Statistic Part 2 Assignment

1. What is hypothesis testing in statistics?
Hypothesis testing is a statistical method for making decisions or inferences about a population based on sample data. It involves testing an assumption (hypothesis) about a population parameter.

---

2. What is the null hypothesis, and how does it differ from the alternative hypothesis?

- **Null Hypothesis (H_0):** A default assumption that there is no effect or no difference.
- **Alternative Hypothesis (H_1 or Ha):** The hypothesis that there is an effect or a difference.

---

3. What is the significance level in hypothesis testing, and why is it important?

- The **significance level (α)** is the probability of rejecting the null hypothesis when it is actually true (Type I error).
- Common values are 0.05, 0.01.
- It defines the threshold for statistical significance.

4. What does a P-value represent in hypothesis testing?
The **P-value** is the probability of obtaining a test statistic as extreme as the one observed, assuming the null hypothesis is true.

5..**How do you interpret the P-value in hypothesis testing?**

- If **$P \le \alpha$**, reject the null hypothesis (evidence is statistically significant).
- If **$P > \alpha$**, fail to reject the null hypothesis (not statistically significant).

---

6. What are Type 1 and Type 2 errors in hypothesis testing?

- **Type I Error (False Positive):** Rejecting a true null hypothesis.
- **Type II Error (False Negative):** Failing to reject a false null hypothesis.

---

7. What is the difference between a one-tailed and a two-tailed test in hypothesis testing?

- **One-tailed test:** Tests for an effect in one direction (e.g., H_1: $\mu > \mu_0$).
- **Two-tailed test:** Tests for any difference (e.g., H_1: $\mu \neq \mu_0$).

---

8. What is the Z-test, and when is it used in hypothesis testing?
Used when:

- The sample size is large (n > 30).
- Population standard deviation is known.
- The data is approximately normally distributed.

9. How do you calculate the Z-score, and what does it represent in hypothesis testing?

Z=x¯−μσ/nZ = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}Z=σ/nx¯−μ

Where:

- x¯\bar{x}x¯: Sample mean
- μ\muμ: Population mean
- σ\sigmaσ: Population standard deviation
- nnn: Sample size
  Z-score tells how many standard deviations the sample mean is from the population

mean.

10. What is the T-distribution, and when should it be used instead of the normal distribution?
Use the **T-distribution** when:

- Sample size is small (n < 30).
- Population standard deviation is unknown.

11. **What is the T-test, and how is it used in hypothesis testing?**
The **T-test** is used to compare sample means to a known value or another sample when population standard deviation is unknown.

12. What is the difference between a Z-test and a T-test?

- **Z-test:** Large sample size, known population standard deviation.
- **T-test:** Small sample size, unknown population standard deviation.

13. What is the relationship between Z-test and T-test in hypothesis testing?
They both test hypotheses about population means, but the **T-test** adjusts for extra uncertainty in small samples. As the sample size increases, the T-distribution approaches the normal (Z) distribution.

---

14. What is a confidence interval, and how is it used to interpret statistical results?
A **confidence interval** gives a range within which the population parameter is likely to fall with a certain confidence level (e.g., 95%).

15 What is the margin of error, and how does it affect the confidence interval?
The **margin of error** defines how far the interval can stretch from the sample estimate. Larger margin = wider interval = more uncertainty.

16.How is Bayes' Theorem used in statistics, and what is its significance?
Bayes' Theorem updates the probability of a hypothesis based on new evidence:

P(H∣E)=P(E∣H)⋅P(H)P(E)P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}P(H∣E)=P(E)P(E∣H)⋅P(H)

It's the foundation of Bayesian statistics, useful when incorporating prior knowledge.

❖ Chi-square & F-distributions

17. What is the Chi-square distribution, and when is it used?
Used for categorical data to test:

- Independence in contingency tables.
- Goodness of fit of observed vs. expected frequencies.

18.What is the Chi-square goodness of fit test, and how is it applied?
It tests if a sample matches an expected distribution:

χ2=∑(O−E)2E\chi^2 = \sum \frac{(O - E)^2}{E}χ2=∑E(O−E)2

Where:

- O = observed frequency
- E = expected frequency

19. What is the F-distribution, and when is it used in hypothesis testing?
Used to compare **variances** between two samples or in **ANOVA**. Skewed right and depends on degrees of freedom.

---

20 What is an ANOVA test, and what are its assumptions?
ANOVA (Analysis of Variance) compares means of 3 or more groups.
Assumptions:

- Normality
- Homogeneity of variances

- Independence of observations

21 -What are the different types of ANOVA tests?

- **One-Way ANOVA:** One independent variable
- **Two-Way ANOVA:** Two independent variables
- **Repeated Measures ANOVA:** Same subjects across conditions

23- What is the F-test, and how does it relate to hypothesis testing?
The **F-test** is used to compare two variances or to test significance in ANOVA. A high F-value suggests group means are significantly different.

Practical Solution

1.Write a Python program to perform a Z-test for comparing a sample mean to a known population mean and interpret the results@. Z-test: Comparing Sample Mean to Population Mean

```
from scipy import stats

import numpy as np


# Known population values

population_mean = 100

population_std = 15
```

```
# Sample data
sample = np.array([102, 98, 101, 99, 97, 100, 103, 104, 96, 98])
sample_mean = np.mean(sample)
sample_size = len(sample)

# Z-test
z_score = (sample_mean - population_mean) / (population_std /
np.sqrt(sample_size))
p_value = 2 * (1 - stats.norm.cdf(abs(z_score)))

print(f"Z-score: {z_score:.3f}")
print(f"P-value: {p_value:.4f}")

# Interpretation
alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: The sample mean is significantly different.")
else:
    print("Fail to reject the null hypothesis: No significant difference found.")
```

---

2. Simulate Data & Perform Hypothesis Testing

python

CopyEdit

```
np.random.seed(42)
data = np.random.normal(loc=102, scale=15, size=50)  # Simulated sample
population_mean = 100


z_score = (np.mean(data) - population_mean) / (15 / np.sqrt(len(data)))
p_value = 2 * (1 - stats.norm.cdf(abs(z_score)))


print(f"Simulated Z-score: {z_score:.3f}, P-value: {p_value:.4f}")
```

---

3. One-Sample Z-test Function

python

CopyEdit

```
def one_sample_z_test(sample, pop_mean, pop_std):
    mean = np.mean(sample)
    n = len(sample)
    z = (mean - pop_mean) / (pop_std / np.sqrt(n))
    p = 2 * (1 - stats.norm.cdf(abs(z)))
```

```python
    return z, p


# Example usage
sample = np.random.normal(102, 15, 30)
z, p = one_sample_z_test(sample, 100, 15)
print(f"Z-score: {z:.2f}, P-value: {p:.4f}")
```

---

4. Two-tailed Z-test with Visualization

python

CopyEdit

```python
import matplotlib.pyplot as plt


def visualize_two_tailed_z(z_score, alpha=0.05):
    x = np.linspace(-4, 4, 1000)
    y = stats.norm.pdf(x)


    plt.figure(figsize=(10, 5))
    plt.plot(x, y)
    plt.fill_between(x, y, where=(x < -stats.norm.ppf(1 - alpha/2)) | (x > stats.norm.ppf(1 - alpha/2)), color='red', alpha=0.3, label='Rejection Region')
    plt.axvline(z_score, color='blue', linestyle='--', label=f'Z-score: {z_score:.2f}')
```

```
    plt.title('Two-Tailed Z-test Decision Region')
    plt.legend()
    plt.grid(True)
    plt.show()

visualize_two_tailed_z(z_score)
```

---

5. Visualize Type I & Type II Errors

python

CopyEdit

```
def visualize_errors(mu0=100, mu1=103, sigma=10, n=30, alpha=0.05):
    se = sigma / np.sqrt(n)
    x = np.linspace(90, 110, 1000)

    null_dist = stats.norm(mu0, se)
    alt_dist = stats.norm(mu1, se)

    critical_value = stats.norm.ppf(1 - alpha, mu0, se)

    plt.plot(x, null_dist.pdf(x), label='H₀ distribution', color='blue')
    plt.plot(x, alt_dist.pdf(x), label='H₁ distribution', color='green')

    plt.fill_between(x, 0, null_dist.pdf(x), where=(x >= critical_value), color='red', alpha=0.3, label='Type I Error')
    plt.fill_between(x, 0, alt_dist.pdf(x), where=(x < critical_value), color='orange', alpha=0.3, label='Type II Error')

    plt.axvline(critical_value, color='black', linestyle='--', label='Critical value')
    plt.title('Type I and Type II Errors')
    plt.legend()
    plt.grid(True)
    plt.show()

visualize_errors()
```

---

6. Independent T-test

python
CopyEdit
```
group1 = np.random.normal(100, 10, 30)
group2 = np.random.normal(105, 10, 30)

t_stat, p_val = stats.ttest_ind(group1, group2)
print(f"T-statistic: {t_stat:.2f}, P-value: {p_val:.4f}")

if p_val < 0.05:
    print("Reject null hypothesis: Groups are significantly different.")
else:
    print("Fail to reject null hypothesis: No significant difference.")
```

---

7. Paired T-test with Visualization

python
CopyEdit
```
before = np.random.normal(100, 10, 20)
after = before + np.random.normal(2, 3, 20)

t_stat, p_val = stats.ttest_rel(before, after)
print(f"Paired T-test: t = {t_stat:.2f}, p = {p_val:.4f}")

plt.plot(before, label='Before')
plt.plot(after, label='After')
plt.title('Before vs. After (Paired Samples)')
plt.legend()
plt.show()
```

---

8. Simulate and Compare Z-test vs T-test

python

```
CopyEdit
sample = np.random.normal(100, 15, 25)

# Z-test
z_stat = (np.mean(sample) - 100) / (15 / np.sqrt(len(sample)))
z_p = 2 * (1 - stats.norm.cdf(abs(z_stat)))

# T-test
t_stat, t_p = stats.ttest_1samp(sample, 100)

print(f"Z-test: Z = {z_stat:.2f}, P = {z_p:.4f}")
print(f"T-test: t = {t_stat:.2f}, P = {t_p:.4f}")
```

---

9. Confidence Interval Function

```
python
CopyEdit, confidence=0.95):
def confidence_interval(sample
    n = len(sample)
    mean = np.mean(sample)
    std_err = stats.sem(sample)
    margin = stats.t.ppf((1 + confidence) / 2, n - 1) * std_err
    return (mean - margin, mean + margin)

# Example
data = np.random.normal(100, 10, 30)
ci = confidence_interval(data)
print(f"95% Confidence Interval: {ci[0]:.2f} to {ci[1]:.2f}")
```