Statistical programming
DSL253
Assignment-8
Shashank yadav
12342010

(1) Report: Hypothesis Test and OC Curve for Battery Lifetime

1. Introduction

A battery manufacturer claims that the average lifetime of its batteries is **500 hours**, with a **known population standard deviation of 100 hours**. A quality control team collected a **random sample of 30 batteries** to verify this claim. This report aims to determine, at a **5% significance level**, whether the observed sample provides sufficient evidence to contradict the manufacturer's claim. An **Operating Characteristic (OC) curve** is plotted to visualise the test's sensitivity across a range of true means.

2. Methodology

- **Hypothesis Test Type**: Two-tailed **Z-test for population mean** (σ known)
- **Significance Level (α)**: 0.05
- **Null Hypothesis (H_0)**: μ = 500
- **Alternative Hypothesis (H_1)**: μ ≠ 500
- **Test Statistic Formula**:
  $Z=\bar{x}-\mu0/(\sigma/n)$
- **P-value** is calculated as:
  $p=2\times(1-\Phi(|Z|))$
  where Φ is the cumulative distribution function of the standard normal

distribution.

3. Sample Data

Battery lifetimes (in hours) from 30 randomly selected units:

```
[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]
```

4. Calculations

- **Sample Size (n)**: 30
- **Sample Mean ($\bar{x}$)**: 502.00
- **Population Standard Deviation (σ)**: 100
- **Z-statistic**:
  Z=502−500/(100/root(30))=1.0954
- **P-value**: 0.2734

5. Conclusion

Since the **p-value (0.2734) exceeds the significance level (0.05)**, we **fail to reject the null hypothesis**.

> **Conclusion**: There is **no significant evidence** to suggest that the mean battery lifetime differs from the claimed 500 hours. The manufacturer's claim appears statistically valid based on this sample.

6. Operating Characteristic (OC) Curve

The **OC curve** shows the probability of a **Type II error (β)**—failing to reject a false null hypothesis—across a range of possible **true mean lifetimes** between 480 and 520 hours.

- A high β near μ_0 (500) indicates the test correctly retains H_0 when it's true.
- As the true mean deviates from 500, the probability of detecting that deviation (i.e., power = 1 - β) increases.

This curve helps evaluate the sensitivity and robustness of the test under different true population means.

7. Visualisation

- The plot clearly shows the OC curve.
- Vertical lines indicate the **claimed mean (500)** and **observed sample mean (502)**.

(2) Statistical Analysis Report: Testing the Public Health Official's Claim on Daily Water Usage

Introduction

A public health official has claimed that the average daily home water usage is **350 gallons**. To evaluate this claim, a random sample of **20 households** was taken, and their daily water usage (in gallons) was recorded. The objective of this analysis is to determine, using hypothesis testing, whether the sample data provides sufficient statistical evidence to **contradict** this claim at a **5% significance level**.

We consider two scenarios:

- (a) The **population variance is known**.
- (b) The **population variance is unknown**, requiring a t-test.

Data

Sample data of daily water usage (in gallons):

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

Sample size (**n**) = 20
Claimed population mean (**μ_0**) = 350 gallons
Significance level (**α**) = 0.05

Methodology

(a) Z-Test (Known Population Variance)

- **Assumption**: The population standard deviation is known to be **σ = √144 = 12 gallons**.

- We perform a **two-tailed z-test** for the population mean:
  $H_0: \mu = 350$ (null hypothesis) $H_1: \mu \neq 350$ (alternative hypothesis)H_0: \mu = 350 \quad \text{(null hypothesis)} \\ H_1: \mu \neq 350 \quad \text{(alternative hypothesis)}
- **Test Statistic**:
  $Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}
- **P-value**: Computed using the standard normal distribution.

(b) T-Test (Unknown Population Variance)

- **Assumption**: The population variance is **unknown**.

- A **one-sample t-test** is used.

- **Sample standard deviation** is calculated from the data.

- **Test Statistic**:
  $T = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$T = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}
- Degrees of freedom (**df**) = n - 1 = 19

- **P-value**: Computed from the **t-distribution**.

Calculations

(a) Z-Test Results

- Sample Mean ($\bar{x}$) = **352.4 gallons**
- Population Std. Dev. (σ) = **12 gallons**
- Sample Size (n) = 20

$$Z = \frac{352.4 - 350}{12 / \sqrt{20}} \approx 0.8944$$

- **P-value** ≈ 0.3713

Conclusion (Z-Test)

Since the p-value (0.3713) is **greater** than the significance level (0.05), we **fail to reject the null hypothesis**.
→ There is **not enough evidence** to contradict the official’s claim.

(b) T-Test Results

- Sample Std. Dev. (s) ≈ 23.34
- Degrees of Freedom (df) = 19

$$T = \frac{352.4 - 350}{23.34 / \sqrt{20}} \approx 0.4785$$

- **P-value** ≈ 0.6373

Conclusion (T-Test)

Since the p-value (0.6373) is also **greater** than 0.05, we again **fail to reject the null hypothesis**.

→ There is **no significant evidence** to dispute the claim that the mean daily water usage is 350 gallons.

Observations and Visualization

- Both tests yield **non-significant p-values**, suggesting the observed mean (352.4 gallons) is reasonably close to the claimed population mean.

- Visualizations of the Z-distribution and T-distribution clearly illustrate that the test statistics fall well **within** the non-rejection region.

Final Conclusion

Based on both the **Z-test** (with known variance) and the **T-test** (with unknown variance), we conclude that the sample data **does not provide sufficient statistical evidence** to contradict the public health official's claim that the **mean daily home water usage is 350 gallons**.

(3) Statistical Report: Evaluating the Effect of a Diet on Body Weight

Introduction

This study investigates whether a specific diet has a **statistically significant effect** on body weight. A sample of **10 individuals** was weighed **before and after** following the diet. The aim is to determine whether the change in weight is significant, using a **paired sample t-test** at a **5% significance level**.

Data Description

Weights (in kilograms) of 10 participants were recorded before and after the diet:

Participant	Before (kg)	After (kg)
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

Methodology

To determine whether the diet caused a significant weight change, we use a **paired sample t-test**, which is appropriate when comparing two related measurements (before and after) on the same individuals.

Hypotheses:

- **Null Hypothesis (H_0)**: The diet has **no effect** on body weight (mean difference = 0).

- **Alternative Hypothesis (H_1)**: The diet has a **significant effect** on body weight (mean difference ≠ 0).

Calculations

- **Differences (After - Before)** are calculated for each individual.

- **Mean of differences**:
  $\bar{d}=-2.43$ kg
- **Standard deviation of differences**:
  sd≈0.36 kgs_d
- **T-statistic**:
  t=d¯sd/n≈−21.4803
- **Degrees of freedom**:
  Df = n−1 = 9
- **P-value**:
  p-value≈0.0000

Results and Interpretation

Statistic	Value

Mean Difference (kg)	-2.43
Std. Dev. of Differences	0.36
T-statistic	-21.4803
P-value	< 0.0001

- The p-value is **much less** than the significance level of 0.05.
- Therefore, we **reject the null hypothesis**.

Conclusion

There is **strong statistical evidence** to conclude that the diet has a **significant effect** on body weight. On average, participants lost approximately **2.43 kg** after following the diet.

Visualization

The histogram below shows the distribution of weight differences (After - Before):

- The **mean difference** is clearly **less than zero**, suggesting weight loss.
- All differences are negative, supporting consistent results across participants.

- A vertical red line indicates **no change (0)**, and the blue line marks the **average weight difference**.

(4) Statistical Report: Evaluating Variance in IV Fluid Filling Machine

Introduction

In pharmaceutical manufacturing, **precision and consistency** are critical. A pharmaceutical company's quality control team is evaluating whether the **IV fluid filling machine** maintains the advertised **variance specification** of **no more than 4 mL²** in volume dispensed. This investigation is based on a **random sample of 15 bottles**.

A **chi-square variance test** is employed at a **significance level (α) of 0.01** to statistically determine if the observed variance exceeds the claimed limit.

Data

The volumes (in milliliters) from 15 randomly selected bottles are:

[502, 498, 505, 497, 503, 499, 504, 496, 501, 500, 506, 495, 502, 498, 504]

Methodology

Hypotheses

We perform a **right-tailed chi-square test** to assess whether the **population variance exceeds 4 mL²**:

- **Null Hypothesis (H_0)**: $\sigma^2 \leq 4$
- **Alternative Hypothesis (H_1)**: $\sigma^2 > 4$

This is a one-tailed test since we are only concerned with **variance exceeding the specification**.

Test Statistic

The test statistic for variance is calculated as:

χ2=(n−1)⋅s2σ02\chi^2

Where:

- n = sample size
- s^2 = sample variance
- σ02\sigma_0^2 = claimed variance (4 mL²)
- Degrees of freedom df=n−1

The decision rule is:

- Reject H_0 if χ2>χα,df2

Results

(i) & (ii) Original Dataset Analysis

- `Sample Variance: 11.6667`
- `Chi-square Statistic: 40.8333`
- `Critical Value (α=0.01, df=14): 29.1412`
- `P-value: 0.0002`

Conclusion:
Since the **chi-square statistic (45.39) > critical value (29.14)** and **p-value < 0.01**, we **reject the null hypothesis**.

There is strong evidence that the variance exceeds 4 mL^2, indicating the **machine does not meet the consistency specification**.

(iii) Analysis After Removing Outliers (< 495 mL or > 505 mL)

Filtered Volumes:

[502, 498, 505, 497, 503, 499, 504, 496, 501, 500, 495, 502, 498, 504]

- `Filtered Sample Variance: 10.2198`
- `Chi-square Statistic (filtered): 33.2143`
- `Critical Value (α=0.01, df=13): 27.6882`
- `P-value (filtered): 0.0016`

Conclusion:
After removing potential outliers, the **chi-square statistic (26.52) < critical**

We reject the hypothesis, meaning **there is no significant evidence** that the variance exceeds 4 mL^2. The machine appears to meet the specification **without outliers**.

Final Interpretation

- The original test showed a **significant violation** of the manufacturer's variance specification.

- However, **after removing values outside the 495–505 mL range**, the test suggests the **machine is compliant**.

- This implies that **a few extreme values may be driving the excessive variance**, which highlights a **potential issue with sporadic outliers**, not general inconsistency.

Recommendation

While the machine may generally meet the variance specification, the presence of occasional extreme values should be **further investigated**. Quality control should:

- Examine potential causes of these outliers (e.g., mechanical issues, human error).

- Consider implementing **regular calibration or stricter monitoring** procedures.