Session 17 - 25

Hypothesis Testing

Prof. Jigar M. Shah

Session 17 - 25

Hypothesis Testing

- Concepts
  - Introduction
  - Null & Alternate Hypothesis
  - Significance Level
  - Type I & Type II Errors
  - Power of Hypothesis Test
  - One-Tailed & Two-Tailed Tests of Hypothesis
  - Exercise
- One Sample Tests
  - Procedure for Hypothesis Testing
  - Hypothesis Testing of Means
    - When Population Standard Deviation, Sigma (σ), is Known
    - When Population Standard Deviation, Sigma (σ), is Unknown
  - Hypotheses Testing of Proportion of Success

Hypothesis Testing

Concepts

- Introduction

Hypothesis	An assumption made about a population parameter
Hypothesis Testing	Testing the validity of hypothesis i.e. testing the validity of the assumption made about a population parameter • Making an assumption of the population parameter (hypothesized population parameter) • Gathering sample data & computing the sample statistic • Determining the difference between the hypothesized population parameter and the actual value of the sample statistic • Judging (objectively) whether difference is significant or not • The smaller the difference, the greater the likelihood that our hypothesized value for the mean is correct

Hypothesis Testing

Concepts

- Introduction

Hypothesis Testing	E.g. Specifications in the contract for a new sports complex, require the aluminum sheets used roofing to be 0.04-inch-thick. If sheets are appreciably thicker than 0.04 inch, the structure would not be able to support the additional weight. If sheets are appreciably thinner than 0.04 inch, roof strength would be inadequate. Of the 10,000 sheets required, a random sample of 100 sheets is drawn & its mean thickness is found to be 0.0408 inch. From past experience, it is believed that these sheets come from a thickness population with a standard deviation of 0.004 inch. On the basis of these data, the decision maker must decide whether the 10,000 sheets meet specifications.

Hypothesis Testing

Concepts

- Introduction

Hypothesis Testing	E.g. • Assuming that aluminum sheets are 0.04 inch thick & that population standard deviation is 0.004 inch, decision maker must decide how likely is it that one would get a sample mean of 0.0408 inch or more from the population i.e. If true mean (of population) is 0.04 inch, and standard deviation is 0.004 inch, then what are the chances of getting a sample mean that differs from 0.04 inch by 0.0008 (0.0408 - 0.04) inch or more? • Need to compute the probability that a random sample with a mean thickness of 0.0408 inch will be selected from the population having a mean thickness of 0.04 inch & standard deviation 0.004 inch • If probability is too low, mean thickness of aluminum sheets ≠ 0.04 inch

Hypothesis Testing

Concepts

- Introduction

Hypothesis Testing	E.g.
	• Hypothesized population mean μ = 0.04 inch, σ = 0.004 inch, what is the probability of getting a sample with mean x = 0.0408 inch? • Standard error $\sigma_{\bar{x}} = \sigma / \sqrt{n}$ = 0.004 / $\sqrt{100}$ = 0.0004 inch • z = (x – μ) / $\sigma_{\bar{x}}$ = (0.0408 – 0.04) / 0.0004 = 0.0008/0.0004 = 2, and thus p(-2 ≤ z ≤ 2) = 0.045 • There is a low probability that a population with true mean of 0.04 inch would produce a sample having sample mean as 0.0408 inch • Difference between sample mean & hypothesized population mean is large • How low or how high is the probability to accept or reject the assumption is the decision maker's choice to make

Hypothesis Testing

Concepts

- Introduction
  - An automobile manufacturer claims that a particular model gets 28 kmpl mileage. The Environmental Protection Agency, using a sample of 49 automobiles of this model, finds the sample mean to be 26.8 kmpl. From previous studies, the population standard deviation is known to be 5 kmpl. Could we reasonably expect (within 2 standard errors) that we could select such a sample if indeed the population mean is actually 28 kmpl?
  - How many standard errors around the hypothesized value should be used to be 99.44 percent certain that hypothesis is accepted when it is true?

Hypothesis Testing

Concepts

- Introduction
  - A grocery store has specially packaged oranges and has claimed a bag of oranges will yield 2.5 liters of juice. After randomly selecting 42 bags, a stacker found the average juice production per bag to be 2.2 liters. Historically, we know the population standard deviation is 0.2 liters. Using this sample and a decision criterion of 2.5 standard errors, could we conclude the store's claims are correct?
  - If a hypothesized value is rejected because it differs from a sample statistic by more than 1.75 standard errors, what is the probability that the hypothesis rejected is in fact true?
  - How many standard errors around the hypothesized value should be used to be 98 percent certain that hypothesis is accepted when it is true?

Hypothesis Testing

Concepts

- Introduction
  - A magazine has asserted that the amount of time PC owners spend on their PCs averages 23.9 hours per week & has a std. dev. of 12.6 hours per week. A random sampling of 81 of its subscribers revealed a sample mean usage of 27.2 hours per week. On the basis of this sample, is it reasonable to conclude (using 2 standard errors as the decision criterion) that the magazine’s subscribers are different from average PC owners?

Hypothesis Testing

Concepts

- Null & Alternate Hypothesis
  - Assumed value i.e., hypothesized value of the population parameter must be stated before beginning sampling procedure

Null Hypothesis H_0	The assumption (about population parameter) to be tested • Will be assumed to be true during the testing of the hypothesis • Will be rejected only if sample data provide substantial contradictory evidence
Alternate Hypothesis H_1 or H_A or H_a	Opposite of what is stated in the null hypothesis • Includes all population values not included in the null hypothesis • Will be selected only if there is strong enough sample evidence to support it • Is deemed to be true if the null hypothesis is rejected

Hypothesis Testing

Concepts

- Null & Alternate Hypothesis

E.g.	To test that the population mean is 500
Null Hypothesis	H_0: μ = 500 (Null hypothesis is that population mean is 500)
Alternate Hypothesis	H_1: μ ≠ 500 (Alternate hypothesis is that population mean is 500)
	OR
Null Hypothesis	H_0: μ = 500 (Null hypothesis is that population mean is 500)
Alternate Hypothesis	H_1: μ > 500 (Alternate hypothesis is that population mean is greater than 500)
	OR
Null Hypothesis	H_0: μ = 500 (Null hypothesis is that population mean is 500)
Alternate Hypothesis	H_1: μ < 500 (Alternate hypothesis is that population mean is less than 500)

Hypothesis Testing

Concepts

- Null & Alternate Hypothesis

Developing / Formulating Null & Alternate Hypotheses	
Testing the status quo or challenging a claim	• Used in process analysis or validating claims • Assuming the status quo or claim to be true (Null Hypothesis is that the status quo is true) • Null Hypothesis is rejected only if sample data provides strong enough evidence to accept the Alternate Hypothesis
Testing a research hypothesis	• Used in development of new products • Assuming (Null Hypothesis) that new product is no better than the original one • Null Hypothesis rejected only if sample data provides strong enough evidence to accept the Alternate Hypothesis (which is also known as Research Hypothesis in this case)

Hypothesis Testing

Concepts

- Null & Alternate Hypothesis

Forms of Hypothesis Test		
$H_0 : \mu \geq \mu_0$	$H_0 : \mu \leq \mu_0$	$H_0 : \mu = \mu_0$
$H_1 : \mu < \mu_0$	$H_1 : \mu > \mu_0$	$H_1 : \mu \neq \mu_0$
One-Tailed Tests		Two-Tailed Tests
H_0 is null Hypothesis	μ is population mean	
H_1 is alternate Hypothesis	μ_0 is hypothesized value of population mean	
Equality part (either ≥, ≤, or =) always appears in Null Hypothesis		

Hypothesis Testing

Concepts

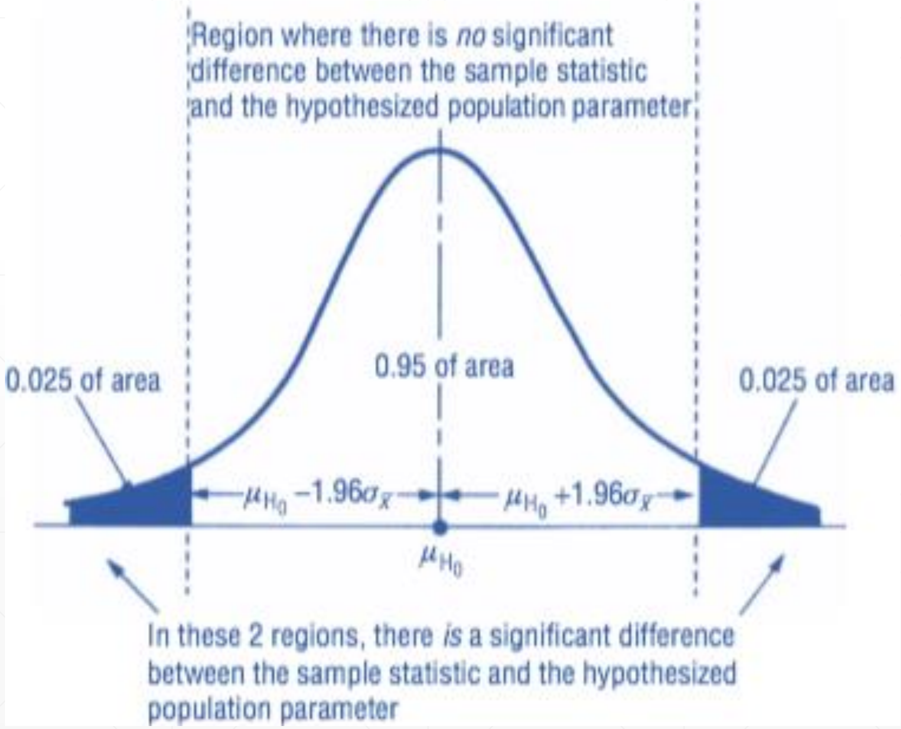
- Significance Level

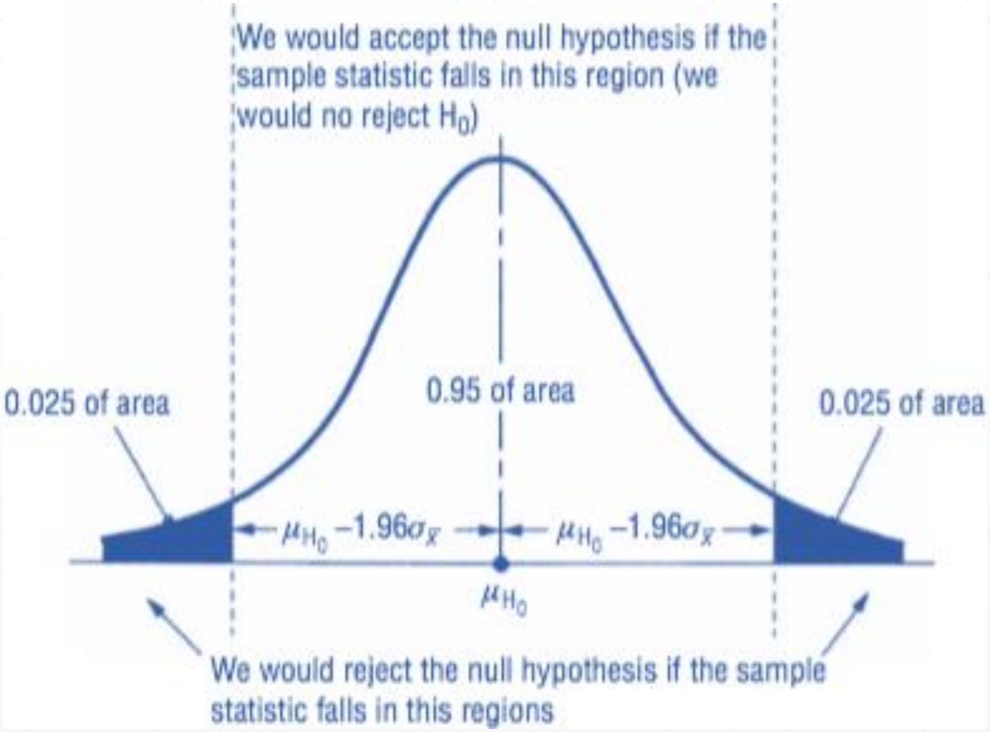
Goal of Hypothesis Test	
The purpose of hypothesis testing is not to question the computed value of the sample statistic but to make a judgment about the difference between that sample statistic and a hypothesized population parameter	
Significance Level	indicates the percentage of sample statistic that is outside certain limits if the hypothesis is assumed to be correct i.e., the probability of rejecting the null hypothesis when it is true as an equality • For 5% level of significance, null hypothesis will be rejected if difference between sample statistic & hypothesized population parameter is so large that it or a larger difference would occur, on an avg., only 5 or fewer times in every 100 samples when hypothesized population parameter is correct • In interval estimation, the confidence level indicates the percentage of sample statistic that falls within the defined confidence interval

Hypothesis Testing

Concepts

- Significance Level





Hypothesis Testing

Concepts

- Significance Level

Even if the sample statistic falls in the acceptable region, this does not prove the null hypothesis H_0 is true; it simply does not provide statistical evidence to reject it

The null hypothesis H_0 is accepted (not proven), when sample data do not cause it to be rejected

Higher the significance level, higher the probability of rejecting the null hypothesis when it is true as an equality

Hypothesis Testing

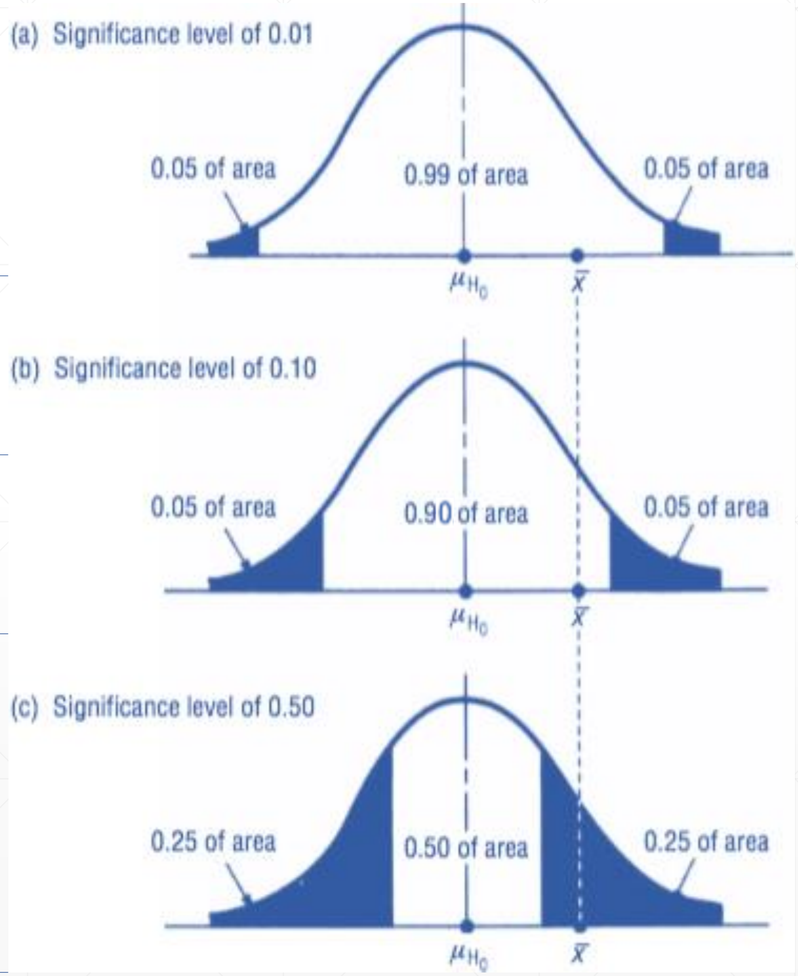
Concepts

- Significance Level

In part (a), the null hypothesis that the population mean is equal to the hypothesized value would be accepted

In part (b), the null hypothesis that the population mean is equal to the hypothesized value would be accepted

In part (c), this same null hypothesis would be rejected because the significance level of 0.50 is so high that the null hypothesis would rarely be accepted when it is not true but, at the same time, it would often be rejected when it is true as an equality



Hypothesis Testing

Concepts

- Type I & Type II Errors

Type I Error	Type II Error
Rejecting a null hypothesis when it is true is called a Type I error, and its probability (also known as the significance level) is symbolized as α (alpha)	Accepting a null hypothesis when it is false is called a Type II error, and its probability is symbolized as β (beta)

Trade-offs between the two Types of Error
Probability of making one type of error can be reduced only by increasing the probability of making the other type of error

		Population Condition	
		H_0 True	H_1 True
Conclusion	Accept H_0	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

Hypothesis Testing

Concepts

- Type I & Type II Errors

Preference for Type I Error	Preference for Type II Error
When the cost (consequence) of accepting a false hypothesis is high (severe) as compared to the cost (consequence) of rejecting a true hypothesis	When the cost (consequence) of rejecting a true hypothesis is high (severe) as compared to the cost (consequence) of accepting a false hypothesis
E.g.: Testing a batch of chemicals to be ok or defective Null Hypothesis: Batch is OK Type I Error: Rejecting the null hypothesis when it is true, i.e., rejecting a batch of chemicals when it is OK. It involves time & trouble of reworking a batch that should have been accepted i.e., unnecessary reworking for an OK batch. Type II Error: Accepting the null hypothesis when it is false, i.e., accepting a batch which is poisonous, the consequences of which would be much sever Cost of Type II Error is much higher of Type I Error So a Type I Error would be preferred over a Type II Error	E.g.: Testing an assembled engine to be ok or defective Null Hypothesis: Engine is OK. Type I Error: Rejecting the null hypothesis when it is true, i.e., rejecting an assembled engine when it is OK. It involves disassembling the entire engine at the factory, the cost of which is much higher. Type II Error: Accepting the null hypothesis when it is false, i.e., accepting an assembled engine which is defective. It involves making required repairs Cost of Type I Error is much higher of Type II Error So a Type II Error would be preferred over a Type I Error

Hypothesis Testing

Concepts

- Power of Hypothesis Test

Power of Hypothesis Test
• Accepting a null hypothesis when it is false is called a Type II error, and its probability is symbolized as β (beta) • 1 - β is therefore, the probability of rejecting a null hypothesis when it is false • A good hypothesis test must reject a null hypothesis when it is false, and hence must provide as large a probability for rejecting a false null hypothesis as possible i.e. a large value of 1 – β
A high value of 1 - β means the hypothesis test is working quite well (that it is rejecting the null hypothesis when it is false)
A low value of 1 - β means the hypothesis test is working very poorly (that it is not rejecting the null hypothesis when it is false)
Because the value of 1 - β is the measure of how well the hypothesis test is working, it is known as the power of the test

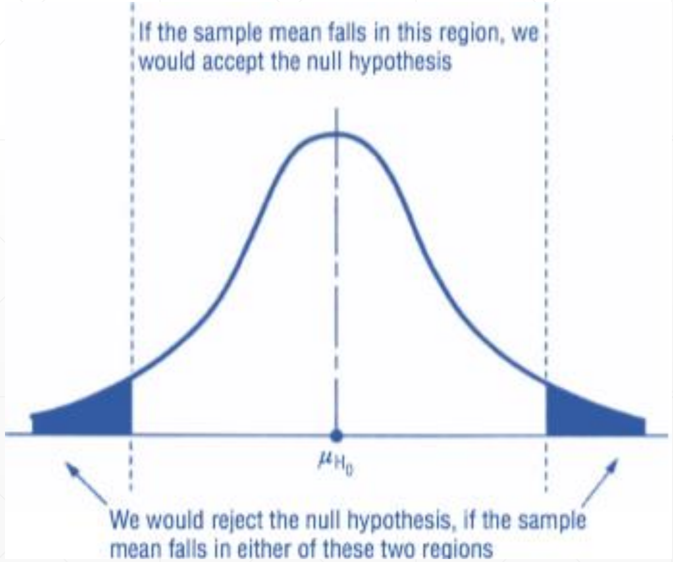
Hypothesis Testing

Concepts

- Two-Tailed & One-Tailed Tests of Hypothesis

Two-Tailed Test of Hypothesis

- Rejects the null hypothesis if the same mean is significantly higher / lower than the hypothesized population mean
- The two-tailed test thus has two rejection regions
- Appropriate when H_0: $\mu = \mu_0$ and H_1: $\mu \neq \mu_0$

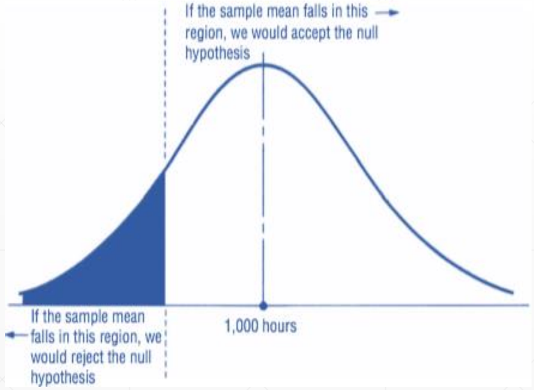


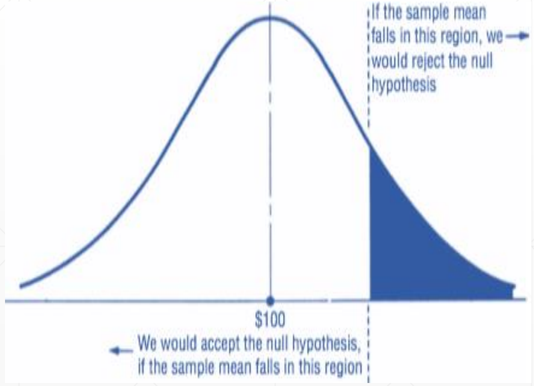
Hypothesis Testing

Concepts

- Two-Tailed & One-Tailed Tests of Hypothesis

Two Types One-Tailed Test of Hypothesis	
Left-Tailed Test of Hypothesis	Right-Tailed Test of Hypothesis
• Left-tailed test (lower-tailed test) rejects the null hypothesis if the same mean is significantly lower than the hypothesized population mean • H_0: $\mu \geq \mu_0$ and H_1: $\mu < \mu_0$	• Right-tailed test (upper-tailed test) rejects the null hypothesis if the same mean is significantly higher than the hypothesized population mean • H_0: $\mu \leq \mu_0$ and H_1: $\mu > \mu_0$





Hypothesis Testing

Concepts

- Exercise
  - A highway safety engineer decides to test the load-bearing capacity of a bridge that is 20 years old. Considerable data are available from similar tests on the same type of bridge. Which is appropriate, a one-tailed or a two-tailed test? If the minimum load-bearing capacity of this bridge must be 10 tons, what are the null and alternative hypotheses?
  - Formulate null & alternative hypotheses to test whether mean annual rainfall in Mumbai, exceeds 90 inches.
  - In a trial, the null hypothesis is that an individual is innocent of a certain crime. Would the legal system prefer to commit a Type I or a Type II error with this hypothesis?

Hypothesis Testing

Concepts

- Exercise
  - The null hypothesis is that the battery for a heart pacemaker has an average life of less than or equal to 300 days, with the alternative hypothesis being that the battery life is more than 300 days. As the quality engineer for battery manufacturer:
    - Would you rather make a Type I or a Type II error?
    - For your preference of error, should you use a high or a low significance level?
  - The statistics department installed energy-efficient lights, heaters, and air conditioners last year. Now they want to determine whether the average monthly energy usage has decreased. Should they perform a one-tailed or two-tailed test? If their previous average monthly energy usage was 3124 kilowatt hours, what are the null and alternative hypotheses?

Hypothesis Testing

One Sample Tests

- Procedure for Hypothesis Testing

Procedure for Hypothesis Testing	
1. Develop the null and alternative hypotheses 2. Specify the level of significance 3. Collect the sample data and compute the value of the test statistic	
p-Value Approach	Critical Value Approach
4. Use the value of test statistic to compute the p-value 5. Reject H_0 if the p-value $\leq \alpha$	4. Use the level of significance to determine the critical value and the rejection rule 5. Use the value of test statistic & the rejection rule to determine whether to reject H_0
6. Interpret the statistical conclusion in the context of the application	
p-Value Approach & Critical Value Approach will always provide the same conclusion	

Hypothesis Testing

One Sample Tests

- Hypothesis Testing of Means
  - When Population Standard Deviation, Sigma (σ), is Known

Procedure for Hypothesis Testing

1. Develop the null and alternative hypotheses
2. Specify the level of significance
3. Collect the sample data and compute the value of the test statistic

p-Value Approach	Critical Value Approach
4. Use the value of test statistic to compute the p-value 5. Reject H_0 if the p-value $\leq \alpha$	4. Use the level of significance to determine the critical value and the rejection rule 5. Use the value of test statistic & the rejection rule to determine whether to reject H_0

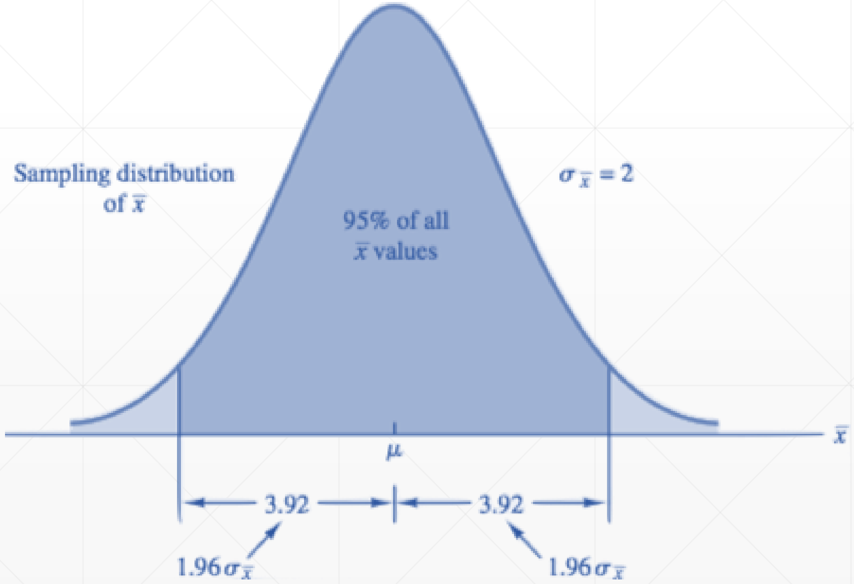
6. Interpret the statistical conclusion in the context of the application

p-Value Approach & Critical Value Approach will always provide the same conclusion

Hypothesis Testing

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



Hypothesis Testing

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

Hypothesis Testing

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

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

Hypothesis Testing

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.

Hypothesis Testing

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.

Hypothesis Testing

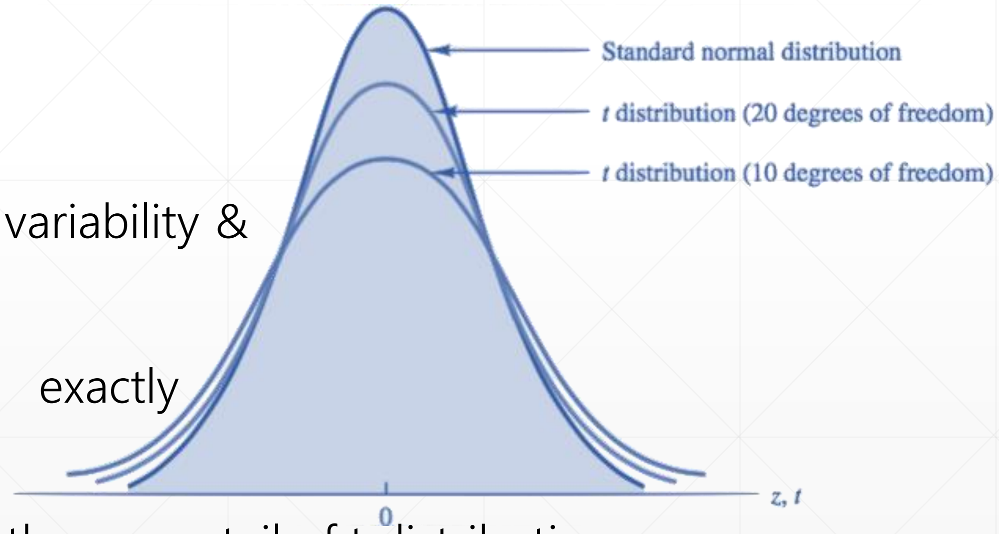
Interval Estimation of Population Mean

- 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

Hypothesis Testing

Interval Estimation of Population Mean

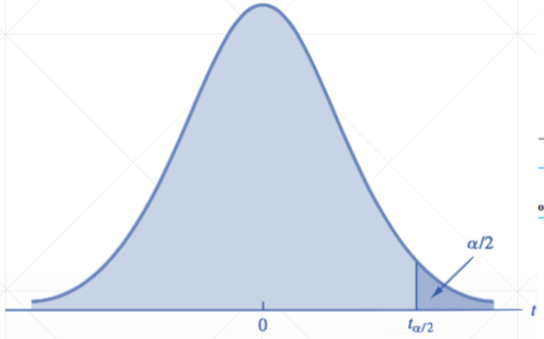
- When Population Standard Deviation, Sigma (σ), is Unknown
  - As the no. of degrees of freedom increases, the difference between the t distribution & the standard normal distribution becomes smaller & smaller
  - Mean of t distribution is zero
  - t distribution with more degrees of freedom exhibits 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



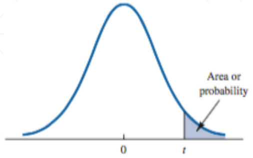
Hypothesis Testing

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

Hypothesis Testing

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

Hypothesis Testing

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 = $\bar{x} \pm (t_{\alpha/2} \times$ 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.

Hypothesis Testing

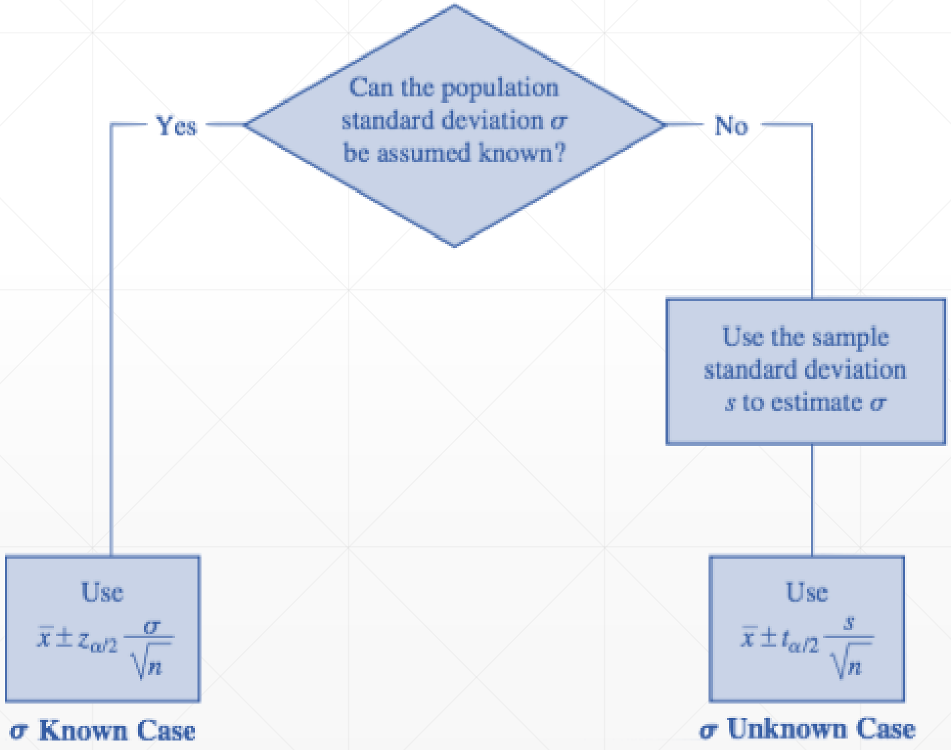
Interval Estimation of Population Mean

- 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.

Hypothesis Testing

Interval Estimation of Population Mean

Summary of Interval Estimation Procedures for a Population Mean



Hypothesis Testing

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 ≅ Range ÷ 4, if the population standard deviation is unknown
where Range is the difference between the largest & the smallest data values

Hypothesis Testing

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?

Hypothesis Testing

Interval Estimation of Population Proportion

- Interval Estimation of Population Proportion

Interval Estimate of Population Proportion, p

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

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

Hypothesis Testing

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.

Hypothesis Testing

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
$\bar{p}$	~ 0.5, if it is not known

Hypothesis Testing

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