

Quiz #3 Demonstration of Statistical Modeling with ldf and MANOVA

```
discriminant groups = method(1,3)
/variables = qzs e1 e2 fin
/statistics = 1 6 7 13
/method = wilks
/pin = .05.
```

← determines selection method
 ← sets the inclusion criterion

Statistical modeling with ldf is a 2-step process:

- Step 1 – determine the subset of variables that will provide maximum discrimination among the groups
- Step 2 – perform a “regular” ldf with that subset, to determine diffuse vs. concentrated structure, optimal weightings, performance of the model, etc.

Step 1 – determining the subset of variables to be included
 -- ldf uses the same “logic” as forward stepwise multiple regression

```
----- Variables not in the analysis after step 0 -----
Variable Tolerance Minimum Tolerance Signif. of F to enter Wilks' Lambda
QZS      1.0000000 1.0000000 .2751 .95573
E1       1.0000000 1.0000000 .0000 .60293
E2       1.0000000 1.0000000 .0009 .78043
FIN      1.0000000 1.0000000 .0000 .68153
```

← E1, E2 & Fin are all "viable predictors"
 ← best single discriminator, will be the first variable to enter the model

At step 1, E1 was included in the analysis.

	Wilks' Lambda	Degrees of Freedom	Signif.
Wilks' Lambda	.60293	1 2	57.0
Equivalent F	18.7689	2	57.0 .0000

```
----- Variables in the analysis after step 1 -----
Variable Tolerance F to remove
E1       1.0000000 18.769
```

```
----- Variables not in the analysis after step 1 -----
Variable Tolerance Minimum Tolerance Signif. of F to enter Wilks' Lambda
E2       .6960419 .6960419 .2508 .57387
FIN      .6838990 .6838990 .0211 .52533
QZS      .6442234 .6442234 .0002 .44398
```

← e2 won't contribute if added
 ← fin would contribute if added
 ← best variable to add to model

At step 2, QZS was included in the analysis.

	Wilks' Lambda	Degrees of Freedom	Signif.
Wilks' Lambda	.44398	1 3	56.0
Equivalent F	23.0099	3	56.0 .0000

```
----- Variables in the analysis after step 2 -----
Variable Tolerance F to remove
E1       .62254 12.332
QZS      .58874 14.233
```

```
----- Variables not in the analysis after step 2 -----
Variable Tolerance Minimum Tolerance Signif. of F to enter Wilks' Lambda
E2       .5965419 .5965419 .3444 .65387
FIN      .6344439 .6344439 .0211 .33254
```

← e2 won't contribute if added
 ← fin would contribute if added

Remember, this is a “stepwise” process, so the continued contribution of each variable to the model is reassessed after each new variable is added. Any variable that does not continue to contribute to the model is dropped (same as in forward stepwise multiple regression)

SPSS prints a nice summary of the stepwise modeling process

Summary Table							
Step	Action		Vars In	Wilks'		Sig.	Label
	Entered	Removed		Lambda	Sig.		
1	E1		1	.60293	.0000		
2	QZS		2	.44398	.0000		
3	FIN		3	.38141	.0000		

Step 2 -- complete an ldf using just the selected subset of variables

Canonical Discriminant Functions									
Fcn	Pct of Variance	Cum Pct	Canonical Corr	After		Wilks' Lambda	Chisquare	DF	Sig
				Fcn	Lambda				
1*	85.17	85.17	.7352	:	0	.3814	53.978	6	.0000
2*	14.83	100.00	.4123	:	1	.8300	10.433	2	.0054

Standardized Canonical Discriminant Function Coefficients

	FUNCTIONS		Structure Matrix:		
	FUNC 1	FUNC 2		FUNC 1	FUNC 2
QZS	-.83826	.59783	E1	.73932*	.27642
E1	1.23735	-.71023	E2	.53921*	.29542
FIN	.00348	1.12066	FIN	.52162	.84797*
			QZS	-.09948	.41155*

Notice that E2 is not in the model, but appears in the structure matrix !!! The variables in this model have been chosen to be "non-redundant". Having the structure weights for the variables that are and are not included in the model allows you to decide whether or not the variables excluded from this "optimal model" are an important part of a "complete description" of the relationships between group membership and these variables. In this case, E2 would certainly be considered part of complete description.

Example of MaxminF selection

```
discriminant groups = method(1,3)
  /variables = qzs e1 e2 fin
  /statistics = 1 6 7 13
  /method = maxminf
  /pin = .05.
```

SPSS Output

Summary Table									
Step	Action		Vars In	Wilks'		Minimum F	Sig.	Between Groups	
	Entered	Removed		Lambda	Sig.				
1	E1		1	.60293	.0000	6.99654	.0105	2	3
2	FIN		2	.52533	.0000	6.80555	.0023	1	2
3	QZS		3	.38141	.0000	5.22672	.0030	2	3

- E1 as entered in the first step, because it had the best overall F (same as in the Wilks, above).
- With E1 entered, the model does worst at discriminating between groups 2 & 3, so on the next step the intent will be to add that variable that maximally increases the pairwise F for those two groups
- FIN is the variable that most helped to further discriminate between groups 2 & 3
- With E1 and FIN entered, the model does worst at discriminating between groups 1 & 2, so on the next step the intent will be to add that variable that maximally increases this pairwise F.
- QZS is the variable that helped to further discriminate between groups 1 & 2
- With E1, FIN and QZS added to the model the poorest pairwise discrimination is between groups 2 & 3
- But alas, there are no more variables that if added would contribute to the model (E2 doesn't make it into the model)
- Notice that for these data the Wilks and maxminf selected the same set of variables, but that the selection order differed.