문제 **1.** 자동기계로 커피병에 **1**파운드의 커피를 채워 넣는 공정에서**,** 실제의 커피 무게가 평균 **1.03kg,** 표준편차 **0.02kg**의 정규분포를 이룬다고 하자**.** 이 때**,** 어느 커피병의 실제 커피 무게가

1) **1kg** 이하일 확률

평균이 1.03이고, 표준편차가 0.02이다.

그리고 1kg은 평균에서 1.5 표준편차만큼 떨어져 있다.

따라서

$P(X \leq 1) \simeq 0.0668$ 이므로
6.68%라고 볼 수 있다.

2) **1.06kg** 이상일 확률

마찬가지로 평균에서 0.03만큼 떨어져 있으므로 1.5 표준편차만큼 떨어져 있다.

1번과 마찬가지로 $P(X >= 1.06) \simeq 0.0668$ 이다.

문제 **2.** 어느 회사 제품의 건전지 수명이 정규분포를 이룬다고 가정하자**.** 만일 **33%**의 건전지 수명이 **45**시간 이하이고 **10%**의 건전지ㅣ 수명이 **65**시간 이상이라고 할 때**,** 이 분포의 평균과 표준편차는 얼마인가**?**

$P(X \leq 45) = 0.33$

$P(X \geq 65) = 0.10$
$P(X \leq 65) = 0.90$

표준 정규 분포표 상에서

$P(Z \leq -0.44) = 0.33$
$P(Z \leq 1.28) = 0.90$

$Z = \frac{X-\mu}{\sigma}$
$X = \mu + Z \cdot \sigma$

따라서

$45 = \mu - 0.44\sigma$
$65 = \mu + 1.28\sigma$

두 식을 정리하면,

$\mu = 50.12$

$$\sigma = 11.63$$

문제 **3-1.** 아래 자료는 특정 공장의 **50**일간 일별 냉장고 생산량을 보여준다**.** 생산량은 정규분포를 따른다고 할 수 있나**?** 평균과 분산은 무엇인가**?**

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

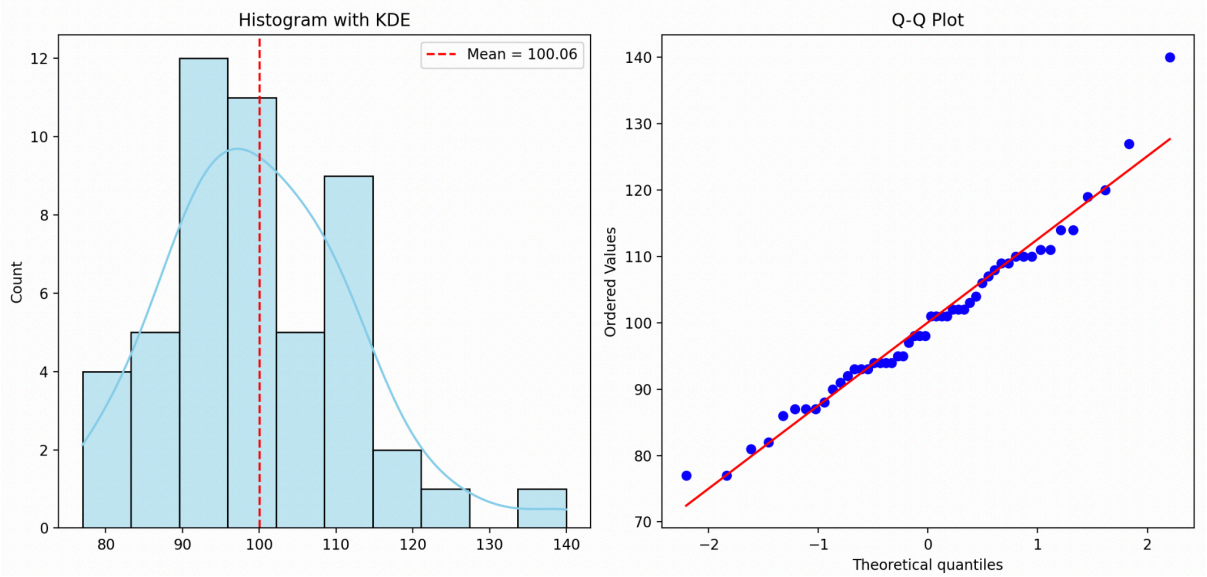
이 문제를 풀기 위해 아래와 같이 파이썬 코드를 작성하였다.

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

# 데이터 입력
data = [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]

# 평균과 분산 계산
mean = np.mean(data)
variance = np.var(data, ddof=1)
std_dev = np.std(data, ddof=1)

# 정규성 검정 (Shapiro-Wilk Test)
shapiro_stat, shapiro_p = stats.shapiro(data)

# 시각화와 함께 텍스트 출력 포함한 전체 코드

# 평균, 분산, 표준편차, p-value 출력
print(f"📊 평균 (Mean): {mean:.2f}")
print(f"📈 분산 (Variance): {variance:.2f}")
print(f"📉 표준편차 (Standard Deviation): {std_dev:.2f}")
print(f"🧪 Shapiro-Wilk 정규성 검정 p-value: {shapiro_p:.4f}")

# 히스토그램과 KDE
plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
sns.histplot(data, kde=True, bins=10, color='skyblue')
plt.axvline(mean, color='red', linestyle='--', label=f'Mean = {mean:.2f}')
plt.title('Histogram with KDE')
plt.legend()
```

```
# Q-Q Plot
plt.subplot(1, 2, 2)
stats.probplot(data, dist="norm", plot=plt)
plt.title("Q-Q Plot")

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

해당 결과는

```
py
(6.86x) (base) kimdawoon@gimdaun-ui-MacBook-Pro ~/Dat
py
📊 평균 (Mean): 100.06
📈 분산 (Variance): 154.55
📉 표준편차 (Standard Deviation): 12.43
🧪 Shapiro-Wilk 정규성 검정 p-value: 0.2203
```



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**.** 신기술에 의해 생산량이 증가했다고 보는 것이 적절한가**?** 이유를 설명하기 위한 적절한 지표를 제시할 수 있는가**?**

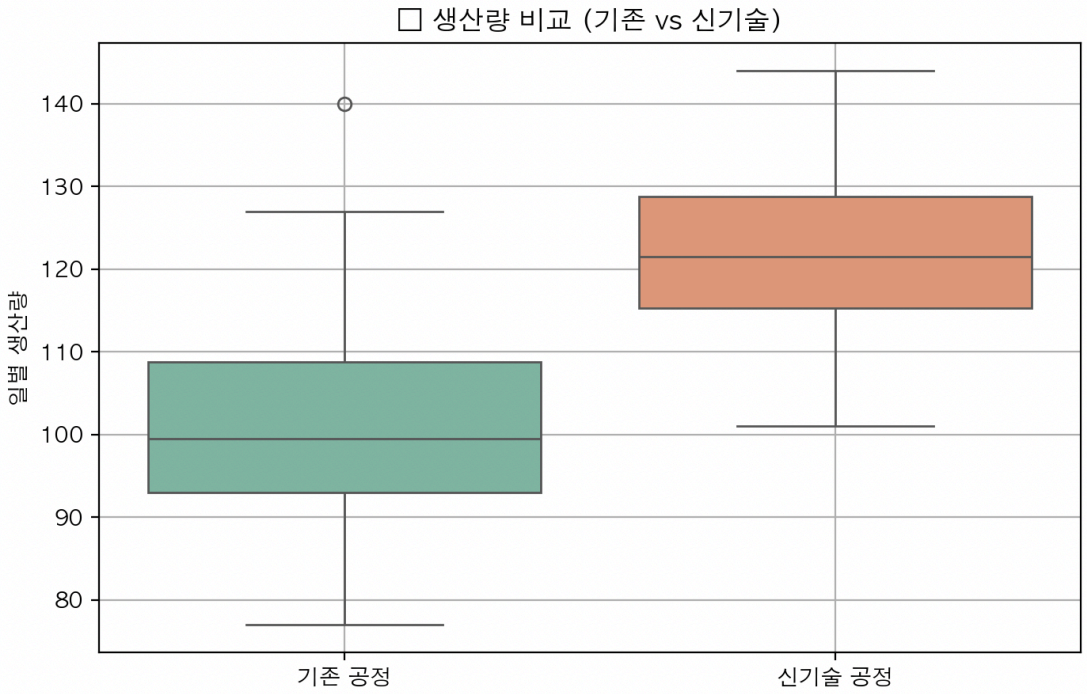
116 122 131 135 139 126 109 113 132 144 103 121 128 128 101 121 122 118 112 117

이 문제를 풀기 위해 아래와 같이 파이썬 코드를 작성하였다.

```
# 재실행: 필요한 패키지 재임포트 및 데이터 재정의
import numpy as np
```

```python
from scipy.stats import ttest_ind
import matplotlib.pyplot as plt
import seaborn as sns

plt.rcParams['font.family'] = 'AppleGothic'
plt.rcParams['axes.unicode_minus'] = False


# 기존 50일 생산량 데이터
old_data = [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]

# 신기술 적용 20일 생산량 데이터
new_data = [116, 122, 131, 135, 139, 126, 109, 113, 132, 144,
            103, 121, 128, 128, 101, 121, 122, 118, 112, 117]

# 평균 계산
old_mean = np.mean(old_data)
new_mean = np.mean(new_data)


# 결과 출력
print("📊 이전 평균 (Old Mean):", round(old_mean, 2))
print("🆕 신기술 평균 (New Mean):", round(new_mean, 2))

# 시각화: 두 그룹 평균 비교
plt.figure(figsize=(8, 5))
sns.boxplot(data=[old_data, new_data], palette="Set2")
plt.xticks([0, 1], ['기존 공정', '신기술 공정'])
plt.title("📦 생산량 비교 (기존 vs 신기술)")
plt.ylabel("일별 생산량")
plt.grid(True)
plt.show()
```

```
    plt.show()
○ (6.86x) (base) kimdawoon@gimdaun-ui-MacBook-Pro ~/DataScience_Stati
py
📊 이전 평균 (Old Mean): 100.06
🆕 신기술 평균 (New Mean): 121.9
problem3_2.py:38: UserWarning: Glyph 128230 (\N{PACKAGE}) missing from
```



기존 공정과 신 공정 후 생산량의 Boxplot을 그려 비교해보았다.
Box plot을 그려보면, 신 공정 후, 생산량이 증가됨을 시각적으로 확인할 수 있다.

문제 **4.** 콜센터에 걸려오는 문의 전화 수는 포아송 분포를 따른다고 가정하고**, 2**분에 평균 **3**통의 전화가 들어온다**.** 최소 **3**분 동안 전화가 한 통도 오지 않을 확률을 포아송 분포와 지수분포를 이용하여 각각 계산하시오**.**

$\lambda = 3/2$ (1분에 걸려올 전화 수)

[포아송 분포 이용]

$$P(N = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

$t = 3, k = 0$

$$P(N = 0) = \frac{(1.5 \cdot 3)^0 e^{-1.5 \cdot 3}}{k!} = e^{-4.5}$$

[지수 분포 이용]

지수분포의 누적분포함수는

$P(T > t) = e^{-\lambda t}$이므로
여기에서도

$t = 3, \lambda = 3/2$ 이므로

$P(T > 3) = e^{-4.5}$

이다.