Chapter 8: Confidence intervals for a single sample

Lianming Wang

University of South Carolina

January 4, 2023

Target parameter is the parameter of interest in the study.

- μ: mean, average
- σ^2: variance
- p: proportion, percentage, fraction, rate

- An **interval estimator** (or **confidence interval**) is a formula that tells us how to use sample data to calculate an interval that estimates a population parameter.
- The **confidence coefficient** is the probability that an interval estimator encloses the population parameter – that is, the relative frequency with which the interval estimator encloses the population parameter when the estimator is used repeatedly a very large number of times.
- The **confidence level** is the confidence coefficient expressed as a percentage.

Section 8.1. Confidence interval for μ, for a normal distribution with known variance

Assume that $X_1, \cdots, X_n$ are an i.i.d. sample from $N(\mu, \sigma^2)$ with σ^2 known.

- The $100(1-\alpha)\%$ confidence interval for μ

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- $Z_{\alpha/2}$ is the upper $\alpha/2$ quantile, to the right of which the area under the standard normal density curve is $\alpha/2$.
- Commonly used $Z_{\alpha/2}$ in constructing confidence intervals

$100(1-\alpha)\%$	α	$Z_{\alpha/2}$
90%	.10	1.645
95%	.05	1.96
99%	.01	2.576

Exercise 1.

The diameter of holes for a bable harness is known to have a normal distribution with $\sigma = 0.01$ inch. A random sample of size 10 yields an average diameter 1.5045 inch. Find a 95% confidence interval for the true mean hole diameter.

Interpretation of a confidence interval

- A confidence interval is a random interval. It changes with random samples.
- For a specific confidence interval, the interval either contains the true value or does not contain the true value.
- The appropriate statement is that the interval $\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ brackets the true value μ with confidence $100(1-\alpha)\%$.
- **A frequency interpretation**: if the same rule is used to constuct confidence intervals for μ for a large number of random samples, about $100(1-\alpha)\%$ of the resulting confidence intervals will contain the true value μ.
- The confidence is based on the whole procedure and it is not for a specific confidence interval.

Sample size determination

- It is observed that the large sample size, the smaller length of the confidence interval.
- $B = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ is called the margin error of a confidence interval.
- If one uses $\bar{x}$ to estimate μ, the estimation error $|\bar{x} - \mu|$ is within B with confidence $100(1-\alpha)\%$.
- One can choose the minimum sample size to control the margin error at a desired level.
- The minimum sample size is

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{B}\right)^2.$$

Exercise 2.

The diameter of holes for a bable harness is known to have a normal distribution with $\sigma = 0.01$ inch. What is the minimum sample size needed in order to estimate the true mean diameter within 0.005 inch at 95% confidence?

Section 8.2. Confidence interval for μ for a large sample

What if the normality assumption does not hold or the variance σ^2 is unknonwn?

Consider a large sample (with sample size $n \geq 30$),

- When σ is known, the $100(1-\alpha)\%$ confidence interval for μ

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- When σ is unknown, the $100(1-\alpha)\%$ confidence interval for μ

$$\bar{x} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}}$$

- The construction is based on the central limit theorem.

Exercise 3.

A chemical company wishes to estimate the mean number of man-hours lost per week due to employee illness. Suppose the company takes a random sample of 40 employees and obtains a sample mean of 3.8 hours and standard deviation of 0.42 hours. Form a 90% confidence interval for the true mean number of man-hours lost per week.

Exercise 4.

To estimate the average service time at a fast food restaurant, a random sample of 35 orders were found to take an average of 72.2 seconds to complete with a standard deviation of 12.8 seconds. Find a 95% confidence interval for the true mean time to complete an order.

Section 8.3. Confidence interval for μ for a small sample

- The assumption of a known variance σ^2 is often unrealistic in real-life applications.
- Need the normality assumption: $(X_1, \cdots, X_n)$ is a iid sample from a population with $N(\mu, \sigma^2)$.
- $100(1-\alpha)\%$ confidence interval for μ when σ is unknown

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

- $t_{\alpha/2,n-1}$ is the cutoff point above which the area under the curve of t distribution with the degree of freedom $n-1$ is $\alpha/2$.

Find critical values of t distribution.

(a) $P(t > t_{.025}) = .025$, where df=10.

(b) $P(t > t_{.01}) = .01$, where df=17.

(c) $P(t > t_{.005}) = .005$, where df=6.

(d) $P(t > t_{.05}) = .05$, where df=13.

Exercise 5.

A research engineer for a tire manufacturer is investigate tire life for a new rubber compound and has built 16 tires and tested them to end-of-life in a road test. The sample mean and standard deviation are 60,139.7 kilometers and 3645.94 kilometers, respectively. Assume that the lifetimes of these tires have a normal distribution.

(a) Find a 90% confidence interval for the true mean tire life.

(b) Find a 95% confidence interval for the true mean tire life.

(c) Find a 99% confidence interval for the true mean tire life.

Exercise 6.

In a pollution study, a random sample of eight air specimens contained an average of 2.26 microgram$/m^3$ of suspended matter with a standard deviation of 0.56 microgram$/m^3$. Find a 99% confidence interval for the mean amount of suspended matter in a 1 m^3 air sample in this area. (Assume the amount of suspended matter is normally distributed.)

Section 8.4 Large-Sample Confidence Interval for a Population Proportion p

- Point estimate $\hat{p} = \frac{X}{n}$, where X is the number of events of interest and n is the total number of events.
- $\hat{p}$ is an unbiased estimator of p.
- The standard deviation of $\hat{p}$ is $\sqrt{\frac{p(1-p)}{n}}$.
- When the sample size n is large (the sample size is large enough if both $n\hat{p} \geq 15$ and $n(1-\hat{p}) \geq 15$), then the sampling distribution of $\hat{p}$ is approximately normal.
- For a large random sample, a $100(1-\alpha)\%$ confidence interval for p is

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

Exercise 7.

A random sample of 400 computer chips is taken from a large lot of chips and 50 of them are found to be defective.

(a) Find a 95% confidence interval for the true proportion of defective chips contained in the lot.

(b) The manager of the lot claims that the true defective rate is only 10%. What will you comment on the manager's claim based on your result in (a)? .

Sample size determination

- The margin error is $B = Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$.
- Solving the equation for n yields

$$n = \left(\frac{Z_{\alpha/2}}{B}\right)^2 p(1-p).$$

- This formula contains unkonwn parameter p. Two reasonable solutions:
  - Use a rough guess or an estimate p_0: $n = \left(\frac{Z_{\alpha/2}}{B}\right)^2 p_0(1-p_0)$.
  - Use the conversative conversion: $n = \frac{1}{4}\left(\frac{Z_{\alpha/2}}{B}\right)^2$.

Exercise 8.

A random sample of 50 suspension helmet used by motorcycle riders and automobile race-car drivers was subjected to an impact test, and on 18 of these helmets some damage was observed.

(a) Find a 95% confidence interval for the true proportion of helmets of this type that would show damage from this test.

(b) Using the point estimate of p obtained from the prelimenary sample of 50 helmets, how many helmets must be tested to be 95% confidence that the error in estimating the true value of p is less than 0.02?

(c) How large must the sample be if we wish to be at least 95% confidence that the error in estimating the true value of p is less than 0.02, regardless of the true value of p?