문제 **1.** 대한민국 **30**대 월소득이 평균 **200**만원**,** 표준편차 **50**만원인 정규분포를 따른다고 가정하자**.** 임의로 **50**명의 **30**대 직장인을 골라 월소득 평균을 구했을 때**,** 평균값이 **250**만원보다 높을 확률은 얼마인가**?**

모집단

$$\mu = 200\text{만원}, \quad \sigma = 50\text{만원}$$

표본 분포

$$n = 50$$

$$\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) = N(200, \frac{50}{\sqrt{50}}) = N(200, 7.071067812)$$

관심 확률

$$P(\overline{X} > 250) = P(Z > \frac{250-200}{7.071067812}) = P(Z > 7.07)$$

$$P(Z > 7.07) = 1 - \Phi(7.07) \approx 7.9 \times 10^{-13}$$

50명 표본 평균이 250만원 이상이 될 확률은 사실 상 0에 가깝다.

문제 **2.** 대한민국 **30**대 월소득이 정규분포를 따른다고 가정하자**.** 임의로 **5**명의 **30**대 직장인을 골라 월소득액을 조사한 결과가 아래와 같다**. 30**대 모집단의 월소득 평균이 **230**만원보다 클 확률은 얼마인가**?**

[180, 230, 220, 270, 200]

표본 갯수: 5개
표본 평균: (180 + 230 + 220 + 270 + 200)/5 = 220
표본 분산: 1150
표본표준편차: $\sqrt{1150} \approx 33.91$
표준오차: $\frac{33.91}{\sqrt{5}} \approx 15.17$

중심극한 정리 근사를 하면,

$\mu \approx N(220, 15.17)$

$P(\mu > 230) = P(Z > \frac{230-\overline{x}}{SE}) = P(Z > 0.659) = 1 - \Phi(0.659) \approx 0.2548$

약 25%

문제 **3.1** 아래자료는 특정 공장의 **50**일간 일별 냉장고 생산량을 보여준다**.** 공장의 실제 생산량 평균이 **105**보다 클 확률은 얼마인가**?**

87 106 87 127 95 114 109 94 111 110 95 87 77 91 119 102 86 110 110 94 140 92 107 101 103 104 111 94 93 94 109 98 102 120 108 93 102 93 77 97 101 82 98 101 98 90 101 88 81 114

표본평균 = 100.06

표본표준편차 = 12.43

표준오차= $\frac{s}{\sqrt{50}} = 1.758$

$$z = \frac{105-100.06}{1.758} = 2.81$$

$$P(Z > 2.81) = 0.0025$$

0.25%

문제 **3.2** 냉장고 생산공정에 있어서 생산량을 증대시켜줄 것으로 기대하는 신기술을 도입하고자 한다**. 20**일 동안 신기술을 적용하여 테스트 해 본 결과 아래와 같은 일별 생산량 자료를 얻었다**.** 신기술의 평균 생산량이 기존 기술의 평균 생산량보다 더 높을 확률은 얼마인가**?**

116 122 131 135 139 126 109 113 132 144 103 121 128 101 121 122 118 112 117

신기술 표본평균 = 121.58

신기술 표본표준편차 = 11.55

신기술 평균 - 기존 평균 = 21.52대

기존 시스템 분산 비중 $\frac{(12.43)^2}{50} = 3.09$

신기술 시스템 분산 비중 $\frac{(11.55)^2}{19} = 7.02$

두 비중의 합은 10.11

평균 차의 표준오차 $\sqrt{10.11} = 3.18$

따라서 두 표본의 평균 차는 21.52인데, 두 표본의 표준오차는 3.18이므로,
약 6.8배 차이가 난다.

따라서 신기술 평균이 기존 평균보다 클 확률은 거의 100%에 가깝다.

문제 **3.3** 신기술의 평균 생산량이 기존 기술의 평균 생산량보다 **20**이상 더 높을 확률은 얼마인가**?**

3-2에서 두 표본의 평균 차는 21.52이다.
평균 차의 표준 오차는 3.18이다.

20 - 21.52 = -1.52

$z = \frac{-1.52}{3.18} \approx -0.48$

$P(Z > -0.48)$ 이면 표준 정규표로 약 0.68

따라서 신기술 평균이 기존 평균보다 20대 이상 더 클 확률은 약 68%이다.

[코드로 확인]

```
import math
import numpy as np

# ---------- 데이터 ----------
old = np.array([
   87,106,87,127,95,114,109,94,111,110,
   95,87,77,91,119,102,86,110,110,94,
   140,92,107,101,103,104,111,94,93,94,
   109,98,102,120,108,93,102,93,77,97,
   101,82,98,101,98,90,101,88,81,114
])

new = np.array([
   116,122,131,135,139,126,109,113,132,144,
   103,121,128,101,121,122,118,112,117
])  # 19개

# ---------- 헬퍼: 표준정규 상부꼬리 ----------
def std_normal_sf(z):
   """Z~N(0,1) 상부꼬리(SF) = 1-CDF"""
   return 0.5 * math.erfc(z / math.sqrt(2))

# ---------- 문제 3-1 ----------
n_old   = len(old)
mean_o  = old.mean()
```

```python
s_old    = old.std(ddof=1)
se_old   = s_old / math.sqrt(n_old)

z31 = (105 - mean_o) / se_old       # mu0 - x̄
p31 = std_normal_sf(z31)

# ---------- 문제 3-2 & 3-3 ----------
n_new    = len(new)
mean_n   = new.mean()
s_new    = new.std(ddof=1)

diff     = mean_n - mean_o
se_diff = math.sqrt(s_old**2 / n_old + s_new**2 / n_new)

# 3-2: Δ > 0
z32 = (0 - diff) / se_diff
p32 = std_normal_sf(z32)

# 3-3: Δ > 20
z33 = (20 - diff) / se_diff
p33 = std_normal_sf(z33)

# ---------- 결과 출력 ----------
print("Problem 3-1  P(μ_old > 105)        ≈ {:.4f}".format(p31))
print("Problem 3-2  P(μ_new - μ_old > 0)  ≈ {:.4f}".format(p32))
print("Problem 3-3  P(μ_new - μ_old > 20) ≈ {:.4f}".format(p33))
```

```
(6.86x) (base) kimdawoon@gimdaun-ui-MacBook-Pro ~/DataScience_Statistic_R_Data/HW/W07 main python problem3.py
Problem 3-1  P(μ_old > 105)        ≈ 0.0025
Problem 3-2  P(μ_new - μ_old > 0)  ≈ 1.0000
Problem 3-3  P(μ_new - μ_old > 20) ≈ 0.6835
```