

H1-2

DV

Contents

Hard exclusion criteria	2
Descriptive statistics	2
Assumption checks	6
Note	12
MANOVA	12
Linear discriminant analysis	13
Evaluating individual predictors	16

```
library(knitr)
library(kableExtra)
opts_chunk$set(dpi = 600, dev = 'pdf', echo = F)
options(digits = 5)

## here() starts at /home/denis/Documents/Poso/faks/istraživanja/inter-
testing-feedback-2018/analyses

## -- Attaching packages -----
----- tidyverse 1.2.1 --

## v ggplot2 3.1.0      v purrr 0.3.1
## v tibble 2.0.1       v dplyr 0.8.0.1
## v tidyr 0.8.3        v stringr 1.4.0
## v readr 1.3.1        v forcats 0.4.0

## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::filter()    masks stats::filter()
## x dplyr::group_rows() masks kableExtra::group_rows()
## x dplyr::lag()       masks stats::lag()

## [conflicted] Will prefer dplyr::filter over any other package

## Parsed with column specification:
## cols(
##   .default = col_double(),
##   when = col_datetime(format = ""),
##   giveFeedback = col_logical(),
```

```
## condition = col_character(),
## kolikoProcitaoText1 = col_character(),
## kolikoProcitaoText2 = col_character(),
## kolikoProcitaoText3 = col_character(),
## readingDeficits = col_character(),
## which = col_character(),
## readingDifficultiesThisExp = col_character(),
## activityFactor = col_character()
## )

## See spec(...) for full column specifications.

## sROC 0.1-2 loaded

## Loading required package: car

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:purrr':
##
##     some

## Loading required package: viridisLite

## Loading required package: sgeostat

## [conflicted] Will prefer dplyr::select over any other package
```

Hard exclusion criteria

The following analyses are going to be conducted on a subset of the collected data which contains 203 cases. First, we will take a look at the data going into this analysis. Then, we will check whether the assumptions for conducting a MANOVA are satisfied. Finally, we will conduct the analyses specified in the `analysis-plan.md` file.

Descriptive statistics

This analysis is going to be run on a subset participants who were in no-feedback conditions. This includes the rereading group, and the two no-feedback test groups.

This leaves us with 122 cases.

```
## Observations: 122
## Variables: 3
## $ activityFactor <chr> "content", "general", "rereading", "content", "...
## $ totalCorrect <dbl> 15, 12, 6, 12, 14, 14, 9, 5, 14, 13, 14, 8, 12,...
## $ totalIntrusors <dbl> 3, 3, 5, 3, 3, 4, 4, 4, 4, 3, 4, 6, 3, 3, 6, 8,...
```

First of all, it's important to mention that the group sizes are imbalanced, but the difference is really small:

```
## # A tibble: 3 x 2
##   activityFactor      n
##   <chr>           <int>
## 1 content         42
## 2 general         40
## 3 rereading       40
```

Here are the descriptives for the whole subset.

```
##           vars    n  mean    sd median trimmed  mad min max range skew
## totalCorrect    1 122 11.40 3.04     11   11.39 2.97   4  19   15 0.04
## totalIntrusors  2 122  4.18 1.98      4    4.10 1.48   0  10   10 0.38
##           kurtosis    se
## totalCorrect    -0.41 0.28
## totalIntrusors  -0.17 0.18
```

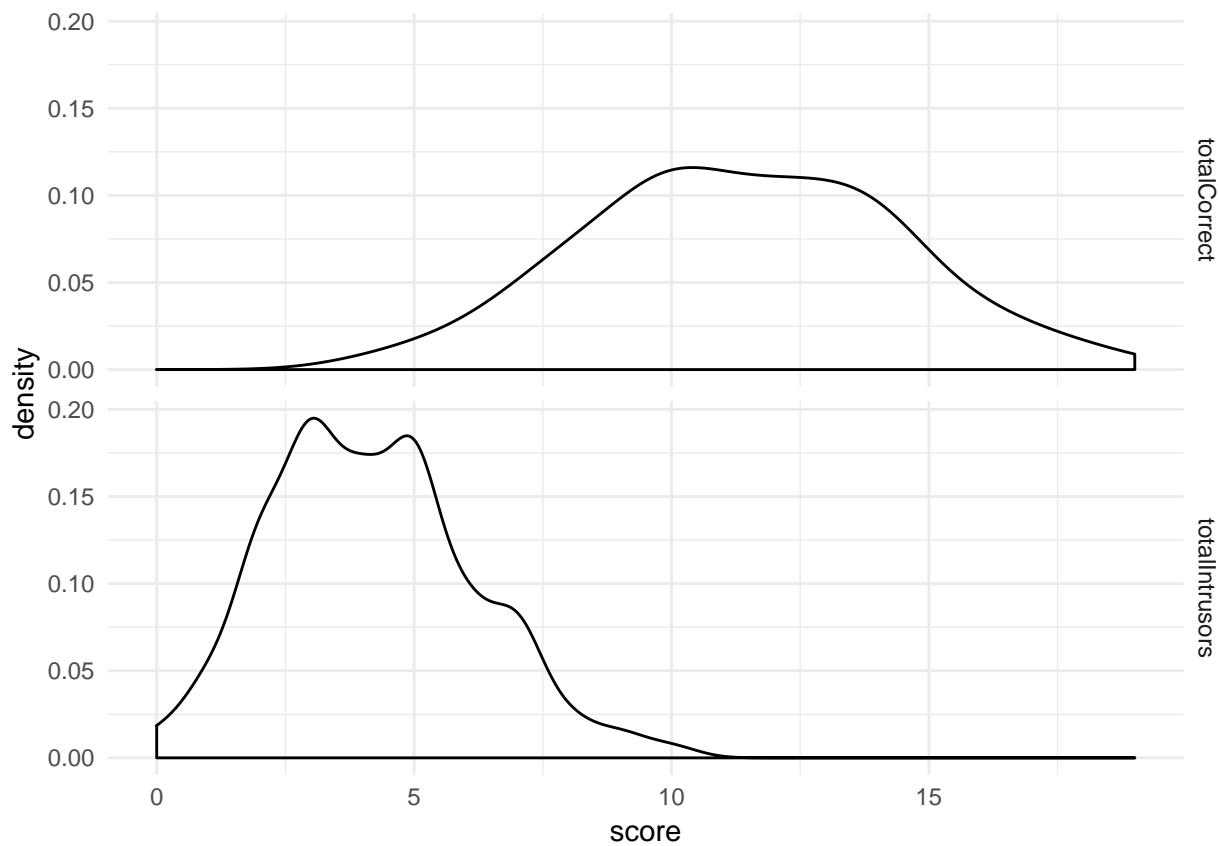
And the descriptives by group:

```
##   condition  n  mean    sd median trimmed  mad min max range    skew
## 1   content 42 12.786 3.0165     12  12.765 2.9652   7  19   12 0.039247
## 2   general 40 10.475 2.8374     10  10.531 2.9652   5  16   11 -
0.053301
## 3 rereading 40 10.875 2.8028     11  10.938 2.9652   4  17   13 -
0.140828
##   kurtosis    se
## 1 -0.77457 0.46546
## 2 -0.98633 0.44863
## 3 -0.25332 0.44316

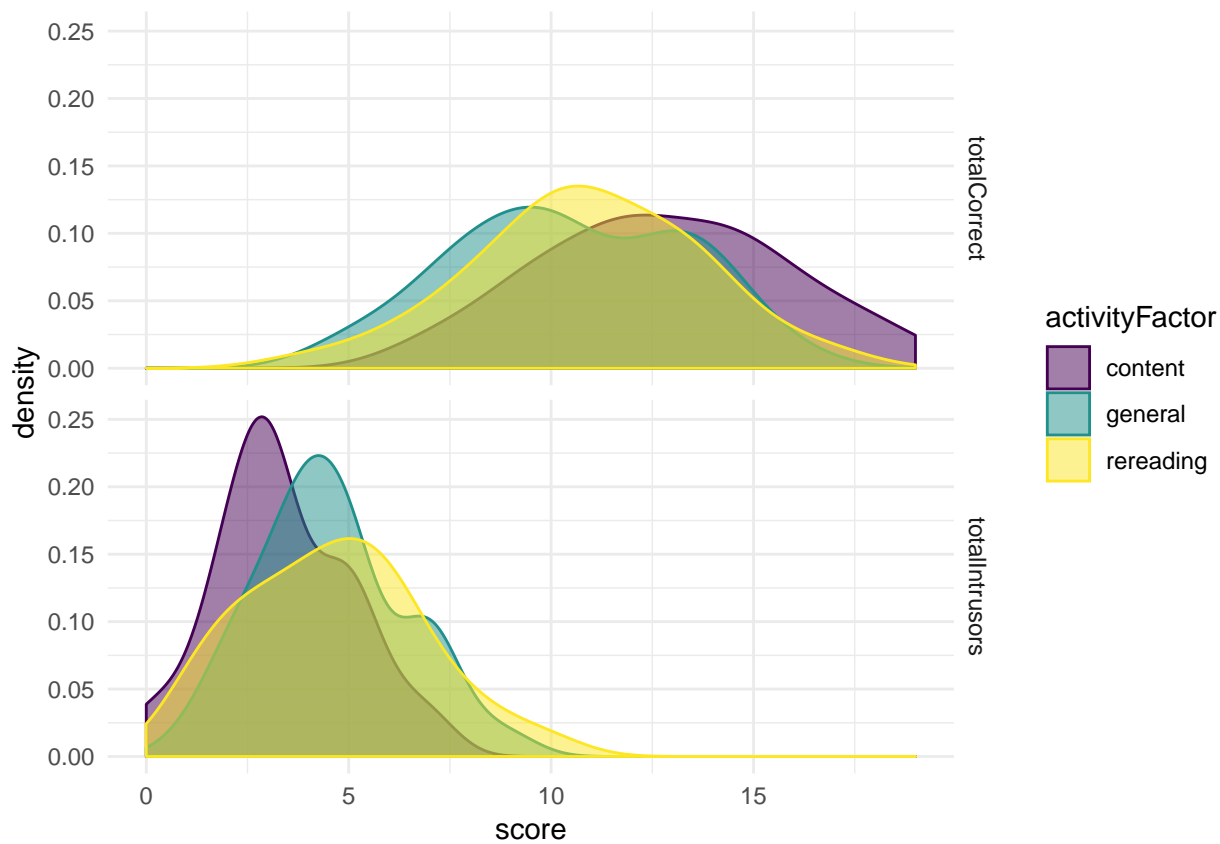
##   condition  n  mean    sd median trimmed  mad min max range    skew
## 1   content 42 3.381 1.6668      3  3.3529 1.4826   0   7    7 0.20311
## 2   general 40 4.575 1.8242      4  4.5312 1.4826   1   9    8 0.32757
## 3 rereading 40 4.625 2.2152      5  4.5312 2.2239   1  10    9 0.27232
##   kurtosis    se
```

```
## 1 -0.38484 0.25719
## 2 -0.48422 0.28843
## 3 -0.53693 0.35025
```

Now, we'll plot the DV distributions, both on the whole subset, and per group.



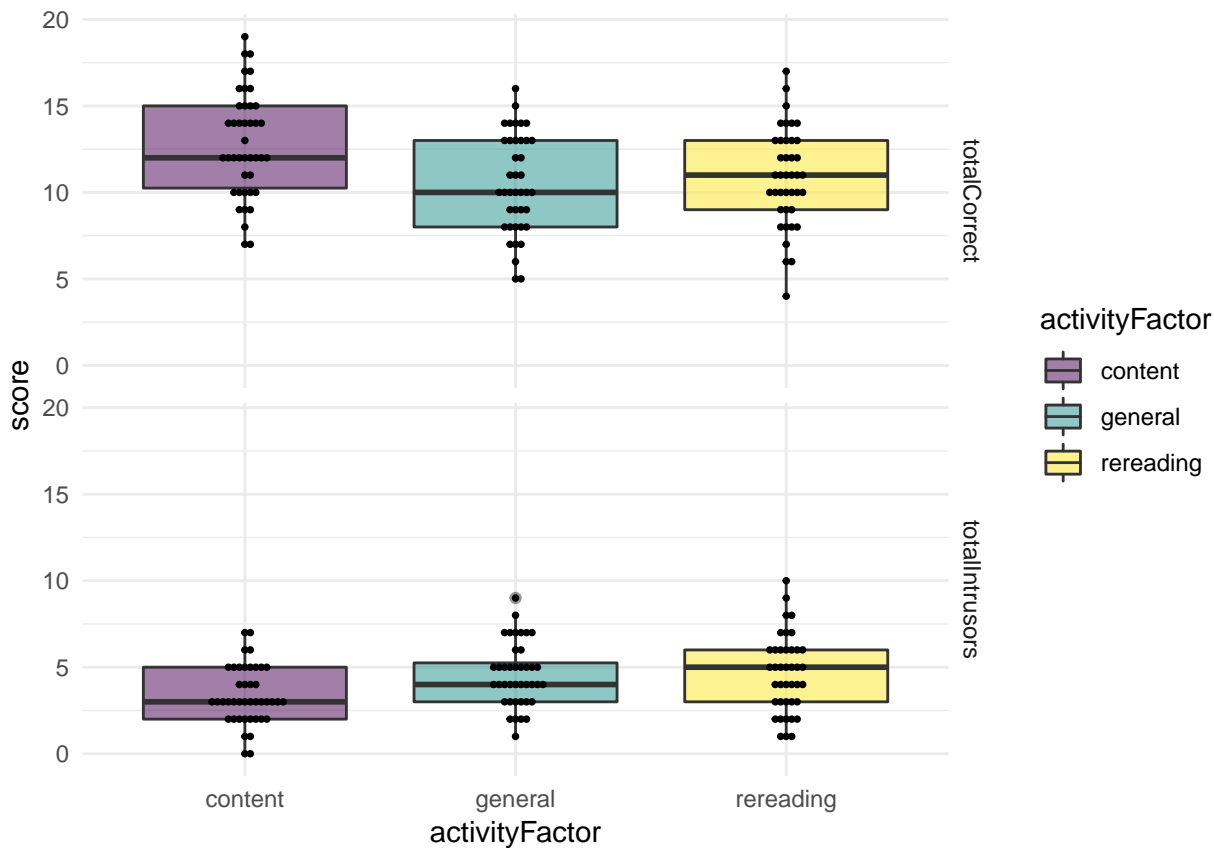
The graph shows that there could be significant deviations from normality. Let's look at the distributions in each group.



The distributions seem to be fairly similar on both dependent variables, and look a bit more normal than on the whole sample.

Next, here are the boxplots for the three groups, and for both DVs.

`stat_bindot()` using `bins = 30`. Pick better value with `binwidth`.



Looks like there could be an outlier on the number of total intrusions in the general knowledge test condition.

Assumption checks

Next, let's look at the correlation between the DVs. First, let's look at the correlation in the whole sub-sample.

```
## Call:psych::corr.test(x = .)
## Correlation matrix
##           totalCorrect totalIntrusions
## totalCorrect           1.00          -0.67
## totalIntrusions        -0.67           1.00
## Sample Size
## [1] 122
## Probability values (Entries above the diagonal are adjusted for multiple test)
##           totalCorrect totalIntrusions
## totalCorrect           0              0
## totalIntrusions        0              0
##
## To see confidence intervals of the correlations, print with the short=FALSE
```

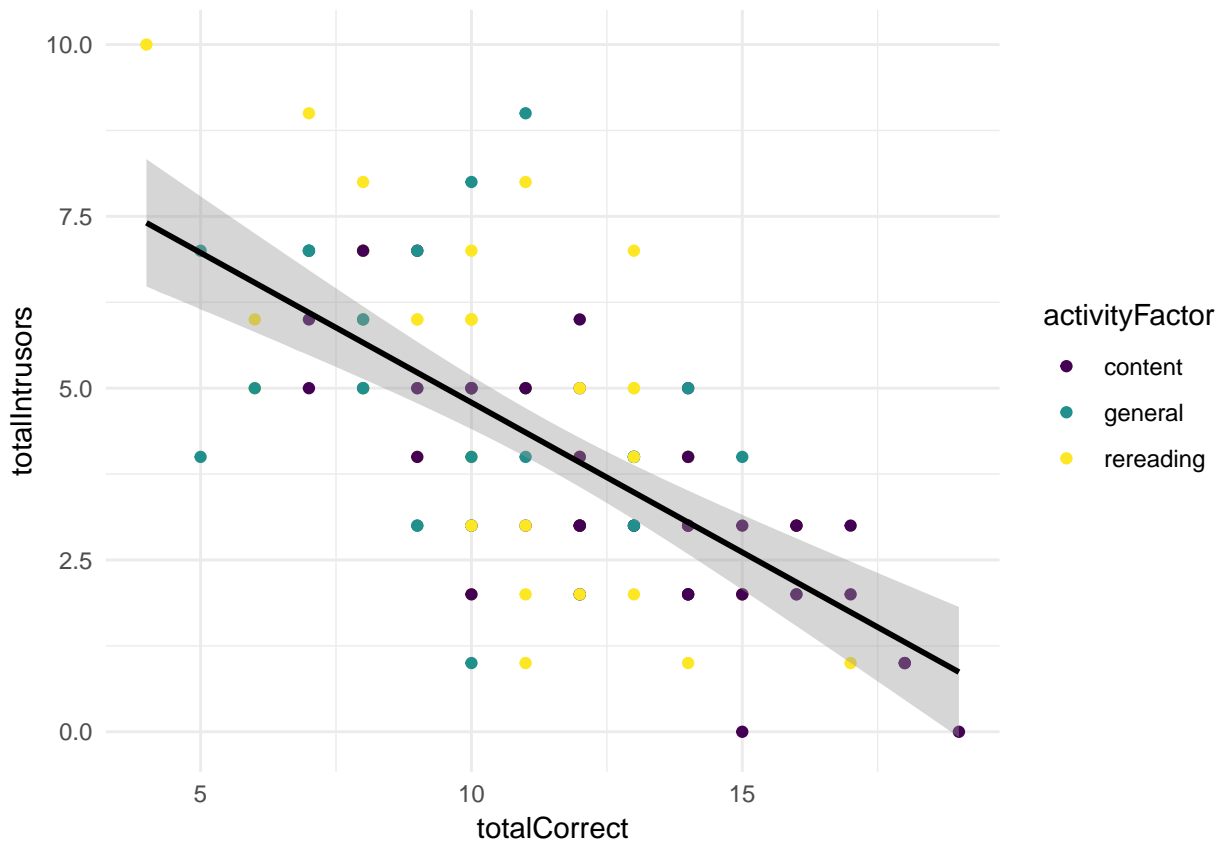
We found a statistically significant correlation of -0.66831. Next, let's look at the correlation

in each group.

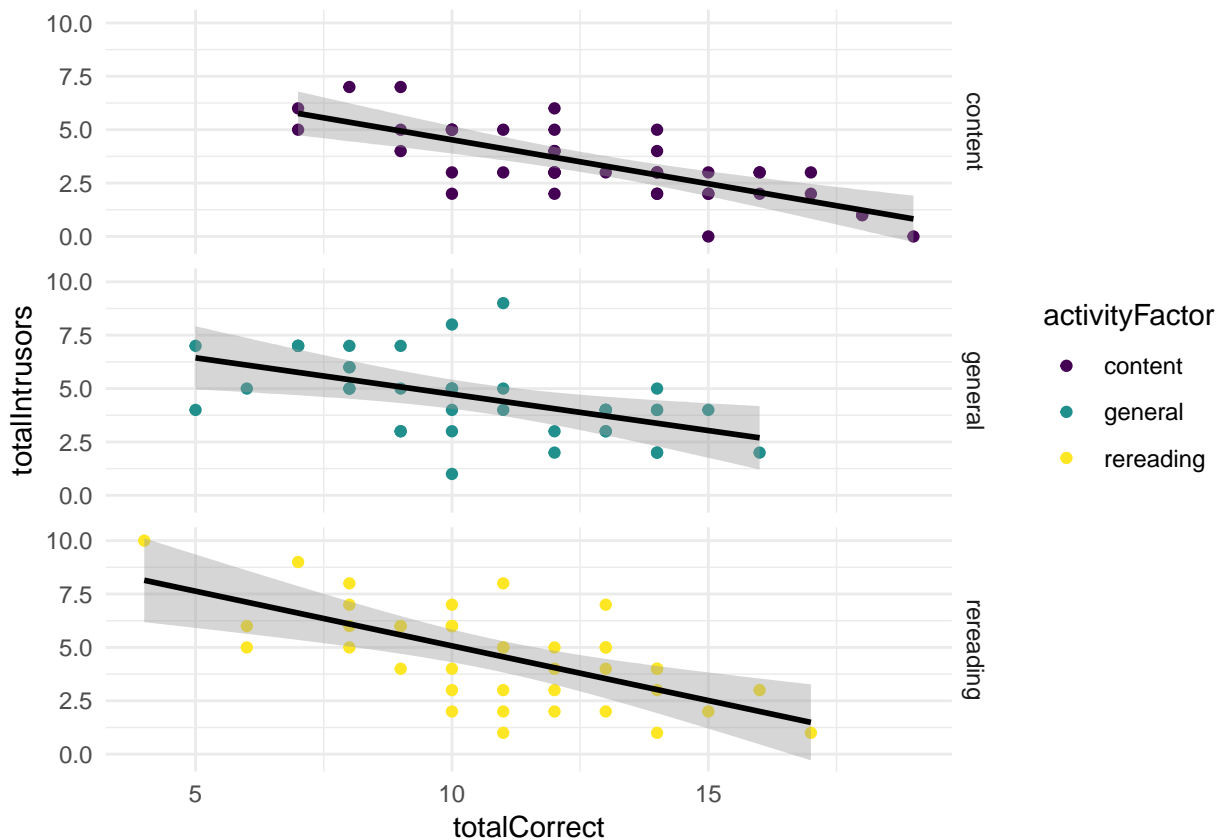
```
## datHardNofeed$activityFactor: content
## Call:FUN(x = data[x, , drop = FALSE])
## Correlation matrix
##               totalCorrect totalIntrusors
## totalCorrect           1.00          -0.74
## totalIntrusors        -0.74           1.00
## Sample Size
## [1] 42
## Probability values (Entries above the diagonal are adjusted for multiple test)
##               totalCorrect totalIntrusors
## totalCorrect           0           0
## totalIntrusors         0           0
##
## To see confidence intervals of the correlations, print with the short=FALSE
## -----
## datHardNofeed$activityFactor: general
## Call:FUN(x = data[x, , drop = FALSE])
## Correlation matrix
##               totalCorrect totalIntrusors
## totalCorrect           1.00          -0.53
## totalIntrusors        -0.53           1.00
## Sample Size
## [1] 40
## Probability values (Entries above the diagonal are adjusted for multiple test)
##               totalCorrect totalIntrusors
## totalCorrect           0           0
## totalIntrusors         0           0
##
## To see confidence intervals of the correlations, print with the short=FALSE
## -----
## datHardNofeed$activityFactor: rereading
## Call:FUN(x = data[x, , drop = FALSE])
## Correlation matrix
##               totalCorrect totalIntrusors
## totalCorrect           1.00          -0.65
## totalIntrusors        -0.65           1.00
## Sample Size
## [1] 40
```

```
## Probability values (Entries above the diagonal are adjusted for multiple test
##
##          totalCorrect totalIntrusors
## totalCorrect          0          0
## totalIntrusors        0          0
##
## To see confidence intervals of the correlations, print with the short=FALSE
```

The correlations in all three groups are fairly similar and statistically significant. To check whether the relationship between the DVs could be described as linear, we'll plot the scatter-plots for the whole sample and for each group.



As can be seen from the plot, the relationship seems to be pretty linear, and the 99% confidence interval of the regression slopes is pretty small. Now, let's take a look at the scatterplots for each group separately.



Again, we can see that the relationships are linear, as well as similar. Next, we will check the multivariate normality assumption using the Henze-Zirkler test.

```
## datHardNoFeed$activityFactor: content
## $multivariateNormality
##           Test           HZ p value MVN
## 1 Henze-Zirkler 0.23726 0.95885 YES
##
## $univariateNormality
##           Test           Variable Statistic    p value Normality
## 1 Shapiro-Wilk totalCorrect      0.9751      0.4830      YES
## 2 Shapiro-Wilk totalIntrusors    0.9465      0.0483      NO
##
## $Descriptives
##           n    Mean Std.Dev Median Min Max 25th 75th      Skew
## totalCorrect 42 12.786 3.0165     12   7  19 10.25  15 0.039247
## totalIntrusors 42 3.381 1.6668      3   0   7  2.00   5 0.203106
##
##           Kurtosis
## totalCorrect -0.77457
## totalIntrusors -0.38484
##
## -----
```

```
## datHardNoFeed$activityFactor: general
## $multivariateNormality
##           Test           HZ p value MVN
## 1 Henze-Zirkler 0.59687 0.23328 YES
##
## $univariateNormality
##           Test           Variable Statistic    p value Normality
## 1 Shapiro-Wilk totalCorrect      0.9646      0.2398      YES
## 2 Shapiro-Wilk totalIntrusors    0.9559      0.1211      YES
##
## $Descriptives
##           n    Mean Std.Dev Median Min Max 25th 75th      Skew
## totalCorrect 40 10.475 2.8374     10   5 16   8 13.00 -0.053301
## totalIntrusors 40 4.575 1.8242     4   1 9   3 5.25 0.327568
##           Kurtosis
## totalCorrect -0.98633
## totalIntrusors -0.48422
##
## -----
## datHardNoFeed$activityFactor: rereading
## $multivariateNormality
##           Test           HZ p value MVN
## 1 Henze-Zirkler 0.2667 0.91818 YES
##
## $univariateNormality
##           Test           Variable Statistic    p value Normality
## 1 Shapiro-Wilk totalCorrect      0.9858      0.8872      YES
## 2 Shapiro-Wilk totalIntrusors    0.9665      0.2770      YES
##
## $Descriptives
##           n    Mean Std.Dev Median Min Max 25th 75th      Skew
## totalCorrect 40 10.875 2.8028     11   4 17   9 13 -0.14083
## totalIntrusors 40 4.625 2.2152     5   1 10   3 6 0.27232
##           Kurtosis
## totalCorrect -0.25332
## totalIntrusors -0.53693
```

The result of Henze-Zirkler's multivariate normality test shows that a statistically significant departure from multivariate normality was not detected. Hence, we will assume that there really is no departure. The points on the Chi-Square Q-Q plot follow the straight line fairly

well, which is also indicative of a normal distribution.

However, the Shapiro-Wilk test for univariate normality indicates a departure in the distribution of the total number of intrusive distractors chosen in one group. It is interesting to notice that the value of the W statistic is close to the maximum of 1, which indicates a close fit to the normal distribution (Salkind, 2007). Also, the test is only marginally significant at the conventional .05 level. Therefore, we will assume that the data in each group is normally distributed.

Now, we'll take a look at the homogeneity of covariance matrices assumption. First, let's take a look at the matrices themselves.

```
## datHardNofeed$activityFactor: content
##               totalCorrect totalIntrusors
## totalCorrect      9.0993      -3.7456
## totalIntrusors    -3.7456       2.7782
## -----
## datHardNofeed$activityFactor: general
##               totalCorrect totalIntrusors
## totalCorrect      8.0506      -2.7417
## totalIntrusors    -2.7417       3.3276
## -----
## datHardNofeed$activityFactor: rereading
##               totalCorrect totalIntrusors
## totalCorrect      7.8558      -4.0224
## totalIntrusors    -4.0224       4.9071
```

The covariance matrices look quite similar. The ratio of largest to smallest variance of the number of correct answers is 1.1583. The same ratio for the number of chosen intrusors is 1.76629. Finally, the ratio of the largest to the smallest covariance between the two DVs is 0.68159. As we can see, the variance ratios are pretty close to one. Still, let's test them with Box's M test.

```
##
## Box's M-test for Homogeneity of Covariance Matrices
##
## data:  datHardNofeed %>% select(., -activityFactor)
## Chi-Sq (approx.) = 9.34, df = 6, p-value = 0.16
```

Box's M returns a non-significant p-value, indicating that we cannot reject the null hypothesis. Field, Miles, and Zoe (2012) and Raykov and Marcoulides (2008) warn that Box's M is extremely sensitive, so we'd expect to find a difference if there really was one. On the other hand, Field et al. (2012) warn that the test can return a non-significant p-value when the assumption

of multivariate normality is not tenable. However, given the results of the Henze-Zirkler multivariate normality test, we suspect that this is not the case. Since all assumptions seem to hold, we can proceed with the planned MANOVA.

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 substantively meaningful in the context of the conducted study. Given the above, no statistical criteria is used for exclusion at this point.

MANOVA

Now that it seems that all the assumptions of a MANOVA are satisfied, let's run the analysis.

```
##                               Df Pillai approx F num Df den Df Pr(>F)
## as.factor(activityFactor)    2  0.126      3.99      4    238 0.0038 **
## Residuals                    119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##                               Df Wilks approx F num Df den Df Pr(>F)
## as.factor(activityFactor)    2  0.875      4.07      4    236 0.0033 **
## Residuals                    119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##                               Df Hotelling-Lawley approx F num Df den Df
## as.factor(activityFactor)    2      0.142      4.16      4    234
## Residuals                    119
##                               Pr(>F)
## as.factor(activityFactor) 0.0028 **
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##                               Df   Roy approx F num Df den Df  Pr(>F)
## as.factor(activityFactor)    2  0.137      8.13      2    119 0.00049 ***
```

```
## Residuals                119
## ---
## 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 analysis.

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, $\xi^2 = 0.06283$. Finally, we will calculate Tatsuoka's (1970; according to Huberty & Olejnik, 2006) extension of the ω^2 to the multivariate case. In this case, $\omega_{mult}^2 = 0.10949$. The adjusted value of the ξ^2 statistic is $\xi_{adj}^2 = 0.04708$

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

Linear discriminant analysis

```
##
## Canonical Discriminant Analysis for as.factor(activityFactor):
##
##      CanRsqr Eigenvalue Difference Percent Cumulative
## 1 0.120185    0.136602      0.1311    96.13      96.13
## 2 0.005469    0.005499      0.1311     3.87     100.00
##
## Class means:
##
##           Can1      Can2
## content      0.5011 -0.01039
## general     -0.3094 -0.08451
## rereading   -0.2168  0.09541
##
## std coefficients:
##           Can1      Can2
## totalCorrect    0.7362  1.065
## totalIntrusors -0.3551  1.245
```

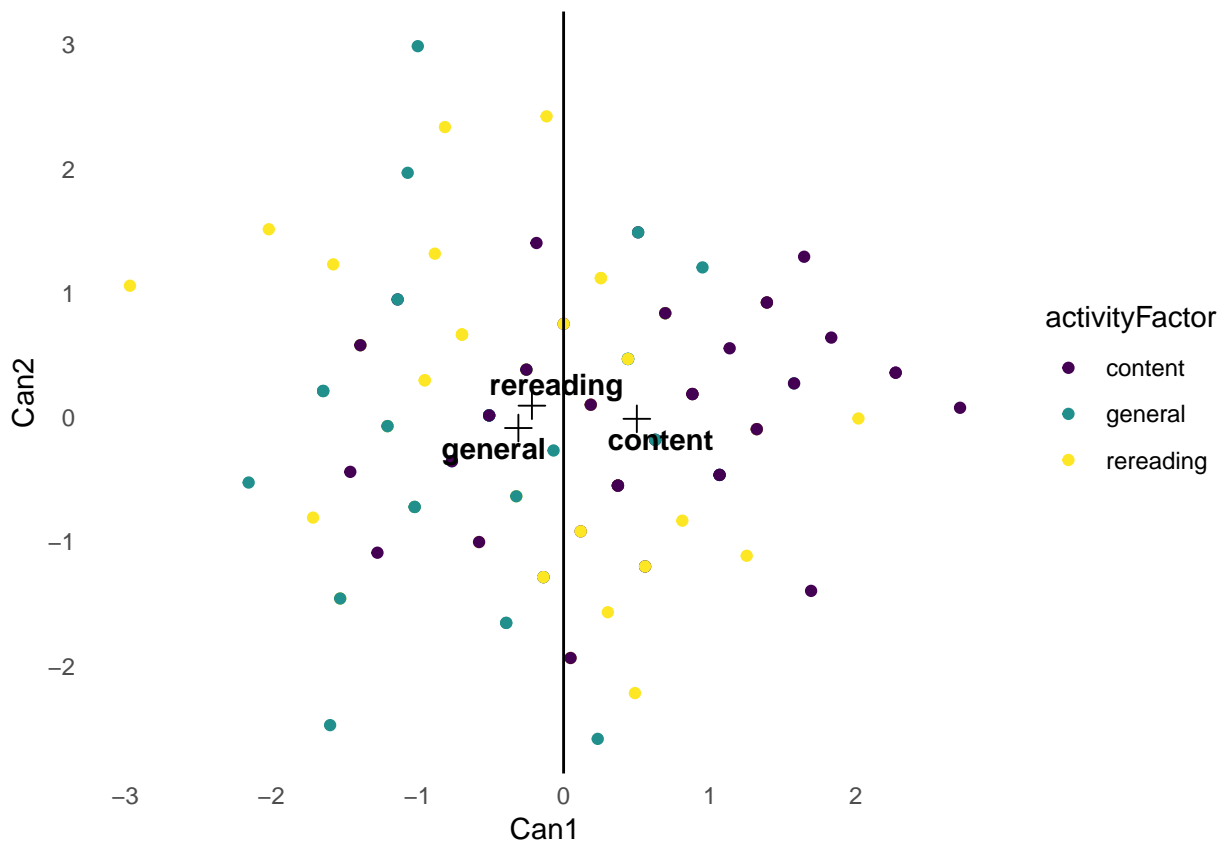
```
##
## ===== print =====
##
## Canonical Discriminant Analysis for as.factor(activityFactor):
##
##      CanRsq Eigenvalue Difference Percent Cumulative
## 1 0.12018      0.1366      0.131   96.13      96.1
## 2 0.00547      0.0055      0.131    3.87     100.0
##
## Test of H0: The canonical correlations 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
```

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.

```
##              Can1      Can2
## totalCorrect  0.25479 0.36845
## totalIntrusors -0.18574 0.65094

##              Can1      Can2
## totalCorrect  0.96582 0.25920
## totalIntrusors -0.83829 0.54523
```

Looking at the structure scores, we can see that both the total number of correct answers and the total 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 is 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 intrusive 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
## 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 0.45902, 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](#)) guidelines, 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
## 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 .

{

References

- Field, A., Miles, J., & Zoe, F. (2012). *Discovering Statistics Using R*. Thousand Oaks, CA: SAGE Publications Ltd.
- Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* (Vol. 498). John Wiley & Sons.
- Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. Routledge.
- Salkind, N. J. (2007). *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: Sage.
- Tabachnick, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics*. Pearson.

}