More on Confidence Intervals

Yifan Zhu

Iowa State University

Outline

$n \geq 25$, σ unknown

- The formula for a 2-sided, $1-\alpha$ CI for a true mean μ from before was:

$$(\overline{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$$

- If $n \geq 25$ and σ is unknown, you can replace σ in the confidence interval formula with the sample standard deviation, $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$.
- The formula for the 2-sided confidence interval becomes:

$$(\overline{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}},\ \overline{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}})$$

- The analogous upper and lower confidence intervals, respectively, are:

$$(-\infty,\ \overline{x} + z_{1-\alpha}\frac{s}{\sqrt{n}})$$

$$(\overline{x} - z_{1-\alpha}\frac{s}{\sqrt{n}},\ \infty)$$

Your turn: breaking strength of wire

- Suppose you are a manufacturer of construction equipment. You make 0.0125 inch wire rope and need to determine how much weight it can hold before breaking so that you can label it clearly.
- Here are breaking strengths, in kg, for 40 sample wires:

100.37	96.31	72.57	88.02	105.89	107.80	75.84	92.73	67.47
94.87	122.04	115.12	95.24	119.75	114.83	101.79	80.90	96.10
118.51	109.66	88.07	56.29	86.50	57.62	74.70	92.53	86.25
82.56	97.96	94.92	62.93	98.44	119.37	103.70	72.40	71.29
107.24	64.82	93.51	86.97					

- The sample mean breaking strength is 91.85 kg and the sample standard deviation is 17.79 kg.
- Using the appropriate 95% confidence interval, try to determine whether the breaking strengths is meet the requirement of 85 kg.

Answers: breaking strength of wire

- Since we want the breaking strengths to be above 85 kg, I choose a lower confidence interval (one with a lower bound).
- $\alpha = 1 - 0.95 = 0.05$, $\overline{x} = 91.85$, $s = 17.79$, and $n = 40$.

$$
\begin{aligned}
&(\overline{x} - z_{1-\alpha}\frac{s}{\sqrt{n}},\ \infty) \\
&\quad = \left(91.85 - z_{1-0.05}\frac{17.79}{\sqrt{40}},\ \infty\right) \\
&\quad = (91.85 - z_{0.95} \cdot 2.81,\ \infty) \\
&\quad = (91.85 - 1.64 \cdot 2.81,\ \infty) \\
&\quad = (87.24,\ \infty)
\end{aligned}
$$

- With 95% confidence, we have shown that the true mean breaking strength is above 87.24 kg. Hence, we meet the 85 kg requirement with 95% confidence.
- What is the maximum confidence level with which we can meet the 85kg requirement?

Answers: breaking strength of wire

- The confidence interval is:

$$(91.85 - z_{1-\alpha} \cdot 2.81,\ \infty)$$

- To meet the requirement, we need

$$\begin{aligned} 91.85 - z_{1-\alpha} \cdot 2.81 &< 85 \\ z_{1-\alpha} &< \frac{6.85}{2.81} \\ &= 2.44 \\ \Phi(z_{1-\alpha}) &< \Phi(2.44) \\ 1 - \alpha &< 0.9926 \end{aligned}$$

- Hence, we could have raised the confidence level up to 99.26 % and still shown that we met the requirement.
- (In hypothesis testing, which will come later, $1 - 0.9926 = 0.0074$ will be a **p-value**.)
- Now, calculate and interpret a 95%, 2-sided confidence interval for the true mean breaking strength. Is there any reason to *disbelieve* that the true mean breaking strength is 94 kg?

Answers: breaking strength of wire

- The two-sided confidence interval is:

$$
\begin{aligned}
&(\overline{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}},\ \overline{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}) \\
&= \left(91.85 - z_{1-0.05/2}\frac{17.79}{\sqrt{40}},\ 91.85 + z_{1-0.05/2}\frac{17.79}{\sqrt{40}}\right) \\
&= (91.85 - z_{0.975}\cdot 2.81,\ 91.85 + z_{0.975}\cdot 2.81) \\
&= (91.85 - 1.96\cdot 2.81,\ 91.85 + 1.96\cdot 2.81) \\
&= (86.34, 97.36)
\end{aligned}
$$

- With 95% confidence, the true mean breaking strength is between 86.34 kg and 97.36 kg.
- Since 94 kg is in the interval, at $\alpha = 0.05$, we have no evidence to dispute the claim that the true mean breaking strength is 94 kg.

Outline

$n < 25, \ \sigma$ unknown

- We need to assume that $X_1, \ldots, X_n$ are not only iid with mean μ and variance σ^2, but also that these random variables are *normally distributed*.
  - We can't use the Central Limit Theorem since $n < 25$.
  - However, the additional assumption makes $\overline{X}$ normally distributed (A linear combination of *independent* normal random variables is normal.)
- We need to use the $t_{n-1,1-\alpha/2}$ instead of $z_{1-\alpha/2}$ in the confidence intervals
  - Although $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, it's a fact that $\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$ since s is random ($(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$).
  - For $n < 25$, the t_{n-1} distribution is *not* close enough to N(0,1).

New confidence interval formulas for $n < 25$, σ unknown

- Two-sided $1-\alpha$ CI:

$$(\overline{x} - t_{n-1,\ 1-\alpha/2}\frac{s}{\sqrt{n}},\ \overline{x} + t_{n-1,\ 1-\alpha/2}\frac{s}{\sqrt{n}})$$

- One-sided lower $1-\alpha$ CI:

$$(\overline{x} - t_{n-1,\ 1-\alpha}\frac{s}{\sqrt{n}},\ \infty)$$

- One-sided upper $1-\alpha$ CI:

$$(-\infty,\ \overline{x} + t_{n-1,\ 1-\alpha}\frac{s}{\sqrt{n}})$$

Your turn: concrete beams

- 10 concrete beams were each measured for flexural strength (MPa):

8.2	8.7	7.8	9.7	7.4
7.8	7.7	11.6	11.3	11.8

- Assuming the flexural strengths are iid normal, calculate and interpret a two-sided 99% CI for the flexural strength of the beams.
- Is the true mean flexural strength below the minimum requirement of 11 MPa? Find out with the appropriate 95% CI.

Answers: concrete beams

- $n = 10,\ \alpha = 0.01$
- $\overline{x} = \frac{1}{10}(8.2 + 8.7 + \cdots + 11.8) = 9.2$
- $s = \sqrt{\frac{1}{10-1}[(8.7-9.2)^2 + (8.7-9.2)^2 + \cdots + (11.8-9.2)^2]} = 1.76$
- The two-sided 99% CI is:

$$
\begin{aligned}
&(\overline{x} - t_{n-1,\ 1-\alpha/2}\frac{s}{\sqrt{n}},\ \overline{x} + t_{n-1,\ 1-\alpha/2}\frac{s}{\sqrt{n}}) \\
&= \left(9.2 - t_{10-1,\ 1-0.01/2}\frac{1.76}{\sqrt{10}},\ 9.2 + t_{10-1,\ 1-0.01/2}\frac{1.76}{\sqrt{10}}\right) \\
&= (9.2 - t_{9,0.995} \cdot 0.556,\ 9.2 + t_{9,0.995} \cdot 0.556) \\
&= (9.2 - 3.250 \cdot 0.556,\ 9.2 + 3.250 \cdot 0.556) \\
&= (7.393,\ 11.007)
\end{aligned}
$$

- We're 99% confident that the true flexural strength of this kind of concrete beam is between 7.393 MPa and 11.007 MPa.

Answers: concrete beams

- I want to know whether the true mean flexural strength is *below* 11 MPa. Hence, I need an *upper* 95% confidence interval (i.e., with an upper bound).

$$\begin{aligned}
(-\infty,\ \overline{x} + t_{n-1,\ 1-\alpha}\frac{s}{\sqrt{n}}) \\
&= (-\infty,\ 9.2 + t_{9,\ 1-0.05}\frac{1.76}{\sqrt{10}}) \\
&= (-\infty,\ 9.2 + t_{9,0.95} \cdot 0.556) \\
&= (-\infty,\ 9.2 + 1.83 \cdot 0.556) \\
&= (-\infty,\ 10.21)
\end{aligned}$$

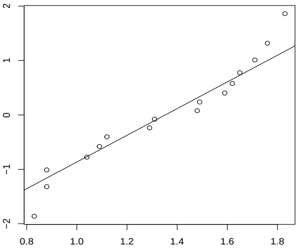
- We're 95% confident that the true mean flexural strength is below 10.21 MPa. That's below 11 MPa. So at $\alpha = 0.05$, we have shown that the true mean flexural strength is also below 11 MPa and the requirement is not met.

Your turn: paint thickness

- Consider the following sample of observations on coating thickness for low-viscosity paint:

0.83	0.88	0.88	1.04	1.09	1.12	1.29	1.31
1.48	1.49	1.59	1.62	1.65	1.71	1.76	1.83

- A normal QQ plot shows that they are close enough to normally distributed.



- Calculate and interpret a two-sided 90% confidence interval for the true mean thickness.

Answers: paint thickness

- $n = 16,\ \alpha = 0.1$
- $\overline{x} = \frac{1}{16}(0.83 + 0.88 + \cdots + 1.83) = 1.35$ mm
- $s = \sqrt{\frac{1}{16-1}[(0.83-1.35)^2 + (0.88-1.35)^2 + \cdots + (1.83-1.35)^2]} = 0.34$ mm
- The two-sided 90% CI is:

$$\begin{aligned}
&(\overline{x} - t_{n-1,\ 1-\alpha/2}\frac{s}{\sqrt{n}},\ \overline{x} + t_{n-1,\ 1-\alpha/2}\frac{s}{\sqrt{n}}) \\
&= \left(1.35 - t_{10-1,\ 1-0.1/2}\frac{0.34}{\sqrt{16}},\ 1.35 + t_{16-1,\ 1-0.1/2}\frac{0.34}{\sqrt{16}}\right) \\
&= (1.35 - t_{15,0.95} \cdot 0.085,\ 1.35 + t_{15,0.95} \cdot 0.085) \\
&= (1.35 - 1.75 \cdot 0.085,\ 1.35 + 1.75 \cdot 0.085) \\
&= (1.201,\ 1.499)
\end{aligned}$$

- We're 90% confident that the true mean thickness is between 1.201 mm and 1.499 mm.

Outline

$n \geq 25$, σ unknown

$n < 25$, σ unknown

Hypothesis Testing with Confidence Intervals

Statistical inference

- **Statistical inference**: using data from the sample to draw conclusions about the population
  - Point estimation (confidence intervals): estimating population parameters and specifying the degree of precision of the estimate.
  - Hypothesis testing: testing the validity of statements about the population that are framed in terms of parameters.

Hypothesis testing

- **Hypothesis testing (significance testing)**: the use of data in the quantitative assessment of the plausibility of some trial value or a parameter.
- You have competing **hypotheses**, or statements, about a population:
  - The **null hypothesis**, denoted H_0 is the proposition that a parameter equals some fixed number.
  - The **alternative hypothesis**, denoted H_a or H_1, is a statement that stands in opposition to the null hypothesis.
  - Examples:

$$\mathrm{H_0}: \mu = \# \qquad \mathrm{H_0}: \mu = \# \qquad \mathrm{H_0}: \mu = \#$$

$$\mathrm{H_a}: \mu > \# \qquad \mathrm{H_a}: \mu < \# \qquad \mathrm{H_a}: \mu \neq \#$$

- Note: $H_a : \mu \neq \#$ makes a **two-sided test**, while $H_a : \mu < \#$ and $H_a : \mu > \#$ make a **one-sided test**.
- The goal is to use the data to debunk the null hypothesis in favor of the alternative:
  - Assume H_0.
  - Try to show that, under H_0, the data are preposterous.
  - If the data are preposterous, reject H_0 and conclude H_a.

Hypothesis testing

- Outcomes of a hypothesis test:

The ultimate decision is in favor of:

The true state of affairs is described by:	H_0	H_a
H_0		Type I error
H_a	Type II error	

- α (the very same α in confidence intervals) is the probability of rejecting H_0 when H_0 is true.
  - α is the Type I Error probability.
  - For honesty's sake, α is fixed before you even *look* at the data.

3 Methods of Hypothesis Testing

1. Confidence intervals
2. Critical values
3. P-values

Formal steps of a hypothesis test using confidence intervals

1. State the hypotheses, H_0 and H_a.
2. State the significance level, α.
3. State the form of the $1 - \alpha$ confidence interval you will use, along with all the assumptions necessary.
   - The confidence interval should contain μ when there is little to no evidence against H_0 and should *not* contain μ when there is strong evidence against H_0.
   - Use one-sided confidence intervals for one-sided tests (i.e., $H_a : \mu < \#$ or $\mu > \#$) and two-sided intervals for two-sided tests ($H_a : \mu \neq \#$).
4. Calculate the $1 - \alpha$ confidence interval.
5. Based on the $1 - \alpha$ confidence interval, either:
   - Reject H_0 and conclude H_a, or
   - Fail to reject H_0.
6. Interpret the conclusion using layman's terms.

Example: breaking strength of wire

- Suppose you are a manufacturer of construction equipment. You make 0.0125 inch wire rope and need to determine how much weight it can hold before breaking so that you can label it clearly.
- Here are breaking strengths, in kg, for 40 sample wires:

100.37	96.31	72.57	88.02	105.89	107.80	75.84	92.73	67.47
94.87	122.04	115.12	95.24	119.75	114.83	101.79	80.90	96.10
118.51	109.66	88.07	56.29	86.50	57.62	74.70	92.53	86.25
82.56	97.96	94.92	62.93	98.44	119.37	103.70	72.40	71.29
107.24	64.82	93.51	86.97					

- Let's conduct a hypothesis test to find out if the true mean breaking strength is above 85 kg.

Example: breaking strength of wire

1. $H_0 : \mu = 85$ kg and $H_a : \mu > 85$ kg, where μ is the true mean breaking strength.
2. $\alpha = 0.05$
3. Since this is a one-sided (lower) test, I will use a lower $1 - \alpha$ confidence interval:

$$\left(\overline{x} - z_{1-\alpha}\frac{s}{\sqrt{n}},\ \infty\right)$$

   I am assuming:

   - The data points $x_1, \ldots x_n$ were iid draws from some distribution with mean μ and some constant variance.
4. From before, we calculated the confidence interval to be $(87.24, \infty)$.
5. With 95% confidence, we have shown that $\mu > 87.24$. Hence, at significance level $\alpha = 0.05$, we have shown that $\mu > 85$. We reject H_0 and conclude H_a.
6. There is enough evidence to conclude that the true mean breaking strength of the wire is greater than 85 kg. Hence, the requirement is met.

Example: concrete beams

- 10 concrete beams were each measured for flexural strength (MPa):

8.2	8.7	7.8	9.7	7.4
7.8	7.7	11.6	11.3	11.8

- At $\alpha = 0.01$, I will test the hypothesis that the true mean flexural strength is 10 MPa.

Example: concrete beams

1. $H_0 : \mu = 10MPa$, $H_a : \mu \neq 10MPa$, where μ is the true mean flexural strength of the beams.
2. $\alpha = 0.01$.
3. Since this is a two-sided test, I will use a two-sided confidence interval:

$$\left(\overline{x} - t_{n-1,\ 1-\alpha}\frac{s}{\sqrt{n}},\ \overline{x} + t_{n-1,\ 1-\alpha}\frac{s}{\sqrt{n}}\right)$$

   I am assuming the data points $x_1, \ldots x_n$ were independently drawn from $N(\mu, \sigma^2)$.
4. From before, we calculated the confidence interval to be (7.393, 11.007).
5. Since 10 MPa is in the interval, we fail to reject H_0.
6. There is not enough evidence to conclude that the true mean flexural strength is different from 10 MPa.

Your turn: paint thickness

- Consider the following sample of observations on coating thickness for low-viscosity paint:

0.83	0.88	0.88	1.04	1.09	1.12	1.29	1.31
1.48	1.49	1.59	1.62	1.65	1.71	1.76	1.83

- Using $\alpha = 0.1$, test the hypothesis that the true mean paint thickness is 1.00 mm.
- Note: the 90% confidence interval for the true mean paint thickness was calculated from before as (1.201, 1.499).

Answers: paint thickness

1. $H_0 : \mu = 1.00$, $H_a : \mu \neq 1.00mm$, where μ is the true mean paint thickness.
2. $\alpha = 0.1$.
3. Since this is a two-sided test, I will use a two-sided confidence interval:

$$\left(\overline{x} - t_{n-1,\ 1-\alpha} \frac{s}{\sqrt{n}},\ \overline{x} + t_{n-1,\ 1-\alpha} \frac{s}{\sqrt{n}} \right)$$

   I am assuming the data points $x_1, \dots x_n$ were independently drawn from $N(\mu, \sigma^2)$.
4. From before, we calculated the confidence interval to be (1.201, 1.499).
5. Since 1.00 mm is not in the interval, we reject H_0 and conclude H_a.
6. There is enough evidence to conclude that the true mean paint thickness is not 1.00 mm.

Next time

1. Review of hypothesis testing with confidence intervals.
2. Hypothesis testing with critical values.
3. Hypothesis testing with p-values.