Session 15 - 16

Statistical Estimation

Prof. Jigar M. Shah

Statistical Estimation

- Introduction
  - Point Estimation
  - Interval Estimation
- Interval Estimation of Population Proportion
  - Interval Estimation of Population Proportion
  - Sample Size Determination
- Interval Estimation of Population Mean
  - When Population Standard Deviation, Sigma (σ), is Known
  - When Population Standard Deviation, Sigma (σ), is Unknown
  - Sample Size Determination

Statistical Estimation

Introduction

- Point Estimation

Point Estimation	Estimating the value of a population parameter based on the value of the corresponding sample statistic
Point Estimator	Sample statistic corresponding to a population parameter
Point Estimate	Value of the sample statistic
Parameter Value	Value of the population parameter

is point estimator of

Sample Statistic	Measure of the Characteristic	Population Parameter
$\bar{x}$	Mean	μ
s	Standard Deviation	σ
p	Proportion of Success	$\bar{p}$

A sample statistic is the point estimator of the corresponding population parameter

Statistical Estimation

Introduction

- Interval Estimation
  - The point estimate provided by the point estimator cannot be expected to provide the exact value of the population parameter
  - Need to include a margin of error with the point estimate in order to get an interval estimate

Interval Estimate	=	Point Estimate	±	Margin of Error

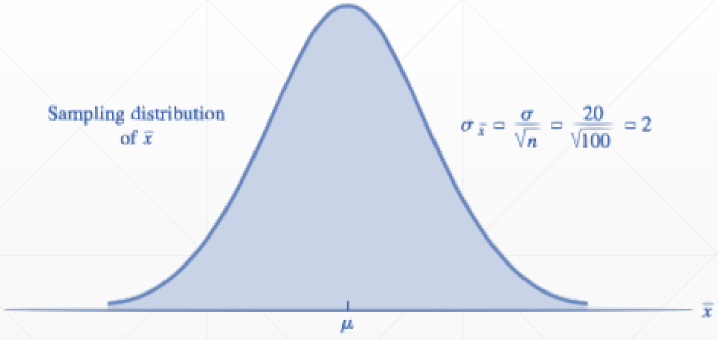
Interval Estimation	Estimating the interval of possible values of a population parameter based on the value of the corresponding sample statistic & the margin of error

Interval Estimate of Population Mean	=	$\bar{x}$	±	Margin of Error
Interval Estimate of Population Proportion	=	$\bar{p}$	±	Margin of Error

Statistical Estimation

Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Known
  - Sampling distribution of $\bar{x}$ shows how the different possible values of $\bar{x}$ are distributed around the population mean μ, & since population standard deviation, sigma (σ), is known, standard error $\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right)$
  - 95% of the values of any normally distributed random variable are within ±1.96 standard deviations of the mean (using data from standard normal table)
  - If σ = 20, n = 100, $\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right) = \left(\frac{20}{\sqrt{100}}\right)$ = 2, and hence, 95% of the values of $\bar{x}$ will be within ±1.96 standard deviations, i.e., within ±1.96 × 2 = ±3.92 of the population mean μ

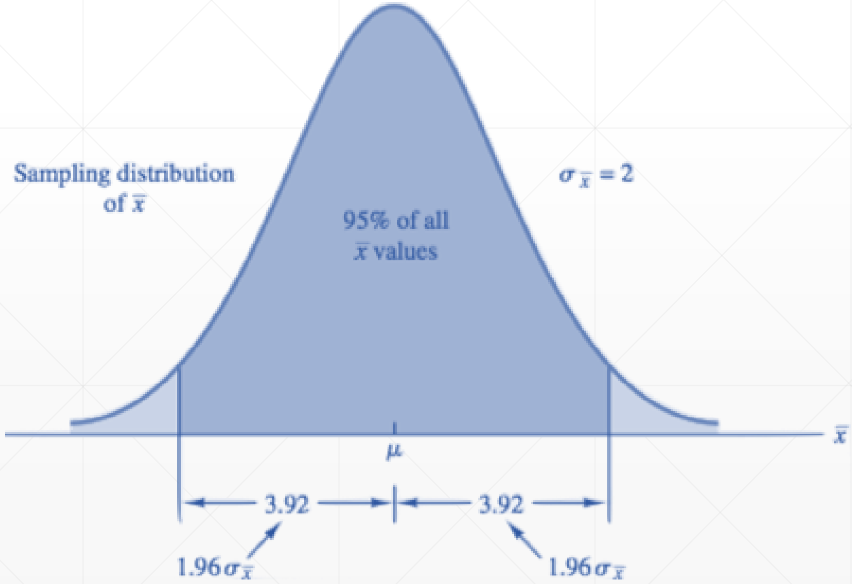


  - If margin of error is equal to ±3.92, Interval Estimate of μ = $\bar{x}$ ± Margin of Error = $\bar{x}$ ± 3.92

Statistical Estimation

Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Known
  - Any sample mean $\bar{x}$ within the darkly shaded region will provide an interval that contains the population mean μ
  - Since 95% of all possible sample means are in the darkly shaded region, 95% of all the intervals formed by subtracting 3.92 from $\bar{x}$ and adding 3.92 to $\bar{x}$ will include the population mean μ
  - Since 95% of all the intervals formed using $\bar{x} \pm 3.92$ will contain will include the population mean μ, it can be said that one is 95% confident that the interval estimate $\bar{x} \pm 3.92$ includes the population mean μ



Statistical Estimation

Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Known
  - If $\bar{x} = 82$,

$$\text{Interval Estimate of } \mu = \bar{x} \pm \text{Margin of Error} = 82 \pm 3.92 = 78.08 \text{ to } 85.92$$

  - Thus, one can be 95% confident that the interval estimate 78.08 to 85.92 includes the population mean μ i.e., the interval has been established at 95% confidence value
  - Value 0.95 is referred to as the confidence coefficient
  - Interval 78.08 to 85.92 is called 95% confidence interval

Statistical Estimation

Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Known

Interval Estimate of Population Mean, μ, When Population Standard, σ, is Known

Interval Estimate = $\bar{x} \pm$ Margin of Error = $\bar{x} \pm (z_{\alpha/2} \times$ Standard Error) = $\bar{x} \pm (z_{\alpha/2} \times \sigma_{\bar{x}}) = \bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

where $(1 - \alpha)$ is the confidence coefficient

$z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ is the margin of error

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

Confidence Level	α	$\alpha/2$	$z_{\alpha/2}$
90%	0.10	0.050	1.645
95%	0.05	0.025	1.960
99%	0.01	0.005	2.576

Values of $z_{\alpha/2}$ for the most commonly used Confidence Intervals

Statistical Estimation

Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Known
  - A simple random sample of 40 items resulted in a sample mean of 25. The population standard deviation is 5.
    - What is the standard error of the mean?
    - At 95% confidence, what is the margin of error?
  - A simple random sample of 50 items from a population with std. dev. 6 resulted in a sample mean of 32.
    - Provide a 90% confidence interval for the population mean.
    - Provide a 95% confidence interval for the population mean.
    - Provide a 99% confidence interval for the population mean.

Statistical Estimation

Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Known
  - The mean monthly rent at assisted-living facilities was reported to have increased 17% over the last five years to $3486. Assume this cost estimate is based on a sample of 120 facilities and, from past studies, it can be assumed that the population standard deviation is $650. Develop the 90% confidence interval estimate, 95% confidence interval estimate, and 99% confidence interval estimate of the population mean monthly rent. What happens to the width of the confidence interval as the confidence level is increased? Does this seem reasonable? Explain.

Statistical Estimation

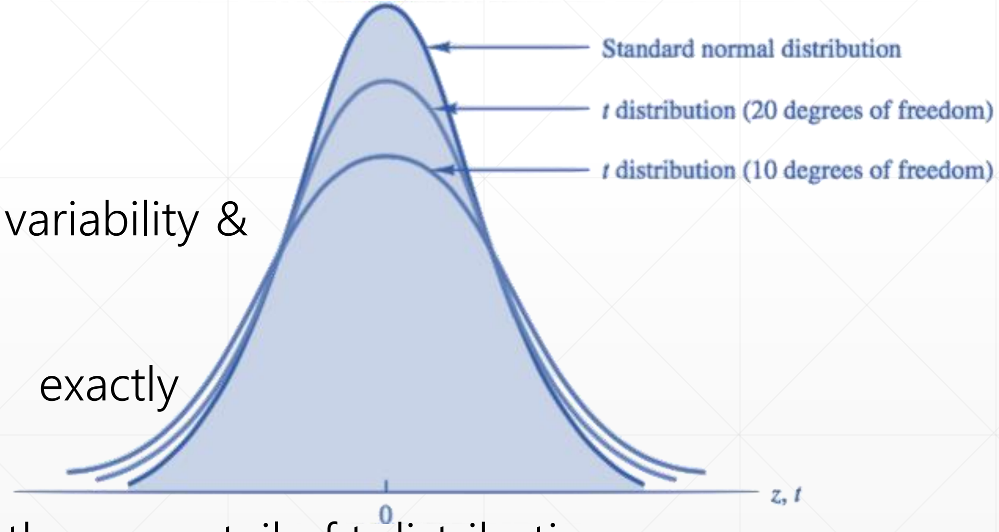
Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Unknown
  - Sample is used to estimate both the mean, μ, as well as the standard deviation, σ, of the population
  - Makes use of a probability distribution known as t distribution
  - t Distribution - a family of similar probability distributions, with a specific t distribution depending on a parameter known as the degrees of freedom
  - Mathematical development of t distribution is based on the assumption of a normal distribution for the population from which the sample is drawn
  - t distribution with one degree of freedom is unique, as is the t distribution with two degrees of freedom, with three degrees of freedom, and so on

Statistical Estimation

Interval Estimation of Population Mean

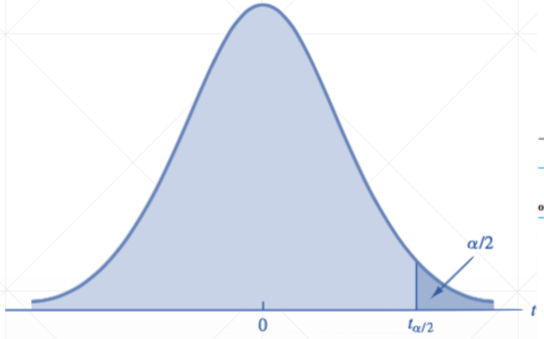
- When Population Standard Deviation, Sigma (σ), is Unknown
  - As the no. of degrees of freedom increases, the difference between the t distribution & the standard normal distribution becomes smaller & smaller
  - Mean of t distribution is zero
  - t distribution with more degrees of freedom exhibits less variability & more closely resembles the standard normal distribution
  - t distribution with an infinite degrees of freedom exactly resembles the z distribution
  - $t_{\alpha/2}$ represents value of t providing an area equal to $\alpha/2$ in the upper tail of t distribution



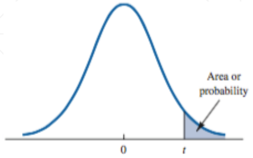
Statistical Estimation

Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Unknown



If the degrees of freedom exceed 100, the infinite degrees of freedom row can be used to approximate the actual t value; in other words, for more than 100 degrees of freedom, the standard normal z value provides a good approximation to the t value



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
30	.854	1.310	1.697	2.042	2.457	2.750
31	.853	1.309	1.696	2.040	2.453	2.744
32	.853	1.309	1.694	2.037	2.449	2.738
33	.853	1.308	1.692	2.035	2.445	2.733
34	.852	1.307	1.691	2.032	2.441	2.728
35	.852	1.306	1.690	2.030	2.438	2.724
36	.852	1.306	1.688	2.028	2.434	2.719
37	.851	1.305	1.687	2.026	2.431	2.715
38	.851	1.304	1.686	2.024	2.429	2.712
39	.851	1.304	1.685	2.023	2.426	2.708
40	.851	1.303	1.684	2.021	2.423	2.704
41	.850	1.303	1.683	2.020	2.421	2.701
42	.850	1.302	1.682	2.018	2.418	2.698
43	.850	1.302	1.681	2.017	2.416	2.695
44	.850	1.301	1.680	2.015	2.414	2.692
45	.850	1.301	1.679	2.014	2.412	2.690
46	.850	1.300	1.679	2.013	2.410	2.687
47	.849	1.300	1.678	2.012	2.408	2.685
48	.849	1.299	1.677	2.011	2.407	2.682
49	.849	1.299	1.677	2.010	2.405	2.680
50	.849	1.299	1.676	2.009	2.403	2.678
51	.849	1.298	1.675	2.008	2.402	2.676
52	.849	1.298	1.675	2.007	2.400	2.674
53	.848	1.298	1.674	2.006	2.399	2.672
54	.848	1.297	1.674	2.005	2.397	2.670
55	.848	1.297	1.673	2.004	2.396	2.668
56	.848	1.297	1.673	2.003	2.395	2.667
57	.848	1.297	1.672	2.002	2.394	2.665
58	.848	1.296	1.672	2.002	2.392	2.663
59	.848	1.296	1.671	2.001	2.391	2.662
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
70	.847	1.294	1.667	1.994	2.381	2.648
71	.847	1.294	1.667	1.994	2.380	2.647
72	.847	1.293	1.666	1.993	2.379	2.646
73	.847	1.293	1.666	1.993	2.379	2.645
74	.847	1.293	1.666	1.993	2.378	2.644
75	.846	1.293	1.665	1.992	2.377	2.643
76	.846	1.293	1.665	1.992	2.376	2.642
77	.846	1.293	1.665	1.991	2.376	2.641
78	.846	1.292	1.665	1.991	2.375	2.640
79	.846	1.292	1.664	1.990	2.374	2.639
80	.846	1.292	1.664	1.990	2.374	2.639
81	.846	1.292	1.664	1.990	2.373	2.638
82	.846	1.292	1.664	1.989	2.373	2.637
83	.846	1.292	1.663	1.989	2.372	2.636
84	.846	1.292	1.663	1.989	2.372	2.636
85	.846	1.292	1.663	1.988	2.371	2.635
86	.846	1.291	1.663	1.988	2.370	2.634
87	.846	1.291	1.663	1.988	2.370	2.634
88	.846	1.291	1.662	1.987	2.369	2.633
89	.846	1.291	1.662	1.987	2.369	2.632
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

Statistical Estimation

Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Unknown
  - For a t distribution with 16 degrees of freedom, find the area, or probability, in each region:
    - To the right of 2.120
    - To the left of 1.337
    - To the left of -1.746
    - To the right of 2.583
    - Between -2.120 & 2.120
    - Between -1.746 & 1.746
  - Find the t values for each of the following cases:
    - Upper tail area of 0.025 with 12 degrees of freedom
    - Lower tail area of 0.05 with 50 degrees of freedom
    - Upper tail area of 0.01 with 30 degrees of freedom
    - Where 90% of the area falls between these two t values with 25 degrees of freedom
    - Where 95% of the area falls between these two t values with 45 degrees of freedom

Statistical Estimation

Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Unknown

Interval Estimate of Population Mean, μ, When Population Standard, σ, is Unknown			
Interval Estimate =	$\bar{x} \pm$ Margin of Error = $\bar{x} \pm$ ($t_{\alpha/2}$ × Estimate of Standard Error) = $\bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$		
where s	is the sample standard deviation	$t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$	is the margin of error
$(1 - \alpha)$	is the confidence coefficient		
$t_{\alpha/2}$	is the t value providing an area of $\alpha/2$ in the upper tail of the corresponding t distribution with $n - 1$ degrees of freedom		

- The mean credit card balance for a sample of 70 households was ₹9312 with a sample standard deviation of ₹4007. Provide a 95% confidence interval for the population mean.

Statistical Estimation

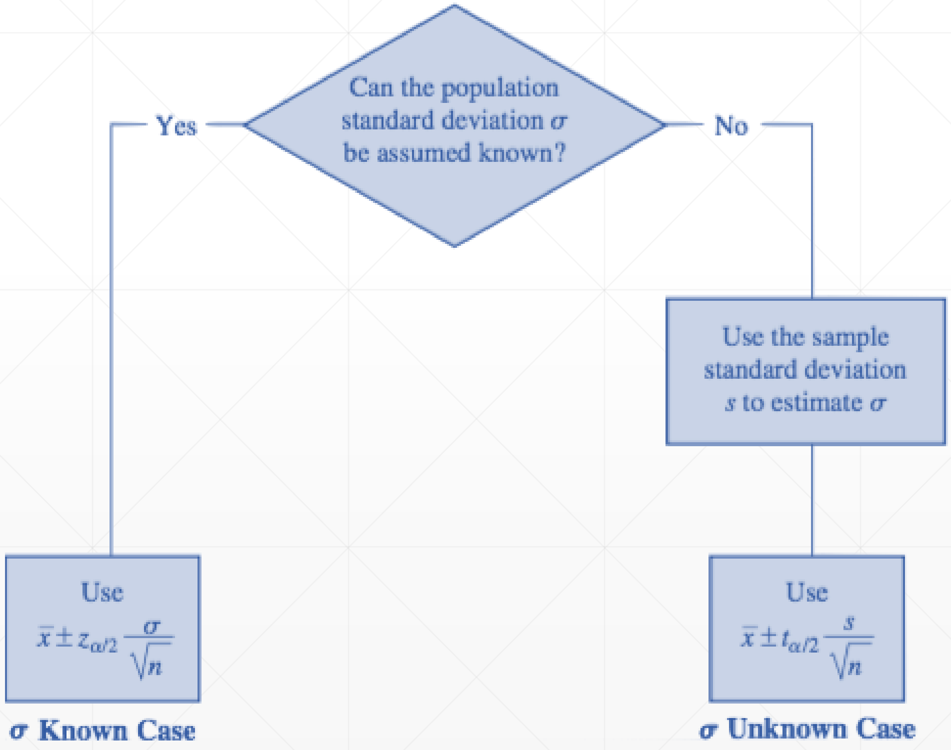
Interval Estimation of Population Mean

- When Population Standard Deviation, Sigma (σ), is Unknown
  - A simple random sample of 50 items from a population with std. dev. 6 resulted in a sample mean of 32.
    - Provide a 90% confidence interval for the population mean.
    - Provide a 95% confidence interval for the population mean.
    - Provide a 99% confidence interval for the population mean.
  - A simple random sample with n = 54 provided a mean of 22.5 & std. dev. of 4.4. Develop the 90%, 95% & 99% confidence intervals for the population mean. What happens to the margin of error & the confidence interval as the confidence level is increased?
  - A sample of 65 weekly sales activity reports shows a mean of 19.5 customer contacts per week with a std. dev. of 5.2. Provide 90% & 95% confident intervals for population mean no. of weekly customer contacts.

Statistical Estimation

Interval Estimation of Population Mean

Summary of Interval Estimation Procedures for a Population Mean



Statistical Estimation

Interval Estimation of Population Mean

- Sample Size Determination

Sample Size for an Interval Estimate of Population Mean, μ

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

where

E	is the desired margin of error	n	is the required sample size
$z_{\alpha/2}$	is the z value providing an area of $\alpha/2$ in the upper tail of the standard normal probability distribution		
σ	is the population standard deviation		
σ	$\cong$ Range ÷ 4, if the population standard deviation is unknown where Range is the difference between the largest & the smallest data values		

Statistical Estimation

Interval Estimation of Population Mean

- Sample Size Determination
  - How large a sample should be selected to provide a 95% confidence interval with a margin of error of 10? Assume the population standard deviation is 40.
  - The range for a set of data is 36.
    - What is the estimated value (planning value) for the population standard deviation?
    - At 95% confidence, how large a sample would provide a margin of error of 3?
    - At 95% confidence, how large a sample would provide a margin of error of 2?

Statistical Estimation

Interval Estimation of Population Proportion

- Interval Estimation of Population Proportion

Interval Estimate of Population Proportion, p

Interval Estimate = $\bar{p} \pm$ Margin of Error = $\bar{p} \pm (z_{\alpha/2} \times \text{Standard Error}) = \bar{p} \pm (z_{\alpha/2} \times \sigma_{\bar{p}}) = \bar{p} \pm z_{\alpha/2}\left(\sqrt{\frac{p(1-p)}{n}}\right)$

where $(1-\alpha)$ is the confidence coefficient

$z_{\alpha/2}\left(\sqrt{\frac{p(1-p)}{n}}\right)$ is the margin of error

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

Since p is unknown, $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \cong \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ and thus, Margin of Error = $z_{\alpha/2}\left(\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right)$

Interval Estimate = $\bar{p} \pm z_{\alpha/2}\left(\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right)$

Statistical Estimation

Interval Estimation of Population Proportion

- Interval Estimation of Population Proportion
  - A simple random sample of 400 individuals provides 100 Yes responses to a question.
    - What is the point estimate of the proportion of population that would provide Yes responses?
    - What is the estimate of standard error of the proportion?
    - Compute the 95% confidence interval for population proportion.
  - According to statistics report, a surprising number of motor vehicles are not covered by insurance. Sample results, consistent with the report, showed 46 of 200 vehicles were not covered by insurance.
    - What is the point estimate of the proportion of vehicles not covered by insurance?
    - Develop a 90% confidence interval for the population proportion.

Statistical Estimation

Interval Estimation of Population Proportion

- Sample Size Determination

Sample Size for an Interval Estimate of Population Proportion, p

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

where

E	is the desired margin of error	n	is the required sample size
$z_{\alpha/2}$	is the z value providing an area of $\alpha/2$ in the upper tail of the standard normal probability distribution		
$\bar{p}$	is the sample proportion		
$\bar{p}$	$\sim$ 0.5, if it is not known		

Statistical Estimation

Interval Estimation of Population Proportion

- Sample Size Determination
  - In a survey, the point estimate of the population proportion is taken as 0.35. How large a sample should be taken to provide a 95% confidence interval with a margin of error of .05?
  - At 95% confidence, how large a sample should be taken to obtain a margin of error of 0.03 for the estimation of a population proportion? Assume that past data are not available for developing the estimate of proportion.

Thank You

Prof. Jigar M. Shah