Chapter 10 - Hypothesis Testing with Two Populations

Contents

1. Comparing Two Population Means: Independent Samples 1
- 1.1. Hypotheses 1
- 1.2. Conditions for the significance test for $\mu_1 - \mu_2$ 1
- 1.3. Properties of Sampling Distribution of $\bar{X}_1 - \bar{X}_2$ 1
- 1.4. Properties of Sampling Distribution $\bar{x}_1 - \bar{x}_2$ 2
- 1.5. Degrees of Freedom (df) 2
- 1.6. p-value and Conclusion 2

1. Comparing Two Population Means: Independent Samples

1.1. Hypotheses

Null Hypothesis:

$$H_0 : \mu_1 - \mu_2 = D_0 \qquad \text{or} \qquad H_0 : \mu_1 = \mu_2 \text{ if } D_0 = 0$$

D_0 is the hypothesized difference between the two means.

1.2. Conditions for the significance test for $\mu_1 - \mu_2$

- Quantatative varaible with μ_1 and μ_2 defined in context
- Data is obtained using randomization (like simple random sampling)
- Both population distributions are approximately normal (or $n_1 + n_2 \geq 20$)
- Populations are independent, which result in independent samples
- Population standard deviations are unknown.

1.3. Properties of Sampling Distribution of $\bar{X}_1 - \bar{X}_2$

- The mean of the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is $\mu_1 - \mu_2$; that is

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$$

- The standard deviation ("standard error") of the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

1.4. Properties of Sampling Distribution $\bar{x}_1 - \bar{x}_2$

Because we don't know population standard deviations, they are estimated using the two sample standard deviations from our independent samples.
The two-sample **test statistic** is calculated as:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where

- $\bar{x}_1$ and $\bar{x}_2$ are the sample means
- s_1 and s_2 are the sample standard deviations
- n_1 and n_2 are the sample sizes of each sample.

The Test statistic gives us how many standard errors $\bar{x}_1 - \bar{x}_2$ is away from D_0

1.5. Degrees of Freedom (df)

The t-table is used to find the p-value based on the degrees of freedom.
The test statistic is approximated using the t-distribution with df as follows:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{1}{n_1-1}\right)\left(\frac{s_1^2}{n_1}\right)^2 + \left(\frac{1}{n_2-1}\right)\left(\frac{s_2^2}{n_2}\right)^2}$$

1.6. p-value and Conclusion

Example:
The Kona Corporation produces coconut milk. They have both a day shift (called the B shift) and a night shift (call the G shift) to do the process. They would like to know if the day shift and the night shift are equally efficient in processing the coconuts. A study is done sampling 9 G shifts and 16 B shifts. The average number of hours to process 100 pounds of coconuts for G shift and B shift are 2 and 3.2 hours. The sample standard deviation for the G and B shift are 0.866 and 1.00. The number of hours to process 100 pounds of coconuts for both shifts are normally distributed. Is there a difference in the mean amount of time for each shift to process 100 pounds of coconuts? Test at the 5% level of significance.

(solved on next page.)

- first get all data from the question

$n_1 = 16 \qquad \bar{x} = 3.2 \qquad s_1 = 1$
$n_2 = 9 \qquad \bar{x} = 2 \qquad s_2 = 0.866$

- determine if it meets conditions

  Simple random sample, both populations quantitative, normal population

- get the t-statistic

  To get the t-statistic, plug data into formula:

$$t = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(3.2 - 2) - 0}{\sqrt{\frac{(1)^2}{16} + \frac{(0.866)^2}{9}}} = 3.142$$

  now we need degrees of freedom

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{1}{n_1 - 1}\right)\left(\frac{s_1^2}{n_1}\right)^2 + \left(\frac{1}{n_2 - 1}\right)\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{(1)^2}{16} + \frac{(0.866)^2}{9}\right)^2}{\left(\frac{1}{16} - 1\right)\left(\frac{1^2}{16}\right)^2 + \left(\frac{1}{9} - 1\right)\left(\frac{0.866^2}{9}\right)^2} \approx 18.847$$

  always round df up meaning 19 degrees of freedom.
  We find our t-statistic on the t-table with a value of 2.861.
  Conclusion:
  $t = 3.142 > 2.861$ on t-table, so $\frac{p\text{-value}}{2} < 0.005$
  p-value $< 0.01 < \alpha = 0.05$
  $\therefore$ reject H_0