8 Statistical Inference

8.1 Estimates for Population Parameters

In statistical inference, the main goal is to predict population parameters depending on sample statistics. In other words, to predict something about the population based on data obtained from samples.

Point Estimate and Interval Estimate

There are two types of estimates for population parameters.

- Point estimate:
  A **point estimate** is a *single number* that is out "best guess" for the parameter.

- Interval estimate:
  An **interval estimate** is an *interval of numbers* within which the parameter value is believed to fall.

Ex 1: Let the population parameter under consideration be the mean height of Texas Tech students.

"Depending on sample data, we predict that the mean height of Texas Tech students is 170 cm".
This is a point estimate.

"Depending on sample data, we predict that the mean height of Texas Tech students is 170 ± 3 cm". i.e. between (167, 173) cm.
This is an interval estimate.

Point Estimate Vs Interval Estimate

- A point estimate does not give us any information as to how accurate our guess could be. i.e. how close the estimate is likely to be to the actual parameter value.

- Hence, an interval estimate is more useful as it incorporates a margin of error which helps us to measure the accuracy of the point estimate.

Point Estimation

The best guess for a population parameter is to use an appropriate sample statistic.

- For a population proportion, use the sample proportion
- For a population mean, use the sample mean

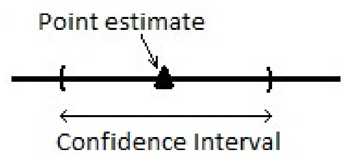
Properties of Point Estimators

1. A good estimator has a sampling distribution that is centered at the population parameter.
2. A good estimator has a small standard error compared to other estimators.

Interval Estimation

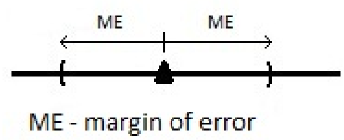
Confidence Interval

A **confidence interval** is an interval containing the most believable values for a parameter. The probability that this method produces an interval that contains the parameter is called the **confidence level**. This is a number chosen to be close to 1, most commonly 0.95.



Margin of Error

The **margin of error** measures how accurate the piont estimate is likely to be in estimating a parameter. It is a multiple of the standard error of the sampling distribution of the estimate, such as $z(se)$.



Therefore, the confidence interval is of the form [point estimate $\pm$ margin of error].

8.2 Constructing a Confidence Interval to Estimate a Population Proportion

In this section, we learn how to construct a confidence interval for a population proportion. The procedure used is as given below.

Confidence Interval for a Population Proportion p

A confidence interval for a population proportion p, using the sample proportion $\hat{p}$ and the standard error se for a sample size of n, is

$$\hat{p} \pm z(se) \text{ where } se = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The value of z depends on the confidence level of interest. For the commonly used confidence levels 95%, 90%, and 99%, the value of z equals 1.96, 1.65 and 2.58 respectively.

The margin of error is given by $me = z(se)$.

For the above confidence interval to be valid, it is required that both $n\hat{p}$ and $n(1-\hat{p})$ are at least 15.

Ex 2: A survey found that 89% of a random sample of 1024 American adults approved of cloning endangered animals. Find the margin of error and the confidence interval for this survey if we want 90% confidence in our estimate of the percentage of American adults who approve of cloning endangered animals.

To calculated the confidence interval, we need to know the sample proportion $\hat{p}$, the relevant z value and the standard error se. To find the standard error we need n. From the question we can figure out that $\hat{p} = 0.89$, and $n = 1024$. The z value corresponding to 90% confidence level is 1.65.

$$se = \sqrt{\frac{0.89(1-0.89)}{1024}} = 0.0098$$

Therefore, the margin of error $me = z(se) = 1.65(0.0098) = 0.0162$.

Then the confidence interval is given by $(\hat{p} - me, \hat{p} + me)$. i.e. $(\hat{p} - z(se), \hat{p} + z(se))$.

$(0.89 - 0.0162, 0.89 + 0.0162) = (0.8738, 0.9062) = (87.38\%, 90.62\%)$

Ex 3: Of 300 items tested, 12 are found to be defective. Construct a 95% confidence interval to estimate the proportion of all such items that are defective.

For this question $\hat{p} = 12/300 = 0.04, n = 300$ and the corresponding z value for 95% confidence interval is $z = 1.96$.

$$se = \sqrt{\frac{0.04(1-0.04)}{300}} = .0113$$

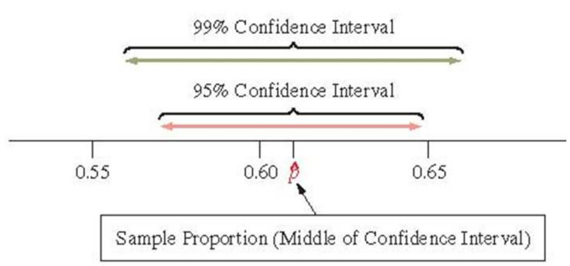
$me = z(se) = 1.96(0.0113) = 0.0222$

Therefore, the confidence interval $= (0.04 - 0.0222, 0.04 + 0.0222) = (0.0178, 0.0622)$

Properties of Margin of Error

The margin of error for a confidence interval:

- Increases as the confidence level increases
- Decreases as the sample size increases

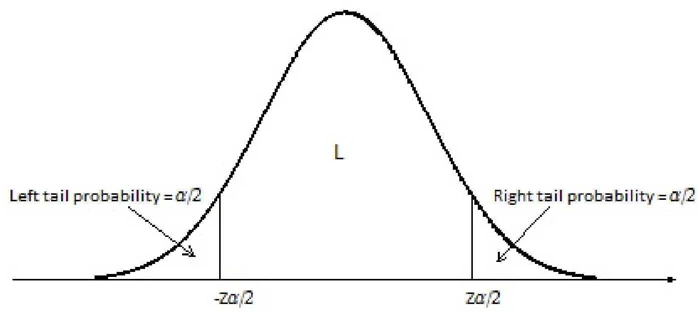


Finding a Confidence Interval for an Arbitrary Confidence Level

Let L be the confidence level that we require expressed in decimal form. This implies that the confidence interval that we create has a probability L to contain the actual population parameter. This implies that we have $1 - L$ error probability.

Let $\alpha = 1 - L$ be the error probability.

Since the distribution is symmetric, the left tail probability is $\dfrac{\alpha}{2}$.



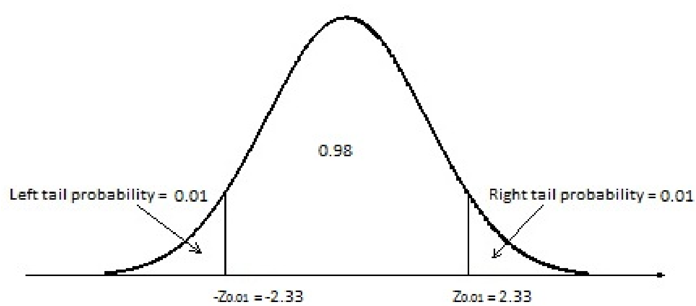
Then $z_{\frac{\alpha}{2}}$ is the absolute value of the z score that has area $\frac{\alpha}{2}$ to its left. Then the confidence interval is given by

$\hat{p} \pm z_{\frac{\alpha}{2}}(se)$ where $se = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Ex 4: Of 346 items tested, 12 are found to be defective. Construct a 98% confidence interval to estimate the proportion of all such items that are defective.

$\hat{p} = 12/346 = 0.0347$, $n = 346$. Therefore $se = \sqrt{\frac{0.0347(1-0.0347)}{346}} = 0.0098$.

To proceed with the calculations, we need the $z_{\frac{\alpha}{2}}$ for a confidence level of 98%.



Therefore, what we need to find is $z_{0.01}$. This can be found be looking up the normal table to find the z-score for which 0.01 area lies to its left and then taking the absolute value. The z-score that corresponds to -2.33. Therefore by taking the absolute value $z_{0.01} = 2.33$.

Now margin of error $me = z_{0.01}(se) = 2.33(0.0098) = 0.0229$.

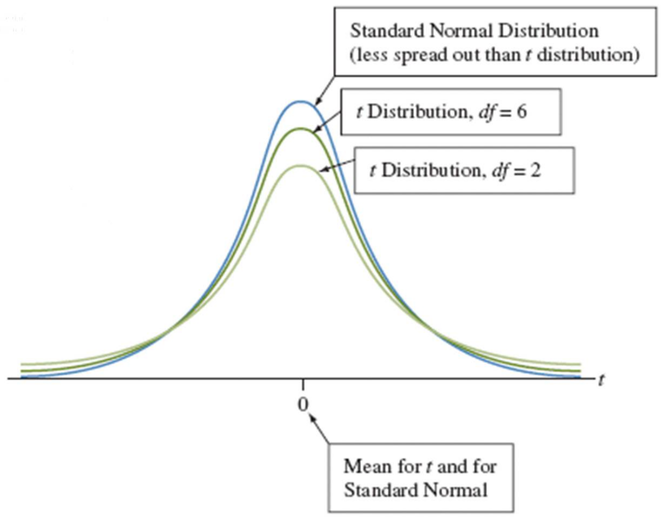
Therefore, the 98% confidence interval $= (0.0347 - 0.0229, 0.0347 + 0.0229) = (0.0118, 0.0576)$.

8.3 Constructing a Confidence Interval to Estimate a Population Mean

The procedure used to construct a confidence interval in this section is similar to the procedure used in the previous section with minor differences. The main difference is that we use a t-distribution for the calculation of the margin of error instead of using the z-distribution.

Properties of the t-distribution

- The t-distribution is bell shpaed and symmetric about 0.
- The probabilities depend on the degrees of freedom, $df = n - 1$. The t-distribution has a slightly different shape for each distinct value of df, and different t-scores apply for each df value.
- The t-distribution has thicker tails and is more spread out than the standard normal distribution (z-distribution).
- When df is 30 or more, the t-distribution is nearly identical to the z-distribution.



Reading the t-table

To read the t-table, we need $df = n - 1$ and the confidence level. Then read the t-score as given below.

- Select the column for the relevant confidence level
- Select the row for the relevant degree of freedom, df
- Read the t-score

Ex 5: Find the t-score for a sample of size 7 and for a confidence level 95%.

$df = n - 1 = 7 - 1 = 6$, Confidence level = 95%.

	Confidence Level					
	80%	90%	95%	98%	99%	99.8%
df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
1	3.078	6.314	12.706	31.821	63.657	318.3
...						
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785

Therefore, from the table, the corresponding t-score is 2.447.

Note: A confidence interval for a population mean is also of the form [point estimate ± margin of error].

Confidence Interval for a Population Mean μ

A confidence interval for the population mean μ, using the sample mean $\bar{x}$ and the standard error se for a sample of size n, is
$\bar{x} \pm t(se)$ where $se = \dfrac{s}{\sqrt{n}}$
(s - sample standard deviation)

The value of t depends on the degree of freedom $df = n - 1$ and the confidence level of interest. This t-score can be read from the t-table as explained above.

The margin of error is given by $me = t(se)$.

Ex 6: Thirty randomly selected students took the statistics final. If the sample mean was 82 and the standard deviation was 12.2, find the margin of error and construct a confidence interval with a confidence level of 99% for the mean score of all students.

To construct a confidence interval for a population mean we need to know $\bar{x}$, t, and se.

From the data we can figure out that $\bar{x} = 82$, $n = 30$ and $s = 12.2$. For $df = n - 1 = 29$ and a confidence level of 99%, the table gives $t = 2.756$.

Then $se = \dfrac{s}{\sqrt{n}} = \dfrac{12.2}{\sqrt{30}} = 2.227$
Therefore, the margin of error $me = t(se) = 2.756(2.227) = 6.138$

Therefore, the confidence interval $= (\bar{x} - me, \bar{x} + me) = (82 - 6.138, 82 + 6.138) = (75.86, 88.14)$

Ex 6: A sociologist develops a test to measure attitudes about public transportation, and 27 randomly selected subjects are given the test. Their mean score is 76.2 and their standard deviation is 21.4. Construct the 95% confidence interval for the mean score of all such subjects.

$\bar{x} = 76.2$, $n = 27$, and $s = 21.4$.

$df = n - 1 = 26$. Therefore, for a confidence level of 95%, $t = 2.056$

$se = \dfrac{s}{\sqrt{n}} = \dfrac{21.4}{\sqrt{27}} = 4.12$
Therefore, $me = t(se) = 2.056(4.12) = 8.47$

Therefore, the confidence interval $= (76.2 - 8.47, 76.2 + 8.47) = (67.7, 84.7)$.