Question 4

Intent of Question

The primary goal of this question is to assess a student's ability to conduct a test of significance by stating the hypotheses of interest, checking the necessary conditions, calculating the test statistic and *p*-value, and making a conclusion in context.

Solution

Step 1: States a correct pair of hypotheses.

Let μ_D denote the mean difference (after – before) in dexterity scores for the population of individuals enrolled in the program.
$H_0 : \mu_D = 0$ versus $H_a : \mu_D > 0$

Step 2: Identifies a correct test (by name or formula) and checks appropriate conditions.

One sample *t*-test or paired *t*-test or $t = \frac{\bar{x}_D}{s_D / \sqrt{n}}$.

We are told that the 12 people are a random sample. Assume that the differences (after – before) are approximately normal. This check may be done with a histogram, dotplot, stem-and-leaf display, or normal probability plot. The student should note that the normal assumption is not unreasonable because the plot displays no obvious skewness or outliers.

Step 3: Correct mechanics, including the value of the test statistic and the *p*-value (or rejection region).

$\bar{x}_D = 0.375$, $s_D = 0.367$
Degrees of freedom $= 12 - 1 = 11$

$$t = \frac{0.375}{\frac{0.367}{\sqrt{12}}} = 3.54$$

$p\text{-value} = 0.002$

Step 4: States a correct conclusion in the context of the problem.

Since the *p*-value is less than 0.05, we can reject the null hypothesis of no difference in favor of the alternative and conclude that, on average, people who completed the program have significantly increased manual dexterity.

Scoring

Each of the four steps is scored as essentially correct (E) or incorrect (I).

Question 4 (continued)

Notes for Step 2:

Although it is not recommended, a one sample confidence interval for μ_D could be used to test the hypotheses in Step 1. An appropriate adjustment to the confidence level must be made since we are conducting a one-sided test. The correct formula is

$$\left(\bar{x}_D - t^*_{n-1,\alpha}\frac{s_D}{\sqrt{n}},\ \infty\right) \Rightarrow \left(0.375 - 1.796 \times \frac{0.367}{\sqrt{12}},\ \infty\right) \Rightarrow (0.1847,\ \infty).$$

The null hypothesis of no change in mean dexterity scores is rejected at the 0.05 level of significance because the right end of this 95 percent one sided confidence interval is above zero. If the *t*-value used to constructing the confidence interval does not match the significance level given in the conclusion, then the maximum score for Step 4 is partially correct (P).

If an incorrect two sample procedure is used, then Step 2 is scored as incorrect. The maximum score for a two sample t procedure is 3.

Incorrect Solutions for Step 2			
Procedure	**df**	**Test Statistic**	***p*-value**
Two sample *t*-test	21.98	$t = 1.05$	0.153
Pooled *t*-test	22	$t = 1.05$	0.152

A response using separate confidence intervals for the two means is also scored as incorrect for Step 2.

Notes for Step 3:

An identifiable minor arithmetic error is Step 3 will not necessarily change a score from essentially correct to incorrect.

If the student argues that the normal distribution is not reasonable, then they may use hypothesis tests for the median.

Other Solutions for Step 3		
Procedure	**Test Statistic**	***p*-value**
Sign Test	$B = 8$	0.0547
Wilcoxon Signed Rank Test	$W = 52$	0.007

If the *p*-value is incorrect but the conclusion in Step 4 is consistent with the computed *p*-value, Step 4 can be considered essentially correct.

Notes for Step 4:

If both an α and a *p*-value are given, the linkage in Step 4 is implied. If no α is given, the solution must be explicit about the linkage by giving a correct interpretation of the *p*-value or explaining how the conclusion follows from the *p*-value.

If the hypotheses are reversed in Step 1 (i.e., $H_0 : \mu_D > 0$ versus $H_a : \mu_D = 0$), then the conclusion also needs to be reversed. Otherwise, both parts should be scored as incorrect (I).

Question 4 (continued)

4. The developers of a training program designed to improve manual dexterity claim that people who complete the 6-week program will increase their manual dexterity. A random sample of 12 people enrolled in the training program was selected. A measure of each person's dexterity on a scale from 1 (lowest) to 9 (highest) was recorded just before the start of and just after the completion of the 6-week program. The data are shown in the table below.

Person	Before Program	After Program
A	6.7	7.8
B	5.4	5.9
C	7.0	7.6
D	6.6	6.6
E	6.9	7.6
F	7.2	7.7
G	5.5	6.0
H	7.1	7.0
I	7.9	7.8
J	5.9	6.4
K	8.4	8.7
L	6.5	6.5
Total	81.1	85.6

Can one conclude that the mean manual dexterity for people who have completed the 6-week training program has significantly increased? Support your conclusion with appropriate statistical evidence.

One-Sample t-Test for Mean Difference

$H_0: \mu_D = 0$ where $\mu_D = \mu_1 - \mu_2$

$H_a: \mu_D < 0$

μ_1 = mean manual dexterity before program

μ_2 = mean manual dexterity after program

Conditions: A random sample of differences in manual dexterity before and after the training program has been selected from the population of differences (given, as a random sample of people enrolled in said program has been selected). The population of differences appears to be approximately normally distributed, as the histogram of sample differences () does not display extreme skewness or outliers.

$$t = \frac{\bar{x}_D - \mu_0}{\frac{s_D}{\sqrt{n_D}}} = \frac{-0.375 - 0}{\frac{0.3671}{\sqrt{12}}} = -3.539$$

p-value = $P(t < -3.539) = 0.00232$

* w/ df = 12 − 1 = 11

Since the pvalue ($p = 0.00232$) does not exceed our selected significance level ($\alpha = 0.01$), we can reject the null hypothesis. Hence, there is sufficient, statistically significant evidence to conclude that the mean manual dexterity for people who have completed the 6-week training program has significantly increased.

GO ON TO THE NEXT PAGE.

4. The developers of a training program designed to improve manual dexterity claim that people who complete the 6-week program will increase their manual dexterity. A random sample of 12 people enrolled in the training program was selected. A measure of each person's dexterity on a scale from 1 (lowest) to 9 (highest) was recorded just before the start of and just after the completion of the 6-week program. The data are shown in the table below.

Person	Before Program	After Program
A	6.7	7.8
B	5.4	5.9
C	7.0	7.6
D	6.6	6.6
E	6.9	7.6
F	7.2	7.7
G	5.5	6.0
H	7.1	7.0
I	7.9	7.8
J	5.9	6.4
K	8.4	8.7
L	6.5	6.5
Total	81.1	85.6

Can one conclude that the mean manual dexterity for people who have completed the 6-week training program has significantly increased? Support your conclusion with appropriate statistical evidence.

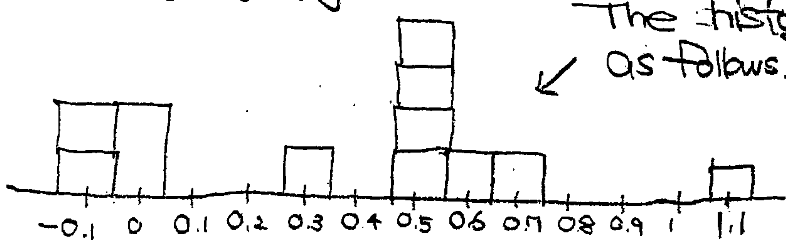
We perform a t-test for the difference between the means.

ho: ~~$\mu_{before} < \mu_{after}$~~ ha: ~~$\mu_{before} \geq \mu_{after}$~~, $\mu_{after} - \mu_{before} < 0$

The set of the differences between measures of dexterity after program and those before program (random sample)

: { 1.1, 0.5, 0.6, 0, 0.7, 0.5, 0.5, −0.1, −0.1, 0.5, 0.3, 0 }

The histogram of the set is as follows:



The distribution of the set has no significant skewness and outliers, so we can assume that population follows normal distribution

$\bar{x}_d = 0.375$ df = 11

$S_d = 0.38$

$$t = \frac{\bar{x}_d - 0}{\frac{S}{\sqrt{n}}} = \frac{0.375 - 0}{\frac{0.38}{\sqrt{12}}} = 3.42$$

GO ON TO THE NEXT PAGE.

The P-value is about 0.997.

Since the P-value is extremely large, much larger than 0.1, 10% significance level, there is little evidence to reject the claim that the training program improves manual dexterity.

4B2

NO TEST MATERIAL ON THIS PAGE

GO ON TO THE NEXT PAGE.

4. The developers of a training program designed to improve manual dexterity claim that people who complete the 6-week program will increase their manual dexterity. A random sample of 12 people enrolled in the training program was selected. A measure of each person's dexterity on a scale from 1 (lowest) to 9 (highest) was recorded just before the start of and just after the completion of the 6-week program. The data are shown in the table below.

Person	Before Program	After Program
A	6.7	7.8
B	5.4	5.9
C	7.0	7.6
D	6.6	6.6
E	6.9	7.6
F	7.2	7.7
G	5.5	6.0
H	7.1	7.0
I	7.9	7.8
J	5.9	6.4
K	8.4	8.7
L	6.5	6.5
Total	81.1	85.6

Can one conclude that the mean manual dexterity for people who have completed the 6-week training program has significantly increased? Support your conclusion with appropriate statistical evidence.

I will use the hypothesis test to ~~m~~ see if the ~~mean~~ mean has increased after the training proggramm. Let μ_1 be the mean before the programm and μ_2 - mean after the programm

H_0: ~~$\mu_1 = \mu_2$~~ $\mu_2 - \mu_1 = 0$

H_A: ~~$\mu_1 < \mu_2$~~ $\mu_2 - \mu_1 > 0$

~~B~~ Before the program: $\bar{X}_1 = 6.758$ $S_x = 0.887$ $n = 12$

After the programm: $\bar{Y} = 7.133$ $S_x = 0.861$ $n = 12$

The statistic for the difference in means $\mu_2 - \mu_1$ is $\bar{X}_2 - \bar{X}_1$. The standard deviation is $\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

~~p-value = Pr (~~ As sample size is small, $(n=12)$, t-statistic is used with $d.f = n-1 = 11$. $t_{(df=11)} = \frac{\bar{X}_2 - \bar{X}_1 - H_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{0.375 - 0}{0.3568} = 1.051$

p-value $= P_r(t > 1.051) = 0.1579$

p-value is high. It is higher than all appropriate significance levels (such as 0.1 and 0.05). Which means we can accept the null hypothesis ~~with~~ at 90% and 95% confidence level.

The was ~~not~~ no mean increace in dexterity after training

GO ON TO THE NEXT PAGE.

Question 4

Sample: 4A
Score: 4

This essay recognizes the paired nature of the data and correctly specifies a one-sample *t*-test for the mean difference. The null hypothesis and appropriate one-sided alternative hypotheses are clearly stated with notation for the population means clearly defined. This enables one to know that the symbol for the mean difference refers to the mean manual dexterity before entering the program minus the mean manual dexterity after completing the program. It is important that the direction of this difference is well defined. This essay addresses the model assumptions underlying the use of the one-sample *t*-test by presenting a histogram of the observed differences to check for outliers and the shape of the distribution of differences and concludes that the data in this small sample present no strong reason to doubt the assumption of a normal distribution for differences. The essay also notes that a simple random sample of subjects was provided in the stem of the question. Simple random sampling could be used to help justify the assumption that the subjects respond independently of each other and are representative of the population from which they were selected. The value of the test statistic, degrees of freedom, and the *p*-values are correctly calculated. A correct conclusion is reached in the context of the problem and justified by comparing the *p*-value to a .01 significance level.

Sample: 4B
Score: 3

A one-sample *t*-test for the mean difference is identified by statement and by formula. However, the null hypothesis and one-sided alternative hypotheses are in the wrong direction. A histogram of the observed differences is used to check for outliers and the shape of the distribution of differences. The standard deviation for the differences is not computed correctly, but the *t*-statistic is correctly evaluated from the incorrect standard deviation. The *p*-value is very large, but it is consistent with the stated null and alternative hypotheses, and the conclusion that is reached is also consistent with the *p*-value and reversed hypotheses. This essay recognizes the paired nature of the data and shows a good understanding of computing a *t*-statistic and reaching a conclusion, but there is some confusion in determining the appropriate null and alternative hypothesis from the context of the problem.

Sample: 4C
Score: 2

This essay fails to recognize that before and after responses should be treated as paired data instead of independent samples. Appropriate null and alternative hypotheses are stated, but a *t*-test for two independent samples is specified. There are no checks of the assumptions of normality and independent samples. The two-sample *t*-test is correctly evaluated, but it yields a *p*-value that is larger than the *p*-value for the correct paired *t*-test because the strong positive correlation between the before and after measurements on these 12 subjects is ignored by the two-sample test. While the conclusion is consistent with the large *p*-value for the two-sample *t*-test, it appears to accept the null hypothesis instead of indicating that the two-sample *t*-test does not provide sufficient evidence to conclude that the training increased mean dexterity.