Interval Estimation

Carlos Hurtado
churtado@richmond.edu

Robins School of Business
University of Richmond

Oct 2, 2019

On the Agenda

1. Population Mean: σ Known
2. Population Mean: σ Unknown
3. Determining the Sample Size
4. Population Proportion

On the Agenda

1. Population Mean: σ Known
2. Population Mean: σ Unknown
3. Determining the Sample Size
4. Population Proportion

Population Mean: σ Known

- A point estimator cannot be expected to provide the exact value of the population parameter
- An interval estimate can be computed by adding and subtracting a margin of error to the point estimate:

$$\text{Point Estimate } + / - \text{ Margin of Error}$$

- The purpose of an interval estimate is to provide information about how close the point estimate is to the value of the parameter

Population Mean: σ Known

- To develop an interval estimate of a population mean, the margin of error must be computed using either
  - the population standard deviation σ, or
  - the sample standard deviation s
- σ is rarely known exactly, but often a good estimate can be obtained based on historical data
- We refer to such cases as the σ known case

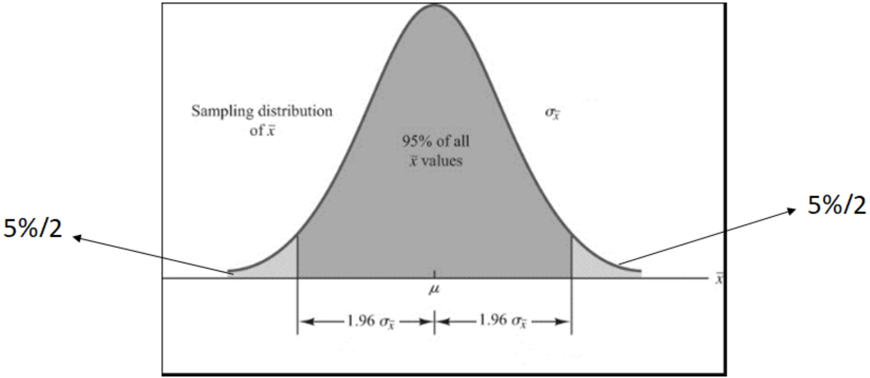
Population Mean: σ Known

- Let's assume we want 95% confidence that the value of the sample mean is within the margin of error
- By the central limit theorem, $\bar{x}$ has a normal distribution
- We want to find a critical value (cv) such that:

$$P(|\bar{x} - \mu| \leq cv) = 95\%$$

- This cv will give us the margin of error

Population Mean: σ Known



Population Mean: σ Known

- Interval estimate of μ

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where:

$\bar{x}$ is the sample mean

$1 - \alpha$ is the confidence level

$z_{\alpha/2}$ is the z value providing an area of $\alpha/2$ in the lower tail of the standard normal distribution

σ is the population standard deviation

n is the sample size

Population Mean: σ Known

- Adequate Sample Size
  - In most applications, a sample size of $n \geq 30$ is adequate
  - If the population distribution is highly skewed or contains outliers, a sample size of 50 or more is recommended

On the Agenda

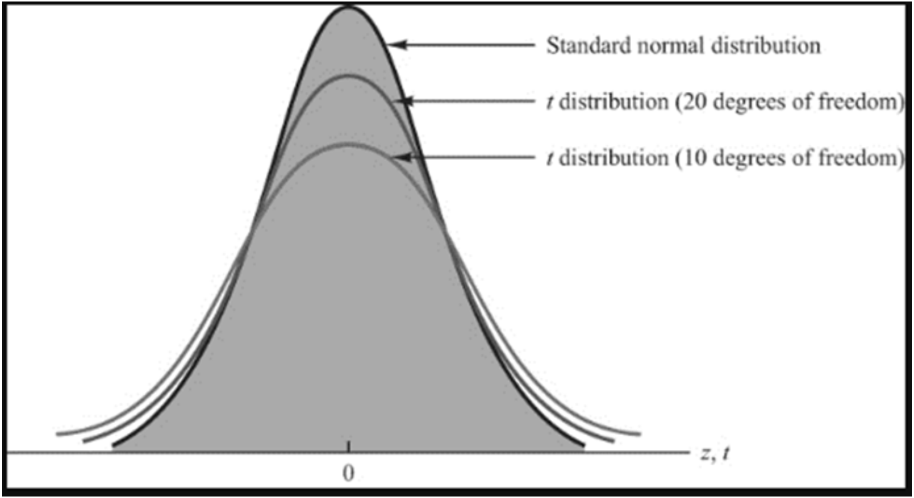
Population Mean: σ Unknown

- In many cases, the population standard deviation σ is unknown
- We use the sample standard deviation s to estimate σ
- This is the $\underline{\sigma \text{ unknown}}$ case
- In this case, the interval estimate for μ is based on the t distribution

Population Mean: σ Unknown

- The t distribution is a family of similar probability distributions
- A specific t distribution depends on a parameter known as the degrees of freedom
- Degrees of freedom refer to the number of independent pieces of information that go into the computation of s
- As the degrees of freedom increase, the difference between the t distribution and the standard normal probability distribution becomes smaller and smaller

Population Mean: σ Known



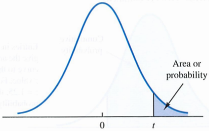
Population Mean: σ Unknown

- For more than 100 degrees of freedom, the standard normal z value provides a good approximation to the t value

- The standard normal z values can be found in the infinite degrees row (labeled ∞) of the t distribution table

Population Mean: σ Known

TABLE 2 *t* DISTRIBUTION



Entries in the table give t values for an area or probability in the upper tail of the t distribution. For example, with 10 degrees of freedom and a .05 area in the upper tail, $t_{.05} = 1.812$.

Degrees of Freedom	Area in Upper Tail .20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
10	.879	1.372	1.812	2.228	2.764	3.169
11	.876	1.363	1.796	2.201	2.718	3.106
12	.873	1.356	1.782	2.179	2.681	3.055
13	.870	1.350	1.771	2.160	2.650	3.012
14	.868	1.345	1.761	2.145	2.624	2.977
15	.866	1.341	1.753	2.131	2.602	2.947
16	.865	1.337	1.746	2.120	2.583	2.921
17	.863	1.333	1.740	2.110	2.567	2.898
18	.862	1.330	1.734	2.101	2.552	2.878
19	.861	1.328	1.729	2.093	2.539	2.861
20	.860	1.325	1.725	2.086	2.528	2.845
21	.859	1.323	1.721	2.080	2.518	2.831
22	.858	1.321	1.717	2.074	2.508	2.819
23	.858	1.319	1.714	2.069	2.500	2.807
24	.857	1.318	1.711	2.064	2.492	2.797
25	.856	1.316	1.708	2.060	2.485	2.787
26	.856	1.315	1.706	2.056	2.479	2.779
27	.855	1.314	1.703	2.052	2.473	2.771
28	.855	1.313	1.701	2.048	2.467	2.763
29	.854	1.311	1.699	2.045	2.462	2.756

Population Mean: σ Unknown

- Interval estimate of μ

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where:

$\bar{x}$ is the sample mean

$1 - \alpha$ is the confidence level

$t_{\alpha/2}$ is the t value providing an area of $\alpha/2$ in the upper tail of the t distribution with n-1 degrees of freedom

s is the sample standard deviation

n is the sample size

Population Mean: σ Unknown

Example: Population Mean and σ Unknown

- Credit card debt for the population of US households
- The credit card balances of a sample of 70 households provided:
  - a mean credit card debt of \$9,312
  - with a sample standard deviation of \$4,007
- What is the 95% confidence interval estimate of the mean credit card debt for the population of US households?

Example: Population Mean and σ Unknown

- At 95% confidence, $\alpha = 5\%$, and $\alpha/2 = 2.5\%$
- We need the $t_{0.025}$ with $n-1 = 69$ degrees of freedom

Degrees of Freedom	Area in Upper Tail					
	.20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	.848	1.296	1.671	2.000	2.390	2.660
61	.848	1.296	1.670	2.000	2.389	2.659
62	.847	1.295	1.670	1.999	2.388	2.657
63	.847	1.295	1.669	1.998	2.387	2.656
64	.847	1.295	1.669	1.998	2.386	2.655
65	.847	1.295	1.669	1.997	2.385	2.654
66	.847	1.295	1.668	1.997	2.384	2.652
67	.847	1.294	1.668	1.996	2.383	2.651
68	.847	1.294	1.668	1.995	2.382	2.650
69	.847	1.294	1.667	1.995	2.382	2.649
⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	.846	1.291	1.662	1.987	2.368	2.632
91	.846	1.291	1.662	1.986	2.368	2.631
92	.846	1.291	1.662	1.986	2.368	2.630
93	.846	1.291	1.661	1.986	2.367	2.630
94	.845	1.291	1.661	1.986	2.367	2.629
95	.845	1.291	1.661	1.985	2.366	2.629
96	.845	1.290	1.661	1.985	2.366	2.628
97	.845	1.290	1.661	1.985	2.365	2.627
98	.845	1.290	1.661	1.984	2.365	2.627
99	.845	1.290	1.660	1.984	2.364	2.626
100	.845	1.290	1.660	1.984	2.364	2.626
∞	.842	1.282	1.645	1.960	2.326	2.576

Example: Population Mean and σ Unknown

- The 95% confidence interval for the mean credit card debt is

$$\bar{x} \pm t_{0.025} \frac{s}{\sqrt{n}}$$

  or

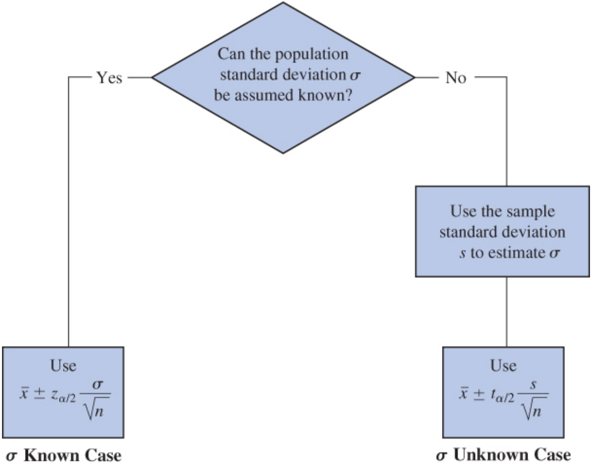
$$9,312 \pm 1.995 \frac{4,007}{\sqrt{n}} = 9,312 \pm 955$$

- In other words: We are 95% confident that the mean credit card debt for the population of US households is between \$8357 and \$10267

Population Mean: σ Unknown

- Adequate Sample Size
  - In most applications, a sample size of $n \geq 30$ is adequate
  - If the population distribution is highly skewed or contains outliers, a sample size of 50 or more is recommended

Summary


Can the population standard deviation σ be assumed known?
Yes
No
Use the sample standard deviation s to estimate σ
Use $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
Use $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$
σ Known Case
σ Unknown Case

On the Agenda

Determining the Sample Size

- Let E denote the desired margin of error
- E is the amount added to and subtracted from the point estimate to obtain an interval estimate
- If a desired margin of error is selected prior to sampling
  - We can determine the sample size necessary to satisfy E!

Determining the Sample Size

- Margin of Error

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Necessary Sample Size

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

Determining the Sample Size

- The Necessary Sample Size equation requires a value for the population standard deviation σ
- If σ is unknown, a preliminary or planning value for s can be used in the equation
  1. Use the estimate of the population standard deviation computed in a previous study
  2. Use a pilot study to select a preliminary study and use the sample standard deviation from the study
  3. Use judgment or a "best guess" for the value of σ

On the Agenda

Population Proportion

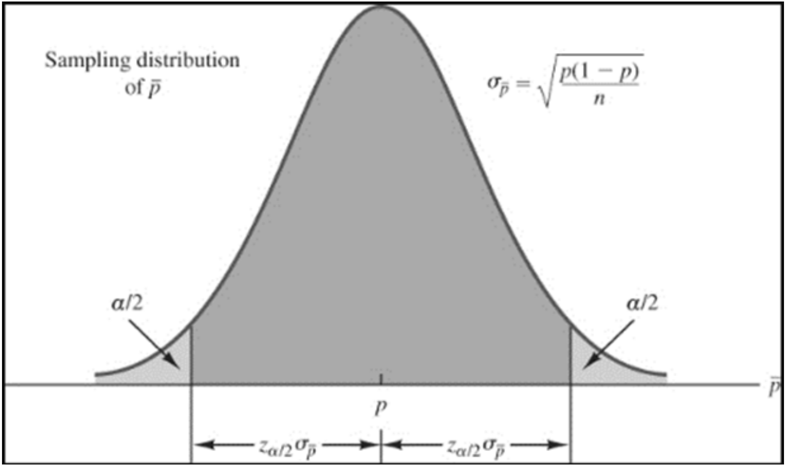
- The general form of an interval estimate of a population proportion is:

$$\bar{p} \pm E$$

- The sampling distribution of $\bar{p}$ plays a key role in computing the margin of error for this interval estimate

- The sampling distribution of $\bar{p}$ can be approximated by a normal distribution whenever $np \geq 5$ and $n(1-p) \geq 5$

Population Proportion



Population Proportion

- The unbiased estimator of p is $\bar{p}$, hence

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

where:

$\bar{p}$ is the sample estimator of the proportion

$1-\alpha$ is the confidence level

$z_{\alpha/2}$ is the z value providing an area of $\alpha/2$ in the lower tail of the standard normal probability distribution

n is the sample size

Sample Size and Population Proportion

- Margin of Error

$$E = z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

- Solving for the necessary sample size n, we get

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

- However, $\bar{p}$ will not be known until after we have selected the sample.
- We will use the planning value p^* for $\bar{p}$

Sample Size and Population Proportion

- Necessary Sample Size

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

- The planning value p^* can be chosen by
  1. Using the sample proportion from a previous sample of the same or similar units, or
  2. Selecting a preliminary sample and using the sample proportion from this sample
  3. Using judgment or a “best guess” for a p^* value
  4. Otherwise, using $p^* = 0.5$

Example: Population Proportion

- Survey of women golfers
- A national survey of 900 women golfers was conducted to learn how women golfers view their treatment at golf courses in United States
- The survey found that 396 of the women golfers were satisfied with their treatment at golf courses
- What is the 95% confidence interval for the proportion of women golfers satisfied with their treatment at golf courses?

Example: Population Proportion

- Survey of women golfers

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

where: $n = 900$, $\bar{p} = 396/900 = 0.44$, and $z_{\alpha/2} = 1.96$

- The 95% confidence interval is:

$$0.44 \pm 1.96\sqrt{\frac{0.44(1-0.44)}{900}} = 0.44 \pm 0.0324$$

- Survey results enable us to state with 95% confidence that between 40.76% and 47.24% of all women golfers are satisfied

Example: Population Proportion

- Survey of women golfers
- Suppose the survey director wants to estimate the population proportion with a margin of error of 0.025 at 95% confidence
- How large a sample size is needed to meet the required precision?
- Note: A previous sample of similar units yielded 0.44 for the sample proportion

Example: Population Proportion

- Survey of women golfers

$$E = z_{\alpha/2}\sqrt{\frac{p^*(1-p^*)}{n}} = 0.025$$

- At the 95% confidence, $z_{0.0125} = 1.96$. We know that $p^* = 0.44$

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p^*(1-p^*) = \left(\frac{1.96}{0.025}\right)^2 0.44(0.56) = 1,514.5$$

- A sample of $1,515$ is needed to reach a desired precision of $\pm\ 0.025$ at 95% confidence