Chapter 6
ONE–SAMPLE ESTIMATION PROBLEMS

NGUYỄN THỊ THU THỦY[1]

SCHOOL OF APPLIED MATHEMATICS AND INFORMATICS
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

HANOI – 2020

[1]Email: thuy.nguyenthithu2@hust.edu.vn

Content

6.1.1 Point Estimate

Definition 1 (Unbiased estimator)

A statistic $\hat{\Theta}$ is said to be an unbiased estimator of the parameter θ if

$$\mu_{\hat{\Theta}} = E[\hat{\Theta}] = \theta.$$

Example 1

Show that S^2 is an unbiased estimator of the parameter σ^2.

6.1.1 Point Estimate

Solution

$$E[S^2] = E\Big[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\Big]$$
$$= \frac{1}{n-1}\Big[\sum_{i=1}^{n} E[X_i - \mu]^2 - nE[\overline{X} - \mu]^2\Big] = \frac{1}{n-1}\Big(\sum_{i=1}^{n}\sigma_{X_i}^2 - \sigma_{\overline{X}}^2\Big).$$

However,

$$\sigma_{X_i}^2 = \sigma^2, \quad \text{for } i = 1, 2, \ldots, n, \quad \text{and } \sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}.$$

Therefore,

$$E[S^2] = \frac{1}{n-1}\Big(n\sigma^2 - n\frac{\sigma^2}{n}\Big) = \sigma^2.$$

6.1.1 Point Estimate

Definition 2 (Most efficient estimator)

If we consider all possible unbiased estimators of some parameter θ, the one with the smallest variance is called the most efficient estimator of θ.

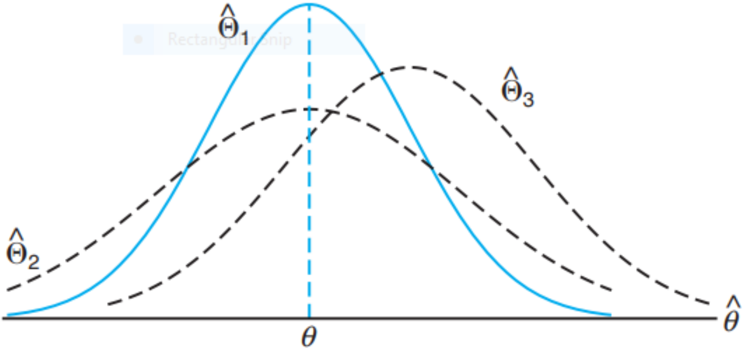

Figure: Sampling distributions of different estimators of θ

6.1.1 Point Estimate

Note

Figure 1 illustrates the sampling distributions of three different estimators, $\hat{\Theta}_1$, $\hat{\Theta}_2$, and $\hat{\Theta}_3$, all estimating θ. It is clear that only $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased, since their distributions are centered at θ. The estimator $\hat{\Theta}_1$ has a smaller variance than $\hat{\Theta}_2$ and is therefore more efficient. Hence, our choice for an estimator of θ, among the three considered, would be $\hat{\Theta}_1$.

6.1.2 Interval Estimation

Interval estimate

An interval estimate of a population parameter θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$, where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on the value of the statistic $\hat{\Theta}$ for a particular sample and also on the sampling distribution of $\hat{\Theta}$.

6.1.2 Interval Estimation

Interpretation of Interval Estimates

- From the sampling distribution of $\hat{\Theta}$ we shall be able to determine $\hat{\Theta}_L$ and $\hat{\Theta}_U$ such that $P(\hat{\Theta}_L < \theta < \hat{\Theta}_U)$ is equal to any positive fractional value we care to specify. If, for instance, we find $\hat{\Theta}_L$ and $\hat{\Theta}_U$ such that

$$P(\hat{\Theta}_L < \theta < \hat{\Theta}_U) = 1 - \alpha,$$

  for $0 < \alpha < 1$, then we have a probability of $1 - \alpha$ of selecting a random sample that will produce an interval containing θ.
- The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$, computed from the selected sample, is called a confidence interval.
- The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$, computed from the selected sample, is called a $100(1 - \alpha)\%$ confidence interval.
- The endpoints, $\hat{\theta}_L$ and $\hat{\theta}_U$, are called the lower and upper confidence limits.

6.1.2 Interval Estimation

Interpretation of Interval Estimates

- When $\alpha = 0.05$, we have a 95% confidence interval, and when $\alpha = 0.01$, we obtain a wider 99% confidence interval.
- The wider the confidence interval is, the more confident we can be that the interval contains the unknown parameter.

Content

Two-Sided Intervals

- Writing $z_{\alpha/2}$ for the z-value which we find an area of $\alpha/2$ under the normal curve, we can see from Figure 2 that

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha,$$

where

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}.$$

- Hence,

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

- So,

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

Two-Sided Intervals

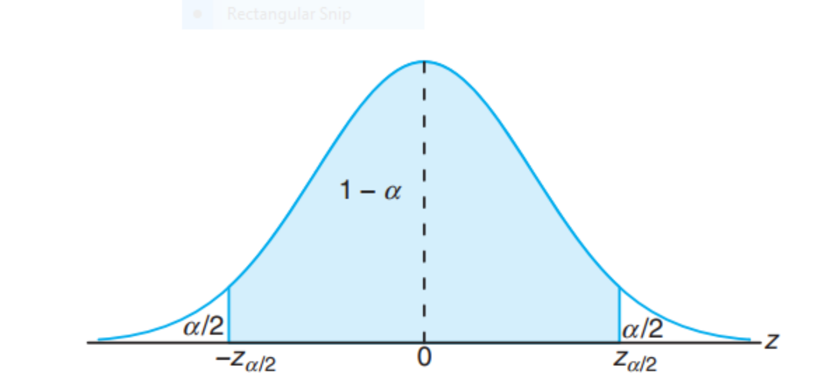

Figure: $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha,$

Two-Sided Intervals

Theorem 1 (Confidence Interval on μ, σ^2 Known)

If $\overline{x}$ is the mean of a random sample of size n from a population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is given by

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

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Two-Sided Intervals

Example 2

The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter.

Solution

- The point estimate of μ is $\overline{x} = 2.6$.
- The z-value leaving an area of 0.025 to the right, and therefore an area of 0.975 to the left, is $z_{0.025} = 1.96$.
- Hence, the 95% confidence interval is

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right)$$

which reduces to $2.50 < \mu < 2.70$.

Two-Sided Intervals

Solution (continuous)

- To find a 99% confidence interval, we find the z-value leaving an area of 0.005 to the right and 0.995 to the left.
- We have $z_{0.005} = 2.58$, and the 99% confidence interval is

$$2.6 - (2.58)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.58)\left(\frac{0.3}{\sqrt{36}}\right)$$

or simply

$$2.471 < \mu < 2.729.$$

Two-Sided Intervals

Table the values of standard normal CDF $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0,1	53983	54380	54776	55172	55567	55962	56356	56749	57142	57535
0,2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0,3	61791	62172	62556	62930	63307	63683	64058	64431	64803	65173
0,4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68739
0,5	69146	69447	69847	70194	70544	70884	71226	71566	71904	72240
0,6	72575	72907	73237	73565	73891	74215	74537	74857	75175	75490
0,7	75804	76115	76424	76730	77035	77337	77637	77935	78230	78524
0,8	78814	79103	79389	79673	79955	80234	80511	80785	81057	81327
0,9	81594	81859	82121	82381	82639	82894	83147	83398	83646	83891
1,0	84134	84375	84614	84850	85083	85314	85543	85769	85993	86214
1,1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1,2	88493	88686	88877	89065	89251	89435	89617	89796	89973	90147
1,3	90320	90490	90658	90824	90988	91149	91309	91466	91621	91774
1,4	91924	92073	92220	92364	92507	92647	92786	92922	93056	93189
1,5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1,6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449
1,7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327
1,8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062
1,9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670

Two-Sided Intervals

Table the values of standard normal CDF $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$

x	0	1	2	3	4	5	6	7	**8**	9
2,0	97725	97778	97831	97882	97932	97982	98030	98077	98124	98169
2,1	98214	98257	98300	98341	98382	98422	99461	98500	98537	98574
2,2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899
2,3	98928	98956	98983	99010	99036	99061	99086	99111	99134	99158
2,4	99180	99202	99224	99245	99266	99285	99305	99324	99343	99361
2,5	99379	99396	99413	99430	99446	99261	99477	99492	**99506**	99520
2,6	99534	99547	99560	99573	99585	99598	99609	99621	99632	99643
2,7	99653	99664	99674	99683	99693	99702	99711	99720	99728	99763
2,8	99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
2,9	99813	99819	99825	99831	99836	99841	99846	99851	99856	99861
3,0	0,99865	3,1	99903	3,2	99931	3,3	99952	3,4	99966	
3,5	99977	3,6	99984	3,7	99989	3,8	99993	3,9	99995	
4,0	999968									
4,5	999997									
5,0	99999997									

Two-Sided Intervals

Theorem 2 (Error)

If $\overline{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed $e = z_{\alpha/2}\sigma/\sqrt{n}$.

Example 3

In Example 2, we are

- 95% confident that the sample mean $\overline{x} = 2.6$ differs from the true mean μ by an amount less than $(1.96)(0.3)/\sqrt{36} = 0.1$ and
- 99% confident that the difference is less than $(2.58)(0.3)/\sqrt{36} = 0.129$.

Two-Sided Intervals

Theorem 3 (Size)

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

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2.$$

Two-Sided Intervals

Example 4

How large a sample is required if we want to be 95% confident that our estimate of μ in Example 2 is off by less than 0.05?

Solution

The population standard deviation is $\sigma = 0.3$. Then, by Theorem 3,

$$n = \left[\frac{(1.96)(0.3)}{0.05}\right]^2 = 138.3.$$

Therefore, we can be 95% confident that a random sample of size 139 will provide an estimate $\overline{x}$ differing from μ by an amount less than 0.05.

One-Sided Confidence Bounds

- From the Central Limit Theorem:

$$P\Big(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_\alpha\Big) = 1-\alpha.$$

One can then manipulate the probability statement much as before and obtain

$$P(\mu > \overline{X} - z_\alpha\sigma/\sqrt{n}) = 1-\alpha$$

- Similar manipulation of

$$P\Big(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > -z_\alpha\Big) = 1-\alpha$$

gives

$$P(\mu < \overline{X} + z_\alpha\sigma/\sqrt{n}) = 1-\alpha$$

One-Sided Confidence Bounds

Theorem 4 (One-Sided Confidence Bounds on μ, σ^2 Known)

If $\overline{X}$ is the mean of a random sample of size n from a population with variance σ^2, the one-sided $100(1-\alpha)\%$ confidence bounds for μ are given by

$$\text{upper one-sided bound: } \quad \overline{x} + z_\alpha \sigma/\sqrt{n};$$

$$\text{lower one-sided bound: } \quad \overline{x} - z_\alpha \sigma/\sqrt{n}.$$

One-Sided Confidence Bounds

Example 5

In a psychological testing experiment, 25 subjects are selected randomly and their reaction time, in seconds, to a particular stimulus is measured. Past experience suggests that the variance in reaction times to these types of stimuli is 4 $\sec^2$ and that the distribution of reaction times is approximately normal. The average time for the subjects is 6.2 seconds. Give an upper 95% bound for the mean reaction time.

Solution

The upper 95% bound is given by

$$\overline{x} + z_\alpha \sigma/\sqrt{n} = 6.2 + (1.645)\sqrt{4/25} = 6.2 + 0.658 = 6.858 \text{ seconds.}$$

Hence, we are 95% confident that the mean reaction time is less than 6.858 seconds.

6.2.2 The Case of σ Unknown

- The random variable $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ has a t-distribution with $n-1$ degrees of freedom.
- The procedure is the same as that with σ known except that σ is replaced by S and the standard normal distribution is replaced by the t-distribution.
- Referring to Figure 3, we can assert that

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha,$$

where $t_{\alpha/2}$ is the t-value with $n-1$ degrees of freedom.

- Substituting for T, we write

$$P\Big(- t_{\alpha/2} < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}\Big) = 1 - \alpha.$$

- Hence

$$P\Big(\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}\Big) = 1 - \alpha.$$

6.2.2 The Case of σ Unknown

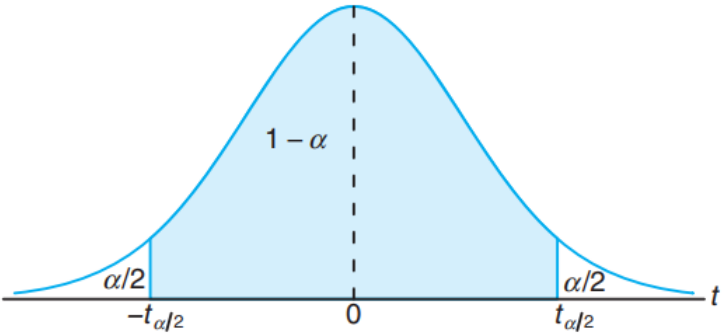

Figure: $P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha,$

6.2.2 The Case of σ Unknown

Theorem 5 (Confidence Interval on μ, σ^2 Unknown)

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

$$\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}},$$

where $t_{\alpha/2}$ is the t-value with $\nu = n-1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

6.2.2 The Case of σ Unknown

Theorem 6 (One-Sided Confidence Bounds on μ, σ^2 Unknown)

Computed one-sided confidence bounds for μ with σ unknown are as the reader would expect, namely

$$\text{upper one-sided bound:} \quad \overline{x} + t_\alpha \frac{\sigma}{\sqrt{n}};$$

$$\text{lower one-sided bound:} \quad \overline{x} - t_\alpha \frac{\sigma}{\sqrt{n}}.$$

They are the upper and lower $100(1-\alpha)\%$ bounds, respectively. Here t_α is the t-value having an area of α to the right.

6.2.2 The Case of σ Unknown

Example 6

The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, and 9.6 liters. Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution.

Solution

- The sample mean and standard deviation for the given data are $\overline{x} = 10.0$ and $s = 0.283$.
- Using Table t-Distribution, we find $t_{0.025} = 2.447$ for $v = 6$ degrees of freedom.
- Hence, the 95% confidence interval for μ is

$$10.0 - (2.447)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10.0 + (2.447)\left(\frac{0.283}{\sqrt{7}}\right),$$

which reduces to $9.74 < \mu < 10.26$.

t-Distribution

d.f. \ α	$0,10$	$0,05$	**0,025**	$0,01$	$0,005$	$0,002$	0.0005
1	3,078	6,314	12,706	31,821	63,526	318,309	363,6
2	1,886	2,920	4,303	6,965	9,925	22,327	31,600
3	1,638	2,353	3,128	4,541	5,841	10,215	12,922
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	4,705	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	2,606	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

t-Distribution

d.f. \ α	$0,10$	$0,05$	**0,025**	$0,01$	$0,005$	$0,002$	0.0005
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	1,796	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
$+\infty$	1,282	1,645	1,960	2,326	2,576	3,090	3,291

6.2.2 The Case of σ Unknown

Concept of a Large-Sample Confidence Interval

Often statisticians recommend that even when normality cannot be assumed, σ is unknown, and $n \geq 30$, s can replace σ and the confidence interval

$$\overline{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

may be used.

6.2.2 The Case of σ Unknown

Example 7

Scholastic Aptitude Test (SAT) mathematics scores of a random sample of 500 high school seniors in the state of Texas are collected, and the sample mean and standard deviation are found to be 501 and 112, respectively. Find a 99% confidence interval on the mean SAT mathematics score for seniors in the state of Texas.

Solution

Since the sample size is large, it is reasonable to use the normal approximation. Using Table the values of standard normal CDF $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$, we find $z_{0.005} = 2.575$. Hence, a 99% confidence interval for μ is

$$501 \pm (2.575)\left(\frac{112}{\sqrt{500}}\right) = 501 \pm 12.9,$$

which yields $488.1 < \mu < 513.9$.

Content

6.3 Estimating Proportion

- A point estimator of the proportion p in a binomial experiment is given by the statistic $\hat{P} = X/n$, where X represents the number of successes in n trials.
- Therefore, the sample proportion $\hat{p} = x/n$ will be used as the point estimate of the parameter p.

6.3 Estimating Proportion

- By the Central Limit Theorem, for n sufficiently large, $\hat{P}$ is approximately normally distributed with mean

$$\mu_{\hat{P}} = E[\hat{P}] = E\left[\frac{X}{n}\right] = \frac{np}{n} = p$$

and variance

$$\sigma^2_{\hat{P}} = \sigma^2_{X/n} = \frac{\sigma^2_X}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}.$$

- Therefore, we can assert that

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha, \quad \text{with } Z = \frac{\hat{P} - p}{\sqrt{pq/n}},$$

and $z_{\alpha/2}$ is the value above which we find an area of $\alpha/2$ under the standard normal curve.

6.3 Estimating Proportion

- Substituting for Z, we write

$$P\left(-z_{\alpha/2} < \frac{\hat{P}-p}{\sqrt{pq/n}} < z_{\alpha/2}\right) = 1-\alpha.$$

- When n is large, very little error is introduced by substituting the point estimate $\hat{p} = x/n$ for the p under the radical sign. Then we can write

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \simeq 1-\alpha.$$

- On the other hand, by solving for p in the quadratic inequality above, $-z_{\alpha/2} < \dfrac{\hat{P}-p}{\sqrt{pq/n}} < z_{\alpha/2}$, we obtain another form of the confidence interval for p with limits

$$\frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n}}{1 + \frac{z^2_{\alpha/2}}{n}} \pm \frac{z_{\alpha/2}}{1 + \frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z^2_{\alpha/2}}{4n^2}}.$$

6.3 Estimating Proportion

Theorem 7 (Large-Sample Confidence Intervals for p)

If $\hat{p}$ is the proportion of successes in a random sample of size n and $\hat{q} = 1 - \hat{p}$, an approximate $100(1-\alpha)\%$ confidence interval, for the binomial parameter p is given by **(method 1)**

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

or by **(method 2)**

$$\frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n}}{1 + \frac{z^2_{\alpha/2}}{n}} - \frac{z_{\alpha/2}}{1 + \frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z^2_{\alpha/2}}{4n^2}} < p < \frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n}}{1 + \frac{z^2_{\alpha/2}}{n}} + \frac{z_{\alpha/2}}{1 + \frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z^2_{\alpha/2}}{4n^2}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

6.3 Estimating Proportion

Note

- When n is small and the unknown proportion p is believed to be close to 0 or to 1, the confidence-interval procedure established here is unreliable and, therefore, should not be used.
- To be on the safe side, one should require both $n\hat{p}$ and $n\hat{q}$ to be greater than or equal to 5.
- Note that although method 2 yields more accurate results, it is more complicated to calculate, and the gain in accuracy that it provides diminishes when the sample size is large enough. Hence, method 1 is commonly used in practice.

6.3 Estimating Proportion

Example 8

In a random sample of $n = 500$ families owning television sets in the city of Hamilton, Canada, it is found that $x = 340$ subscribe to HBO. Find a 95% confidence interval for the actual proportion of families with television sets in this city that subscribe to HBO.

6.3 Estimating Proportion

Solution

- The point estimate of p is $\hat{p} = 340/500 = 0.68$. We find $z_{0.025} = 1.96$. Therefore, using method 1, the 95% confidence interval for p is

$$0.68 - 1.96\sqrt{\frac{(0.68)(0.32)}{500}} < p < 0.68 + 1.96\sqrt{\frac{(0.68)(0.32)}{500}},$$

  which simplifies to $0.6391 < p < 0.7209$.

- If we use method 2, we can obtain

$$\frac{0.68 + \frac{1.96^2}{(2)(500)}}{1 + \frac{1.96^2}{500}} \pm \frac{1.96}{1 + \frac{1.96^2}{500}}\sqrt{\frac{(0.68)(0.32)}{500} + \frac{1.96^2}{(4)(500)^2}} = 0.6786 \pm 0.0408,$$

  which simplifies to $0.6378 < p < 0.7194$.

- Apparently, when n is large (500 here), both methods yield very similar results.

6.3 Estimating Proportion

Theorem 8 (Error)

If $\hat{p}$ is used as an estimate of p, we can be $100(1-\alpha)\%$ confident that the error will not exceed $e = z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.

Example 9

In Example 8, we are 95% confident that the sample proportion $\hat{p} = 0.68$ differs from the true proportion p by an amount not exceeding 0.04.

6.3 Estimating Proportion

Theorem 9 (Choice of Sample Size)

If $\hat{p}$ is used as an estimate of p, we can be $100(1-\alpha)\%$ confident that the error will be less than a specified amount e when the sample size is approximately

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{e^2}.$$

6.3 Estimating Proportion

Example 10

How large a sample is required if we want to be 95% confident that our estimate of p in Example 8 is within 0.02 of the true value?

Solution

Let us treat the 500 families as a preliminary sample, providing an estimate $\hat{p} = 0.68$. Then, by Theorem 9,

$$n = \frac{(1.96)^2(0.68)(0.32)}{(0.02)^2} = 2089.8 \simeq 2090.$$

Therefore, if we base our estimate of p on a random sample of size 2090, we can be 95% confident that our sample proportion will not differ from the true proportion by more than 0.02.

6.3 Estimating Proportion

Theorem 10 (Choice of Sample Size)

If $\hat{p}$ is used as an estimate of p, we can be at least $100(1-\alpha)\%$ confident that the error will not exceed a specified amount e when the sample size is

$$n = \frac{z_{\alpha/2}^2}{4e^2}.$$

Example 11

How large a sample is required if we want to be at least 95% confident that our estimate of p in Example 8 is within 0.02 of the true value?

6.3 Estimating Proportion

Solution

Unlike in Example 10, we shall now assume that no preliminary sample has been taken to provide an estimate of p. Consequently, we can be at least 95% confident that our sample proportion will not differ from the true proportion by more than 0.02 if we choose a sample of size

$$n = \frac{(1.96)^2}{(4)(0.02)^2} = 2401.$$

Comparing the results of Examples 10 and 11, we see that information concerning p, provided by a preliminary sample or from experience, enables us to choose a smaller sample while maintaining our required degree of accuracy.

Content

Problems

Problem 6.1

An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with a standard deviation of 40 hours. If a sample of 30 bulbs has an average life of 780 hours, find a 96% confidence interval for the population mean of all bulbs produced by this firm.

Problem 6.2

The heights of a random sample of 50 college students showed a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. (a) Construct a 98% confidence interval for the mean height of all college students. (b) What can we assert with 98% confidence about the possible size of our error if we estimate the mean height of all college students to be 174.5 centimeters?

Problems

Problem 6.3

A machine produces metal pieces that are cylindrical in shape. A sample of pieces is taken, and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. Find a 99% confidence interval for the mean diameter of pieces from this machine, assuming an approximately normal distribution.

Problem 6.4

The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint: 3.4 2.5 4.8 2.9 3.6 2.8 3.3 5.6 3.7 2.8 4.4 4.0 5.2 3.0 4.8 Assuming that the measurements represent a random sample from a normal population, find a 95% prediction interval for the drying time for the next trial of the paint.

Problems

Problem 6.5

In a random sample of 1000 homes in a certain city, it is found that 228 are heated by oil. Find 99confidence intervals for the proportion of homes in this city that are heated by oil using both methods presented on page 297.

Problem 6.6

(a) A random sample of 200 voters in a town is selected, and 114 are found to support an annexation suit. Find the 96% confidence interval for the fraction of the voting population favoring the suit. (b) What can we assert with 96% confidence about the possible size of our error if we estimate the fraction of voters favoring the annexation suit to be 0.57?

Problems

Problem 6.7

A geneticist is interested in the proportion of African males who have a certain minor blood disorder. In a random sample of 100 African males, 24 are found to be afflicted. (a) Compute a 99% confidence interval for the proportion of African males who have this blood disorder. (b) What can we assert with 99% confidence about the possible size of our error if we estimate the proportion of African males with this blood disorder to be 0.24?