Test for equivalence, non-inferiority and superiority

Some notes about the **Confidence Interval** for population parameter:

- Estimate $\pm$ Critical value $\times$ SE(Estimator). The critical value is usually based on the standard normal table (z score) or t-table for a given α level. Here critical value $\times$ SE is called **margin of error**.
- A confidence interval contains a range of reasonable estimates of the population parameter. A 95% CI means we are 95% confident that it contains the true value of the population parameter.
- The usage of the confidence interval
  - hypothesis test. e.g., when testing for $\mu \neq \mu_0$, we reject the two sided hypothesis at level α if and only if the $100(1-\alpha)\%$ confidence interval for μ does not contain μ_0.
  - estimate sample size of a test for given margin of error and α level.
  - test for equivalence, non-inferiority and superiority of clinical treatments or medicines for certain acceptable difference (δ).

Some notes about clinical trials with active controls:

- When a valid treatment exists, it may be unethical to use a placebo as control.
- The existing treatment can be used as control in trials of new therapies.
- The goal of such trials may be to show that the new therapy is no worse than, equivalent to, or superior to the existing therapy.
- When a variant drug is introduced, the goal may show that the new medication is not inferior to the existing therapy in terms of the rate of adverse events.

1 Equivalence, non-inferiority and superiority of two proportions

Testing an equivalence hypothesis.

- Let p_1 and p_2 denote the success rates in the treatment and active control groups of a clinical trial.
- The hypotheses are
  $H_0 : |p_1 - p_2| \geq \delta$ (the two treatments are NOT equivalent)

  $H_a : |p_1 - p_2| < \delta$ (the two treatments are equivalent)

  where $\delta > 0$ is the predetermined **margin of equivalence**.
- The null hypothesis is the hypothesis of non-equivalence (p_1 and p_2 differ by at least δ), the alternative is the hypothesis of equivalence (p_1 and p_2 differ by less than δ).
- The equivalence test finds significant evidence against H_0 at level α if and only if the $100(1-\alpha)\%$ confidence interval for $p_1 - p_2$ is contained within the interval $(-\delta, \delta)$.

  That is, declare equivalence of the two treatments only when the confidence interval is contained in $(-\delta, \delta)$.

Example 1: A clinical trial was carried out to assess the efficacy of a new vaccine for pertussis. The active control was the current perussis vaccine, and the primary null hypothesis was that the proportion of subjects with anti-pertussis immune responses in group 1 (treatment) is non-equivalent to the proportion of subjects with an immune response in group 2 (control), with the alternative being that the response rates are equivalent. (In this case an immune response was defined as being a 4-fold increase in antibody titre at 4 weeks post vaccination.)

Denoting the sample size n_i and number of subjects exhibiting an immune response x_i in group i, the data were as follows:

i	n_i	X_i
1	200	163
2	197	175

- The hypotheses to be tested are

  $H_0 : |p_1 - p_2| \geq \delta$

  $H_a : |p_1 - p_2| < \delta$

  where p_1 and p_2 denote the proportions of individuals exhibiting an immune response in groups 1 and 2, and δ is the margin of non-equivalence.

- The non-equivalence margin was taken as $\delta = .15$, and there is significant evidence against the null hypothesis of non-equivalance if the confidence interval for $p_1 - p_2$ is contained in the interval (-.15,.15).

- We wish to test at level $\alpha = .05$.

- The estimated immune responses and their differences are $\hat{p}_1 = 163/200 = .815$, $\hat{p}_2 = 175/197 = .888$, $\hat{p}_1 - \hat{p}_2 = -.073$

- The standard error of $\hat{p}_1 - \hat{p}_2$ is

$$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{200} + \frac{\hat{p}_2(1-\hat{p}_2)}{197}} = .035.$$

- The 95% confidence interval for $p_1 - p_2$ is $-.073 \pm 1.96(.035)$ or (-.143,-.004).
- The confidence interval is contained in the interval $(-\delta, \delta) = (-.15, .15)$, so we can conclude that the two vaccines are equivalent.

Testing a non-inferiority hypothesis.

- Let p_1 and p_2 denote the success rates in the treatment and active control groups of a clinical trial, where **a high success rate represents a good outcome.**
- The goal of the trial is to show that the new therapy (group 1) is non-inferior to the active control (group 2), in terms of success rate.
- The null hypothesis is that the treatment is inferior to the control, and the alternative hypothesis is that treatment is non-inferior to control.
- The formal hypotheses are

  $H_0 : p_1 \leq p_2 - \delta$ (treatment 1 inferior to treatment 2)

  $H_a : p_1 > p_2 - \delta$ (treatment 1 non-inferior to treatment 2)

  where $\delta > 0$ is the predetermined **margin of non-inferiority**.
- The alternative admits the possiblity that the success rate for the treatment may be less than for the control, but not by more than δ.
- The hypotheses can be written
  $H_o : p_1 - p_2 \leq -\delta$
  $H_a : p_1 - p_2 > -\delta$
- The usual approach to testing is to find significant evidence against H_0 at level α (i.e. **declare non-inferiority of treatment 1**) if and only if the left hand endpoint of the two sided $100(1-2\alpha)\%$ confidence interval for $p_1 - p_2$ is greater than $-\delta$.
- **The confidence level** $100(1-2\alpha)\%$ **is used instead of** $100(1-\alpha)\%$ **because the alternative hypothesis is one sided.**

Example 2: a non-inferiority test for the pertussis vaccine data.

- The hypotheses are then

  $H_0 : p_1 \leq p_2 - \delta$

  $H_a : p_1 > p_2 - \delta$, or $p_1 - p_2 > -\delta$.

- The alternative admits p_1 being less than p_2, but only up to the margin δ.

- Assume that we are again testing at level $\alpha = .05$.

- In this case we find statistically significant evidence against H_0 if the lower end of the 90% confidence interval on $p_1 - p_2$ is greater than $-\delta$.

- Here we need $z_{.05} = -1.645$, and the 90% confidence interval is $-.073 \pm 1.645(.035)$ or (-.132,-.015).

- Because the lower end of the 90% confidence interval for $p_1 - p_2$ is greater than -.15, we declare that the new vaccine (treatment 1) is non-inferior to the old (treatment 2).

Testing a superiority hypothesis.

- Here the goal is to show that treatment 1 is superior to treatment 2 (i.e. $p_1 > p_2 + \delta$).
- The hypotheses are

  $H_0 : p_1 \leq p_2 + \delta$ (treatment 1 not superior to treatment 2)

  $H_a : p_1 > p_2 + \delta$ (treatment 1 superior to treatment 2)

  where $\delta > 0$ is the pre-determined **margin of superiority**.
- The hypotheses can be written

  $H_0 : p_1 - p_2 \leq \delta$

  $H_a : p_1 - p_2 > \delta$
- This looks exactly like the non-inferiority test, except for the sign on the difference under the hypotheses.
- The usual approach to testing is to find significant evidence against H_0 at level α (i.e. **declare superiority of treatment 1**) if and only if the left hand endpoint of the two sided $100(1-2\alpha)\%$ confidence interval for $p_1 - p_2$ is greater than δ.
- **The confidence level** $100(1-2\alpha)\%$ **is used instead of** $100(1-\alpha)\%$ **because the alternative hypothesis is one sided.**

Example 3: a superiority test for the pertussis vaccine data.

- The hypotheses are

  $H_0 : p_1 \leq p_2 + \delta.$

  $H_a : p_1 > p_2 + \delta.$

- The superiority hypothesis (the alternative) states that the immune response rate for the new vaccine exceeds that for the old vaccine by at least δ.

- We find statistically significant evidence against H_0 if the lower limit of the $100(1 - 2\alpha)\%$ confidence interval on $p_1 - p_2$ is greater than δ.

- Because the lower end of the 90% confidence interval for $p_1 - p_2$, being (-.132,-.015), is NOT greater than .15, we do not have sufficient evidence to declare superiority of the new vaccine.

2 Equivalence, noninferiority or superiorty of means of normal populations

Procedures are essentially identical to testing for differences of proportions. Just use the pooled t confidence interval for the difference of two means in place of the CI for difference of two proportions.

- **Testing an equivalence hypothesis.**

  $H_0 : |\mu_A - \mu_B| \geq \delta$ (A and B are NOT equivalent)
  $H_A : |\mu_A - \mu_B| < \delta$ (A and B are equivalent)

  where $\delta > 0$ is the predetermined margin of equivalence.

  The null hypothesis is the hypothesis of non-equivalence, the alternative is the hypothesis of equivalence.

  The usual approach to testing is to reject H_0 at level α if and only the $100(1-\alpha)\%$ confidence interval for $\mu_A - \mu_B$ is contained within the interval $(-\delta, \delta)$.

- **Testing a noninferiority hypothesis.**

  Suppose that large values of the mean represent a favourable outcome, and that the goal is to show that treatment A is non-inferior to B (ie. $\mu_A > \mu_B - \delta$).

  The hypotheses are
  $H_0 : \mu_A \leq \mu_B - \delta$ (A inferior to B)
  $H_A : \mu_A > \mu_B - \delta$ or $\mu_A - \mu_B > -\delta$ (noninferiority of A)

  where $\delta > 0$ is the predetermined margin of non-inferiority.

The usual approach to testing is to reject H_0 at level α if and only if the **left hand endpoint of the two sided** $100(1-2\alpha)\%$ confidence interval for $\mu_A - \mu_B$ is greater than $-\delta$.

- **Testing a superiority hypothesis.**

  Suppose that large values of the mean represent a favourable outcome, and that the goal is to show that treatment A is superior to B ($\mu_A > \mu_B + \delta$).

  The hypotheses are

  $H_0 : \mu_A \leq \mu_B + \delta$. (A is NOT superior to B)

  $H_A : \mu_A > \mu_B + \delta$ or $\mu_A - \mu_B > \delta$ (A superior to B)

  where $\delta > 0$ is the predetermined margin of superiority.

  The usual approach to testing is to reject H_0 at level α if and only if the **left hand endpoint of the two sided** $100(1-2\alpha)\%$ confidence interval for $\mu_A - \mu_B$ is greater than δ.

Example: A study was carried out to compare two population means, μ_1 and μ_2. A 95% confidence interval for $\mu_1 - \mu_2$ was calculated as (-.06, 2.46), and a 90% confidence interval was calculated as (.16, 2.24).

- When carrying out an equivalence test of $H_0 : |\mu_1 - \mu_2| \geq 1$ vs $H_A : |\mu_1 - \mu_2| < 1$ at level .05, would you reject the null hypothesis? Why?

  No, because the 95% CI for $\mu_1 - \mu_2$ is NOT contained in (-1, 1).

- Assuming that a large value of the mean is a favourable outcome, when carrying out a non-inferiority test of $H_0 : \mu_1 \leq \mu_2 - 1$, vs $H_A : \mu_1 > \mu_2 - 1$ at level .05, would you conclude that treatment 1 (which gave mean μ_1) is noninferior to treatment 2 (which gave mean μ_2). Why?

  Yes, because the lower end of the 90% confidence interval for $\mu_1 - \mu_2$ is GREATER THAN -1.

- Assuming that a large value of the mean is a favourable outcome, when carrying out a superiority test of $H_0 : \mu_1 \leq \mu_2 + 1$, vs $H_A : \mu_1 > \mu_2 + 1$ at level .05, would you conclude that treatment 1 (which gave mean μ_1) is superior to treatment 2 (which gave mean μ_2). Why?

  No, because the lower end of the 90% confidence interval for $\mu_1 - \mu_2$ is SMALLER THAN 1.