Chapter7 Inferences based on a single sample:

Estimation with Confidence Intervals

In this chapter, our goal is to ____________ the value of an unknown population parameter of interest, such as a population mean and a proportion from a binomial population, and to assess the ____________ of the estimate.

7.1 Identifying the target parameter

Definition 7.1

The unknown population parameter (e.g., mean or proportion) that we are interested in estimating is called the **target parameter.**

Determining the target parameter

Mean, average -- _________________________

Proportion; percentage; fraction; rate ----------- _________________________

A __________________: is a rule or formula that tells us how to use the sample data to calculate a single number that can be used as an estimate of the population parameter.
(the value of a sample statistic is used to estimate the parameter.)

_______________________is a point estimate of population mean μ .

______________________ is a point estimate of population proportion p .

Example1: To estimate the average math test score for all 5^{th} graders in an elementary school, a sample of 25 students yields the average math score is 78,
The sample mean ____ is a point estimate of the average math test score for all 5^{th} graders.

Example2: To estimate the proportion of all voters in a state will vote candidate A, a sample of 1000 voters yields 560 will vote candidate A,
The sample proportion _________ is a point estimate of the proportion of all voters in this state will vote this candidate.

(Since using point estimators to estimate target parameters, we cannot assign any level of certainty with those point estimators. To fix this drawback, we can use interval estimator.)

An __________________ (or _________________) is a formula that tells us how to use sample data to calculate an interval that estimates a population parameter.

The ___________________ is the probability that an interval estimator encloses the population

parameter. That is, the relative frequency with which similar constructed intervals enclose the population parameter when the estimator is used repeatedly a very large number of times.
The ________________ is the confidence coefficient expressed as a percentage.

*If the confident level is 95%, then in the _______ run, _______ of similar constructed confidence intervals will contain population mean μ and _______ will not.

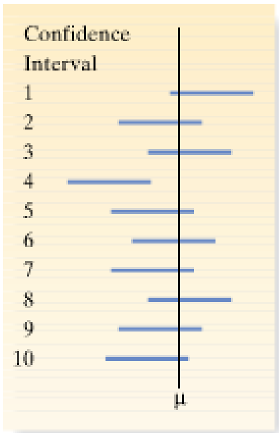


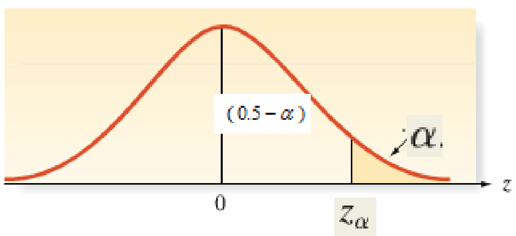
FIGURE 7.3
Confidence intervals for μ: 10 samples

7.2 Large-sample confidence interval for a population mean

First let's introduce _______ notation.

Definition 7.4
The value z_α is defined as the value of the standard normal random variable z such that the area α will lie to its right. In other words, $P(z > z_\alpha) = \alpha$.

(To find z_α, we need use the standard normal curve z table (Table IV. P884). Review the examples in class note CH.5: using z-table to find z values.)
Now we want to find a z value z_α, we need find the area (or probability) between 0 and z_α first, it equals $(0.5-\alpha)$, then we can check this probability $(0.5-\alpha)$ in the entries of the table, then back to corresponding z value.



For example:

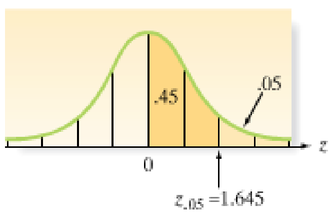


TABLE 7.2 Commonly Used Values of $z_{\alpha/2}$

Confidence Level $100(1-\alpha)$	α	$\alpha/2$	$z_{\alpha/2}$
90%	.10	.05	1.645
95%	.05	.025	1.96
99%	.01	.005	2.575

- Conditions required for a valid large-sample confidence interval for μ :

  1. A ________________ is selected from the target population.
  2. The sample size n is __________. (_______)

- Large-sample 100(1-α)% confidence interval for μ

- **Interpret:** We can be ________ confident that _____________lies between the lower and upper bounds of the confidence interval.

- **Note:** Very wide C.I is not useful, to reduce the width of the C.I, we can:
  1. _____________ sample size.
  2. _____________ confidence level.

Example1. A random sample of 90 bags of a kind bread produced a mean weight is 25.9oz and a standard deviation 2.7oz.

1. Give a point estimate of the mean weight per bag of this kind bread.

2. Find a 90% confidence interval to estimate the mean weight of this kind bread.

3. Interpret the result of the confidence interval.

Example2. Unoccupied seats on flights cause airlines to lose revenue**.** Suppose a large airline wants to estimate its average number of unoccupied seats per flight over the past year. 225 flights are randomly selected and results to mean number of unoccupied seats per flight is 11.6, and standard deviation is 4.1.

a. Give a point estimate of μ , the mean number of unoccupied seats per flight.

b. Calculate a 95% confidence interval to estimate the mean number of unoccupied seats per flight μ .

c. Interpret the result of the confidence interval.

7.3 Small-sample confidence interval for a population mean

When we make the inference based on small sample, _________ can't be applied and the sample standard deviation can't be a good approximation for population standard deviation.

If the sample mean $\bar{x}$ based on a small sample from normal population, We define statistic:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

T-distribution:

1. ______ shaped and symmetric with mean ____;
2. It is more ____________ than standard normal variable z, it is associated with its degree of freedom, _______________.
3. As the degree of freedom ____________, t-curve is close to ____________.

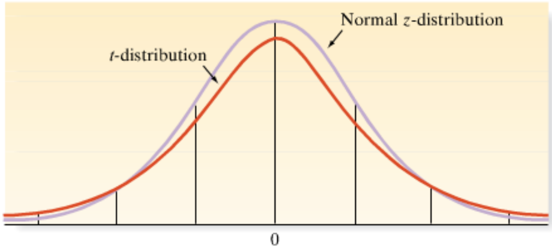


FIGURE 7.7
Standard normal (z) distribution and t-distribution with 4 df

*Since degrees of freedom are related to the sample size n, it is helpful to think of the number of degrees of freedom as the amount of information in the sample available for estimating the target parameter.

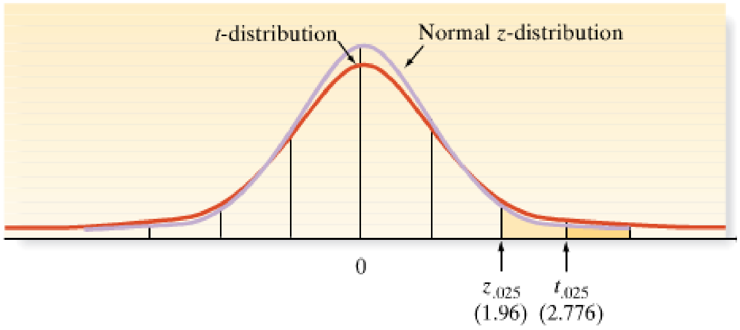
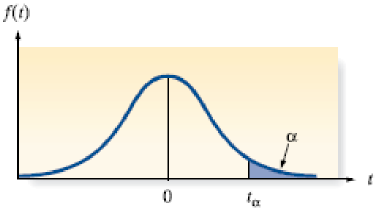


FIGURE 7.8
The $t_{.025}$ value in a t-distribution with 4 df, and the corresponding $z_{.025}$ value

TABLE VI Critical Values of t



Degrees of Freedom	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
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
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
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

Source: This table is reproduced with the kind permission of the Trustees of Biometrika from E. S. Pearson and H. O. Hartley (eds.), *The Biometrika Tables for Statisticians,* Vol. 1, 3d ed., Biometrika, 1966.

Table VI (P796), Critical Values of t.

For example:

- **Conditions required for a valid small-sample confidence interval for** μ :

1. A _____________ sample is selected from the target population.
2. The population has a probability distribution that is approximately _________ .
3. The sample size n is _________. (_________)

- **Small-sample 100(1-α)% confidence interval for μ**

________________________________ where $t_{\alpha/2}$ is based on (n - 1) degree of freedom

Interpret: We can be 100(1-α)% confident that μ lies between the lower and upper bounds of the confidence interval.

Example1. A random sample of 28 light bulbs had a mean life of 497 hours and a standard deviation of 25 hours. Assume the life times of light bulbs are normally distributed. Construct a 90% confidence interval for the mean life μ of all light bulbs of this type. Interpret the result.

1. Give a point estimate of the mean life time of this kind light bulb.

2. Construct a 90% confidence interval to estimate the mean life time of this kind light bulb.

3. Interpret the result of the confidence interval.

Example2. Some quality-control experiments require destructive sampling in order to measure some particular characteristic of the product. The cost of the destructive sampling often dictates small samples.

Suppose a manufacturer of printers for personal computers wishes to estimate the mean number of characters printed before the printhead fails.15 printheads are randomly selected and the mean number of characters printed before the printhead fails is 1.239 millions and standard deviation is 0.193 millions.

Q: Calculate a 95% confidence interval for the number of characters printed before the printhead fails. **Interpret the result**.

7.4 large-sample confidence interval for a population proportion p

We are interested in estimating the ____________ or __________ of some group with certain characteristic. (for example, to estimate the percentage of people in favor of a candidate)

A point estimate of p is the _______________________ .

- Sampling distribution of $\hat{p}$:

1. The mean of the sampling distribution of $\hat{p}$ is p. $\hat{p}$ is unbiased estimator of p. _______

2. the standard deviation of the sampling distribution of $\hat{p}$ is ______________.

3. for large samples, the sampling distribution of $\hat{p}$ is approximately __________. A sample size is considered large if _________________________.

- Conditions required for a valid large-sample confidence interval for p:

1. A _______________ sample is selected from the target population;
2. The sample size n is large. (This condition will be satisfied if both ___________________)

- Large-sample 100(1-α)% confidence interval for p

Interpret: We can be 100(1-α)% confident that ________________lies between the lower and upper bounds of the confidence interval.

Example1. The Bureau of economic and business research at the University of Florida conducts quarterly surveys to gauge consumer sentiment in the sunshine state. Suppose that BEBR randomly samples 484 consumers and finds that 257 are optimistic about the state of the economy.

1. Calculate a point estimate of the proportion of Florida consumers who are optimistic about the state of the economy.

2. Calculate a 90% confidence interval to estimate the proportion of Florida consumers who are optimistic about the state of the economy.

3. Interpret the result.

4. Based on the confidence interval, can BEBR infer that the majority of Florida consumers are optimistic about the economy?

Example2. **Gun Control,** In a survey conducted by Louis Harris of LH research, 1250 U.S. adults were polled regarding their views on banning handgun sales, 650 favored a ban.

1. Calculate a point estimate of the proportion of all U.S. adults in favor of banning handgun sales.

2. Use a 95% confidence interval to estimate the proportion of all U.S. adults in favor of banning handgun sales.

3. Interpret the result.

4. Based on this 95% confidence interval, can we infer that the majority of U.S. adults in favor of banning handgun sales?

7.5 **Determining the Sample size**

How large a sample is necessary to make an accurate estimate for a population mean?

$$\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Margin of error (<u>sampling error</u>):

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

- **Minimum Sample Size Needed for an Interval Estimate of the Population Mean**

Example: A scientist wishes to estimate the average depth of a river. He wants to be 99% confident that the estimate is accurate within 2 feet. From a previous study, the standard deviation of the depths measured was 4.38 feet. What is the minimum sample size the scientist should pick up?

Therefore, to be 99% confident that the estimate is within 2 feet of the true mean depth, the scientist needs at least a sample of ______ measurements.

- **Minimum Sample Size Needed for Interval Estimate of a Population Proportion**

Example1: A researcher wishes to estimate, with 95% confidence, the proportion of people who own a home computer. A previous study shows that 40% of those interviewed had a computer at home. The researcher wishes to be accurate within 2% of the true proportion. Find the minimum sample size necessary.

Example2: The same researcher wishes to estimate the proportion of executives who own a car phone. She wants to be 90% confident and be accurate within 5% of the true proportion.

Find the minimum sample size necessary.

Learning objective of Chapter 7:

1. Determine the target population parameter (mean μ or proportion p).
2. Using z-table and t-table find the critical value ($z_{\alpha/2}$ *and* $t_{\alpha/2}$)
3. Find and calculate a point estimate, a confidence interval estimate of a population mean, give an interpretation of the confidence interval.
4. Find and calculate a point estimate, a confidence interval estimate of a population proportion, give an interpretation of the confidence interval.
5. Determine the minimum sample size needed for an interval estimate of the population mean and population proportion.