

### 4.5.3 Kanonická korelační analýza

Kanonická korelační analýza je vícerozměrná metoda, která se používá ke zkoumání závislosti mezi dvěma skupinami proměnných. První ze dvou skupin se považuje za *soubor závisle proměnných*  $y$  a druhá za *soubor nezávisle proměnných*  $x$ . Toto rozdělení je ale čistě účelové z důvodu výkladu a nemá žádný vliv na řešení problému. Jde v podstatě o rozšíření metody vícenásobné lineární regrese a korelační analýzy. Zatímco ve vícenásobné lineární regresi hledáme nejlepší kombinaci  $m$  nezávisle proměnných  $x_1, x_2, \dots, x_m$  k výpočtu *jediné* závisle proměnné  $y$  hledáme v kanonické korelační analýze lineární vztah  $U_1 = a_1 y_1 + a_2 y_2 + \dots + a_p y_p$  mezi skupinou  $p$ , čili *více než jediné*, závisle proměnných  $y_1, y_2, \dots, y_p$ , a dále lineární vztah  $V_1 = b_1 x_1 + b_2 x_2 + \dots + b_p x_p$  mezi skupinou  $m$  nezávisle proměnných  $x_1, x_2, \dots, x_p$ . Podstata metody spočívá v tom, že se v každé skupině proměnných vyhledávají koeficienty  $a$  a  $b$  tak, aby pro všech  $n$  objektů vyčíslené kanonické proměnné  $U_i$  a  $V_i$ ,  $i = 1, \dots, n$ , vykazovaly maximální párový korelační koeficient. Po jejich nalezení se pak hledají další lineární kombinace čili kanonické proměnné  $U_2$  a  $V_2$ , které mají druhý největší korelační koeficient za podmínky, že  $U_2$  a  $V_2$  jsou nekorelované s prvními kanonickými proměnnými  $U_1$  a  $V_1$ .

V kanonické korelační analýze jsou koeficienty  $a$  a  $b$  hledány tak, aby maximalizovaly korelaci mezi proměnnými  $U_1$  a  $V_1$ . Po nalezení nejlepších odhadů  $a$  a  $b$  se  $U_1$  nazývá *první kanonická proměnná* závisle proměnných  $y$  a  $V_1$  *první kanonická proměnná* nezávisle proměnných  $x$ . Obě kanonické proměnné mají průměr roven nule. Korelace mezi  $U_1$  a  $V_1$  se nazývá *první kanonická korelace* a čtverec této korelace je nazýván *vlastní číslo*.

První kanonická korelace je tudíž největší možná korelace mezi lineárními kombinacemi závisle proměnných  $y$  a lineárními kombinacemi nezávisle proměnných  $x$ . První kanonická korelace představuje analogii vícenásobnému korelačnímu koeficientu ve vícenásobné lineární regresi mezi *jedinou* závisle proměnnou  $y$  a souborem nezávisle proměnných  $x$ . Rozdíl proti vícenásobné lineární regresi je pouze v tom, že u kanonické korelace je *několik* závisle proměnných  $y$  a dále je nutno navíc hledat lineární kombinaci mezi nimi.