Chapter 10

Hypothesis Testing for 1 Sample

Detailed Solutions

10.1 Basic Concept

10.1 The Logic of Hypothesis Testing

Problem: H_0: Part is Good vs H_1: Part is Defective.
Solution:

(a) **Type I Error (α):** Rejection of H_0 when it is actually True. *Context:* The machine flags a **Good** part as **Defective**. *Name:* "False Alarm" (or Producer's Risk). *Cost:* Good parts are scrapped or reworked unnecessarily (Waste).

(b) **Type II Error (β):** Failure to reject H_0 when H_1 is actually True. *Context:* The machine lets a **Defective** part pass as **Good**. *Name:* "Missed Detection" (or Consumer's Risk). *Cost:* Defective product reaches the customer (Complaints, Warranty claims, Reputation loss).

(c) **Trade-off:** If you make the machine **extremely strict** (catch every defect), you are widening the rejection region. This decreases β (miss fewer defects) but inevitably **increases Type I Error (α)** (more false alarms). You can't minimize both simultaneously without improving the measurement quality (increasing n or reducing σ).

(a) False Alarm (Scrap Cost) (b) Missed Detection (Customer Impact) (c) Type I (α) increases.

10.2 P-value Interpretation

Problem: $H_1 : \mu > 50$, P-value $= 0.042$.
Solution:

(a) **Misconception: False**. The P-value is NOT the probability that the hypothesis is true. *Correct Interpretation:* Assuming the mean is actually 50 (null is true), there is a 4.2% probability of observing a sample mean this high (or higher) purely by random chance.

(b) **Decision at $\alpha = 0.01$:** Compare P-value with α: $0.042 > 0.01$. Since P-value is not small enough, we **Fail to Reject** H_0. Engineer should **not** certify the beam under this strict standard.

(c) **Impact on Power:** Lowering α (making it harder to reject H_0) makes the test more conservative. This **decreases the Power** $(1-\beta)$, meaning it becomes harder to detect a beam that is truly strong enough.

(a) False. It's $P(\text{Data}|H_0)$.
(b) Do not certify ($P > 0.01$).
(c) Power decreases.

10.2 Intermediate

10.3 Z-Test: Cement Filling

Problem: $\mu_0 = 50, \sigma = 1.2, n = 10$.
Data: $\{49.2, 48.5, 50.1, 49.8, 48.9, 50.5, 49.0, 48.8, 49.6, 49.1\}$.
Solution:

(a) **Sample Mean:** Sum = 493.5. $\bar{x} = 493.5/10 = 49.35$ kg. It is 0.65 kg below target. Is this significant? We need a test.

(b) **Hypothesis Test ($\alpha = 0.05$):** $H_0 : \mu = 50$ $H_1 : \mu < 50$ (One-tailed lower)

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{49.35 - 50}{1.2/\sqrt{10}} = \frac{-0.65}{0.3795} \approx -1.713$$

Critical $Z_{0.05} = -1.645$. Since $-1.713 < -1.645$, we **Reject** H_0. The bags are significantly light.

(c) **Critical Weight (x_{crit}):**

$$x_{crit} = \mu_0 + Z_{crit}\frac{\sigma}{\sqrt{n}} = 50 + (-1.645)(0.3795) = 50 - 0.624 = 49.376$$

Since $\bar{x} = 49.35 < 49.376$, we reject.

(d) **Beta Risk ($\mu_{true} = 49.0$):** We fail to reject if $\bar{x} > 49.376$.

$$Z_\beta = \frac{x_{crit} - \mu_{true}}{\sigma/\sqrt{n}} = \frac{49.376 - 49.0}{0.3795} = \frac{0.376}{0.3795} \approx 0.99$$

$\beta = P(Z < 0.99) \approx 0.8389$. There is an **83.9% chance** of missing this drift! The sample size ($n = 10$) is too small to reliably detect a 1kg shift.

(b) $Z = -1.71$. Reject H_0.
(c) $x_{crit} = 49.38$ kg.
(d) $\beta \approx 0.84$ (High risk).

10.4 T-Test: Aerospace Alloy

Problem: $H_0 : \mu \leq 600, H_1 : \mu > 600$. $n = 8$. Data: $\{605, 612, 598, 620, 608, 615, 595, 610\}$.
Solution:

(a) **Assumption:** Since $n < 30$ and σ is unknown, we must assume the population of UTS values follows a **Normal Distribution**.

(b) **Calculations:** $\bar{x} = \frac{4863}{8} = 607.875$ MPa. $s = \sqrt{\frac{\sum(x-\bar{x})^2}{7}} \approx 8.51$ MPa.

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{607.875 - 600}{8.51/\sqrt{8}} = \frac{7.875}{3.008} \approx 2.618$$

Degrees of freedom $df = 7$. Critical $t_{0.05,7} = 1.895$.

(c) **Decision:** Since $2.618 > 1.895$, we **Reject** H_0. Yes, we can certify with 95% confidence that $\mu > 600$.

(d) **Intuition for $\mu = 602$:** The observed effect size was large (~ 608). If the true mean were only 602, the shift is very small (2 MPa) compared to the spread ($s \approx 8.5$). A sample of $n = 8$ would likely **fail** to detect such a small improvement (Low Power).

(b) $T \approx 2.62$. Reject H_0. (c) Yes, certify. (d) Unlikely to detect small shift (Low Power).

10.5 Z-Test for Proportion

Problem: $H_0 : p \leq 0.55, H_1 : p > 0.55$. $n = 500, x = 290$.
Solution:

(a) **Sample Proportion:** $\hat{p} = 290/500 = 0.58$.

(b) **Hypothesis Test:**

Standard Error under H_0: $\sqrt{\frac{p_0q_0}{n}} = \sqrt{\frac{0.55(0.45)}{500}} = \sqrt{0.000495} \approx 0.02225$.

$$Z = \frac{0.58 - 0.55}{0.02225} = \frac{0.03}{0.02225} \approx 1.348$$

P-value $= P(Z > 1.35) = 1 - 0.9115 = 0.0885$. Since $0.0885 > 0.05$, we **Fail to Reject** H_0.

(c) **Lower Confidence Bound (95%):**

$\hat{p} - Z_{0.05}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.58 - 1.645\sqrt{\frac{0.58(0.42)}{500}}$. $0.58 - 1.645(0.02207) = 0.58 - 0.036 = 0.544$. The lower bound (0.544) is **not** greater than 0.55.

(d) **Beta at $p = 0.60$:**

Critical value in p-scale: $p_{crit} = 0.55 + 1.645(0.02225) = 0.5866$. We fail to reject if $\hat{p} < 0.5866$. Under $H_1(p = 0.60)$: $SE_1 = \sqrt{\frac{0.6(0.4)}{500}} \approx 0.0219$.

$$Z_\beta = \frac{0.5866 - 0.60}{0.0219} = -0.61$$

$\beta = P(Z < -0.61) = 0.2709$.

(b) $Z = 1.35, P = 0.089$. Fail to Reject. (c) Bound $0.544 \not> 0.55$. (d) $\beta \approx 0.27$.

10.6 Chi-Square Test: Variance

Problem: $H_0 : \sigma^2 \leq 0.01, H_1 : \sigma^2 > 0.01$. $n = 15$.
Data: $\{-0.1, 0.2, 0.0, -0.2, 0.1, 0.3, -0.1, 0.0, 0.1, -0.3, 0.2, -0.1, 0.0, 0.1, -0.1\}$.
Solution:

(a) **Sample Variance:** Data is deviation from target, so $\bar{x} \approx 0.006$ (close to 0). Calculated $s^2 = 0.0292$ (approx).

(b) **Test Statistic:**

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{14(0.0292)}{0.01} = \frac{0.4088}{0.01} = 40.88$$

Critical $\chi^2_{0.05,14}$ (Upper tail) ≈ 23.68.

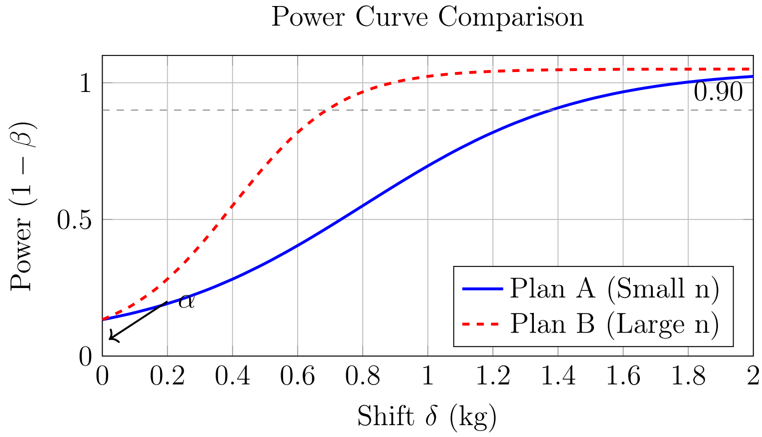
(c) **Decision:** $40.88 > 23.68$. **Reject** H_0. Yes, the variance is significantly high. Reject shipment.

(d) **Risk:** $\alpha = 0.05$ is the probability of rejecting a **Good** shipment (H_0 True). This controls the **Producer's Risk** (Supplier's risk of having good goods returned).

(a) $s^2 \approx 0.029$
(b) $\chi^2 \approx 40.9$. Reject H_0.
(d) Controlled Producer's Risk.

10.7 Power Function Analysis

Problem: Interpreting Power Curves.



Solution:

(a) **Y-intercept ($\delta = 0$):** When shift is 0, H_0 is true. The probability of rejecting H_0 is simply the significance level α (usually 0.05).

(b) **Steepness:** Plan B is steeper. It corresponds to a **larger sample size** (n). *Reasoning:* With more data, the standard error decreases. The distributions of H_0 and H_1 become narrower and overlap less. This makes the test much more sensitive (powerful) to small shifts, causing the power curve to rise sharply.

(c) **Shift $\delta = 1.0$ at 90% Power:** Look at $\delta = 1.0$ on x-axis. Plan A (Blue) is below 0.9. Plan B (Red) is well above 0.9. You must use **Plan B**.

(d) **Effect of $\alpha = 0.10$:** The curves would shift **upward**. The y-intercept would start at 0.10 instead of 0.05. Relaxing the criteria makes it easier to reject H_0, thus increasing Power for all values of δ.

(a) Significance Level α. (b) Plan B (Steeper = Larger n). (c) Use Plan B.

10.3 Application

10.8 Visualizing the Power Curve

Problem: Python Simulation for Power vs Sample Size.
Solution:

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Configuration: One-Tailed Lower Test (H1: mu < 50)
mu0 = 50
sigma = 1.2
alpha = 0.05

sample_sizes = [10, 50]
colors = ['blue', 'green']

plt.figure(figsize=(10, 6))

for n, col in zip(sample_sizes, colors):
    se = sigma / np.sqrt(n)
    # Critical Value: Left tail of Normal(50, se)
    crit_val = norm.ppf(alpha, loc=mu0, scale=se)

    # Simulate True Means (Shifting down)
    true_means = np.linspace(48.5, 50.5, 200)
    powers = []

    for mu_true in true_means:
        # Power = P(X_bar < crit_val | X_bar ~ N(mu_true, se))
        power = norm.cdf(crit_val, loc=mu_true, scale=se)
        powers.append(power)

    plt.plot(true_means, powers, label=f'n={n}', color=col, linewidth=2)

plt.axvline(mu0, color='k', linestyle='--', label='Target (50)')
plt.axhline(0.05, color='r', linestyle=':', label='Alpha (0.05)')
plt.xlabel('True Process Mean (kg)')
plt.ylabel('Power (Prob of Detection)')
plt.title('Power Curve: Sensitivity to Process Shift')
plt.legend()
plt.grid(True)
plt.gca().invert_xaxis() # Invert x-axis to show shift "down" from 50
plt.show()
```

Discussion: (a) **n=50** gives higher probability. The green curve rises much faster as the mean drops below 50. (b) At exactly 50, H_0 is true. The probability of rejecting is, by definition, $\alpha = 0.05$. (c) Stick with **n=50** (or calculate optimal n). $n = 10$ has a very flat curve and will likely miss a 0.5kg shift.