Multivariate *t* Test: Hotelling’s *T2*

You have two groups and two or more outcome variables. You want to create a weighted linear combination (a canonical variate) of the outcome variables that maximizes the difference between the two groups. You are using SPSS. I am using the [Howell data set](http://core.ecu.edu/psyc/wuenschk/SPSS/Howell-Variables.htm). The MANOVA procedure requires that you use syntax:

Manova

addsc iq gpa BY gender(1 2)

/discrim raw stan corr alpha(1)

/print signif(mult univ eigen dimenr)

homogeneity(boxm) error(corr)

/noprint param(estim)

/method=unique

/error within+residual

/design .

88 cases accepted.

0 cases rejected because of out-of-range factor values.

0 cases rejected because of missing data.

2 non-empty cells.

1 design will be processed.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Multivariate test for Homogeneity of Dispersion matrices

Boxs M = 12.05156

F WITH (6,30012) DF = 1.92520, P = .073 (Approx.)

Chi-Square with 6 DF = 11.55360, P = .073 (Approx.)

If this is way significant, you have a problem with an assumption.

\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* A n a l y s i s o f V a r i a n c e -- Design 1 \* \* \* \* \*

EFFECT .. gender

Multivariate Tests of Significance (S = 1, M = 1/2, N = 41 )

Test Name Value Exact F Hypoth. DF Error DF Sig. of F

Pillais .11724 3.71876 3.00 84.00 .015

Hotellings .13281 3.71876 3.00 84.00 .015

Wilks .88276 3.71876 3.00 84.00 .015

Roys .11724

Note.. F statistics are exact.

The genders differ significantly on the canonical variate.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Eigenvalues and Canonical Correlations

Root No. Eigenvalue Pct. Cum. Pct. Canon Cor.

1 .13281 100.00000 100.00000 .34241

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

EFFECT .. gender (Cont.) **These are equivalent to multiple t tests.**

Univariate F-tests with (1,86) D. F.

Variable Hypoth. SS Error SS Hypoth. MS Error MS F Sig. of F

addsc 418.21894 13006.86061 418.21894 151.24257 2.76522 .100

iq 127.97045 14541.01818 127.97045 169.08161 .75686 .387

gpa 7.12656 57.43290 7.12656 .66782 10.67131 .002

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

EFFECT .. gender (Cont.)

Raw discriminant function coefficients

Function No.

Variable 1

addsc -.00169

iq -.02337

gpa 1.35210

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Standardized discriminant function coefficients

Function No.

Variable 1

addsc -.02074

iq -.30387

gpa 1.10494

The standardized canonical variate = -.021\*Zaddsc -.304\*Ziq + 1.10\*Zgpa

These are essentially beta weights and suffer from the same problems as do beta weights. The values may be greatly affected by redundancy and their can be suppressor effects.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Correlations between DEPENDENT and canonical variables

Canonical Variable

Variable 1

addsc -.49203

iq .25742

gpa .96658

The are often called “loadings.” For these data scoring high on the canonical variate is strongly associated with GPA, somewhat associated with IQ, and negatively associated with ADDSC.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

In a discriminant function analysis we flip the MANOVA around to predict group membership from the continuous variables.

DISCRIMINANT

/GROUPS=gender(1 2)

/VARIABLES=addsc iq gpa

/ANALYSIS ALL

/PRIORS SIZE

/STATISTICS=TABLE

/CLASSIFY=NONMISSING POOLED.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Eigenvalues** | | | | |
| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
| 1 | .133a | 100.0 | 100.0 | .342 |
| a. First 1 canonical discriminant functions were used in the analysis. | | | | |

The correlation between gender and the canonical variate (usually called a discriminant function in this context) is .342.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Wilks' Lambda** | | | | |
| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
| 1 | .883 | 10.537 | 3 | .015 |

The canonical correlation is significant

|  |  |
| --- | --- |
| **Standardized Canonical Discriminant Function Coefficients** | |
|  | Function |
| 1 |
| addsc | -.021 |
| iq | -.304 |
| gpa | 1.105 |

Same as we got with MANOVA

|  |  |
| --- | --- |
| **Structure Matrix** | |
|  | Function |
| 1 |
| gpa | .967 |
| addsc | -.492 |
| iq | .257 |
| Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions  Variables ordered by absolute size of correlation within function. | |

Same as we got with MANOVA.

|  |  |
| --- | --- |
| **Functions at Group Centroids** | |
| gender | Function |
| 1 |
| 1 | -.279 |
| 2 | .465 |
| Unstandardized canonical discriminant functions evaluated at group means | |

These are group means on the canonical variate. The mean is higher for girls than for boys.

**Classification Statistics**

Here we wish to predict, for each subject, what is the gender.

|  |  |  |  |
| --- | --- | --- | --- |
| **Prior Probabilities for Groups** | | | |
| gender | Prior | Cases Used in Analysis | |
| Unweighted | Weighted |
| 1 | .625 | 55 | 55.000 |
| 2 | .375 | 33 | 33.000 |
| Total | 1.000 | 88 | 88.000 |

There are more boys than girls in the sample, so if we just predicted “boy” for every case we would be correct 62.5% of the time. We might be able to do better if predict not only from the base rates of gender but also from scores on the canonical variate.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Resultsa** | | | | | |
|  |  | gender | Predicted Group Membership | | Total |
|  |  | 1 | 2 |
| Original | Count | 1 | 49 | 6 | 55 |
| 2 | 18 | 15 | 33 |
| % | 1 | 89.1 | 10.9 | 100.0 |
| 2 | 54.5 | 45.5 | 100.0 |
| a. 72.7% of original grouped cases correctly classified. | | | | | |

Well, 72.7% is a bit better.

**Binary Logistic Regression** – to predict group membership but with fewer assumptions than with discriminant function analysis.

LOGISTIC REGRESSION VARIABLES gender

/METHOD=ENTER addsc iq gpa

/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

|  |  |
| --- | --- |
| **Dependent Variable Encoding** | |
| Original Value | Internal Value |
| 1 | 0 |
| 2 | 1 |

**Block 0: Beginning Block This is an intercept only (base rate) model.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Tablea,b** | | | | | |
|  | Observed | | Predicted | | |
|  | gender | | Percentage Correct |
|  | 1 | 2 |
| Step 0 | gender | 1 | 55 | 0 | 100.0 |
| 2 | 33 | 0 | .0 |
| Overall Percentage | |  |  | 62.5 |
| a. Constant is included in the model. | | | | | |
| b. The cut value is .500 | | | | | |

Same as with DFA.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables in the Equation** | | | | | | | |
|  | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step 0 | Constant | -.511 | .220 | 5.382 | 1 | .020 | .600 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables not in the Equation** | | | | | |
|  | | | Score | df | Sig. |
| Step 0 | Variables | addsc | 2.741 | 1 | .098 |
| iq | .768 | 1 | .381 |
| gpa | 9.714 | 1 | .002 |
| Overall Statistics | | 10.317 | 3 | .016 |

**Block 1: Method = Enter Here we include the predictor variables.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Omnibus Tests of Model Coefficients** | | | | |
|  | | Chi-square | df | Sig. |
| Step 1 | Step | 11.233 | 3 | .011 |
| Block | 11.233 | 3 | .011 |
| Model | 11.233 | 3 | .011 |

The model is significant.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Summary** | | | |
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 105.202a | .120 | .163 |
| a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001. | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Tablea** | | | | | |
|  | Observed | | Predicted | | |
|  | gender | | Percentage Correct |
|  | 1 | 2 |
| Step 1 | gender | 1 | 48 | 7 | 87.3 |
| 2 | 17 | 16 | 48.5 |
| Overall Percentage | |  |  | 72.7 |
| a. The cut value is .500 | | | | | |

Classification just as good as with the DFA.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables in the Equation** | | | | | | | |
|  | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step 1a | addsc | -.003 | .028 | .010 | 1 | .920 | .997 |
| iq | -.020 | .024 | .670 | 1 | .413 | .981 |
| gpa | 1.061 | .393 | 7.288 | 1 | .007 | 2.890 |
| Constant | -1.104 | 3.509 | .099 | 1 | .753 | .331 |
| a. Variable(s) entered on step 1: addsc, iq, gpa. | | | | | | | |

Only GPA had a significant unique effect. Exp(B) is an odds ratio. Each increase of 1 point in GPA was associated with the odds of being female nearly tripling.

With the binary logistic regression one can include dichotomous predictors in the model.

[Karl L. Wuensch](http://core.ecu.edu/psyc/wuenschk/klw.htm), March, 2021.