



Corrigé de l'examen CT de statistiques

Voir le fichier ROXY0.xls disponible sous spiral.

Pour plus de lisibilité tous les tableaux et les figures sont renvoyés à la fin du corrigé, à partir de la page 6.

Correction de l'exercice 1.

On étudie la variable catégorielle Q6_niveau. Voir le chapitre 3.

- Avec Rcommander, on peut dénombrer les catégories ; on trouve

Confirmée	Débrouillée	Débutante	Experte
126	134	42	16

Il y a en tout 4 catégories. Les fréquences associées sont :

Q6_niveau	Confirmée	Débrouillée	Débutante	Experte
	39.622642	42.138365	13.207547	5.031447

Voir le tableau 1 page 6.

- Avec Rcommander, on peut tracer le camembert et le diagramme en barre. Voir les figures 1 page 6 et 2 page 7.

Correction de l'exercice 2.

- On étudie la variable catégorielle Q5_lieu.

– Avec Rcommander, on peut dénombrer les catégories ; on trouve

Pistes	Pistes et Bords de pistes
254	64

Il y a en tout 2 catégories. Les fréquences associées sont :

Q5_lieu	Pistes	Pistes et Bords de pistes
	79.87421	20.12579

Voir le tableau 2 page 6.

– Avec Rcommander, on peut tracer le camembert et le diagramme en barre. Voir les figures 3 page 8 et 4 page 8.

- Croisons maintenant les deux variables catégorielles Q6_niveau et Q5_lieu.

Voir chapitre 7.

On définit le tableau croisé avec Rcommander.

Voir le tableau 4 page 7.

On obtient

Pearson's Chi-squared test

```
data: xtable2
```

```
X-squared = 31.6087, df = 3, p-value = 6.328e-07
```

On peut donc faire une statistique du chi-carré et on obtient $X^2 = 31.6087$ On obtient donc une taille d'effet

$$w = \sqrt{\frac{X^2}{n}} = \sqrt{\frac{31.6087}{318}} = 0.315275$$

Pour la probabilité critique, on obtient $p = 6.328e-07$.

La valeur de la taille d'effet w est égale à 0.3153 ; ainsi, au vu des seuils conventionnels proposés par Cohen (0.1,0.3,0.5), la liaison entre les deux variables peut être considérée comme forte. Par ailleurs, la probabilité critique est égale à 6.328e-07, inférieure au seuil de 0.05 ; ainsi, la liaison entre les deux variables peut être considérée comme statistiquement significative.

On conclut donc que le lieu de la pratique dépend du niveau de pratique.

Correction de l'exercice 3.

- On étudie la variable catégorielle Q8_renonce.

Avec Rcommander, on peut dénombrer les catégories ; on trouve

Mauvaise matériel	Mauvaise météo	Mauvaise neige	NA
15	189	34	4
Ne renonce jamais	Trop de monde		
52	24		

Il y a en tout 6 catégories. Les fréquences associées sont :

Q8_renonce

Mauvaise matériel	Mauvaise météo	Mauvaise neige	NA
4.716981	59.433962	10.691824	1.257862
Ne renonce jamais	Trop de monde		
16.352201	7.547170		

Voir le tableau 3 page 7.

- Comme dans l'exercice 3, on croise les deux variables catégorielles Q8_renonce et Q5_niveau.

On définit le tableau croisé avec Rcommander.

Voir le tableau 5 page 7.

On obtient

Pearson's Chi-squared test

```
data: xtable3
```

```
X-squared = 29.8586, df = 15, p-value = 0.01244
```

On peut donc faire une statistique du chi-carré et on obtient $X^2 = 29.8586$ On obtient donc une taille d'effet

$$w = \sqrt{\frac{X^2}{n}} = \sqrt{\frac{29.8586}{318}} = 0.306423$$

Pour la probabilité critique, on obtient $p = 0.01244$.

La valeur de la taille d'effet w est égale à 0.3064 ; ainsi, au vu des seuils conventionnels proposés par Cohen (0.1,0.3,0.5), la liaison entre les deux variables peut être considérée comme forte. Par

ailleurs, la probabilité critique est égale à 0.01244, inférieure au seuil de 0.05 ; ainsi, la liaison entre les deux variables peut être considérée comme statistiquement significative.

On conclut donc que le comportement dépend du niveau de pratique.

Correction de l'exercice 4.

On croise maintenant la variable catégorielle `Q6_niveau` définie dans l'exercice 1 et la variable numérique `âge`.

Voir le chapitre 5.

- (1) On peut tracer une collection de boîte de dispersion (par groupe) : voir figure 5 page 9.

On peut aussi faire des statistiques par groupe :

	mean	sd	0%	25%	50%	75%	100%	n	NA
Confirmée	37.73810	14.85082	11	24.00	39.0	46.00	75	126	0
Débrouillée	37.51880	13.64796	14	26.00	39.0	46.00	75	133	1
Débutante	36.73810	11.35902	15	29.25	36.0	43.75	58	42	0
Experte	40.68750	10.78715	13	37.00	40.5	45.50	63	16	0

- (2) Confirmons cela grâce à \mathbb{R} . En utilisant l'ajustement de modèle, on obtient un rapport de corrélation RC égal à 0.0031 et une probabilité critique égale à 0.8055.

La valeur du rapport de corrélation RC est égale à 0.003129 ; ainsi, au vu des seuils conventionnels proposés par Cohen (0.01,0.05,0.15), la liaison entre les deux variables peut être considérée comme faible. Par ailleurs, la probabilité critique est égale à 0.8055, supérieure au seuil de 0.05 ; ainsi, la liaison entre les deux variables peut être considérée comme statistiquement non significative.

On peut donc conclure qu'il n'y a pas de différence selon les pratiques, puisque la liaison est faible et statistiquement non significative.

Correction de l'exercice 5.

On étudie la variable catégorielle `Q29_txt_PCS`.

- Avec Rcommander, on la convertit tout d'abord en une variable catégorielle appelée `fact_Q29_txt_PCS`.

on peut alors dénombrer les catégories ; on trouve

```

1  2  5  31  33  34  35  37  38  42  43  45  46  47  48  52  53  54  55  56
1  18  1  14  11  16  4  48  18  13  20  4  8  4  4  14  4  22  10  3
61  73  81  84  85 NaN
1  1  2  64  10  3

```

Il y a en tout 26 catégories. Les fréquences associées sont :

fact_Q29_txt_PCS	1	2	5	31	33	34	35
	0.3144654	5.6603774	0.3144654	4.4025157	3.4591195	5.0314465	1.2578616
		37	38	42	43	45	46
	15.0943396	5.6603774	4.0880503	6.2893082	1.2578616	2.5157233	1.2578616
		48	52	53	54	55	56
	1.2578616	4.4025157	1.2578616	6.9182390	3.1446541	0.9433962	0.3144654
		73	81	84	85	NaN	
	0.3144654	0.6289308	20.1257862	3.1446541	0.9433962		

Voir le tableau 6 page 10.

- Avec Rcommander, on peut tracer le camembert et l'histogramme. Voir les figures 6 page 11 et 7 page 11.
- Ces deux graphes sont illisibles à cause du trop grand nombre de CSP. Comme indiqué dans le chapitre 3, on peut apporter des améliorations à cette analyse.

– On peut tracer un graphe en point en ordonnant les catégories en tapant :

```
dotchart(sort(table(fact_Q29_txt_PCS)))
```

On obtient le graphe de la figure 8 page 12, beaucoup plus lisible.

Sur ce graphe, on se rend compte que deux catégories sont beaucoup plus nettement représentées que les autres ; il s'agit des numéros 84 et 37, correspondant aux élèves étudiants et cadres administratifs et commerciaux d'entreprise.

– On pourrait donc mettre à part ces deux catégories et regrouper les autres par sous-groupes. Examinons de plus près le tableau 6 page 10. On peut dorénavant déjà conserver les catégories 84, 37 et NaN (correspondant aux non réponses, toujours à conserver). Rappelons les grandes catégories données dans l'énoncé :

- 1 Agriculteurs exploitants
- 2 Artisans, commerçants et chefs d'entreprise
- 3 Cadre et professions intellectuelles supérieures
- 4 Professions intermédiaires
- 5 Employés
- 6 Ouvriers
- 7 Retraités
- 8 Autres personnes sans activité professionnelle

On constate que les catégories 1, 2 et 5 apparaissent telles quelles sans précisions. On les conserve donc. Enfin, on regroupera les catégories restantes en les regroupant de la façon suivante :

- 3 Les cadres et professions intellectuelles supérieures sauf 37 définissent la catégorie 3
- 4 Les professions intermédiaires définissent la catégorie 4
- 5 Les employés restants définissent la catégorie 5
- 6 Les ouvriers définissent la catégorie 6
- 7 Les retraités définissent la catégorie 7
- 8 Les autres personnes sans activité professionnelle sauf 84 définissent la catégorie 8

Dans la fenêtre de commande, on pourra taper les instructions suivantes :

```
test <- logical(length(ROXY0$Q29_txt_PCS))
test <- !test
new_Q29_txt_PCS <- numeric(length(ROXY0$Q29_txt_PCS))
testlo <- is.nan(ROXY0$Q29_txt_PCS)
new_Q29_txt_PCS[testlo] <- NaN
test[testlo] <- FALSE
aconserver <- c(37, 84, 1, 2, 5)
for (i in 1:length(aconserver)) {
  testlo <- (ROXY0$Q29_txt_PCS == aconserver[i]) & test
  new_Q29_txt_PCS[testlo] <- aconserver[i]
  test[testlo] <- FALSE
}
```

```
}
new_Q29_txt_PCS[test] <- floor(ROXY0$Q29_txt_PCS[test]/10)
```

La variable `new_Q29_txt_PCS` contient ce que l'on cherchait !

– On fait comme précédemment avec cette nouvelle variable.

```
1  2  3  4  5  6  7  8 37 84 NaN
1 18 63 53 54  1  1 12 48 64  3
```

Il y a en tout 11 catégories. Les fréquences associées sont :

```
fact_Q29_txt_PCS
      1          2          3          4          5          6          7
0.3144654 5.6603774 19.8113208 16.6666667 16.9811321 0.3144654 0.3144654
      8          37          84          NaN
3.7735849 15.0943396 20.1257862 0.9433962
```

Voir le tableau 7 page 10.

Voir les figures 9, 10 et 11.

Ensemble des tableaux et des figures

	Q6_niveau
Confirmée	39.62
Débrouillée	42.14
Débutante	13.21
Experte	5.03

TAB. 1. Le tableau des fréquence de la variable niveau de pratique

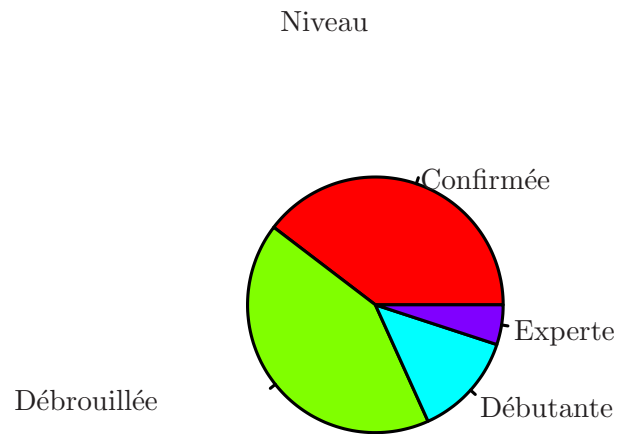


FIG. 1. La camembert de la variable niveau

	Q5_lieu
Pistes	79.87
Pistes et Bords de pistes	20.13

TAB. 2. Le tableau des fréquence de la variable lieu de pratique

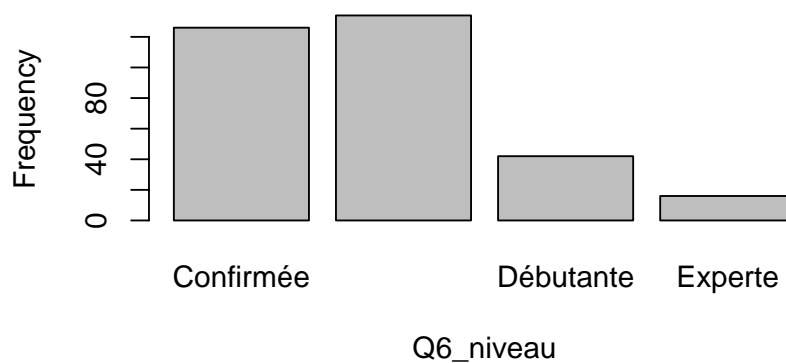


FIG. 2. le graphe en barre de la variable niveau

	Q8_renonce
Mauvaise matériel	4.72
Mauvaise météo	59.43
Mauvaise neige	10.69
NA	1.26
Ne renonce jamais	16.35
Trop de monde	7.55

TAB. 3. Le tableau des fréquence de la variable de renoncement

	Pistes	Pistes et Bords de pistes
Confirmée	85	41
Débrouillée	120	14
Débutante	40	2
Experte	9	7

TAB. 4. Le tableau croisé lieu/niveau

	Mauvaise matériel	Mauvaise météo	Mauvaise neige	NA	Ne renonce jamais	Trop de monde
Confirmée	2	82	11	0	24	7
Débrouillée	9	78	15	3	23	6
Débutante	4	19	5	1	4	9
Experte	0	10	3	0	1	2

TAB. 5. Le tableau croisé comportement de renoncement/niveau

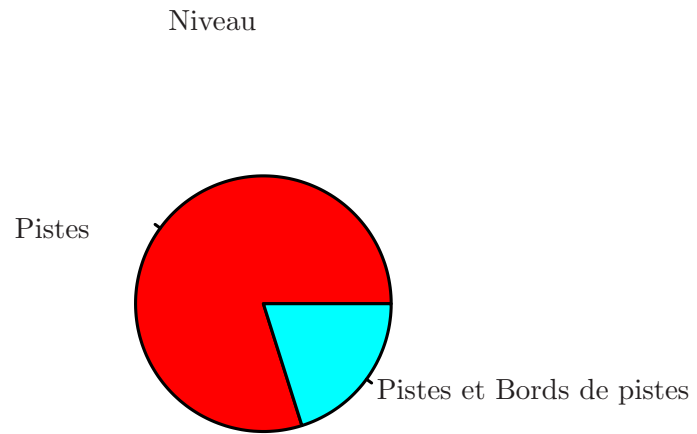


FIG. 3. La camembert de la variable lieu

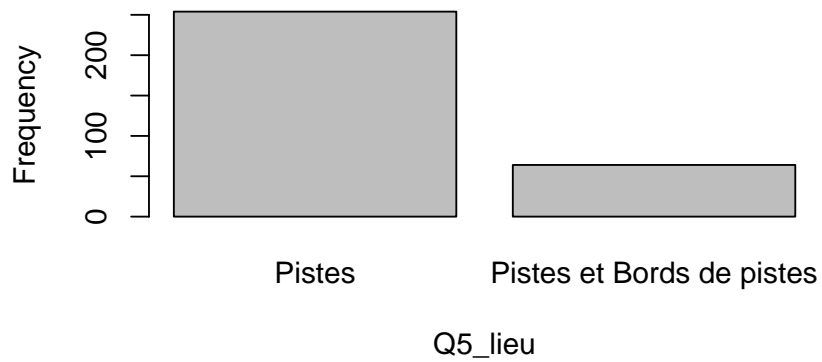


FIG. 4. le graphe en barre de la variable lieu

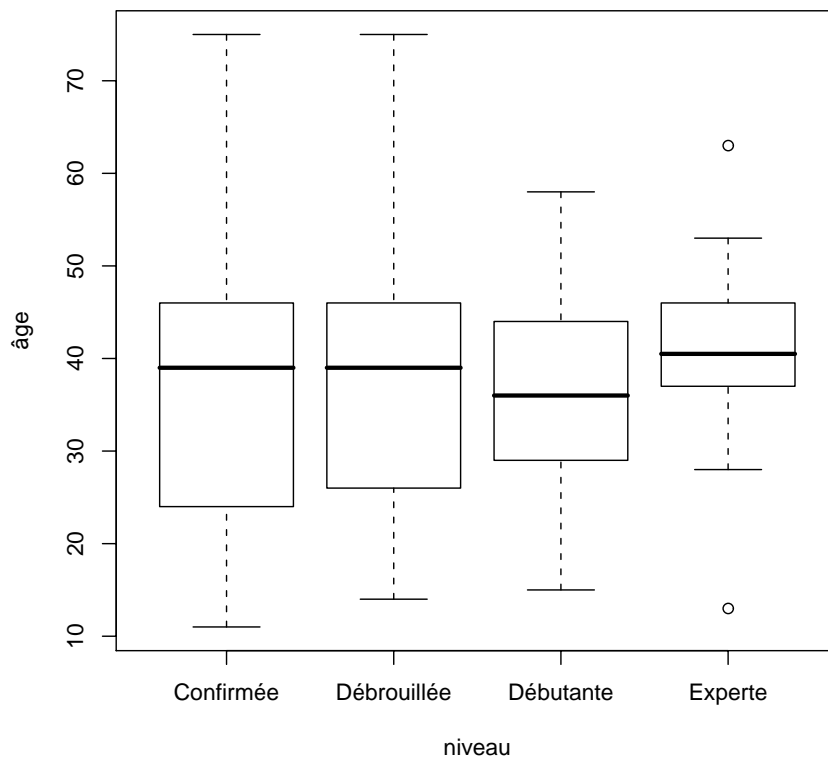


FIG. 5. collection de boîte de dispersion par groupe.

fact_Q29_txt_PCS	
1	0.31
2	5.66
5	0.31
31	4.40
33	3.46
34	5.03
35	1.26
37	15.09
38	5.66
42	4.09
43	6.29
45	1.26
46	2.52
47	1.26
48	1.26
52	4.40
53	1.26
54	6.92
55	3.14
56	0.94
61	0.31
73	0.31
81	0.63
84	20.13
85	3.14
NaN	0.94

TAB. 6. Le tableau des fréquence de la variable CSP

fact_Q29_txt_PCS	
1	0.31
2	5.66
3	19.81
4	16.67
5	16.98
6	0.31
7	0.31
8	3.77
37	15.09
84	20.13
NaN	0.94

TAB. 7. Le tableau des fréquence de la nouvelle variable simplifiée CSP

CSP

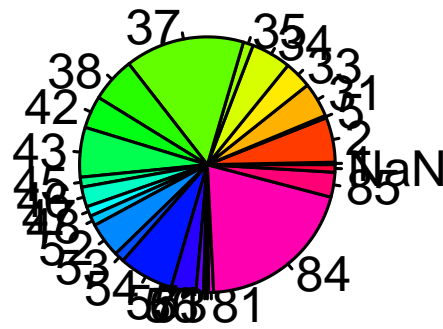


FIG. 6. La camembert de la variable CSP

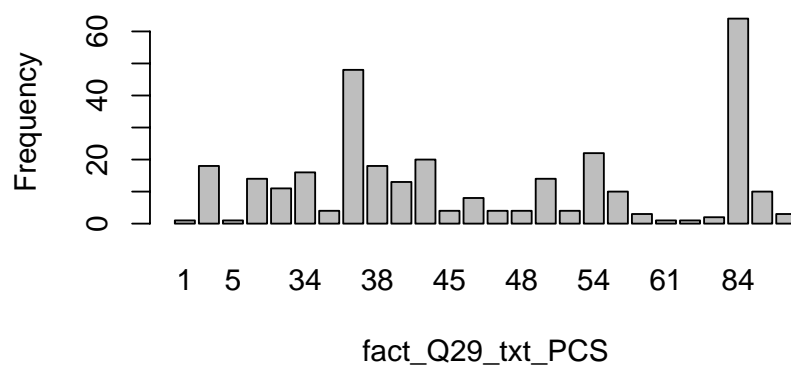


FIG. 7. le graphe en barre de la variable CSP

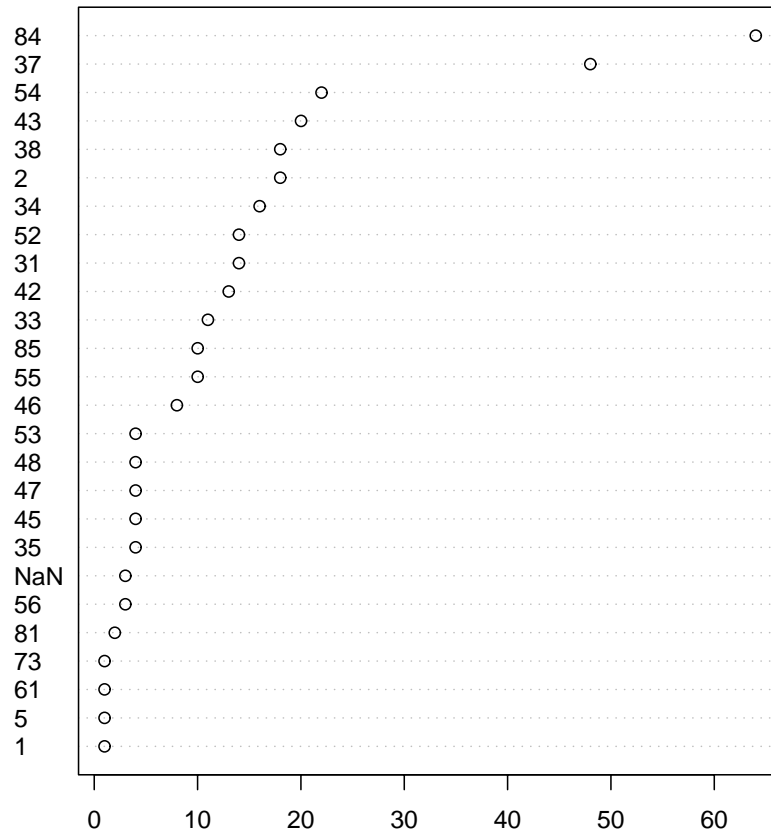


FIG. 8. le graphe en points de la variable CSP

CSP simplifiées

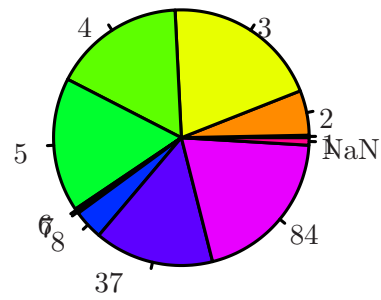


FIG. 9. La camembert de la nouvelle variable simplifiée CSP

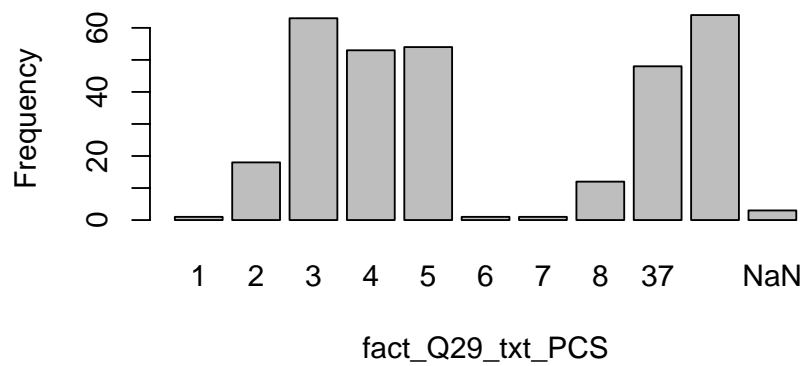


FIG. 10. le graphe en barre de la nouvelle variable simplifiée CSP

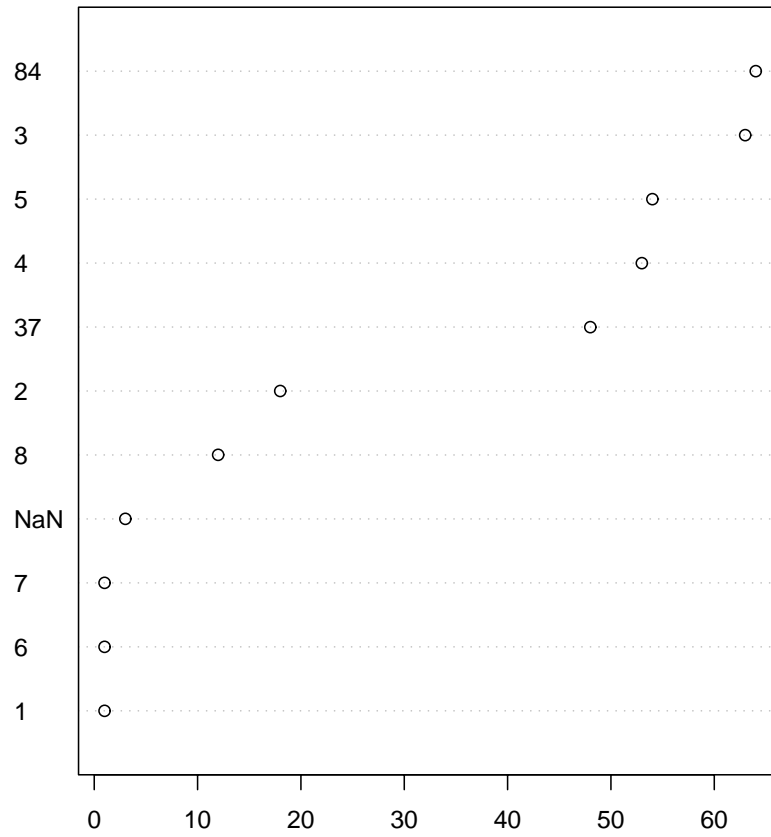


FIG. 11. le graphe en points de la variable simplifiée CSP