Lecture 8

STATISTICAL INTERVALS FOR A SINGLE SAMPLE

FPT Education

FPT UNIVERSITY

Department of Mathematics

Võ Văn Nam

Contents

1. Introduction to Point estimator

Point estimator

- A **point estimate** of a parameter θ is a single number that can be regarded as a sensible value for θ. A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic is called the **point estimator** of θ.

- A point estimator $\hat{\theta}$ is said to be an **unbiased estimator** of θ if $E(\hat{\theta}) = \theta$ for every possible value of θ. If $\hat{\theta}$ is not unbiased, the difference $E(\hat{\theta}) - \theta$ is called the bias of $\hat{\theta}$.

FPT Education FPT UNIVERSITY

Example

An automobile manufacturer has developed a new type of bumper, which is supposed to absorb impacts with less damage than previous bumpers. The manufacturer has used this bumper in a sequence of 25 controlled crashes against a wall, each at 10 mph, using one of its compact car models.

Let X be the number of crashes that result in no visible damage to the automobile. The parameter to be estimated is p the proportion of all such crashes that result in no damage. If X is observed to be $x = 15$, the most reasonable estimator and estimate are

$$\text{estimator } \hat{p} = \frac{X}{n}, \quad \text{estimate} = \frac{x}{n} = 0.6$$

Confidence interval

- A **confidence interval** estimate for μ with $100(1-\alpha)\%$ confidence level is an interval of the form $\ell \leq \mu \leq u$, where ℓ and u are computed from the sample data, such that

$$P(\ell \leq \mu \leq u) = 1 - \alpha,$$

where ℓ: lower limit; u: upper limit.

- A confidence interval provides additional information about variability.

Question. How to find ℓ and u?

Critical value z_α

Critical value z_α (or a percentage point) is the value such that $P(Z > z_\alpha) = \alpha$.

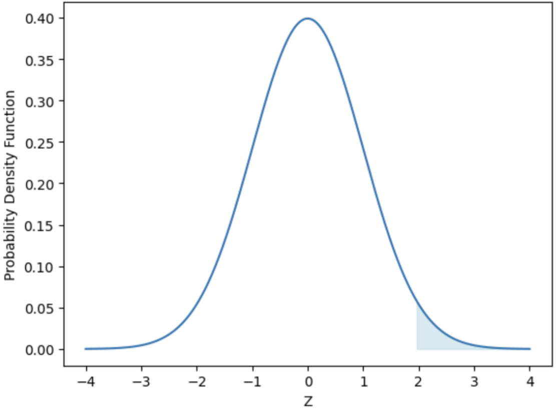
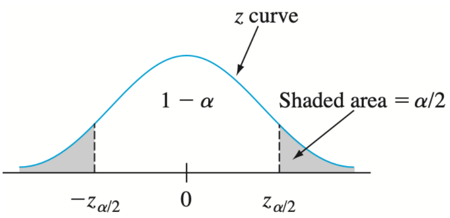


Figure: $z_{0.025} = 1.96$ since shaded area $= 0.025$

2. Confidence interval for the population mean μ

Confidence interval for μ



If σ is known then by CLT, we have

$$P(-z_{\alpha/2} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}) = 1-\alpha,$$

or equivalently,

$$P(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha.$$

Confidence interval for μ (σ is known)

Theorem (C.I. for μ if σ is 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 $100(1-\alpha)\%$ C.I. on μ is given by

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

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

Example

In a sample of 36 randomly selected women, it was found that their mean height was 65.3 inches. From previous studies, it is assumed that the standard deviation of all womens heights $\sigma = 2.5$ inches. Construct a 90%, 95% confidence intervals for the mean height of all women.

Answer. 90% Confidence Interval: (64.61, 65.99)
95% Confidence Interval: (64.48, 66.12)

Remarks

- The error of estimation: $|\bar{x} - \mu| \leq z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
- To reduce the error, we can increase sample size n.
- If we want to be $100(1 - \alpha)\%$ confident that the error will not exceed a specified amount E the required sample size should be

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

Example. A nurse at a local hospital is interested in estimating the birth weight of infants. How large a sample must she select if she desires to be 98% confident that the true mean is within 4 ounces of the sample mean? Assume that the standard deviation of the birth weights is known to be 6 ounces and $z_{0.01} = 2.326$.

FPT Education
FPT UNIVERSITY

One-sided confidence interval for μ (σ is known)

Theorem (One-sided C.I. for μ if σ is known)

- *A $100(1-\alpha)\%$ upper-confidence bound for μ is*

$$\mu \leq \bar{x} + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

- *A $100(1-\alpha)\%$ lower-confidence bound for μ is*

$$\mu \geq \bar{x} - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

FPT UNIVERSITY

Example

The diameter of holes for a cable harness is known to have a normal distribution with $\sigma = 0.01$ inch. A random sample of size 15 yields an average diameter of 1.5 inches. Find a 98% lower-confidence bound for the population mean. Given that $z_{0.02} = 2.054$

Choose the best answer.

A. 1.495 **B.** 1.505 **C.** 1.506 **D.** 1.494

Confidence interval for μ (σ is unknown)

Question. What if we do NOT know the population standard deviation σ?

$\rightarrow$ Replace the population standard deviation by a sample standard deviation s.
$\rightarrow$ Use t-distribution instead of normal distribution.

Question. What is t-distribution?

The **t-distribution**, also known as the **Student's t-distribution**, is a probability distribution that is used in statistical inference for estimating population parameters when the sample size is small and the population standard deviation is unknown.

Student's t-distribution

- The t-distribution is similar to the normal distribution but has heavier tails, making it more suitable for small sample sizes. It is parameterized by the **degrees of freedom** (df), which is related to the sample size. The t-distribution approaches the standard normal distribution as the degrees of freedom increase.

- The probability density function (PDF) of the t-distribution is given by:

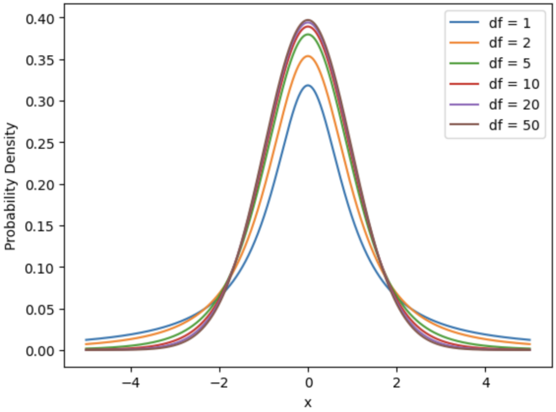
$$f(t; df) = \frac{\Gamma\left(\frac{df+1}{2}\right)}{\sqrt{df\pi}\,\Gamma\left(\frac{df}{2}\right)}\left(1 + \frac{t^2}{df}\right)^{-\frac{df+1}{2}}$$

where:

- t is the random variable following the t-distribution,
- df is the degrees of freedom,
- Γ is the gamma function.

Student's t-distribution (cont')

Key properties of the t-distribution include:



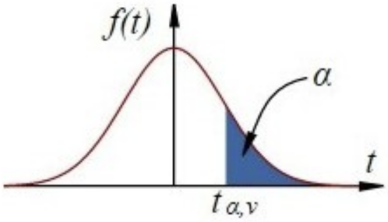
1. *Symmetry:* The t-distribution is symmetric around zero.
2. *Mean and Median:* The mean of the t-distribution is zero, and the median is also zero.
3. *Tails:* As the degrees of freedom increase, the t-distribution approaches the standard normal distribution, and the tails become less heavy.

Confidence interval for μ (σ is unknown)

Theorem (t-distribution)

Suppose $X_1, X_2, ..., X_n$ *is a random sample of size* n *taken from a population has normal distribution with mean* μ. *Let* $\bar{x}, s^2$ *be the sample mean and sample variance, respectively.*

Then, $T = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ *is approximated by a t-distribution with degree of freedom* $df = n - 1$.



$t_{\alpha,df}$ = T.INV.2T$(2\alpha, df)$ (in Excel)

Confidence interval for μ (σ is unknown)

Theorem (C.I. for μ if σ is unknown)

Two-sided $100(1-\alpha)\%$ *confidence interval on* μ *is:*

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

.

Example. The weight distribution of all products of a company has a normal distribution. A random sample of products has the following weights (in kg):

$$1.9 \quad 2.0 \quad 2.0 \quad 2.1 \quad 1.8 \quad 2.2 \quad 1.8$$

Construct a 95% confidence interval for the true average weight of all products.

One-sided confidence interval for μ (σ is unknown)

Theorem (One-sided C.I. for μ if σ is unknown)

- *A* $100(1-\alpha)\%$ *upper-confidence bound for* μ *is*

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

- *A* $100(1-\alpha)\%$ *lower-confidence bound for* μ *is*

$$\mu \geq \bar{x} - t_{\alpha,n-1} \cdot \frac{s}{\sqrt{n}}$$

Example (continues the previous example). Construct a 95% lower-confidence bound for the average weight of all products.

3. Confidence interval for the population proportion p

Confidence interval for p

Theorem

A random sample of size n has been taken from a population, and x observations in this sample belong to a class of interest.

Then, $\hat{p} = \dfrac{x}{n}$ is a point estimator of population proportion p.

For n large, $\dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ is approximated by the standardnormal distribution Z.

Confidence interval for p

Theorem (C.I. for p)

Two-sided $100(1-\alpha)\%$ *confidence interval on the proportion* p *of the population is*

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

.

Example. Of 1000 randomly selected cases of lung cancer, 750 resulted in death within 10 years. Calculate a 95% two-sided confidence interval of the death rate from lung cancer. Given that $z_{0.025} = 1.96$.

FPT Education
FPT UNIVERSITY

Remarks

The required sample size that the error estimating $|\hat{p} - p|$ not exceed E is:

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

If a previous estimate $\hat{p}$ is known, change $p(1-p)$ by $\hat{p}(1-\hat{p})$. Else, use

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

Example. (cont') What sample size is needed to be 95% confident that the error in estimating the true value of p is less than 4%?

One-sided confidence interval for p

Theorem (One-sided C.I. for p)

- *A* $100(1-\alpha)\%$ *upper-confidence bound for* p *is* $p \leq \hat{p} + z_\alpha \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
- *A* $100(1-\alpha)\%$ *lower-confidence bound for* p *is* $p \geq \hat{p} - z_\alpha \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Example. A survey of 250 homeless people showed that 47 were veterans. Construct a 95% upper confidence bound for the proportion of homeless people who are veterans.

FPT Education
FPT UNIVERSITY

Appendix: Gamma Function

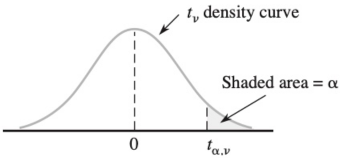
- The **gamma function**, denoted by $\Gamma(x)$, is a mathematical function that generalizes the concept of a factorial to non-integer values. It is defined for all complex numbers except for non-positive integers, where it is undefined. The gamma function is a fundamental component in various branches of mathematics, including analysis, number theory, and probability.
- The gamma function is defined as:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt$$

- For positive integers n, $\Gamma(n) = (n-1)!$. The gamma function extends the factorial function to real and complex numbers.
- In Python, you can use the 'gamma' function from the 'scipy.special' module to compute the gamma function.

FPT Education
FPT UNIVERSITY

Table A.5 Critical Values for t Distributions



	α						
v	.10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781

10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659

30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582
38	1.304	1.686	2.024	2.429	2.712	3.319	3.566
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.262	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Thank you!

namvv14@fe.edu.vn