Comparing two means

The Big Idea:

Suppose a simple random sample of size \( n_1 \) is taken from one population, and another simple random sample of size \( n_2 \) is taken from another. We can estimate the means of these two populations by forming \( \bar{x}_1 \) and \( \bar{x}_2 \), respectively. Our basic questions are: is there a difference between the two population means, and if so, what is it?

Formulas

Just as in the single sample regime, there are two cases, one where the standard deviations are known, and one where they are not. Since we generally do not know the standard deviations, we focus on the case that they are need to be estimated from the data.

- A an approximate \( C \) level confidence interval for \( \mu_1 - \mu_2 \) is

\[
\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.
\]

Here, \( t^* \) is the value for the \( T(k) \) density curve with area \( C \) between \( -t^* \) and \( t^* \), where \( k \) is the smaller of \( n_1 - 1 \) and \( n_2 - 1 \).

- To calculate \( P \)-values for a null hypothesis of the form \( H_0 : \mu_1 = \mu_2 \), first form the test statistic

\[
t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.
\]

Let \( k \) be the smaller of \( n_1 - 1 \) and \( n_2 - 1 \). Then

- if \( H_a : \mu_1 < \mu_2 \), the approximate \( P \)-value is the area under a \( T(k) \) curve to the left of \( t \).
- if \( H_a : \mu_1 > \mu_2 \), the approximate \( P \)-value is the area under a \( T(k) \) curve to the right of \( t \).
- if \( H_a : \mu \neq \mu_0 \), the approximate \( P \)-value is the area under a \( T(k) \) curve to the right of \( |t| \) and to the left of \(-|t|\).

Practice Problem

1. (7.64 in text) Assume \( \bar{x}_1 = 100, \bar{x}_2 = 110, s_1 = 18, s_2 = 15, n_1 = 50, n_2 = 40 \). Find a 95% confidence interval for the difference in the corresponding values of \( \mu \). Also find a 99% confidence interval.

\[
t^* \text{ taken from } \chi^2(39) \text{ curve. } \chi^2(40) \text{ is the closest one on table, so use that. } t^* \text{ for } 95\% \text{ C.I. is } t^* = 2.021
\]

So: 95% C.I. : \( \bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = [10 \pm 7.03] \)

for 99% C.I., \( t^* = 2.704 \). So 99% C.I. : \([[-10 \pm 9.41]]\)