Supplementary material - linear discriminant analysis

DV

Contents

In the results section of the paper, Roy-Bargmann stepdown analyses are used to dive deeper into the nature of the differences identified by the MANOVAs. As some authors (eg. Field, Miles, & Zoe, 2012; Salkind, 2007; Tabachnick & Fidell, 2012) recommend conducting a linear discriminant analysis as a follow-up to MANOVA, we are reporting these results here.

Hard exclusion criteria

As in the paper itself, the following analyses are going to be conducted on a subset of the collected data which contains 203 cases. We will conduct the analyses specified in the `analysis-plan.md` file and follow them up with linear discriminant analyses.

Interpolated activity effect

Again, we will first conduct a MANOVA with the total number of correct answers and total number of intrusive distractors chosen as dependent variables and the type of interpolated activity as the independent variable.

Note

A decision was made not to check the univariate and multivariate outliers at this point. Regarding the univariate outliers - the boxplots point to only one case which could be an outlier. The scatterplots show no point that's obviously different from the rest. As for the multivariate outliers, Tabachnick and Fidell (2012) warn that the Mahalanobis distance can produce false negatives or false positives. Furthermore, deleting a set of outliers and rerunning the analysis can reveal yet another set of outliers — without a clear-cut and absolute criterion, exclusions are somewhat arbitrary. Finally, cases were excluded based on criteria that are more or less substantivelly meaningful in the context of the conducted study. Given the above, no statistical criteria is used for exclusion at this point.

MANOVA

Here's the otuput of `R`'s `manova` function:

```
##
## Type II MANOVA Tests:
##
## Sum of squares and products for error:
##                totalCorrect totalIntrusors
## totalCorrect            993           -417
## totalIntrusors         -417            435
##
## ------------------------------------------
##
## Term: as.factor(activityFactor)
##
## Sum of squares and products for the hypothesis:
##                totalCorrect totalIntrusors
## totalCorrect          125.9          -70.5
## totalIntrusors        -70.5           41.0
##
## Multivariate Tests: as.factor(activityFactor)
##                  Df test stat approx F num Df den Df  Pr(>F)
## Pillai            2     0.126     3.99      4    238 0.00376 **
## Wilks             2     0.875     4.07      4    236 0.00327 **
## Hotelling-Lawley  2     0.142     4.16      4    234 0.00285 **
## Roy               2     0.137     8.13      2    119 0.00049 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As can be seen from the resulting output, Pillai's V indicates that the three groups differ significantly along the linear combination of the two DVs. The other three reported statistics point to the same conclusion. Therefore, we'll proceed with conducting a linear discriminant analyis.

We can look at $1 - \Lambda$ as an extension to the univariate η^2 (Huberty & Olejnik, 2006). In our case, the multivariate η^2 is 0.125, which represents the proportion of total variance associated with the activity type IV. Further, we can calculate the effect size index ξ^2, which is based on Pillai's test statistic, and represents the mean squared canonical correlation (Huberty & Olejnik, 2006):

$$\xi^2 = \frac{U}{r},$$

where r is the number of variates (2, in our case). Therefore, $\xi^2 = 0.063$. Finally, we will calculate Tatsuoka's (1970; according to Huberty & Olejnik, 2006) extension of the ω^2 to the multivariate case.

In this case, $\omega^2_{mult} = 0.109$. The adjusted value of the ξ^2 statistic is $\xi^2_{adj} = 0.047$

Now, let's take a closer look at the nature of our effect, using linear discriminant analysis.

Linear discriminant analysis

We are using the `candisc` function from the eponymous package (Friendly & Fox, 2017) to conduct the LDA.

```
##
## Canonical Discriminant Analysis for as.factor(activityFactor):
##
##     CanRsq Eigenvalue Difference Percent Cumulative
## 1 0.12018     0.1366      0.131   96.13       96.1
## 2 0.00547     0.0055      0.131    3.87      100.0
##
## Test of H0: The canonical correlations in the
## current row and all that follow are zero
##
##   LR test stat approx F numDF denDF Pr(> F)
## 1        0.875     4.07     4   236  0.0033 **
## 2        0.995     0.65     1   119  0.4202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Class means
```

```
##             Can1    Can2
## content    0.501 -0.0104
## general   -0.309 -0.0845
## rereading -0.217  0.0954
```

```
## Raw coefficients
```

```
##                  Can1  Can2
## totalCorrect    0.255 0.368
## totalIntrusors -0.186 0.651
```

```
## Standardized coefficients
```

```
##                  Can1 Can2
## totalCorrect    0.736 1.06
## totalIntrusors -0.355 1.24
```

From the above output we can see that the first variate explains most of the variance. Furthermore, Wilks' lambda values inform us that the groups are separated only on the first variate, so that's the only one we'll interpret. Also, we can see that the variation in the grouping variable is almost exclusively explained by the first variate.

Looking at the structure scores, we can see that both the total number of correct answers and the total

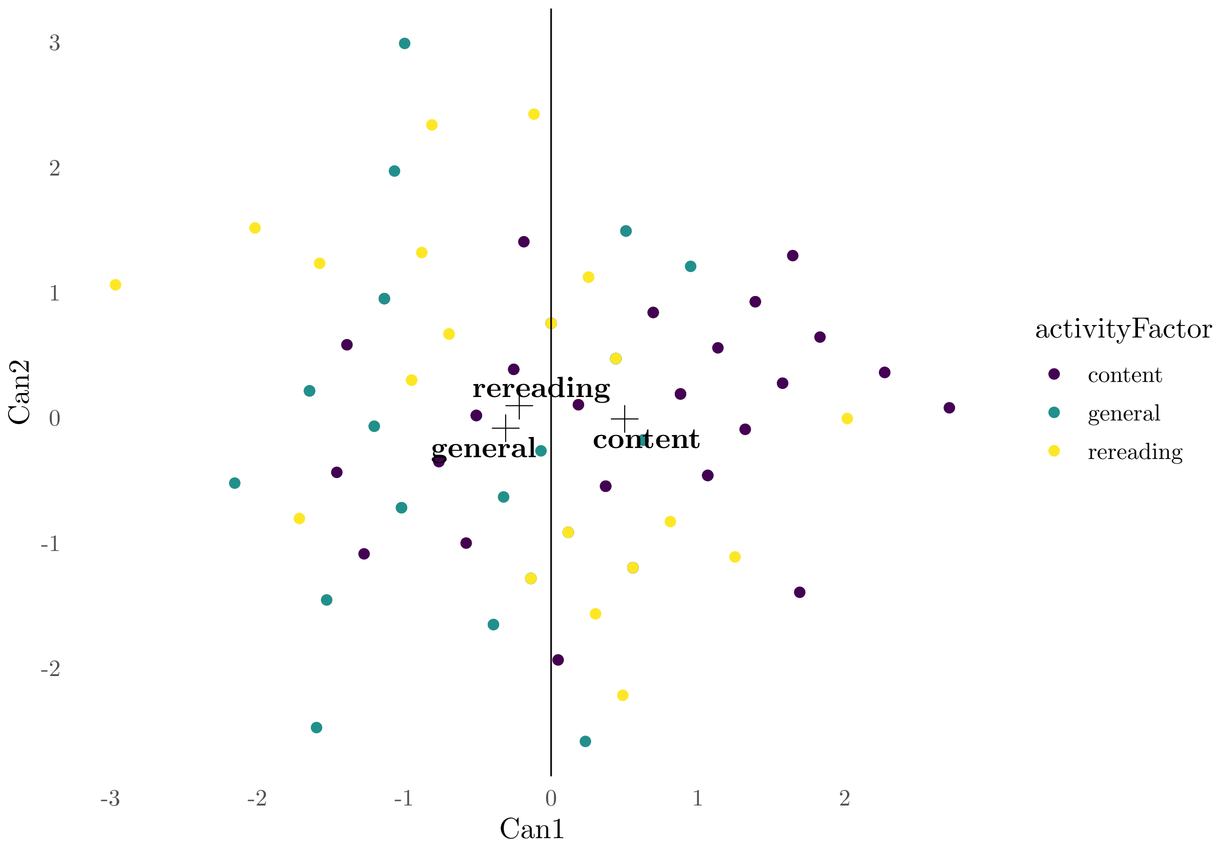

Figure 1: Plot showing the cases' location on the two variates. Group means on the variates are marked by crosses. The vertical line marks the 0 on the first variate.

number of intrusive distractors chosen share a lot of variance with the first variate. The first variate is almost completely defined by the total number of correct answers, but the contribution of the number of chosen intrusors in also considerable. This could be due to the relatively high correlation between those two variables.

To assess the ability of the LDA model to discriminate group membership based on the number of correct answers to the questions and the number of chosen instrusive distractors, we'll re-train the model and evaluate it's error rate using the leave-one-out cross-validation technique.

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction  content general rereading
##   content        27      16        17
##   general        11      19        13
##   rereading       4       5        10
##
## Overall Statistics
##
##                Accuracy : 0.459
##                  95% CI : (0.368, 0.552)
##     No Information Rate : 0.344
```

```
##     P-Value [Acc > NIR] : 0.00572
##
##                   Kappa : 0.185
##  Mcnemar's Test P-Value : 0.00577
##
## Statistics by Class:
##
##                      Class: content Class: general Class: rereading
## Sensitivity                   0.643          0.475            0.250
## Specificity                   0.588          0.707            0.890
## Pos Pred Value                0.450          0.442            0.526
## Neg Pred Value                0.758          0.734            0.709
## Precision                     0.450          0.442            0.526
## Recall                        0.643          0.475            0.250
## F1                            0.529          0.458            0.339
## Prevalence                    0.344          0.328            0.328
## Detection Rate                0.221          0.156            0.082
## Detection Prevalence          0.492          0.352            0.156
## Balanced Accuracy             0.615          0.591            0.570
```

As can be seen from the table, the total LOOCV accuracy is 0.459, which is significantly above the no information rate (which is taken to be the largest class percentage in the data). According to the Landis & Koch (1977; as reported in Salkind, 2007) guidlines, this represents only a slight agreement between the predicted and actual classes. Next, we'll drill into the individual predictors to see which are useful for discriminating between different groups.

Evaluating individual predictors

Tabachnick and Fidell (2012) describe the process of sequential discriminant analysis, where predictors are entered one-by-one, and the improvement in classification accuracy is monitored. Therefore, we'll fit an LDA model containing only the number of correct answers as a predictor. Then, we will compare this model's LOOCV accuracy to that of the full model (reported at the end of the previous section).

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction  content general rereading
##   content        29      15        16
##   general        11      22        18
##   rereading       2       3         6
##
## Overall Statistics
##
##                Accuracy : 0.467
##                  95% CI : (0.376, 0.56)
```

```
##     No Information Rate : 0.344
##     P-Value [Acc > NIR] : 0.00334
##
##                   Kappa : 0.198
##  Mcnemar's Test P-Value : 5.87e-05
##
## Statistics by Class:
##
##                      Class: content Class: general Class: rereading
## Sensitivity                   0.690          0.550           0.1500
## Specificity                   0.613          0.646           0.9390
## Pos Pred Value                0.483          0.431           0.5455
## Neg Pred Value                0.790          0.746           0.6937
## Precision                     0.483          0.431           0.5455
## Recall                        0.690          0.550           0.1500
## F1                            0.569          0.484           0.2353
## Prevalence                    0.344          0.328           0.3279
## Detection Rate                0.238          0.180           0.0492
## Detection Prevalence          0.492          0.418           0.0902
## Balanced Accuracy             0.651          0.598           0.5445
```

As can be seen from the second confusion matrix, the accuracy of this model is actually somewhat higher than in the full model, as is Cohen's κ. Importantly, we notice that adding the total number of intrusors to the model doesn't significantly increase the accuracy of the model (the 95% confidence intervals for the accuracies of the two models completely overlap).

Multivariate contrasts

We've planned to contrast the two test groups with the rereading group, and the two test groups with each other. That's what we'll do here.

```
##           test vs rereading content vs general
## content                   1                  1
## general                   1                 -1
## rereading                -2                  0
```

```
##                                 totalCorrect totalIntrusors
## (Intercept)                           11.379          4.194
## activityFactortest vs rereading        0.252         -0.216
## activityFactorcontent vs general       1.155         -0.597
```

Now that we've set up the model, let's run the contrasts. The first contrast is between the two test groups (content and general knowledge) and the rereading group.

```
##
## Sum of squares and products for the hypothesis:
##                totalCorrect totalIntrusors
```

```
## totalCorrect            15.3          -13.1
## totalIntrusors         -13.1           11.3
##
## Sum of squares and products for error:
##                totalCorrect totalIntrusors
## totalCorrect             993           -417
## totalIntrusors          -417            435
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df Pr(>F)
## Pillai            1     0.026     1.57      2    118   0.21
## Wilks             1     0.974     1.57      2    118   0.21
## Hotelling-Lawley  1     0.027     1.57      2    118   0.21
## Roy               1     0.027     1.57      2    118   0.21
```

As can be seen from the test statistics, no significant difference is found between the two test groups and the rereading group. Next, we'll look at the contrast between the content test group and the general knowledge test group.

```
##
## Sum of squares and products for the hypothesis:
##                totalCorrect totalIntrusors
## totalCorrect           109.4          -56.5
## totalIntrusors         -56.5           29.2
##
## Sum of squares and products for error:
##                totalCorrect totalIntrusors
## totalCorrect             993           -417
## totalIntrusors          -417            435
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df Pr(>F)
## Pillai            1     0.102     6.73      2    118 0.0017 **
## Wilks             1     0.898     6.73      2    118 0.0017 **
## Hotelling-Lawley  1     0.114     6.73      2    118 0.0017 **
## Roy               1     0.114     6.73      2    118 0.0017 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This contrast is statistically significant, indicating that the two groups differ on the linear combination of the number of correct answers and number of intrusive distractors chosen. We'll calculate the same effect size indices as for the omnibus model.

The multivariate η^2 is 0.102. The effect size index ξ^2 is 0.051. Finally, we will calculate Tatsuoka's (1970; according to Huberty & Olejnik, 2006) extension of the ω^2 to the multivariate case. In this case, $\omega^2_{mult} = 0.087$. The adjusted value of the ξ^2 statistic is $\xi^2_{adj} = 0.035$

Contrast LDA

Again, to further investigate the nature of the difference between the content and general knowledge test group, we'll conduct a linear discriminant analysis to try and find the variate that best discriminates these two groups.

```
##
## Type II MANOVA Tests: Pillai test statistic
##                Df test stat approx F num Df den Df Pr(>F)
## activityFactor  1     0.148     6.85      2     79 0.0018 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Canonical Discriminant Analysis for activityFactor:
##
##   CanRsq Eigenvalue Difference Percent Cumulative
## 1 0.1479     0.1735                100        100
##
## Class means:
##
## [1]  0.4015 -0.4216
##
##  std coefficients:
##   totalCorrect totalIntrusors
##         0.7210        -0.3722
```

```
##                   Can1
## totalCorrect    0.964
## totalIntrusors -0.851
```

Again, we see that both predictors are highly correlated with the discriminant function, albeit with different signs. Let's look at the LOOCV prediction accuracy.

```
## Confusion Matrix and Statistics
##
##            Reference
## Prediction content general
##    content      28      16
##    general      14      24
##
##                Accuracy : 0.634
##                  95% CI : (0.52, 0.738)
##     No Information Rate : 0.512
##     P-Value [Acc > NIR] : 0.0175
##
##                   Kappa : 0.267
```

```
##  Mcnemar's Test P-Value : 0.8551
##
##             Sensitivity : 0.667
##             Specificity : 0.600
##          Pos Pred Value : 0.636
##          Neg Pred Value : 0.632
##              Prevalence : 0.512
##          Detection Rate : 0.341
##    Detection Prevalence : 0.537
##       Balanced Accuracy : 0.633
##
##        'Positive' Class : content
##
```

With both predictors, the prediction accuracy is significantlly above the no information rate.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction content general
##    content      29      15
##    general      13      25
##
##                Accuracy : 0.659
##                  95% CI : (0.546, 0.76)
##     No Information Rate : 0.512
##     P-Value [Acc > NIR] : 0.00523
##
##                   Kappa : 0.316
##  Mcnemar's Test P-Value : 0.85011
##
##             Sensitivity : 0.690
##             Specificity : 0.625
##          Pos Pred Value : 0.659
##          Neg Pred Value : 0.658
##              Prevalence : 0.512
##          Detection Rate : 0.354
##    Detection Prevalence : 0.537
##       Balanced Accuracy : 0.658
##
##        'Positive' Class : content
##
```

Furthermore, the prediction accuracy doesn't drop significantly when we omit the total number of intrusors.

References

Field, A., Miles, J., & Zoe, F. (2012). *Discovering Statistics Using R*. Thousand Oaks, CA: SAGE Publications Ltd.

Friendly, M., & Fox, J. (2017). *Candisc: Visualizing Generalized Canonical Discriminant and Canonical Correlation Analysis*.

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* (Vol. 498). John Wiley & Sons.

Salkind, N. J. (2007). *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: Sage.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics*. Pearson.