t Test Example

This second example of the independent-groups (independent-samples, between-subjects) t test compares the number of new terms recalled after simply reading (*reading only*) information from a course text vs. reading plus retrieval (*retrieval practice*),[1] but, this time, notice that the variability among terms recalled is much lower. In the previous example given in class, the two means were the same, 6 vs. 9, for reading only and retrieval practice, respectively. The variances in the first example were 10 for reading only and 4 for retrieval practice (s = 3.16 and s = 2), and the calculated t value was not significant, $t(8) = -1.80$, ns.[2]

Reading vs. Retrieval Example

Reading Only	Retrieval Practice
4	9
4	8
6	10
8	8
8	10

$$\bar{Y}_1 = 6 \qquad \bar{Y}_2 = 9$$
$$s^2 = 4 \qquad s^2 = 1$$

Compute t

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{s_{\bar{Y}_1 - \bar{Y}_2}} = \frac{6-9}{1} = \frac{3}{1} = -3$$

Compute the pooled variance

$$s^2_{pooled} = \left[\frac{n_1 - 1}{n_1 + n_2 - 2}\right] s_1^2 + \left[\frac{n_2 - 1}{n_1 + n_2 - 2}\right] s_2^2$$
$$= \left[\frac{5-1}{5+5-2}\right]4 + \left[\frac{5-1}{5+5-2}\right]1$$
$$= \left[\frac{4}{8}\right]4 + \left[\frac{4}{8}\right]1$$
$$= 2 + .5 = 2.5$$

[Formulas are on pp. 129-130 of Myers, Well, & Lorch (2010). A simpler formula for s^2_{pooled}, Eq 6.8 on p. 130, can be used if group sample sizes are equal.]

$$s_{pooled} = \sqrt{s^2_{pooled}} = \sqrt{2.5} = 1.58$$

Compute the 95% confidence interval

$$CI = \left(\bar{Y}_1 - \bar{Y}_2\right) \pm \left(t_{df,\alpha/2} s_{\bar{Y}_1 - \bar{Y}_2}\right)$$
$$= -3 \pm \left[(-2.306)(1)\right]$$
$$= -3 \pm -2.306$$
$$= -5.306, -.649$$

$t_{df,\alpha/2}$ is the critical value of t from the *t*-table with the appropriate df for a two-tailed test indicated by $\alpha/2$.

Note that it is traditional to report positive values for the *t* and the accompanying confidence interval, because which direction you subtract is arbitrary. So, most authors would report a *t*-value of 3.00 here and CI of .65,5.31.

Compute the standard error of the differences

$$s_{\bar{Y}_1 - \bar{Y}_2} = s_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$
$$= 1.58\sqrt{\frac{1}{5} + \frac{1}{5}}$$
$$= 1.58\sqrt{.4}$$
$$= (1.58)(.63)$$
$$= 1.00$$

Compute Cohen's d standardized effect size measure

$$\hat{d} = \left|\frac{\bar{Y}_1 - \bar{Y}_2}{s_{pooled}}\right|$$
$$= \left|\frac{6-9}{1.58}\right|$$
$$= \left|\frac{3}{1.58}\right| = 1.89$$

[1] These are hypothetical results, but they are based in part on actual research about studying practices, such as Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*(6018), 772-775.

[2] The use of *ns* for "nonsignificant" is no longer indicated by the current version of the APA style manual. The exact p-value is supposed to be given (unless p < .001) whether the statistic is significant or not. In the case of hand computation, there is not a choice.

The degrees of freedom for a between-groups t test are $df = n_1 + n_2 - 2 = 10 - 2 = 8$. The two-tailed critical value of t (obtained from the Table C.3 in the back of the text) if $\alpha = .05$ is equal to 2.306. If the negative sign for the calculated *t*-value is ignored, the calculated value exceeds the critical value, so the difference is significant at the $\alpha = .05$ level.

Computer Examples

Below SPSS and R are used to compare the two groups (variable name is `type`, for learning type—reading only or retrieval practice) on the number of new terms recalled (`recall`), the dependent variable.

Syntax

```
t-test groups=type(0,1)
  /vars=recall.

*side-by-side (or above and below) plots (output suppressed to save space).
examine variables=recall by type
  /plot=histogram
  /statistics=none
  /nototal.

examine variables=recall by type
  /plot=boxplot
  /statistics=none
  /nototal.
```

SPSS Menus Steps

Analyze→ Compare Means→ Independent Samples t-test. Move over the test variable and grouping variable to the appropriate boxes. "Test Variable" is the dependent variable and "Grouping Variable" is the variable indicating the two groups (e.g., treatment and control groups). Click "Define groups": Place the two values of the grouping variable into the boxes for Group 1 and Group 2 (values are 0 and 1 in my data set). Click "Continue". "Ok."

SPSS Output

Group Statistics

	type learning strategy	N	Mean	Std. Deviation	Std. Error Mean
recall number of new terms recalled	0 reading only	5	6.00	2.000	.894
	1 retrieval practice	5	9.00	1.000	.447

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
recall number of new terms recalled	Equal variances assumed	3.200	.111	-3.000	8	.009	.017	-3.000	1.000	-5.306	-.694
	Equal variances not assumed			-3.000	5.882	.012	.025	-3.000	1.000	-5.459	-.541

Independent Samples Effect Sizes

				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
recall number of new terms recalled	Cohen's d	1.581	-1.897	-3.400	-.320
	Hedges' correction	1.752	-1.713	-3.069	-.289
	Glass's delta	1.000	-3.000	-5.315	-.618

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control (i.e., the second) group.

The raw "point estimate" of Cohen's *d* is the most commonly reported effect size measure. See Section 6.7.1 in the Myers et al. (2010) text for information on the Hedges and Glass corrections.

R

```
#clears the previous results from the console
> cat("\014")
#clears active data frame from previous analyses
> rm(d)>
>
> library(haven)
> d = read_sav("c:/jason/spsswin/uvclass/ttest.sav")
```

```
> library(lessR)
> #obtain t-test
> ttest(recall ~ type)

> #lessR generates side-by-side density plots by default
>
> #barchart by groups
> #Plot(recall, by1=type, vbs.plot="barchart")
>
> #Histogram by groups (code given in handout)
> Histogram(recall, by1=type)

> #boxplot by groups (code given in handout)
> #Plot(recall, by1=type, vbs.plot="boxplot")

------ Describe ------

recall for type 1:  n.miss = 0,  n = 5,  mean = 9.000,  sd = 1.000
recall for type 0:  n.miss = 0,  n = 5,  mean = 6.000,  sd = 2.000

Mean Difference of recall:  3.000

Weighted Average Standard Deviation:   1.581


------ Assumptions ------

Note: These hypothesis tests can perform poorly, and the
      t-test is typically robust to violations of assumptions.
      Use as heuristic guides instead of interpreting literally.

Null hypothesis, for each group, is a normal distribution of recall.
Group 1  Shapiro-Wilk normality test:  W = 0.821,  p-value = 0.119
Group 0  Shapiro-Wilk normality test:  W = 0.821,  p-value = 0.119

Null hypothesis is equal variances of recall, homogeneous.
Variance Ratio test:  F = 4.000/1.000 = 4.000,  df = 4;4,  p-value = 0.208
Levene's test, Brown-Forsythe:  t = -1.789,  df = 8,  p-value = 0.111


------ Infer ------

--- Assume equal population variances of recall for each type

t-cutoff for 95% range of variation: tcut =  2.306
Standard Error of Mean Difference: SE =  1.000

Hypothesis Test of 0 Mean Diff:  t-value = 3.000,  df = 8,  p-value = 0.017

Margin of Error for 95% Confidence Level:  2.306
95% Confidence Interval for Mean Difference:  0.694 to 5.306


--- Do not assume equal population variances of recall for each type

t-cutoff: tcut =  2.459
Standard Error of Mean Difference: SE =  1.000

Hypothesis Test of 0 Mean Diff:  t = 3.000,  df = 5.882, p-value = 0.025

Margin of Error for 95% Confidence Level:  2.459
95% Confidence Interval for Mean Difference:  0.541 to 5.459
```

Write-up Example

A t test was used to compare reading only and retrieval practice on the number of new terms recalled. On average, retrieval practice recalled more words (*M* = 9.00, *SD* = 2) than reading only recalled (*M* = 6.00, *SD* = 1). This difference was statistically significant, $t(8) = 3.00$, $p = .017$, 95% CI[.69,5.31], indicating that the higher average recall for students using retrieval practice than what would have been expected due to chance. Moreover, the difference between the two groups was large, $d = 1.89$.[3]

[3] Citation for Cohen's d is considered unnecessary, as it is regarded as a common statistic. For less common statistics or newer statistics, a citation and sometimes a phrase to describe the statistic is needed. Confidence limits are usually recommended—just remember that that they are constructed around the mean difference, not the *t*-value. According to the APA Manual 7th Edition, in the first use, use 95% CI[,], but on subsequent uses drop 95%.