Math 170S

Lecture Notes Section 7.1 *†

Confidence intervals for one mean

Instructor: Swee Hong Chan

NOTE: Materials that appear in the textbook but do not appear in the lecture notes might still be tested. Please send me an email if you find typos.

*Version date: Wednesday 28th October, 2020, 11:19.

†This notes is based on Hanbaek Lyu's and Liza Rebrova's notes from the previous quarter, and I would like to thank them for their generosity. "*Nanos gigantum humeris insidentes* (I am but a dwarf standing on the shoulders of giants)".

1 Funny story

The friendly instructor (FI) were hired by the casino to check if a player had cheated in a coin-flipping game. During the investigation, the FI flipped the coin 20 times; 5 of them Heads, and 15 of them Tails. So the FI declared,

"I estimate that the success probability for this coin is $\widehat{p} = \frac{5}{20} = 0.25$."

The casino therefore filed a lawsuit against the player. At the court, the friendly instructor gave the following expert testimony:

"The success probability of the coin is between 0.04 and 0.46 with confidence of at least 95%. There is therefore at least a chance of 95% that the player had tampered with the coin.

2 Setting of the problem

- **Object:** X is **any** random variable with **unknown mean** μ.
- **Input:** Sample values $X_1, \ldots, X_n$ for X. Note that $X_1, \ldots, X_n$ are random variables here.
- **Output:** A range of values that μ can take, with high probability.

The output is usually phrased as

"If I repeat the experiment 100 times, we would have that (approximately) 95 of the times the difference between $\overline{\mathrm{X}}$ and μ is less than 5".

The sentence above is too long; we can rephrase the sentence above mathematically as

$$P\left[\,\overline{\mathrm{X}}-5 \ \leq\ \mu\ \leq\ \overline{\mathrm{X}}+5\,\right] \approx 0.95.$$

Now the sentence above is too technical. We thus compromise by using the semi-mathematical sentence:

"μ is contained in the interval $[\overline{\mathrm{X}}-5, \overline{\mathrm{X}}+5]$ with 95% confidence."

3 Confidence interval

For any positive real number ε and α, the sentence

"μ is contained in the interval $[\overline{\mathrm{X}}-\varepsilon, \overline{\mathrm{X}}+\varepsilon]$ with confidence $1-\alpha$"

means

$$P\left[\,\overline{\mathrm{X}}-\varepsilon \;\le\; \mu \;\le\; \overline{\mathrm{X}}+\varepsilon\,\right] \;\approx\; 1-\alpha.$$

The interval $[\overline{\mathrm{X}}-\varepsilon, \overline{\mathrm{X}}+\varepsilon]$ is called the **confidence interval**, and $1-\alpha$ is called the **confidence coefficient**.

Note that ε and α depends on each other. The smaller ε is, the larger α is. In particular

- α is equal to 0 when ε is equal to ∞:

  "μ is contained in the interval $[\overline{\mathrm{X}} - \infty, \overline{\mathrm{X}} + \infty]$ with confidence 1".

- α is equal to 1 when ε is equal to 0 for continuous RVs:

  "μ is contained in the interval $[\overline{\mathrm{X}} - 0, \overline{\mathrm{X}} + 0]$ with confidence 0".

4 Confidence interval: Case I

Object: X is a random variable with **unknown mean** μ but with **known variance** σ^2.

Given: Confidence constant $1-\alpha$.

Question: Compute ε.

Solution: ε can be computed by the formula:

$$\varepsilon \;=\; z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}, \tag{1}$$

where $z_{\alpha/2}$ is the real number such that

$$P[N(0,1) \geq z_{\alpha/2}] \;=\; \alpha/2.$$

The value $z_{\alpha/2}$ can be computed from the Table V in Appendix B. (Textbook Time)

5 Confidence interval: Example

Let X be the length of life of a 60-watt light bulb made by Friendly Instructor Company.

Suppose that X is a normal random variable with mean μ and variance 1296.

If a random sample of $n = 25$ bulbs is tested until they burn out, yielding sample mean of $\overline{\mathrm{x}} = 1478$ hours, compute the 60% confidence interval.

Answer:

Given: α is 0.4 since it is for 60% confidence.

$$\varepsilon \;=\; z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} \;=\; z_{0.2}\frac{\sqrt{1296}}{\sqrt{25}} \;=\; 0.842\frac{36}{5} \;=\; 6.0624\,.$$

The confidence interval is then

$$\begin{aligned}\left[\,\overline{\mathrm{x}}-\varepsilon\,,\,\overline{\mathrm{x}}+\varepsilon\right] \;&=\left[1478-6.0624\,,\,1478+6.0624\right]\\ &=\left[1471.9376\;,\;1484.0624\right].\end{aligned}$$

Remark 1. To be confident of the confidence interval, we need n to be fairly large, and $n \geq 30$ is usually sufficient for all purposes. When we know that X is a nice random variable (e.g., normal), $n \geq 5$ is sufficient.

Remark 2. The formula for ε in (3) comes from the central limit theorem (from Math 170E), see the textbook for the detailed proof.

6 Confidence interval: Case II

Object: X is a **normal** random variable with **unknown mean** μ and **unknown variance**.

Given: Confidence constant $1 - \alpha$.

Question: Compute ε.

7 The error ε: Case II-a

Assumption: $n - 1 \leq 30$.

Solution: Compute the sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{\mathrm{x}})^2.$$

Then ε can be computed by

$$\varepsilon \;=\; t_{\alpha/2}(n-1)\, \frac{s}{\sqrt{n}}, \tag{2}$$

where $t_{\alpha/2}(n-1)$ is the real number such that

$$P[T \geq t_{\alpha/2}(n-1)] \;=\; \alpha/2,$$

where T is the student's t distribution with $n-1$ degrees of freedom, and can be computed from Table VI in Appendix B. (Textbook time)

8 The error ε: Case II-b

Assumption: $n - 1 > 30$.

Solution: ε can be computed by the formula:

$$\varepsilon \;=\; z_{\alpha/2}\,\frac{s}{\sqrt{n}}, \tag{3}$$

where s^2 is the sample variance and $z_{\alpha/2}$ is the real number such that

$$P[N(0,1) \geq z_{\alpha/2}] \;=\; \alpha/2.$$

9 Confidence interval: Example

Let X be the amount of butterfat (in pounds) produced by a typical cow in Mikage farm.

Assume that X is a normal random variable with unknown mean μ and unknown variance σ^2.

The following is butter fat production for $n = 20$ cows:

$$
\begin{array}{cccccccccc}
481 & 537 & 513 & 583 & 453 & 510 & 570 & 500 & 457 & 555 \\
618 & 327 & 350 & 643 & 499 & 421 & 505 & 637 & 599 & 392.
\end{array}
$$

Give a 90% confidence interval for μ.

Answer: We first compute the sample mean $\overline{\mathrm{x}}$ and sample variance s^2 (Exercise time)

$$\overline{\mathrm{x}} = 507.50; \qquad s = 89.75; \qquad \alpha = 0.10\,.$$

So our estimate for μ is the sample mean 507.50. Then

$$t_{\alpha/2}(n-1) \;=\; t_{0.05}(19) \;=\; 1.729\,.$$

The constant ε is then given by

$$\varepsilon \;=\; t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} \;=\; 1.729\,\frac{89.75}{\sqrt{20}} \;\approx\; 34.70. \qquad (4)$$

So our confidence interval is

$$\begin{aligned}[\overline{\mathrm{x}}-\varepsilon\,,\,\overline{\mathrm{x}}+\varepsilon] \;&= [507.50 - 34.70\,,\, 507.50 + 34.70] \\ &= [472.80\,,\, 542.20]\,.\end{aligned}$$

Remark 3. The formula for ε is derived using student's t distribution, see the textbook for the proof.

Remark 4. For this unknown variance case, X **must be** normal RVs. The same conclusion cannot be made for general RVs.

Remark 5. For the approximation to be decent, we need $n \geq 50$ even for specialized purpose.

10 One-sided confidence interval: lower bound

One-sided confidence interval can come in two flavors, lower and upper bound.

For the lower bound, we say

"μ is greater than $\overline{\mathrm{X}} - \varepsilon$ with confidence $1 - \alpha$"

to mean

$$P[\mu \geq \overline{\mathrm{X}} - \varepsilon] \approx 1 - \alpha.$$

In this case, the **one-sided confidence interval** is $[\overline{\mathrm{X}} - \varepsilon, \infty)$.

11 One-sided confidence interval: upper bound

For the upper bound, we say

"μ is less than $\overline{\mathrm{X}} + \varepsilon$ with confidence $1 - \alpha$"

to mean

$$P[\mu \leq \overline{\mathrm{X}} + \varepsilon] \approx 1 - \alpha.$$

In this case, the **one-sided confidence interval** is $(-\infty, \overline{\mathrm{X}} + \varepsilon]$.

12 One-sided confidence interval: Formulas

Object: X is **any** random variable with **unknown** mean μ and **known** variance σ^2

Formula: ε is given by

$$\varepsilon = z_\alpha \, \frac{\sigma}{\sqrt{n}}.$$

Pay attention: $z_{\alpha/2}$ for two sided confidence intervals is replaced with z_α here.

Object: X is a **normal** random variable with **unknown** mean μ and **unknown** variance σ^2

Formula: ε is given by

$$\varepsilon = \begin{cases} t_\alpha(n-1)\,\frac{s}{\sqrt{n}}; & \text{if } n-1 \le 30; \\ z_\alpha\,\frac{s}{\sqrt{n}}; & \text{if } n-1 > 30. \end{cases}$$

Pay attention: $t_{\alpha/2}(n-1)$ for two sided confidence intervals is replaced with $t_\alpha(n-1)$. The same with $z_{\alpha/2}$ and z_α.

13 One-sided confidence interval: Example

Let X be the length of life of a 60-watt light bulb made by Friendly Instructor Company.

Suppose that X is a normal random variable with mean μ and variance 1296.

If a random sample of $n = 25$ bulbs is tested until they burn out, yielding sample mean of 1478 hours, compute the lower bound for μ with 60% confidence interval.

Answer: The variance is **known**, so

$$\varepsilon = z_\alpha \frac{\sigma}{\sqrt{n}} \;=\; z_{0.4}\,\frac{\sqrt{1296}}{\sqrt{25}} \;=\; (0.253)\frac{36}{5} \;=\; 1.8216,$$

and the one-sided confidence interval is

$$\left[\bar{\mathrm{x}} - \varepsilon, +\infty\right) \;=\; \left[1478 - 1.8216, +\infty\right) \;=\; \left[1476.1784\,,\; +\infty\right).$$

14 One-sided confidence interval: Example

Let us return to the Mikage farm example.

Give an upper bound for μ and a 90% confidence interval for μ.

Answer: Mikage farm example is **normal** with **unknown** variance.

Given: $\bar{x} = 507.50$, $s = 89.75$, $\alpha = 0.1$.

So

$$t_\alpha(n-1) \; = \; t_{0.10}(19) \; = \; 1.328 .$$

The constant ε is then given by

$$\varepsilon \; = \; t_\alpha(n-1)\frac{s}{\sqrt{n}} \; = \; (1.328)\,\frac{89.75}{\sqrt{20}} \; \approx \; 26.65.$$

The one-sided confidence interval for **upper bound** is

$$(-\infty, \bar{x}+\varepsilon] \; = \; (-\infty, 507.50+26.65] \; = \; (-\infty, 534.15],$$

as desired.