v)

12. color = "green"
13. fig, ax = plt.subplots(figsize=(7, 3.6))
14. ax.plot(x, pdf, color=color, lw=1.2)
```

```
15. # k ~ -t_0.05 사이 영역 (면적 0.045)
16. mask = (x >= k) & (x <= b)
17. ax.fill_between(x[mask], pdf[mask], color=color,
    alpha=0.3)
18. ax.vlines(b, 0, t_pdf(np.array(b), v), colors=color,
    lw=1)
19. ax.vlines(k, 0, t_pdf(np.array(k), v), colors=color,
    lw=1)

20. # 0 위치 점선
21. ax.vlines(0, 0, pdf.max(), colors="black",
    linestyles="--", lw=1)

22. # 라벨
23. ax.text(-2.5, 0.008, "0.045", fontsize=10)
24. ax.text(3.08, 0, r"$t$", fontsize=12, va="center")

25. for s in ["top", "right", "left"]:
26.     ax.spines[s].set_visible(False)
27. ax.set_yticks([])
28. ax.set_xticks([k, b, 0])
29. ax.set_xticklabels([r"$k$", r"$-t_{0.05}$", "0"])
30. ax.tick_params(axis="x", length=0, labelsize=11)
31. ax.set_xlim(-3.05, 3.0)
32. ax.set_ylim(0, pdf.max() * 1.05)

33. plt.tight_layout()
34. plt.show()
```

부록#3

- 그래프 코드
  - <그림 9> 전형적인 F 분포

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt

4.  def f_pdf(x, d1, d2):
5.      c = (math.lgamma((d1 + d2) / 2) - math.lgamma(d1
    / 2) - math.lgamma(d2 / 2)
6.           + (d1 / 2) * math.log(d1 / d2))
7.      return np.exp(c + (d1 / 2 - 1) * np.log(x) - (d1
    + d2) / 2 * np.log1p(d1 * x / d2))

8.  x = np.linspace(1e-4, 3.2, 1000)
9.  y1 = f_pdf(x, 10, 30)
10. y2 = f_pdf(x, 6, 10)

11. fig, ax = plt.subplots(figsize=(7, 4))
12. ax.plot(x, y1, color="green", lw=1.2)   # d.f. = (10,
    30)
13. ax.plot(x, y2, color="black", lw=1.2)   # d.f. = (6,
    10)
```

```
14. # 라벨
15. ax.text(1.15, y1.max() * 0.9, "d.f. = (10, 30)",
    color="green", fontsize=10)
16. ax.text(0.55, y1.max() * 0.4, "d.f. = (6, 10)",
    color="black", fontsize=10)
17. ax.text(3.25, 0, r"$f$", fontsize=12, va="center")

18. # 축 스타일: 왼쪽, 아래 축만
19. ax.spines["top"].set_visible(False)
20. ax.spines["right"].set_visible(False)
21. ax.set_yticks([])
22. ax.set_xticks([0])
23. ax.tick_params(axis="x", length=0, labelsize=11)
24. ax.set_xlim(0, 3.2)
25. ax.set_ylim(0, y1.max() * 1.05)

26. plt.tight_layout()
27. plt.show()
```

부록#3

- 그래프 코드
  - <그림 10> F 분포의 역수 관계 (1/2)

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt

4.  def f_pdf(x, d1, d2):
5.      c = (math.lgamma((d1 + d2) / 2) - math.lgamma(d1
    / 2) - math.lgamma(d2 / 2)
6.           + (d1 / 2) * math.log(d1 / d2))
7.      return np.exp(c + (d1 / 2 - 1) * np.log(x) - (d1
    + d2) / 2 * np.log1p(d1 * x / d2))

8.  def f_upper(alpha, d1, d2):
9.      """오른쪽 면적이 alpha인 f_alpha(d1, d2)
    (수치적분)"""
10.     g = np.logspace(-6, 6, 400000)
11.     p = f_pdf(g, d1, d2)
12.     cdf = np.concatenate([[0], np.cumsum((p[1:] +
    p[:-1]) / 2 * np.diff(g))])
13.     return np.interp(1 - alpha, cdf, g)

14. v1, v2 = 6, 10
15. lo = 1 / f_upper(0.05, v2, v1)    # f_0.95(6,10) = 1
    / f_0.05(10,6) = 0.246
16. hi = f_upper(0.05, v1, v2)        # f_0.05(6,10) =
    3.22
```

```
17. x = np.linspace(1e-4, 4.5, 1500)
18. y = f_pdf(x, v1, v2)

19. color = "green"
20. fig, ax = plt.subplots(figsize=(8, 4))
21. ax.plot(x, y, color=color, lw=1.3)

22. # 양쪽 꼬리 (각각 면적 0.05)
23. for mask, edge in [(x <= lo, lo), (x >= hi, hi)]:
24.     ax.fill_between(x[mask], y[mask], color=color,
    alpha=0.35)
25.     ax.vlines(edge, 0, f_pdf(np.array(edge), v1, v2),
    colors=color, lw=1)

26. # 면적 라벨
27. ax.annotate("0.05", xy=(lo * 0.6, 0.03),
    xytext=(0.45, 0.35),
28.             fontsize=11,
    arrowprops=dict(arrowstyle="->", lw=0.8))
29. ax.annotate("0.05", xy=(hi + 0.3, 0.01), xytext=(hi
    + 0.1, 0.15),
30.             fontsize=11,
    arrowprops=dict(arrowstyle="->", lw=0.8))
```

부록#3

- 그래프 코드
  - <그림 10> F 분포의 역수 관계 (2/2)

```
# 관계식
ax.text(4.4, y.max() * 0.95,
        r"$f_{0.95}(6, 10) = \dfrac{1}{f_{0.05}(10, 6)} = \dfrac{1}{4.06} = 0.246$",
        ha="right", va="top", fontsize=12)
ax.text(4.4, y.max() * 0.72, r"$v_1 = 6,\ v_2 = 10$", ha="right", va="top", fontsize=11)

ax.text(4.58, 0, r"$f$", fontsize=12, va="center")
ax.spines["top"].set_visible(False)
ax.spines["right"].set_visible(False)
ax.set_yticks([])
ax.set_xticks([0, lo, hi])
ax.set_xticklabels(["0", "\n" + rf"$f_{{0.95}} = {lo:.3f}$", rf"$f_{{0.05}} = {hi:.2f}$"])
ax.tick_params(axis="x", length=0, labelsize=10)
ax.set_xlim(0, 4.5)
ax.set_ylim(0, y.max() * 1.08)

plt.tight_layout()
plt.show()
```

부록#3

- 그래프 코드
  - <그림 11> 분위수그림

```
1.  import numpy as np
2.  import matplotlib.pyplot as plt

3.  # 예시 데이터
4.  data = [12, 15, 18, 19, 21, 22, 24, 25, 27, 30,
5.           31, 33, 35, 36, 38, 40, 42, 45, 47, 50]

6.  # 1. 데이터 오름차순 정렬
7.  data = np.sort(data)

8.  # 2. 데이터 개수 및 순위
9.  n = len(data)
10. i = np.arange(1, n + 1)

11. # 3. 순위에 대응하는 비율 계산
12. f_i = (i - 3/8) / (n + 1/4)

13. # 4. 가로선 형태의 분위수 그림
14. plt.figure(figsize=(8, 5))
```

```
15. # 각 데이터값에 짧은 가로선 표시
16. line_width = 0.035
17. plt.hlines(
18.     y=data,
19.     xmin=f_i - line_width / 2,
20.     xmax=f_i + line_width / 2,
21.     color="green",
22.     linewidth=2
23. )

24. plt.xlabel("f")
25. plt.ylabel("data")
26. plt.xlim(0, 1)
27. plt.grid(axis="y", linestyle="--", alpha=0.3)
28. plt.tight_layout()
29. plt.show()
```

부록#3

- 그래프 코드
  - <그림 12> 정규분위수그림

```
1.  import numpy as np
2.  import matplotlib.pyplot as plt
3.  from statistics import NormalDist

4.  # Sample data: 30 observations from N(50, 5^2)
5.  rng = np.random.default_rng(7)
6.  data = np.sort(rng.normal(50, 5, 30))
7.  n = len(data)

8.  # Theoretical quantiles: standard normal quantile of
    f_i = (i - 3/8) / (n + 1/4)
9.  i = np.arange(1, n + 1)
10. f = (i - 3 / 8) / (n + 1 / 4)
11. q = np.array([NormalDist().inv_cdf(p) for p in f])

12. # Reference line: y = x¯ + s * q
13. xbar, s = data.mean(), data.std(ddof=1)
14. qq = np.array([q.min() - 0.3, q.max() + 0.3])
```

```
15. color = "green"
16. fig, ax = plt.subplots(figsize=(6, 5))
17. ax.scatter(q, data, color=color, marker="_", s=200,
    linewidths=2,   # ← 여기만 바꿈
18.            zorder=3, label="Observed data")
19. ax.plot(qq, xbar + s * qq, color="black", lw=1,
    ls="--",
20.         label=rf"Reference line $y = \bar{{x}} +
    s\,q$")

21. ax.set_xlabel("Standard normal quantile
    $q_{0,1}(f_i)$")
22. ax.set_ylabel("Observed value $y_{(i)}$")
23. ax.legend(frameon=False)
24. ax.spines["top"].set_visible(False)
25. ax.spines["right"].set_visible(False)

26. plt.tight_layout()
27. plt.show()
```

부록#3

- 그래프 코드
  - <그림 13> 정규분포를 따르지 않는 정규확률그림

```
1.  import numpy as np
2.  import matplotlib.pyplot as plt
3.  from statistics import NormalDist

4.  Z = NormalDist()

5.  # Not normal: exponential sample (right-skewed)
6.  rng = np.random.default_rng(7)
7.  data = np.sort(rng.exponential(5, 30))
8.  n = len(data)

9.  i = np.arange(1, n + 1)
10. f = (i - 3 / 8) / (n + 1 / 4)                # cumulative probability
11. z = np.array([Z.inv_cdf(p) for p in f])      # plotted on normal scale

12. # reference line: z = (x - xbar) / s
13. xbar, s = data.mean(), data.std(ddof=1)
14. xx = np.array([data.min(), data.max()])
```

```
15. fig, ax = plt.subplots(figsize=(6, 5))
16. ax.scatter(data, z, color="green", s=25, zorder=3, label="Observed data")
17. ax.plot(xx, (xx - xbar) / s, color="black", lw=1, ls="--",
18.         label="Normal reference line")

19. # y-axis: normal probability scale (labels in %)
20. probs = [0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.99]
21. ax.set_yticks([Z.inv_cdf(p) for p in probs])
22. ax.set_yticklabels([f"{p * 100:g}" for p in probs])
23. ax.set_ylim(Z.inv_cdf(0.005), Z.inv_cdf(0.995))
24. ax.grid(axis="y", lw=0.4, alpha=0.5)

25. ax.set_title("Data not following a normal distribution", fontsize=12)
26. ax.set_xlabel("Observed value")
27. ax.set_ylabel("Cumulative probability (%)")
28. ax.legend(frameon=False, loc="upper left")
29. ax.spines["top"].set_visible(False)
30. ax.spines["right"].set_visible(False)

31. plt.tight_layout()
32. plt.show()
```

부록#3

- 그래프 코드
  - <그림 14> 정규분포를 따르는 정규확률그림

```
1.  import numpy as np
2.  import matplotlib.pyplot as plt
3.  from statistics import NormalDist

4.  Z = NormalDist()

5.  # Normal: N(50, 5^2) sample
6.  rng = np.random.default_rng(7)
7.  data = np.sort(rng.normal(50, 5, 30))
8.  n = len(data)

9.  i = np.arange(1, n + 1)
10. f = (i - 3 / 8) / (n + 1 / 4)              #
    cumulative probability
11. z = np.array([Z.inv_cdf(p) for p in f])    #
    plotted on normal scale

12. # reference line: z = (x - xbar) / s
13. xbar, s = data.mean(), data.std(ddof=1)
14. xx = np.array([data.min(), data.max()])
```

```
15. fig, ax = plt.subplots(figsize=(6, 5))
16. ax.scatter(data, z, color="green", s=25, zorder=3,
    label="Observed data")
17. ax.plot(xx, (xx - xbar) / s, color="black", lw=1,
    ls="--",
18.         label="Normal reference line")

19. # y-axis: normal probability scale (labels in %)
20. probs = [0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90,
    0.95, 0.99]
21. ax.set_yticks([Z.inv_cdf(p) for p in probs])
22. ax.set_yticklabels([f"{p * 100:g}" for p in probs])
23. ax.set_ylim(Z.inv_cdf(0.005), Z.inv_cdf(0.995))
24. ax.grid(axis="y", lw=0.4, alpha=0.5)

25. ax.set_title("Data following a normal distribution",
    fontsize=12)
26. ax.set_xlabel("Observed value")
27. ax.set_ylabel("Cumulative probability (%)")
28. ax.legend(frameon=False, loc="upper left")
29. ax.spines["top"].set_visible(False)
30. ax.spines["right"].set_visible(False)

31. plt.tight_layout()
32. plt.show()
```