Multivariate statistics: Assignment 1

Team 27: **Raïsa Carmen** *s0204278* **Marco Chi Chung Fong** *stnumber*
Wenting Jiang *s0204278* **Chin Wei Ma** *r0877202*

1 Task 1

1.1 CFA to construct a measurement model for the Attitude items

There are 9 attitude items that are scored on a five-point Likert scale. To conduct CFA on the attitude items using the covariance matrix, we first center the data.

1.1.1 A simple 3-factor model

We first conduct a simple confirmatory factor analysis, assuming each item only has a loading on the concept it aims to measure (organic, packaging, and cruelty free). We will assume the the three latent variables are correlated and the factor loading of the first indicator of each latent variable is fixed to 1. We fit the model on standardized data. The first columns in Table 3 shows several performance measures for the model. It shows that the currently proposed 3-factor model is not a good fit. The chi-squared goodness of fit tests indicate that the constraints imposed by the model are not supported ($p < 0.001$). The cutoff for a good model for CFI and TLI (cutoff > 0.95) and for RMSEA and SRMR (cutoff < 0.08) are also not satisfied. On the other hand, composite reliability measures the reliability of the factor scores. We can see that the composite reliability values are high (Table 1), therefore, the factors are measured in a reliable way. Figure 1 in the appendix shows a graphical representation of the model, including all loadings, correlations and variances.

In the standardized solution, the standardized loadings represent correlations between a variable and a factor (Table 1). All standardized loadings are above 0.7. Therefore, the squared loadings are higher than 0.5. This reflects a sufficient reliability of the indicator variables. Since all the standardized loadings are positive and significant, there is convergent validity.

and the error variances indicate the proportion of the variance in a variable that cannot be explained by the model (Table 1).

```
#We first standardize the variables
cosmetics_std <- scale(cosmetics, center = TRUE, scale = FALSE)
covmat1 <- cov(cosmetics_std[,1:9])
simplemodel1 <-
'organic = ~1*A_organic1 + A_organic2 + A_organic3
  packaging = ~1*A_packaging1 + A_packaging2 + A_packaging3
  crueltyfree = ~1*A_crueltyfree1 + A_crueltyfree2 + A_crueltyfree3
  organic ~~ organic
  packaging ~~ packaging
  crueltyfree ~~ crueltyfree
  organic ~~ packaging
  organic ~~ crueltyfree
  packaging ~~ crueltyfree'
fit1 <- cfa(simplemodel1, sample.cov = covmat1, sample.nobs = nrow(cosmetics))
sum_fit1 <- summary(fit1, fit.measure = T)
sum_fit1_std <- standardizedSolution(fit1)
```

Table 1: The solution of the simple model for the attitudes.

std_loading	value
organic =~ A_organic1	0.87 (0.80, 0.94)***
organic =~ A_organic2	0.73 (0.63, 0.82)***
organic =~ A_organic3	0.72 (0.62, 0.81)***
packaging =~ A_packaging1	0.84 (0.78, 0.91)***
packaging =~ A_packaging2	0.79 (0.72, 0.87)***
packaging =~ A_packaging3	0.80 (0.73, 0.88)***
crueltyfree =~ A_crueltyfree1	0.91 (0.87, 0.96)***
crueltyfree =~ A_crueltyfree2	0.79 (0.72, 0.86)***
crueltyfree =~ A_crueltyfree3	0.86 (0.81, 0.92)***

	std_error.variance	value
10	organic~~organic	1.00 (1.00, 1.00)
11	packaging~~packaging	1.00 (1.00, 1.00)
12	crueltyfree~~crueltyfree	1.00 (1.00, 1.00)
13	organic~~packaging	0.74 (0.63, 0.84)***
14	organic~~crueltyfree	0.60 (0.48, 0.73)***
15	packaging~~crueltyfree	0.72 (0.63, 0.82)***
16	A_organic1~~A_organic1	0.24 (0.12, 0.36)***
17	A_organic2~~A_organic2	0.47 (0.34, 0.61)***
18	A_organic3~~A_organic3	0.48 (0.35, 0.62)***
19	A_packaging1~~A_packaging1	0.29 (0.18, 0.40)***
20	A_packaging2~~A_packaging2	0.37 (0.25, 0.49)***
21	A_packaging3~~A_packaging3	0.35 (0.24, 0.47)***
22	A_crueltyfree1~~A_crueltyfree1	0.17 (0.08, 0.25)***
23	A_crueltyfree2~~A_crueltyfree2	0.38 (0.26, 0.49)***
24	A_crueltyfree3~~A_crueltyfree3	0.25 (0.16, 0.35)***

factor	reliability
organic	0.817
packaging	0.855
crueltyfree	0.892

1.1.2 A 3-factor model with correlated error terms

Since the simple 3-factor model does not seem to perform well, we alter the model by including correlated error terms for all pairs of items that focus on the same aspect. We also impose equal residual correlations for all pairs of items that focus on the same aspect.

```
corrmodel1 <-
'organic = ~1*A_organic1 + A_organic2 + A_organic3
  packaging = ~1*A_packaging1 + A_packaging2 + A_packaging3
  crueltyfree = ~1*A_crueltyfree1 + A_crueltyfree2 + A_crueltyfree3
  A_organic1 ~~c*A_packaging1
  A_organic1 ~~c*A_crueltyfree1
  A_packaging1 ~~c*A_crueltyfree1

  A_organic2 ~~d*A_packaging2
  A_organic2 ~~d*A_crueltyfree2
  A_packaging2 ~~d*A_crueltyfree2

  A_organic3 ~~e*A_packaging3
  A_organic3 ~~e*A_crueltyfree3
  A_packaging3 ~~e*A_crueltyfree3

  organic ~~ organic
  packaging ~~ packaging
  crueltyfree ~~ crueltyfree

  organic ~~ packaging
  organic ~~ crueltyfree
  packaging ~~ crueltyfree
  '
fit1corr <- cfa(corrmodel1, sample.cov = covmat1, sample.nobs = nrow(cosmetics))
sum_fit1corr <- summary(fit1corr, fit.measure = T)
sum_fit1_std_corr <- standardizedSolution(fit1corr)
```

1.1.3 Conclusion

An anova test between the two models shows that the model with correlated error terms is significantly better (p-value < 0.001).

Since, however, the performance measures (second column in Table 3) shows less-than-perfect fit, we look at the residual correlations in the model with correlated error terms for all pairs of attitude items that focus on the same aspect and notice that 7 (19.44%) of all correlations are larger than 0.05 or smaller than -0.05 (this was 27.7% in the simple model). Three of the largest residual correlations involved the correlations between A_organic3, A_packaging3, and A_crueltyfree3 which leads us to believe that the assumption that these correlations are equal does not hold. Indeed, a model that relaxes this assumption has a good TLI (0.967), CFI (0.983), RMSEA (0.073), and SRMR (0.031). The Chi-square goodness of fit test still has a p-value of 0.018.

1.2 CFA to construct a measurement model for the Behavior-Intention items

There are 9 behavior-intention items that are scored on a five-point Likert scale. As with the attitude items, we we fit a CFA on the covariance matrix of the centered dataset.

1.2.1 A simple 3-factor model

Table 3 shows, in the third column) that all performance metrics, except for SRMSR, indicate that this simple model does not fit the data well. Nevertheless, composite reliability (Table 2) is high for all three latent variables.

```
#We first standardize the variables
covmat1 <- cov(cosmetics_std[,10:18])
simplemodel1 <-
'organic = ~1*BI_organic1 + BI_organic2 + BI_organic3
  packaging = ~1*BI_packaging1 + BI_packaging2 + BI_packaging3
  crueltyfree = ~1*BI_crueltyfree1 + BI_crueltyfree2 + BI_crueltyfree3
  organic ~~ organic
  packaging ~~ packaging
  crueltyfree ~~ crueltyfree
  organic ~~ packaging
  organic ~~ crueltyfree
  packaging ~~ crueltyfree'
fit1 <- cfa(simplemodel1, sample.cov = covmat1, sample.nobs = nrow(cosmetics))
sum_fit1 <- summary(fit1, fit.measure = T)
sum_fit1_std <- standardizedSolution(fit1)
```

1.2.2 A 3-factor model with correlated error terms

Since the simple 3-factor model does not seem to perform well, we alter the model by including correlated error terms for all pairs of items that focus on the same aspect. We also impose equal residual residual correlations for all pairs of items that focus on the same aspect.

```
corrmodel1 <-
'organic = ~1*BI_organic1 + BI_organic2 + BI_organic3
  packaging = ~1*BI_packaging1 + BI_packaging2 + BI_packaging3
  crueltyfree = ~1*BI_crueltyfree1 + BI_crueltyfree2 + BI_crueltyfree3

  BI_organic1 ~~c*BI_packaging1
  BI_organic1 ~~c*BI_crueltyfree1
  BI_packaging1 ~~c*BI_crueltyfree1
  BI_organic2 ~~d*BI_packaging2
  BI_organic2 ~~d*BI_crueltyfree2
  BI_packaging2 ~~d*BI_crueltyfree2
  BI_organic3 ~~e*BI_packaging3
  BI_organic3 ~~e*BI_crueltyfree3
  BI_packaging3 ~~e*BI_crueltyfree3

  organic ~~ organic
  packaging ~~ packaging
  crueltyfree ~~ crueltyfree
  organic ~~ packaging
  organic ~~ crueltyfree
  packaging ~~ crueltyfree
  '
fit1corr <- cfa(corrmodel1, sample.cov = covmat1, sample.nobs = nrow(cosmetics))
sum_fit1corr <- summary(fit1corr, fit.measure = T)
sum_fit1_std_corr <- standardizedSolution(fit1corr)
```

Table 2: The standardized solution of the simple model for the behavior-intent items.

std_loading	value
organic =~ BI_organic1	0.89 (0.84, 0.93)***
organic =~ BI_organic2	0.90 (0.85, 0.94)***
organic =~ BI_organic3	0.84 (0.79, 0.90)***
packaging =~ BI_packaging1	0.88 (0.83, 0.92)***
packaging =~ BI_packaging2	0.89 (0.85, 0.93)***
packaging =~ BI_packaging3	0.87 (0.82, 0.91)***
crueltyfree =~ BI_crueltyfree1	0.92 (0.88, 0.95)***
crueltyfree =~ BI_crueltyfree2	0.92 (0.89, 0.95)***
crueltyfree =~ BI_crueltyfree3	0.94 (0.91, 0.97)***

	std_error.variance	value
10	organic~~organic	1.00 (1.00, 1.00)
11	packaging~~packaging	1.00 (1.00, 1.00)
12	crueltyfree~~crueltyfree	1.00 (1.00, 1.00)
13	organic~~packaging	0.88 (0.82, 0.93)***
14	organic~~crueltyfree	0.78 (0.71, 0.86)***
15	packaging~~crueltyfree	0.83 (0.77, 0.90)***
16	BI_organic1~~BI_organic1	0.22 (0.14, 0.29)***
17	BI_organic2~~BI_organic2	0.20 (0.12, 0.27)***
18	BI_organic3~~BI_organic3	0.29 (0.20, 0.38)***
19	BI_packaging1~~BI_packaging1	0.23 (0.15, 0.31)***
20	BI_packaging2~~BI_packaging2	0.21 (0.13, 0.28)***
21	BI_packaging3~~BI_packaging3	0.25 (0.17, 0.33)***
22	BI_crueltyfree1~~BI_crueltyfree1	0.16 (0.10, 0.22)***
23	BI_crueltyfree2~~BI_crueltyfree2	0.16 (0.10, 0.22)***
24	BI_crueltyfree3~~BI_crueltyfree3	0.12 (0.07, 0.17)***

factor	reliability
organic	0.908
packaging	0.910
crueltyfree	0.946

Table 3: Performance measure for the different models

	Attitudes		Behavior-intention	
parameter	simple model	with correlated error terms	simple model	with correlated error terms
user model Chisq. (df)	120.89 (24)***	56.74 (21)***	147.81 (24)***	26.78 (21)
baseline model Chisq. (df)	906.01 (36) ***	906.01 (36) ***	1478.43 (36) ***	1478.43 (36) ***
comparative fit index (CFI)	0.889	0.959	0.914	0.996
Tucker-Lewis index (TLI)	0.833	0.93	0.871	0.993
RMSEA (ll,ul)	0.16 (0.14, 0.19)***	0.11 (0.07, 0.14)**	0.19 (0.16, 0.21)***	0.04 (0.00, 0.09)
Standardized root mean square residual	0.057	0.042	0.033	0.02

1.2.3 Conclusion

An anova test between the two models shows that the model with correlated error terms for all pairs of Behavior-Intention items that focus on the same aspect is significantly better (p-value < 0.001).

The performance measures (column 3 and 4 in Table 3) show a good fit and all residual correlations are between -0.05 and 0.05 (the simpler model had 0 (0%) residual correlations between -0.05 and 0.05). For the simple model We shall thus keep this model as the final model.

1.3 Structural equation model to evaluate the impact of attitude on behavior intention

We first fit a structural equation model on the covariance matrix of all items.

- A_organic, A_packaging, and A_crueltyfree are related to the attitude items with a model with correlated error terms for pairs of items that focus on the same aspects. For statements that focus on "the right thing to do" or "pleasant", there are equal correlations. As discussed in section 1.1.3, we relax the constraint of equal residual correlations for items that focus on the fact that purchasing sustainable cosmetics is "a must".
- BI_organic, BI_packaging, and BI_crueltyfree are related to the attitude items with a model with correlated error terms for pairs of items that focus on the same aspects. As discussed in section 1.2.3, a model that imposes the constraint of equal residual correlations for all pairs of items that focus on the same aspect has a good fit and will be used here.

Structural relations are added to assess the effect of (1) Att_organic on BI_organic, (2) Att_packaging on BI_packaging and (3) Att_crueltyfree on BI_crueltyfree.

```
cormat <- cov(cosmetics_std)
sem1 <- 'BI_organic = ~1*BI_organic1 + BI_organic2 + BI_organic3
  BI_packaging = ~1*BI_packaging1 + BI_packaging2 + BI_packaging3
  BI_crueltyfree = ~1*BI_crueltyfree1 + BI_crueltyfree2 + BI_crueltyfree3
  BI_organic1 ~~c*BI_packaging1
  BI_organic1 ~~c*BI_crueltyfree1
  BI_packaging1 ~~c*BI_crueltyfree1
  BI_organic2 ~~d*BI_packaging2
  BI_organic2 ~~d*BI_crueltyfree2
  BI_packaging2 ~~d*BI_crueltyfree2
  BI_organic3 ~~e*BI_packaging3
  BI_organic3 ~~e*BI_crueltyfree3
  BI_packaging3 ~~e*BI_crueltyfree3
  BI_organic ~~ BI_organic
  BI_packaging ~~ BI_packaging
  BI_crueltyfree ~~ BI_crueltyfree
  BI_organic ~~ BI_packaging
  BI_organic ~~ BI_crueltyfree
  BI_packaging ~~ BI_crueltyfree

  A_organic = ~1*A_organic1 + A_organic2 + A_organic3
  A_packaging = ~1*A_packaging1 + A_packaging2 + A_packaging3
  A_crueltyfree = ~1*A_crueltyfree1 + A_crueltyfree2 + A_crueltyfree3
  A_organic1 ~~a*A_packaging1
  A_organic1 ~~a*A_crueltyfree1
  A_packaging1 ~~a*A_crueltyfree1
  A_organic2 ~~b*A_packaging2
  A_organic2 ~~b*A_crueltyfree2
  A_packaging2 ~~b*A_crueltyfree2
```

```
  A_organic3 ~~A_packaging3
  A_organic3 ~~A_crueltyfree3
  A_packaging3 ~~A_crueltyfree3
  A_organic ~~ A_organic
  A_packaging ~~ A_packaging
  A_crueltyfree ~~ A_crueltyfree
  A_organic ~~ A_packaging
  A_organic ~~ A_crueltyfree
  A_packaging ~~ A_crueltyfree

  #structural model
  BI_organic ~A_organic
  BI_packaging ~A_packaging
  BI_crueltyfree ~A_crueltyfree
  '
fitsem1 <- sem(sem1, sample.cov = cormat, sample.nobs = nrow(cosmetics))
sum_sem1 <- summary(fitsem1)
sum_sem1_std <- standardizedSolution(fitsem1)
```

With a test statistics of 145.01 with 118 degrees of freedom, the chi-square p-value is 0.046 which means we can reject the null hypothesis that the model fits well.

The structural equation model shows that all correlations between latent variables are positive and highly significant. The unstandardized and standardized regression coefficients are shown in respectively the first and second column of Table 4.

- an increase of one unit in attitude_organic increases the behavior intention with 0.68.
- an increase of one unit in attitude_packaging increases the behavior intention with 0.684.
- an increase of one unit in attitude_crueltyfree increases the behavior intention with 0.716.

These population regression coefficients are quite similar so we next test a model that imposes that all three regression coefficients are the same.

1.3.1 Equal population regression coefficients

To fit a model with equal population regression coefficients, we replace the *structural model* part in the previous SEM description with the expression below and re-fit the model.

```
'  #structural model
  BI_organic ~p*A_organic
  BI_packaging ~p*A_packaging
  BI_crueltyfree ~p*A_crueltyfree'
```

With a test statistics of 146.18 with 120 degrees of freedom, the chi-square p-value is 0.052 which means we cannot reject the null hypothesis that the model fits well.

Since an anova test for the two SEMs has a p-value of 0.557, we cannot reject the null hypothesis that the models are the same. Nevertheless, the chi-square test was slightly better so we prefer this simpler model with equal population regression coefficients of the structural model. The unstandardized and standardized regression coefficients are shown in respectively the third and fourth column of Table 4.

- an increase of one unit in attitude_organic increases the behavior intention with 0.666.
- an increase of one unit in attitude_packaging increases the behavior intention with 0.695.
- an increase of one unit in attitude_crueltyfree increases the behavior intention with 0.714.

Table 4: Population regression coefficients in both SEMs.

Regression coefficient	General SEM		Equal population regression coefficients	
	unstandardized	standardized	unstandardized	standardized
BI_organic~ A_organic	0.87 ***	0.68 ***	0.81 ***	0.67 ***
BI_packaging~ A_packaging	0.76 ***	0.68 ***	0.81 ***	0.70 ***
BI_crueltyfree~ A_crueltyfree	0.82 ***	0.72 ***	0.81 ***	0.71 ***

2 Task 2

2.1 Canonical correlation analysis

After standardizing both X and Y variables with zero mean and unit variance, we proceed with inspecting the squared canonical correlations. We can see that the first canonical variate u1 (based on a linear combination of X variables) accounts for 23.4% of the variance in the canonical variate t1 (based on a linear combination of Y variables). The canonical variate u2 accounts for 5.2% of the variance in the canonical variate t2, etc.

```
zbenefits <- benefits
zbenefits[, 2:14] <- scale(zbenefits[, 2:14], scale = TRUE, center = TRUE)


cancor.out <- cancor(cbind(SL_pensioners, SL_unemployed, SL_old_gvntresp,
                           SL_unemp_gvntresp)
                     ~ SB_strain_economy + SB_prevent_poverty +
                       SB_equal_society + SB_taxes_business +
                       SB_make_lazy + SB_caring_others +
                       unemployed_notmotivated + SB_often_lessthanentitled +
                       SB_often_notentitled, data = zbenefits)
#print summary results
summary(cancor.out)
```

```
##
## Canonical correlation analysis of:
##   9   X  variables:  SB_strain_economy, SB_prevent_poverty, SB_equal_society, SB_taxes_business, S
##   with    4   Y  variables:  SL_pensioners, SL_unemployed, SL_old_gvntresp, SL_unemp_gvntresp
##
##      CanR   CanRSQ   Eigen percent    cum                          scree
## 1 0.48323 0.233515 0.30466 79.8465  79.85 ******************************
## 2 0.22817 0.052061 0.05492 14.3939  94.24 *****
## 3 0.13741 0.018883 0.01925  5.0442  99.28 **
## 4 0.05218 0.002723 0.00273  0.7155 100.00
##
## Test of H0: The canonical correlations in the
## current row and all that follow are zero
##
##      CanR LR test stat approx F numDF   denDF   Pr(> F)
## 1 0.48323      0.71092   32.719    36 12357.1 < 2.2e-16 ***
```

```
## 2 0.22817      0.92751   10.477    24  9565.8 < 2.2e-16 ***
## 3 0.13741      0.97845    5.163    14  6598.0 8.545e-10 ***
## 4 0.05218      0.99728    1.501     6  3300.0    0.1735
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Raw canonical coefficients
##
##    X  variables:
##                                  Xcan1      Xcan2     Xcan3     Xcan4
## SB_strain_economy         -0.0909717  0.4172121  0.564470 -0.059128
## SB_prevent_poverty         0.0779679 -0.0254661 -0.329579 -0.125299
## SB_equal_society           0.1279718  0.3828047 -0.585296 -0.097459
## SB_taxes_business         -0.0850983  0.0972611 -0.067364 -0.947887
## SB_make_lazy              -0.3819813  0.0411048 -0.206351  0.231770
## SB_caring_others           0.0069064  0.0060264  0.128499 -0.149934
## unemployed_notmotivated   -0.4933957 -0.1393655 -0.333507  0.134556
## SB_often_lessthanentitled  0.2525276 -0.6831611  0.127790 -0.360191
## SB_often_notentitled      -0.1393188 -0.4867982 -0.255268  0.146316
##
##    Y  variables:
##                         Ycan1     Ycan2    Ycan3    Ycan4
## SL_pensioners      0.220475  0.651836 -0.28265  0.78198
## SL_unemployed     -0.526682  0.156985 -0.64871 -0.63976
## SL_old_gvntresp   -0.098433 -0.599184 -0.55693  0.72377
## SL_unemp_gvntresp  0.764899  0.057483 -0.33698 -0.71784
```

```r
#compute redundancies
R2tu <- cancor.out$cancor^2
R2tu <- cancor.out$cancor^2
VAFYbyt <- apply(cancor.out$structure$Y.yscores^2, 2, sum)/3
redund <- R2tu*VAFYbyt
round(cbind(R2tu,VAFYbyt,redund,total = cumsum(redund)), 4)
```

```
##         R2tu VAFYbyt redund  total
## Ycan1 0.2335  0.3799 0.0887 0.0887
## Ycan2 0.0521  0.4266 0.0222 0.1109
## Ycan3 0.0189  0.3635 0.0069 0.1178
## Ycan4 0.0027  0.1633 0.0004 0.1182
```

```r
#print canonical loadings
round(cancor.out$structure$X.xscores, 2)
```

```
##                           Xcan1 Xcan2 Xcan3 Xcan4
## SB_strain_economy         -0.54  0.27  0.44 -0.27
## SB_prevent_poverty         0.22  0.10 -0.53 -0.18
## SB_equal_society           0.33  0.33 -0.73 -0.15
## SB_taxes_business         -0.45  0.12  0.01 -0.85
## SB_make_lazy              -0.80 -0.02 -0.02 -0.05
## SB_caring_others          -0.56 -0.06  0.07 -0.21
## unemployed_notmotivated   -0.80 -0.19 -0.26 -0.02
## SB_often_lessthanentitled  0.30 -0.73  0.06 -0.36
## SB_often_notentitled      -0.56 -0.47 -0.19  0.00
```

```r
round(cancor.out$structure$Y.yscores, 2)
```

```
##                   Ycan1 Ycan2 Ycan3 Ycan4
## SL_pensioners      0.18  0.81 -0.36  0.42
## SL_unemployed     -0.61  0.31 -0.65 -0.32
## SL_old_gvntresp    0.11 -0.71 -0.60  0.34
## SL_unemp_gvntresp  0.85 -0.11 -0.42 -0.30
```

To investigate the amount of variance in Y that is accounted for by X, we compute the redundancies from the output. We can see from below that u1, which is the first pair of canonical variates, accounts for 8.9% of the variance in the Y variables, and the second and the third pair of canonical variates accounts for an additional 2.2% and 0.7% of the variance in the Y variables. As the additional variance accounted for by the forth canonical variate is negligible, we can say that the total amount of variance in Y that can be accounted for by X is 11.8%. As the additional variance accounted for by the last canonical variates u4 is rather small (0.0004), this suggests that not all pairs of canonical variates are significant. We can use Wilk's Lambda for a formal test.

The p-value of 0.1735 suggests that there is not enough evidence to reject the null hypothesis: p(u4, t4) = 0. Therefore, the last canonical correlation is zero. As the other p-values are very small, we reach the same conclusion that the first three pairs of canonical correlations are significant. In particular, the canonical correlation of the first pair (r = 48.3%) and the second pair (r = 22.8%) are stronger while the canonical correlation of the third pair (r = 13.7%) is weaker, therefore we focus on the first two pairs for interpretation.

To interpret canonical variates, we look at the canonical loadings, which summarizes the correlations between the canonical variates and the respective original variables. We can observe that the first covariate u1 is strongly negatively correlated with X variables (e.g.SB_make_lazy, unemployed_notmotivated) suggesting that the unemployed are too lazy to find jobs. The first covariate t1 is strongly positively correlated with SL_unemp_gvntresp which reflects the standard living for the employed. Not surprising, having high standard of living when unemployed is likely associated with more laziness to find jobs.

The second covariate u2 is strongly negatively correlated with the X variable SB_often_lessthanentitled, which indicates many with very low incomes get less benefit than legally entitled to. The second covariate t2 is strongly positively correlated with the Y variable SL_pensioners which reflects the standard living for the pensioners, and is strongly negatively assoicated with the Y variable SL_old_gvntresp which reflects the standard of living for the old.

2.2 Split-half approach

```r
train <- benefits[seq(2,3310, by = 2), ]
valid <- benefits[seq(1,3310, by = 2), ]
train[,2:14] <- scale(train[, 2:14], center = TRUE, scale = TRUE)
valid[,2:14] <- scale(valid[, 2:14], center = TRUE, scale = TRUE)

#conduct CCA on training data
cancor.train <- cancor(cbind(SL_pensioners, SL_unemployed, SL_old_gvntresp,
                             SL_unemp_gvntresp)
                       ~ SB_strain_economy + SB_prevent_poverty +
                         SB_equal_society + SB_taxes_business +
                         SB_make_lazy + SB_caring_others +
                         unemployed_notmotivated + SB_often_lessthanentitled +
                         SB_often_notentitled, data = train)
#conduct CCA on validation data
cancor.valid <- cancor(cbind(SL_pensioners, SL_unemployed, SL_old_gvntresp,
                             SL_unemp_gvntresp)
```

```r
                           ~ SB_strain_economy + SB_prevent_poverty +
                             SB_equal_society + SB_taxes_business +
                             SB_make_lazy + SB_caring_others +
                             unemployed_notmotivated + SB_often_lessthanentitled +
                             SB_often_notentitled, data = valid)
# canonical variates calibration set
train.X1 <- cancor.train$score$X
train.Y1 <- cancor.train$score$Y
# compute canonical variates using data of calibration set and coefficients
# estimated on validation set
train.X2 <- as.matrix(train[,6:14]) %*% cancor.valid$coef$X
train.Y2 <- as.matrix(train[,2:5]) %*% cancor.valid$coef$Y
round(cor(train.Y1,train.Y2),3)
```

```
##        Ycan1  Ycan2  Ycan3  Ycan4
## Ycan1 -0.985  0.121 -0.148  0.044
## Ycan2 -0.057 -0.989 -0.116 -0.036
## Ycan3  0.146  0.083 -0.973 -0.145
## Ycan4  0.069  0.006 -0.130  0.988
```

```r
round(cor(train.X1,train.X2),3)
```

```
##        Xcan1  Xcan2  Xcan3  Xcan4
## Xcan1 -0.985 -0.013 -0.058 -0.100
## Xcan2  0.040 -0.893 -0.219  0.283
## Xcan3  0.031  0.027 -0.557 -0.206
## Xcan4 -0.091  0.100  0.072  0.257
```

```r
round(cor(train.X1,train.Y1),3)
```

```
##       Ycan1 Ycan2 Ycan3 Ycan4
## Xcan1 0.482 0.000 0.000 0.000
## Xcan2 0.000 0.244 0.000 0.000
## Xcan3 0.000 0.000 0.145 0.000
## Xcan4 0.000 0.000 0.000 0.046
```

```r
round(cor(train.X2,train.Y2),3)
```

```
##       Ycan1  Ycan2 Ycan3  Ycan4
## Xcan1 0.468 -0.067 0.065 -0.026
## Xcan2 0.019  0.215 0.022  0.011
## Xcan3 0.019  0.043 0.089  0.016
## Xcan4 0.040 -0.076 0.027  0.011
```

```r
round(cor(train.Y2,train.Y2),3)
```

```
##        Ycan1  Ycan2 Ycan3 Ycan4
## Ycan1  1.000 -0.050 0.001 0.006
## Ycan2 -0.050  1.000 0.014 0.034
## Ycan3  0.001  0.014 1.000 0.010
## Ycan4  0.006  0.034 0.010 1.000
```

```r
round(cor(train.X2,train.X2),3)
```

```
##        Xcan1  Xcan2  Xcan3 Xcan4
## Xcan1  1.000 -0.037 -0.047 0.020
```

```
## Xcan2 -0.037  1.000  0.024 0.017
## Xcan3 -0.047  0.024  1.000 0.035
## Xcan4  0.020  0.017  0.035 1.000
```

We then utilized the split-half approach for conducting canonical correlation analysis on both validation and training dataset to examine the validity of the results. From the reliability of the canonical variates for Y and X variables, we can observe that the Xcan1 and Ycan1 contributes 0.482 and 0.468 in the comparisons of the elements which are considered reliable compared to their counterparts. The off-diagonal elements are close to 0 which is considered as normal because the canonical variates compared are based on different datasets so there should be very less correlations.

3 Appendix

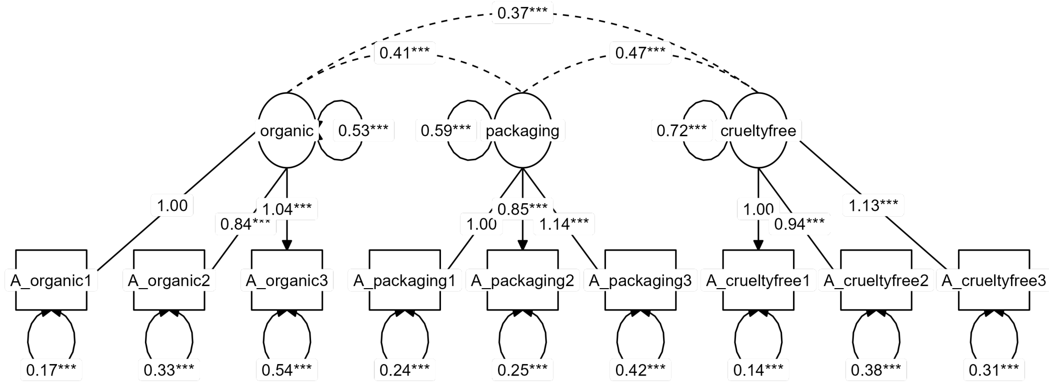

Figure 1: A graphical representation of the simple model for the attitudes.

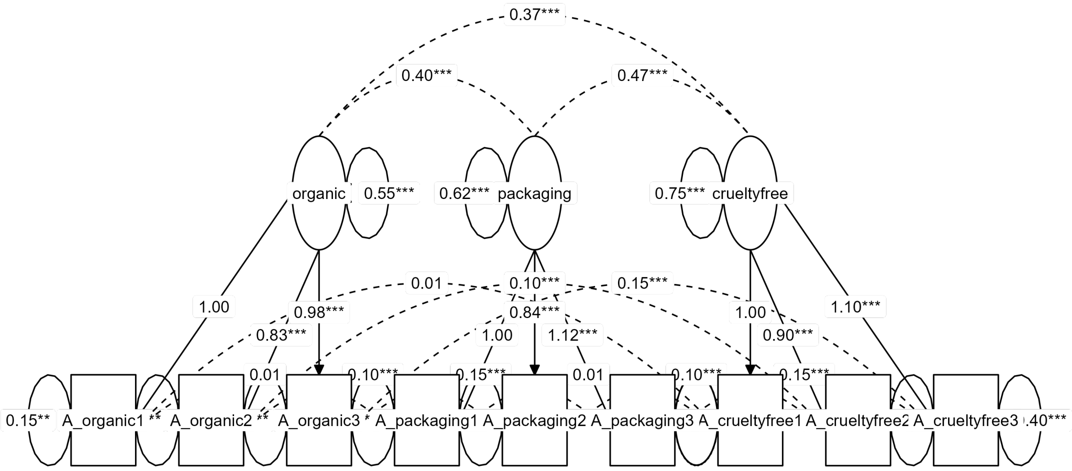

Figure 2: A graphical representation of the model for the attitudeswith correlated error terms for all pairs of items that focus on the same aspect.

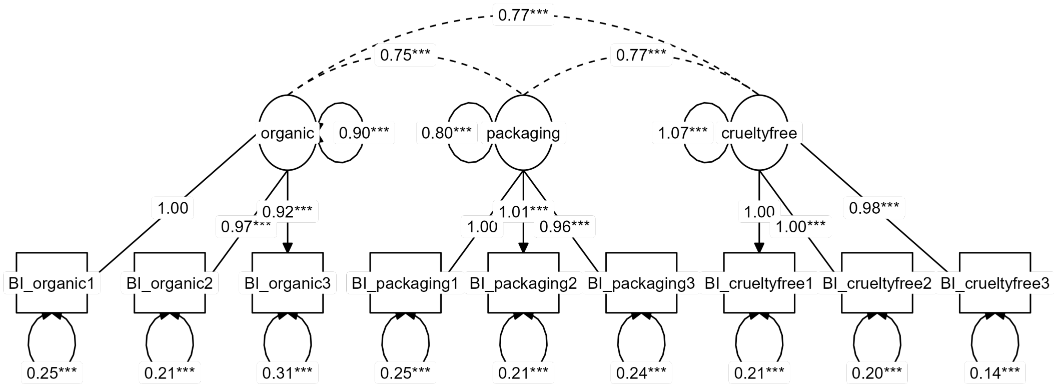

Figure 3: A graphical representation of the simple model for the behavior-intent items.

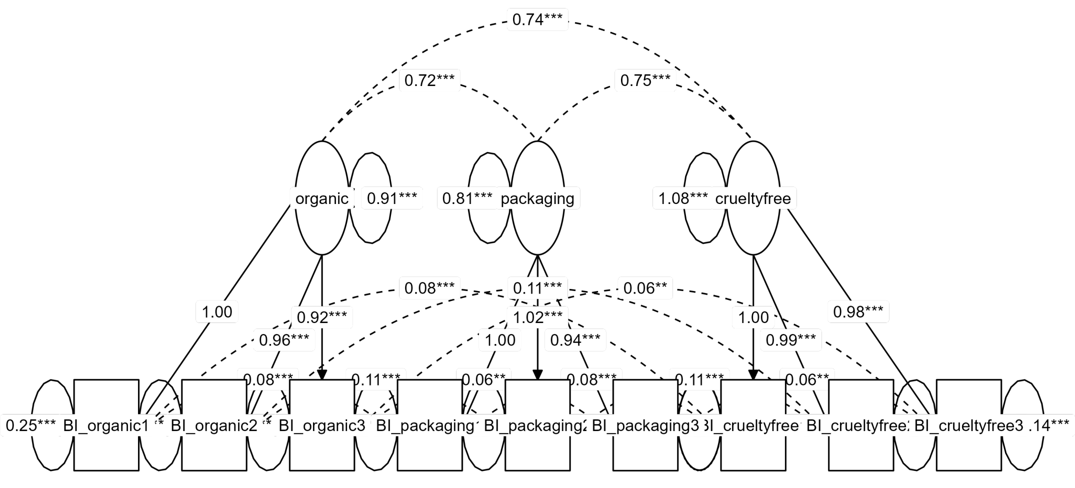

Figure 4: A graphical representation of the model with correlated error terms for the behavior-intent items that focus on the same aspect.