Frequently Asked Questions

1. What are the steps to be followed for the construction of the C.I for the unknown population mean with known and unknown variance of the population?

Answer:

Steps to obtain confidence interval for mean

- First obtain the point estimate of μ, that is, the sample mean .
- If *n* is large, then the Central Limit Theorem can be used and y bar is normally distributed with mean μ and standard deviation sigma by root n
- Select a confidence level. The most common level is 95%.
- Then the confidence interval is [$\bar{y}$ - $Z_{\alpha/2}$ $\sqrt{\sigma^2/n}$, $\bar{y}$ + $Z_{\alpha/2}$ $\sqrt{\sigma^2/n}$]
- If the population variance is unknown the CI given by

$$\left[\ \bar{y} - t_{\alpha}(n-1)\sqrt{s^2/n}\ ,\ \bar{y} + t_{\alpha}(n-1)\sqrt{s^2/n}\ \right]$$

where $s^2 = \frac{\sum_{i=1}^{n}(yi-\bar{y})^2}{n-1}$ and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

2. Find 90% C.I for the mean of the normal variable with σ=2 and a sample of size 8 gave the values 9, 14, 10, 12, 7, 13, 11, 12

Answer:

Given σ=2 and n =8

100 (1- α) % C.I for the population mean μ when the variance is known as σ^2 is given by

$$\left[\ \bar{y} - Z_{\alpha/2}\sqrt{\sigma^2/n}\ ,\ \bar{y} + Z_{\alpha/2}\sqrt{\sigma^2/n}\ )\ \right]$$

Given 100(1- α)%=90% implies 1- α=0.90 implies α=0.1 then α/2= 0.05

From the table of Standard Normal probabilities we get $Z_{\alpha/2} = Z_{0.05} = 1.64$

$\bar{y} = 11$ and √n= 2.828427

Therefore, 90% Confidence Interval for the mean of the distribution

$$\left[\ \bar{y} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ,\ \bar{y} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ \right]$$

$$=\left[\ 11-1.64\frac{2}{2.828427}\ ,\ 11+1.64\frac{2}{2.828427}\ \right]=[\ 9.840345,\ 12.15966]$$

Therefore 90% Confidence Interval for the mean of the distribution is

[9.840345, 12.15966]

3. Two random samples of sizes 10 and 12 from normal populations having the same variance gave $\bar{x} = 20, \bar{y} = 24$, $s1^2$ =25 and $s2^2$ =36. Find 90% Confidence limits for the difference of populations means.

Answer:

Given m=10, n=12, $\bar{x}$ = 20, $\bar{y}$ = 24, s12 =25 and s22 =36, 1-α=0.90

100 (1- α) % C.I for the difference of means of two populations with unknown but common Standard Deviations is given by

[($\bar{x}$ - $\bar{y}$)- t_α(m+n-2) $S_p\sqrt{\frac{1}{m}+\frac{1}{n}}$, ($\bar{x}$ - $\bar{y}$)+ t_α(m+n-2) $S_p\sqrt{\frac{1}{m}+\frac{1}{n}}$]

Select a confidence level. In this analysis, the confidence level is defined for us in the problem. We are working with a 90% confidence level.

t_α(m+n-2)= $t_{0.1}$(20)=1.73

$$S_p = \sqrt{\frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}}$$

$$S_p = \sqrt{\frac{(10-1)(25) + (12-1)(36)}{10+12-2}}$$

=5.5723

By substituting the above in the interval we get

[(20-24)-(1.73)(5.5723) $\sqrt{\frac{1}{10}+\frac{1}{12}}$,(20-24)+(1.73)(5.5723) $\sqrt{\frac{1}{10}+\frac{1}{12}}$]

[-8.127635, 0.127635]

Therefore, the 90% confidence interval is -8.127635 to 0.127635

4. A sample of size 10 from a normal distribution with S.D 3 is as follows. Establish 90% confidence interval for the mean of the distribution 6.5, 5.5, 4.8, 5.6, 4, 4.7, 10.5, 8, 7.3, 6.8

Answer:
$Xi \sim N(\mu,\sigma^2)$ Given σ=3 and n =10

100 (1- α) % C.I for the population mean μ when the variance is known as σ^2 is given by

$$[\ \bar{y} - Z_{\alpha/2} \sqrt{\sigma^2/n}\ ,\ \bar{y} + Z_{\alpha/2} \sqrt{\sigma^2/n}\ )\ ]$$

Given 100(1- α)%=90% implies 1- α=0.90 implies α=0.1 then α/2= 0.05

From the table of Standard Normal probabilities we get $Z_{\alpha/2} = Z_{0.05} = 1.64$

$\bar{y} = 6.37$ and $\sqrt{n}$= 3.16
Therefore 90% Confidence Interval for the mean of the distribution

$$[\ \bar{y} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\ ,\ \bar{y} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\ ]$$

$$=[\ 6.37-1.64 \frac{3}{3.16}\ ,\ 6.37+1.64 \frac{3}{3.16}\ ]=[4.814,\ 7.9258]$$

Therefore 90% Confidence Interval for the mean of the distribution is

[4.814, 7.9258]

5. An auditor is faced with a population of 1000 vouchers and wants to estimate the average value of the population. A sample of 50 vouchers is selected with average voucher amount of $1076.39, standard deviation of $273.62. Set up the 95% confidence interval estimate of the average amount for the population of vouchers.

Answer:

Given , n=50, $\bar{y} = \$1076.39$ and s =$273.62

95% Confidence Interval for the population mean μ is

$$[\ \bar{y} - t_\alpha(n-1) \sqrt{s^2/n}\ ,\ \bar{y} + t_\alpha(n-1) \sqrt{s^2/n}\ ]$$

1 -α = 0.95 which implies α=0.05 and since n is greater than 30

$t_\alpha(n-1) = 1.96$**.**

By substituting the given values in the above we get

[1076.39–(1.96) 273.62/√50 , 1076.39+(1.96) 273.62/√50]

=[1026.39 , 1126.39]

The 95% confidence interval for the population average amount of the vouchers is between 1026.39 and 1126.39

6. The local baseball team conducts a study to find the amount spent on refreshments at the ball park. Over the course of the season they gather simple random samples of 50 men and 100 women. For men, the average expenditure was $20, with a standard deviation of $3. For women, it was $15, with a standard deviation of $2. What is the 99% confidence interval for the spending difference between men and women? Assume that the two populations are independent and normally distributed.

Answer:

Identify a sample statistic. Since we are trying to estimate the difference between population means, we choose the difference between sample means as the sample statistic. Thus, x_1 - x_2=

$20 - $15 = $5.

Find standard error. The standard error is an estimate of the standard deviation of the difference between population means. We use the sample standard deviations to estimate the standard error (SE).

$$\text{S.E}(\bar{x} - \bar{y}) = \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

$$\text{S.E}(\bar{x} - \bar{y}) = \sqrt{\frac{3^2}{50} + \frac{2^2}{100}} = 0.47$$

Find critical value. The critical value is a factor used to compute the margin of error. Because the sample sizes are large enough, we express the critical value as a z score.

$Z_{0.01/2} = Z_{0.05} = 2.58$

Compute margin of error (ME): ME = critical value * S.E

$$= Z_{\alpha/2} \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

= 2.58 * 0.47 = 1.21

Therefore, 100 (1- α) % C.I for the difference of means of two populations with unknown Standard Deviations is given by

$$\left[(\bar{x} - \bar{y}) - Z_{\alpha/2}\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}\ ,\ (\bar{x} - \bar{y}) + Z_{\alpha/2}\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}\ \right]$$

=[5-1.21, 5+1.21]= [3.79, 6.21]

Therefore, the 99% confidence interval is \$3.79 to \$6.21. That is, we are 99% confident that men outspend women at the ballpark by about \$5 $\pm$ \$1.21.

7. The following information was obtained from 83 boys and 95 girls selected from their college regarding their performance in the final examination

Marks

	A.M.	S.D.
Boys	30.92	7.81
Girls	29.21	11.56

Establish a 95% confidence interval for the difference between the average marks of the boys and the girls.

Answer:

We have to establish 95% Confidence interval for the difference between the average performance of boys and girls where the population variances are unknown

100 (1- α) % C.I for the difference of means of two populations with unknown Standard Deviations is given by

$$\left[(\bar{x} - \bar{y}) - Z_{\alpha/2}\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}\ ,\ (\bar{x} - \bar{y}) + Z_{\alpha/2}\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}\ \right]$$

Because the sample sizes are large enough, we express the critical value as a z score.

Where, m=83, n=95 ; $\bar{x}$=30.92 , $\bar{y}$=29.21, S_1=7.81 S_2=11.56

Given 100(1- α)%=95% implies 1- α=0.95 implies α=0.05 then α/2= 0.025

From the table of Standard Normal probabilities we get $Z_{\alpha/2} = Z_{0.025}$=1.96

By substituting in the above expression we get

$$[(30.92-29.21)-1.96\sqrt{\frac{(7.81)^2}{83}+\frac{(11.56)^2}{95}},\quad (30.92-29.21)+1.96\sqrt{\frac{(7.81)^2}{83}+\frac{(11.56)^2}{95}}]$$

= [-1.15828, 4.578279]

Therefore, 95 % C.I for the difference of means of two populations with unknown but common Standard Deviations is given by

[-1.15828, 4.578279]

8. Establish 95% confidence interval for the mean height of college students (in inches) 154.4, 152.2, 152.1, 162.9, 159.7, 158.5 151.4, 153.9, 153.4, 150.8, 150.9, 157.7

Answer:
$X_i \sim N(\mu,\sigma^2)$ Given n =11

Given 100(1- α)%=95% implies 1- α=0.95 implies α=0.05

From the table of Students t distribution the value of t for (n-1) =(12-1) = 11 degrees of freedom at 0.05 level of significance is $t_{\alpha}(n-1) = t_{0.05}(11) = 2.20$

$\bar{y} = 154.825$ and $s^2 = \frac{\sum_{i=1}^{n}(yi-\bar{y})^2}{n-1} = 15{,}5875$ and s= 3.948101 and √n=3.464102

100 (1- α) % C.I for the population mean μ when the variance is unknown is given by$[\bar{y} - t_{\alpha}(n-1)\frac{s}{\sqrt{n}}, \bar{y} + t_{\alpha}(n-1)\frac{s}{\sqrt{n}}]$

where $s^2 = \frac{\sum_{i=1}^{n}(yi-\bar{y})^2}{n-1}$ and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

$=[154.825 - 2.20\frac{3.948101}{3.464102}, 154.825 + 2.20\frac{3.948101}{3.464102}]$

=[152.3176 157.3324]

Therefore 95% Confidence Interval for the mean of the distribution is
=[152.3176 157.3324]

9. A Simple Random Sample of size 25 was drawn from a population which has the mean value of 50 and its s value is 8. Set up 95% interval estimate for the population mean μ.

Answer:

Therefore, 100 (1- α) % C.I for the population mean $\bar{Y}$ when the variance is unknown under SRSWR is given by

$$[\ \bar{y} - t_\alpha(n\text{-}1)\sqrt{s^2/n}\ ,\ \bar{y} + t_\alpha(n\text{-}1)\sqrt{s^2/n}\ ]$$

From the table of students t distribution t(24) = 2.0639 at 5% level of significance. Hence the interval is

$$[\ 50 - 2.0639\sqrt{64/25}\ ,\quad 50 + 2.0639\sqrt{64/25}\ ]$$

$[46.69, 53.30]$

Hence, an Interval estimate for the population mean μ is $[46.69, 53.30]$

10. An IQ test was administered to 5 persons before and after they were trained. The obtained results are as under. Establish a 95% confidence interval for the difference between the average I.Q

I.Q before training: 110 120 123 132 125

I.Q after training: 120 118 125 136 121

Answer:

The given set of observations is correlated variables. Hence we find di for the given values

xi	yi	di=xi-yi	di^2
110	120	-10	100
120	118	2	4
123	125	-2	4
132	136	-4	16
125	121	4	16
Total		-10	140

Therefore, 100 (1- α) % C.I for the population mean θ in case of correlated variables is given by

$$[\ \bar{d} - t_\alpha(n\text{-}1)\frac{s_d}{\sqrt{n}}\ ,\ \bar{d} + t_\alpha(n\text{-}1)\frac{s_d}{\sqrt{n}}\ ] \text{ where } s_d^2 = \frac{\sum(d_i - \bar{d})^2}{n-1} \text{ and } \bar{d} = \frac{\sum d_i}{n}$$

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-10}{5} = -2$$

$$s_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} = \frac{\sum di^2 - n\bar{d}^2}{n-1} = 30$$ and $s_d =$ 5.477226

100 (1- α) %=95% which implies α=0.05

From the table of t-distribution we get $t_\alpha(n-1) = t_{0.05}(4)=2.78$

Therefore 90 % C.I for the difference in mean IQ is given by

$$[\bar{d} - t_\alpha(n-1)\frac{s_d}{\sqrt{n}}, \bar{d} + t_\alpha(n-1)\frac{s_d}{\sqrt{n}}]$$

$$[-2 - 2.78\frac{5.477226}{\sqrt{5}}, -2 + 2.78\frac{5.477226}{\sqrt{5}}] = [-8.80958, 4.809581]$$

Hence, 95% Confidence Interval for the difference in the mean IQ is

[-8.80958, 4.809581]

11. Suppose that simple random samples of college freshman are selected from two universities - 15 students from school A and 20 students from school B. On a standardized test, the sample from school A has an average score of 1000 with a standard deviation of 100. The sample from school B has an average score of 950 with a standard deviation of 90. What is the 90% confidence interval for the difference in test scores at the two schools, assuming that test scores came from normal distributions with common variances in both schools?

Answer:

Given m=15, n=20, $\bar{x}$ = 1000, $\bar{y}$ = 950, s1=100, s2=90, 1-α=0.90

$$\bar{x} - \bar{y} = 1000 - 950 = 50$$

100 (1- α) % C.I for the difference of means of two populations with unknown but common Standard Deviations is given by

$$[(\bar{x} - \bar{y}) - t_\alpha(m+n-2)\, S_p\sqrt{\frac{1}{m}+\frac{1}{n}},\ (\bar{x} - \bar{y}) + t_\alpha(m+n-2)\, S_p\sqrt{\frac{1}{m}+\frac{1}{n}}]$$

Select a confidence level. In this analysis, the confidence level is defined for us in the problem. We are working with a 90% confidence level.

$t_\alpha(m+n-2)= t_{0.1}(33)=1.282$

$$S_p = \sqrt{\frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}}$$

$$S_p = \sqrt{\frac{(15-1)(100)^2 + (20-1)(90)^2}{15+20-2}}$$

=2.8598

By substituting the above in the interval we get

$$[50-(1.282)(2.8598)\sqrt{\frac{1}{15}+\frac{1}{20}},50+(1.282)(2.8598)\sqrt{\frac{1}{15}+\frac{1}{20}}]$$

[48.74755, 51.25245]

Therefore, the 90% confidence interval is 48.74755 to 51.25245.

12. Marks in a test before and after coaching of 9 students are given below. Assuming that the difference in marks is from Normal distribution find a 90% Confidence Interval for the difference in the mean performance.

Marks before coaching: 60 12 38 45 19 72 70 50 48

Marks after coaching: 62 28 40 41 30 75 70 54 50

Answer:

The given set of observations is correlated variables. Hence, we find di for the given values

xi	yi	di=xi-yi	di^2
60	62	-2	4
12	28	-16	256
38	40	-2	4
45	41	4	16
19	30	-11	121
72	75	-3	9

70	70	0	0
50	54	-4	16
48	50	-2	4
Total		-36	430

Therefore,100 (1- α) % C.I for the population mean θ in case of correlated variables is given by

$[\overline{d}$ - t_α(n-1) $\frac{s_d}{\sqrt{n}}$, $\overline{d}$ + t_α(n-1) $\frac{s_d}{\sqrt{n}}$] where $s_d^2 = \frac{\sum (d_i - \overline{d})^2}{n-1}$ and $\overline{d} = \frac{\sum d_i}{n}$

$$\overline{d} = \frac{\sum d_i}{n} = \frac{-36}{9} = -4$$

$$s_d^2 = \frac{\sum (d_i - \overline{d})^2}{n-1} = \frac{\sum di^2 - n\overline{d}^2}{n-1} = 35.75$$ and $s_d = 5.9791$

100 (1- α) %=90% which implies α=0.1

From the table of t-distribution we get $t_\alpha(n-1) = t_{0.1}(8)=1.86$

Therefore, 90 % C.I for the difference in mean performance is given by

$[\overline{d}$ - t_α(n-1) $\frac{s_d}{\sqrt{n}}$, $\overline{d}$ + t_α(n-1) $\frac{s_d}{\sqrt{n}}$]

$[-4 - 1.86 \frac{5.9791}{\sqrt{9}}, -4 + 1.86 \frac{5.9791}{\sqrt{9}},\ ] = [-7.707,-0.2930]$

Hence, 90% Confidence Interval for the difference in the mean performance is [-7.707,-0.2930]

13. Two fertilizers X and Y are tested on 10 plots respectively. The yields in different plots is given below. Establish a 95% confidence interval for the difference between the average yields

    Fertilizer A: 16 17 12 14 8 15 9 14 8 10

Fertilizer B: 13 22 15 12 14 18 8 21 10 17

Answer:

We have to establish 95% Confidence interval for the difference between the average sales of the two salesmen where the population variances are unknown

100 (1- α) % C.I for the difference of means of two populations with unknown but common Standard Deviations is given by

$$[(\bar{x} - \bar{y}) - t_\alpha(m+n-2)\ S_p\sqrt{\frac{1}{m}+\frac{1}{n}},\ (\bar{x} - \bar{y}) + t_\alpha(m+n-2)\ S_p\sqrt{\frac{1}{m}+\frac{1}{n}}]$$

Where m=n=10 ; $\bar{x} = \frac{\sum_{i=1}^{7} xi}{m} = \frac{123}{10} = 12.3$ and $\bar{y} = \frac{\sum_{i=1}^{7} yi}{n} = \frac{150}{10} = 15$

$$S_p = \sqrt{\frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}} = \sqrt{\frac{\sum(xi-\bar{x})^2 + \sum(yi-\bar{y})^2}{m+n-2}} =$$

$$S_p = \sqrt{\frac{102.1+186}{18}} = 4.000694$$

Given 100(1- α)%=95% implies 1- α=0.95 implies α=0.05

From the table of Students t distribution for (m+n-2) = 18 degrees freedom we get $t_{0.05}(18)=2.10$

By substituting in the above expression we get

$$[(12.3-15) - (2.10)(4.000694)\sqrt{\frac{1}{10}+\frac{1}{10}},\ (12.3-15) + (2.10)(4.000694)\sqrt{\frac{1}{10}+\frac{1}{10}}] = [-6.45725, 1.057246]$$

Therefore, 95 % C.I for the difference of means of two populations with unknown but common Standard Deviations is given [-6.45725,1.057246]

14. A social worker is faced with a population of 1000 families and wants to estimate the average expenditure of the families of the population. A sample of 50 families is selected with average expenditure of \$1076.39, standard deviation of \$273.62. Set up the 95% confidence interval estimate of the average expenditure of the families in the population.

Answer:

Given , n=50, $\overline{y} = \$1076.39$ and s =$273.62

95% Confidence Interval for the population mean μ is

[$\bar{y}$ –tα(n-1) $\sqrt{s^2/n}$, $\bar{y}$ + tα(n-1) $\sqrt{s^2/n}$]

1 -α = 0.95 which implies α=0.05 and since n is greater than 30

$t_\alpha(n-1) = 1.96$.

By substituting the given values in the above we get

[1076.39–(1.96) 273.62/√50 , 1076.39+(1.96) 273.62/√50]

=[1026.39, 1126.39]

The 95% confidence interval for the population average expenditure of the families is between $1026.39 and $1126.39

15. A random sample of 25 students is selected from an University to estimate an average time spent by a student in a library in a month. A sample data reveals that a sample average=50 hours and a standard deviation is 8 Set up 95% Confidence Interval estimate for the average time spent by the students of University in a library.

Answer:

Given , n=25, $\overline{y} = 50$ and s =8

95% Confidence Interval for the population mean μ is

[$\bar{y}$ –tα(n-1) $\sqrt{s^2/n}$, $\bar{y}$ + $t_\alpha(n-1)$ $\sqrt{s^2/n}$]

1 -α = 0.95 which implies α=0.05 and $t_\alpha(n-1) = t_{0,05}(24)=2.06$ from the table of students t distribution

By substituting the given values in the above we get

[50–(2.06) 8/√25 , 50+(2.06) 8/√25]

=[46.704, 53.296]

The 95% confidence interval for the average time spent by the students of an university in a library is between 46.70 hours and 53.30 hours.