CANONICAL CORRELATION

(Or how I learned to stop worrying, and snoop the data)

Jonathan D. Herzberger
COM 631
April 15, 2013

PART I: MODEL

Welcome to the madcap world of canonical correlation, where we basically throw multiple variables into a gladiatorial arena, looking for patterns of linear relationships between the two sets. When the dust settles, which variables will be teamed up?

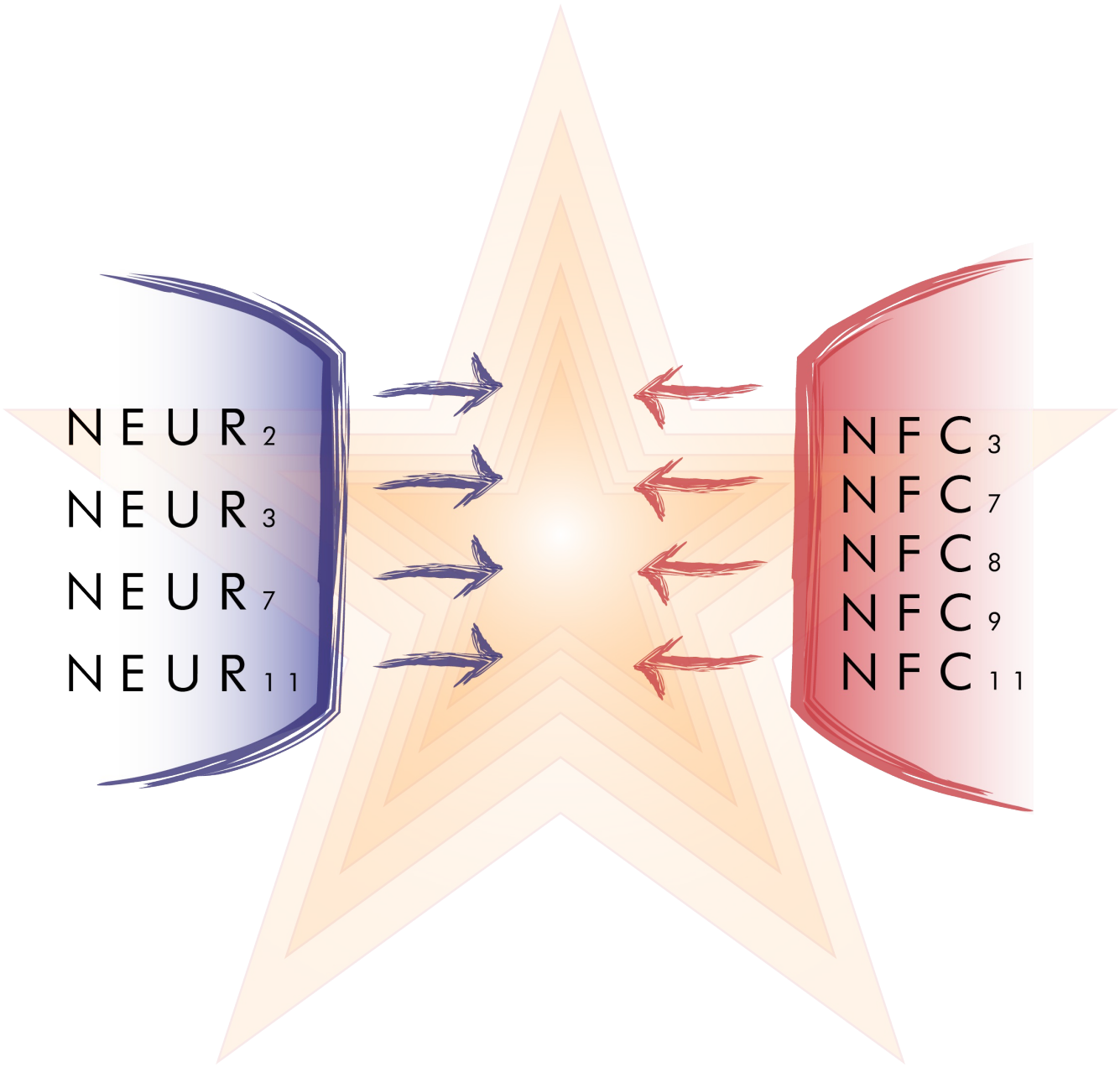
TODAY'S CONTESTANTS

Today, we'll be looking for patterns of linear relationships between two distinct sets of variables from the 2011 Neuendorf et al. Music and Film Experiment. Let's meet them, shall we?

SET ONE: NEUROTICISM	SET TWO: NEED FOR COGNITION
Neuroticism 2: Do you ever feel 'just miserable' for no reason?	Need For Cognition 3: Thinking is not my idea of fun.
Neuroticism 3: Are you an irritable person?	Need For Cognition 7: I only think as hard as I have to.
Neuroticism 7: Are you a worrier?	Need For Cognition 8: I prefer to think about small, daily projects than long-term ones.
Neuroticism 11: Do you often feel lonely?	Need For Cognition 9: I like tasks that require little thought once I've learned them.
	Need For Cognition 11: I really enjoy a task that involves coming up with new solutions

Note: Set one variables were measured via a dummy-coded binary response wherein 0=no, and 1=yes. The Need for Cognition questions were answered via 5-point Likert-type scale wherein 1 = extremely uncharacteristic of you (not at all like you), 2 = somewhat uncharacteristic, 3 = uncertain, 4 = somewhat characteristic, and 5 = extremely characteristic of you (very much like you). In addition, please note that questions 3, 7, 8 and 9 are worded negatively; these were not reverse-coded for this analysis.

Our Model



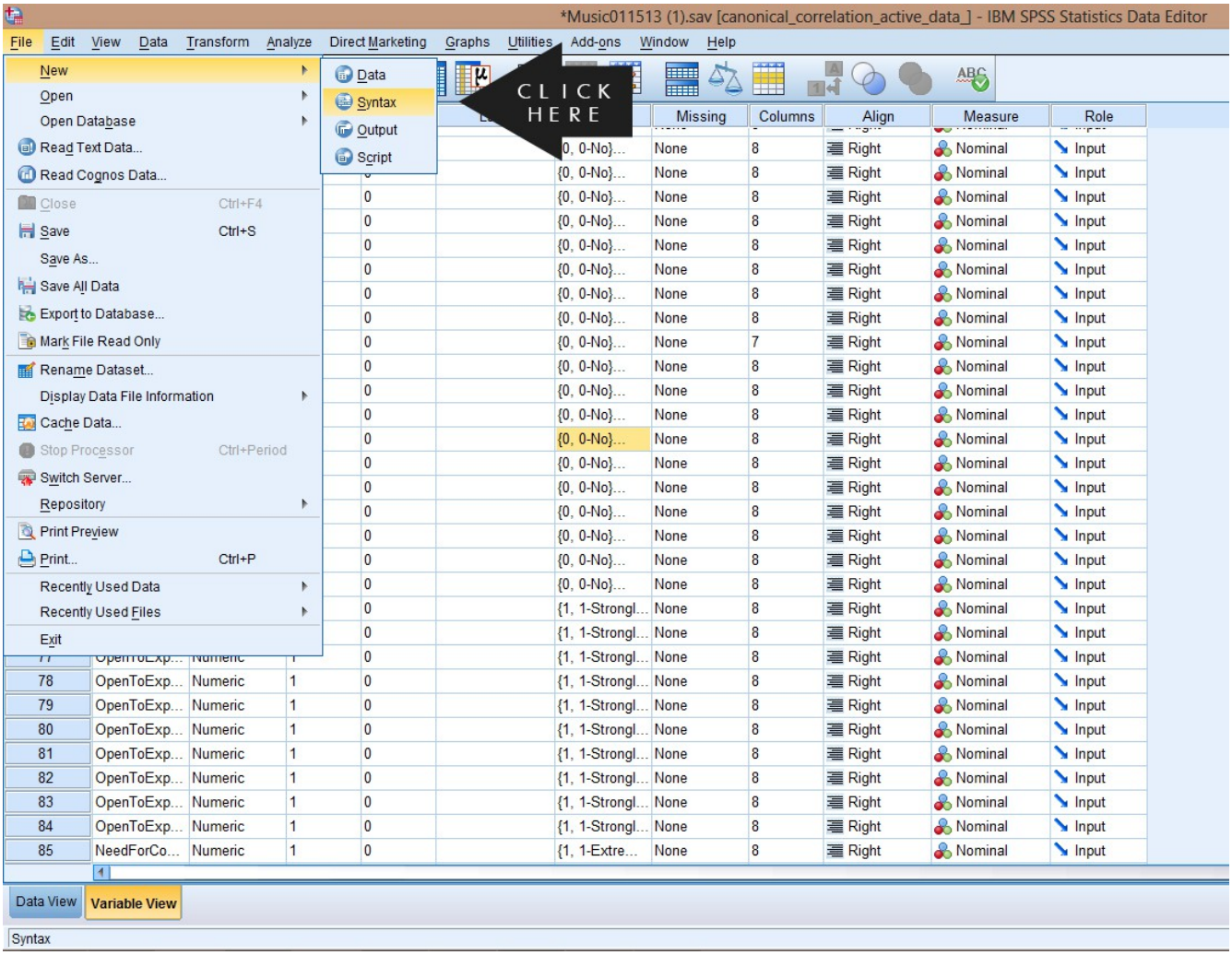
PART II: RUNNING SPSS

Unlike other procedures, Canonical Correlation can only be run via syntax. And unfortunately, unlike other syntax procedures, it can be very iffy from version to version. Yours should look something like:

INCLUDE 'C:\Program Files\IBM\SPSS\Statistics\21\Samples\English\Canonical correlation.sps'.
CANCORR SET1=NameOfVariable1 NameOfVariable2/
SET2=NameOfVariable3 NameOfVariable4/.

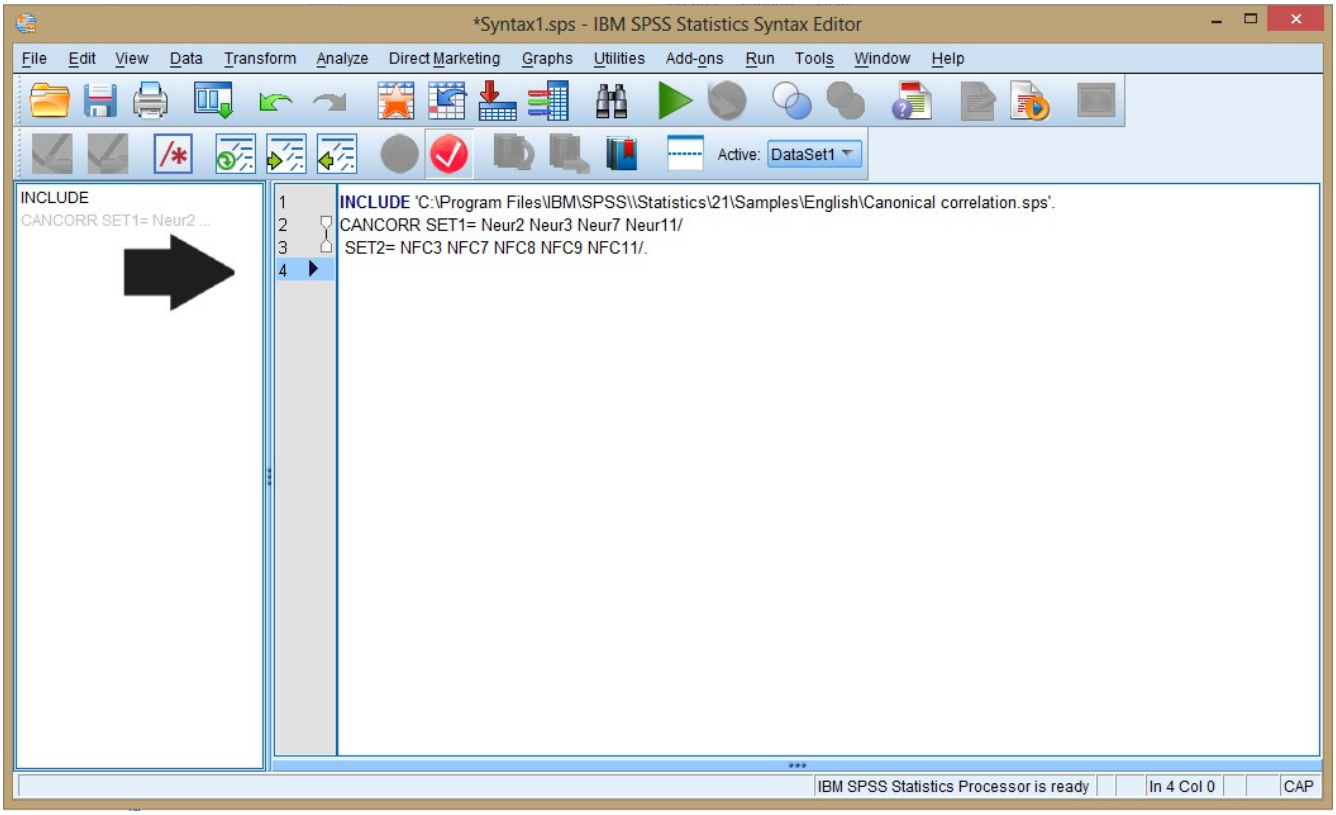
- with the qualifier that this may not work right off the bat, and you may need to try some different approaches to get it to work in your version of SPSS. For now, see Dr. Neuendorf's handout for additional relevant details.

Click **FILE → NEW → SYNTAX**



Cool! I bet we can put all sorts of things in there!

Now that we have a shiny new syntax box, let's put stuff in it!



Enter in the syntax as seen here, substituting your variable sets for the dummy variables included below. Remember - this can be tricky. SPSS 16 seems to hate Canonical Correlation, and the syntax will have different hiccups depending upon your edition. Be smart, try different things, but as always, your mileage may vary. Still, you'll want to put in some variant of:

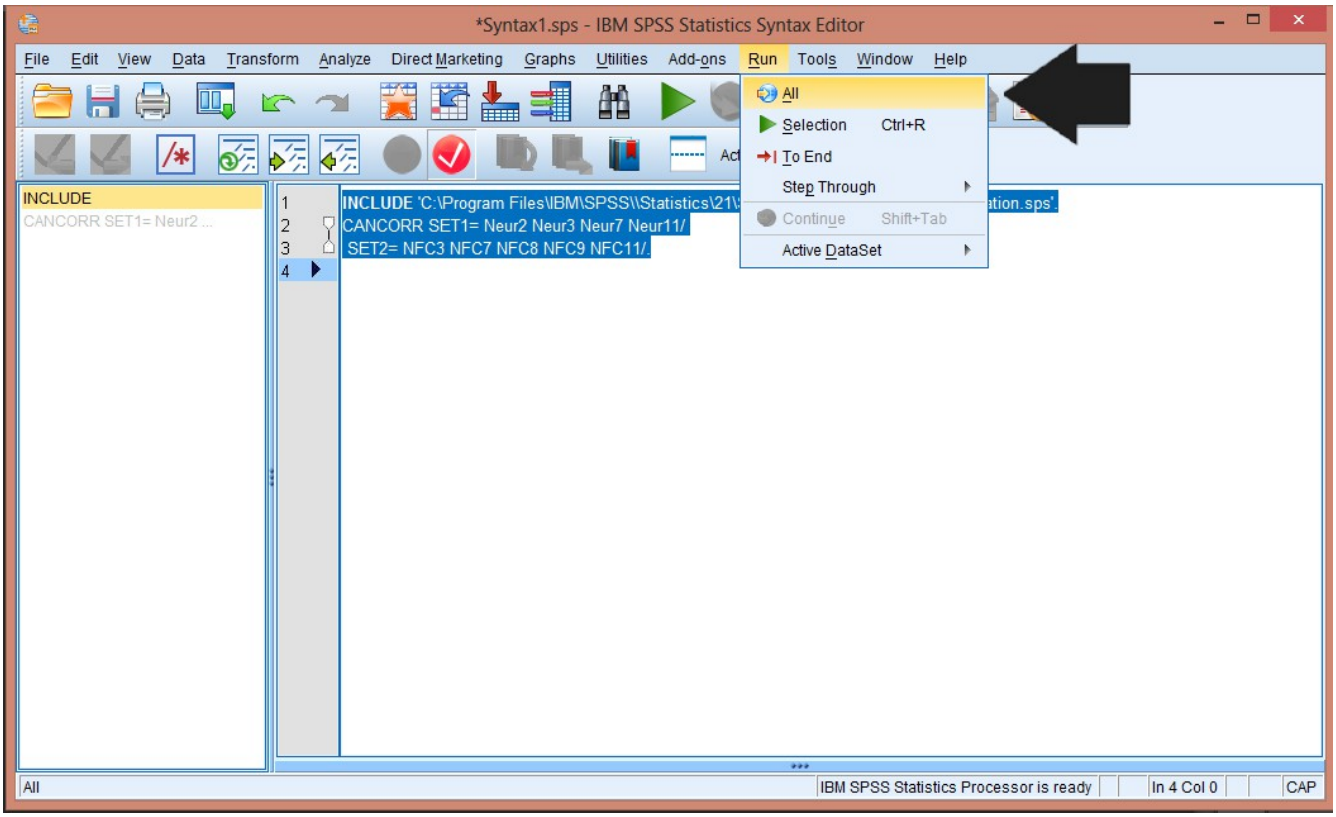
```
INCLUDE 'C:\Program Files\IBM\SPSS\Statistics\21\Samples\English\Canonical
correlation.sps'.
CANCORR SET1=IndieVariable1 IndieVariable2/
 SET2=DeepVariable1 DeepVariable2/.
```

Worth keeping in mind: the SPSS Macro that we're calling on here is tailored to older versions – it attempts to rewrite files that are simply not present in SPSS 21. So, if you're using v.21, be prepared for an entire tree's worth of error messages like the following:

```
Error # 2302 in column 12.  Text: canonical_correlation_AB_.sav
The specified file does not exist and cannot be erased.
Execution of this command stops.
```

Just be advised – your data is probably in there! Somewhere...

Then, click **RUN → ALL**



Exciting, no?

PART III: SPSS OUTPUT

```
GET
  FILE='C:\Users\Jonathan\Documents\Academic\CSU\Graduate\631
Multivariate\Canonical Correlation\Work in progress\Music011513 (1).sav'.
DATASET NAME DataSet1 WINDOW=FRONT.

SAVE OUTFILE='C:\Users\Jonathan\Documents\Canonical Music.sav'
  /COMPRESSED.
DATASET ACTIVATE DataSet1.

SAVE OUTFILE='C:\Users\Jonathan\Documents\Canonical Music.sav'
  /COMPRESSED.
INCLUDE 'C:\Program Files\IBM\SPSS\\Statistics\21\Samples\English\Canonical
correlation.sps'.
  18  0  * Canonical correlation.sps.  This version allows long variable names and
uses datasets.Canonical correlation.sps.
  19  0
  20  0  preserve.
  21  0  set printback=off.
 724  0  RESTORE.
 725  0
 727  0  * End of INSERT and INCLUDE nesting level 01.
CANCORR SET1= Neur2 Neur3 Neur7 Neur11/
 SET2= NFC3 NFC7 NFC8 NFC9 NFC11/.
```

Matrix

```
Run MATRIX procedure:


Correlations for Set-1
         Neur2   Neur3   Neur7  Neur11
Neur2   1.0000   .3565   .1973   .4220
Neur3    .3565  1.0000   .0769   .1292
Neur7    .1973   .0769  1.0000   .1890
Neur11   .4220   .1292   .1890  1.0000


Correlations for Set-2
         NFC3    NFC7    NFC8    NFC9   NFC11
NFC3   1.0000   .3672   .5074   .3295  -.4239
NFC7    .3672  1.0000   .3699   .3109  -.1765
NFC8    .5074   .3699  1.0000   .3649  -.1751
NFC9    .3295   .3109   .3649  1.0000  -.1646
NFC11  -.4239  -.1765  -.1751  -.1646  1.0000


Correlations Between Set-1 and Set-2
          NFC3    NFC7    NFC8    NFC9   NFC11
Neur2    .1759   .1942   .2709   .3030  -.2188
```

Neur3	.2643	.1163	.2374	.1146	-.1710
Neur7	-.0016	.2110	.1133	.2086	-.0532
Neur11	-.0097	.0075	.0454	.2647	.0632

Canonical Correlations

1	.430
2	.334
3	.209
4	.085

Test that remaining correlations are zero:

	Wilk's	Chi-SQ	DF	Sig.
1	.687	35.672	20.000	.017
2	.843	16.207	12.000	.182
3	.949	4.956	6.000	.550
4	.993	.694	2.000	.707

Standardized Canonical Coefficients for Set-1

	1	2	3	4
Neur2	-.811	-.120	.169	-.832
Neur3	-.202	-.595	-.540	.678
Neur7	-.409	.357	.589	.644
Neur11	.176	.772	-.761	.169

Raw Canonical Coefficients for Set-1

	1	2	3	4
Neur2	-1.614	-.239	.337	-1.656
Neur3	-.426	-1.252	-1.138	1.428
Neur7	-.829	.723	1.193	1.304
Neur11	.374	1.638	-1.615	.359

Standardized Canonical Coefficients for Set-2

	1	2	3	4
NFC3	.238	-.585	-.804	.580
NFC7	-.310	.079	.804	.673
NFC8	-.473	-.277	-.029	-.154
NFC9	-.448	.844	-.466	-.156
NFC11	.458	.328	-.395	.810

Raw Canonical Coefficients for Set-2

	1	2	3	4
NFC3	.196	-.481	-.660	.476
NFC7	-.252	.064	.654	.548
NFC8	-.354	-.207	-.022	-.115
NFC9	-.349	.657	-.363	-.121
NFC11	.507	.363	-.437	.895

Canonical Loadings for Set-1

	1	2	3	4
Neur2	-.889	.064	-.229	-.392

Neur3	-.500	-.511	-.533	.453
Neur7	-.551	.433	.437	.564
Neur11	-.269	.712	-.648	.027

Cross Loadings for Set-1

	1	2	3	4
Neur2	-.383	.021	-.048	-.033
Neur3	-.215	-.171	-.112	.039
Neur7	-.237	.145	.091	.048
Neur11	-.116	.238	-.136	.002

Canonical Loadings for Set-2

	1	2	3	4
NFC3	-.458	-.558	-.509	.354
NFC7	-.618	-.034	.423	.638
NFC8	-.711	-.294	-.240	.190
NFC9	-.714	.521	-.426	.055
NFC11	.569	.472	-.115	.498

Cross Loadings for Set-2

	1	2	3	4
NFC3	-.197	-.186	-.107	.030
NFC7	-.266	-.011	.089	.054
NFC8	-.306	-.098	-.050	.016
NFC9	-.307	.174	-.089	.005
NFC11	.245	.158	-.024	.042

Redundancy Analysis:

Proportion of Variance of Set-1 Explained by Its Own Can. Var.

	Prop Var
CV1-1	.354
CV1-2	.240
CV1-3	.237
CV1-4	.169

Proportion of Variance of Set-1 Explained by Opposite Can.Var.

	Prop Var
CV2-1	.066
CV2-2	.027
CV2-3	.010
CV2-4	.001

Proportion of Variance of Set-2 Explained by Its Own Can. Var.

	Prop Var
CV2-1	.386
CV2-2	.179
CV2-3	.138
CV2-4	.164

```
Proportion of Variance of Set-2 Explained by Opposite Can. Var.
               Prop Var
CV1-1              .072
CV1-2              .020
CV1-3              .006
CV1-4              .001

------ END MATRIX -----
```

Create

Created Series

SET_NUM		Series Name	Case Number of Non-Missing Values		N of Valid Cases	Creating Function
			First	Last		
1.00	1	VARSEQ	1	4	4	CSUM(VARSEQ)
2.00	1	VARSEQ	5	9	5	CSUM(VARSEQ)

```
The canonical scores have been written to the active dataset.
Also, a file containing an SPSS Scoring program has been written
To use this file, GET a system file with the SAME variables
that were used in the present analysis.  Then use an INSERT command
to run the scoring program.
For example :

GET FILE anotherfilename
INSERT FILE= 'canonical_correlation_scoring_syntax_.sps' .
EXECUTE.
```

PART IV: TABLING RESULTS

Canonical correlation 1

SET 1	LOADING		SET 2	LOADING
Neur2	-.889 *	Rc = .430	NFC3	-.458
Neur3	-.500		NFC7	-.618 *
Neur7	-.551 *	Rc^2 = .185	NFC8	-.711 *
Neur11	-.269		NFC9	-.714 *
			NFC11	.569 *

CV1-1=.354 CV2-1=.386

Wilk's Lambda: .687
Chi-Square: 35.67
Degrees of Freedom: 20
p = .017

Canonical correlation 2

SET 1	LOADING		SET 2	LOADING
Neur2	.064	Rc=.334	NFC3	-.558 *
Neur3	-.511		NFC7	-.034
Neur7	.433	Rc^2=.111	NFC8	-.294
Neur11	.712 *		NFC9	.521
			NFC11	.472

CV1-2=.240 CV2-2=.179

Wilk's Lambda: .843
Chi-Square: 16.207
Degrees of Freedom: 12
p = .182

* - Significant loading at $p < .05$ (Hair et al.)

PART V: WRITEUP OF RESULTS

So! A canonical correlation analysis was performed, exploring the relationship between two sets of variables; namely, measures of neuroticism, and measures of need for cognition.

From our analysis, we find one significant canonical correlation. In this first function, our Rc = .430, which is to say that the two variates have 18.5% shared variance. Wilk's Lambda was found to be significant via a chi-square test with 20 degrees of freedom, and $p = .017$.

To examine canonical root loadings, Hair et al.'s factor loading guidelines were used. Loadings greater than .55 were found to be significant, given the sample size of 101. To this effect, two of the four variables included in set 1 have significant loadings in CV1-1.

The significant variables are:

- Neur2 (Do you ever feel 'just miserable' for no reason?) (loading = -.889)
- Neur7 (Are you a worrier?) (loading = -.551)

Also, CV1-1 accounts for 35.4% of the variance across the variables in set 1.

In set two, four of the five variables had significant loadings in CV2-1. Specifically:

- NFC7 (I only think as hard as I have to) (loading = -.618)
- NFC8 (I prefer to think about small, daily projects than long-term ones) (loading = -.711)
- NFC9 (I like tasks that require little thought once I've learned them) (loading = -.714)
- NFC11 (I really enjoy a task that involves coming up with new solutions) (loading = .569)

Additionally, CV2-1 accounts for 38.6% of the variance across the variables in in set #2.

In our second function, Rc = .334, is non-significant ($p = .182$), and will therefore not be further examined.

In the first canonical function, both of the loadings in set one are negative. In set two, all of the loadings are negative, except the positively worded NFC11. This indicates that the less that someone feels miserable for no reason, or is a worrier, the less they'll only think as hard as they have to, the less they prefer to think about small projects instead of long-term ones, the less they like tasks that require little thought once they've learned them, and the more they really enjoy tasks that involve coming up with new solutions.

In summary, the less misery and worrying one experiences, the more likely one is to think about things deeply. Neat!