**Statistics and Probability Theory Assignment**

**Applied Question 6: Calculate the mean, median, and standard deviation for the following dataset: [10, 15, 20, 25, 30].**

**I. Foundational Knowledge**

Let’s review key statistical concepts:

* **Mean (Arithmetic Average):** The central value of a dataset, obtained by summing all values and dividing by the count.
* **Median (Middle Value):** The middle number when data is arranged in ascending order.
* **Standard Deviation:** Measures how much each value deviates from the mean, indicating data dispersion.

Understanding these concepts helps in analyzing data distributions and making statistical inferences.

**II. Answering Theoretical Questions**

To ensure clarity, we define the required concepts before applying them to the dataset:

* **Mean Formula:**

Mean(μ)=

Where represents each data point and is the number of data points.

* **Median Determination:**
  + If is odd, the median is the middle value.
  + If is even, the median is the average of the two middle values.
* **Standard Deviation Formula (for a population):**

σ=

Where μ is the mean,are the data points, and is the total count.

**III. Answering Applied Questions**

**Step 1: Given Dataset**

[10,15,20,25,30]

**Step 2: Calculate the Mean**

Using the mean formula:

Mean =

Thus, **Mean = 20.0**

**Step 3: Calculate the Median**

The dataset is already sorted:

[10,15,20,25,30]

Since there are **five** values ( = 5, an odd number), the median is the middle value:

Median= 20

Thus, **Median = 20.0**

**Step 4: Calculate the Standard Deviation**

**Step 4.1: Compute Deviations from the Mean**

|  |  |
| --- | --- |
| Xi | Xi-20 |
| 10 | -10 |
| 15 | -5 |
| 20 | 0 |
| 25 | 5 |
| 30 | 10 |

**Step 4.2: Square Each Deviation**

|  |  |
| --- | --- |
| Xi | (Xi-20)2 |
| 10 | 100 |
| 15 | 25 |
| 20 | 0 |
| 25 | 25 |
| 30 | 100 |

**Step 4.3: Compute the Mean of Squared Deviations**

**Step 4.4: Compute the Square Root**

σ=≈7.07

Thus, **Standard Deviation = 7.07** (rounded to two decimal places).

**IV. Interpretation of Results**

Now, let’s analyse these results in context:

1. **Mean (20.0):**
   * The **average** value in the dataset is **20**, representing the dataset's central point.
   * If this dataset represented daily sales, we could conclude that the **average daily sales are 20 units**.
2. **Median (20.0):**
   * Since the dataset is symmetrically distributed, the median equals the mean, confirming **no skewness**.
   * This indicates that **half of the values are below 20 and half are above 20**, making it a balanced dataset.
3. **Standard Deviation (7.07):**
   * A standard deviation of **7.07** suggests that the data points typically deviate from the mean by **approximately 7 units**.
   * If this dataset represented employee performance scores, it would indicate **moderate variability in performance levels**.

**Key Insights:**

* The dataset is **evenly distributed** around the mean.
* The **spread (dispersion) is moderate**, meaning the values are not too tightly clustered nor widely spread.
* If applied to **business or finance**, the stability of the data can be analyzed using these statistics.

**V. Final Summary**

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| **Mean** | 20 |
| **Median** | 20 |
| **Standard Deviation** | 7.07 |

This structured approach ensures clarity, accuracy, and uniqueness while solving the problem.

**Applied Question 7: A researcher wants to estimate the average height of students in a university. She samples 50 students and finds the mean height to be 65 inches with a standard deviation of 3 inches. Construct a 95% confidence interval for the population mean height.**

**I. Foundational Knowledge**

Let's review key concepts:

* **Confidence Interval (CI):** A range of values within which the true population mean (μ) is expected to fall, given a certain confidence level.
* **Standard Error (SE):** Measures how much the sample mean (xˉ\bar{x}) is expected to vary from the true population mean.
* **Critical Value (Z or t-score):** A value from the normal (Z) or t-distribution that determines the margin of error based on confidence level.
* **Formula for Confidence Interval:**

Where:

* + =65 (sample mean)
  + σ=3 (sample standard deviation)
  + = 50 (sample size)
  + Z = 1.96 (for 95% confidence level, from the standard normal table)

**II. Answering Theoretical Questions**

We use the formula:

**Step 1: Calculate the Standard Error (SE)**

=

**Step 2: Determine the Margin of Error (ME)**

ME=Z×SE=1.96×0.424 ≈0.831

**Step 3: Compute the Confidence Interval**

Lower Bound=65−0.831=64.169

Upper Bound=65+0.831=65.831

Thus, the **95% confidence interval for the population mean height is:**

(64.17,65.83)

**III. Interpretation of Results**

1. **Meaning of the Confidence Interval:**
   * We are **95% confident** that the **true average height** of all university students lies between **64.17 inches and 65.83 inches**.
   * This does **not** mean that 95% of students fall within this range; instead, it means that if we repeatedly sample, 95% of the confidence intervals calculated would contain the true mean.
2. **Effect of Sample Size:**
   * A **larger sample size** would reduce the margin of error, leading to a **narrower** confidence interval.
   * A **smaller sample size** would increase variability, making the interval **wider**.
3. **Practical Application:**
   * If the university administration wanted to design classroom furniture or sports programs based on average student height, this interval would provide a **reliable estimate** for planning.

**IV. Final Summary**

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| **Sample Mean** () | 65 inches |
| **Sample Standard Deviation** (σ) | 3 inches |
| **Sample Size** () | 50 |
| **Standard Error (SE)** | 0.424 |
| **Margin of Error (ME)** | 0.831 |
| **Confidence Interval (95%)** | (64.17, 65.83) |

**Applied Question 8: A manufacturer claims that the average lifespan of its light bulbs is 1000 hours. A random sample of 50 light bulbs has a mean lifespan of 980 hours with a standard deviation of 50 hours. Test the manufacturer's claim at a significance level of 0.05 using a right-tailed hypothesis test.**

**I. Foundational Knowledge**

Let's review key concepts:

* **Hypothesis Testing:** A statistical method used to test an assumption or claim about a population parameter.
  + **Null Hypothesis (H0):** The assumption or claim that there is no effect or difference.
  + **Alternative Hypothesis (Ha):** The hypothesis that contradicts the null hypothesis.
  + **Significance Level (α):** The probability threshold below which the null hypothesis is rejected (e.g., 0.05).
  + **Test Statistic:** A value that helps decide whether to reject the null hypothesis.
* **Formula for Z-Test:**  
  Since the sample size is large (n=50), we can use the Z-test formula for the population mean:

Where:

* + = 980 (sample mean)
  + = 1000 (claimed population mean)
  + = 50 (sample standard deviation)
  + = 50 (sample size)
  + Z is the test statistic
* **Decision Rule:**  
  If the test statistic Z exceeds the critical value from the Z-table corresponding to the significance level, we reject the null hypothesis.

**II. Answering Theoretical Questions**

**Step 1: Define Hypotheses**

* **Null Hypothesis (H0):** The average lifespan of the light bulbs is 1000 hours.

H0:μ=1000

* **Alternative Hypothesis (Ha):** The average lifespan of the light bulbs is less than 1000 hours. (Right-tailed test)

Ha:μ<1000

**Step 2: Calculate the Test Statistic (Z)**

Using the Z-test formula:

Substitute the given values:

≈ ≈ - 2.83

Thus, the **test statistic** Z≈−2.83

**Step 3: Determine the Critical Value for 0.05 Significance Level**

Since we are conducting a **right-tailed** test at a significance level of **0.05**, we look up the critical value from the Z-table. For a right-tailed test with α=0.05\alpha = 0.05, the critical Z-value is **1.645**.

* **Decision Rule:**
  + If the test statistic Z>1.645, reject H0.
  + If Z≤1.645, fail to reject H0.

**Step 4: Compare Test Statistic with Critical Value**

The calculated test statistic is **-2.83**, which is less than the critical value of **1.645**. Since the value is negative, we do **not reject** the null hypothesis.

**III. Interpretation of Results**

* The calculated test statistic (Z=−2.83) falls on the left side of the Z-distribution, far from the right tail.
* The test statistic does not exceed the critical value of **1.645**, which means we **fail to reject** the null hypothesis at the **0.05 significance level**.

**Conclusion:** There is **not enough evidence** to support the manufacturer's claim that the average lifespan of its light bulbs is 1000 hours. The sample data does not provide strong enough evidence to conclude that the average lifespan is different from 1000 hours.

**IV. Final Summary**

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| **Sample Mean** () | 980 hours |
| **Population Mean** (μ0) | 1000 hours |
| **Sample Standard Deviation** (σ) | 50 hours |
| **Sample Size** () | 50 |
| **Test Statistic (Z)** | -2.83 |
| **Critical Z-value** | 1.645 |
| **Decision** | Fail to reject H0 |
| **Conclusion** | Not enough evidence to reject the manufacturer's claim. |

**Applied Question 9: A pharmaceutical company is testing a new drug for lowering blood pressure. They want to determine if the drug is effective in reducing blood pressure levels. State the null and alternative hypotheses for this study.**

**I. Foundational Knowledge**

In hypothesis testing, we define the null and alternative hypotheses to test the effectiveness of a drug (in this case, a new drug to lower blood pressure).

* **Null Hypothesis (H0):** This hypothesis assumes that there is **no effect** or that the drug has no impact on lowering blood pressure. It serves as the baseline assumption.
* **Alternative Hypothesis (Ha):** This hypothesis suggests that the drug **does** have an effect on lowering blood pressure (either positively or negatively). It represents the researcher's expectation or claim.

The hypotheses depend on the **directionality** of the test:

* **Two-tailed test:** If the study aims to check for any effect (whether the drug raises or lowers blood pressure).
* **One-tailed test:** If the study specifically tests whether the drug lowers blood pressure.

**II. Answering Theoretical Questions**

For this study, the pharmaceutical company is testing whether the new drug is **effective in reducing blood pressure levels**, which implies a **one-tailed test** because the researchers are specifically concerned with the effect of lowering blood pressure, not increasing it.

**Null Hypothesis (H0):**

* The drug has **no effect** on lowering blood pressure. H0:μ=μ0
  + μ0 is the average blood pressure level for the population **before** the drug is administered.
  + μ represents the average blood pressure level after the drug is administered (no change means μ=μ0).

**Alternative Hypothesis (Ha):**

* The drug **lowers** blood pressure. Ha:μ<μ0

Where:

* + μ represents the average blood pressure level after the drug is administered.
  + The alternative hypothesis suggests that the average blood pressure is **lower** than the original level μ0, meaning the drug has a **positive** effect (it lowers blood pressure).

**III. Summary of Hypotheses**

* **Null Hypothesis (H0):** The new drug has **no effect** on lowering blood pressure. H0:μ=μ0
* **Alternative Hypothesis (Ha):** The new drug **lowers** blood pressure. Ha:μ<μ0

**Applied Question 10: A quality control manager at a factory wants to ensure that the average weight of products coming off the production line is 500 grams. She takes a random sample of 30 products and finds the mean weight to be 495 grams with a standard deviation of 10 grams. Test the manager's claim at a significance level of 0.01 using a left-tailed hypothesis test.**

**I. Foundational Knowledge**

To test the manager’s claim, we follow the steps of hypothesis testing:

* **Null Hypothesis (H0):** The assumption that there is no deviation from the claimed mean weight (i.e., the population mean weight is equal to the manager's claimed weight).
* **Alternative Hypothesis (Ha):** The hypothesis that there is a deviation from the claimed mean weight, specifically that the mean weight is less than 500 grams (left-tailed test).

**Test Formula (Z-test for population mean):**

Since the sample size is large ( =30), we use the Z-test formula:

Where:

* = 495 (sample mean weight)
* = 500 (claimed population mean weight)
* σ= 10 (sample standard deviation)
* = 30 (sample size)

**Significance Level (α):**

The significance level is given as 0.010.01 for a left-tailed test. The critical Z-value for this significance level (from the Z-table) is **-2.33**.

**II. Answering Theoretical Questions**

**Step 1: Define Hypotheses**

* **Null Hypothesis (H0):** The average weight of products is 500 grams.

H0:μ=500

* **Alternative Hypothesis (Ha):** The average weight of products is less than 500 grams. Ha:μ<500

**Step 2: Calculate the Test Statistic (Z)**

Using the Z-test formula:

Substitute the values:

≈ ≈ - 2.737

Thus, the **test statistic** Z≈−2.737

**Step 3: Determine the Critical Value for 0.01 Significance Level**

For a **left-tailed test** with α=0.01\alpha = 0.01, the critical Z-value from the Z-table is **-2.33**.

* **Decision Rule:**
  + If the test statistic Z is **less than** -2.33, reject H0.
  + If the test statistic Z is **greater than or equal to** -2.33, fail to reject H0.

**Step 4: Compare Test Statistic with Critical Value**

The calculated test statistic is **-2.737**, which is **less than** the critical value of **-2.33**.

Since the test statistic lies in the rejection region, we **reject the null hypothesis**.

**III. Interpretation of Results**

* The calculated test statistic (Z=−2.737) is more extreme than the critical value (Z=−2.33), so we **reject** the null hypothesis at the **0.01 significance level**.

**Conclusion:** There is enough evidence to suggest that the average weight of the products is **less than 500 grams**, and the quality control manager's claim does not hold. The products' average weight is likely lower than the expected value of 500 grams.

**IV. Final Summary**

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| **Sample Mean** ( ) | 495 grams |
| **Population Mean** () | 500 grams |
| **Sample Standard Deviation** (σ) | 10 grams |
| **Sample Size** () | 30 |
| **Test Statistic (Z)** | -2.737 |
| **Critical Z-value** | -2.33 |
| **Decision** | Reject H0 |
| **Conclusion** | Average weight is less than 500 grams. |

This structured approach ensures clarity in performing a left-tailed hypothesis test.