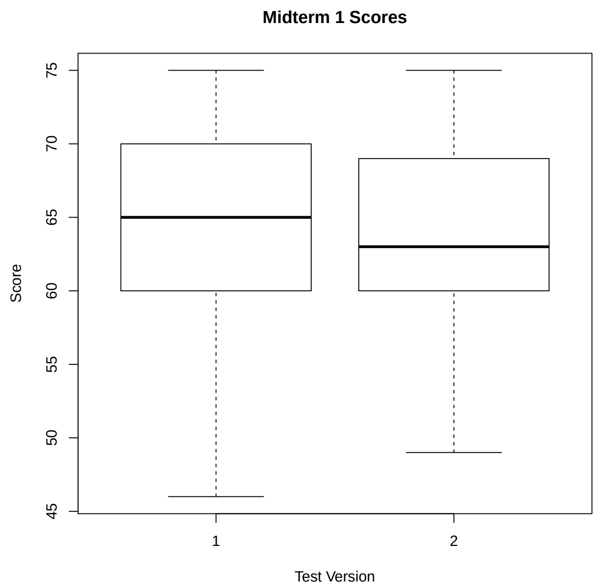
Comparing Two Populations – Solutions

STAT-UB.0103 – Statistics for Business Control and Regression Models

Comparing Two Populations

1. Here are boxplots of the scores from midterm 1 for the two versions of the test (outliers have been removed). Is there evidence that one test was harder than the other?



For the first version, 37 students took the test; the mean score was 64.4 and the standard deviation was 7.0. For the second version, 38 students took the test; the mean score was 63.9 and the standard deviation was 6.8.

What does it mean for one test to be harder than another? Be as precise as possible.

2. This is a continuation of the previous problem (testing whether or not one test was harder than another). Suppose you want to perform a hypothesis test at level 5%.

   (a) What are the populations?

   > **Solution:** Population 1: the scores from version 1 if you could give the test to all students in the class.
   >
   > Population 2: the scores from version 1 if you could give the test to all students in the class.

   (b) What are the null and alternative hypotheses?

   > **Solution:**
   >
   > $$\begin{aligned} H_0 &: \mu_1 = \mu_2 \quad \text{(same mean difficulty)} \\ H_1 &: \mu_1 \neq \mu_2 \quad \text{(one test is harder than the other)} \end{aligned}$$

   (c) What are the samples?

   > **Solution:** The scores for the students who actually took the two tests.

   (d) What is the test statistic?

   > **Solution:**
   >
   > $$\begin{aligned} \bar{x}_1 - \bar{x}_2 &= 64.4 - 63.9 \\ &= 0.5 \\ \mathrm{SE}(\bar{x}_1 - \bar{x}_2) &= \sqrt{\frac{(7.0)^2}{37} + \frac{(6.8)^2}{38}} \\ &= 1.6 \\ z &= \frac{\bar{x}_1 - \bar{x}_2}{\mathrm{SE}(\bar{x}_1 - \bar{x}_2)} \\ &= \frac{0.5}{1.6} \\ &= 0.3 \end{aligned}$$

   (e) What is the rejection region?

   > **Solution:** Reject H_0 if $|z| > 2$.

(f) What is the result of the test?

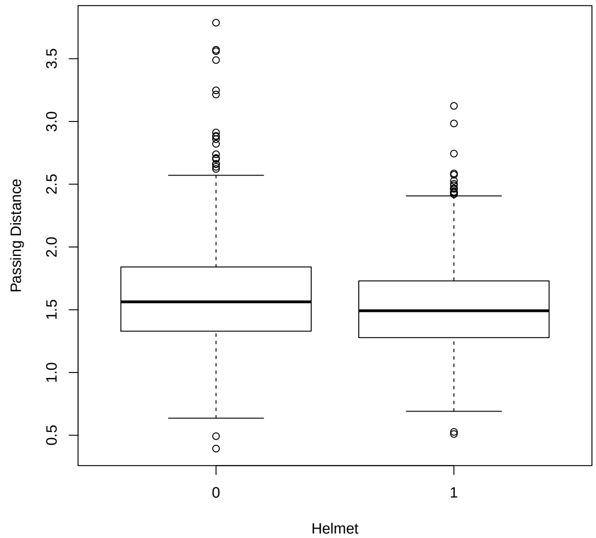
Solution: Since $|0.3| \leq 2$, we do not reject H_0. We do not have significant test that one test was harder than the other.

(g) Approximately what is the p-value?

Solution:

$$2\Phi(-0.3) \approx 0.76$$

3. Here are boxplots of the passing distances (in meters) for a bike rider with and without a helmet. Is there evidence that the passing distance differs when the rider has a helmet?



Here are the sample statistics for the passing distance without a helmet: $n_1 = 1206$, $\bar{x}_1 = 1.61$, $s_1 = 0.405$. Here are the sample Here are the sample statistics for the passing distance with a helmet: $n_2 = 1149$, $\bar{x}_2 = 1.52$, $s_2 = 0.354$.

Formulate the problem as a hypothesis test, using significance level 5%.

(a) What are the populations?

> **Solution:** Population 1: all passing distances while not wearing a helmet.
> Population 2: all passing distances while wearing a helmet.

(b) What are the null and alternative hypotheses?

> **Solution:**
>
> $$H_0 : \mu_1 = \mu_2 \quad \text{(same mean distance for both populations)}$$
> $$H_a : \mu_1 \neq \mu_2.$$

(c) What are the samples?

> **Solution:** All recorded passing distances.

(d) What is the test statistic?

Solution:

$$
\begin{aligned}
\bar{x}_1 - \bar{x}_2 &= 1.61 - 1.52 \\
&= 0.09 \\
\mathrm{SE}(\bar{x}_1 - \bar{x}_2) &= \sqrt{\frac{(0.405)^2}{1206} + \frac{(0.354)^2}{1149}} \\
&= 0.016 \\
z &= \frac{\bar{x}_1 - \bar{x}_2}{\mathrm{SE}(\bar{x}_1 - \bar{x}_2)} \\
&= \frac{0.09}{0.016} \\
&= 5.6
\end{aligned}
$$

(e) What is the rejection region?

Solution: Reject H_0 if $|z| > 2$.

(f) What is the result of the test?

Solution: Reject H_0 since $|5.6| > 2$.

Confidence Intervals

4. Find a 95% confidence interval for the difference in mean score between the two midterm versions.

> **Solution:** Above, we computed
> $$\bar{x}_1 - \bar{x}_2 = 0.5$$
> and
> $$\mathrm{SE}(\bar{x}_1 - \bar{x}_2) = 1.6.$$
> Thus, a 95% confidence interval for $\mu_1 - \mu_2$ is
> $$\begin{aligned} (\bar{x}_1 - \bar{x}_2) \pm 2 \cdot \mathrm{SE}(\bar{x}_1 - \bar{x}_2) \\ 0.5 \pm 2(1.6) \\ 0.5 \pm 3.2. \end{aligned}$$
> With 95% confidence, the difference in mean score between the two versions is between -2.7 and 3.7.

5. Find a 95% confidence interval for the difference in passing difference with and without a helmet. *Hint:* $\bar{x}_1 - \bar{x}_2 = 0.09$ *and* $SE(\bar{x}_1 - \bar{x}_2) = 0.016$.

> **Solution:**
> $$0.09 \pm 2(0.016)$$
> With 95% confidence, the difference in population means is between 0.58 and 1.22 meters.