Canonical Correlation

simple correlation -- \( y_1 = x_1 \)

multiple correlation -- \( y_1 = x_1 \ x_2 \ x_3 \)

canonical correlation -- \( y_1 \ y_2 \ y_3 = x_1 \ x_2 \ x_3 \)

• The “most multivariate” of the correlation models

• All other models (including ANOVAs, ANCOVAs and MANOVA and LDF) are part of this model

Uses and Alternatives for Canonical Correlation

Just because there are multiple criterion variables and multiple predictors doesn’t mean that a canonical correlation is the best analysis for the data!

• If your research hypotheses/questions are about correlations between specific criterion variables and specific predictor variables ...

  • use the specified set of Pearson's correlations

  • use Hotelling’s t or Rosenthal’s Z to test hypotheses about which correlations are larger than others (which predictors are more correlated with which criterion variables)

• If your research hypotheses/questions are about differences between the best models to understand/predict specific criterion variables

  • build the specified set of multiple regression models

  • use “cross-validation” procedures and Hotelling’s/Rosenthal’s tests to compare multiple regression models across the criterion variables

Let’s take a look at how canonical correlation “works”, to help understand when to use it (instead of simple or multiple reg)…

Start with multiple y and x variables

\( y_1 \ y_2 \ y_3 = x_1 \ x_2 \ x_3 \)

• construct a “canonical variate” as the combination of y variables

  \( CV_{y1} = b_1 y_1 + b_2 y_2 + b_3 y_3 \)

• construct a “canonical variate” as the combination of x variables

  \( CV_{x1} = b_1 x_1 + b_2 x_2 + b_3 x_3 \)

• The canonical correlation is the correlation of the canonical variates

\( R_c = r_{cvy1, cvx1} \)
1st thing -- The weights used to construct $CV_{y1}$ and $CV_{x1}$ are chosen to produce the largest possible $R_c$.

- This is different from simple correlation
  - weighting for $X_1$ is chosen so $y'$ mimics $y$
- This is different from multiple correlation
  - weighting for $x_1$, $x_2$, etc chosen to maximize the correlation of $y'$ with $y$
- In canonical correlation the weightings for the $Y$ variables and for the $X$ variables are chosen simultaneously to maximize the correlation between the constructed $X$ variate and the constructed $Y$ variate.

2nd thing -- There can be multiple canoniicals.

- max # is the smaller of the number of criterion & number of predictors
- like ldf, have to decide if we have a concentrated or diffuse structure

So, canonical correlation is used (instead of comparisons among selected simple and/or multiple correlations) when …

- research questions are about which combination(s) of criterion variables are related to which combination(s) or predictors
- research questions are about whether the set of variables have a concentrated or a diffuse correlational structure

Examples:

- have a set of interrelated criterion variables (e.g., five measures of depression) and want to know if they all share the same relationship with the predictors (a concentrated structure)
- have two sets of predictor variables and want to show that one set is most useful for understanding the one criterion variable and the other set is most useful for understanding a different set of criterion variables

Steps of inspecting and describing a canonical correlation...

- Determine the number of canonical variates and correlations
  - maximum number is smaller of # predictors or # criterion
  - each $R_c$ is tested for $H_0$: $R_c = 0$ (same test as for MR)
- Evaluate the strength of each canonical relationship
  - $R_c^2$ is “shared variance” between that x and y variate
- Interpret each canonical variate
  - Structure weights tell which variables related to which variates
  - Standardized regression weights tell about the unique contribution of variables to variates
Depicting and Describing “Shared Variance” in Canonical Correlation

There are 3 kinds of “shared variances” in cancorr:

- \( \text{RC}^2 \)
  - squared canonical correlation
  - variance shared between the corresponding Y and X canonical variates at

- PC
  - variance shared between a canonical variate and the set of variables from which it is constructed
  - an index of how well that canonical variate “represents” the set of variables
  - PC is computed as the average structure weight for that canonical variate

- RC
  - redundancy coefficient
  - variance shared between a canonical variate from one set of variables and the other set of variables
  - an index of how well a canonical variate “represents” the other set of variables

We start with dependent (y) variables and covariate/predictor (x) variables

- a canonical variate is constructed for each set so that the variates are as highly correlated as possible
- the \( \text{RC}^2 \) tells the variance shared between the canonical variates
- the PC tells the variance shared between each canonical variate and the variables from which it was constructed

\[
\text{CV}_{y1} = b_1 y_1 + b_2 y_2 + b_3 y_3 + a
\]

\[
\text{CV}_{x1} = b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + a
\]

We can also ask how much variance is shared between the CV of one set of variables (the covariate/predictor variables) and the other set of variables (the dependent variables) — a redundancy coefficient

- shown is “how much variance among the dependent variables (Ys) is accounted for by the covariate/predictor canonical variate (CV\(_{x1}\))
- a redundancy coefficient is calculated as \(\text{RC}^2 \times \text{PC} \) of the variables involved (\(\text{RC}^2 \times \text{PC}_{y1}\), here)
About the size of the redundancy coefficient

- the “old” story was that a small redundancy coefficient meant the canonical correlation analysis was uninteresting or misleading

- the “new” story a redundancy coefficient can be “small” for two different reasons that have very different interpretations
  - “good” -- the CV is “strongly” composed of a minority of the variables in that set (few vars, but with large weights)
    - often denotes specificity among the variables -- a prerequisite for a diffuse structure
  - “yucky” -- the CV is “weakly” composed of a majority of the variables in that set (many variables, but small weights)
    - often denotes that the construct related to the “other vars” is not well-represented by these vars, and so must be “put together” using small portions of many vars