Question 1

A manufacturer claims that the average lifetime of its batteries is 500 hours, with a known standard deviation of 100 hours. A quality control team tests 30 randomly selected batteries, obtaining the following lifetimes:

495, 520, 510, 505, 480, 500, 515, 495, 510, 505,
490, 515, 495, 505, 500, 510, 485, 495, 500, 520,
510, 495, 505, 500, 515, 505, 495, 510, 500, 495

Perform a hypothesis test at a 5% significance level to determine whether the mean battery lifetime differs significantly from 500 hours. Also, compute the p-value and plot the Operating Characteristic (OC) curve for a range of true means.

Introduction

In quality control and product validation, statistical hypothesis testing is a crucial tool to verify manufacturer claims. In this problem, we assess the validity of a manufacturer's claim regarding the average lifetime of its batteries. Given a known population standard deviation and a random sample of battery lifetimes, we perform a hypothesis test to determine whether the observed data provides sufficient evidence to conclude that the true mean differs from the advertised value. Additionally, we analyze the test's power characteristics by plotting the Operating Characteristic (OC) curve, which illustrates the probability of accepting the null hypothesis for a range of true mean values. Since the population standard deviation is known, we perform a one-sample *z-test* for the mean. The process involves formulating the null hypothesis (H_0), which states that the true mean is 500 hours, and the alternative hypothesis (H_1), which posits that the true mean is different from 500 hours. We use the sample mean, sample size, and standard deviation to compute the *z-statistic*, which is then used to calculate the *p-value*. The p-value helps determine whether the observed data provides enough evidence to reject the null hypothesis at a given significance level (in this case, $\alpha = 0.05$).

Furthermore, we plot the *Operating Characteristic (OC) Curve*, which shows the probability of accepting the null hypothesis for various true mean values. This curve helps assess the test's sensitivity to deviations from the claimed mean and is a valuable tool in evaluating the performance of the hypothesis test under different conditions.

Data

Given with question .

Methodology

As both mean and the standard devaition are given so calculate the mean of the given data and then check whether it lies between the range we get by subtracting mean and dividing significance level from the sample and the calculating probablity according to significance level given.
The code used was as follows :-

```
import math
import numpy as np
from scipy.stats import norm
l=[495, 520, 510, 505, 480, 500, 515, 495, 510, 505, 490, 515, 495, 505, 500, 510, 485, 495, 500, 520, 510,
495, 505, 500, 515, 505, 495, 510, 500, 495]
l=np.array(l)
mu1=500
std=100
k=len(l)
z=((l.mean()-mu1)*(math.sqrt(k)))/std
print(f"the range of 5%error significance is between {-196/math.sqrt(k) + mu1} and {196/math.sqrt(k) + mu1}")
print(f"the mean of the dataset is{l.mean()}")
print(f"as mean lies between the give range hence hypothesis is accepted")
print(f"the p value is as follows {2-2*(norm.cdf(z))}")
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

mu_0 = 500
sigma = 100
n = 30
alpha = 0.05

se = sigma / np.sqrt(n)
z_crit = norm.ppf(1 - alpha / 2)
acceptance_lower = mu_0 - z_crit * se
acceptance_upper = mu_0 + z_crit * se

mu_values = np.linspace(470, 530, 200)
beta_values = norm.cdf(acceptance_upper, loc=mu_values, scale=se) - \
              norm.cdf(acceptance_lower, loc=mu_values, scale=se)

plt.figure(figsize=(10, 6))
plt.plot(mu_values, beta_values, label='OC Curve', color='blue')
plt.axvline(mu_0, color='gray', linestyle='--', label='
```

```
    Claimed Mean (      = 500)')
plt.title("Operating Characteristic (OC) Curve")
plt.xlabel("True Mean (  )")
plt.ylabel("P(Fail to Reject  H  )")
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()
```

Code Explanation

The provided Python code performs a hypothesis test for the population mean using a known standard deviation (i.e., a one-sample *z-test*). Here's a step-by-step explanation of the code:

- First, the sample data of battery lifetimes is stored in a list and converted into a NumPy array for efficient computation.
- The claimed population mean $\mu_0 = 500$ and known standard deviation $\sigma = 100$ are specified. The sample size n is also determined.
- The test statistic z is calculated using the formula:

  $$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

  where $\bar{x}$ is the sample mean. This z-value helps quantify how far the sample mean is from the claimed population mean in terms of standard errors.
- A 5% significance level is used to determine the acceptance region. The code prints the corresponding range:

  $$\mu_0 \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

  where $z_{\alpha/2} \approx 1.96$ for a two-tailed test at $\alpha = 0.05$.
- The sample mean is compared with the acceptance region. If it lies within the interval, the null hypothesis is not rejected.
- Finally, the *p-value* is calculated using the cumulative distribution function (CDF) of the standard normal distribution. The p-value represents the probability of obtaining a result as extreme or more extreme than the observed sample mean, assuming the null hypothesis is true:

  $$\text{p-value} = 2 \cdot (1 - \Phi(|z|))$$

  where Φ is the standard normal CDF.

OC Curve Code Explanation

The following Python code is used to generate the Operating Characteristic (OC) Curve, which illustrates the probability of failing to reject the null hypothesis for various values of the true population mean μ.

- The claimed mean $\mu_0 = 500$, known standard deviation $\sigma = 100$, and sample size $n = 30$ are defined. The significance level is set to $\alpha = 0.05$.
- The standard error of the mean is calculated using:

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}}$$

- The critical z-value corresponding to the two-tailed test is computed using:

$$z_{\alpha/2} = \texttt{norm.ppf}(1 - \alpha/2)$$

- The acceptance region for the null hypothesis is determined:

$$[\mu_0 - z_{\alpha/2} \cdot \mathrm{SE},\ \mu_0 + z_{\alpha/2} \cdot \mathrm{SE}]$$

- A range of potential true means (μ values from 470 to 530) is considered. For each true mean, the probability of failing to reject H_0 (i.e., the power complement β) is calculated using:

$$\beta(\mu) = P(\text{accept } H_0 \mid \mu) = \Phi\left(\frac{\text{upper} - \mu}{\mathrm{SE}}\right) - \Phi\left(\frac{\text{lower} - \mu}{\mathrm{SE}}\right)$$

  where Φ is the cumulative distribution function (CDF) of the normal distribution.

- The resulting OC curve is plotted, showing how the probability of Type II error (β) changes as the true mean varies. A vertical line is added at $\mu_0 = 500$ to indicate the claimed mean.
- The curve visually demonstrates that the probability of accepting H_0 is highest when the true mean equals the claimed mean, and it decreases as the true mean deviates from μ_0.

Result

The results obtained were as follows :
1.the range of 5 percent error significance is between 464.2154595763291 and 535.7845404236708
2.The mean is 502.6666666666667
3.as mean lies between the give range hence hypothesis is accepted

4.the p value is as follows 0.8838745372567058

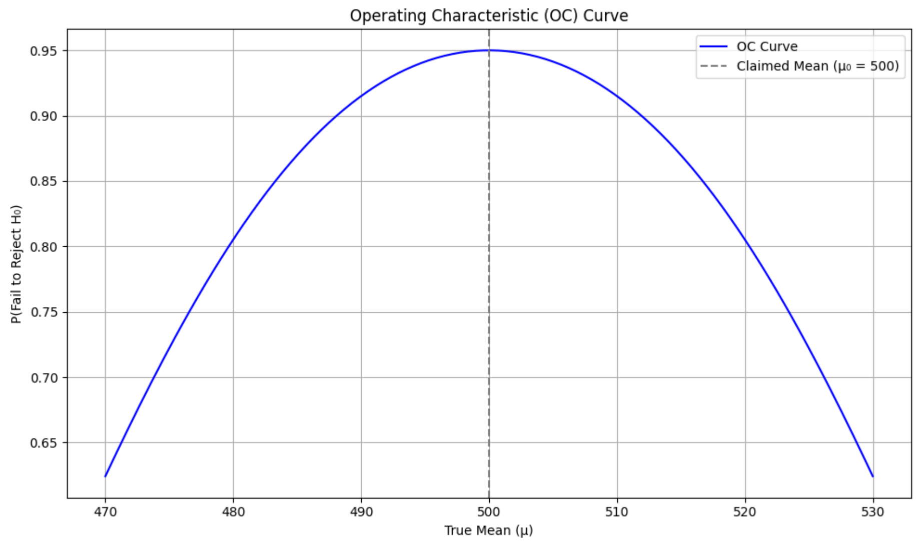

Figure 1: Operating Charecterisitc (OC) curve

Question 2

A public health official claims that the mean home water use is 350 gallons per day. To verify this claim, a study was conducted on 20 randomly selected homes, and their daily water usage (in gallons) was recorded as follows:

340, 344, 362, 375, 356, 386, 354, 364, 332, 402,
340, 355, 362, 322, 372, 324, 318, 360, 338, 370

Consider the following two cases:

[(a)]**Known Population Variance**

Assume that the population variance is known to be $\sigma^2 = 144$. Perform a hypothesis test at a 5% significance level to assess whether the sample data provides sufficient evidence to contradict the official's claim that the mean home water usage is 350 gallons per day.

Instructions:

1. • Clearly state the null and alternative hypotheses:

H_0: $\mu = 350$ (The mean water usage is 350 gallons per day)

H_1: $\mu \neq 350$ (The mean water usage is different from 350 gallons per day)

- Calculate the test statistic using the z-test formula for a known population variance:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

  where $\bar{x}$ is the sample mean, $\mu_0 = 350$ is the claimed mean, $\sigma = \sqrt{144} = 12$, and $n = 20$.
- Compute the *p-value* associated with the calculated z-value.
- Compare the p-value to the significance level $\alpha = 0.05$ and state your conclusion:
  - If p-value $< \alpha$, reject H_0.
  - If p-value $\geq \alpha$, fail to reject H_0.

Unknown Population Variance:
Now, suppose that the population variance is unknown. Perform a one-sample *t-test* at a 5% significance level using the given sample to determine whether the data provides sufficient evidence to contradict the official's claim.

- State the null and alternative hypotheses.
- Calculate the sample mean and sample standard deviation.
- Compute the test statistic using the *t-distribution*.
- Determine the p-value.
- State your conclusion.

The objective is to determine whether the sample data provides enough statistical evidence to reject the official's claim about the population mean water usage. We approach this by performing hypothesis testing under two scenarios:

(a) When the population variance is known.

(b) When the population variance is unknown.

We use a significance level of $\alpha = 0.05$ for both tests. Each case involves formulating hypotheses, computing the test statistic, determining the p-value, and making a conclusion based on the results.

Data

Given in the question

Methodology

The code used is as follows:-

```
import numpy as np
ph=[340, 344, 362, 375, 356, 386, 354, 364, 332, 402, 340, 355, 362, 322, 372, 324, 318, 360, 338, 370
]
ph=np.array(ph)
mu=350
std=12
k=len(ph)

e=(1.96)*(std)/math.sqrt(k)
print(f"the range of 5%error significance is between {-e + mu} and {e+ mu}")
if ph.mean()>=-e+mu and ph.mean()<=e+mu:
  print(f"as {ph.mean()} lies between the range hence the claim is excepted")
else:
  print(f"as {ph.mean()} lies outside the range hence the claim is rejected")
from scipy.stats import t
ph=[340, 344, 362, 375, 356, 386, 354, 364, 332, 402, 340, 355, 362, 322, 372, 324, 318, 360, 338, 370
]
ph=np.array(ph)
mu=350
k=len(ph)
sample_std = np.std(ph, ddof=1)
t_stat = (ph.mean() - mu) / (sample_std / np.sqrt(k))
p_value = 2 * (1 - t.cdf(abs(t_stat), df=n-1))
print("Sample Mean:", round(ph.mean(), 4))
print("Sample Standard Deviation:", round(sample_std, 4))
print("t-Statistic:", round(t_stat, 4))
print("p-value:", round(p_value, 4))
if p_value < 0.05:
    print("Conclusion: Reject the null hypothesis (significant evidence against the claim).")
else:
    print(f"Conclusion: Fail to reject the null hypothesis (not enough evidence to contradict the claim) therefore 350 might be the mean ")
```

Hypothesis Testing and Confidence Interval Analysis

The given code performs two tasks: 1. It computes the confidence interval for a population mean given the sample data and 2. It performs a hypothesis test to evaluate the claim that the population mean is 350.

Step 1: Confidence Interval for the Population Mean We are given a sample of 20 data points (stored in the array *ph*) and we want to check whether the sample mean lies within a specified range of values based on a 5% error margin.

1. The sample data is stored in an array *ph*, and the population mean (μ) is given as 350 with a standard deviation (σ) of 12. 2. The number of data points is denoted by $k = 20$.

The error margin e is calculated as:

$$e = 1.96 \times \frac{\sigma}{\sqrt{k}}$$

Where 1.96 is the critical value for a 95% confidence level. The error margin reflects the range within which the true mean is expected to lie with 95% confidence.

The confidence interval is then given by:

$$(\mu - e, \mu + e)$$

We check if the sample mean lies within this range:

$$\text{If } \mu_{\text{sample}} \in (\mu - e, \mu + e) \quad \text{then the claim is accepted, otherwise it is rejected.}$$

Step 2: Hypothesis Testing Next, the code performs a two-tailed hypothesis test to evaluate the null hypothesis that the population mean is 350 against the alternative hypothesis.

The null hypothesis is:

$$H_0 : \mu = 350$$

The alternative hypothesis is:

$$H_1 : \mu \neq 350$$

To conduct the hypothesis test, the following steps are performed:

1. The sample mean and sample standard deviation are calculated.

$$\mu_{\text{sample}} = \frac{1}{k} \sum_{i=1}^{k} ph_i$$

$$s = \sqrt{\frac{1}{k-1} \sum_{i=1}^{k} (ph_i - \mu_{\text{sample}})^2}$$

2. The test statistic, a t-statistic, is computed:

$$t = \frac{\mu_{\text{sample}} - \mu}{\frac{s}{\sqrt{k}}}$$

Where $\mu = 350$ is the population mean, s is the sample standard deviation, and k is the number of observations.

3. The p-value is computed using the cumulative distribution function (CDF) of the t-distribution with $k - 1$ degrees of freedom:

$$p\text{-value} = 2 \times (1 - \text{CDF}(|t|))$$

The decision rule is: - If the p-value is less than 0.05 (significance level), we reject the null hypothesis. - If the p-value is greater than or equal to 0.05, we fail to reject the null hypothesis.

Step 3: Conclusion The conclusion is drawn based on the p-value:

$$\text{If } p\text{-value} < 0.05 \quad \text{then we reject } H_0 \quad \text{(significant evidence against the claim).}$$

$$\text{If } p\text{-value} \geq 0.05 \quad \text{then we fail to reject } H_0 \quad \text{(not enough evidence to contradict the claim).}$$

In this case, if the p-value is larger than 0.05, we conclude that 350 could indeed be the population mean, as there is not enough evidence to reject the null hypothesis.

Results

The results obtained were as follows :-
1.the range of 5 percent error significance is between 344.7407681169205 and 355.2592318830795.
2.the p value is as follows :0.3166392991960294
3.as 353.8 lies between the range hence the claim is excepted Part(b)
Sample Mean: 353.8
Sample Standard Deviation: 21.8478
t-Statistic: 0.7778
p-value: 0.443
Conclusion: Fail to reject the null hypothesis (not enough evidence to contradict the claim).

Question 3

Introduction

A nutritionist aims to evaluate whether a newly designed diet plan significantly impacts body weight. To investigate this, a sample of 10 individuals is selected, and their weights are measured before and after adhering to the diet for one month. The goal is to statistically determine if the observed changes in weight are significant using a paired t-test at the 5% significance level.

Problem Statement

The recorded weights (in kilograms) of the participants before and after the diet plan are given below:

Participant	**Before (kg)**	**After (kg)**	**Difference (Before - After)**
1	85.2	82.5	
2	78.5	75.8	
3	92.3	90.1	
4	80.0	77.2	
5	88.7	85.4	
6	76.4	74.5	
7	90.5	87.6	
8	84.1	81.3	
9	79.0	76.8	
10	86.2	83.0	

Hypothesis Test

We perform a **paired t-test** to assess whether there is a statistically significant mean difference in weights before and after the diet.

Let d_i represent the difference in weight for the i-th participant ($\text{Before}_i - \text{After}_i$).

- Null Hypothesis (H_0): The diet plan has no significant effect on body weight, i.e., the mean of the differences $\mu_d = 0$.
- Alternative Hypothesis (H_1): The diet plan has a significant effect on body weight, i.e., $\mu_d \neq 0$.

This is a **two-tailed test** at a significance level of $\alpha = 0.05$.

Steps

To perform the paired t-test:

1. Compute the differences $d_i = \text{Before}_i - \text{After}_i$.
2. Compute the mean difference $\bar{d}$.
3. Compute the standard deviation of the differences s_d.

4. Compute the t-statistic using:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

   where $n = 10$ is the number of participants.

5. Determine the p-value from the t-distribution with $n - 1 = 9$ degrees of freedom.

6. Compare the p-value with $\alpha = 0.05$ to draw a conclusion.

Data

Given with question

Methodology

The code used ias folowws-

```python
import numpy as np
from scipy.stats import t


before = np.array([85.2, 78.5, 92.3, 80.0, 88.7, 76.4, 90.5,
     84.1, 79.0, 86.2])
after  = np.array([82.5, 75.8, 90.1, 77.2, 85.4, 74.5, 87.6,
     81.3, 76.8, 83.0])


diff = before - after


mean_diff = np.mean(diff)
std_diff = np.std(diff, ddof=1)
n = len(diff)
t_stat = mean_diff / (std_diff / np.sqrt(n))


p_value = 2 * (1 - t.cdf(abs(t_stat), df=n-1))


print("Mean Difference:", round(mean_diff, 4))
print("Standard Deviation of Differences:", round(std_diff,
    4))
print("t-Statistic:", round(t_stat, 4))
print("p-value:", round(p_value, 4))
```

```python

if p_value < 0.05:
    print("Conclusion: Reject  H       The diet has a 
        significant effect.")
else:
    print("Conclusion: Fail to reject  H       No 
        significant effect detected.")
```

- `before` and `after` contain the body weights (in kg) of 10 individuals before and after the diet, respectively.
- `diff = before - after` computes the paired differences d_i, which are the changes in weight for each individual.
- `mean_diff` is the sample mean of the differences:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

- `std_diff` is the sample standard deviation of the differences:

$$s_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2}$$

- `t_stat` calculates the t-statistic using the formula:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

  where $n = 10$ is the number of participants.

- `p_value` computes the two-tailed p-value from the t-distribution with $n - 1 = 9$ degrees of freedom:

$$\text{p-value} = 2 \cdot (1 - \text{CDF}_t(|t|))$$

- The code checks whether the p-value is less than the significance level $\alpha = 0.05$:
  - If `p_value < 0.05`, we reject the null hypothesis $H_0 : \mu_d = 0$, concluding that the diet plan has a statistically significant effect.
  - Otherwise, we fail to reject the null hypothesis, indicating that there is insufficient evidence to conclude that the diet plan has a significant effect.

Result

The output obtained were as follows:-
1.Mean Difference: 2.67
2.Standard Deviation of Differences: 0.4473
3.t-Statistic: 18.8745
4.p-value: 0.0
5.Conclusion: Reject H → The diet has a significant effect.

Question 4

The quality control team at a pharmaceutical company needs to verify whether their IV fluid filling machine maintains the advertised consistency in volume. The manufacturer claims that the variance of the fluid volume is at most $\sigma^2 \leq 4$ mL^2. A random sample of 15 bottles has been collected with the following volumes (in mL):

$$\{502, 498, 505, 497, 503, 499, 504, 496, 501, 500, 506, 495, 502, 498, 504\}$$

Objectives

(i) Calculate the sample variance of the dataset.

(ii) Perform a chi-square test at a significance level of $\alpha = 0.01$ to determine whether the machine violates the variance specification.

(iii) Investigate how removing potential outliers (volumes less than 495 mL or greater than 505 mL) affects the conclusion.

Data

Given in question

Methodology

```
import numpy as np
from scipy.stats import chi2
volumes = np.array([502, 498, 505, 497, 503, 499, 504, 496,
    501, 500, 506, 495, 502, 498, 504])
sigma_sq_0 = 4
alpha = 0.01

def variance_test(data, sigma_sq_0, alpha):
```

```
    n = len(data)
    sample_var = np.var(data, ddof=1)
    chi_stat = (n - 1) * sample_var / sigma_sq_0
    p_value = 1 - chi2.cdf(chi_stat, df=n - 1)
    critical_value = chi2.ppf(1 - alpha, df=n - 1)

    print("Sample Variance:", round(sample_var, 4))
    print("Chi-square Statistic:", round(chi_stat, 4))
    print("Critical Value (        .       ):", round(
        critical_value, 4))
    print("p-value:", round(p_value, 4))

    if chi_stat > critical_value:
        print("Conclusion: Reject  H        Variance exceeds
            specification.")
    else:
        print("Conclusion: Fail to reject  H        Variance
            is within acceptable limits.")

print("=== Full Dataset ===")
variance_test(volumes, sigma_sq_0, alpha)


filtered_volumes = volumes[(volumes >= 495) & (volumes <=
    505)]

print("\n=== After Removing Outliers (<495 or >505) ===")
variance_test(filtered_volumes, sigma_sq_0, alpha)
```

Explanation

1. The dataset `volumes` contains 15 IV fluid volumes in mL.
2. The null hypothesis and alternative hypothesis are defined as:

$$H_0 : \sigma^2 \leq 4 \quad \text{(variance is acceptable)}$$
$$H_1 : \sigma^2 > 4 \quad \text{(variance exceeds specification)}$$

3. The function `variance_test` performs the test using the formula:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

   where:

   - s^2 is the sample variance
   - $\sigma_0^2 = 4$ is the claimed variance

- n is the sample size

4. The critical value is obtained from the chi-square distribution at the 99% confidence level (i.e., significance level $\alpha = 0.01$):

$$\chi^2_{\text{critical}} = \chi^2_{1-\alpha,\, df=n-1}$$

5. The `p-value` is calculated as:

$$p = P(\chi^2 \geq \chi^2_{\text{stat}}) = 1 - \text{CDF}(\chi^2_{\text{stat}})$$

6. If the test statistic exceeds the critical value, we reject H_0, concluding that the variance is too high.

7. The program runs the test:
   - First using the full dataset.
   - Then using a filtered dataset where outliers outside [495, 505] are removed.

8. This comparison investigates whether the removal of extreme values significantly alters the conclusion of the hypothesis test.

Conclusion

The results obtained were as follows:-

Using the Full Dataset

- **Sample Variance:** 11.6667
- **Chi-square Statistic:** 40.8333
- **Critical Value ($\chi^2_{0.99,\,14}$):** 29.1412
- **p-value:** 0.0002
- **Conclusion:** Reject $H_0 \Rightarrow$ Variance exceeds specification.

After Removing Outliers (Volumes ¡ 495 or ¿ 505)

- **Sample Variance:** 10.2198
- **Chi-square Statistic:** 33.2143
- **Critical Value ($\chi^2_{0.99,\,12}$):** 27.6882
- **p-value:** 0.0016
- **Conclusion:** Reject $H_0 \Rightarrow$ Variance exceeds specification.