Comparing Two Population Means: Independent Sampling (Revisited)

(*See pp.352-364*)

On Thursday, November 20, we compared a new method of teaching reading to the current standard method. Our comparison was based on the results of a reading test given at the end of a learning period. Of a random sample of 22 learners, 10 were taught by the new method and 12 were taught by the standard method. All 22 learners were taught by qualified instructors under similar conditions. The results we considered are shown in the table below. (The data is in the CD file **READING**.)

C3	C4
NewScore	StandardScore
80	79
76	73
70	72
80	62
66	76
85	68
79	70
71	86
81	75
76	68
	73
	66

We found that the mean score on the post test for those taught by the new method was $\mu_1 = 76.4$ and the mean score on the post test for those taught by the standard method was $\mu_2 = 72.33$. We wondered if the difference in the two means is sufficient to allow us to make a valid comparison of the relative effectiveness of the two methods.

In this case we considered two population means with small ($n < 30$) samples; so we used the *t*-distribution.

To use the t-distribution, both sampled populations must be approximately normal with equal variances, and the random samples must be selected independently of each other.

Our hypotheses were as follows:

H_0: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 \neq 0$

We used our data to estimate the true mean difference between the test scores of the two methods. We constructed a 95% **($\alpha = 0.05$)** confidence interval for ($\mu_1 - \mu_2$). MINITAB output is shown below. (p. 428) (*In our notes for November 20, we showed output for a TI graphing calculator*.)

Two sample T for SCORE

METHOD	N	Mean	StDev	SE Mean
NEW	10	76.40	5.83	1.8
STD	12	72.33	6.34	1.8

95% CI for mu (NEW) - mu (STD): (-1.4, 9.5)

T-Test mu (NEW) = mu (STD) (vs not =): T = 1.55 P = 0.14 DF = 20 ($n_1 + n_2 - 2$) Both use Pooled StDev = 6.12

We estimated with (with 95% confidence) the mean test score for the new method was anywhere between 1.4 points less than, to 9.5 points more than, the mean test score for the standard method.

There was insufficient evidence to indicate that $\mu_1 - \mu_2 \neq 0$ because the 95% CI included 0 as a possible value. Alternatively, a test of hypothesis proceeded as follows. We used a two-tailed test in this case.

Our rejection region: $t < -t_{\alpha/2} = -t_{0.025} = -2.086$ or $t > t_{\alpha/2} = t_{0.025} = 2.086$ based on 20 df.

Since the observed value of *t* (1.55) was not in the rejection region we did not reject H_0. As above, we could not reject the premise that the two populations have the same mean. (At the 0.05 confidence level, we could not conclude that the new method was better.)

Comparing Two Population Means: Paired Difference Experiments
(See pp. 370-379)

In the previous example, suppose it was possible to measure the reading aptitudes or "reading IQ's" of the learners *before* they were subjected to a teaching method. In this case we find eight pairs of learners with similar reading aptitudes, and ne member of each pair is randomly assigned to the standard method while the other is assigned to the new method. The data are shown in the table below. In this case, do the data support the hypothesis that the mean test score for the learners taught by the new method is greater than the mean test score for those taught by the standard method?

PAIR	NEW	STANDARD	DIFFERENCE
1	77	72	5
2	74	68	6
3	82	76	6
4	73	68	5
5	87	84	3
6	69	68	1
7	66	61	5
8	80	76	4

Our hypotheses are as follows:

H_0: $(\mu_1 - \mu_2) = 0$

H_a: $(\mu_1 - \mu_2) > 0$

In this case we cannot use the two-sample *t*-test because the assumption of independent samples is violated. We randomly chose pairs of pretest scores, thus we have not independently chosen the samples for the two methods.

Our approach is to consider the differences in test scores as a random sample of differences for all pairs of matched by aptitude learners past and present. We then use this sample to make inferences about the true mean of the population differences which is equal to the difference $(\mu_1 - \mu_2)$. So, our test becomes

$$H_0:\ \mu_d = 0 \qquad (\mu_1 - \mu_2) = 0$$

$$H_a:\ \mu_d > 0 \qquad (\mu_1 - \mu_2) > 0$$

The test statistic is a one-sample *t*-test since we are now analyzing a single small sample of differences.

Test statistic: $t = \dfrac{\bar{x}_d - 0}{s_d/\sqrt{n_d}}$

where $\bar{x}_d = Sample\ mean\ difference$

$s_d = Sample\ standard\ deviation\ of\ differences$

$n_d = Number\ of\ differences = Number\ of\ pairs$

Assumptions: The distribution population of differences in test scores is approximately normal. The sample differences are randomly selected from the population differences.

Rejection region: At a significance level of $\alpha = 0.05$ we will reject H_0 if $t > t_{0.05} = 1.895$ where $t_{0.05}$ is based on $(n_d - 1) = 7$ degrees of freedom.

We can calculate $n_d = 8$, $\bar{x}_d = 4.375$, $s_d = 1.685$. So, $t = \frac{4.375}{1.685/\sqrt{8}} \approx 7.34$.

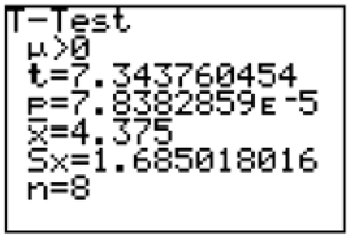
Since the value of *t* falls in the rejection region, we conclude (at the $\alpha = 0.05$ significance level) that the population mean test score for learners taught by the new method exceeds the population mean score for those taught by the standard method. We can also reach this conclusion by noting that the p-value of the test is much smaller than $\alpha = 0.05$. (See the MINITAB output below.)

```
Paired T for NEW - STANDARD

             N     Mean    StDev  SE Mean
NEW          8  76.0000   6.9282   2.4495
STANDARD     8  71.6250   7.0089   2.4780
Difference   8  4.37500  1.68502  0.59574


95% lower bound for mean difference: 3.24632
T-Test of mean difference = 0 (vs > 0): T-Value = 7.34  P-Value = 0.000
```

Instructions for doing this test with a TI graphing calculator can be found on pp. 378-379.



The hypothesis testing procedures for the difference between two means in a paired difference experiment are summarized on page 374 for both large and small n. Rather than using the t-statistic, we may use the z-statistic in the case of a large ($n_d \geq 30$) sample size.

Exercise 1: Comparing Two Population Means: Independent Sampling (LM7_9)
Independent random samples from normal populations produced the following results:

Sample 1: 1.2, 3.1, 1.7, 2.8, 3.0　　　　Sample 2: 4.2, 2.7, 3.6, 3.9

a. Do the data provide sufficient evidence to indicate that $\mu_1 > \mu_2$? Test, using $\alpha = 0.10$.
b. Find a 90% confidence interval for $(\mu_1 - \mu_2)$.
c. Which procedure a. or b. provides more information about $(\mu_1 - \mu_2)$?

Exercise 2. Comparing Two Population Means: Paired Difference Experiments (LM&_33)
Data for a random sample of six paired observations are shown below.

Population 1	7	3	9	6	4	8
Population 2	4	1	7	2	4	7
Difference						

a. Calculate the differences between each pair of observations by subtracting observation 2 from observation 1. Calculate $\bar{x}_d$ and s_d^2.
b. If μ_1 and μ_2 are the means of populations 1 and 2 respectively, express μ_d in terms of μ_1 and μ_2.
c. Form a 95% confidence interval for μ_d.
d. Test the null hypothesis H_0: $\mu_d = 0$ against the alternative hypothesis H_a: $\mu_d \neq 0$. Use $\alpha = 0.05$.