Statistical Inference

Instructor: Mohammadreza A. Dehaqani

Pouya HajiMohammadi Gohari, Parmis Bathayan

Fall 2024

Homework 3

- If you have questions, email the HW Authors or use the class group we are ready to help!
- Please check the course page for key submission guidelines and late policy details to avoid issues.
- For computational problems, your grade depends heavily on how well you analyze your results. Always include explanations alongside your code.
- This course is your launchpad for tackling real-world challengesplease ,dive in, explore, and push your knowledge beyond the classroom!
- As we mentioned in class, you'll have a quick (5-10 minute) in-person (or virtual!) hand-in session to help us check your understanding of the work you've submitted. For each assignment, about 25 students will be randomly chosen by an algorithm designed to ensure fairness. This algorithm will make sure you only present around 2 times during the term to keep things stress-free. However, if we notice inconsistencies between your work and what you present, the algorithm will adjust, increasing the chances you'll be selected again. Think of it as a dynamic process that adapts based on your performanceensuring everyone gets a fair shot!

Question 1: Say Hello to Neyman-Pearson

Let $X^{'} = (X_1, \cdots, X_n)$ denote a random sample from the distribution that has the pdf

$$f(x;\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2}\right), \ -\infty < x < \infty$$

1. Assume null and alternative hypothesis as follows:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta = \theta_1$$

   Derive the likelihood ratio and identify the test statistic.

2. Express the rejection region in terms of the test statistic.
3. Compute the power under specified parameters:

$$\theta_0 = 0, \ \theta_1 = 1, \ \alpha = 0.05$$

Question 2: Wald Test

Let $X_1, \cdots, X_n \sim Poisson(\lambda)$.

1. From MLE, find $\hat{\lambda}$.
2. What is the standard deviation of the estimated point($\hat{\lambda}$)?

3. Suppose we test the following hypothesis:

$$H_0 : \lambda = \lambda_0$$
$$H_1 : \lambda \neq \lambda_1$$

If we use the statistic:

$$\frac{\hat{\lambda} - \lambda_0}{\hat{se}}$$

What is the distribution of this statistic under H_0? Express the rejection area in terms of the test statistic.

4. (Programming) Let λ_0, $n = 20$ and $\alpha = 0.05$. Simulate $X_1, \cdots, X_n \sim Possion(\lambda_0)$ and perform above test. Repeat many times and count how often you reject the null. How close is the Type I error rate to 0.05?

Question 3: $f_0(x)$ vs $f_1(x)$

Consider two p.d.f's $f_0(x)$ and $f_1(x)$ that are defined as follows:

$$f_0(x) = \begin{cases} \frac{3}{2} & \text{for } 0 \leq x \leq \frac{2}{3} \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_1(x) = \begin{cases} \frac{9}{2}x & \text{for } 0 \leq x \leq \frac{2}{3} \\ 0 & \text{otherwise} \end{cases}$$

Suppose that a single observation X is taken from a distribution for which the p.d.f $f(x)$ is either $f_0(x)$ or $f_1(x)$, and the following simple hypothesis are to be tested:

$$H_0 : \ f(x) = f_0(x),$$
$$H_1 : \ f(x) = f_1(x)$$

1. Describe a test procedure for which the following equation would be minimum:

$$\alpha + \beta$$

where α and β are the probability of type I and II errors, respectively.

2. (Programming) Consider 1000 equally spaced decision boundaries c in the range $[0.\frac{3}{2}]$, for each decision boundary c compute the α and β. Plot the power of the test as a function of the decision boundary c and compare it to the theoretical part.

3. A medical diagnostic test is performed to detect a rare disease. The disease is serious, and a missed diagnosis (α) can result in severe consequences, while a false alarm (β) may lead to unnecessary but harmless additional tests. The test result is based on an observed value X, drawn from a distribution $f(x)$, which is either:

$$H_0 : \ f(x) = f_0(x) \text{ (patient is healthy)},$$
$$H_1 : \ f(x) = f_1(x) \text{ (patient has the disease)}$$

In medical diagnostics, minimizing false negatives (β) is prioritized, as the consequences of missing a diagnosis are far more severe than unnecessary follow-up tests. To address this, we assign risks to each type of error:

- R_1: Risk associated with α
- R_2: Risk associated with β

The expected risk is defined as:

$$\text{Expected Risk} = R_1\alpha + R_2\beta$$

Provide a procedure to minimize the expected risk by selecting an optimal decision boundary c^*(Derive the expression for the optimal decision boundary c^* based on R_1 and R_2).

Question 4: Uniformly Most Powerful

Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with pdf $f(x;\theta) = \theta x^{\theta-1}, 0 < x < 1$, zero elsewhere, where $\theta > 0$. Show the likelihood ratio has the statistic $\prod_{i=1}^{n} X_i$. Use this to determine the UMP[1] test for $H_0 : \theta = \theta^{'}$ against $H_1 : \theta < \theta^{'}$, for fixed $\theta^{'}$.

Question 5: Generalized Likelihood Ratio(Optional)

1. Let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ be independent random samples from the distributions $N(\theta_1, \theta_3)$ and $N(\theta_2, \theta_4)$, respectively. Show that the likelihood ratio for testing $H_0 : \theta_1 = \theta_2, \theta_3 = \theta_4$ against all alternatives is given by

$$\frac{\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}\right]^{n/2}\left[\frac{\sum_{j=1}^{m}(y_j-\bar{y})^2}{m}\right]^{m/2}}{\left\{\left[\sum_{i=1}^{n}(x_i-u)^2+\sum_{j=1}^{m}(y_j-u)^2\right]/(n+m)\right\}^{(n+m)/2}},$$

   where $\mu = \frac{n\bar{x}+m\bar{y}}{m+n}$.

2. Let the independent random variables X and Y have distributions that are $N(\theta_1, \theta_3)$ and $N(\theta_2, \theta_4)$, where the means θ_1 and θ_2 and common variance θ_3 are unknown. If $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ denote independent random samples from these distribution. Show that the likelihood ratio for testing $H_0 : \theta_1 = \theta_2$, unspecified, and θ_3 unspecified, can be based on the test statistic T(t-distribution) with n+m-2 degrees of freedom.

Question 6: Error Types

One generates a number X from a uniform distribution on the interval $[0, \theta]$. A hypothesis test is performed to test $H_0 : \theta = 2$ against $H_A : \theta \neq 2$ by rejecting H_0, if $X \leq 0.1$ or $X \geq 1.9$.

1. Compute the probability of a type I error. Illustrate the rejection region on a plot of the uniform probability density function when $\theta = 2$.
2. Compute the probability of a type II error if the true value of θ is 2.5. Show the calculations and illustrate how the shift in θ affects the overlap between the rejection and acceptance regions.
3. Suppose we change the decision rule to reject H_0 if $X \leq c_1$ or $X \geq c_2$, where c_1 and c_2 are adjustable thresholds. Derive appropriate values for c_1 and c_2 such that the type I error probability is exactly $\alpha = 0.05$. Explain how the new α affects the power of the test.
4. Discuss the impact of increasing θ on the type II error probability. Provide a mathematical argument to support your answer.
5. Derive an upper bound for the type I error probability α using an appropriate inequality studied in the course (e.g., for distributions with finite variance). Compare this bound to the exact value of α from part 3 and discuss its significance.
6. Identify and use an appropriate inequality to derive an upper bound for the type II error probability β when $\theta = 2.5$. If needed, feel free to search the web for relevant inequalities, and justify your choice. Discuss how such bounds are useful in hypothesis testing, especially when exact calculations are impractical. **(Optional)**

[1] Uniformly Most Powerful

Question 7: Warm Up t-test (Optional)

Given a two-sample t-test to determine if there is a significant difference between the means of two populations:

$$\text{Null Hypothesis: } H_0 : \mu_1 = \mu_2, \quad \text{Alternative Hypothesis: } H_1 : \mu_1 \neq \mu_2$$

Follow the instructions below to prove that H_0 is rejected if and only if the confidence interval for the difference between the sample means does not include zero.

(a) Derive the test statistic assuming $H_0 : \mu_1 = \mu_2$, and identify the conditions under which H_0 is rejected.

(b) Derive the confidence interval formula for the difference $\mu_1 - \mu_2$, and explain the role of the critical value in determining the interval's endpoints.

(c) Prove that the confidence interval excludes zero if and only if the t-test rejects H_0. Provide a clear explanation of the equivalence between these two approaches and interpret the result in the context of hypothesis testing.

Question 8: t-test

We perform a t-test for the null hypothesis $H_0 : \mu = 10$ at significance level $\alpha = 0.05$ by means of a dataset consisting of $n = 16$ elements with sample mean 11 and sample variance 4.

1. Should we reject the null hypothesis in favor of $H_A : \mu \neq 10$?
2. What if we test against $H_A : \mu > 10$?
3. Explain why the results of the two tests in parts 1 and 2 differ, even though the null hypothesis is the same. Discuss the relationship between the critical regions for the one-tailed and two-tailed tests by providing illustrations of the t-distribution.

Question 9: Error Types

One generates a number x from a uniform distribution on the interval $[0, \theta]$. A hypothesis test is performed to test $H_0 : \theta = 2$ against $H_A : \theta \neq 2$ by rejecting H_0, if $x \leq 0.1$ or $x \geq 1.9$.

1. Compute the probability of a type I error. Illustrate the rejection region on a plot of the uniform probability density function when $\theta = 2$.
2. Compute the probability of a type II error if the true value of θ is 2.5. Show the calculations and illustrate how the shift in θ affects the overlap between the rejection and acceptance regions.
3. Suppose we change the decision rule to reject H_0 if $x \leq c_1$ or $x \geq c_2$, where c_1 and c_2 are adjustable thresholds. Derive appropriate values for c_1 and c_2 such that the type I error probability is exactly $\alpha = 0.05$. Explain how the new α affects the power of the test.
4. Discuss the impact of increasing θ on the type II error probability. Provide a mathematical argument to support your answer.

Question 10: An Apple a Day... (Programming Question)

An agricultural researcher is studying the average number of apples consumed per day in a population. They have conducted preliminary research and found that the populations mean daily consumption is approximately 2 apples, with a standard deviation of 0.5 apples. To explore how confidence intervals behave under repeated sampling, the researcher designs an experiment as follows:

1. **Define the Population and Sample:** Assume the population distribution of daily apple consumption is normal with a mean (μ) of 2 apples per day and a standard deviation (σ) of 0.5 apples. Imagine that the researcher draws a random sample of 12 individuals from this population.

2. **Estimate the Confidence Interval (CI):** Calculate a 95% confidence interval for the sample mean of daily apple consumption for each sample of 12 individuals. Repeat this sampling process 100 times, each time recalculating the 95% confidence interval.

3. **Analyze the Confidence Interval Coverage:** Determine the percentage of times (out of 100) that the calculated confidence interval actually contains the true population mean of 2 apples per day.

4. **Compare Different Methods for CI Estimation:**

   (a) Assume two cases:

      i. The population standard deviation (σ) is known.

      ii. The population standard deviation is unknown.

   (b) For each case, calculate the 95% confidence intervals using the following approaches:

      i. Known population variance (σ): Use the Z-distribution to construct the confidence interval.

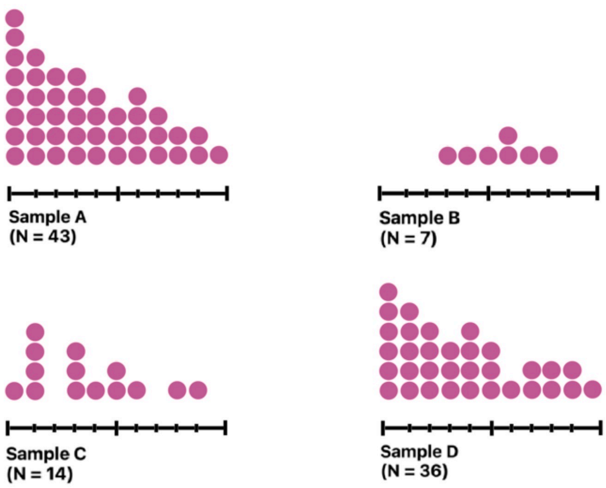
      ii. Unknown population variance: Use both the Z-distribution (with sample standard deviation s) and the T-distribution (with sample standard deviation s).

   (c) Compare the results. Observe and comment on whether and why the confidence intervals constructed with the Z-distribution, using s instead of σ, are narrower than expected.

Question 11: Short Answers - Big Concepts

Instructions: For each scenario below, analyze the situation, consider the statistical goals, and select the best testing approach. Use your understanding of hypothesis testing, sample conditions, and t-test types to choose the most appropriate answers.

1. Researchers have gathered four samples with varying characteristics. Assess each sample and determine which ones meet the conditions required for performing a t-test.



2. A restaurant advertises that its burgers weigh 250 grams. To verify this claim, the manager randomly selects 16 burgers, finding that their weights are roughly symmetric with no outliers. The manager wants to use these sample data to conduct a t-test about the mean. As a manager, based on the business goals and implications of customer satisfaction, would you test a simple or a composite hypothesis?

3. A fitness scientist wants to test the effectiveness of a new program designed to improve flexibility in senior citizens. Participants were evaluated on flexibility using a standard scale both before and after completing the program. What kind of test, null, and alternative hypothesis are most appropriate for this scenario?

Question 12: Speed Camera Error Rate

Suppose three speed cameras are set up along a stretch of road to catch people driving over the speed limit of 40 miles per hour. Each speed camera is known to have a normal measurement error modeled on $N(0, 5^2)$. For a passing car, let $\bar{x}$ be the average of the three readings. Our default assumption for a car is that it is not speeding.

1. Find the distribution of $\bar{x}$. Describe the above story in the context of hypothesis testing (write down H_0 and H_A).

2. The police would like to set a threshold on $\bar{x}$ for issuing tickets so that no more than 4% of law-abiding, non-speeders are mistakenly given tickets. This means that they set the threshold conservatively so that no more than 4% of drivers going exactly 40 mph get a ticket.
   (a) Use the hypothesis testing description in part 1 to help determine what threshold they should set.
   (b) Illustrate the distribution and the rejection area.
   (c) If possible, calculate the probability that a person getting a ticket was not speeding.
   (d) Suppose word gets out about the speed trap, and no one speeds along it anymore. What percentage of tickets are given in error?

3. What is the power of this test with the alternative hypothesis that the car is traveling at 45 miles per hour? How many cameras are needed to achieve a power of 0.9 with $\alpha = 0.04$? (Optional)

Question 13: Statistically Delicious Coffee

Imagine you're a data scientist hired by a coffee shop chain that wants to optimize customer satisfaction by improving coffee quality and drink preparation speed. The coffee shop wants to test two things:

- Whether there is a difference in preparation times between the coffee shop branches A and B.
- Whether a new coffee recipe B tastes better than the original coffee recipe A among customers.

1. You design an experiment to collect data for both objectives. Based on the nature of each comparison, what statistical tests would be appropriate to determine if there are significant differences for each case?

2. The coffee shop suspects Branch A might be faster at preparing drinks than Branch B. You measure the preparation time (in minutes) of a random sample of 12 orders from each branch. Based on your answer in part 1, test if there is a statistically significant difference in preparation times between Branch A and Branch B. (Assume equal variances, and let $\alpha = 0.05$). The data is shown in the table below:

3. The coffee shop also wants to test whether the new coffee recipe (Recipe B) tastes better than the original recipe (Recipe A). They conduct a taste test with 12 customers, each of whom rates both Recipe A and Recipe B on a scale of 1-10. Based on your answer in part 1, test if there is a statistically significant improvement in taste ratings with Recipe B compared to Recipe A. (Let $\alpha = 0.05$). The data is shown in the table below:

Order	Branch A	Branch B
1	6	7
2	7	6
3	7	7
4	8	8
5	6	9
6	9	7
7	5	10
8	7	9
9	8	9
10	7	9
11	5	7
12	7	9

Table 1: Preparation times for Branch A and Branch B.

Customer	Recipe A	Recipe B
1	8	9
2	7	6
3	7	7
4	6	8
5	9	7
6	6	7
7	5	9
8	8	9
9	8	10
10	7	9
11	5	9
12	7	6

Table 2: Taste test ratings for Recipe A and Recipe B.