STAB57: An Introduction to Statistics

Shahriar Shams

Week 7 (Test of Hypothesis)

Winter 2023

Recap of Week 6

- Idea of interval estimation using Likelihood func.
- Definition of Confidence Interval (CI)
- CI for parameters of Normal dist
  - CI for μ, (σ^2 known)
  - CI for μ, (σ^2 unknown)
  - CI for σ^2
- MLE based Confidence Intervals
- One-sided Confidence Intervals
- Few definitions related to CI and interpreting CI

Learning goals for this week

- Idea of test of hypothesis and types of hypothesis
- Two approaches:
  - Critical value approach
  - p-value approach
- Type-1, Type-2 error and Power of a test.
- Test of hypothesis using Confidence Interval
- One sided test

These are selected topics from Evans and Rosenthal: chapter 6.3 and John A. Rice: Chap 9.2, 9.3

Section 1

Idea of test of hypothesis and types of hypothesis

Test of hypothesis

- Suppose we are interested in $\psi(\theta)$
- In point and interval estimation we try to guess the value of $\psi(\theta)$ based on the sample observations.
- In test of hypothesis we start with a hypothetical statement like $\psi(\theta) = \psi_0$
- We call this null hypothesis, H_0
- The idea is to check whether our observed data supports H_0 or not.

Test of hypothesis: a numerical example

- Suppose, we are interested in the average income of all Canadians (μ)
- We want to test $H_0 : \mu = \$35,000$
- We collect 10K (representative samples) individuals and get their income data.
- We calculate the sample mean ($\bar{x}$) and here are few scenarios:
  - scenario-1: $\bar{x} = 35,100$
  - scenario-2: $\bar{x} = 35,500$
  - scenario-3: $\bar{x} = 36,000$

    ...
  - scenario-10: $\bar{x} = 50,000$
- In which scenario you will reject H_0?
- In other words: in which scenario the sample mean looks surprising to you if you believe the H_0 to be true?

Null vs. Alternative hypothesis

- **Null Hypothesis, H_0:** the hypothesis that we want to test.
  - For example, in the previous slide we wanted to test whether $\mu = \$35,000$

- **Alternative Hypothesis** (written as H_a): The alternative values of the the parameter of interest
  - Often this is what we are trying to prove as a researcher.
  - For example, we might say
    $H_a : \mu > \$35,000$ or
    $H_a : \mu < \$35,000$ or
    $H_a : \mu \neq \$35,000$ or simply
    $H_a : \mu = \$40,000$

Simple vs. Composite hypothesis

- **Simple hypothesis:** when a hypothesis involves only a single value from the parameter space. e.g. $\mu = \$35,000$
- **Composite hypothesis:** when a hypothesis involves more than one values from the parameter space. e.g. $\mu > \$35,000$ or $\mu \neq \$35,000$
- In practice, often we test simple null against composite alternative hypothesis.

Section 2

Two approaches of hypothesis testing

Significance level

- Due to uncertainty, often we reject H_0 even though it could be true.
- Clearly this is a mistake!
- We assign a (preferably) small predefined probability of making this mistake.
- We call this level of significance and denote it by α

Subsection 1

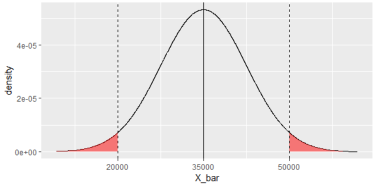
Critical region approach

Test statistic, $T(X)$

- It's a quantity that simultaneously serves few purposes:
  - It summarizes the sample data through an estimator
  - When H_0 is true, it has a known distribution
  - And under that distribution it's possible to find some areas that has probability α
- The pivots that we used in constructing confidence intervals are good examples of test statistic.

Critical region, $R_\alpha(T)$

- A region of the distribution of the test statistic such that we will reject H_0 if $T(X) \in R_\alpha(T)$
- Example: for the numerical example of average income of all Canadians, we can reject the hypothesis $H_0 : \mu = \$35,000$ if $\bar{x} < 20000$ or $\bar{x} > 50000$ (these are made up numbers)



- Here, $\bar{x} < 20000$ and $\bar{x} > 50000$ constitutes the rejection region.
- We need to make sure that

$$P[T(X) \in R_\alpha(T) | H_0 \; true] = \alpha$$

Testing $H_0 : \mu = \mu_0$ when $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ [σ^2 is known]

- Null Hypothesis, $H_0 : \mu = \mu_0$
- Test statistic, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$
- If H_0 is true ie. $\mu = \mu_0$ then $\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$
- Rejection region: $(-\infty, z_{\frac{\alpha}{2}}) \cup (z_{1-\frac{\alpha}{2}}, \infty)$
- We reject H_0 if $\quad \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < z_{\frac{\alpha}{2}} \quad$ or $\quad \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > z_{1-\frac{\alpha}{2}}$

- Intuition: we reject the null hypothesis when the test statistic falls in the lower probability area of the distribution under the null.
- In Naive words: If μ_0 is the true mean then $\bar{X}$ shouldn't be too far from μ_0

Numerical example of critical region approach

Exercise-6.3.1 (E&R):
$(4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3) \overset{iid}{\sim} N(\mu, \sigma_0^2)$ with $\sigma_0^2 = 0.5$
Test $H_0 : \mu = 5$ at level of significance, $\alpha = 0.05$

1. $\bar{x} = \frac{1}{10}(4.7 + 5.5 + ... + 5.3) = 4.88$
2. test statistic, $T(X) = \frac{4.88-5}{\frac{\sqrt{0.5}}{\sqrt{10}}} = -0.537$
3. given level of significance, $\alpha = 0.05$
4. Rejection region, $(-\infty, -1.96) \cup (1.96, \infty)$
5. Since, test statistic value -0.537 does not fall in to the rejection area, we fail to reject H_0

Note: We never say we accept H_0.
We failed to prove that H_0 is wrong $\not\Rightarrow$ H_0 is right!

Other cases...

Testing $H_0 : \mu = \mu_0$; $X_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$ [σ^2 is unknown]

- The frame work remains same with two changes:
  1. Test statistic, $\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{(n-1)}$
  2. Rejection regions are calculated based on a t-distribution

$$R_\alpha(T) = (-\infty, t_{\frac{\alpha}{2}(df=n-1)}) \cup (t_{1-\frac{\alpha}{2}(df=n-1)}, \infty)$$

Exercise-6.3.2 (E&R): (4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3) $\stackrel{iid}{\sim} N(\mu, \sigma^2)$ with both μ and σ^2 unknown
Test $H_0 : \mu = 5$ at level of significance, $\alpha = 0.05$

1. $\bar{x} = 4.88$ and $s = 0.696$
2. Test statistic, $T = \frac{4.88-5}{0.696/\sqrt{10}} = -0.545$
3. Rejection regions$=(-\infty, -2.262) \cup (2.262, \infty)$
4. Fail to reject H_0

Other cases... (cont...)

Testing $H_0 : \sigma^2 = \sigma_0^2$; $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$

- The frame work remains same with two changes:
  1. Test statistic, $\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{(n-1)}$
  2. Rejection regions are calculated based on a χ^2-distribution

$$R_\alpha(T) = (-\infty, \chi^2_{\frac{\alpha}{2}(df=n-1)}) \cup (\chi^2_{1-\frac{\alpha}{2}(df=n-1)}, \infty)$$

Exercise-6.3.2 (E&R): (4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3) $\overset{iid}{\sim} N(\mu, \sigma^2)$ with both μ and σ^2 unknown
Test $H_0 : \sigma^2 = 0.5$ at level of significance, $\alpha = 0.05$
You do it...

Subsection 2

p-value approach

A numeric example first...

$(4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3) \overset{iid}{\sim} N(\mu, \sigma_0^2)$ with $\sigma_0^2 = 0.5$
Test $H_0 : \mu = 5$

- Let's revisit the example we did on slide 15
- α was given to be 0.05
- Let's re calculate the rejection region for some other values of α
  - $\alpha = 0.9 \implies R_\alpha = (-\infty, -0.126) \cup (0.126, \infty) \implies \textit{reject } H_0$
  - $\alpha = 0.8 \implies R_\alpha = (-\infty, -0.253) \cup (0.253, \infty) \implies \textit{reject } H_0$
  - $\alpha = 0.6 \implies R_\alpha = (-\infty, -0.524) \cup (0.524, \infty) \implies \textit{reject } H_0$
  - $\alpha = 0.592 \implies R_\alpha = (-\infty, -0.536) \cup (0.536, \infty) \implies \textit{reject } H_0$
  - $\alpha = 0.5 \implies R_\alpha = (-\infty, -0.674) \cup (0.674, \infty) \implies \textit{fail to reject } H_0$
- 0.592 (approx.) is the smallest α at which H_0 would be rejected.

p-value [Rice p-334]

- **Def 1:** Its the smallest level of significance at which H_0 would be rejected based on the observed data.
- **Def 2:** Its the probability of observing the result as or more extreme than that actually observed if H_0 is true.
- In naive words, p-value suggests how surprising the observed sample is if we assume H_0 to be true.
- Conventionally we compare p-value to 0.01, 0.05 or 0.1
- If p-value is less than a predefined cut-off we reject H_0

Calculating p-value

- **for z-test**

$$2\left[1-\Phi\left(|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}|\right)\right]$$

where Φ is the CDF of a standard normal distribution.

$(4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3) \overset{iid}{\sim} N(\mu, \sigma_0^2)$ with $\sigma_0^2 = 0.5$
Test $H_0 : \mu = 5$
From slide 15, test statistic = -0.537
p-value=$2 * (1 - pnorm(0.537)) \approx 0.5912$

- **for t-test**

$$2\left[1-G\left(|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}|\right)\right]$$

where G is the CDF of a $t_{(n-1)}$ distribution.

From slide 16, test statistic = -0.545
p-value=$2 * (1 - pt(0.545, df = 9)) \approx 0.5989$

Section 3

Type-1, Type-2 error and Power of a test

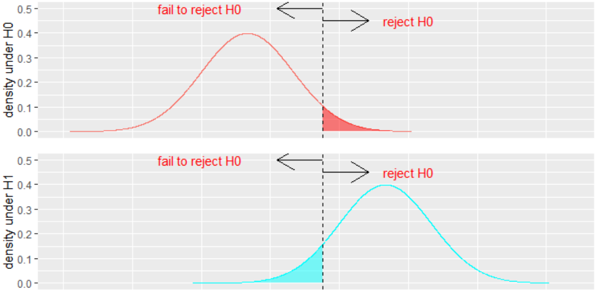
Type-1, Type-2 error and Power of a test

	fail to reject H_0	**reject H_0**
H_0 **true**	Correct decision	type-1 error
H_0 **false**	type-2 error	Correct decision

- **P[Type-1 error]=** $\alpha = P[\text{reject } H_0 | H_0 \text{ true}]$
- **P[Type-2 error]=** $\beta = P[\text{fail to reject } H_0 | H_0 \text{ false}]$
- **Power of a test=** $1 - \beta = P[\text{reject } H_0 | H_0 \text{ false}]$

Type-1 and Type-2 error using graphs

- Suppose we are testing two simple hypotheses:
  $H_0 : \mu = 1 \ vs. \ H_1 : \mu = 4$
- Only one of them can be true (and there are no other options)

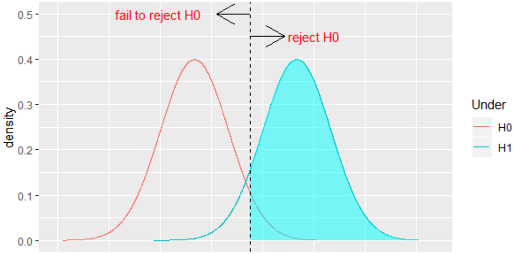


- Type-1 error: The area shaded in red on the left figure
- Type-2 error: The area shaded in cyan on the right figure

Note: For a given sample size, decreasing one type will increase the other!

Power of a test using graph

- Power of a test = 1 - P[type-2 error]



- Power is calculated using the density under H_1
- So in this example, instead of $\mu = 4$, if H_1 changes to $\mu = 5$ we will have a different power.
- When we have a composite H_1 like $\mu \neq 1$, we will have a power function (a function that takes μ as an argument and calculates power for each μ)

Numerical example

Suppose we have $N(\mu, \sigma^2)$ populations with unknown μ and $\sigma = 3$
We want to test $H_0 : \mu = 1 \;\; vs. \;\; H_1 : \mu = 4$ at $\alpha = 0.05$
we decide to take $n = 9$ observations.
Calculate P[type-2 error] and the power.

1. $var[\bar{X}] = \frac{\sigma^2}{n} = \frac{3^2}{9} = 1$
2. Under H_0: $\bar{X} \sim N(1, 1)$
3. Under H_1: $\bar{X} \sim N(4, 1)$
4. $R_\alpha = \bar{X}$ satisfying $\frac{\bar{X}-1}{1} > z_{0.95} \implies \bar{X} > 1 + z_{0.95} \implies \bar{X} > 2.645$
5. Power $= P[\bar{X} > 2.645$ Under the $H_1] \implies P[Z > \frac{2.645-4}{1}] = 0.912$
6. P[type-2 error] $= 1 - 0.912 = 0.088$

Homework: change the H_1, try $\mu = 3, 5, 6, 7$ etc... and calculate the power in each case.

Section 4

Test of hypothesis using Confidence Interval

A simple way of testing hypothesis

- In *week* $-$ 6 we learned how to construct γ-level Confidence intervals.
- We kept the γ part of the distribution and discarded the corners ($1-\gamma$ portion).
- In test of hypothesis, we define the corners as the <u>rejection region</u>
- Intuitively we are doing the same task!
- Let's set $\alpha = 1 - \gamma$
- Constructing a γ level confidence interval for μ and checking whether μ_0 is inside or not is equivalent of testing the hypothesis of $\mu = \mu_0$ at $(1-\gamma)$ level of significance.

A numeric example

- Last week (slide-16), we calculated the 95% CI for μ as (4.442,5.318)
- If we want to test $\mu = 5$ at $\alpha = 5\%$, we would fail to reject the hypothesis since 5 is inside the interval.
- Which is the same conclusion we reached this week (on slide-15)

Section 5

One sided test

One sided test (E&R page 337)

- When testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, we define our rejection region on both sides.
- When testing $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$, intuitively we define our rejection region on the right side only.
- Similarly, when testing $H_0 : \mu = \mu_0$ against $H_1 : \mu < \mu_0$, we define our rejection region on the left side only.

One sided p-value (for Z-test)

- On slide 21, we calculated p-value keeping in mind the two sides of the rejection region.
- When testing $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$,

$$p - value = 1 - \Phi\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)$$

- When testing $H_0 : \mu = \mu_0$ against $H_1 : \mu < \mu_0$,

$$p - value = \Phi\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)$$

- Similar idea for t, χ^2 or other tests.

One sided test using one sided CI

I will leave it for you to figure this out...

hint: having a α level rejection region on the right, is same as constructing $(1-\alpha)$ level left sided CI and vice versa

Section 6

Testing using large sample property of MLE

Question: 1

Can we construct a test for testing $H_0 : \theta = \theta_0$ using the fact that

$$\frac{\hat{\theta} - \theta_0}{\sqrt{1/nI(\theta_0)}} \xrightarrow{D} N(0, 1)$$

Question: 2

Can we construct a test using the variable $S(\theta_0)$ (score evaluated at θ_0)

- What is the distribution of this variable under $H_0 : \theta = \theta_0$
- What is the mean?
- What is the variance?

We will learn more on these two ideas along with Likelihood ratio test on week-9

Discussion

Statistical significance vs. Practical Significance

Homework (Non-credit)

Evans and Rosenthal

Exercise: 6.3.1-6.3.6, 6.3.11, 6.3.14

John A. Rice

Exercise 9: 1, 3, 5, 9