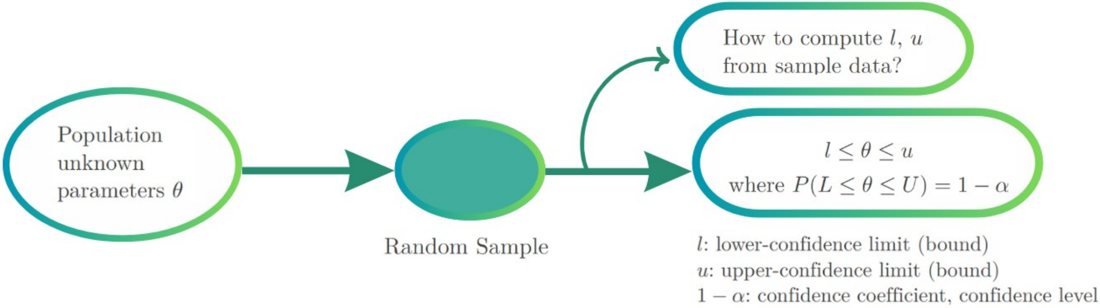
Chapter 8:
Statistical Interval for a Single Sample

MAS291 - STATISTICS & PROBABILITY

Ly Anh Duong

duongla3@fe.edu.vn

Table of Contents



Population
unknown
parameters θ
Random Sample
How to compute l, u
from sample data?
$l \leq \theta \leq u$
where $P(L \leq \theta \leq U) = 1 - \alpha$
l: lower-confidence limit (bound)
u: upper-confidence limit (bound)
$1 - \alpha$: confidence coefficient, confidence level

Table of Contents

2 Confidence Interval on the Mean of a Normal Distribution, Variance known

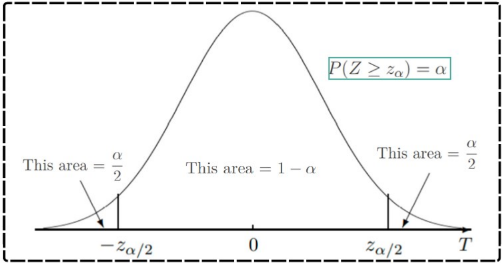
Confidence Interval on the Mean of a Normal Distribution, Variance known

Problem: Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with unknown mean μ and known variance σ^2.
1. A confidence interval estimate for μ is an interval of the form $L \leq \mu \leq U$.
2. If $P(L \leq \mu \leq U) = 1 - \alpha$, $(0 \leq \alpha \leq 1)$, then
   - $[L, U]$ is called random confidence interval (khoảng tin cậy)
   - $1 - \alpha$ is called the confidence coefficient, level of confidence, confidence level (mức tin cậy).

$100(1-\alpha)\%$ CI of A standard normal distribution

If population has normal distribution with mean μ and variance σ^2. Take a sample size n, $X_1, X_2, ..., X_n$ and the sample mean $\overline{X} = \dfrac{X_1 + ... + X_n}{n} \sim \mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$

$$\Rightarrow Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1). \text{ Then}$$



$$P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) = 1 - \alpha$$

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Confidence Interval on the Mean of a Normal Distribution, Variance known

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1-\alpha)\%$ CI on μ is given by:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

Confidence Interval on the Mean of a Normal Distribution, Variance known - Example 8.1

ASTM Standard E23 deines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (J) on specimens of A238 steel cut at $60^o C$ are as follows: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, and 64.3. Assume that impact energy is normally distributed with $\sigma = 1$J. Find a 95% CI for μ, the mean impact energy.

Confidence Interval on the Mean of a Normal Distribution, Variance known - Solution

We want to find a 95% CI for μ, the mean impact energy. The required quantities are $z_{\alpha/2} = z_{0.025} = 1.96$, $n = 10$, $\sigma = 1$, and $\bar{x} = 64.46$. The resulting 95% CI is:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$64.46 - 1.96\frac{1}{\sqrt{10}} \leq \mu \leq 64.46 + 1.96\frac{1}{\sqrt{10}}$$

$$63.84 \leq \mu \leq 65.08$$

Practical Interpretation: Based on the sample data, a range of highly plausible values for mean impact energy for A238 steel at 60°C is $63.84J \leq \mu \leq 65.08J$.

Interpreting a Confidence Interval

If an infinite number of random samples are collected and a $100(1-\alpha)\%$ confidence interval for is computed from each sample, $100(1-\alpha)\%$ of these intervals will contain the true value of μ.

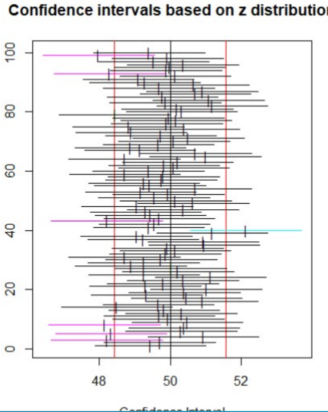
We don't know if the statement is true for this specific sample, but the method used to obtain the interval $[l, u]$ yields correct statements $100(1-\alpha)\%$ of the time.

Simulated confidence intervals

2 Confidence Interval on the Mean of a Normal Distribution, Variance known

100 samples of size 25 were generated from a norm (mean = 50, sd = 4) distribution, and each sample was used to find a 95% confidence interval for the population mean.

The 100 CIs are represented above by horizontal lines, and the respective sample means are denoted by vertical slashes. CIs that "cover" the true $\mu = 50$ are plotted in black; those that fail to cover are plotted in a lighter color. In the plot we see that 7 of the simulated intervals out of the 100 failed to cover $\mu = 50$, which is a success rate of 93%. If the number of generated samples were to increase from 100 to 1000 to 10000,... then we would expect our success rate to approach the exact value of 95%.

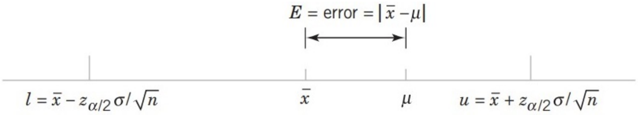


Choice of sample size for specific error on the mean, variance known

If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error $|\bar{x}-\mu|$ will not exceed a specified amount E when the sample size is:

$$n = \left\lceil \left(\frac{z_{\alpha/2}\sigma}{E} \right)^2 \right\rceil$$

Note: From the $100(1-\alpha)\%$ CI: $\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$



Choice of sample size - Exercise

To illustrate the use of this procedure, consider the CVN test described in Example 8-1 and suppose that we want to determine how many specimens must be tested to ensure that the 95% CI on μ for A238 steel cut at $60^o C$ has a length of at most 1.0J

Choice of sample size - Exercise

To illustrate the use of this procedure, consider the CVN test described in Example 8-1 and suppose that we want to determine how many specimens must be tested to ensure that the 95% CI on μ for A238 steel cut at $60^o C$ has a length of at most 1.0J

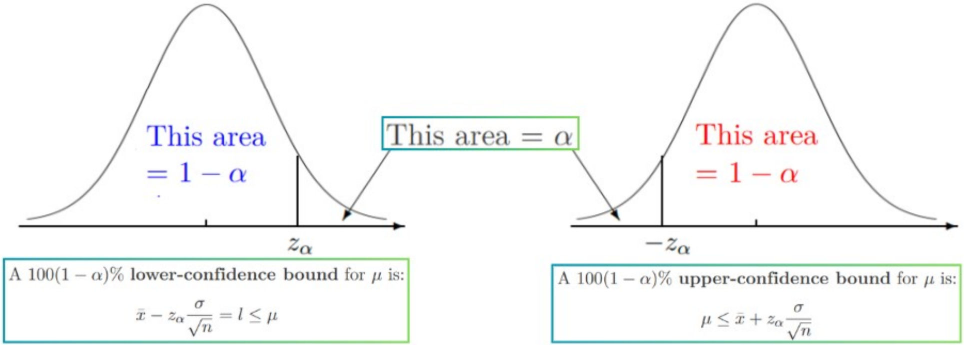
Solution

Using equation 8-6 with $E = 0.5, \sigma = 1, z_{\alpha/2} = 1.96$. The required sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{(1.96).1}{0.5}\right)^2 = 15.37$$

Therefore, n must be 16.

One-Sided Confidence Bounds on the Mean, Variance Known



In case, population distribution is NOT normal, its mean can be estimated by using CLT as follows.

When n is large, the quantity: $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$ has an approximate standard normal distribution. Consequently,

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

is a large-sample confidence interval for μ, with confidence level of approximately $100(1-\alpha)\%$.

Exercise

An article in the 1993 volume of the Transactions of the American Fisheries Society reports the results of a study to investigate the mercury contamination in largemouth bass. A sample of fish was selected from 53 Florida lakes, and mercury concentration in the muscle tissue was measured (ppm). The mercury concentration values were

1.230	1.330	0.040	0.044	1.200	0.270
0.490	0.190	0.830	0.810	0.710	0.500
0.490	1.160	0.050	0.150	0.190	0.770
1.080	0.980	0.630	0.560	0.410	0.730
0.590	0.340	0.340	0.840	0.500	0.340
0.280	0.340	0.750	0.870	0.560	0.170
0.180	0.190	0.040	0.490	1.100	0.160
0.100	0.210	0.860	0.520	0.650	0.270
0.940	0.400	0.430	0.250	0.270	

The summary statistics for these data are as follows:

Variable	*N*	Mean	Median	StDev
Concentration	53	0.5250	0.4900	0.3486

Minimum	Maximum	*Q1*	*Q3*
0.0400	1.3300	0.2300	0.7900

Solution:

We see that $n = 53, s = 0.3486, \overline{x} = 0.525$, $z_{0.025} = 1.96$. The approximation 95% CI on μ is

$$\overline{x} - z_{0.025}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{0.025}\frac{s}{\sqrt{n}}$$

$$0.5250 - 1.96\frac{0.3486}{\sqrt{53}} \leq \mu \leq 0.5250 + 1.96\frac{0.3486}{\sqrt{53}}$$

$$0.4311 \leq \mu \leq 0.6189$$

Table of Contents

t distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable: $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$

has a t distribution

with $k = n - 1$ degrees of freedom.

The t probability density function is:

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\,\Gamma\left(\frac{k}{2}\right)} \cdot \frac{1}{\left[\left(\frac{x^2}{k}\right) + 1\right]^{(k+1)/2}},$$

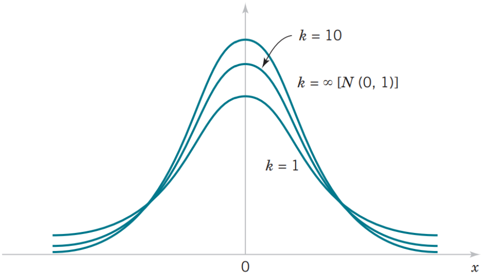


FIGURE 8.4

Probability density functions of several *t* distributions.

t Confidence Interval on Mean

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval on μ is given by:

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

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the t distribution with $n-1$ degrees of freedom.

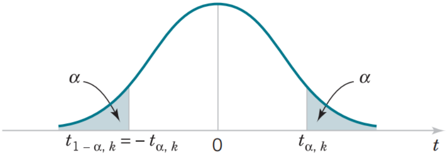


FIGURE 8.5

Percentage points of the *t* distribution.

Remark. One-sided confidence bounds on the mean are found by replacing $t_{\alpha/2,n-1}$ with $t_{\alpha,n-1}$.

t Confidence Interval on Mean - Exercise

An article in the Journal of Materials Engineering [“Instrumented Tensile Adhesion Tests on Plasma Sprayed Thermal Barrier Coatings” (1989, Vol. 11(4), pp. 275–282)] describes the results of tensile adhesion tests on 22 U-700 alloy specimens. The load at specimen failure is as follows (in megapascals):

19.8	10.1	14.9	7.5	15.4	15.4
15.4	18.5	7.9	12.7	11.9	11.4
11.4	14.1	17.6	16.7	15.8	
19.5	8.8	13.6	11.9	11.4	

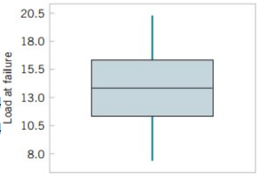


FIGURE 8.6

Box-and-whisker plot for the load at failure data.

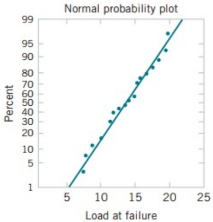


FIGURE 8.7

Normal probability plot of the load at failure data.

Construct 95% CI for μ.

t Confidence Interval on Mean - Solution

The sample mean is $\bar{x} = 13.71$, and the sample standard deviation is $s = 3.55$. Figures 8.6 and 8.7 show a box plot and a normal probability plot of the tensile adhesion test data, respectively. These displays provide good support for the assumption that the population is normally distributed. We want to find a 95% CI on μ. Since $n = 22$, we have $n - 1 = 21$ degrees of freedom for t, so $t_{0.025,21} = 2.080$. The resulting CI is:

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

$$13.71 - 2.080(3.55)/\sqrt{22} \leq \mu \leq 13.71 + 2.080(3.55)/\sqrt{22}$$

$$12.14 \leq \mu \leq 15.28$$

Table of Contents

Confidence Interval on σ^2 and σ of a Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2, and let S^2 be the sample variance. Then the random variable:

$$X^2 = \frac{(n-1)S^2}{\sigma^2}$$

has a chi-square (χ^2) distribution with $n-1$ degrees of freedom.
The probability density function of a χ^2 random variable is:

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}, \quad x > 0$$

Percentage point of the χ^2 distribution

4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

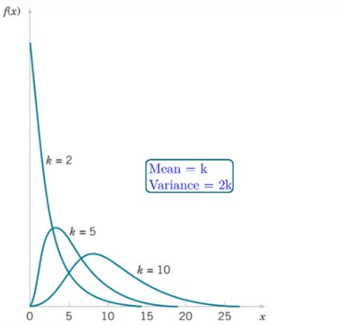


FIGURE 8.8

Probability density functions of several χ^2 distributions.

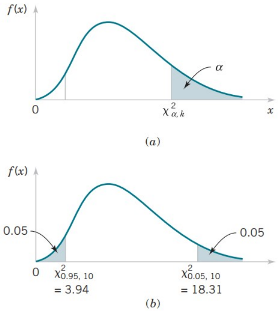


FIGURE 8.9

Percentage point of the χ^2 distribution. (a) The percentage point $\chi^2_{\alpha,k}$. (b) The upper percentage point $\chi^2_{0.05,10} = 18.31$ and the lower percentage point $\chi^2_{0.95,10} = 3.94$.

If s^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance σ^2, then a $100(1-\alpha)\%$ confidence interval on σ^2 is:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \tag{1}$$

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. A confidence interval for σ has lower and upper limits that are the square roots of the corresponding limits in Equation 1.

One-Sided Confidence Bounds on the Variance

The $100(1-\alpha)\%$ lower and upper confidence bounds on σ^2 are:

$$\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}} \leq \sigma^2 \quad \text{and} \quad \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}}$$

respectively.

Confidence Interval on the Variance - Exercise

An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153^2$ (fluid ounce). If the variance of fill volume is too large, an unacceptable proportion of bottles will be under- or overfilled. We will assume that the fill volume is approximately normally distributed.
Find 95% upper confidence bound on the variance.

Confidence Interval on the Variance - Exercise

An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153^2$ (fluid ounce). If the variance of fill volume is too large, an unacceptable proportion of bottles will be under- or overfilled. We will assume that the fill volume is approximately normally distributed.
Find 95% upper confidence bound on the variance.

Solution

We have $s^2 = 0.0153^2$ and $\chi_{0.95,19} = 10.117$. Then

$$\sigma^2 \leq \frac{(n-1)s^2}{\chi_{0.95,19}} = \frac{(19)(0.0153)^2}{10.117} = 0.0287$$

Practical Interpretation: At the 95% level of confidence, the data indicate that the process standard deviation could be as large as 0.17 fluid ounce. The process engineer or manager now needs to determine if a standard deviation this large could lead to an operational problem with under-or over filled bottles.

Table of Contents

5 Large-Sample Confidence Interval for a Population Proportion

Normal Approximation for a Binomial Proportion

5 Large-Sample Confidence Interval for a Population Proportion

- p: a population proportion
- $\widehat{P} = X/n$: a point estimator of p
- When n is large enough, $X/n \sim \text{Normal}(\text{ mean} = p, \text{ variance} = p(1-p)/n)$, if p is not too close to either 0 or 1.
- Requirement for approximation: $np, n(1-p) \geq 5$.

If n is large, the distribution of

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

is approximately standard normal.

Approximate Confidence Interval on a Binomial Proportion

If $\hat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest, an approximate $100(1-\alpha)\%$ confidence interval on the proportion p of the population that belongs to this class is:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

Approximate Confidence Interval on a Binomial Proportion - Exercise

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. Therefore, a point estimate of the proportion of bearings in the population that exceeds the roughness specification is $\hat{p} = \dfrac{x}{n} = 10/85 = 0.12$.

Find 95% two-sided confidence interval for p.

Approximate Confidence Interval on a Binomial Proportion - Exercise

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. Therefore, a point estimate of the proportion of bearings in the population that exceeds the roughness specification is $\hat{p} = \dfrac{x}{n} = 10/85 = 0.12$.

Find 95% two-sided confidence interval for p.

Solution

$$\hat{p} - z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.12 - 1.96\sqrt{\frac{0.12(0.88)}{85}} \leq p \leq 0.12 + 1.96\sqrt{\frac{0.12(0.88)}{85}}$$

which simplifies to:

$$0.0509 \leq p \leq 0.2243$$

Choice of Sample Size

Sample Size for a Specified Error ($E = |p - \hat{P}| \leq z_{\alpha/2}\sqrt{p(1-p)/n}$) on a Binomial Proportion

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p})$$

An upper bound on n is given by (Note that $p(1-p) \leq 0.25$)

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (0.25)$$

Choice of Sample Size - Exercise

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. How large a sample is required

a) If we want to be 95% confident that the error in using $\hat{p}$ to estimate p is less than 0.05?

b) If we wanted to be at least 95% confident that our estimate $\hat{p}$ of the true proportion p was within 0.05 regardless of the value of p?

Choice of Sample Size - Exercise

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. How large a sample is required
a) If we want to be 95% confident that the error in using $\hat{p}$ to estimate p is less than 0.05?
b) If we wanted to be at least 95% confident that our estimate $\hat{p}$ of the true proportion p was within 0.05 regardless of the value of p?

Solution

a)

$$n = \left(\frac{z_{0.025}}{E}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.05}\right)^2 0.12(0.88) \approx 163$$

b)

$$n = \left(\frac{z_{0.025}}{E}\right)^2 (0.25) = \left(\frac{1.96}{0.05}\right)^2 (0.25) \approx 385$$

Approximate One-Sided Confidence Bounds on a Binomial Proportion

The approximate $100(1-\alpha)\%$ lower and upper confidence bounds are:

$$\hat{p} - z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \quad \text{and} \quad p \le \hat{p} + z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

respectively.

Summary (marginal error ϵ)

CI for μ:

Case	Condition	ϵ	CI
1	$(i).\ n < 30,\ X \sim \mathcal{N}(\mu, \sigma^2)$ or $(i').\ n \geq 30$ and σ known	$z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	$[\bar{x} - \epsilon, \bar{x} + \epsilon]$
2	$n \geq 30$ and σ^2 unknown	$z_{\alpha/2}\frac{s}{\sqrt{n}}$	
3	$n < 30,\ X \sim \mathcal{N}(\mu, \sigma^2)$ and σ^2 unknown	$t_{\alpha/2,\,n-1}\frac{s}{\sqrt{n}}$	

CI for p:

Condition	ϵ	CI
$n\hat{p} \geq 5$ và $n(1-\hat{p}) \geq 5$	$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$[\hat{p} - \epsilon, \hat{p} + \epsilon]$

CI for σ^2: $\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2;n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2;n-1}}\right]$

CI for σ: $\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2;n-1}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2;n-1}}}\right]$

Q&A

Thank you for listening!