## Difference Between Means: Key Concepts

i) Assumptions for confidence interval on the difference between Means

To compute a confidence interval (CI) for the difference between means, we assume:

1) Equal variance: The two populations have the same variance (homogeneity of variance)
2) Normality: The populations are normally distributed
3) Independence: Each value in the sample is independent of the others

Minor violations of assumptions 1 and 2 do not significantly affect the results.

ii) formula for confidence interval on Difference between Means

The CI for the difference between two mean is given by

$$\text{lower limit} = M_1 - M_2 - (t_{CL} \cdot S_{M_1-M_2})$$

$$\text{Upper limit} = M_1 - M_2 + (t_{CL} \cdot S_{M_1-M_2})$$

3 Example calculation (Animal Research Study)

step1 : Compute the standard error of the difference between means

$$S_{M_1-M_2} = \sqrt{\frac{2 \times MSE}{n}}$$

Given data

females $n_1 = 17$, $M_1 = 5.353$, $S_1^2 = 2.743$

Males $n_2 = 17$, $M_2 = 3.882$, $S_2^2 = 2.985$

Mean Squared Error (MSE).

$$MSE = \frac{S_1^2 + S_2^2}{2} = \frac{2.743 + 2.985}{2} = 2.864$$

Compute Standard Error

$$S_{M_1-M_2} = \sqrt{\frac{2 \times 2.864}{17}} = 0.5805$$

step 2 find the t-value for a 95% Confidence interval (df=32)

from the t-table, for df = 32, $t(0.95) = 2.037$

Step3 Compute the confidence Interval

• Difference in Means

$$M_1 - M_2 = 5.353 - 3.882 = 1.471$$

. Calculate limits

$$\text{lower limit} = 1.471 - (2.037 \times 0.5805) = 0.29$$

$$\text{Upper limit} = 1.471 + (2.037 \times 0.5805) = 2.65$$

Thus, the 95% Confidence interval is

$$0.29 \le \mu_f - \mu_m \le 2.65$$

This suggests that in the population, the difference between female and male attitudes toward animal research is likely between 0.29 and 2.65 points

4) formatting Data for computer Analysis

Many statistical programs require data to be formatted with:

One column for group labels (e.g 1 = female, 2 = Male)

One column for scores

Example

| G (Group) | Y (score) |
|---|---|
| 1 | 3 |
| 1 | 4 |
| 1 | 5 |
| 2 | 5 |
| 2 | 6 |
| 2 | 7 |

This structure allows software to perform a t-test efficiently

Confidence intervals for pearson's correlation ($\rho$)

1) Standard Error of Fisher's z' Transformation

The sampling distribution of pearson's correlation coefficient ($r$) is not normally distributed. To compute a confidence interval for the population correlation $\rho$, we use Fisher's z' transformation, which converts $r$ into a variable that is approximately normally distributed. The standard error of z' is given by

$$SE_{z'} = \frac{1}{\sqrt{N-3}}$$

where N is the sample size

2) Steps to compute a confidence interval for $\rho$

1. Convert Pearson's $r$ to $z'$ using Fisher's transformation
2. Compute a confidence interval in term of $z'$
3. Convert the confidence interval back to $r$

3) Example Calculation (Animal Research Study)

Given data:

- Sample size $= N = 34$
- Sample correlation $r = -0.654$
- Fisher's $z'$ transformation of $r$: $z' = -0.78$

Step 1: compute the standard error of $z'$

$$SE_{z'} = \frac{1}{\sqrt{34-3}} = 0.18$$

Step 2: compute the 95% confidence interval for $z'$

for a 95% confidence level, the $z$ score is 1.96

Lower limit of $z' = -0.78 - (1.96 \times 0.18) = -1.13$

Upper limit of $z' = -0.78 + (1.96 \times 0.18) = -.43$

Step 3: Convert $z'$ confidence limits back to $r$

using a conversion calculator

$z' = -1.13$ corresponds to $r = -0.81$

$z' = -0.43$ corresponds to $r = -0.40$

Thus, the 95% confidence interval for $\rho$ is

$$-0.81 \le \rho \le 0.40$$

4 Compute a 99% confidence Interval

For a 99% confidence level, the zscore is 2.58

Lower limit of $z' = -0.78 - (2.58 \times 0.18) = -1$

Upper limit of $z' = -0.78 + (2.58 \times 0.18) = -0$

Converting back to r

$z' = -1.24$ corresponds to $r = -0.84$

$z' = -0.32$ corresponds to $r = -0.31$

Thus, the 99% confidence interval for p is

$$-0.84 \le p \le 0.31$$

As expected, the 99% confidence interval is wider than the 95% confidence interval, reflecting greater uncertainity

Confidence Interval for a population proportion

1) Estimating the population proportion

In a two-person election, a candidate comission a poll to estimate public support.

Sample size $N = 500$

Favorable responses $x = 260$

Sample proportion

$$p = \frac{x}{N} = \frac{260}{500} = 0.52$$

2) Standard Error of the proportion

The standard error of a sample proportion is estimated as

$$SE_p = \sqrt{\frac{p(1-p)}{N}}$$

Substituting values: $SE_p = \sqrt{\frac{0.52(1-0.52)}{500}} = 0.0223$

3 Compute the Condidence Interval (95 %)

Using the z score for 95% condidence ($z = 1.96$), the condidence interval is: $p \pm z \times SE_p$

$0.52 \pm (1.96 \times 0.0223)$

$0.52 \pm 0.0437$

To correct for continuity, we subtract add 0.5 / N (0.01):

Lower limit $= 0.52 - 0.0437 - 0.001 = 0.475$

Upper limit $= 0.52 + 0.0437 + 0.001 = 0.565$

Thus, the 95% condidence interval for $\pi$ is

$0.475 \leq \pi \leq 0.565$

4 Margin of Error

The condidence interval suggests that between 47.5% and 56.5% of voters support the candidate.

Import note

The 4.5% margin of error applies to the percentage favoring the candidate, not the difference between candidates

The margin of error for the difference between two candidates is $\sqrt{2} \times 4.5\% = 6.36\%$

This is often misreported in media polls!