Confidence Intervals

Difference in Means

Grinnell College

October 18, 2024

We saw when looking at histograms or boxplots we could compare means or medians to see if groups were different.

- maybe try to estimate the *difference* between 2 pop. means?

We just saw how to estimate a single pop. mean, let's take what we know to figure this out.

Example – Waggle Dance

Honeybee scouts investigate new home or food source options; the scouts communicate the information to the hive with a "waggle dance."

Scientists took bees to an island with only two possible options for nesting: one of very high quality and one of low quality.

They recorded:

- quality of the sites
- number of times a bee performed the dance (circuits)
- distance to the nesting sites
- duration of waggle dance

Research question: How is the number of waggle circuits related to quality of a nesting site?

- estimate the difference in pop. mean number of waggle circuits for each nesting site

Notation

2 groups $\rightarrow$ need to keep track of info separately for each of them

Group 1:

- μ_1 = pop. mean for group 1
- $\bar{x}_1$ = sample mean for group 1
- s_1 = std. dev. for group 1
- n_1 = sample size for group 1

Group 2:

- μ_2 = pop. mean for group 2
- $\bar{x}_2$ = sample mean for group 2
- s_2 = std. dev. for group 2
- n_2 = sample size for group 2

Note: Sometimes we may use A/B for subscripts or use letters that include more context about the groups

CI for Difference in Means

Our **point estimate** for $\mu_1 - \mu_2$ is unsurprisingly $\overline{x}_1 - \overline{x}_2$

Our **SE** formula is more complicated:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Our **df** value is a little different too

- df $=$ min(n_1, n_2) - 1

CI for Difference in Means

Putting this all together...

95% CI for difference in population means:

$$\overline{x}_1 - \overline{x}_2 \pm t_{(.975, df)} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

100(1-α)% CI for difference in population means:

$$\overline{x}_1 - \overline{x}_2 \pm t_{(1-\alpha/2, df)} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Note: df $=$ min(n_1, n_2) - 1

Difference in Means – Interpretation

CI Interpretation is a little more involved when we are looking for a difference in means

- include context
- mention in some way the order we are comparing means
- a positive value for the CI indicates μ_1 is larger than μ_2
  - $\mu_1 - \mu_2 > 0 \rightarrow \mu_1 > \mu_2$
- a negative value for the CI indicates μ_1 is larger than μ_2
  - $\mu_1 - \mu_2 < 0 \rightarrow \mu_1 < \mu_2$

"We are 100(1-α)% confident that (the difference in population means) is between (lower value) and (upper value)."

Difference in Means – Interpretation

"We are 100(1-α)% confident that (the difference in population means) is between (lower value) and (upper value)."

Example: Suppose we have a 90% CI of (-12.3, 24.8)

"We are 90% confident that the difference in population means is between -12.3 and 24.8."

OR

"We are 90% confident that the μ_1 is between 12.3 *lower* and 24.8 *higher* than μ_2"

OR

"We are 90% confident that the pop. mean for group 1 is between 12.3 *lower* and 24.8 *higher* than the pop. mean for group 2."

Difference in Means – Conditions

In order to make a 100(1-α)% CI for the difference in pop. means we need the following to all be true:

- There was a random sample for both groups
- $n_1 \geq 30$
- $n_2 \geq 30$
- the *groups* must be independent of each other
  - ask: do the values from one group influence values for another?
  - this is not the same thing as saying both groups behave differently