2026/03/12, 2026 확률 기초 세미나

이공학도를 위한 확률 및 통계학

-8장 확률표본과 표본분포-

민 윤 홍(yunhong@pel.sejong.ac.kr)

세종대학교 프로토콜공학연구실

목 차

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

목 차

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

확률표본(Random Sample)

- 모집단(Population)
  - 정의
    - 관심이 있는 대상과 관련된 모든 관측가능한 값의 집합
      - 모집단에서 각각의 관측값은 확률분포 $f(x)$를 가지는 확률변수 X의 값이 됨
        - e.g., 공정에서 제품의 불량을 1, 양호한 경우를 0이라고 하면, 해당 모집단의 관측값 X는 $b(x; 1, p) = p^x q^{1-x}, x = 0, 1$의 분포를 따르는 베르누이확률변수가 됨
  - 종류
    - 유한모집단
      - 조사 대상 전체의 수가 고정되어 있거나 셀 수 있는 집단
        - e.g., PEL 인원들의 키
    - 무한모집단
      - 조사 대상 전체의 수가 끊임없이 생성되거나 무한히 반복하는 집단
        - e.g., 주사위를 무한히 던져서 생성될 수 있는 숫자들의 조합

확률표본

- 표본(Sample)
  - 정의
    - 모집단의 부분집합
  - 필요성
    - 모집단을 구성하는 전체를 관측하기 불가능할 때 사용함
      - e.g., 전구의 평균 수명을 결정하기 위해 모든 전구를 조사하는 일
  - 특징
    - 모집단의 성질을 잘 반영할 수 있도록 해야 함
      - 일관되게 과대 혹은 과소추정할 경우 편의(bias) 혹은 편향되어 있다고 판단함

확률표본

- 정의
  - 모집단의 모든 구성 요소가 표본으로 추출될 확률이 동등하도록 무작위로 뽑은 표본

- 특징
  - 추출된 표본들은 상호 독립적이고 동일한 모집단에서 추출되었으므로 동일한 분포를 가짐
  - 서로 독립인 n개의 확률변수 $X_1, X_2, \ldots, X_n$이 동일한 확률분포 $f(x)$를 따를 때, $X_1, X_2, \ldots, X_n$을 모집단 $f(x)$로부터의 크기 n인 확률표본이라고 함
    - 결합확률분포는 $f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$으로 표현됨

목 차

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

대표적 통계량

- 대표본(Large Sample)
  - 정의
    - 표본의 크기가 $n \geq 30$인 표본
  - 필요성
    - 미지의 모집단에 대한 정보를 얻기 위해 대표본을 추출하여 추론할 수 있음
      - e.g., 커피 애호가들 중 스타벅스를 선호하는 집단
  - 특징
    - 추론하고자 하는 정보를 p라고 하면 대표본을 통해 추론한 정보는 $\hat{p}$으로 표기함
      - $\hat{p}$은 표본에 따라 달라지는 값임

대표적 통계량

- 표본의 중심위치
  - 정의
    - 모집단에서 추출한 표본 데이터가 어디에 집중되어 있는지 나타내는 대푯값
  - 종류
    - 표본평균(Sample Mean)
      - 모집단에서 추출한 표본 데이터들의 평균
      - $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $(X_1, X_2, \ldots, X_i$는 표본 데이터) 로 표기
    - 표본중앙값(Sample Median)
      - 모집단에서 추출한 표본들의 중앙값 (n = 표본의 개수)
      - $\bar{x} = \begin{cases} x_{(n+1)/2}, & n\text{이 홀수일 때} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right), & n\text{이 짝수일 때} \end{cases}$
    - 표본최빈값(Sample Mode)
      - 모집단에서 가장 많이 발생한 표본값

대표적 통계량

- 표본의 중심위치
  - 예제 8.1

관측값들이 다음과 같을 때, 표본평균, 표본중앙값, 표본최빈값을 각각 구하여라.

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

- 표본평균 $\bar{X} = \frac{0.32+0.53+0.28+0.37+0.47+0.43+0.36+0.42+0.38+0.43}{10} = 0.399$
- 표본중앙값
  - 관측값들을 오름차순으로 정리하면
    $$0.28, 0.32, 0.36, 0.37, \mathbf{0.38}, \mathbf{0.42}, 0.43, 0.43, 0.47, 0.53$$
  - 표본의 개수가 10개 이므로 $\frac{1}{2} \times (0.38 + 0.42) = 0.4$
- 표본최빈값은 가장 많이 발생한 값이므로 0.43이 됨

대표적 통계량

- 표본의 산포
  - 정의
    - 모집단에서 추출한 표본 데이터들이 대푯값을 중심으로 얼마나 흩어져 있는지를 나타내는 값
  - 종류 (1/3)
    - 표본분산(Sample Variance)
    - 표본표준편차(Sample Standard Deviation)
    - 표본범위(Sample Range)

대표적 통계량

- 표본의 산포
  - 종류 (2/3)
    - 표본분산
      - 모집단에서 추출한 표본의 흩어진 정도를 나타내는 통계량
      - $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 로 표기
      - $S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2\right]$로도 표기
      - 증명

표본분산의 정의에 의하여

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

$\bar{X}$를 $\sum_{i=1}^{n}\frac{X_i}{n}$으로 치환하고 분모 분자에 n을 곱하면

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

대표적 통계량

- 표본의 산포
  - 종류 (3/3)
    - 표본표준편차(Sample Standard Deviation)
      - 모집단에서 추출한 표본이 표본평균을 중심으로 얼마나 퍼져 있는지를 나타내는 측정값
      - $S = \sqrt{S^2}$ 로 표기
    - 표본범위(Sample Range)
      - 모집단에서 최댓값과 최솟값의 차이
      - $R = X_{max} - X_{min}$

대표적 통계량

- 표본의 산포
  - 예제 8.2

세종대학교에 있는 상가 중 임의로 4곳을 선택하여 200g 병커피의 가격을 비교한 결과 지난 달보다 12원, 15원 17원, 20원씩 인상되었다. 가격인상에 대한 확률분포의 평균과 분산을 구하여라.

- 표본평균 $\bar{x} = \frac{12+15+17+20}{4} = 16$원이므로, 표본분산 s^2은 다음과 같이 구할 수 있음
  - $s^2 = \frac{1}{3}\sum_{i=1}^{4}(x_i - 16)^2 = \frac{(12-16)^2+(15-16)^2+(17-16)^2+(20-16)^2}{3} = \frac{34}{3}$
  - 평균 16을 기준으로 인상액이 제곱을 기준으로 $\frac{34}{3}$만큼 흩어져 있음을 알 수 있음
    - 표준편차 $s \approx 3.37$로 평균 16을 기준으로 3.37만큼 흩어져 있음을 알 수 있음

- 만약 건국대학교에 있는 상가들의 인상액이 평균 16원이고 표준편차가 $\frac{43}{3}$인 경우
  - 평균 인상액은 동일하나 인상액이 제곱을 기준으로 $\frac{43}{3}$만큼 흩어져 있음을 알 수 있음
  - 건대 표준편차 $s_2 \approx 3.7$로 평균 16을 기준으로 3.7만큼 흩어져 있음을 알 수 있음
  - 결과적으로, 세종대학교보다 건국대학교 상가들의 인상액이 더 불규칙적임을 알 수 있음

대표적 통계량

- 표본의 산포
  - 예제 8.3

어느 날 한강에서 임의로 선정된 6명의 낚시꾼들에 의해 잡힌 광어의 수는 각각 3, 4, 5, 6, 6, 7마리 일 때, 표본분산과 표본범위를 구하여라.

- $\sum_{i=1}^{6} x_i^2 = 171, \sum_{i=1}^{6} x_i = 31, S = 6$이므로
  - $s^2 = \frac{1}{(6)(5)}[(6)(171) - (31)^2] = \frac{13}{6}$
- $X_{max} = 7, X_{min} = 3$이므로 표본범위는 $7 - 3 = 4$

목 차

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

표본평균의 분포

- 표본분포(Sample Distribution)
  - 정의
    - 통계량의 확률분포
      - 동일한 모집단에서 같은 크기의 표본을 반복적으로 추출했을 때 얻어지는 표본 통계량
        - e.g., 선거에서 특정후보에 대한 지지율을 알아보기 위해 통행하는 사람들의 의견을 물어본 후 얻어지는 통계량
  - 필요성
    - 모집단 전체를 조사할 수 없을 때, 추출한 표본 통계량을 바탕으로 모집단이 나타내는 확률적 성질을 추정할 수 있음

표본평균의 분포

- 정의
  - 표본평균 $\bar{X}$의 확률분포
- 특징
  - 모집단 X와 동일한 정규분포를 따름
  - 표본평균 $\bar{X}_i = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$의 평균과 분산은 다음과 같음
    - $\mu_{\bar{X}} = \frac{1}{n}\left(\underbrace{\mu + \mu + \cdots \mu}_{n}\right) = \mu$
    - $\sigma_{\bar{X}}^2 = \frac{1}{n^2}\left(\underbrace{\sigma^2 + \sigma^2 + \cdots \sigma^2}_{n}\right) = \frac{\sigma^2}{n}$

표본평균의 분포

- 중심극한정리(CLT, Central Limit Theorem)
  - 정의
    - 모집단의 분포와 상관없이, 표본의 크기 n이 $n \geq 30$ 이라면, 표본평균들의 분포가 정규분포에 근사하다는 정리
  - 수식

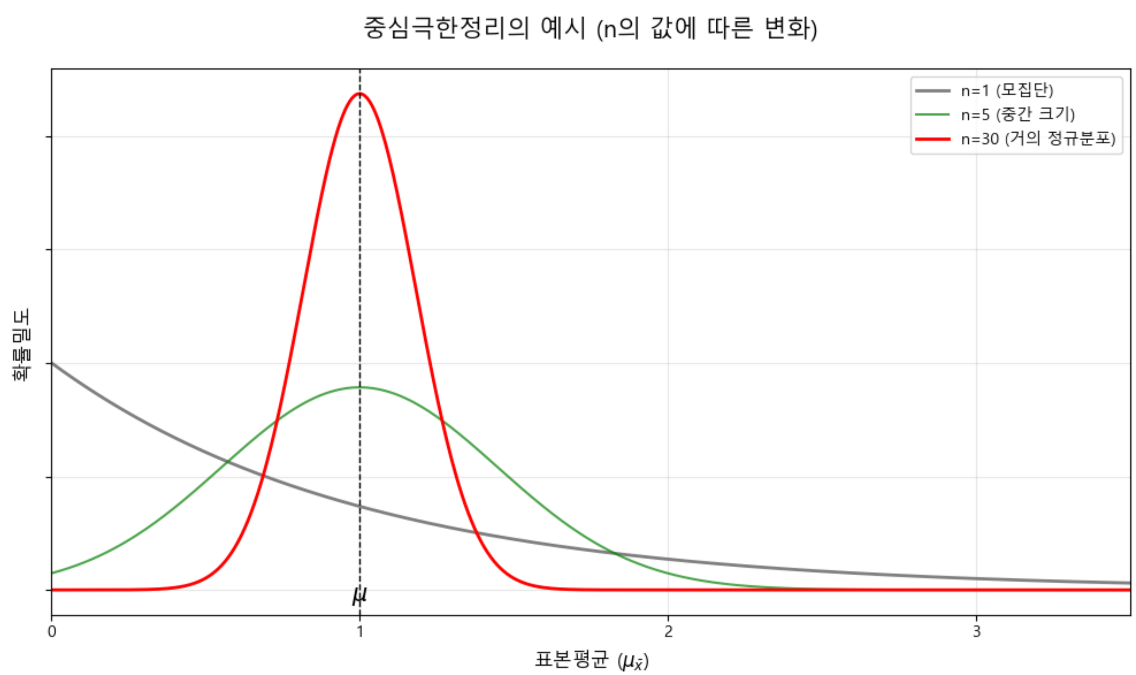
평균이 μ, 분산이 σ^2인 모집단으로부터 크기 n인 확률표본을 추출했을 때, 표본의 평균 $\bar{X}$에 대해서

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

는 $n \geq 30$ 일 때 표준정규분포 $n(z; 0,1)$에 접근한다.

표본평균의 분포

- 중심극한정리
  - 특징
    - n이 커짐에 따라 정규분포에 근접함
    - $\bar{X}$의 평균은 표본에 상관없이 항상 μ임
    - $\bar{X}$의 분산은 n이 커질수록 작아짐



<그림 1> 중심극한정리 그래프

표본평균의 분포

- 중심극한정리
  - 예제 8.4

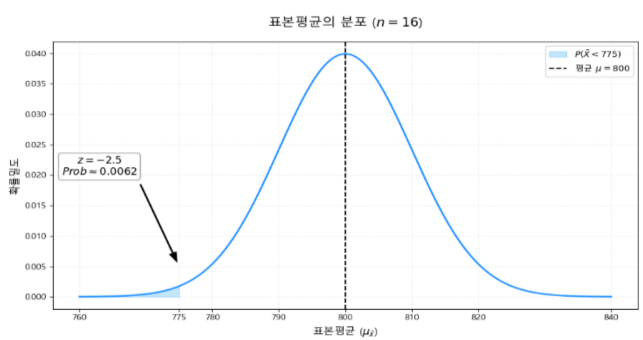
한 공장에서 생산되는 전구의 수명은 평균이 800시간이고, 표준편차가 40시간인 정규분포를 따른다고 할 때, 임의로 추출한 16개의 전구의 평균수명($\bar{X}$)이 775시간 미만일 확률을 구하여라.

- $\bar{X}$의 분포는 $\mu_{\bar{X}} = 800, \sigma_{\bar{X}} = \frac{40}{\sqrt{16}} = 10$ 인 정규분포를 따름
- 구하고자 하는 확률값은 $\bar{X} < 775$이고 $\bar{x} = 775$에 대응하는 표준정규계수 z는 다음과 같음

$$z = \frac{775 - 800}{10} = -2.5$$

- 구하고자 하는 확률값은

$$P(\bar{X} < 775) = P(Z < -2.5) = 0.0062$$



<그림 2> 예제 8.4 그래프

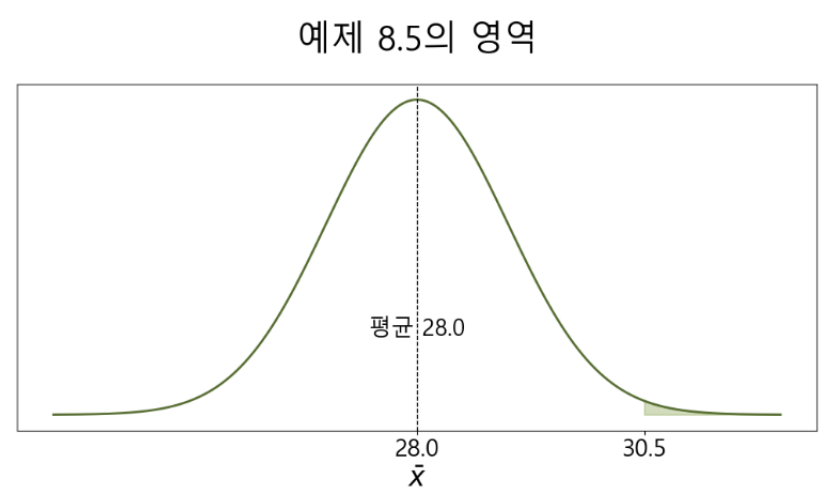
표본평균의 분포

- 중심극한정리
  - 예제 8.5

세종대학교와 상명대학교 사이를 운행하는 셔틀버스의 운행시간은 평균 28분, 표준편차 5분의 분포를 따른다고 한다. 어느 한 주 동안 40번의 운행이 있었다고 할 때, 평균 운행시간이 30분보다 길 확률은 얼마인가?
평균시간은 분 단위로 반올림하여 측정된다고 가정한다.

- 확률분포의 평균과 분산은 $\mu_{\bar{X}} = 28, \sigma_{\bar{X}} = 3, n = 40$ 임
- 시간은 분 단위로 반올림하여 측정하니 $\bar{x} \geq 30.5$를 구하면 됨
- 따라서

$P(\bar{X} > 30) = P\left(\frac{\bar{X}-28}{5/\sqrt{40}} \geq \frac{30.5-28}{5/\sqrt{40}}\right)$
$= P(Z \geq 3.16)$(소수 둘째 자리까지 반올림)
$= 0.0008$



<그림 3> 예제 8.5 그래프

표본평균의 분포

- 두 표본평균 차이의 분포
  - 정의
    - 독립인 두 모집단에서 크기가 다른 표본을 추출했을 때, 표본의 평균의 차이가 나타내는 분포
  - 특징
    - 두 표본의 크기가 각각 30 이상이면 두 모집단의 분포에 관계 없이 정규분포에 근사함
    - $\bar{X}_1 - \bar{X}_2$의 분포는 근사적으로 $N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$을 따름
    - $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$은 근사적으로 표준정규분포를 따름

표본평균의 분포

- 두 표본평균 차이의 분포
  - 예제 8.6

A, B 두 종류의 페인트를 사용하여 각각 18개의 시편에 페인트칠을 한 후 각각의 건조시간을 측정하였다. 두 모집단의 표준편차는 똑같이 1.0이고, 평균건조시간이 같다고 가정했을 때, $P(\bar{X}_A - \bar{X}_B > 1.0)$의 값을 구하여라.
단, $n_A = n_B = 18$이고 $\bar{X}_A$와 $\bar{X}_B$는 각 표본의 평균건조시간을 의미한다.

- $\bar{X}_A - \bar{X}_B$의 분포는 평균과 분산이 각각

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_A - \mu_B = 0$$

$$\sigma^2_{\bar{X}_A - \bar{X}_B} = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}$$

- 구하고자 하는 확률값을 구하기 위해 $\bar{X}_A - \bar{X}_B$를 표준정규화시키면,

$$z = \frac{1 - (\mu_A - \mu_B)}{\sqrt{1/9}} = \frac{1 - 0}{\sqrt{1/9}} = 3.0$$

- 따라서

$$P(Z > 3.0) = 1 - P(Z < 3.0) = 1 - 0.9987 = 0.0013$$

표본평균의 분포

- 두 표본평균 차이의 분포
  - 예제 8.7

A, B 두 회사의 브라운관이 있다. A회사 제품의 수명은 **평균 6.5년, 표준편차가 0.9년,** B회사 제품의 수명은 **평균 6년, 표준편차가 0.8년**이라고 하자. **A회사에서 36개**의 확률표본을 추출하고, **B회사에서는 49개**의 확률표본을 추출했을 때, A회사제품의 표본평균이 B회사 제품의 표본평균보다 적어도 1년 이상 길 확률을 구하라.

- $\bar{X}_A - \bar{X}_B$의 분포는 근사적으로 해석한다고 가정
- 평균은 $\mu_{\bar{X}_A - \bar{X}_B} = 6.5 - 6.0 = 0.5$, 표준편차는 $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{0.81}{36} + \frac{0.64}{49}} = 0.189$인 정규분포를 따름
- $\bar{x}_A - \bar{x}_B = 1.0$에 대응하는 z값은 다음과 같음

$$z = \frac{1.0 - 0.5}{0.189} = 2.65$$

- 따라서

$$P(\bar{X}_A - \bar{X}_B \geq 1.0) = P(Z > 2.65) = 1 - P(Z < 2.65) = 1 - 0.9960 - 0.0040$$

목 차

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

표본분산의 분포

- 필요성
  - 모집단에 표본을 추출하여 평균을 구했을 때, 분산 S^2의 분포를 파악해 표본의 변동성에 대한 정보를 얻을 수 있음

- 특징
  - 분산이 σ^2인 정규모집단으로부터 크기 n인 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 따른다.

표본분산의 분포

- 예제 8.8

어느 회사에서 제조한 배터리의 수명이 평균은 3년, 표준편차는 1년이라고 주장할 때, 5개의 배터리를 임의로 추출하여 시험한 결과 수명이 각각 1.9년, 2.4년, 3.0년, 3.5년, 4.2년이었다. 이 결과를 가지고 회사가 주장하는 표준편차가 1년이라는 것을 믿을 수 있는가 보이시오. 배터리의 수명은 정규분포를 따른다고 가정한다.

- 5개의 배터리의 표본 데이터를 구해 보면

$$\bar{X} = 3,$$

$$S^2 = \frac{(5)(1.9^2 + 2.4^2 + 3^2 + 3.5^2 + 4.2^2) - (15)^2}{(5)(4)} = \frac{(5)(48.26) - (15)^2}{(5)(4)} = 0.815,$$

$$S = 0.903$$

- 따라서

$$\chi^2 = \frac{(4)(0.815)}{1} = 3.26$$

이 되고, 이는 자유도 5-1=4인 카이제곱분포를 따르는 값이 됨

- 자유도가 4인 경우 χ^2의 값의 95%가 0.484와 11.143 사이에 오게 되므로 $\sigma^2 = 1$로 계산된 χ^2의 값 3.26으로 표준편차가 1년이라는 주장은 합당하다고 볼 수 있음

목 차

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

t 분포

- 정의
  - 표본이 적은 통계량에서 정규분포 대신 사용되는 확률분포
- 필요성
  - 모집단이 정규분포라는 것만 알고 모 표준편차는 모를 때, 표본을 통해 모평균을 추정할 때 사용함

t 분포

- 수식

Z와 V가 각각 표준정규확률변수와 자유도 v인 카이제곱확률변수이고, Z와 V도 서로 독립일 때

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

의 확률밀도 함수는

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

가 되고, 이는 자유도 v인 t 분포를 따른다.

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

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

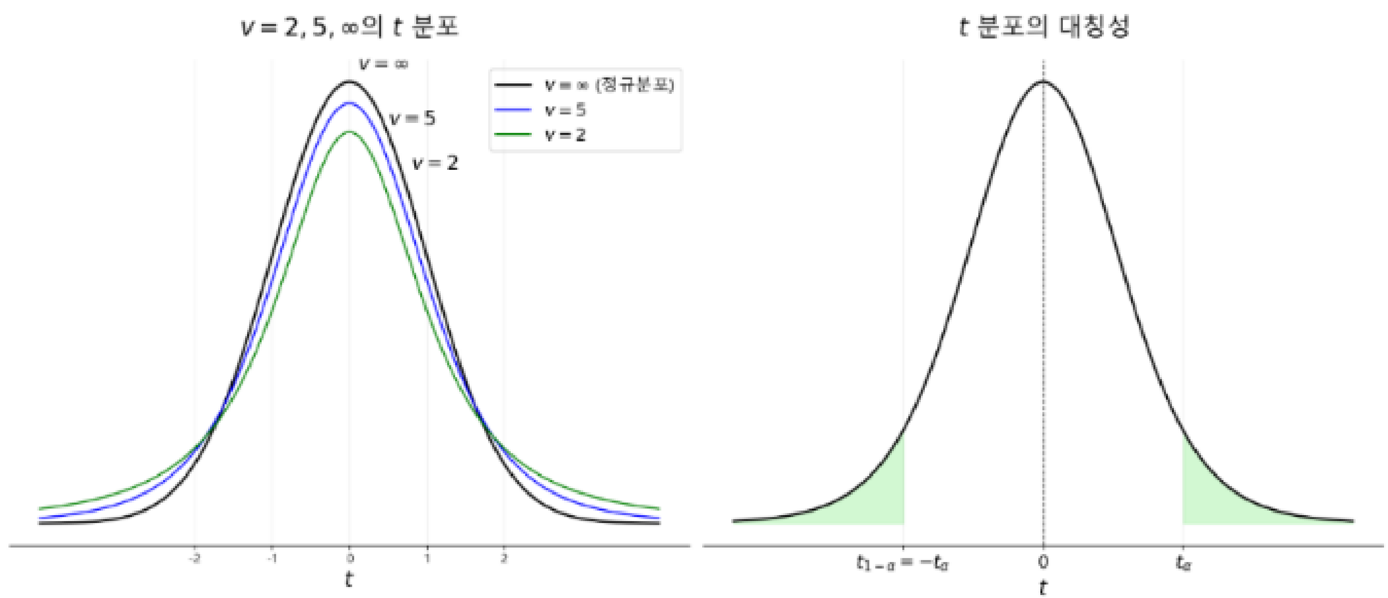
t 분포

- 카이제곱분포(Chi-Squared Distribution)
  - 정의
    - k개의 서로 독립적인 표준정규분포를 따르는 확률변수를 각각 제곱한 후 더하여 얻어지는 분포
  - 특징
    - k의 자유도*를 가짐
      - k에 따라 분포의 모양이 달라짐
      - k가 작을수록 중심이 0에 치우치고 비대칭성이 커짐
      - k가 커질수록 중심이 오른쪽으로 이동하고 정규분포에 근접함
    - t 분포에서 분포의 모양을 결정
      - 표본 평균을 사용하는 t 분포는 마지막 데이터 한 개가 평균값을 맞추기 위해 강제로 결정되기 때문임

*자유도(Degrees of Freedom)
통계적 결정을 내릴 때, 독립적으로 자유롭게 변할 수 있는 값의 개수

t 분포

- 특징
  - 표준정규분포와 같이 원점을 중심으로 대칭을 이룸
  - 자유도v가 커질수록 표준정규분포에 접근함
    - $v \leq 30$이면 표준정규분포보다 평평함
    - $v > 30$이면 표준정규분포와 비슷함



<그림 4> 자유도에 따른 t 분포 그래프와 t 분포의 대칭성

t 분포

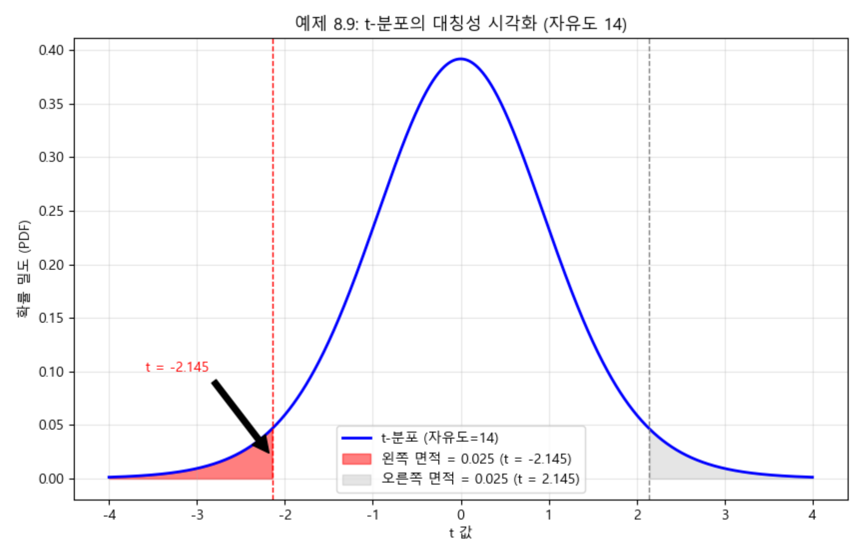
- 예제 8.9

자유도 14인 t 분포에서 왼쪽 면적이 0.025가 되는 t값을 구하여라.

- $v = 14$를 t 분포 수식에 대입하면

$$\frac{\Gamma(15/2)}{\Gamma(7)\sqrt{14\pi}} = \left(\frac{13}{2}\right)! \times \frac{1}{6! \times \sqrt{14\pi}} \times \left(1 + \frac{t^2}{14}\right)^{-15/2} = 0.025$$

- $t_{0.025} = 2.145$
- 구하고자 하는 값은 오른쪽 면적이 아닌 왼쪽 면적이 0.025가 되는 t값이므로 대칭성을 이용, 2.145의 음수인 -2.145임
- 따라서 $t_{0.975} = -t_{0.025} = -2.145$



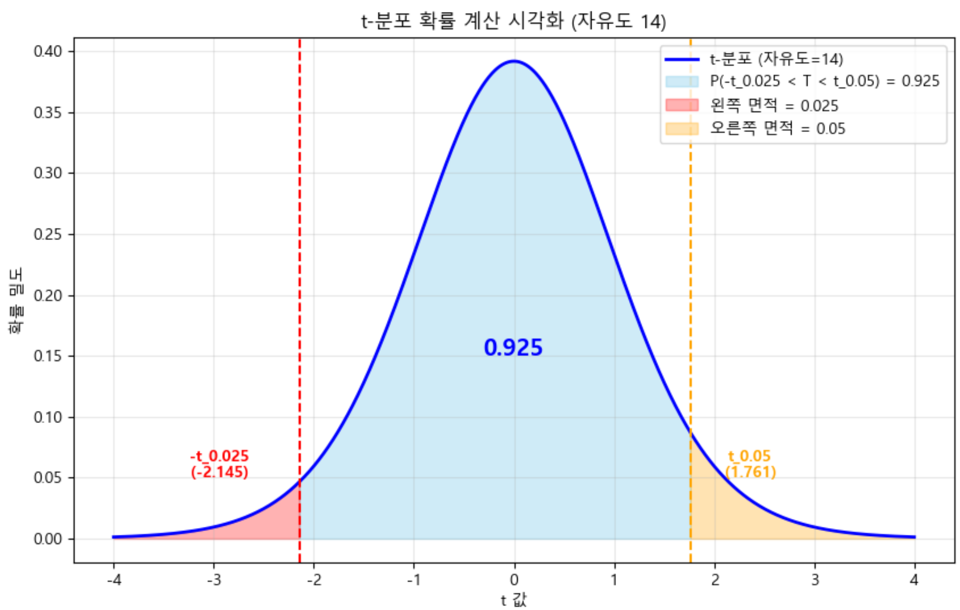
<그림 5> 예제 8.9 그래프

- # 예제 8.10

$P(-t_{0.0025} < T < t_{0.05})$를 구하여라.

- $t_{0.05}$는 t 분포곡선에서 오른쪽 면적이 0.05이고, $-t_{0.025}$는 왼쪽 면적이 0.025 이므로 나머지 면적은 $1 - 0.05 - 0.25 = 0.925$
- 따라서

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



<그림 6> 예제 8.10 그래프

t 분포

- 예제 8.11

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

- 자유도 14인 t 분포에서 1.761은 $t_{0.05}$에 대응됨
- $-t_{0.05} = -1.761$이므로 $k = -t_\alpha$라고 할 때, $0.045 = 0.05 - \alpha$이므로

$$\alpha = 0.005$$

- 따라서, 자유도 14일 때 이를 만족하는 값은

$$k = -t_{0.005} = -2.977$$

이 되고, $P(-2.977 < T < -1.761) = 0.045$ 이 됨

[표 1] 자유도 v와 α값에 따른 t 분포표 ($v = 11 \sim 15$)

v값	α값 (t_α)				
	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

- 예제 8.12

어떤 화공기사는 어느 배치공정의 수율이 원재료의 리터당 500g 이라고 주장하고 있다. 이를 입증하기 위해 매월 **<u>25개의 배치를 추출</u>**하여 시험을 하는데, 결과로 계산된 t값이 $-t_{0.06}$와 $t_{0.05}$사이에 있으면 주장이 타당하기로 가정하자.
<u>25개의 배치의 시험결과 표본평균은 518g 이고 표준편차는 40g</u> 이었다면 어떤 결론을 낼 수 있겠는가? 단, 모집단은 근사적으로 정규분포를 따른다고 가정한다.

- 시험결과를 가지고 t값을 계산하면

$$t = \frac{518 - 500}{40/\sqrt{25}} = 2.25$$

- 자유도가 24일 때 $t_{0.05} = 1.711$이고, 계산된 t값이 $t_{0.05}$보다 크므로 실제의 수율은 500g 이상임

v값	α값 (t_α)				
	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

[표 2] 자유도 v와 α값에 따른 t 분포표 ($v = 21 \sim 25$)

목 차

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

F 분포

- 정의
  - 두 독립적인 카이제곱 변수 U, V를 각각의 자유도 v_1, v_2로 나눈 뒤, 그 비율로 정의되는 확률분포
- 필요성
  - 카이제곱변수의 비율을 통해 두 가지 이상의 표본집단의 분산을 비교하거나 모집단의 분산을 추정할 때 쓰임
    - 두 가지 이상의 표본평균들이 동일한 모평균을 가진 집단에서 추출되었는지 아닌지를 판단함

F 분포

- 수식

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

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

의 확률밀도 함수는

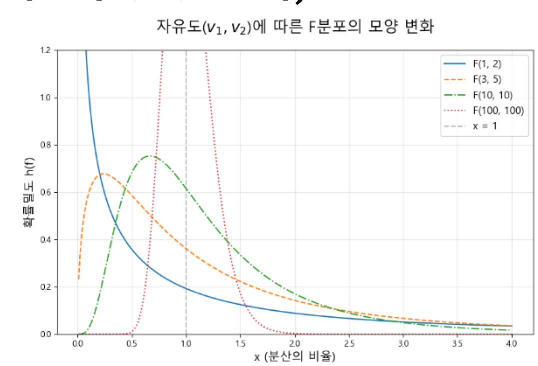
$$h(f) = \begin{cases} \dfrac{\Gamma\left[\dfrac{v_1 + v_2}{2}\right]\left(\dfrac{v_1}{v_2}\right)^{v_1/2}}{\Gamma(v/2)\Gamma(v_2/2)} \times \dfrac{f^{(v_1/2)-1}}{(1 + v_1 f/v_2)^{(v_1+v_2)/2}}, & f > 0 \\ 0, & f \leq 0 \end{cases}$$

가 되고, 이는 자유도 v_1, v_2인 F 분포를 따른다.

F 분포

- 특징 (1/2)
  - F 분포곡선의 형태는 두 자유도 v_1, v_2에 의해 결정됨
  - F 분포에서 자유도는 분자에 있는 카이제곱확률변수의 자유도를 우선으로 사용함
  - 분자에 있는 자유도를 사용한 후 분모에 있는 카이제곱확률변수의 자유도를 사용함
  - 만약 어느 한 쪽의 면적값을 알고 있다면, 분자와 분모의 자유도를 바꾼 뒤 역수를 취해 반대쪽 면적값을 알 수 있음
    - 자유도 v_1, v_2에서 f_a값을 $f_a(v_1, v_2)$로 나타낼 때,

    $f_{(1-\alpha)}(v_1, v_2) = \dfrac{1}{f_\alpha(v_2, v_1)}$ 임



<그림 7> 자유도에 따른 F분포 그래프

F 분포

- 특징 (2/2)
  - 모분산이 각각 σ_1^2, σ_2^2인 모집단에서 크기 n_1, n_2인 표본을 추출했을 때, 표본분산은 각각의 자유도를 따르는 F 분포를 이루게 됨
    - 두 집단의 실제 분산이 같은가를 판별할 때 사용함
    - χ_1^2과 χ_2^2은 각각 자유도가 $v_1 = n_1 - 1, v_2 = n_2 - 1$임
      - $\chi_1^2 = \frac{(n_1-1)S_1^2}{\sigma_1^2}, \chi_2^2 = \frac{(n_2-1)S_2^2}{\sigma_2^2}$
      - $\chi_1^2 = U, \chi_2^2 = V$로 치환해서 정리하면 다음과 같음
        - $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$

F 분포

• 예제 8.13

모집단 A에서 15개를, 모집단 B에서 10개를 뽑아 두 모집단의 표본분산비율이 α보다 적을 확률이 0.05 일 때, α값을 구하여라. 각 모집단의 조건은 다음과 같다.

- 모집단 A의 평균은 4, 분산은 16
- 모집단 B의 평균은 12, 분산은 48
- 두 모집단의 F 분포는 $F(14,9)$를 따른다.

- 문제를 식으로 나타내면 $P\left[\frac{S_A^2}{S_B^2} \le \alpha\right] = 0.05$ 인 α를 찾아야 함
- F분포를 사용하기 위해 양변에 $\frac{\sigma_B^2}{\sigma_A^2}$를 곱하면 $P\left[\frac{S_A^2\sigma_B^2}{S_B^2\sigma_A^2} \le \frac{\sigma_B^2}{\sigma_A^2}\alpha\right] = P(F \le 3\alpha) = 0.05$
- 여사건을 사용하면 $P[F > f_{0.95}(14,9)] = 0.95$이므로 $P[F \le f_{0.95}(14,9)] = 0.05$
- F분포의 특징을 이용하면 $3\alpha = f_{0.95}(14,9) = \frac{1}{f_{0.05}(9,14)} = \frac{1}{2.65}$
- $\alpha \cong 0.1258$

[표 3] 자유도 v_2, v_1와 α값에 따른 F 분포표 ($v_2 = 12 \sim 15, v_1 = 8 \sim 12$)

v2 / v1	8	9	10	12
12	2.85	2.80	2.75	2.69
14	2.70	2.65	2.60	2.53
15	2.64	2.59	2.54	2.48

목 차

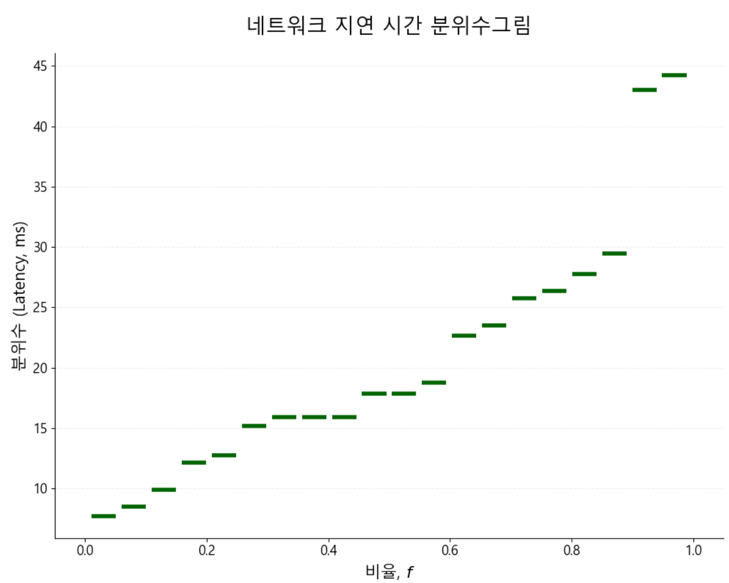
- 확률표본
- 대표적 통계량
- 표본평균의 분포
- 표본분산의 분포
- t 분포
- F 분포
- 분위수와 확률 그림
- 부록

분위수와 확률 그림

- 분위수(Quantile)
  - 정의
    - 분위수를 $q(f)$라고 했을 때, 자료값 중에서 $q(f)$보다 작거나 같은 자료값이 차지하는 비율이 f가 되는 값
      - 확률분포에서 확률변수의 구간을 나누는 기준이 되는 수
      - f는 0과 1사이의 값
        - e.g., $q(0.25)$: 하위 25% 지점의 값, 하위사분위수
          $q(0.5)$: 자료의 중간값, 표본중앙값
  - 필요성
    - 정규분포 크기를 벗어나거나, 산포가 큰 상황에서 분위수가 대푯값으로 사용됨
      - e.g., 연봉 데이터를 조사할 때 상당한 고연봉자가 있을 경우

분위수와 확률 그림

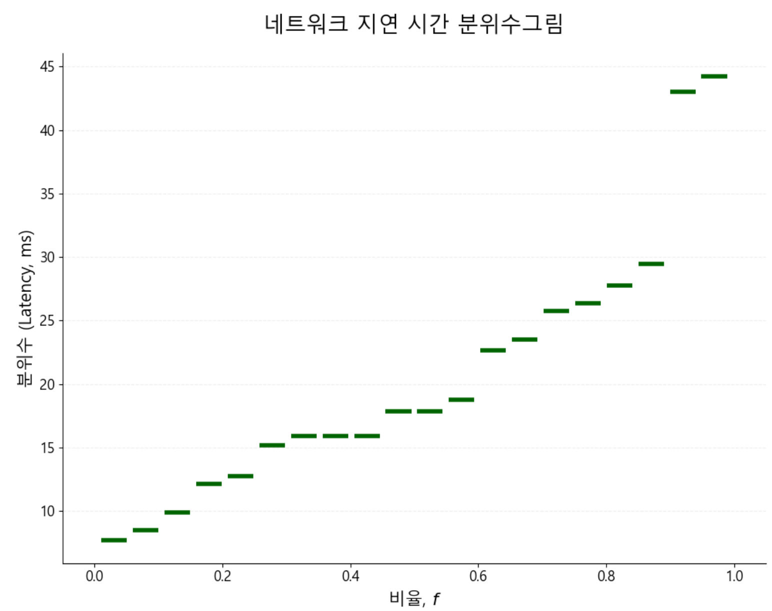
- 분위수그림(Quantile Plot)
  - 정의
    - 전체 데이터의 분포 형태와 밀집도를 나타낸 그림
  - 필요성
    - 표본 데이터가 특정 확률분포를 실제로 따르는지 시각적으로 검증함



<그림 8> 네트워크 지연 시간에 대한 분위수그림

분위수와 확률 그림

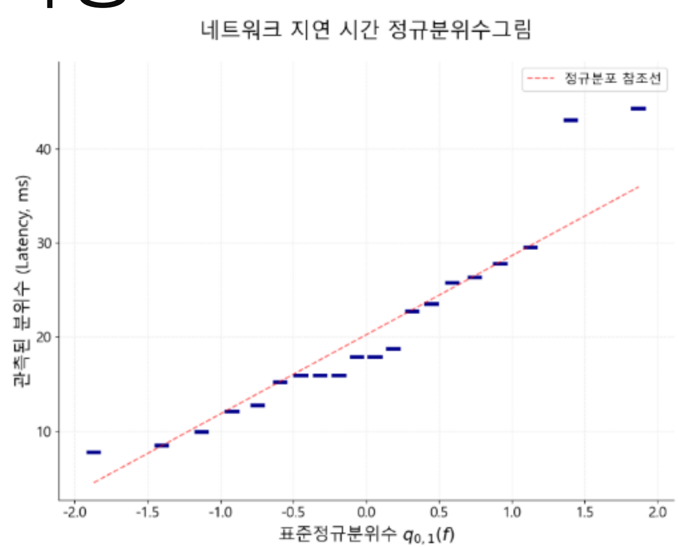
- 분위수그림
  - 특징
    - x축은 f를, y축은 분위수(데이터값)를 표시
    - 데이터를 순서대로 나열한 후, 순위에 맞는 비율 f_i를 계산하여 점으로 표시
      - $f_i = \frac{i-3/8}{n+1/4}$, n =조사한 데이터 수, $i = 1, 2, \ldots, n$



<그림 8> 네트워크 지연 시간에 대한 분위수그림

분위수와 확률 그림

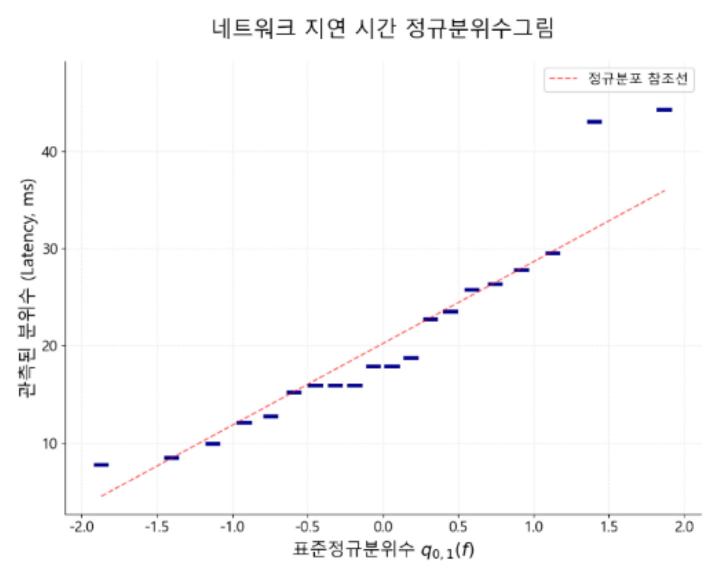
- 정규분위수그림(Normal Quantile-Quantile Plot)
  - 정의
    - 표본 데이터의 분위수와 정규분포를 비교하는 그림
  - 필요성
    - 데이터와 정규분포 참조선을 비교해 데이터가 정규 분포를 따르는지 검증 가능



<그림 9> 네트워크 지연 시간에 대한 정규분위수그림

분위수와 확률 그림

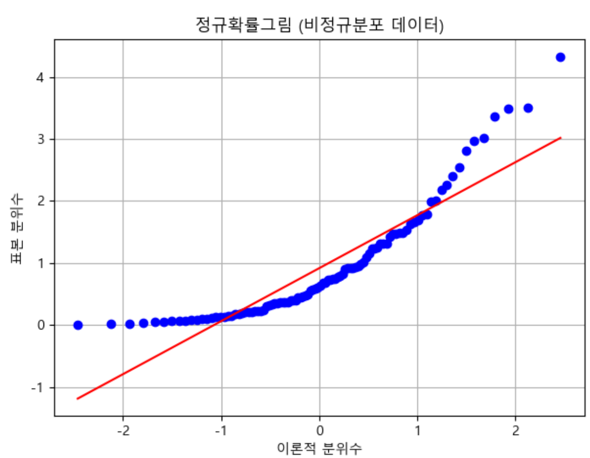
- 정규분위수그림
  - 특징
    - x축은 모집단의 평균을, 기울기는 표준편차를 나타냄
    - 정규분포의 분위수를 정확히 계산하는 식은 복잡하여 근사식으로 계산함
      - 근사식은 $q_{\mu,\sigma}(f) = \mu + \sigma\{4.91[f^{0.14} - (1 - f)^{0.14}]\}$
        - $q_{\mu,\sigma}(f)$은 평균이 μ, 표준편차가 σ인 정규분포에서 누적확률이 f가 되는 지점임



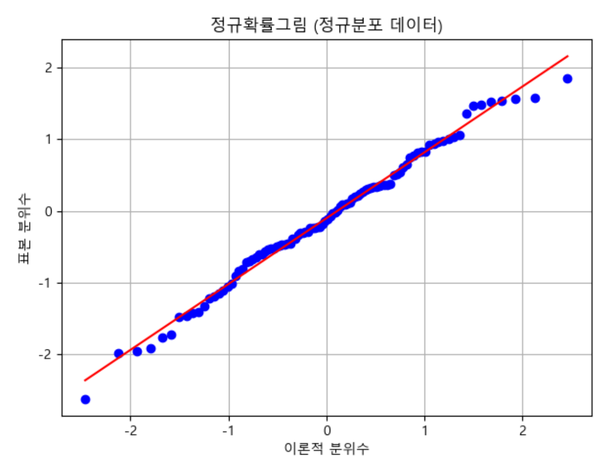
<그림 9> 네트워크 지연 시간에 대한 정규분위수그림

분위수와 확률 그림

- 정규확률그림(Normal Probability Plot)
  - 정의
    - 표본 데이터가 정규분포를 따르는지 시각적으로 검토하기 위해 표본 분위수를 정규분포의 이론적 분위수와 비교하여 그린 그림
  - 필요성
    - 표본 데이터가 정규분포를 따르는지 직관적으로 평가가 가능함



<그림 10> 정규분포를 따르지 않는 정규확률그림



<그림 11> 정규분포를 따르는 정규확률그림

Thanks!

민 윤 홍(yunhong@pel.sejong.ac.kr)

부록#1

• 표준정규분포표

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

- t 분포표 (1/2)

	α						
v	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

부록#2

- t 분포표 (2/2)

	α						
v	**0.02**	**0.015**	**0.01**	**0.0075**	**0.005**	**0.0025**	**0.0005**
1	15.894	21.205	31.821	42.433	63.656	127.321	636.578
2	4.849	5.643	6.965	8.073	9.925	14.089	31.600
3	3.482	3.896	4.541	5.047	5.841	7.453	12.924
4	2.999	3.298	3.747	4.088	4.604	5.598	8.610
5	2.757	3.003	3.365	3.634	4.032	4.773	6.869
6	2.612	2.829	3.143	3.372	3.707	4.317	5.959
7	2.517	2.715	2.998	3.203	3.499	4.029	5.408
8	2.449	2.634	2.896	3.085	3.355	3.833	5.041
9	2.398	2.574	2.821	2.998	3.250	3.690	4.781
10	2.359	2.527	2.764	2.932	3.169	3.581	4.587
11	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	2.197	2.336	2.528	2.661	2.845	3.153	3.850
21	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	2.158	2.291	2.473	2.598	2.771	3.057	3.689
28	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	2.150	2.282	2.462	2.586	2.756	3.038	3.660
30	2.147	2.278	2.457	2.581	2.750	3.030	3.646
40	2.123	2.250	2.423	2.542	2.704	2.971	3.551
60	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	2.054	2.170	2.326	2.432	2.576	2.807	3.290

부록#3

- F 분포표 (1/2)

	$f_{0.05}(v_1, v_2)$								
	v_1								
v_2	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

부록#3

- F 분포표 (2/2)

	$f_{0.05}(v_1, v_2)$									
	v_1									
v_2	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

부록#4 그래프 Python 코드

- <그림 1> 중심극한정리

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
from scipy.stats import expon, norm

# 한글 폰트 설정
mpl.rcParams['font.family'] = 'Malgun Gothic'
mpl.rcParams['axes.unicode_minus'] = False

# 1. 모집단 및 파라미터 설정
mu_true = 1.0  # 평균 위치
x = np.linspace(0, 3.5, 1000)

plt.figure(figsize=(10, 6))

# --- (1) n=1 (모집단) ---
plt.plot(x, expon.pdf(x, scale=mu_true),
         label='n=1 (모집단)', color='gray', linestyle='-', linewidth=2)
# --- (2) n=5 (중간 크기) ---
n_mid = 5
std_mid = mu_true / np.sqrt(n_mid)
plt.plot(x, norm.pdf(x, loc=mu_true, scale=std_mid),
         label=f'n={n_mid} (중간 크기)', color='green', alpha=0.7)
# --- (3) n=30 (거의 정규분포) ---
n_large = 30
std_large = mu_true / np.sqrt(n_large)
plt.plot(x, norm.pdf(x, loc=mu_true, scale=std_large),
         label=f'n={n_large} (거의 정규분포)', color='red', linewidth=2)

plt.axvline(mu_true, color='black', linestyle='--', linewidth=1)
plt.text(mu_true, -0.05, '$\mu$', ha='center', fontsize=15)

# 축 및 레이블 설정
plt.title('중심극한정리의 예시 (n의 값에 따른 변화)', fontsize=15, pad=20)
plt.xlabel('표본평균 ($\mu_{\\bar{x}}$)', fontsize=12)
plt.ylabel('확률밀도', fontsize=12)

plt.xticks([0, 1, 2, 3])
plt.xlim(0, 3.5)
plt.legend()
plt.grid(alpha=0.3)
plt.gca().set_yticklabels([])
plt.tight_layout()
plt.show()
```

부록#4 그래프 Python 코드

- <그림 2> 예제 8.4 그래프

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
from scipy.stats import norm
# 한글 폰트 및 마이너스 깨짐 설정
mpl.rcParams['font.family'] = 'Malgun Gothic'
mpl.rcParams['axes.unicode_minus'] = False
# 1. 파라미터 설정
mu = 800              # 평균
sigma = 40            # 표준편차
n = 16                # 표본 크기
se = sigma / np.sqrt(n)  # 표준오차 = 10
target_x = 775        # 기준값

# x축 범위 설정
x = np.linspace(760, 840, 1000)
y = norm.pdf(x, mu, se)

plt.figure(figsize=(10, 6))
# 2. 정규분포 곡선 및 확률 영역 색칠
plt.plot(x, y, color='dodgerblue', linewidth=2)
x_fill = np.linspace(760, target_x, 100)
plt.fill_between(x_fill, norm.pdf(x_fill, mu, se), color='lightskyblue', alpha=0.5, label=f'$P(\\bar{{X}} < {target_x})$')
# 3. 평균 800 지점에만 수직 점선 유지
plt.axvline(mu, color='black', linestyle='--', linewidth=1.5, label=f'평균 $\mu = {mu}$')

z_val = (target_x - mu) / se
prob = norm.cdf(target_x, mu, se)
plt.annotate(f'$z = {z_val}$\n$Prob \\approx {prob:.4f}$',
             xy=(target_x, 0.005), xytext=(target_x - 12, 0.02),
             arrowprops=dict(facecolor='black', shrink=0.05, width=1, headwidth=8),
             fontsize=12, fontweight='bold', ha='center',
             bbox=dict(boxstyle="round,pad=0.3", fc="white", ec="gray", alpha=0.8))

plt.title(f'표본평균의 분포 ($n={n}$)', fontsize=15, pad=20)
plt.xlabel('표본평균 ($\mu_{\\bar{x}}$)', fontsize=12)
plt.ylabel('확률밀도', fontsize=12)
plt.xticks([760, 775, 780, 790, 800, 810, 820, 840])
plt.grid(alpha=0.2, linestyle='--')
plt.legend()
plt.tight_layout()
plt.show()
```

부록#4 그래프 Python 코드

- <그림 3> 예제 8.5 그래프

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
from scipy.stats import norm

mpl.rcParams['font.family'] = 'Malgun Gothic'
mpl.rcParams['axes.unicode_minus'] = False

# mpl.rcParams['font.size'] = 14

mu, sigma, target_x = 28.0, 1.0, 30.5
x = np.linspace(24, 32, 1000)
y = norm.pdf(x, mu, sigma)

plt.figure(figsize=(10, 6))
plt.plot(x, y, color='darkolivegreen', linewidth=2)
plt.fill_between(np.linspace(target_x, 32, 100),
norm.pdf(np.linspace(target_x, 32, 100), mu, sigma),
color='olivedrab', alpha=0.3)
plt.axvline(mu, color='black', linestyle='--',
linewidth=1)
```

```python
plt.title('예제 8.5의 영역', fontsize=30, pad=30) #
plt.xlabel('$\\bar{x}$', fontsize=24)
plt.xticks([28.0, 30.5], fontsize=21)
plt.text(mu, 0.1, '평균 28.0', fontsize=20,
ha='center')

plt.gca().set_yticks([])
plt.tight_layout()
plt.show()
```

부록#4 그래프 Python 코드

- <그림 4> 자유도에 따른 t 분포 그래프

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
from scipy.stats import t, norm

# 1. 한글 폰트 및 마이너스 깨짐 설정
mpl.rcParams['font.family'] = 'Malgun Gothic'
mpl.rcParams['axes.unicode_minus'] = False

# x축 범위 설정
x = np.linspace(-4, 4, 1000)

fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(16, 7))

ax1.plot(x, norm.pdf(x), 'k-', lw=2, label='$v = \infty$ (정규분포)')
ax1.plot(x, t.pdf(x, df=5), 'b-', lw=1.5, label='$v = 5$')
ax1.plot(x, t.pdf(x, df=2), 'g-', lw=1.5, label='$v = 2$')

# 텍스트 주
ax1.text(0.1, 0.41, '$v = \infty$', fontsize=16)
ax1.text(0.5, 0.36, '$v = 5$', fontsize=16)
ax1.text(0.8, 0.32, '$v = 2$', fontsize=16)
```

```
ax1.set_title('$v=2, 5, \infty$의 $t$ 분포', fontsize=20, pad=20)
ax1.set_xlabel('$t$', fontsize=18)
ax1.set_xticks([-2, -1, 0, 1, 2])
ax1.set_yticks([]) # y축 눈금 제거
ax1.legend(fontsize=14)
ax1.grid(alpha=0.2)
```

부록#4 그래프 Python 코드

- <그림 4-2> *t* 분포의 대칭성

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
from scipy.stats import t, norm

# 1. 한글 폰트 및 마이너스 깨짐 설정
mpl.rcParams['font.family'] = 'Malgun Gothic'
mpl.rcParams['axes.unicode_minus'] = False

# x축 범위 설정
x = np.linspace(-4, 4, 1000)

# 서브플롯 생성 (1행 2열)
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(16, 7))
df_fixed = 10 # 대칭성을 보여주기 위한 임의의 자유도
y_t = t.pdf(x, df_fixed)
ax2.plot(x, y_t, 'k-', lw=2)

# 양쪽 꼬리 임계값
t_alpha = 1.812 # 자유도 10일 때 t_0.05 값 예시
x_left = np.linspace(-4, -t_alpha, 100)
x_right = np.linspace(t_alpha, 4, 100)
```

```python
# 양쪽 꼬리 색칠
ax2.fill_between(x_left, t.pdf(x_left, df_fixed), color='lightgreen', alpha=0.4)
ax2.fill_between(x_right, t.pdf(x_right, df_fixed), color='lightgreen', alpha=0.4)

ax2.axvline(0, color='black', linestyle='--', linewidth=1)
ax2.set_xticks([-t_alpha, 0, t_alpha])
ax2.set_xticklabels(['$t_{1-\\alpha} = -t_{\\alpha}$', '0', '$t_{\\alpha}$'], fontsize=15)

ax2.set_title('$t$ 분포의 대칭성', fontsize=20, pad=20)
ax2.set_xlabel('$t$', fontsize=18)
ax2.set_yticks([])
ax2.grid(alpha=0.2)

# 테두리 정리
for ax in [ax1, ax2]:
    ax.spines['top'].set_visible(False)
    ax.spines['right'].set_visible(False)
    ax.spines['left'].set_visible(False)

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

부록#4 그래프 Python 코드

- <그림 5> 예제 8.9 그래프

```python
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
import matplotlib

matplotlib.rc('font', family='Malgun Gothic')
plt.rcParams['axes.unicode_minus'] = False

df = 14  # 자유도 v = 14
alpha = 0.025  # 왼쪽 꼬리 확률

t_right = stats.t.ppf(1 - alpha, df)  # 약 2.145
t_left = stats.t.ppf(alpha, df)       # 약 -2.145

x = np.linspace(-4, 4, 1000)
y = stats.t.pdf(x, df)

plt.figure(figsize=(10, 6))
plt.plot(x, y, 'b-', lw=2, label=f't-분포 (자유도={df})')
x_fill_left = np.linspace(-4, t_left, 100)
plt.fill_between(x_fill_left, stats.t.pdf(x_fill_left, df), color='red', alpha=0.5,
                 label=f'왼쪽 면적 = {alpha} (t = {t_left:.3f})')
x_fill_right = np.linspace(t_right, 4, 100)
plt.fill_between(x_fill_right, stats.t.pdf(x_fill_right, df), color='gray', alpha=0.2,
                 label=f'오른쪽 면적 = {alpha} (t = {t_right:.3f})')

plt.axvline(t_left, color='red', linestyle='--', lw=1)
plt.axvline(t_right, color='gray', linestyle='--', lw=1)

plt.annotate(f'\n t = {t_left:.3f}', xy=(t_left, 0.02), xytext=(t_left-1.5, 0.1),
             arrowprops=dict(facecolor='black', shrink=0.05), fontsize=10, color='red')

plt.title(f"예제 8.9: t-분포의 대칭성 시각화 (자유도 {df})")
plt.xlabel("t 값")
plt.ylabel("확률 밀도 (PDF)")
plt.legend()
plt.grid(True, alpha=0.3)
plt.show()
```

부록#4 그래프 Python 코드

- <그림 6> 예제 8.10 그래프

```python
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
import matplotlib

matplotlib.rc('font', family='Malgun Gothic')
plt.rcParams['axes.unicode_minus'] = False

plt.rcParams['axes.unicode_minus'] = False

df = 24
t_left_val = stats.t.ppf(0.025, df)  # -t_0.025
t_right_val = stats.t.ppf(1 - 0.05, df)  # t_0.05

x = np.linspace(-4, 4, 1000)
y = stats.t.pdf(x, df)

plt.figure(figsize=(10, 6))
plt.plot(x, y, 'b-', lw=2, label=f't-분포 (자유도={df})')

x_fill = np.linspace(t_left_val, t_right_val, 100)
y_fill = stats.t.pdf(x_fill, df)
plt.fill_between(x_fill, y_fill, color='skyblue', alpha=0.4, label='P(-t_0.025 < T < t_0.05) = 0.925')
x_left = np.linspace(-4, t_left_val, 100)
plt.fill_between(x_left, stats.t.pdf(x_left, df), color='red', alpha=0.3, label='왼쪽 면적 = 0.025')
x_right = np.linspace(t_right_val, 4, 100)
plt.fill_between(x_right, stats.t.pdf(x_right, df), color='orange', alpha=0.3, label='오른쪽 면적 = 0.05')
plt.axvline(t_left_val, color='red', linestyle='--')
plt.axvline(t_right_val, color='orange', linestyle='--')

plt.text(t_left_val - 0.8, 0.05, f'-t_0.025\n({t_left_val:.3f})', color='red', fontweight='bold', ha='center')
plt.text(t_right_val + 0.6, 0.05, f't_0.05\n({t_right_val:.3f})', color='orange', fontweight='bold', ha='center')
plt.text(0, 0.15, '0.925', fontsize=15, ha='center', color='blue', fontweight='bold')

plt.title(f"t-분포 확률 계산 시각화 (자유도 {df})")
plt.xlabel("t 값")
plt.ylabel("확률 밀도")
plt.legend(loc='upper right')
plt.grid(True, alpha=0.3)
plt.show()
```

부록#4 그래프 Python 코드

- <그림 7> 자유도에 따른 F 분포 그래프

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
import scipy.stats as stats

# 1. 한글 폰트 및 마이너스 기호 설정
mpl.rcParams['font.family'] = 'Malgun Gothic'
mpl.rcParams['axes.unicode_minus'] = False

# x축 범위 설정 (F분포는 항상 양수이므로 0부터 시작)
xx = np.linspace(0.01, 4, 1000)

plt.figure(figsize=(10, 6))

# 2. 자유도(v1, v2) 조합에 따른 F분포 그래프
plt.plot(xx, stats.f(1, 2).pdf(xx), ls="-",
label="F(1, 2)")
plt.plot(xx, stats.f(3, 5).pdf(xx), ls="--",
label="F(3, 5)")
plt.plot(xx, stats.f(10, 10).pdf(xx), ls="-.",
label="F(10, 10)")
plt.plot(xx, stats.f(100, 100).pdf(xx), ls=":",
label="F(100, 100)") # 수렴 확인용 추가
# 기준선 및 디자인
plt.axvline(1, color='gray', ls="--", alpha=0.5,
label="x = 1")
plt.ylim(0, 1.2) # 가독성을 위해 y축 범위 제한
plt.xlabel("x (분산의 비율)", fontsize=14)
plt.ylabel("확률밀도 h(f)", fontsize=14)
plt.title("자유도($v_1, v_2$)에 따른 F분포의 모양
변화", fontsize=18, pad=20)
plt.legend(fontsize=12)
plt.grid(alpha=0.3)

plt.show()
```

부록#4 그래프 Python 코드

- <그림 8> 네트워크 지연 시간에 대한 분위수그림

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl

# 1. 한글 폰트 설정
mpl.rcParams['font.family'] = 'Malgun Gothic'
mpl.rcParams['axes.unicode_minus'] = False

# 2. 네트워크 Latency 예시 데이터 (n=20으로 작게 설정하여 개별 데이터 강조)
np.random.seed(42)
n = 20
latency_data = np.random.lognormal(mean=3, sigma=0.5, size=n)
sorted_latency = np.sort(latency_data)

# 3. 비율 f_i 계산 (교재 수식 준수)
i = np.arange(1, n + 1)
f_i = (i - 3/8) / (n + 1/4)

plt.figure(figsize=(10, 8))
# 4. 짧은 가로선 사용
for x, y in zip(f_i, sorted_latency):
    plt.hlines(y, x - 0.02, x + 0.02, colors='darkgreen', lw=4) # 짧은 가로 막대

# 5. 그래프 디자인
plt.title('네트워크 지연 시간 분위수그림', fontsize=20, pad=20)
plt.xlabel('비율, $f$', fontsize=16)
plt.ylabel('분위수 (Latency, ms)', fontsize=16)

plt.xticks(np.arange(0, 1.1, 0.2), fontsize=13)
plt.yticks(fontsize=13)
plt.xlim(-0.05, 1.05)
plt.grid(axis='y', alpha=0.2, linestyle='--')

plt.gca().spines['top'].set_visible(False)
plt.gca().spines['right'].set_visible(False)

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

부록#4 그래프 Python 코드

- <그림 9> 네트워크 지연 시간에 대한 정규분위수 그림

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
from scipy.stats import norm

mpl.rcParams['font.family'] = 'Malgun Gothic'
mpl.rcParams['axes.unicode_minus'] = False

np.random.seed(42)
n = 20
latency_data = np.random.lognormal(mean=3, sigma=0.5, size=n)
sorted_latency = np.sort(latency_data)

i = np.arange(1, n + 1)
f_i = (i - 3/8) / (n + 1/4)
theoretical_quantiles = norm.ppf(f_i)

plt.figure(figsize=(10, 8))
# 4. 정규분위수그림
for x, y in zip(theoretical_quantiles, sorted_latency):
    plt.hlines(y, x - 0.05, x + 0.05, colors='darkblue', lw=4)
z_q1, z_q3 = norm.ppf([0.25, 0.75])
y_q1, y_q3 = np.percentile(sorted_latency, [25, 75
slope = (y_q3 - y_q1) / (z_q3 - z_q1)
intercept = y_q1 - slope * z_q1
plt.plot(theoretical_quantiles, slope * theoretical_quantiles + intercept,
         color='red', linestyle='--', alpha=0.6, label='정규분포 참조선')


# 5. 그래프 디자인
plt.title('네트워크 지연 시간 정규분위수그림', fontsize=22, pad=25)
plt.xlabel('표준정규분위수 $q_{0,1}(f)$', fontsize=18)
plt.ylabel('관측된 분위수 (Latency, ms)', fontsize=18)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.grid(alpha=0.2, linestyle='--')
plt.legend(fontsize=14)
plt.gca().spines['top'].set_visible(False)
plt.gca().spines['right'].set_visible(False)
plt.ylim(min(sorted_latency) - 5, max(sorted_latency) + 5)  # Y축 여백 조절
plt.tight_layout()
plt.show()
```

부록#4 그래프 Python 코드

- <그림 10> 정규분포를 따르지 않는 정규확률그림

```python
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
import matplotlib

matplotlib.rc('font', family='Malgun Gothic')
plt.rcParams['axes.unicode_minus'] = False

np.random.seed(42)
data = np.random.normal(0, 1, 100)

plt.figure(figsize=(7, 5))
stats.probplot(data, dist="norm", plot=plt)
plt.title("정규확률그림 (정규분포 데이터)")
plt.xlabel("이론적 분위수")
plt.ylabel("표본 분위수")

plt.grid(True)
plt.show()
```

부록#4 그래프 Python 코드

- <그림 11> 정규분포를 따르는 정규확률그림

```python
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
import matplotlib

matplotlib.rc('font', family='Malgun Gothic')
plt.rcParams['axes.unicode_minus'] = False

np.random.seed(42)
data = np.random.exponential(scale=1, size=100)

plt.figure(figsize=(7, 5))
stats.probplot(data, dist="norm", plot=plt)

plt.title("정규확률그림 (비정규분포 데이터)")
plt.xlabel("이론적 분위수")
plt.ylabel("표본 분위수")

plt.grid(True)
plt.show()
```