Paired T test

Example (experimental study)

Drunk driving is one of the main causes of car accidents. Interviews with drunk drivers who were involved in accidents and survived revealed that one of the main problems is that drivers do not realize that they are impaired, thinking “I only had 1-2 drinks … I am OK to drive.”

A sample of 20 drivers was chosen, and their reaction times in an obstacle course were measured before and after drinking two beers. The purpose of this study was to check whether drivers are impaired after drinking two beers.

y_B: reaction time

y_A: reaction time

How would you evaluate whether the reaction time was modifed after the two beers?

y_B Before	y_A After
6.25	6.85
2.96	4.78
4.95	5.57
3.94	4.01
4.85	5.91
4.81	5.34
6.6	6.09
5.33	5.84
5.19	4.19
4.88	5.75
5.75	6.25
5.26	7.23
3.16	4.55
6.65	6.42
5.49	5.25
4.05	5.59
4.42	3.96
4.99	5.93
5.01	6.03
4.69	3.72

Dependent vs independent samples

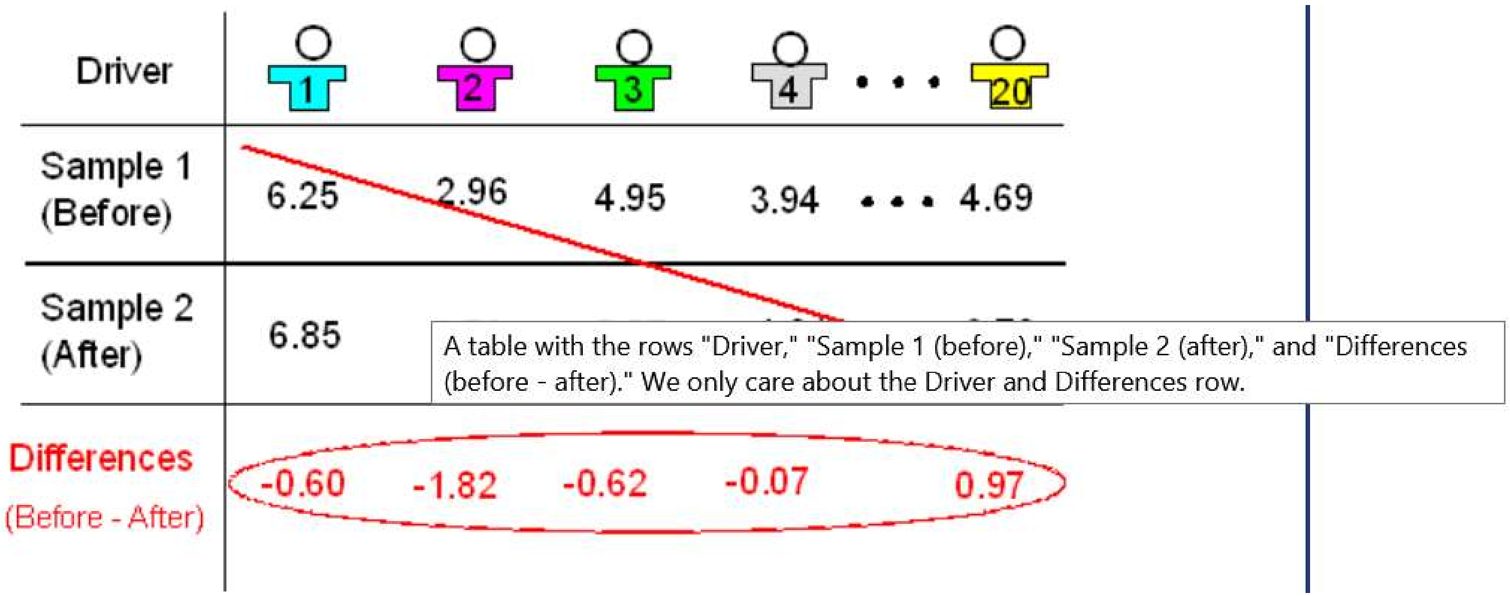
- Two samples are **independent** if the sample values from one population are not related to or somehow naturally paired or matched with the sample values from the other population.
- Two samples are **dependent** (or consist of matched pairs) if the sample values are somehow matched, where the matching is based on some inherent relationship. (That is, each pair of sample values consists of two measurements from the same subject—such as before>after data—or each pair of sample values consists of matched pairs—such as husband>wife data—where the matching is based on some meaningful relationship. Caution: "Dependence" does not require a direct cause>effect relationship.)

Good Experimental Design

- Suppose we want to test the effectiveness of a drug designed to lower blood pressure. It would be better to use before>after measurements from a single group of subjects treated with the drug than to use measurements from one group of subjects who were not treated with the drug and a separate group who were treated. **The advantage of using matched pairs (before>after measurements) is that we reduce extraneous variation, which could occur with the two different independent samples.** This strategy for designing an experiment can be generalized by the following design principle:
- **When designing an experiment or planning an observational study, using dependent samples with matched pairs is generally better than using two independent samples.**

Did the beers affect the reaction time Y?

Considering the difference **within subject** allows to account for the variability in the outcome not due to the exposure.



Thus, instead of considering μ_B and μ_A separately, we can focus on the average of the within subject differences $\bar{d} = -0.5015$ which is an estimate of the (unknown) $\delta_{before-after}$ difference in population before and after the beer assumption

Did the beers affect the reaction time Y?

$H_0: \delta = 0$

$H_1: \delta \neq 0$

$\bar{d} = -0.5015$

$s_d = 0.8686$

$n = 20$

$$t = \frac{-0.5015}{0.8686/\sqrt{20}} = -2.5821$$

α=0.05 Critical values: $t_{df=20-1,0.025} = \pm 2.093$

Reject the null hypothesis: two beers affect the reaction time of the drivers

P-value=0.0183 (by the table 0.01<p-value<0.02)

Did the beers affect the reaction time Y?

$H_0: \delta = 0$

$H_1: \delta \neq 0$

$\bar{d} = -0.5015$

$s_d = 0.8686$

$n = 20$

$$95\%CI: -0.5015 \pm 2.093 * \frac{0.8686}{\sqrt{20}}$$

$$95\%CI: [-0.095; -0.908]$$

We are 95% confident that after drinking two beers, the true mean increase in total reaction time of drivers is between 0.1 and 0.9 of a second.

If the same data were from two independent samples?

40 drivers were recruited and randomly divided into two groups in which reaction times in an obstacle course after drinking were measured. The first group drank two glasses of water and the second group drank two beers.

The purpose of this study was to test whether the attention state of drivers was altered after drinking two beers.

y_1:

y_2:

y_1 No beer	y_2 beer
6.25	6.85
2.96	4.78
4.95	5.57
3.94	4.01
4.85	5.91
4.81	5.34
6.6	6.09
5.33	5.84
5.19	4.19
4.88	5.75
5.75	6.25
5.26	7.23
3.16	4.55
6.65	6.42
5.49	5.25
4.05	5.59
4.42	3.96
4.99	5.93
5.01	6.03
4.69	3.72

If the same data were from two independent samples?

$H_0: \delta = 0$

$H_1: \delta \neq 0$

$\bar{x}_1 = 4.9615$ $\quad \bar{x}_2 = 5.4630$

$s_1 = 0.9709$ $\quad s_2 = 0.9829$

$n_1 = 20$ $\quad n_2 = 20$

$$s = \sqrt{\frac{19*0.9709^2+19*0.9829^2}{38}} = 0.9769$$

$$t = \frac{-0.5015}{0.9769 * \sqrt{\frac{1}{20}+\frac{1}{20}}} = -1.62$$

α=0.05 Valore critico: $t_{df=40-2,0.025} = \pm 2.021$

I do not reject the null hypothesis: I do not have sufficient evidence that two beers affect the reaction time of drivers P-value=0.113 (t-student t table 0.1<p-value<0.2)

Summary on hypothesis test

	OUTCOME	
	CONTINUOUS (e.g. Z-score BMI)	CATEGORICAL- BINARY (e.g. disease yes/no)
Compare a sample with a reference value	One-sample t-test (if Normal or n>30)	One-sample test for proportions (if np and n(1-p) ≥5)
Compare 2 independent groups (ex. std/trt)	Unpaired t-test (if Normal or n>30)	Test for proportions (if np and n(1-p)≥5)
Compare 2 dependent groups (ex.pre/post)	Paired t-test (if difference is Normal or n>30)	McNeamer test

ATTENTION! We have only seen the basic tests!

These tests are not suitable to compare more than two groups or if requirements are not satisfied

Exercise

The soporific effect of a new drug, F2, has been tested with respect to the same effect of an already known drug, F1, on a group of 8 volunteers. First F1 was given and the next night F2.

The results, assessed in terms of additional number of hours of sleep with respect to an average value per subject, already known, are in the table.

Test if the drugs have the same effect with an appropriate test ($\alpha = 0.05$) and build a confidence interval for the difference between the two effects

ID	F_1	F_2	d_i	$(d_i - đ)^2$
1	+0.4	+0.6		
2	+0.3	+0.5		
3	+0.9	+0.7		
4	+0.4	+0.6		
5	+1	+0.9		
6	+1	+1.1		
7	+1	+1.5		
8	+1	+2.1		

Exercise

$$\begin{cases} H_0 : \delta = 0 \\ H_1 : \delta \neq 0 \end{cases}$$

$$\bar{d} = \frac{2}{8} = 0.25$$

$$s_d^2 = \frac{1.14}{7} = 0.163 \Rightarrow s_d = 0.4$$

$$t = \frac{0.25}{0.4/\sqrt{8}} = \frac{0.25}{0.14} = 1.77$$

Since $t_{7;0.975}$=2.365 ⇒ $|t|<t_0$ not reject H_0

ID	F_1	F_2	$d_i = F_2 - F_1$	$(d_i-đ)^2$
1	+0.4	+0.6	+0.2	0.0025
2	+0.3	+0.5	+0.2	0.0025
3	+0.9	+0.7	-0.2	0.2025
4	+0.4	+0.6	+0.2	0.0025
5	+1	+0.9	-0.1	0.1225
6	+1	+1.1	+0.1	0.0225
7	+1	+1.5	+0.5	0.0625
8	+1	+2.1	+1.1	0.7225
			2	1.14

$$I.C._{95\%} = 0.25 \pm 2.365 \cdot 0.14 = [-0.08; 0.58]$$

The confidence interval includes 0 thus with confidence of 95% we can say that the two drugs do not have a different efficacy.

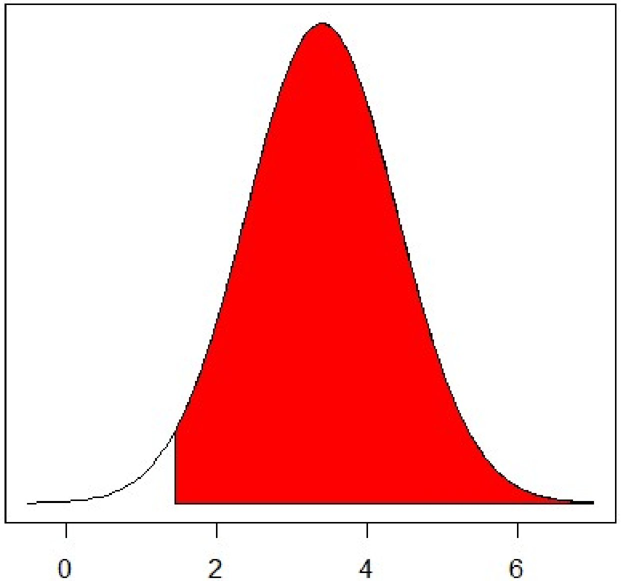
Power assessment

1- β = P(Deciding for H_1 | given that H_1 is true $\delta \neq 0$)

Let us assume that $H_1: \delta_{before-afte} = \Delta = 0.15$ is true, this implies

$$\bar{D}_{b-a} \sim N\left(\Delta;\, \sigma_d \sqrt{\frac{1}{n}}\right) \qquad Z = \frac{\bar{d}_{b-a}}{\sigma_d * \sqrt{\frac{1}{n}}} \sim N\left(\frac{\Delta}{\sigma_d * \sqrt{\frac{1}{n}}};\, 1\right)$$

standardised mean difference



The red area represents the chance (88%) of rejecting H_0 if H_1 is true

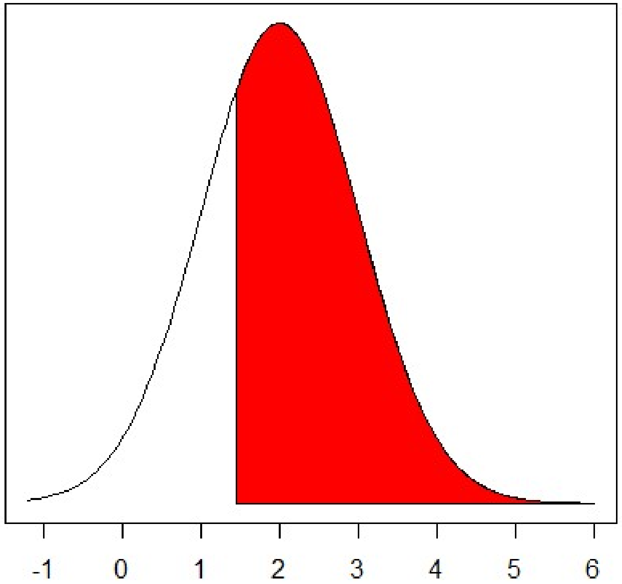
Power assessment

Let us consider an unmatched design

Let us assume that $H_1: \mu_{before} - \mu_{after} = \Delta = 0.15$ is true, this implies

$$T = \frac{\bar{Y}_b - \bar{Y}_a}{\sigma * \sqrt{\frac{1}{n_b} + \frac{1}{n_a}}} \sim N\left(\frac{\Delta}{\sigma * \sqrt{\frac{1}{n_b} + \frac{1}{n_a}}}; 1\right)$$

standardised difference between sampling means



The red area (63%) represents the chance of rejecting H_0 if H_1 is true.
The chance is reduced if we ignore the matching!