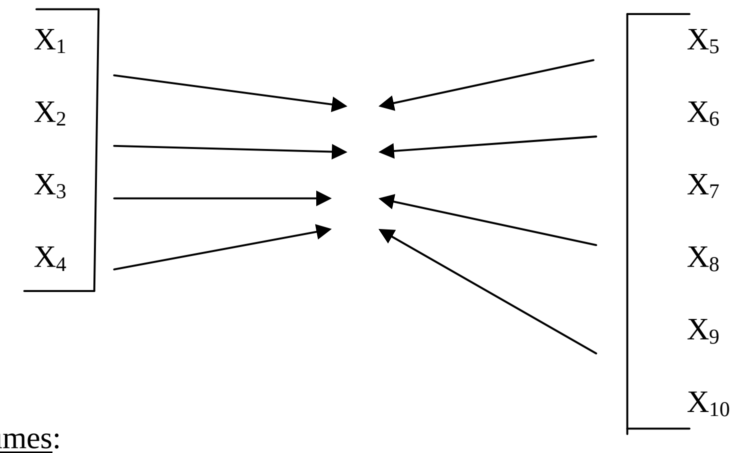
Neuendorf
Canonical Correlation

<u>The Model</u>



<u>Assumes</u>:

1. Metric (interval/ratio) data for all variables. As with Pearson=s r, there is no true distinction between IVs and DVs; SPSS calls the two groups of variables SET 1 and SET 2 instead.

2. Multivariate normality (required for test of statistical significance).

3. LinearityBin 2 ways: Both in the creation of canonical variates (CVs), and in the correlational relationship(s) between those CVs.

4. Canonical variates are orthogonal within a given analysis (uncorrelated, like orthogonal factors)--we are trying to identify statistically independent patterns of linkage between two sets of variables. There can be as many CV pairs as the smallest k (k_1 = number of variables in SET 1; k_2 = number of variables in SET 2).

<u>Statistics</u>:

1. Standardized canonical coefficients (like betas; thus, they are partials). CV1-1 is the first canonical variate for SET 1, and CV2-1 is the first canonical variate for SET 2:

   $[CV1\text{-}1 = ß_1X_{z1} + ß_2X_{z2} + ß_3X_{z3} + ß_4X_{z4}]$ ----> R_c <---- $[CV2\text{-}1 = ß_5Y_{z5} + ß_6Y_{z6} + ß_7Y_{z7}]$

2. Unstandardized canonical coefficientsBSPSS calls them Araw@.

3. Canonical loadings--correlations between individual variables and their canonical variate(s) (CVs). Proposed by some as better than coefficients (betas) in interpreting the various canonical variates (CVs). We can check for significance of the loadings in the same fashion as for factor analysis--see Table 3-2 on p. 128 of Hair et al.

4. Canonical correlation (R_c or CC)--the simple zero-order correlation between a pair of canonical variates (e.g., CV1-1 and CV2-1 above). Its significance is tested via a Wilks= lambda, with a chi-square transformation. Squaring the R_c gives us the proportion of variance shared by the two canonical variates.

5. Cross loadings--correlations between individual variables and the CV(s) for the other set (i.e., on the opposite side).

6. Redundancy analysis--looks at the variance shared between each CV and the corresponding sets of individual variables, both on its Aown@ set, and on the Aother side.@ For example, let's suppose that CV2-1 for a set of 3 variables represents 68% of the total variance in that set (in SPSS, we would find this under Redundancy Analysis, under AProportion of Variance of Set 2 Explained by Its Own Canonical Variate.@). Let=s also assume that the R_c for CV1-1 and CV2-1 is .48 (squared, this is .23, representing a shared variance between CV1-1 and CV2-1 of 23%). The redundancy figure for this is .68 x .23 = .156. This means that 15.6% of the total variance for the three variables in SET 2 is explained by the CV1-1 for SET 1. This can be done for both Asides@ across all CV pairs. The attached AFigure 1" includes redundancy analysis.

Other notes:

1. Just a question: Why do canonical correlation? One answer: As an exploratory technique to discover patterns of linear relationships between two sets of variables, sort of like a two-sided factor analysis. It=s not really appropriate for hypothesis testing in most cases. (There *are* some exceptions.)

2. In SPSS, full canonical correlation analysis is unavailable by clicking through dialog boxes. You must use a macro that has written by SPSS users, and is provided with SPSS (at least through version 15.0; we're not so sure about 16.0). If you do not have the file resident in your SPSS program folder, you can get it via email from your instructor. The file is named Canonical Correlation.sps. Running canonical correlation requires that you use very simple syntax, as shown below. Or you can find Acanonical correlation macro@ in the SPSS Help menu, swipe the three lines of syntax needed, right click, and Acopy.@ Go to a syntax box, paste, and *replace* @varlist1" and Avarlist2" with your two variable lists. The three lines of syntax as indicated in the Help dialogue box are:

```
INCLUDE 'Canonical correlation.sps'.

CANCORR SET1=varlist1 /
```

```
SET2=varlist2 / .
```

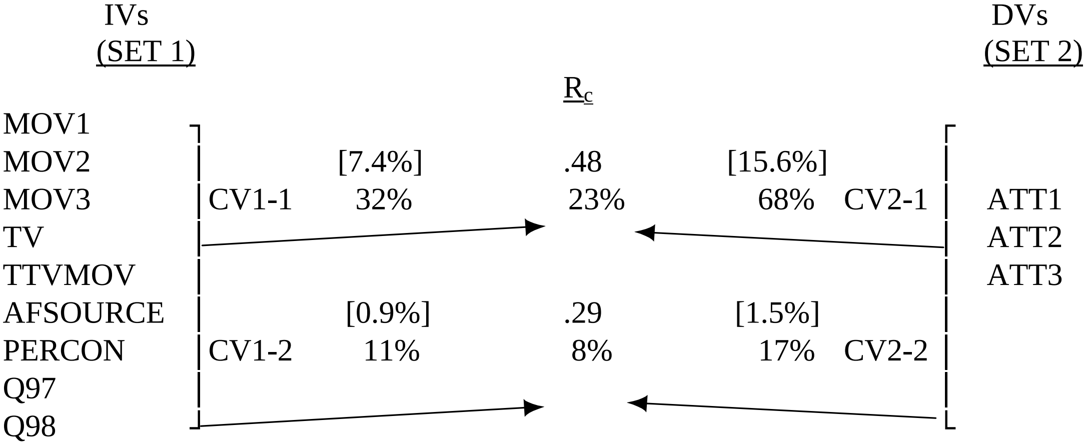
So, your syntax may look something like the following example:

```
INCLUDE 'C:\Program Files\SPSS\Canonical correlation.sps'.
CANCORR SET1=Q5,Q6,Q7,Q8,Q9/
        SET2=Q21,Q22,Q23,Q34.
```

Do not be alarmed when the macro procedure writes canonical variate scores automatically at the end of your data set. Also, do not be alarmed if in running canonical correlation, your data file is automatically given a new name. Simply do *not* save it after running if you don=t want the CV scores.

Also, do not be alarmed if SPSS seems to spazz out while running the CANCORR macroBvarious matrices will flash by on your screen. This is normal.

Figure 1. Redundancy Analysis Example.



NOTE: The bracketed numbers show the portion of Aredundancy analysis@ that indicates what proportion of the variance of a set is shared with the canonical variate on the Aother side.@

Question: What=s missing from Figure 1 that will give us full information for interpreting the findings?

SPECIAL NOTE: We=ve had some problems in the past getting CC to run in SPSS. I have been successful in all cases if I try some combination of the following four guidelines (oddly, different combinations work for different installations of SPSS). So, if you=re having trouble, try the following guidelines:

1. Specify the entire path for the macro. That is, replace ACanonical correlation.sps.@ with AC:\Program Files\SPSS\Canonical correlation.sps.@ (If you=re not on a COM machine, the path may be different.)
2. Make sure the SET1 and SET2 subcommands are on different lines.
3. Make sure the syntax lines are single spaced (in Help, they=re displayed as double spaced, as shown on p. 2).
4. Delete the last slash. Slashes typically are dividers between subcommands, and there doesn=t seem to be any reason for the final one here.

References:
Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation*. Beverly Hills, CA: Sage Publications.