I Párový t-test

I.I Úvod

Párový t-test se používá k porovnání středních hodnot dvou populací. Porovnáváme dva vzorky, přičemž vzorky z jednoho pozorování mohou být spárovány se vzorky z druhého pozorování. Jedná se například o:

- pozorování před a po na stejném objektu (například testy studentů před a po absolvování kurzu);
- porovnání dvou různých metod měření nebo dvou různých léčeb, které jsou aplikovány na stejný subjekt (například měření krevního tlaku pomocí stetoskopu a jiného přístroje.)

I.II Návod na párový t-test

Mějme vzorek n studentů, kterým byl dán test před a po absolvování určitého kurzu. Chceme zjistit, zda výuka vedla ke zlepšení znalostí a dovedností studentů (tedy zda mají vyšší skóre z testu). Výsledky testovaného vzorku studentů nám pomohou udělat závěry o vlivu absolvovaného kurzu.

Nechť x je skóre z testu před kurzem a y je skóre z testu po kurzu.

Abychom mohli otestovat nulovou hypotézu, že rozdíl středních hodnot je nulový (tj. $H_0 : \mu_x - \mu_y = 0$), použijeme následující postup:

1. Vypočítáme rozdíl ($z_i = y_i - x_i$) mezi dvěma pozorováními pro každý pár. Musíme rozlišovat mezi pozitivním a negativním výsledkem.
2. Vypočítáme průměr rozdílů m_z.
3. Vypočítáme směrodatnou odchylku průměrů s_z, a použijeme ji pro výpočet směrodatné chyby průměru rozdílů, $SE(m_z) = \frac{s_z}{\sqrt{n}}$.
4. Vypočítáme testovací statistiku pomocí vzorce $T = \frac{m_z - 0}{SE(m_z)} = \frac{m_z}{\frac{s_z}{\sqrt{n}}}$, za platnosti nulové hypotézy má tato statistika t-rozložení (studentovo rozložení) s $n-1$ stupni volnosti.
5. Použijeme tabulky pro kvantily studentova rozložení, abychom porovnali hodnotu T-statistiky s $t_{1-\alpha/2}(n-1)$ rozdělením. Na základě distribuční funkce studentova rozložení můžeme získat p-hodnotu pro párový t-test.

Poznámka:
Aby byl tento test prokazatelný, je třeba, aby rozdíly z_i pocházely alespoň z **limitně** normálního rozložení. Proto se nedoporučuje používat tento test, máme-li extrémně odlehlé hodnoty.
Můžeme si všimnout, že zavedením rozdílů z_i a úpravou nulové hypotézy $H_0 : \mu_z = 0$ (kde $\mu_z = \mu_x - \mu_y$) dostaneme jednovýběrový t-test.

Příklad:
Použijeme výše uvedený příklad pro $n = 20$ studentů. Dostali jsme následující hodnoty:

student	skóre před kurzem	skóre po kurzu	rozdíl
1	18	22	+4
2	21	25	+4
3	16	17	+1
4	22	24	+2
5	19	16	-3
6	24	29	+5
7	17	20	+3
8	21	23	+2
9	23	19	-4
10	18	20	+2
11	14	15	+1
12	16	15	-1
13	16	18	+2
14	19	26	+7
15	18	18	0
16	20	24	+4
17	12	18	+6
18	22	25	+3
19	15	19	+4
20	17	16	-1

Nejdříve ověříme předpoklad normality pro „rozdíl“, protože máme malý rozsah náhodného výběru použijeme Shapiro-Wilkův test p=0,725, tedy nezamítáme nulovou hypotézu, že data pocházejí z normálního rozložení na hladině významnosti 0,05.
Výpočtem průměrů a směrodatné odchylky dostaneme, že $m_z = 2.05$ a $s_z = 2.837$.

Proto $SE(m_z) = \frac{s_z}{\sqrt{n}} = \frac{2.837}{\sqrt{20}} = 0.634$. A dostáváme tedy, že

$$t_0 = \frac{2.05}{0.634} = 3.231 \quad \text{s 19 stupni volnosti.}$$

Na základě nulové a alternativní hypotézy vypočteme kritický obor nebo p hodnotu.
Mějme nulovou hypotézu $H_0 : \mu_z = 0$ proti alternativní hypotéze $H_1 : \mu_z \neq 0$.
Potom kritický obor bude ve tvaru $W = (-\infty, -t_{1-\alpha/2}(n-1) \cup \langle t_{1-\alpha/2}(n-1), \infty) = (-\infty; -2,093\rangle \cup \langle 2,093; \infty)$, kvantil studentova rozložení vyhledáme v tabulkách. Pomocí softwaru můžeme vypočítat p-hodnotu dostaneme $p = 0,004395$.
Protože $t_0 \in W$ (resp. $p < 0,05$), zamítáme nulovou hypotézu o nulovosti rozdílu středních hodnot mezi výsledky studentů před a po absolvování kurzu ve prospěch alternativy, že rozdíl mezi výsledky studentů před a po absolvování kurzu je různý od nuly na hladině významnosti 0,05.

Mohli bychom také testovat nulovou hypotézu $H_0 : \mu_z \leq 0$ proti alternativě $H_1 : \mu_z > 0$.
Což můžeme převést na $H_0 : \mu_x - \mu_y = \leq 0$ a to je ekvivalentní $H_0 : \mu_x \leq \mu_y$.
Tedy dostáváme $H_0 : \mu_x \leq \mu_y$ proti $H_1 : \mu_x > \mu_y$.
Potom kritický obor bude ve tvaru $W = \langle t_{1-\alpha}(n-1), \infty) = \langle 1,7291; \infty)$, kvantil studentova rozložení vyhledáme v tabulkách. Pomocí softwaru můžeme vypočítat p-hodnotu dostaneme $p = 0,002197$.
Protože $t_0 \in W$ (resp. $p < 0,05$), zamítáme nulovou hypotézu ve prospěch alternativy, že výsledky studentů před absolvování kurzu jsou horší než výsledky studentů po absolvování kurzu na hladině významnosti 0,05. Můžeme tedy předpokládat, že v průměru vede absolvování kurzu ke zlepšení výsledků.

I.III Interval spolehlivosti pro průměr rozdílů

V předchozím příkladě byla odhadovaná změna $+2$ body. I když se jedná o statisticky významný rozdíl, jde o poměrně malou změnu. Bylo by dobré vypočítat interval spolehlivosti rozdílu středních hodnot, který nám řekne, v jakém rozmezí nejspíše bude ona změna ležet.

95% oboustranný interval spolehlivosti pro rozdíl středních hodnot je

$$m_z \pm t_{1-\alpha}(n-1) * \frac{s_z}{\sqrt{n}}$$

kde $t_{1-\alpha}(n-1)$ je 97,5% kvantil t-rozložení při $n-1$ stupních volnosti.

V našem příkladě:
Máme průměr 2,05. 97,5% kvantil t-rozložení s 19 stupni volnosti je 2,093. 95% interval spolehlivosti pro změnu je proto

$$2.05 \pm (2.093 \cdot 0.634) = 2.05 \pm 1.33 = (0.72; 3.38).$$

To potvrzuje fakt, že i když jsou rozdíly statisticky významné, jsou relativně nízké. Můžeme si však na 95 % být jisti, že skutečná střední hodnota leží někde mezi hodnotou o něco menší než jeden bod a hodnotou o něco větší než tři body.