Homework for t-tests -- one sample, two independent samples, and correlated samples

Formulas

One sample t-test: $t = \frac{(\overline{X} - \mu)}{(s/\sqrt{n})}$

Two independent samples t-test: $t = \frac{\overline{X}_1 - \overline{X}_2}{SE_d} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

Correlated samples t-test: $t = \frac{\overline{X}_1 - \overline{X}_2}{SE_d} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} - 2(r_{12})\left(\frac{s_1}{\sqrt{n_1}}\right)\left(\frac{s_2}{\sqrt{n_2}}\right)}}$

where SE_d is the standard error of the difference between means.

For correlated samples t-test, the formula above can be tedious, so a simpler formula can also be used:

$$t = \frac{\overline{d} - \mu_d}{\sqrt{\frac{s_d^2}{N}}} = \frac{\overline{d}}{\sqrt{\frac{s_d^2}{N}}} = \frac{\overline{d}}{SE_d}$$

In this formula, $\overline{d}$ is the mean of the difference between pairs of scores

$$\overline{d} = \frac{\Sigma d}{N}$$

and SE_d is the standard error of the difference

$$SE_d = \sqrt{\frac{s_d^2}{N}}$$

where s_d^2 is the variance of the difference scores and is calculated just like a regular variance, except in this case it is the variance of the difference scores

$$s_d^2 = \frac{\sum (d - \overline{d})^2}{N - 1}$$

In short, the correlated t-test is simply the mean of the difference between two sets of scores, $\overline{d}$, divided by the standard error of that difference, SE_d, i.e., $t = \overline{d}/SE_d$. These formulas will be illustrated in the exercises below.

Instructions

For each of the problems below, provide the following:

(a) Identification of which statistical test should be used
(b) Written null hypothesis
(c) Symbolic null hypothesis
(d) Written non-directional hypothesis
(e) Symbolic non-directional hypothesis
(f) Obtained (calculated) statistic
(g) df, critical t statistics (taken from table in back of text), and decision
(h) Written results in APA style. See "Reporting statistical outcomes" on the course web page for examples.

I recommend that the first four exercises be calculated by hand so you become more familiar with these test procedures.

1. A principal is concerned that disciplinary referrals are on the rise. Last year at the principal's middle school, the average number of discipline referrals across all teachers during the first month was 5.5. This year, the total number of referrals during the first month from each teacher is recorded below. Is there any evidence that discipline referrals have changed from last year? (Calculate this one by hand.)

Teacher	Number of Discipline Referrals Made
Betty	10
LaTonya	2
Sue	8
Mary	7
Fred	6
John	2
Wilma	9
Jim	3
Saundra	8
Wendel	10
Marie	12

2. Male and Female students from a graduate class were asked their opinion about a particular textbook. The scale on which they replied ranged from 1=strongly unfavorable to 5=strongly favorable. Which group appears to have more favorable opinions regarding the textbook, males or females? (Calculate this one by hand.)

Male Opinion Scores		Female Opinion Scores	
5	5	3	1
5	4	3	2
4	3	1	1

3. Below are incoming freshmen mathematics SAT scores for both males and females. Is there any evidence of a difference between them? (Calculate this one by hand.)

Males	Females
M = 585	M = 610
s = 86.7	S = 99.8
n = 526	n = 531

4. Ladies at Weight Watchers wish to know whether participating in Weight Watchers seems to have any effect on weight change. Below are data collected from several participants before and after attending Weight Watchers for a given period of time. Does it appear that weight change results from attending Weight Watchers? (Calculate this one by hand.)

Weight Watcher Participant	Weight before	Weight after
Betty	185	163
LaTonya	136	145
Sue	115	118
Mary	195	188
Freda	145	140
Johnnie	133	136
Wilma	119	123
Jimmie	153	161
Saundra	138	151
Wendella	109	113
Marie	181	165

5. A school counselor is interested in learning whether small-group counseling helps freshmen students become more identified with school during their first year of high school. That is, does small-group counseling help students feel more attached to school; do they perceive school to be more important to them. At the beginning of the school year 15 freshmen students are randomly selected and required to attend small-group counseling sessions for six weeks. Prior to attending, an instrument designed to measure their level of academic identification with school is administered to each student (higher scores on this instrument indicate greater levels of academic identification). At the end of counseling sessions six week later, the same instrument is re-administered to each student. Is there any evidence students in small-group counseling show levels of academic identification that differ from levels taken at the beginning of the school year? Data below show levels of academic identification for students before (start) and after (end) small-group counseling.

Student	Start	End	Student	Start	End	Student	Start	End
John	2	25	Brenda	11	33	Gaylord	5	19
Jerry	6	33	Bonnie	8	25	Ginger	9	33
Jeff	5	25	Bridget	9	23	Gigi	8	35
James	10	21	Buffy	20	27	George	12	22
Jimmy	11	35	Bertha	11	29	Georj	6	19

6. One-group before-after type experimental designs such as that used in #5 above are notoriously problematic and poor. Such designs should never be used for serious decision making. It is extremely important to have comparison groups so changes within groups can be compared. Given this, counselors interested in performing a small-group intervention identical to the type used in #5 above revised that study to include two groups of students. One group received the small-group counseling, and the other did not. Measurement of school identification was taken from both groups after six weeks of school (six weeks of counseling for the treatment group). Is there any evidence that those in counseling have levels of academic identification that differ from those not receiving counseling?

Students who received small-group counseling

Student	Identification Score	Student	Identification Score	Student	Identification Score
John	25	Brenda	33	Gaylord	19
Jerry	33	Bonnie	25	Ginger	33
Jeff	25	Bridget	23	Gigi	35
James	21	Buffy	27	George	22
Jimmy	35	Bertha	29	Georj	19

Students who did not received small-group counseling

Student	Identification Score	Student	Identification Score	Student	Identification Score
Denver	26	Frank	19	Zoila	26
Doug	19	Frances	28	Zelda	21
Deidra	33	Freddie	33	Zoraida	22
Debbie	35	Florence	29	Zola	28
Dave	25	Felicia	30	Zora	31

7. In the 1990s I used a test in EDUR 7130 (Introduction to Educational Research) in which students, across many class sections, average about 89. That class was taught face-to-face. Since 2000 all sections of that class are taught on-line rather than face-to-face. Recently I began using, in my on-line classes, the same test that I used in the 1990s for face-to-face classes. Below are actual scores from one on-line class from that test. Is there any evidence that student achievement, as measured by this one test, differs from typical student achievement in face-to-face classes?

Test Scores from On-line Students

93.02	95.35	81.40
97.67	83.72	81.40
97.67	97.67	86.05
100.00	93.02	79.07
95.35	93.02	81.40
90.70	86.05	76.74
93.02	74.42	77.91
86.05	86.05	65.12
88.37	83.72	76.74

Answers

1. A principal is concerned that disciplinary referrals are on the rise. Last year at the principal's middle school, the average number of discipline referrals across all teachers during the first month was 5.5. This year, the total number of referrals during the first month from each teacher is recorded below. Is there any evidence that discipline referrals have changed from last year? (Calculate this one by hand.)

(a) Identification of which statistical test should be used

one sample t-test

(b) Written null hypothesis

The number of disciplinary referrals from the first month of school this year will equal the average from the first month of school last year (M = 5.5).

(c) Symbolic null hypothesis

H_0: $\mu = 5.5$

(d) Written non-directional hypothesis

The number of disciplinary referrals from the first month of school this year will differ from the average from the first month of school last year (M = 5.5).

(e) Symbolic non-directional hypothesis

H_0: $\mu \neq 5.5$

(f) Obtained (calculated) statistic

One sample t-test: $t = \dfrac{(\overline{X} - \mu)}{(s/\sqrt{n})}$

Mean for these data is M = 7.00
Standard deviation for these data is s = SD = 3.4059
Sample size is n = 11

Calculated t is:

$$t = \frac{(\overline{X} - \mu)}{(s/\sqrt{n})} = \frac{(7.00 - 5.5)}{(3.4059/\sqrt{11})} = \frac{1.5}{(3.4059/3.3166)} = \frac{1.5}{1.0269} = 1.46$$

(g) df, critical t statistics (taken from table in back of text), and decision

df = n – 1 = 10
critical t at $\alpha = .05$ is ± 2.228

Since 1.46 is less than 2.228 (does not fall in the rejection region), fail to reject. Note also that the p-value, if statistical software is used to perform this test, is greater than .05 ($p > .05$, $p = .175$), so fail to reject null.

(h) Written results in APA style. See "Reporting statistical outcomes" on the course web page for examples.

The number of disciplinary referrals for the first month of this year does not appear to differ statistically, at the .05 level, from the average number of referrals during the first month of last school year (M = 7.00, SD = 3.41, 95% CI –0.79 to 3.79, n = 11, $\underline{t} = 1.46$, $\underline{df} = 10$, $\underline{p} = .175$). Although this year's mean is slightly higher, this finding does not differ from what one might expect from chance and therefore these data do not support the principal's concern that number of disciplinary referrals is on the rise.

2. Male and Female students from a graduate class were asked their opinion about a particular textbook. The scale on which they replied ranged from 1=strongly unfavorable to 5=strongly favorable. Which group appears to have more favorable opinions regarding the textbook, males or females? (Calculate this one by hand.)

(a) Identification of which statistical test should be used

two independent samples t-test

(b) Written null hypothesis

There will be no difference in opinion between males and females regarding the textbook.

(c) Symbolic null hypothesis

H_0: $\mu_M = \mu_F$

(d) Written non-directional hypothesis

There will be a difference in opinion between males and females regarding the textbook.

(e) Symbolic non-directional hypothesis

H_1: $\mu_M \neq \mu_F$

(f) Obtained (calculated) statistic

Two independent samples t-test: $t = \dfrac{\overline{X}_1 - \overline{X}_2}{SE_d} = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Males
M = 4.333, s = 0.817, n = 6

Females
M = 1.833, s = 0.983 , n = 6

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{4.333_M - 1.833_F}{\sqrt{\dfrac{0.817_M^2}{6_M} + \dfrac{0.983_F^2}{6_F}}} = \frac{4.333_M - 1.833_F}{\sqrt{\dfrac{0.667_M}{6_M} + \dfrac{0.966_F}{6_F}}} = \frac{4.333_M - 1.833_F}{\sqrt{0.111 + 0.161}} =$$

$$\frac{4.333_M - 1.833_F}{0.522} = \frac{2.50}{0.522} = 4.79$$

(g) df, critical t statistics (taken from table in back of text), and decision

df = n – 2 = 10
critical t at α = .05 is ± 2.228

Since 4.79 is greater than 2.228 (falls in the rejection region), reject the null. Note also that the p-value, if statistical software is used to perform this test, is less than .05 ($p < .05$, $p = .001$), so reject null.

(h) Written results in APA style. See "Reporting statistical outcomes" on the course web page for examples.

Table 1: Results of t-test and Descriptive Statistics of Text Opinion by Sex

Outcome	Group						95% CI for Mean Difference	t	df
	Male			Female					
	M	SD	n	M	SD	n			
Text Opinion	4.33	0.82	6	1.83	0.98	6	1.34, 3.66	4.79*	10

* $p < .05$.

Results of this study show that there is a statistically significant difference in opinions about the text between males and females. As the means in Table 1 demonstrate, males held more favorable opinions about the text than did females.

3. Below are incoming freshmen mathematics SAT scores for both males and females. Is there any evidence of a difference between them? (Calculate this one by hand.)

(a) Identification of which statistical test should be used

two independent samples t-test

(b) Written null hypothesis

There will be no difference in mathematics SAT scores between males and females.

(c) Symbolic null hypothesis

H_0: $\mu_M = \mu_F$

(d) Written non-directional hypothesis

There will be a difference in mathematics SAT scores between males and females.

(e) Symbolic non-directional hypothesis

H_1: $\mu_M \neq \mu_F$

(f) Obtained (calculated) statistic

Two independent samples t-test: $t = \frac{\overline{X}_1 - \overline{X}_2}{SE_d} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{585_M - 610_F}{\sqrt{\frac{86.7_M^2}{526_M} + \frac{99.8_F^2}{531_F}}} = \frac{585_M - 610_F}{\sqrt{\frac{7516.89_M}{526_M} + \frac{9960.04_F}{531_F}}} = \frac{585_M - 610_F}{\sqrt{14.29 + 18.757}}$$

$$= \frac{585_M - 610_F}{5.749} = -4.35$$

(g) df, critical t statistics (taken from table in back of text), and decision

df = n – 2 = 1055
critical t at $\alpha = .05$ is ± 1.96

Since -4.35 is less than a negative -1.96 (falls in the rejection region), reject the null. Note also that the p-value, if statistical software is used to perform this test, is less than .05 ($p < .05$, $p = .001$), so reject null.

(h) Written results in APA style. See "Reporting statistical outcomes" on the course web page for examples.

Table 1: Results of t-test and Descriptive Statistics for SAT by Sex

Outcome	Group						95% CI for Mean Difference	t	df
	Male			Female					
	M	SD	n	M	SD	n			
SAT Math. scores	585	86.7	526	610	99.8	531	-36.29, -13.71	-4.35*	1055

* $p < .05$.

Results of this study show that there is a statistically significant difference in SAT scores between male and female students. Females, on average, had the higher SAT mathematics scores.

4. Ladies at Weight Watchers wish to know whether participating in Weight Watchers seems to have any effect on weight change. Below are data collected from several participants before and after attending Weight Watchers for a given period of time. Does it appear that weight change results from attending Weight Watchers? (Calculate this one by hand.)

Weight Watcher Participant	Weight before	Weight after
Betty	185	163
LaTonya	136	145
Sue	115	118
Mary	195	188
Freda	145	140
Johnnie	133	136
Wilma	119	123
Jimmie	153	161
Saundra	138	151
Wendella	109	113
Marie	181	165

(a) Identification of which statistical test should be used

Correlated samples t-test

(b) Written null hypothesis

There will be no difference in weight measurement before and after joining weight watchers.

(c) Symbolic null hypothesis

H_0: $\mu_b = \mu_a$

(d) Written non-directional hypothesis

There will be a difference in weight measurement before and after joining weight watchers.

(e) Symbolic non-directional hypothesis

H_1: $\mu_b \neq \mu_a$

(f) Obtained (calculated) statistic

$$t = \frac{\overline{d} - \mu_d}{\sqrt{\frac{s_d^2}{N}}} = \frac{\overline{d}}{\sqrt{\frac{s_d^2}{N}}} = \frac{\overline{d}}{SE_d}$$

Before	After	Difference, d	$\bar{d}$	$d-\bar{d}$	$(d-\bar{d})^2$
185	163	22	0.545	21.455	460.317
136	145	-9	0.545	-9.545	91.10703
115	118	-3	0.545	-3.545	12.56703
195	188	7	0.545	6.455	41.66703
145	140	5	0.545	4.455	19.84703
133	136	-3	0.545	-3.545	12.56703
119	123	-4	0.545	-4.545	20.65703
153	161	-8	0.545	-8.545	73.01703
138	151	-13	0.545	-13.545	183.467
109	113	-4	0.545	-4.545	20.65703
181	165	16	0.545	15.455	238.86

M_{before} = 146.273; SD_{before} = 29.306 ; M_{after} = 145.727; SD_{after} = 22.808

$\bar{d} = 0.545$; SS_d = 1174.737; s_d^2 = SS_d/n-1 = 1174.737/10 = 117.4737

$$t = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{N}}} = \frac{0.545}{\sqrt{\frac{117.4737}{11}}} = \frac{0.545}{\sqrt{10.679}} = \frac{0.545}{3.268} = 0.167$$

(g) df, critical t statistics (taken from table in back of text), and decision

df = n – 1 = 10
critical t, at .05, is ± 2.228

Since 0.167 is not larger than 2.228, fail to reject

95%CI = $\bar{d}$ ± (t_{crit})(SE_d)
95%CI = 0.545 ± (2.228)(3.268)
95%CI = 0.545 ± 7.281 = -6.74, 7.83

(h) Written results in APA style. See “Reporting statistical outcomes” on the course web page for examples.

Table 1: Descriptive Statistics and t-test Results for Weight

Outcome	Time							
	Before Weight Watchers		After Weight Watchers			95% CI for Mean Difference	t	df
	M	SD	M	SD	n			
Weight	146.27	29.31	145.73	22.81	11	-6.74, 7.83	0.17	10

* $p < .05$

Note --- correlation not calculated with direct difference approach, so remove from table.

There is not a statistically significant difference in women's weight before or after joining Weight Watchers. Results show that weight loss does not occur after joining Weight Watchers, at least for the 11 women sampled.

5. A school counselor is interested in learning whether small-group counseling helps freshmen students become more identified with school during their first year of high school. That is, does small-group counseling help students feel more attached to school; do they perceive school to be more important to them. At the beginning of the school year 15 freshmen students are randomly selected and required to attend small-group counseling sessions for six weeks. Prior to attending, an instrument designed to measure their level of academic identification with school is administered to each student (higher scores on this instrument indicate greater levels of academic identification). At the end of counseling sessions six week later, the same instrument is re-administered to each student. Is there any evidence students in small-group counseling show levels of academic identification that differ from levels taken at the beginning of the school year? Data below show levels of academic identification for students before (start) and after (end) small-group counseling.

Student	Start	End	Student	Start	End	Student	Start	End
John	2	25	Brenda	11	33	Gaylord	5	19
Jerry	6	33	Bonnie	8	25	Ginger	9	33
Jeff	5	25	Bridget	9	23	Gigi	8	35
James	10	21	Buffy	20	27	George	12	22
Jimmy	11	35	Bertha	11	29	Georj	6	19

(a) Identification of which statistical test should be used

correlated samples t-test

(b) Written null hypothesis

There will be no difference in level of academic identification before and after small group counseling.

(c) Symbolic null hypothesis

H_0: $\mu_b = \mu_a$

(d) Written non-directional hypothesis

There will be a difference in level of academic identification before and after small group counseling.

(e) Symbolic non-directional hypothesis

H_1: $\mu_b \neq \mu_a$

(f) Obtained (calculated) statistic

t = -10.96

(g) df, critical t statistics (taken from table in back of text), and decision

df = 15-1 = 14
$t_{crit} = \pm 2.145$
Reject null since -10.96 is less than -2.145

(h) Written results in APA style. See "Reporting statistical outcomes" on the course web page for examples.

Table 1: Descriptive Statistics and t-test Results for Academic Identification

Outcome	Time								
	Before Counseling		After Counseling			95% CI for Mean Difference			
	M	SD	M	SD	n		r	t	df
Academic Identification	8.87	4.17	26.93	5.71	15	-21.60, -14.53	0.19	-10.96*	14

* $p < .05$

There is a statistically significant difference in level of academic identification displayed by students before and after receiving small group counseling. After small group counseling, students displayed a much higher level of academic identification with school than before small group counseling.

6. One-group before-after type experimental designs such as that used in #5 above are notoriously problematic and poor. Such designs should never be used for serious decision making. It is extremely important to have comparison groups so changes within groups can be compared. Given this, counselors interested in performing a small-group intervention identical to the type used in #5 above revised that study to include two groups of students. One group received the small-group counseling, and the other did not. Measurement of school identification was taken from both groups after six weeks of school (six weeks of counseling for the treatment group). Is there any evidence that those in counseling have levels of academic identification that differ from those not receiving counseling?

Students who received small-group counseling

Student	Identification Score	Student	Identification Score	Student	Identification Score
John	25	Brenda	33	Gaylord	19
Jerry	33	Bonnie	25	Ginger	33
Jeff	25	Bridget	23	Gigi	35
James	21	Buffy	27	George	22
Jimmy	35	Bertha	29	Georj	19

Students who did not received small-group counseling

Student	Identification Score	Student	Identification Score	Student	Identification Score
Denver	26	Frank	19	Zoila	26
Doug	19	Frances	28	Zelda	21
Deidra	33	Freddie	33	Zoraida	22
Debbie	35	Florence	29	Zola	28
Dave	25	Felicia	30	Zora	31

(a) Identification of which statistical test should be used

two independent samples t-test

(b) Written null hypothesis

There will be no difference in level of school identification between those students who received small group counseling and those students who did not receive small group counseling.

(c) Symbolic null hypothesis

H_0: $\mu_{yes} = \mu_{no}$

(d) Written non-directional hypothesis

There will be a difference in level of school identification between those students who received small group counseling and those students who did not receive small group counseling.

(e) Symbolic non-directional hypothesis

H_1: $\mu_{yes} \neq \mu_{no}$

(f) Obtained (calculated) statistic

$t = -0.034$

(g) df, critical t statistics (taken from table in back of text), and decision

df = n – 2 = 30 – 2 = 28
t_{crit} = ±2.048
Fail to reject null since -0.034 is not less than -2.048

(h) Written results in APA style. See "Reporting statistical outcomes" on the course web page for examples.

Table 1: Results of t-test and Descriptive Statistics for School Identification by Counseling

Outcome	Small Group Counseling						95% CI for Mean Difference	t	df
	Yes			No					
	M	SD	n	M	SD	n			
School Identification	26.93	5.71	15	27.00	5.09	15	-4.11, 3.98	-0.03	28

* $p < .05$.

Results of this study show that there is no statistically significant difference in level of school identification between those students who received small group counseling and those who did not. Contrary to the previous study's findings, results of this study show that small group counseling does not appear to show differences in school identification for students.

7. In the 1990s I used a test in EDUR 7130 (Introduction to Educational Research) in which students, across many class sections, average about 89. That class was taught face-to-face. Since 2000 all sections of that class are taught on-line rather than face-to-face. Recently I began using, in my on-line classes, the same test that I used in the 1990s for face-to-face classes. Below are actual scores from one on-line class from that test. Is there any evidence that student achievement, as measured by this one test, differs from typical student achievement in face-to-face classes?

Test Scores from On-line Students

93.02	95.35	81.40
97.67	83.72	81.40
97.67	97.67	86.05
100.00	93.02	79.07
95.35	93.02	81.40
90.70	86.05	76.74
93.02	74.42	77.91
86.05	86.05	65.12
88.37	83.72	76.74

(a) Identification of which statistical test should be used

One sample t-test

(b) Written null hypothesis

Average achievement on EDUR 7130 test scores for on-line students will not differ from the average of test scores for face-to-face classes (M = 89).

(c) Symbolic null hypothesis

H_0: $\mu = 89$

(d) Written non-directional hypothesis

Average achievement on EDUR 7130 test scores for on-line students will differ from the average of test scores for face-to-face classes ($\mu = 89$).

(e) Symbolic non-directional hypothesis

H_0: $\mu \neq 89$

(f) Obtained (calculated) statistic

$t = -1.40$

(g) df, critical t statistics (taken from table in back of text), and decision

$df = n - 1 = 27 - 1 = 26$
critical t at $\alpha = .05$ is ± 2.056

Since -1.40 is not less than -2.056 (does not fall in the rejection region), so fail to reject. Note also that the p-value, if statistical software is used to perform this test, is greater than .05 ($p > .05$, $p = .175$), so fail to reject null.

(h) Written results in APA style. See "Reporting statistical outcomes" on the course web page for examples.

The average test scores for students on-line ($M = 86.69$, $SD = 8.59$, $n = 27$) does not differ, statistically, from the average of face-to-face students ($M = 89$; $t = -1.40$, $df = 26$, $p = .18$, 95%CI = -5.71 to 1.09). These findings suggest that test performance seems to be similar between the two instructional formats.