Paired samples t-test

Elias Zintzaras, M.Sc., Ph.D.

Professor in Biomathematics-Biometry
Department of Biomathematics
School of Medicine
University of Thessaly

Institute for Clinical Research and Health Policy Studies
Tufts University School of Medicine
Boston, MA, USA

Theodoros Mprotsis, MSc, PhD
Teacher & Research Fellow
(http://biomath.med.uth.gr)
University of Thessaly
Email: tmprotsis@uth.gr

Hemoglobin change after administration of EPO

Suppose we want to compare the hemoglobin change after administration of erythropoietin (EPO) in 9 patients. The data (g/L) are as follows:

Subject	Before	After
1	135	160
2	126	157
3	165	153
4	122	165
5	162	155
6	122	160
7	116	165
8	136	170
9	168	157

At the end of the study, we aim to determine whether there was a significant change in hemoglobin levels by comparing the values before and after the administration of erythropoietin (EPO)

Note that the data for each patient are appeared as a pair (baseline, 3 months)

Before the administration of EPO

Hypothesis;

After the administration of EPO

Mean of the matched differences is zero.

$$\mu_d = 0$$

Subject 1	hgb_{before}	hgb_{after}
Subject 2	hgb_{before}	hgb_{after}
Subject 3	hgb_{before}	hgb_{after}
	⋮	⋮

Independence;

These two measures are NOT independent of each other.

They are MATCHED.

Null hypothesis

The **null hypothesis** is that there is no change in hemoglobin levels before and after erythropeitin administration. The mean of the differences is 0.

$$H_0: \mu_d = 0 \qquad H_a: \mu_d \neq 0$$

Significance level: $a = 0.05$

Since this is a matched pairs test with $\boldsymbol{n = 9}$, we will use the t-distribution with 8 degrees of freedom (n-1)

Decision rule:

Since $a = 0.05$ and we are using the t-distribution, the H_0 will be rejected if the test statistic is > 2.31 (**5% point of the t-distribution with n-1=8 degrees of freedom**) or < -2.31

Descriptive Statistics

On **average**, the change in **hemoglobin** is **21.11** units, which is greater than **zero**. This suggests that hemoglobin (hgb) is **increasing** due to erythropoietin (EPO).

However, the variability is quite large (**CV=115%**), the sample size is small (**n=9**), and thus, the error of the mean is considered quite large (**SE=8.11**)

Subject	hgb_{before}	hgb_{after}	$hgb_{before} - hgb_{after}$
1	135	160	-25
2	126	157	-31
3	165	153	12
4	122	165	-43
5	162	155	7
6	122	160	-38
7	116	165	-49
8	136	170	-34
9	168	157	11
		N	9
		Average	-21.11
		Standard Deviation (SD)	24.34
		Coefficient of variation (CV)	-115
		Standard error ($SE = \frac{s}{\sqrt{n}}$)	8.11

Examine the effectiveness of the drug

In order to examine the effectiveness of the drug, we have to answer the following question:

How **confident** we are that the **average improvement** (21.11) is **significant**, i.e. different from zero?

Alternatively, how **confident** we are that the improvement is not **due to chance** (i.e. not random)?

Subject	hgb_{before}	hgb_{after}	$hgb_{before} - hgb_{after}$
1	135	160	-25
2	126	157	-31
3	165	153	12
4	122	165	-43
5	162	155	7
6	122	160	-38
7	116	165	-49
8	136	170	-34
9	168	157	11
		N	9
		Average	-21.11
		Standard Deviation (SD)	24.34
		Coefficient of variation (CV)	-115
		Standard error ($SE = \frac{s}{\sqrt{n}}$)	8.11

Examine the effectiveness of the drug

In order to test whether the average improvement (21.11) is significant, we must consider the **average** in conjunction with the **variability** (standard deviation) and the **size** of the trial, i.e. the **standard error**

Subject	hgb_{before}	hgb_{after}	$hgb_{before} - hgb_{after}$
1	135	160	$\mathbf{-25}$
2	126	157	$\mathbf{-31}$
3	165	153	$\mathbf{12}$
4	122	165	$\mathbf{-43}$
5	162	155	$\mathbf{7}$
6	122	160	$\mathbf{-38}$
7	116	165	$\mathbf{-49}$
8	136	170	$\mathbf{-34}$
9	168	157	$\mathbf{11}$
		N	$\mathbf{9}$
		Average	$\mathbf{-21.11}$
		Standard Deviation (SD)	$\mathbf{24.34}$
		Coefficient of variation (CV)	$\mathbf{-115}$
		Standard error ($SE = \frac{s}{\sqrt{n}}$)	$\mathbf{8.11}$

Statistical Testing of the mean difference

Then, we could test whether the average is significant using the **t-test**

$$t = \frac{average\ difference}{SE}$$

$$t = \frac{\bar{d}}{SE}$$

$$t = \frac{21.11}{8.11}$$

$$t = 2.60$$

If we take the difference (after - before) the sign is ignored.

Subject	hgb_{before}	hgb_{after}	$hgb_{before} - hgb_{after}$
1	135	160	-25
2	126	157	-31
3	165	153	12
4	122	165	-43
5	162	155	7
6	122	160	-38
7	116	165	-49
8	136	170	-34
9	168	157	11
		N	9
		Average	-21.11
		Standard Deviation (SD)	24.34
		Coefficient of variation (CV)	-115
		Standard error ($SE = \frac{s}{\sqrt{n}}$)	8.11

Significance of the mean difference

$$t = \frac{average\ difference}{SE} \qquad t = \frac{\bar{d}}{SE}$$

$$t = \frac{21.11}{8.11} \qquad t = 2.60$$

Now, we have to answer the following question:

How confident we are that t=2.60 is significant (i.e. different from zero) or how confident we are that t=2.60 is not due to chance (i.e. not random)?

If t=2.60 is significant then, we conclude that the average improvement is significant (i.e. different from zero)

Subject	hgb_{before}	hgb_{after}	$hgb_{before} - hgb_{after}$
1	135	160	-25
2	126	157	-31
3	165	153	12
4	122	165	-43
5	162	155	7
6	122	160	-38
7	116	165	-49
8	136	170	-34
9	168	157	11
		N	**9**
		Average	**-21.11**
		Standard Deviation (SD)	**24.34**
		Coefficient of variation (CV)	**-115**
		Standard error ($SE = \frac{s}{\sqrt{n}}$)	**8.11**

Significance of the mean difference

In the **t-distribution** table on the right, we find the critical value $\boldsymbol{t}$, for the significance level $a = 0.05$ and $n - 1 = 9 - 1 = 8$ degrees of freedom

The value is $\mathbf{2.31}$

The value $\mathbf{2.31}$ serves as the **threshold** to determine whether the t-test result ($t = 2.60$) is statistically significant for a study with 8 degrees of freedom (9 subjects)

	Percentage points of the t distribution		
	p-value		
df (=n-1)	0.05	0.01	0.001
1	12.71	63.66	636.62
2	4.3	9.92	31.6
3	3.18	5.84	12.92
4	2.78	4.6	8.61
5	2.57	4.03	6.87
6	2.45	3.71	5.96
7	2.36	3.5	5.41
8	2.31	3.36	5.04
9	2.26	3.25	4.78
10	2.23	3.17	4.59
20	2.09	2.85	3.85
30	2.04	2.75	3.65
40	2.02	2.7	3.55
120	1.98	2.62	3.37
∞	1.96	2.58	3.29

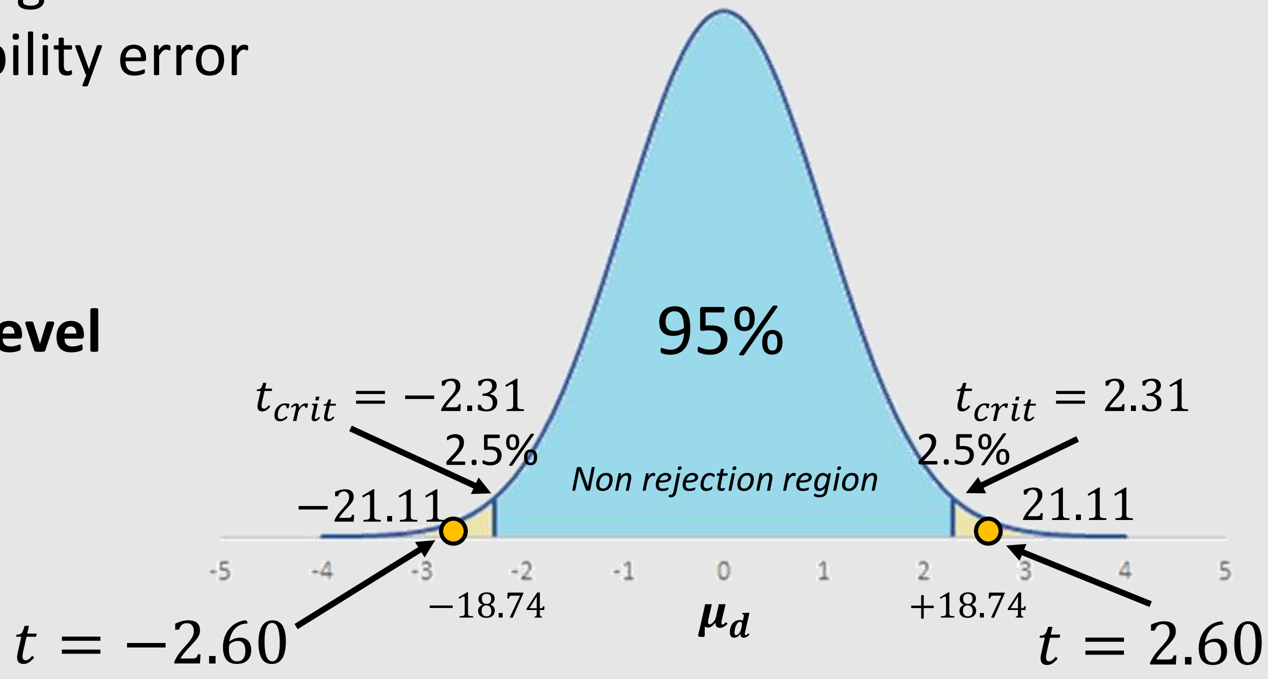
Result

The value of the t-test, which is $\mathbf{2.60}$, is larger than

$\mathbf{2.31}$ (or less than $-\mathbf{2.31}$)

Thus, the **mean difference** (21.11) in hemoglobin is **statistically significant**, with a small probability error ($\boldsymbol{P < 0.05}$)

The **p-value** is also called **significant level**



95% Confidence Interval (CI)

The **95% confidence interval (CI)** for the mean difference between the two groups (before and after) for paired observations provides the range in which the **true mean difference** is expected to lie, with **95% confidence**

$$(\text{mean difference} - \mathrm{t} \cdot \mathrm{SE}, \text{mean difference} + \mathrm{t} \cdot \mathrm{SE})$$

where **t** is the $\mathbf{5\%}$ **point of the t-distribution** with $n - 1 = 9 - 1 = 8$ degrees of freedom, which is $\mathbf{2.31}$

95 % Confidence Interval (CI)

(mean difference - t · SE, mean difference + t · SE)

Therefore, the 95% confidence interval (CI) for the mean difference is:

$$(21.11 - 2.31 \cdot 8.11, 21.11 + 2.31 \cdot 8.11)$$

or

$$(2.40, 39.8)$$

Conclusion

According to our matched sample data, the **95% confidence interval estimate** of the difference between the matched paired means is

$$2.40 - 39.8 \text{ g/L}$$

Thus, with 95% confidence, the true mean value of the differences lies within this interval.

Does this interval contain the hypothetical mean difference of zero?
No. Zero lies outside this confidence interval, indicating that the true mean of the differences is significantly different from zero.

CONCLUSION:
We reject the null hypothesis (H_0) that the mean of the before/after hemoglobin differences is **zero**. The test statistic exceeds the critical value t_{crit} for 8 degrees of freedom, and the sample mean of the differences $(\bar{d})$ is significantly different from zero. Therefore, the drug (EPO) is effective, and the change in hemoglobin ranges from 2.40 to 39.8 g/L