

Rapport de Recherche :
La Rotation des Axes en Analyse en Composantes
Principales (ACP)

ANASS BENAHMED ESBD N:11

OUALISSAOUI MOHAMMAD AF N: 31

18 novembre 2025

Table des matières

1	Introduction	3
2	Le Problème de l'Interprétabilité et la Structure Simple	3
2.1	La complexité des composantes brutes	3
2.2	Le concept de "Structure Simple"	3
3	Fondements Mathématiques de la Rotation	3
3.1	Ce qui est préservé et ce qui change	3
3.2	Types de Rotation	4
3.2.1	Rotations Orthogonales	4
3.2.2	Rotations Obliques	4
4	Implémentation et Cas Pratique : Analyse de la Satisfaction Client	4
4.1	Préparation des Données	4
4.2	Vérification de l'Adéquation des Données	5
4.3	Détermination du Nombre de Composantes à Retenir	6
4.3.1	Méthode 1 : Le Critère du Coude (Scree Plot)	6
4.3.2	Méthode 2 : L'Analyse Parallèle	7
4.3.3	Décision Finale sur le Nombre de Composantes	8
4.4	Extraction, Visualisation et Rotation des Composantes	8
4.4.1	Analyse des Composantes Principales SANS Rotation	8
4.4.2	Analyse des Composantes Principales SANS Rotation	9
4.4.3	Application de la Rotation Varimax et Interprétation	9
4.4.4	Analyse Biplot APRES Rotation (Variables et Individus)	10

1 Introduction

L'Analyse en Composantes Principales (ACP) est une technique fondamentale de réduction de dimensionnalité. Son objectif est de transformer un ensemble de variables corrélées en un nouvel ensemble de variables décorréées, appelées **composantes principales**. Ces composantes sont ordonnées de telle sorte que la première capture la plus grande part de la variance des données, la deuxième la plus grande part de la variance restante, et ainsi de suite.

Si l'ACP est mathématiquement optimale pour la compression de l'information (variance), elle présente une limite majeure : **l'interprétabilité**. Les composantes principales sont des combinaisons linéaires de *toutes* les variables d'origine, rendant leur signification concrète souvent difficile à cerner.

C'est ici qu'intervient la **rotation factorielle**. Il s'agit d'une technique post-ACP visant à transformer les axes factoriels pour obtenir une structure plus simple et plus facilement interprétable, sans altérer la qualité globale de la solution factorielle.

2 Le Problème de l'Interprétabilité et la Structure Simple

2.1 La complexité des composantes brutes

Une composante principale C_k est définie comme :

$$C_k = a_{k1}X_1 + a_{k2}X_2 + \dots + a_{kp}X_p \quad (1)$$

où les X_j sont les variables d'origine (centrées-réduites) et les a_{kj} sont les coefficients (loadings). En général, la plupart des coefficients a_{kj} sont non nuls, ce qui signifie que chaque composante est influencée par presque toutes les variables, rendant son interprétation ambiguë.

2.2 Le concept de "Structure Simple"

La rotation vise à atteindre une "structure simple" (concept de Thurstone, 1947), qui respecte les principes suivants :

1. Chaque variable doit avoir une forte corrélation (loading élevé) avec une seule composante et des corrélations faibles avec les autres.
2. Chaque composante doit être définie par un sous-ensemble de variables qui sont fortement corrélées avec elle, et peu ou pas avec les autres composantes.

Une telle structure facilite l'étiquetage et la compréhension de la signification sous-jacente de chaque composante.

3 Fondements Mathématiques de la Rotation

La rotation s'applique à la **matrice des saturations** (loadings), notée A , de dimensions $p \times m$, où p est le nombre de variables et m est le nombre de composantes retenues. Une rotation consiste à multiplier cette matrice A par une **matrice de rotation** T de dimensions $m \times m$ pour obtenir une nouvelle matrice de saturations A_{rot} :

$$A_{rot} = A \cdot T \quad (2)$$

3.1 Ce qui est préservé et ce qui change

Une propriété fondamentale de la rotation est qu'elle s'effectue au sein du sous-espace défini par les m composantes retenues.

- **Variance totale expliquée** : La somme des variances expliquées par les m composantes reste **inchangée** après rotation. La rotation ne fait que redistribuer cette variance entre les nouvelles composantes.
- **Communalités** : La communalité d'une variable (la part de sa variance expliquée par les m composantes) est **préservée**. Pour une variable j , la communalité h_j^2 est :

$$h_j^2 = \sum_{k=1}^m a_{jk}^2 = \sum_{k=1}^m (a_{rot})_{jk}^2 \quad (3)$$

- **Variance par composante** : La variance expliquée par *chaque* composante individuelle est modifiée. La première composante après rotation n'est plus celle qui maximise la variance ; elle est optimisée pour la simplicité de structure.

3.2 Types de Rotation

Il existe deux grandes familles de rotations :

3.2.1 Rotations Orthogonales

Elles contraignent les nouveaux axes à rester perpendiculaires (décorrélés) les uns par rapport aux autres. La matrice de rotation T est une matrice orthogonale, c'est-à-dire que $T'T = TT' = I$.

- **Varimax** : C'est la méthode la plus populaire. Elle vise à maximiser la variance des carrés des saturations *au sein de chaque colonne* de la matrice A_{rot} . Cela a pour effet de simplifier les composantes en poussant les loadings vers 0 ou ± 1 . Le critère Varimax V à maximiser est :

$$V = \sum_{k=1}^m \left[p \sum_{j=1}^p \left(\frac{a_{jk}^2}{h_j^2} \right)^2 - \left(\sum_{j=1}^p \frac{a_{jk}^2}{h_j^2} \right)^2 \right] \quad (4)$$

(Cette version est la Varimax "normalisée", qui pondère par les communalités).

3.2.2 Rotations Obliques

Elles autorisent les nouveaux axes à être corrélés entre eux. Elles sont utiles lorsque les facteurs sous-jacents sont théoriquement corrélés (par ex., anxiété et dépression).

- **Promax / Oblimin** : Ces méthodes tentent également d'atteindre une structure simple, mais en relâchant la contrainte d'orthogonalité. L'interprétation est plus complexe car il faut aussi analyser la matrice de corrélation entre les nouvelles composantes.

4 Implémentation et Cas Pratique : Analyse de la Satisfaction Client

Pour illustrer la démarche d'une ACP rotationnée, nous utilisons le jeu de données "*Airline Passenger Satisfaction*" de Kaggle. L'objectif est d'identifier les grandes dimensions structurantes de la satisfaction des passagers à partir d'une série de questions spécifiques.

4.1 Préparation des Données

La première étape consiste à charger, nettoyer et sélectionner les variables pertinentes pour notre analyse.

```
1 # --- 1. Chargement des bibliotheques ---
2 library(dplyr)
3 library(psych)
```

```

4 library(janitor)
5 library(ggplot2) # Pour les futurs graphiques
6
7 # --- 2. Chargement des donnees ---
8 # La methode file.choose() ouvre une fenetre pour selectionner le fichier
9 data_raw <- read.csv(file.choose())
10
11 # --- 3. Nettoyage des noms de colonnes ---
12 # Transforme "Inflight.wifi.service" en "inflight_wifi_service"
13 data_clean <- clean_names(data_raw)
14
15 # --- 4. Selection des variables de satisfaction ---
16 # On cree une liste des variables a conserver pour l'analyse
17 satisfaction_vars <- c(
18   "inflight_wifi_service", "departure_arrival_time_convenient",
19   "ease_of_online_booking", "gate_location", "food_and_drink",
20   "online_boarding", "seat_comfort", "inflight_entertainment",
21   "on_board_service", "leg_room_service", "baggage_handling",
22   "checkin_service", "inflight_service", "cleanliness"
23 )
24
25 # Creation du tableau de donnees final, propre et specialise
26 efa_data <- data_clean %>%
27   select(all_of(satisfaction_vars)) %>%
28   # On retire les notes '0' (non applicable) qui fausseraient les correlations
29   filter(across(everything(), ~ . != 0)) %>%
30   # On supprime les lignes avec des valeurs manquantes
31   na.omit()

```

Listing 1 – Chargement et préparation des données en R

4.2 Vérification de l'Adéquation des Données

Avant de procéder à l'ACP, il est indispensable de vérifier si les données sont "factorisables". Pour cela, nous utilisons deux tests statistiques : le test de KMO et le test de sphéricité de Bartlett.

```

> print(kmo_result)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = efa_data)
Overall MSA = 0.79
MSA for each item =

```

inflight_wifi_service	departure_arrival_time_convenient	ease_of_online_booking
0.77	0.73	0.72
gate_location	food_and_drink	online_boarding
0.73	0.84	0.73
seat_comfort	inflight_entertainment	on_board_service
0.83	0.77	0.83
leg_room_service	baggage_handling	checkin_service
0.89	0.82	0.73
inflight_service	cleanliness	
0.79	0.82	

FIGURE 1 – Caption

```

1 # Test de Kaiser-Meyer-Olkin (KMO)
2 # On attend un indice MSA (Measure of Sampling Adequacy) > 0.6
3 kmo_result <- KMO(efa_data)
4 print(kmo_result)
5
6 # Test de sphericite de Bartlett
7 # On attend une p-value tres faible (< 0.05)
8 bartlett_result <- cortest.bartlett(efa_data)
9 print(bartlett_result)

```

Listing 2 – Tests de KMO et de Bartlett

```

> print("Resultat du test de Bartlett :")
[1] "Resultat du test de Bartlett :"
> print(bartlett_result)
$chisq
[1] 574659.9

$ p.value
[1] 0

$df
[1] 91

```

FIGURE 2 – Enter Caption

Les résultats obtenus sont excellents :

- L'indice **KMO global (Overall MSA)** est de 0.79, ce qui est considéré comme "bon" et largement supérieur au seuil de 0.6.
- Le **test de Bartlett** est hautement significatif, avec une p-value de 0. Cela indique que la matrice de corrélation est significativement différente d'une matrice identité, confirmant la présence de corrélations structurées entre nos variables.

Les deux tests nous donnent donc le feu vert pour continuer l'analyse.

4.3 Détermination du Nombre de Composantes à Retenir

La décision la plus critique dans une ACP est de déterminer le nombre de composantes à extraire. Une extraction insuffisante peut masquer des structures importantes, tandis qu'une extraction excessive peut introduire du bruit statistique. Pour prendre une décision robuste, nous allons utiliser deux méthodes complémentaires : le critère du coude (Scree Plot), une méthode visuelle, et l'analyse parallèle, une méthode statistique objective.

4.3.1 Méthode 1 : Le Critère du Coude (Scree Plot)

```

1 # On effectue une ACP simple pour extraire les valeurs propres
2 acp_simple <- prcomp(efa_data, scale. = TRUE)
3 valeurs_propres <- acp_simple$sdev^2
4 scree_data <- data.frame(
5   Composante = 1:length(valeurs_propres),
6   Valeur_Propre = valeurs_propres
7 )
8
9 # On trace le graphique avec ggplot2
10 library(ggplot2)
11 ggplot(scree_data, aes(x = Composante, y = Valeur_Propre)) +
12   geom_line(color = "blue", size = 1) +
13   geom_point(color = "blue", size = 3) +
14   geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
15   labs(
16     title = " boulis  des Valeurs Propres (Crit re du Coude)",
17     subtitle = "La ligne rouge repr sente le seuil de Kaiser (valeur propre =
18     1)",
19     x = "Num ro de la Composante Principale",
20     y = "Valeur Propre (Variance expliqu e)"
21   ) +
22   scale_x_continuous(breaks = 1:length(valeurs_propres)) +
23   theme_minimal()

```

Listing 3 – Génération du Scree Plot en R

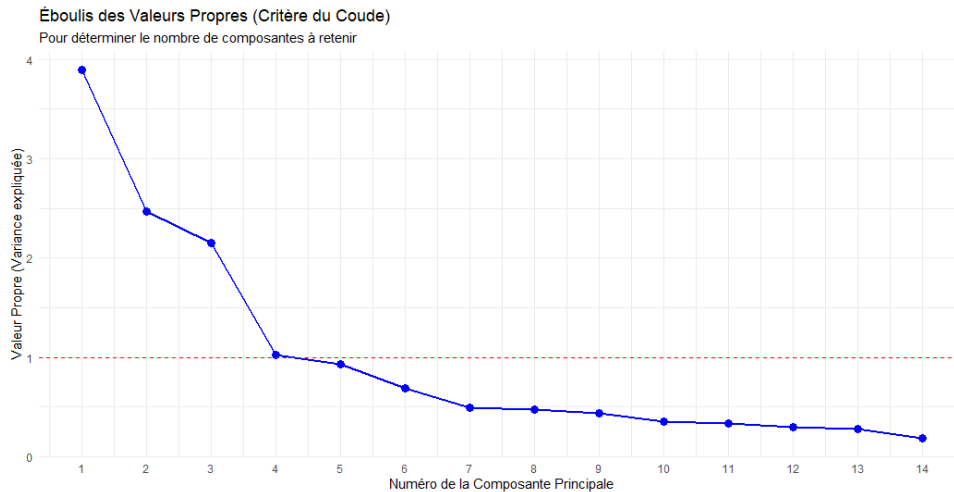


FIGURE 3 – Graphique du critère du coude (Scree Plot).

Interprétation du graphique (Figure 3) : L'éboulis des valeurs propres montre une forte décroissance pour les trois premières composantes. Un "coude" très net apparaît au niveau de la quatrième composante, où la courbe commence à s'aplatir. De plus, on observe que seules trois composantes ont une valeur propre supérieure à 1 (seuil de Kaiser, matérialisé par la ligne rouge). Les deux approches suggèrent donc de retenir **3 composantes**.

4.3.2 Méthode 2 : L'Analyse Parallèle

```
1 # Execution de l'analyse parallèle pour une ACP (pc = Principal Components)
2 parallel_analysis_acp <- fa.parallel(efa_data, fa = "pc")
```

Listing 4 – Analyse parallèle pour une ACP (fa="pc")

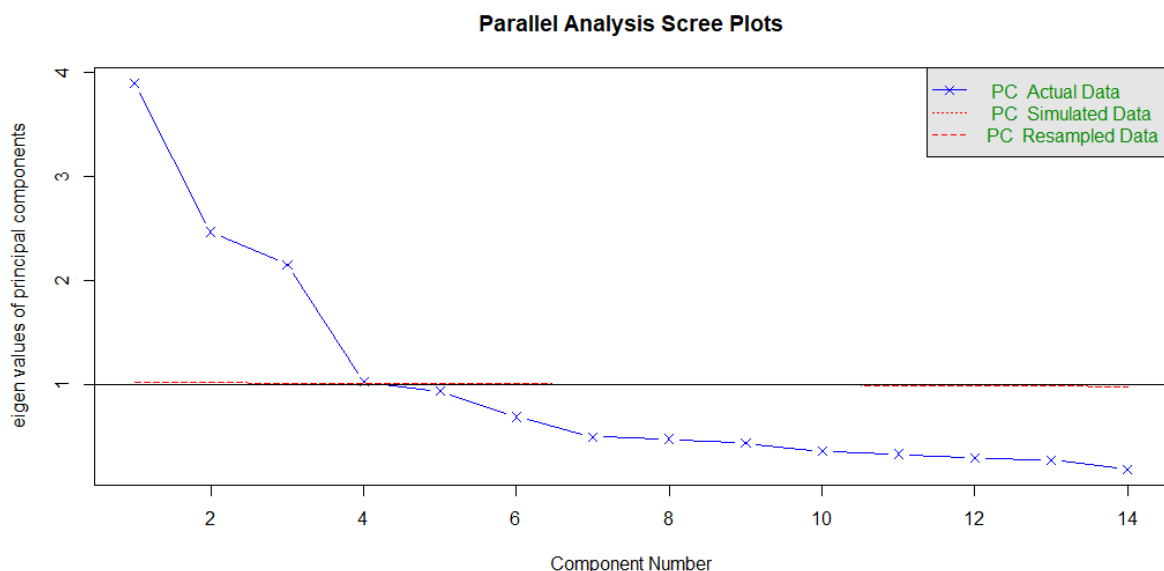


FIGURE 4 – Résultat de l'analyse parallèle.

Interprétation du graphique (Figure 4) : L'analyse parallèle confirme le résultat précédent de manière objective. On observe que les trois premières valeurs propres des données réelles

(courbe bleue) sont nettement supérieures à celles générées par le hasard (courbe rouge). La quatrième valeur propre, en revanche, est déjà en dessous de ce seuil.

4.3.3 Décision Finale sur le Nombre de Composantes

Les deux méthodes, l'une visuelle (critère du coude) et l'autre statistique (analyse parallèle), convergent vers la même conclusion : la structure des données est expliquée de manière optimale par **3 composantes principales**. Nous allons donc poursuivre l'analyse en extrayant et en rotationnant 3 composantes.

4.4 Extraction, Visualisation et Rotation des Composantes

Après avoir déterminé qu'il faut retenir 3 composantes, nous procédons à l'analyse en trois temps :

1. Nous visualisons d'abord la structure des composantes brutes, **sans rotation**, pour mettre en évidence le problème d'interprétabilité.
2. Nous appliquons ensuite une **rotation Varimax** pour obtenir une structure simple.
3. Enfin, nous interprétons les composantes rotationnées pour en extraire des insights stratégiques.

4.4.1 Analyse des Composantes Principales SANS Rotation

Nous commençons par réaliser une ACP standard et nous visualisons la projection des variables sur le premier plan factoriel (Composante 1 vs Composante 2).

```
1 # On utilise l'objet 'acp_simple' que nous avons déjà calculé
2 library(factoextra)
3 fviz_pca_var(acp_simple,
4             axes = c(1, 2),
5             repel = TRUE,
6             title = "Cercle de Corrélation AVANT Rotation")
```

Listing 5 – Visualisation de l'ACP brute avec factoextra

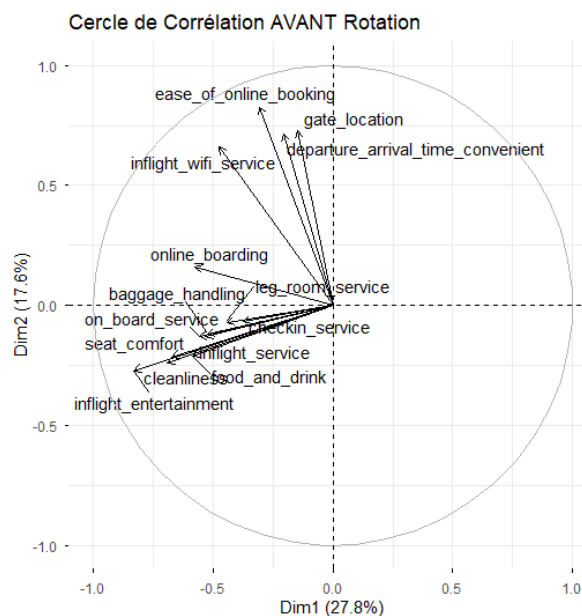


FIGURE 5 – Projection des variables sur le premier plan factoriel SANS rotation.

Interprétation du graphique (Figure 5) : Comme attendu, la structure brute est difficile à interpréter. Les variables ne s'alignent pas clairement sur les axes. Ce graphique illustre parfaitement la nécessité d'une rotation pour clarifier la structure.

4.4.2 Analyse des Composantes Principales SANS Rotation

Pour mettre en évidence le problème d'interprétabilité de l'ACP brute, nous visualisons la projection des variables et des individus sur le premier plan factoriel. En raison du grand nombre de passagers, nous affichons ces derniers sous la forme d'un nuage de points semi-transparent pour observer leur densité.

```
1 # On realise une ACP simple pour la visualisation
2 acp_simple <- prcomp(efa_data, scale. = TRUE)
3
4 # On genere le biplot avec la bibliotheque factoextra
5 library(factoextra)
6 fviz_pca_biplot(acp_simple,
7                 axes = c(1, 2), repel = TRUE,
8                 geom.ind = "point", pointshape = 21, pointsize = 1.5,
9                 col.ind = "gray50", alpha.ind = 0.3,
10                 col.var = "red2",
11                 title = "Biplot de l'ACP - AVANT Rotation")
```

Listing 6 – Biplot de l'ACP brute (avant rotation)

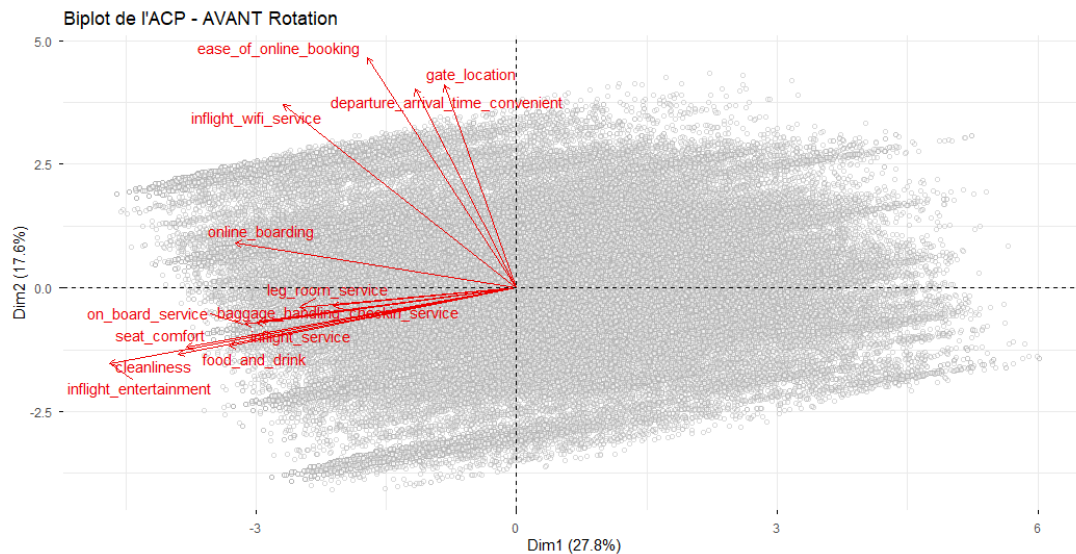


FIGURE 6 – Biplot sur les composantes brutes (AVANT rotation).

Interprétation du biplot (Figure 6) : Le biplot avant rotation confirme la difficulté d'interprétation. Le nuage de points représentant les passagers est dense et étiré en diagonale, ce qui montre une forte corrélation générale mais ne permet pas de distinguer des profils clairs le long des axes Dim.1 et Dim.2. Cette complexité visuelle démontre de manière éclatante la nécessité d'une rotation pour obtenir une structure simple et intelligible.

4.4.3 Application de la Rotation Varimax et Interprétation

Pour remédier au problème d'interprétabilité mis en évidence précédemment, nous appliquons une rotation orthogonale **Varimax** sur les 3 composantes extraites. L'objectif est de faire pivoter les axes pour qu'ils s'alignent au mieux avec les groupes de variables, simplifiant ainsi leur structure.

```

1 # Nombre de composantes a extraire
2 nombre_de_composantes <- 3
3
4 # Execution de l'ACP suivie de la rotation
5 acp_rottee_result <- principal(efa_data,
6                               nfactors = nombre_de_composantes,
7                               rotate = "varimax")
8
9 # Affichage des resultats (loadings)
10 print(acp_rottee_result$loadings, cutoff = 0.4, sort = TRUE)

```

Listing 7 – ACP avec rotation Varimax

La visualisation des variables sur le nouveau plan factoriel rotationné (Figure 7) montre une transformation spectaculaire.

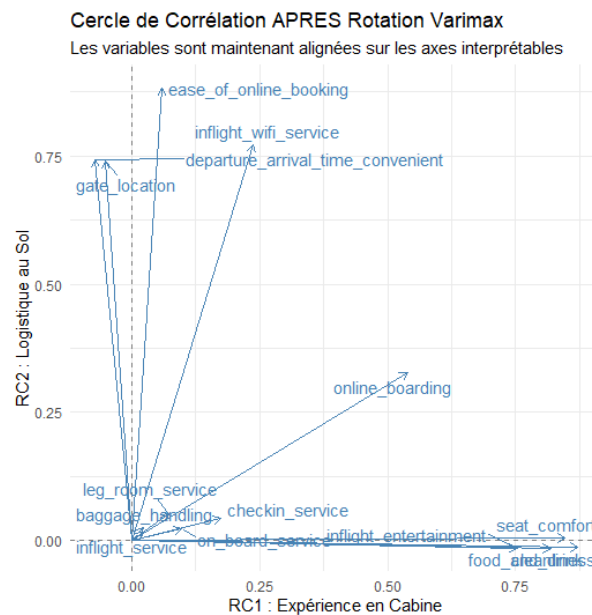


FIGURE 7 – Plan factoriel des variables APRES rotation Varimax.

4.4.4 Analyse Biplot APRES Rotation (Variables et Individus)

Pour achever notre analyse visuelle, nous superposons le nuage de points des passagers sur le plan factoriel rotationné. Ce biplot final est l'outil le plus puissant pour l'interprétation, car il relie directement les profils des individus aux dimensions de la satisfaction que nous avons identifiées.

```

1 # --- Preparation des donnees ---
2 scores_rottees_df <- as.data.frame(acp_rottee_result$scores)
3 loadings_rottees_df <- as.data.frame(unclass(acp_rottee_result$loadings))
4
5 # On calcule un facteur d'echelle pour la lisibilite
6 echelle_biplot <- max(abs(scores_rottees_df[,1:2])) / max(abs(loadings_rottees_df
7   [,1:2])) * 0.8
8
9 # --- Creation du Biplot avec ggplot2 ---
10 library(ggplot2); library(ggrepel)
11 ggplot() +
12   # Couche 1 : Le nuage de points des passagers (tres transparents)
13   geom_point(data = scores_rottees_df, aes(x = RC1, y = RC2),
14             color = "blue", alpha = 0.05) +
15   # Couche 2 : Les fleches pour les variables

```

```

15 geom_segment(data = loadings_rotees_df,
16             aes(x = 0, y = 0, xend = RC1 * echelle_biplot, yend = RC2 *
17               echelle_biplot),
18             arrow = arrow(length = unit(0.2, "cm")), color = "red2") +
19 # Couche 3 : Les etiquettes des variables
20 geom_text_repel(data = loadings_rotees_df,
21               aes(x = RC1 * echelle_biplot, y = RC2 * echelle_biplot, label
22                 = rownames(loadings_rotees_df)),
23               color = "red2", fontface = "bold") +
24 # Couche 4 : Les axes et les titres
25 geom_hline(yintercept = 0, linetype = "dashed", color = "gray50") +
26 geom_vline(xintercept = 0, linetype = "dashed", color = "gray50") +
27 coord_fixed(ratio = 1) +
28 labs(title = "Biplot APRES Rotation Varimax",
29       subtitle = "Projection des passagers et des variables sur les axes
30       interpr tables",
31       x = "RC1 : Exp rience en Cabine",
32       y = "RC2 : Logistique et Commodit ") +
33 theme_minimal()

```

Listing 8 – Génération du Biplot final après rotation

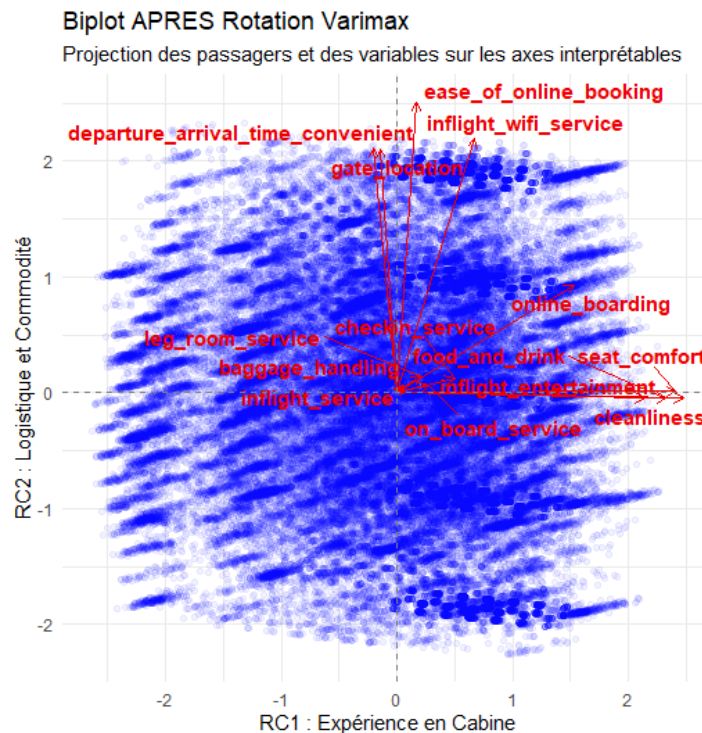


FIGURE 8 – Biplot final après rotation, superposant variables et passagers.

Analyse du biplot (Figure 8) : Le biplot final confirme la clarté de la structure obtenue après rotation. Le nuage de points des passagers (en bleu) est maintenant distribué le long des axes RC1 et RC2, qui ont une signification métier claire. Cette visualisation permet d'associer directement la position d'un passager (son profil) aux dimensions de la satisfaction (les flèches rouges). Par exemple, un passager situé à droite du graphique est une personne qui a fortement apprécié son expérience en cabine.