I Dvouvýběrový t-test

I.I Úvod

Dvouvýběrový t-test se používá pro srovnání středních hodnot dvou populací. V textu bude dále používáno následující označení:

skupina	velikost výběru	výběrový průměr	výběrová směrodatná odchylka
1	n_1	m_1	s_1
2	n_2	m_2	s_2

I.II Návod na dvouvýběrový t-test

Pro otestování nulové hypotézy, že střední hodnoty dvou populací, μ_1 a μ_2, jsou stejné (tj. $H_0 : \mu_1 = \mu_2$):

1. Vypočítáme rozdíl dvou průměrů, $m_1 - m_2$.
2. Vypočítáme **sdruženou výběrovou směrodatnou odchylku** jako
$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$
3. Vypočítáme směrodatnou chybu v rozdílech mezi průměry
$$SE(m_1 - m_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$
4. Vypočítáme T-testovací statistiku určenou vzorcem
$$T = \frac{m_1 - m_2}{SE(m_1 - m_2)}.$$
Za platnosti nulové hypotézy má statistika (studentovo) t-rozdělení s $n_1 + n_2 - 2$ stupni volnosti.
5. Použijeme tabulky pro t-rozdělení, abychom porovnali hodnotu testovací statistiky s kvantilem studentova rozdělením. Na základě distribuční funkce studentova rozdělení získáme p-hodnotu pro dvouvýběrový t-test.

Poznámka:
Pro dvouvýběrový t-test by měly výběry **zhruba** pocházet z normálního rozdělení a jejich rozptyly by měly být **limitně** stejné. Pokud jsou rozptyly různé, musíme použít

$$SE(m_1 - m_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Poté $\frac{m_1 - m_2}{SE(m_1 - m_2)} \sim N(0,1)$ pokud n_1 a n_2 jsou rozumně velké.

Jinak $\frac{m_1 - m_2}{SE(m_1 - m_2)} \sim t_{n'}$, kde $n' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1-1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2-1)}}$ zaokrouhlené dolů na celé číslo.

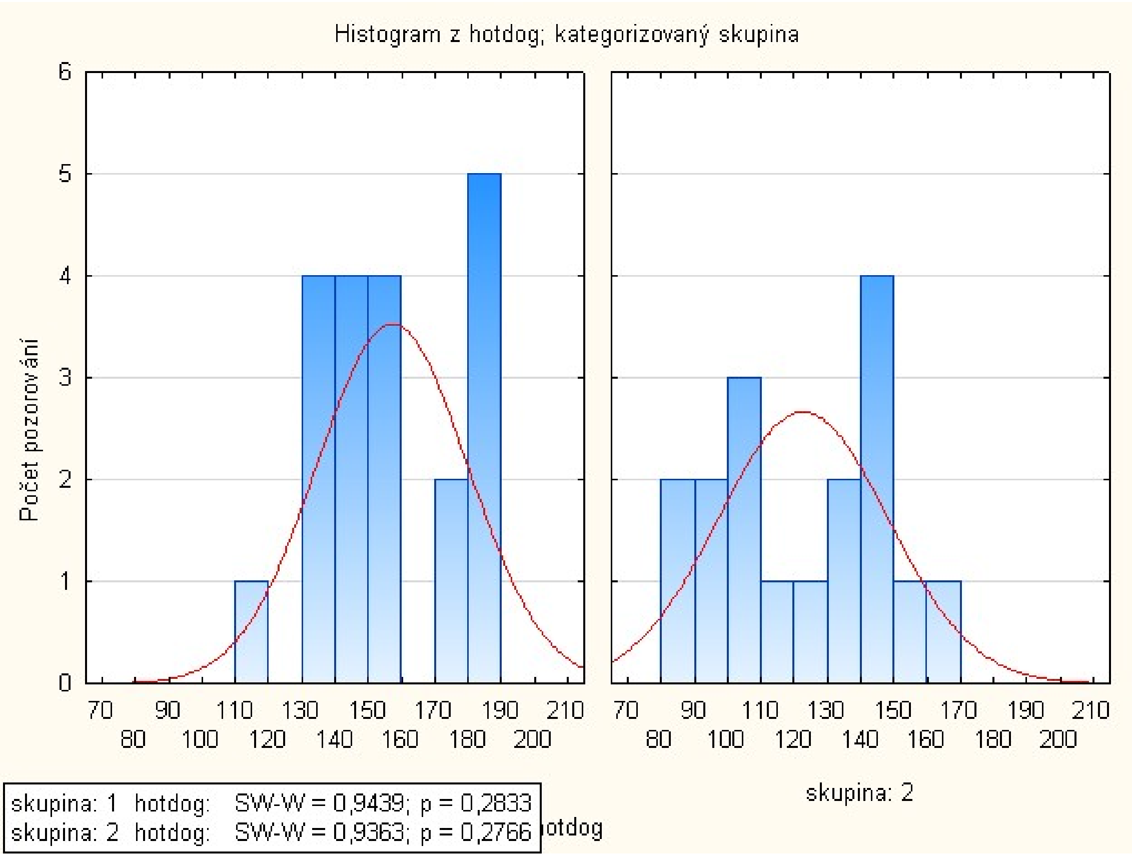
Příklad:
(Data jsou získaná z Moore a McCabe – Úvod do praktické statistiky)

Americký časopis, Consumer Reports, uskutečnil průzkum, kolik kalorií a sodíku obsahují různé značky hotdogů. Byly zkoumány tři různé typy hotdogů, pro naše účely porovnáme pouze hovězí a drůbeží hotdogy. Výsledky níže představují obsah kalorií v různých značkách hovězího a drůbežího hotdogu.

Hovězí hotdogy:
186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190, 157, 131, 149, 135, 132.

Drůbeží hotdogy:
129, 132, 102, 106, 94, 102, 87, 99, 170, 113, 135, 142, 86, 143, 152, 146, 144.

Před provedením t-testu bychom měli zkontrolovat, zda mají oba vzorky zhruba normální rozložení. Proto se podíváme na histogram dat nebo na krabicový graf. Nejsou zde žádné odlehlé



hodnoty a data rozumně připomínají tvar normálního rozdělení navíc p-hodnoty Shapiro-Wilkova testu jsou obě větší než 0,05, tudíž můžeme provést t-test. Nejdříve vypočítáme výběrový průměr a směrodatnou odchylku každé skupiny, jak to uvádí tabulka níže.

skupina	velikost výběru	výběrový průměr	výběrová směrodatná odchylka
hovězí	20	156.85	22.64
drůbeží	17	122.47	25.48

Tudíž máme celkem $m_1 - m_2 = 156.85 - 122.47 = 34.38$.

Směrodatné odchylky jsou přibližně stejné, proto můžeme vypočítat sdruženou výběrovou směrodatnou odchylku.

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{19 \cdot 22.64^2 + 16 \cdot 25.48^2}{35}} = 23.98.$$

Nyní můžeme vypočítat $SE(m_1 - m_2)$ jako

$$s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 23.98 \cdot \sqrt{\frac{1}{20} + \frac{1}{17}} = 7.91.$$

A nyní hodnotu T-statistiky jako

$$T = \frac{m_1 - m_2}{SE(m_1 - m_2)} = \frac{34.38}{7.91}$$

Na základě nulové a alternativní hypotézy vypočteme kritický obor nebo p hodnotu.
Mějme nulovou hypotézu $H_0 : \mu_x - \mu_y = 0$ proti alternativní hypotéze $H_1 : \mu_x - \mu_y \neq 0$.
Potom kritický obor bude ve tvaru $W = (-\infty, -t_{1-\alpha/2}(n_1+n_2-2) \cup \langle t_{1-\alpha/2}(n_1+n_2-2), \infty) =$ $= (-\infty; -2,03 \cup \langle 2,03; \infty)$, kvantil studentova rozložení vyhledáme v tabulkách. Pomocí softwaru můžeme vypočítat p-hodnotu dostaneme $p = 0,000114$.
Protože $t_0 \in W$ (resp. $p < 0,05$), zamítáme nulovou hypotézu o rovnosti středních hodnot obsahů kalorií hovězích a kuřecích hotdogů ve prospěch alternativy, že střední hodnoty obsahů kalorií hovězích a kuřecích hotdogů se liší na hladivě významnosti 0,05.
Na základě krabicového grafu bychom také mohli mít zájem testovat nulovou hypotézu::
$H_0 : \mu_x - \mu_y \leq 0$ proti alternativě $H_1 : \mu_x - \mu_y > 0$.
Což můžeme převést na $H_0 : \mu_x \leq \mu_y$.
Tedy dostáváme $H_0 : \mu_x \leq \mu_y$ proti $H_1 : \mu_x > \mu_y$.
Potom kritický obor bude ve tvaru $W = \langle t_{1-\alpha}(n_1 + n_2 - 2), \infty) = \langle 1,6896; \infty)$, kvantil studentova rozložení vyhledáme v tabulkách. Pomocí softwaru můžeme vypočítat p-hodnotu dostaneme $p = 0,000057$.
Protože $t_0 \in W$ (resp. $p < 0,05$), zamítáme nulovou hypotézu ve prospěch alternativy, že střední hodnota obsahu kalorií v hovězích hotdocích je vyšší než střední hodnota obsahu kalorií v kuřecích hotdocích na hladině významnosti 0,05. Proto zde máme přesvědčivý důkaz, že počet kalorií v drůbežím hotdogu je menší než počet kalorií v hovězím hotdogu.

I.III Interval spolehlivosti pro průměr rozdílů

Dále by bylo vhodné spočítat interval spolehlivosti rozdílu středních hodnot, který určí, v jakém rozmezí bude skutečný rozdíl ležet s pravděpodobností 95%.
95% oboustranný interval spolehlivosti se spočítá jako

$$(m_1 - m_2) \pm t^* \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{nebo ekvivalentně} \quad (m_1 - m_2) \pm t^* \cdot SE(m_1 - m_2),$$

kde t^* je 97,5% kvantil t-rozdělení při $n_1 + n_2 - 2$ stupních volnosti.

Z našeho příkladu:
Rozdíl průměrů je 34.38 kalorií. 97,5% kvantil t-rozdělení s 35 stupni volnosti je 2.03. 95% interval spolehlivosti je tak

$$34.38 \pm (2.03 \cdot 7.91) = 34.38 \pm 16.06 = (18.3, 50.4).$$

Můžeme tedy s 95% jistotou potvrdit, že střední hodnota obsahu kalorií v hovězím hotdogu je mezi přibližně 18 a 50 kaloriemi vyšší než střední hodnota obsahu kalorií u drůbežích hotdogů.