STA255 Review Topics

Fan Wang

May 10, 2025

Concepts: Estimating Parameters

Parameter Estimation:

- Using sample data to estimate population parameters such as mean and variance.
- Two primary methods:
  1. **Method of Moments (MoM):** Matches sample moments (mean, variance, etc.) to population moments.
  2. **Maximum Likelihood Estimation (MLE):** Finds parameter values that maximize the likelihood of observing the given data.

Detailed Explanation: Method of Moments (MoM)

Steps:

- Compute the first k sample moments: $m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$.
- Set these equal to the first k population moments.
- Solve the resulting equations for the parameters.

Example: Uniform Distribution

- Let $X_1, X_2, \ldots, X_n \sim U(a, b)$.
- Sample mean: $\bar{X} = \frac{a+b}{2}$.
- Solve for a, b using additional moments.

Detailed Explanation: Maximum Likelihood Estimation (MLE) (1)

Steps:

- Define the likelihood function: $L(\theta) = f(X_1, X_2, \ldots, X_n; \theta)$.
- Take the log of the likelihood: $\ell(\theta) = \ln(L(\theta))$.
- Differentiate $\ell(\theta)$ with respect to θ and set to zero.
- Solve for θ.

Detailed Explanation: Maximum Likelihood Estimation (MLE) (2)

Example: Poisson Distribution

- Let $X \sim Pois(\lambda)$.
- Log-likelihood:

$$\ell(\lambda) = \sum_{i=1}^{n} \left(X_i \ln(\lambda) - \lambda - \ln(X_i!)\right).$$

- Differentiate the log-likelihood:

$$\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^{n} \left(\frac{X_i}{\lambda} - 1\right).$$

- Set $\frac{\partial \ell}{\partial \lambda} = 0$:

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Practice Question 1: Method of Moments for Exponential Distribution

Problem:

- Let $X \sim Exp(\lambda)$, where $\lambda > 0$ is the rate parameter.
- Use the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ to estimate λ.

Solution:

- Population mean of the exponential distribution is $\mu = \frac{1}{\lambda}$.
- Set sample mean equal to population mean: $\bar{X} = \frac{1}{\lambda}$.
- Solve for λ:
$$\lambda = \frac{1}{\bar{X}}.$$
- Final estimate: $\hat{\lambda} = \frac{1}{\bar{X}}$, where $\bar{X}$ is the observed sample mean.

Practice Question 2: MLE for Normal Distribution (1)

Problem:

- Let $X \sim N(\mu, \sigma^2)$, where σ^2 is known.
- Find the maximum likelihood estimate (MLE) for μ.
- Given data: $X_1, X_2, \ldots, X_n$.

Step 1: Likelihood Function

- The probability density function for X is:

$$f(X; \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - \mu)^2}{2\sigma^2}\right).$$

- The likelihood function is:

$$L(\mu) = \prod_{i=1}^{n} f(X_i; \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - \mu)^2}{2\sigma^2}\right).$$

- Take the log of the likelihood:

$$\ell(\mu) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2.$$

Practice Question 2: MLE for Normal Distribution (2)

Step 2: Solve for μ

- Differentiate the log-likelihood with respect to μ:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu).$$

- Set $\frac{\partial \ell}{\partial \mu} = 0$:

$$\sum_{i=1}^{n} (X_i - \mu) = 0 \implies \mu = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Final Answer:

- The maximum likelihood estimate (MLE) of μ is:

$$\hat{\mu} = \bar{X},$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is the sample mean.

Concepts: Hypothesis Testing (1)

What is Hypothesis Testing?

- A formal method to test claims about population parameters.
- Involves two hypotheses:
  1. **Null Hypothesis (H_0)**: Assumes no effect or difference.
  2. **Alternative Hypothesis (H_1)**: Contradicts H_0.

Concepts: Hypothesis Testing (2)

Key Concepts:

- **Test Statistic**: A function of the sample data used to make decisions about H_0.
- ***p*-value**: The probability of observing a result as extreme or more extreme than the test statistic, assuming H_0 is true.
- α: The significance level, representing the total probability of a Type I error.
- **Type I Error (α)**: Rejecting a true null hypothesis.
- **Type II Error (β)**: Failing to reject a false null hypothesis.

Relationship Between *p*-value and Critical Values (1)

Why Are $p < \alpha$ and Critical Values Equivalent?

- *p*-value and critical values are two equivalent methods for making decisions in hypothesis testing.
- The *p*-value method:
  p-value $=$ Probability of obtaining the observed test statistic or more extreme values under H_0.

Relationship Between p-value and Critical Values (2)

The Critical Value Method:

- Critical values define the boundary of the rejection region.

  $Z_{\text{critical}} =$ The value of the test statistic that defines the rejection region

- Decision Rule:
  1. If $p < \alpha$, reject H_0.
  2. If the test statistic falls in the rejection region (e.g., $Z \geq Z_{\text{critical}}$), reject H_0.
- Both approaches lead to the same conclusion.

Practice Question 1: Testing Mean of Battery Life (1)

A battery manufacturer claims that the average lifespan of their batteries is 98 hours. A quality control team tests a random sample of 50 batteries and finds that the sample has an average lifespan of 100 hours. The population standard deviation of the battery lifespans is known to be 10 hours. Conduct a hypothesis test at a 1% significance level to determine whether the average lifespan of the batteries is greater than claimed by the manufacturer.

Practice Question 1: Testing Mean of Battery Life (2)

Step 1: State the Hypotheses

The null hypothesis (H_0) assumes that the average lifespan of the batteries is 98 hours:

$$H_0 : \mu = 98.$$

The alternative hypothesis (H_1) states that the average lifespan of the batteries is greater than 98 hours:

$$H_1 : \mu > 98.$$

Step 2: Set the Significance Level

The significance level is given as 1% ($\alpha = 0.01$).

Practice Question 1: Testing Mean of Battery Life (3)

Step 3: Compute the Test Statistic

The test statistic for a one-sample Z-test is calculated using the formula:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}},$$

where $\bar{X}$ is the sample mean, μ_0 is the population mean under H_0, σ is the population standard deviation, and n is the sample size.

Substituting the given values:

$$Z = \frac{100 - 98}{10/\sqrt{50}} = \frac{2}{1.414} \approx 1.41.$$

Practice Question 1: Testing Mean of Battery Life (4)

Step 4: Determine the Critical Value and Decision Rule

For a one-tailed test at the 1% significance level, the critical Z-value is:

$$Z_{\text{critical}} = 2.33.$$

The decision rule is as follows:

- If $Z \geq 2.33$, reject H_0.
- Otherwise, fail to reject H_0.

Since $Z = 1.41$, which is less than Z_{critical}, we fail to reject H_0.

Practice Question 1: Testing Mean of Battery Life (5)

Step 5: Conclusion

Based on the test statistic ($Z = 1.41$) and the critical value ($Z_{\text{critical}} = 2.33$), there is insufficient evidence to conclude that the average battery lifespan is greater than 98 hours. At the 1% significance level, we fail to reject the null hypothesis and cannot refute the manufacturer's claim.

Concepts: Confidence Intervals

What are Confidence Intervals?

- A confidence interval provides a range of values that is likely to contain the true value of a population parameter.
- It is used to give an interval estimation of parameters instead of a single-point estimation.

Confidence Level:

- The confidence level represents the proportion of intervals, calculated from repeated random samples, that would contain the true parameter value.
- For example, a 95% confidence level means that 95% of such intervals would capture the true parameter if the experiment were repeated many times.

Detailed Explanation: Confidence Intervals (1)

Formulas:

- For a population mean with a known standard deviation (σ):
$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$
where $Z_{\alpha/2}$ is the critical value from the standard normal distribution corresponding to $\alpha/2$.
- For a population mean with an unknown standard deviation (σ), use the sample standard deviation (s):
$$\bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}},$$
where $t_{\alpha/2,n-1}$ is the critical value from the t-distribution with $n-1$ degrees of freedom.
- For a population proportion:
$$\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

Detailed Explanation: Confidence Intervals (2)

What is α?

- α is the significance level, representing the total probability of Type I error.
- The confidence level is $1 - \alpha$. For example, a 95% confidence level implies $\alpha = 0.05$.
- In a two-tailed interval, $\alpha/2$ represents the probability in each tail of the distribution.

Critical Values:

- $Z_{\alpha/2}$: The critical value from the standard normal distribution (e.g., for $\alpha = 0.05$, $Z_{0.025} = 1.96$).
- $t_{\alpha/2,n-1}$: The critical value from the t-distribution (depends on $n-1$ degrees of freedom and $\alpha/2$).

Detailed Explanation: Confidence Intervals (3)

Example 1: Known Standard Deviation (Heights of Students)

- A sample of 40 students has a mean height $\bar{X} = 160$ cm, and the population standard deviation is $\sigma = 10$.
- Compute a 95% confidence interval:

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 160 \pm 1.96 \frac{10}{\sqrt{40}}.$$

- Simplify:

$$160 \pm 1.96 \times 1.58 \approx 160 \pm 3.1.$$

- Final interval: $[156.9, 163.1]$.

Detailed Explanation: Confidence Intervals (4)

Example 2: Unknown Standard Deviation (Weights of Boxes)

- A sample of 25 boxes has a mean weight $\bar{X} = 50$ kg and a sample standard deviation $s = 4$ kg.
- Compute a 95% confidence interval:

$$\bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}},$$

  where $t_{\alpha/2,n-1}$ is the critical value for $n-1 = 24$ degrees of freedom and $\alpha/2 = 0.025$. From the t-table, $t_{0.025,24} \approx 2.064$.
- Simplify:

$$50 \pm 2.064 \times \frac{4}{\sqrt{25}} = 50 \pm 2.064 \times 0.8 = 50 \pm 1.65.$$

- Final interval: $[48.35, 51.65]$.

Practice Question: Confidence Interval for Proportion (1)

Problem:

- In a survey of 200 voters, 60% favor a policy. Compute a 90% confidence interval for the population proportion.

Step 1: Define the Parameters

- Sample proportion ($\hat{p}$) $= 0.60$.
- Sample size (n) $= 200$.
- Significance level (α) $= 0.10$, so $\alpha/2 = 0.05$.
- Critical value $Z_{\alpha/2} = 1.645$ (from the standard normal table).

Step 2: Write the Formula

- The formula for the confidence interval is:

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Practice Question: Confidence Interval for Proportion (2)

Step 3: Substitute the Values

- Substitute the values into the formula:

$$0.60 \pm 1.645\sqrt{\frac{0.60(1-0.60)}{200}}.$$

- Calculate further:

$$0.60 \pm 1.645 \times 0.0346 = 0.60 \pm 0.057.$$

Final Confidence Interval:

- Confidence interval:

$$[0.60 - 0.057, 0.60 + 0.057] = [0.543, 0.657].$$

Interpretation:

- We are 90% confident that the true proportion of voters who favor the policy lies between 54.3% and 65.7%.