

Utilisation des packages *lavaan* et *psych* de R appliquée à la validation psychométrique d'un questionnaire d'abus de substances (DAST)

R à Québec 2017

Charles-Édouard Giguère

Stéphane Potvin

Centre intégré
universitaire de santé
et de services sociaux
de l'Est-de-
l'Île-de-Montréal

Québec 

Signature

Le contenu de cette présentation se trouvent sur:

<https://sites.google.com/site/cegiguere/presentation-r-a-quebec>

Cet atelier présente les syntaxes de l'article suivant.

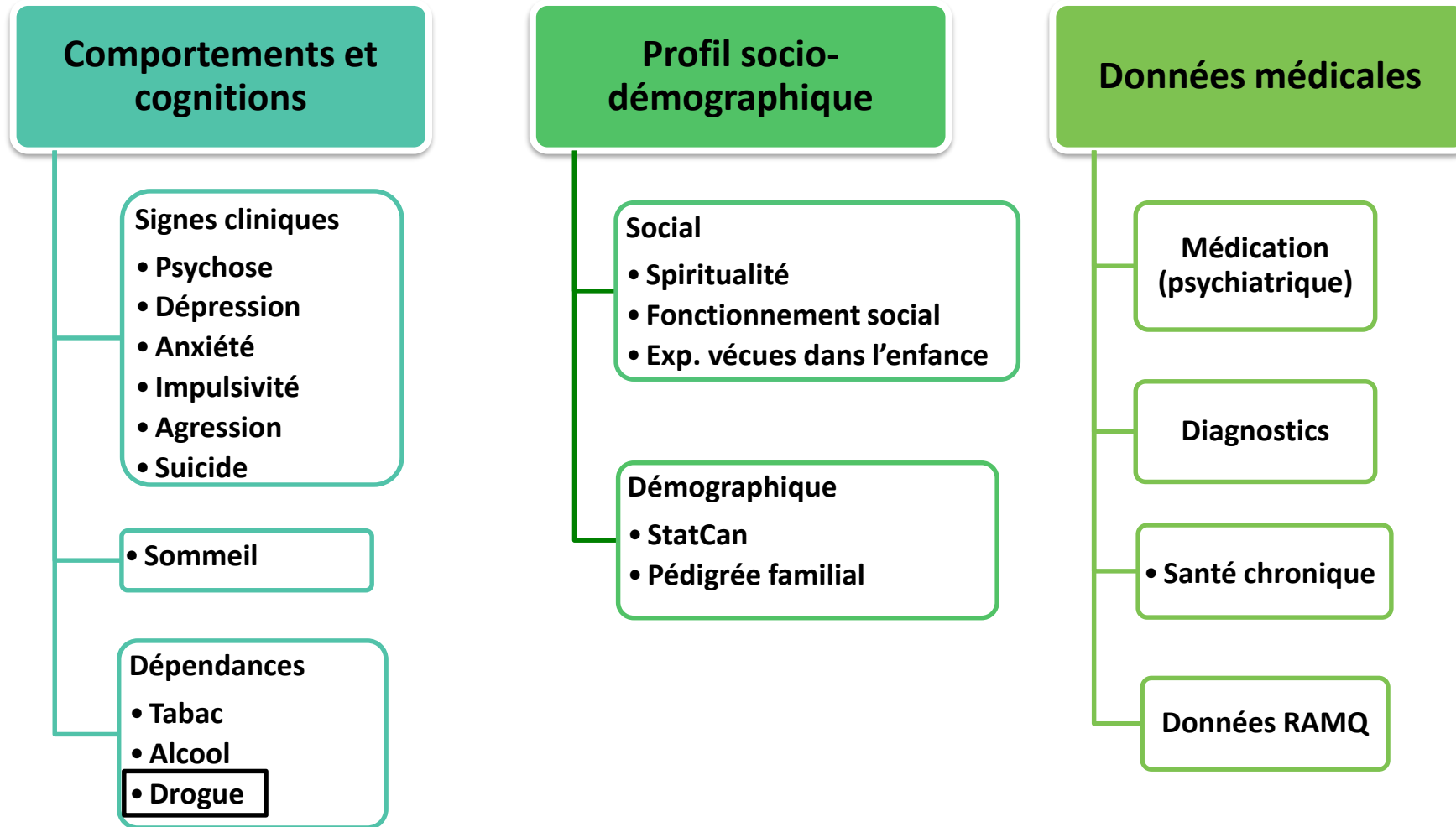
Giguère, C.E. et Potvin S. (2017). **The Drug Abuse Screening Test preserves its excellent psychometric properties in psychiatric patients evaluated in an emergency setting.** *Journal of addictive behaviors.*

Banque Signature (CentreSignature.ca)



Données psychosociales et cliniques

16 questionnaires auto-administrés, scientifiquement validés



Choix de questionnaires basé sur les dimensions et l'approche transnosologique de DSM5 / rDOC NIMH

23 Biomarqueurs

sélectionnés en collaboration avec 8 experts internationaux

- Hémoglobine glyquée
- Cholestérol
- HDL-Cholestérol
- LDL-Cholestérol (calculé)
- Triglycérides
- Glucose
- Insuline

Métaboliques

- 17 β Estradiol
- Testostérone
- Progestérone
- Cortisol

Hormonaux

- Interleukine-6 (IL-6)
- C-Reactive Protein
- Insulin-like growth factor
- TNF-alpha
- IL-1-beta
- Endocannabinoïdes

Inflammatoires
et immunitaires

- Toxoplasmose
(IgM et IgG)

Infectieux

- BPC, pesticides
organochlorés, BFR
(retardateurs de flammes)
- Métaux lourds - Plomb,
Manganèse
- Métaux lourds - Mercure

Toxicologiques

- ADN
- ARN

Génétiques

Le *Drug Abuse Screening Test (DAST-10)*

Les questions suivantes portent sur votre consommation de drogues (sans inclure votre consommation d'alcool ou de tabac) au cours des 12 derniers mois.

1. Avez-vous fait usage de drogues autres que les médicaments administrés à des fins médicales?
2. Utilisez-vous **plus d'une drogue** à la fois?
3. Pouvez-vous **interrompre** votre consommation de drogues au moment voulu?
4. Avez-vous déjà perdu connaissance ou eu des récurrences (« **flashbacks** ») après avoir pris une drogue?
5. Votre consommation de drogues suscite-t-elle un sentiment de **culpabilité** de votre part?
6. Votre épouse ou vos parents ont-ils déjà **critiqué** votre consommation de drogues?
7. Avez-vous **négligé votre famille** en raison de votre consommation de drogues?
8. Vous êtes-vous déjà engagé dans des **activités illégales** afin d'obtenir de la drogue?
9. Avez-vous déjà éprouvé des **symptômes de sevrage** (malaises) après avoir interrompu votre consommation de drogues?
10. Avez-vous eu des **problèmes médicaux** attribuables à votre consommation de drogues (par exemple, les pertes de mémoire, l'hépatite, les convulsions, les saignements, etc.).

Cohérence interne (α de Chronbach)

Afin de calculer la cohérence interne de l'instrument nous avons utiliser la librairie *psych* (Revelle, William 2017).

- Lee J. Chronbach (1957) défini une mesure α qui estime la cohérence interne d'un instrument:

$$\alpha = \frac{K}{K - 1} \left(1 - \frac{\sum_{i=1}^K \Sigma_{ii}}{\sum_{i=1}^K \sum_{j=1}^K \Sigma_{ij}} \right)$$

où K est le nombre d'items et Σ est la matrice de covariance de K items dont on veut effectuer la somme.

Cohérence interne (α de Chronbach)

```
library(dplyr)
```

```
library(psych)
```

```
### Alpha de Chronbach.
```

```
### Intervalle de confiance traditionnel.
```

```
alpha(select(DAST.df,  
            matches("DAST10[01][0-9]n")))
```

```
### Intervalle de confiance par rééchantillonnage (1000 échantillons).
```

```
alpha(select(DAST.df, matches("DAST10[01][0-9]n")),  
       n.iter = 1000)
```


Cohérence interne (α de Chronbach)

raw_alpha	std.alpha	G6(sm)	average_r	S/N	ase	mean	sd
0.88	0.88	0.88	0.42	7.2	0.0058	0.25	0.3

lower	alpha	upper	95% confidence boundaries	
0.87	0.88	0.89		

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(sm)	average_r	S/N	alpha	se
DAST1001n	0.86	0.86	0.86	0.41	6.3	0.0065	
DAST1002n	0.87	0.87	0.87	0.42	6.4	0.0064	
DAST1003n	0.90	0.90	0.89	0.49	8.7	0.0050	
DAST1004n	0.86	0.86	0.87	0.41	6.3	0.0065	
DAST1005n	0.86	0.86	0.87	0.41	6.3	0.0065	
DAST1006n	0.86	0.86	0.86	0.40	6.0	0.0070	
DAST1007n	0.86	0.86	0.86	0.40	5.9	0.0069	
DAST1008n	0.87	0.87	0.87	0.42	6.4	0.0064	
DAST1009n	0.86	0.86	0.86	0.40	6.0	0.0068	
DAST1010n	0.86	0.86	0.87	0.41	6.3	0.0065	

Cohérence interne (α de Chronbach)

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
DAST1001n	912	0.73	0.71	0.68	0.63	0.44	0.50
DAST1002n	912	0.69	0.69	0.64	0.60	0.20	0.40
DAST1003n	911	0.29	0.30	0.17	0.15	0.22	0.41
DAST1004n	912	0.71	0.72	0.68	0.63	0.20	0.40
DAST1005n	912	0.72	0.71	0.68	0.64	0.28	0.45
DAST1006n	912	0.80	0.79	0.78	0.73	0.34	0.48
DAST1007n	912	0.80	0.80	0.78	0.73	0.24	0.43
DAST1008n	912	0.68	0.69	0.64	0.60	0.17	0.37
DAST1009n	912	0.78	0.78	0.76	0.71	0.26	0.44
DAST1010n	912	0.70	0.71	0.67	0.63	0.19	0.39

Non missing response frequency for each item

	0	1	miss
DAST1001n	0.56	0.44	0
DAST1002n	0.80	0.20	0
DAST1003n	0.78	0.22	0
DAST1004n	0.80	0.20	0
DAST1005n	0.72	0.28	0
DAST1006n	0.66	0.34	0
DAST1007n	0.76	0.24	0
DAST1008n	0.83	0.17	0
DAST1009n	0.74	0.26	0
DAST1010n	0.81	0.19	0

Cohérence interne (α de Chronbach)

```
### Création d'une variable index.
```

```
DAST.df$index <- 1:dim(DAST.df)[1]
```

```
### Fonction pour extraire l'estimateur et l'intervalle de confiance
```

```
### de l'alpha de Chronbach.
```

```
alpha.ci.est <- function(index){
```

```
  ax <- suppressMessages(
```

```
    suppressWarnings(
```

```
      alpha(select(DAST.df[index,], matches("DAST10[01][0-9]n")),
```

```
        warnings=FALSE)))
```

```
  c(alpha = ax$total$raw_alpha,
```

```
    alpha.ci.lo = ax$total$raw_alpha - qnorm(0.975)*ax$total$ase,
```

```
    alpha.ci.hi = ax$total$raw_alpha + qnorm(0.975)*ax$total$ase)
```

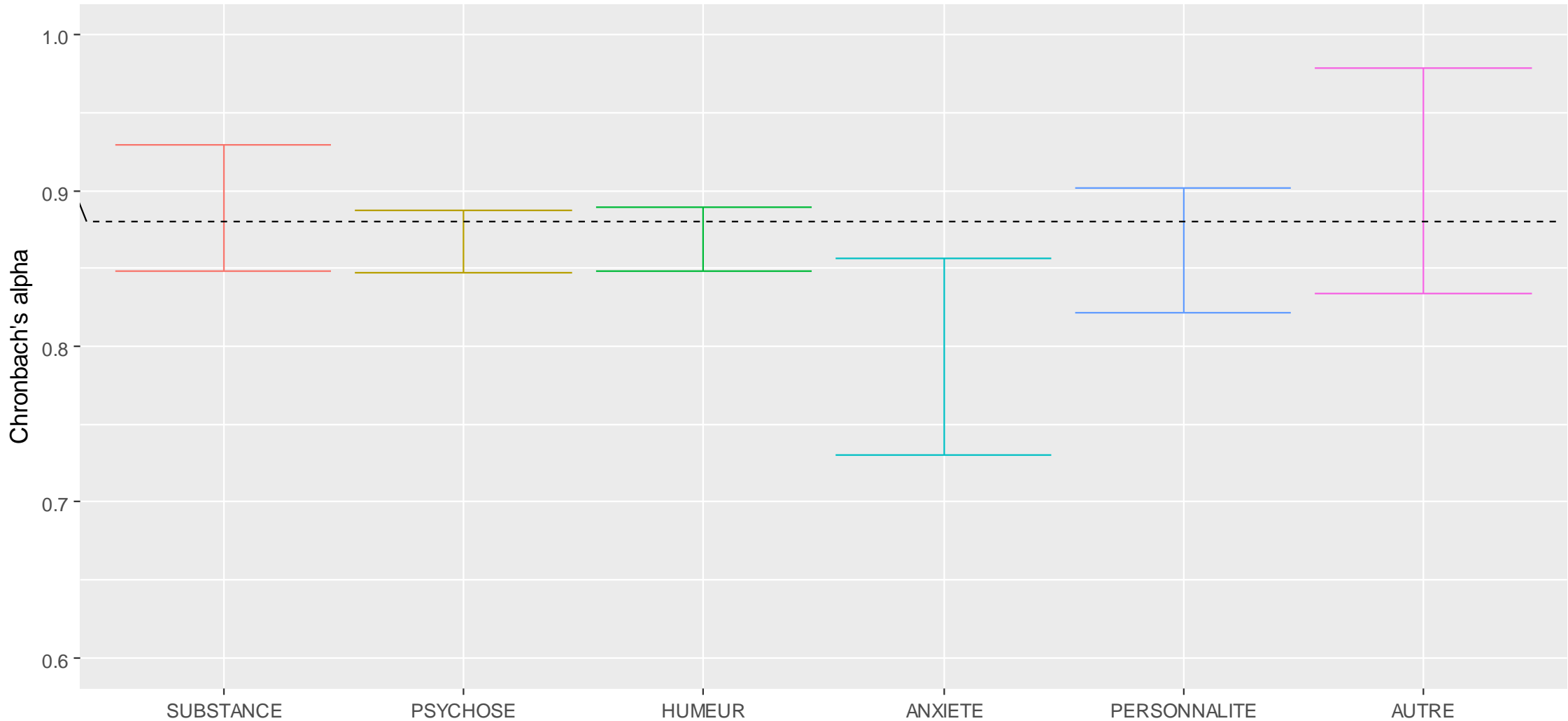
```
}
```

Cohérence interne (α de Chronbach)

```
### Graphique des intervalles de confiance des alphas par diagnostic.
```

```
DAST.df %>%  
  group_by(CIMDX) %>%  
  summarise(alpha      = alpha.ci.est(index)[1],  
            alpha.lo   = alpha.ci.est(index)[2],  
            alpha.hi   = alpha.ci.est(index)[3]) %>%  
  ggplot(aes(x = CIMDX,  
            y = alpha,  
            ymin = alpha.lo,  
            ymax = alpha.hi,  
            color = CIMDX)) +  
  geom_errorbar(show.legend = FALSE) +  
  ylim(c(0.6,1)) +  
  ylab("Chronbach's alpha") +  
  xlab("")
```

Cohérence interne (α de Chronbach)

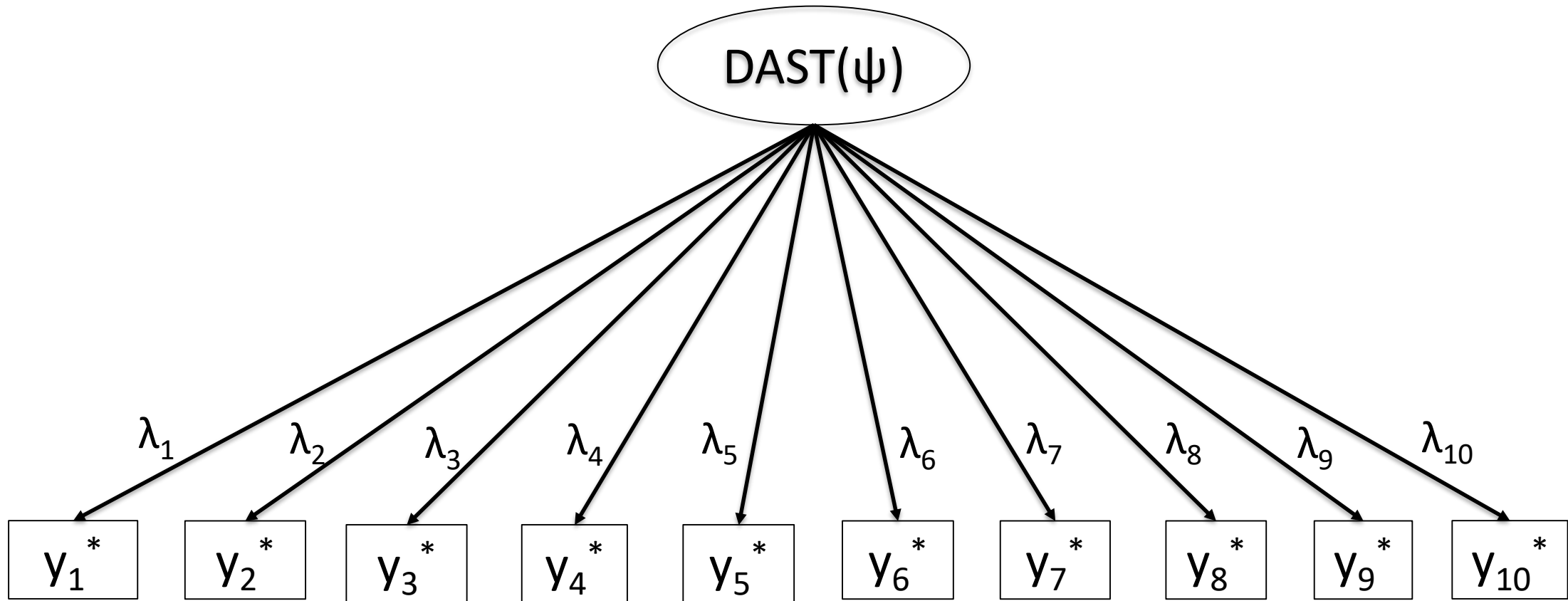


Analyse factorielle confirmatoire (CFA)

- Afin de confirmer que le DAST est bien un modèle à un facteur nous avons utilisé la librairie *lavaan*.
- Soit l'item k dichotomique (0/1) du participant i , Y_{ik} , voici comment le modèle CFA est modélisé :

$$y_{ik} = \varphi(y_{ik}^*, \tau_k) = \begin{cases} 1 & \text{si } y_{ik}^* > \tau_k \\ 0 & \text{si } y_{ik}^* \leq \tau_k \end{cases}$$
$$y_{ik}^* = \beta_{0k} + \lambda_k \psi_i + \epsilon_{ik}$$

Analyse factorielle confirmatoire (CFA)



Analyse factorielle confirmatoire (CFA)

Syntaxes des modèles de *lavaan*.

Création d'un facteur.

Facteur =~ It1 + It2 + ...

Régression de y sur x.

y ~ x

Définition des (co)variances.

y ~~ y

y ~~ x1 + x2

Intercept du modèle

y ~ 1

Définition des seuils (thresholds).

y | t1 t2

Paramètres des modèles de *lavaan*

Fixer un coefficient.

Facteur =~ 1*It1 + It2 + ...

Libérer un coefficient.

Facteur =~ NA*It1 + ...

Définir un label.

y = (lab)*x

y = c(lab.g1,lab.g2)*x

Définir un nouveau paramètre.

Dif.lab := lab.g1 - lab.g2

Définir une contrainte.

lab >1

Analyse factorielle confirmatoire (CFA)

```
### Modèle d'analyse factorielle confirmatoire
```

```
CFA.DAST <- '
```

```
DAST =~ DAST1001f + DAST1002f + DAST1003f + DAST1004f + DAST1005f +  
        DAST1006f + DAST1007f + DAST1008f + DAST1009f + DAST1010f  
,
```

```
### Ajustement du modèle.
```

```
cfaDAST <- cfa(CFA.DAST, DAST.df,  
              ordered = sprintf("DAST10%02df",1:10),  
              parameterization = "theta",  
              estimator = "WLSMV")
```

```
### Sommaire du modèle.
```

```
summary(cfaDAST, standard = TRUE, fit = TRUE)
```

Analyse factorielle confirmatoire (CFA)

Number of observations	Used 911	Total 912	Root Mean Square Error of Approximation:			
Estimator	DWLS	Robust	RMSEA	0.031	0.050	
Minimum Function Test Statistic	65.789	115.029	90 Percent C.I.	0.019	0.043	0.04 0.060
Degrees of freedom	35	35	P-value RMSEA <= 0.05	0.998	0.473	
P-value (Chi-square)	0.001	0.000	Robust RMSEA	NA		
Scaling correction factor		0.590	90 Percent Confidence Interval	NA	NA	
Shift parameter		3.468	Standardized Root Mean Square Residual:			
for simple second-order correction (Mplus variant)			SRMR	0.049	0.049	
Model test baseline model:			Weighted Root Mean Square Residual:			
Minimum Function Test Statistic	20789.799	12373.083	WRMR	1.094	1.094	
Degrees of freedom	45	45	Parameter Estimates:			
P-value	0.000	0.000	Information		Expected	
User model versus baseline model:			Standard Errors		Robust.sem	
Comparative Fit Index (CFI)	0.999	0.994				
Tucker-Lewis Index (TLI)	0.998	0.992				
Robust Comparative Fit Index (CFI)		NA				
Robust Tucker-Lewis Index (TLI)		NA				

Analyse factorielle confirmatoire (CFA)

Latent Variables:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
DAST =~						
DAST1001f	1.000				2.009	0.895
DAST1002f	0.699	0.106	6.568	0.000	1.404	0.814
DAST1003f	0.122	0.034	3.539	0.000	0.245	0.238
DAST1004f	0.771	0.119	6.503	0.000	1.549	0.840
DAST1005f	0.776	0.122	6.364	0.000	1.559	0.842
DAST1006f	1.326	0.259	5.111	0.000	2.663	0.936
DAST1007f	1.199	0.196	6.107	0.000	2.409	0.924
DAST1008f	0.743	0.117	6.363	0.000	1.492	0.831
DAST1009f	0.992	0.157	6.320	0.000	1.993	0.894
DAST1010f	0.766	0.121	6.311	0.000	1.538	0.838

Analyse factorielle confirmatoire (CFA)

Thresholds:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
DAST1001f t1	0.325	0.097	3.350	0.001	0.325	0.145
DAST1002f t1	1.445	0.101	14.290	0.000	1.445	0.838
DAST1003f t1	0.808	0.049	16.550	0.000	0.808	0.785
DAST1004f t1	1.553	0.114	13.566	0.000	1.553	0.842
DAST1005f t1	1.104	0.098	11.286	0.000	1.104	0.596
DAST1006f t1	1.146	0.164	6.993	0.000	1.146	0.403
DAST1007f t1	1.857	0.173	10.728	0.000	1.857	0.712
DAST1008f t1	1.729	0.119	14.478	0.000	1.729	0.962
DAST1009f t1	1.411	0.130	10.851	0.000	1.411	0.633
DAST1010f t1	1.626	0.116	14.014	0.000	1.626	0.886

Analyse factorielle confirmatoire (CFA)

Variances:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.DAST1001f	1.000				1.000	0.199
.DAST1002f	1.000				1.000	0.337
.DAST1003f	1.000				1.000	0.943
.DAST1004f	1.000				1.000	0.294
.DAST1005f	1.000				1.000	0.292
.DAST1006f	1.000				1.000	0.124
.DAST1007f	1.000				1.000	0.147
.DAST1008f	1.000				1.000	0.310
.DAST1009f	1.000				1.000	0.201
.DAST1010f	1.000				1.000	0.297
DAST	4.035	1.044	3.863	0.000	1.000	1.000

Analyse factorielle confirmatoire (CFA)

Scales y*:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
DAST1001f	0.446				0.446	1.000
DAST1002f	0.580				0.580	1.000
DAST1003f	0.971				0.971	1.000
DAST1004f	0.542				0.542	1.000
DAST1005f	0.540				0.540	1.000
DAST1006f	0.351				0.351	1.000
DAST1007f	0.383				0.383	1.000
DAST1008f	0.557				0.557	1.000
DAST1009f	0.449				0.449	1.000
DAST1010f	0.545				0.545	1.000

Test d'invariance (selon le sexe)

- En utilisant la librairie *lavaan*, nous avons aussi testé que le DAST est invariant au sexe (Millsap R. and Yun-Tein J. 2004).
- L'invariance d'une mesure au sexe veut dire que les différences de sexe sur les items sont expliqués uniquement par le facteur latent.
- La comparaison pour un modèle probit dichotomique se fait entre trois modèles.
 - Modèle configural (seuils égaux)
 - Modèle d'invariance faible (seuils égaux et coefficients égaux)
 - Modèle d'invariance strict (seuils égaux, coefficients égaux et variances égales). Ce modèle requiert la paramétrisation *theta* qui utilise un paramètre d'échelle variable et fixe la variance à 1.

Test d'invariance (selon le sexe)

```
CFA.DAST.SX.config <- '  
DAST =~ c( 1, 1)*DAST1001f + c(f21,f22)*DAST1002f +  
        c(f31,f32)*DAST1003f + c(f41,f42)*DAST1004f +  
        c(f51,f52)*DAST1005f + c(f61,f62)*DAST1006f +  
        c(f71,f72)*DAST1007f + c(f81,f82)*DAST1008f +  
        c(f91,f92)*DAST1009f + c(fX1,fX2)*DAST1010f  
DAST1001f | c(t011,t011)*t1  
DAST1002f | c(t021,t021)*t1  
DAST1003f | c(t031,t031)*t1  
DAST1004f | c(t041,t041)*t1  
DAST1005f | c(t051,t051)*t1  
DAST1006f | c(t061,t061)*t1  
DAST1007f | c(t071,t071)*t1  
DAST1008f | c(t081,t081)*t1  
DAST1009f | c(t091,t091)*t1  
DAST1010f | c(t101,t101)*t1  
DAST ~~ NA*DAST  
DAST ~ c(0,NA)*1  
DAST1001f ~~ c(1, 1)*DAST1001f  
DAST1002f ~~ c(1,NA)*DAST1002f  
DAST1003f ~~ c(1,NA)*DAST1003f  
DAST1004f ~~ c(1,NA)*DAST1004f  
DAST1005f ~~ c(1,NA)*DAST1005f  
DAST1006f ~~ c(1,NA)*DAST1006f  
DAST1007f ~~ c(1,NA)*DAST1007f  
DAST1008f ~~ c(1,NA)*DAST1008f  
DAST1009f ~~ c(1,NA)*DAST1009f  
DAST1010f ~~ c(1,NA)*DAST1010f  
'
```


Test d'invariance (selon le sexe)

```
CFA.DAST.SX.weak <- '  
DAST =~ c( 1, 1)*DAST1001f + c(f21,f21)*DAST1002f +  
        c(f31,f31)*DAST1003f + c(f41,f41)*DAST1004f +  
        c(f51,f51)*DAST1005f + c(f61,f61)*DAST1006f +  
        c(f71,f71)*DAST1007f + c(f81,f81)*DAST1008f +  
        c(f91,f91)*DAST1009f + c(fX1,fX1)*DAST1010f  
DAST1001f | c(t011,t011)*t1  
DAST1002f | c(t021,t021)*t1  
DAST1003f | c(t031,t031)*t1  
DAST1004f | c(t041,t041)*t1  
DAST1005f | c(t051,t051)*t1  
DAST1006f | c(t061,t061)*t1  
DAST1007f | c(t071,t071)*t1  
DAST1008f | c(t081,t081)*t1  
DAST1009f | c(t091,t091)*t1  
DAST1010f | c(t101,t101)*t1  
DAST ~~ NA*DAST  
DAST ~ c(0,NA)*1  
DAST1001f ~~ c(1, 1)*DAST1001f  
DAST1002f ~~ c(1,NA)*DAST1002f  
DAST1003f ~~ c(1,NA)*DAST1003f  
DAST1004f ~~ c(1,NA)*DAST1004f  
DAST1005f ~~ c(1,NA)*DAST1005f  
DAST1006f ~~ c(1,NA)*DAST1006f  
DAST1007f ~~ c(1,NA)*DAST1007f  
DAST1008f ~~ c(1,NA)*DAST1008f  
DAST1009f ~~ c(1,NA)*DAST1009f  
DAST1010f ~~ c(1,NA)*DAST1010f  
'
```

Test d'invariance (selon le sexe)

```
CFA.DAST.SX.strict <- '  
DAST =~ c( 1, 1)*DAST1001f + c(f21,f21)*DAST1002f +  
        c(f31,f31)*DAST1003f + c(f41,f41)*DAST1004f +  
        c(f51,f51)*DAST1005f + c(f61,f61)*DAST1006f +  
        c(f71,f71)*DAST1007f + c(f81,f81)*DAST1008f +  
        c(f91,f91)*DAST1009f + c(fX1,fX1)*DAST1010f  
DAST1001f | c(t011,t011)*t1  
DAST1002f | c(t021,t021)*t1  
DAST1003f | c(t031,t031)*t1  
DAST1004f | c(t041,t041)*t1  
DAST1005f | c(t051,t051)*t1  
DAST1006f | c(t061,t061)*t1  
DAST1007f | c(t071,t071)*t1  
DAST1008f | c(t081,t081)*t1  
DAST1009f | c(t091,t091)*t1  
DAST1010f | c(t101,t101)*t1  
DAST ~~ NA*DAST  
DAST ~ c(0,NA)*1  
DAST1001f ~~ c(1, 1)*DAST1001f  
DAST1002f ~~ c(1, 1)*DAST1002f  
DAST1003f ~~ c(1, 1)*DAST1003f  
DAST1004f ~~ c(1, 1)*DAST1004f  
DAST1005f ~~ c(1, 1)*DAST1005f  
DAST1006f ~~ c(1, 1)*DAST1006f  
DAST1007f ~~ c(1, 1)*DAST1007f  
DAST1008f ~~ c(1, 1)*DAST1008f  
DAST1009f ~~ c(1, 1)*DAST1009f  
DAST1010f ~~ c(1, 1)*DAST1010f  
'
```

Test d'invariance (selon le sexe)

Modèle configural (seuils égaux)

```
cfaDAST1 <- cfa(CFA.DAST.SX.config, DAST.df, group = "SEXE",  
               parameterization = "theta", estimator = "wlsmv")
```

Modèle d'invariance faible (seuils égaux et coefs (loadings) égaux).

```
cfaDAST2 <- cfa(CFA.DAST.SX.weak, DAST.df, group = "SEXE",  
               parameterization = "theta", estimator = "wlsmv")
```

Modèle d'invariance strict (seuils égaux, coefs égaux, vars égales)

```
cfaDAST3 <- cfa(CFA.DAST.SX.strict, DAST.df, group = "SEXE",  
               parameterization = "theta", estimator = "wlsmv")
```

Test d'invariance (selon le sexe)

```
> lavTestLRT(cfaDAST1, cfaDAST2,  
             method = "satorra.bentler.2010")
```

Scaled Chi Square Difference Test (method = "satorra.bentler.2010")

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
cfaDAST1	70			87.875			
cfaDAST2	79		116.308		13.584	9	0.1379

Test d'invariance (selon le sexe)

```
> lavTestLRT(cfaDAST2, cfaDAST3,  
             method = "satorra.bentler.2010")
```

Scaled Chi Square Difference Test (method = "satorra.bentler.2010")

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
cfaDAST2	79			116.31			
cfaDAST3	88			125.26	4.8066	9	0.8508

Validité de convergence

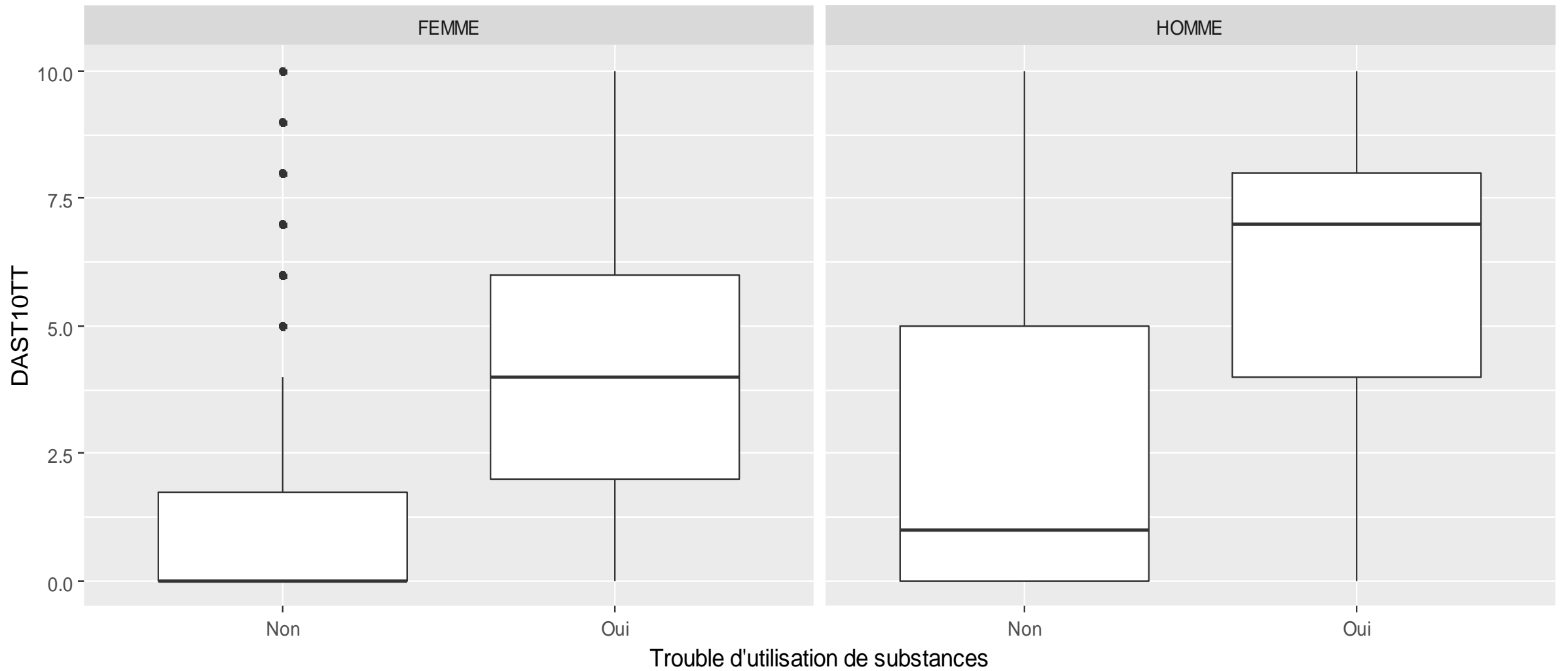
- Nous avons comparé le niveau du DAST sur l'indicateur d'abus de substance par le psychiatre.
- Le psychiatre a donné le diagnostic indépendamment de la réponse du patient au DAST.
- La comparaison a été faite chez les hommes et les femmes séparément.

Validité de convergence

Convergence entre le niveau du questionnaire.

```
ggplot(DAST.df, aes(y = DAST10TT,  
                   x = factor(CIMDR,  
                              labels = c("Non", "Oui")))) +  
  facet_grid(. ~ SEXE) + geom_boxplot() +  
  xlab("Trouble d'utilisation de substances")
```

Validité de convergence



Merci de votre attention

Équipe de la banque Signature
<http://CentreSignature.ca>



Merci à l'équipe de la banque Signature
Coordonnatrice:

Nathe François

Coordonnatrice du laboratoire:

Jannick Boissonneault

Technicienne du laboratoire:

Helen Findlay

Technicienne administrative:

Josée Feltrin

Infirmières:

Mimi Dumont

Carole Feltrin

Nabil Khiari

Françoise Martha Ulysse

Assistante de recherche:

Patricia Gravel

et à tous les participants à la Banque Signature