

REPUBLIQUE DU SENEGAL

Un Peuple - Un But - Une Foi

Agence nationale de la Statistique et de la démographie

Ecole nationale de la Statistique et de l'Analyse économique Pierre Ndiaye

ÉVALUATION

Projet statistique avec R

Rédigé par

Ameth FAYE

Élève Ingénieur Statisticien Économiste

Année académique 2024-2025

Sous la supervision de

M. Aboubacar HEMA

Data scientist

Table des matières

Table des matières	2
Liste des figures	3
Liste des tables	4
0. Entrée en matière	5
I- 0.1. Importation des packages nécessaires et définition des options globales	5
II- 0.2. Importation des bases de données	5
I. Analyse de consistance des bases de données	5
II. Analyse des données et calcul d'indicateurs	8
III- II.1. Analyse socio-démographique des ménages	8
IV- II.2. Score de consommation des ménages	9
IV-1. II.2.1. Analyse descriptive des variables qui composent le SCA	9
IV-2. II.2.2. Calcul du score de consommation	10
IV-3. II.3. Tableau des poids par groupe alimentaire	10
V- II.3. Indice réduit des stratégies de survie	12
V-1. II.3.1. Inspection des variables	12
V-2. II.3.2. Calcul de l'indice réduit des stratégies de survie (rCSI)	13
V-3. II.3.3. Tableau des poids du rCSI	13
V-4. II.3.4. Visualisation de la distribution du rCSI suivant les niveaux administratifs	13
VI- II.4. Stratégies d'adaptation aux moyens d'existence	15
VI-1. II.4.1. Inspection des variables	15
VI-2. II.4.2. Visualisation de la distribution du LHC-SI suivant les niveaux administratifs	16

Liste des figures

Figure : Carte du SCA moyen par région (Tchad)	11
Figure : Carte du rCSI moyen par région (Tchad)	14
Figure : Carte du rCSI moyen par département (Tchad).....	15
Figure : Carte de la catégorie LhCSI majoritaire par région (Tchad)	16

Liste des tables

Table : Synthèse descriptive des composantes du rCSI	12
Table : Poids attribués à chaque stratégie pour le calcul du rCSI.....	13

0. Entrée en matière

I- 0.1. Importation des packages nécessaires et définition des options globales

Dans cette partie, nous procéderons à l'importation des packages nécessaires et à la définition des options globales.

II- 0.2. Importation des bases de données

I. Analyse de consistance des bases de données

La base de données principale est la base de référence pour l'analyse de consistance. Elle contient 100 variables et 8950 observations. Quant à la MAD contient des informations supplémentaires sur les répondants et a 26 variables et 2206 observations. Nous allons voir les valeurs manquantes dans les deux bases de données.

Table 1: Synthèse des variables présentant des valeurs manquantes

Base de données	Variable	Nb de NA	% de NA
Principale	hhh_main_activity	8,950	100.0%
Principale	hhh_matrimonial	8,950	100.0%
Principale	fcs_pr_s_rf	8,950	100.0%
Principale	fcs_fruit_s_rf	8,677	96.9%
Principale	fcs_veg_s_rf	7,938	88.7%
Principale	fcs_fruit_org	7,687	85.9%
Principale	hh_source_income	7,238	80.9%
Principale	hdds_ch	5,659	63.2%
Principale	hdds_fruit_org	5,496	61.4%
Principale	hdds_pr_egg	5,353	59.8%
Principale	hdds_fruit_oth	5,198	58.1%
Principale	hdds_veg_org	4,877	54.5%
Principale	hdds_pr_meat_o	4,786	53.5%
Principale	fcs_dairy_s_rf	4,045	45.2%
Principale	hdds_pr_fish	3,504	39.2%
Principale	hdds_veg_gre	3,094	34.6%
Principale	hhh_edu	3,024	33.8%
Principale	hdds_pr_meat_f	2,728	30.5%
Principale	hdds_dairy	2,552	28.5%
Principale	fcs_pulse_s_rf	2,516	28.1%
Principale	hdds_veg_oth	2,348	26.2%
Principale	fcs_veg_gre	2,345	26.2%
Principale	fcs_veg_org	2,281	25.5%
Principale	fcs_pr_meat_f	1,749	19.5%
Principale	fcs_pr_meat_o	1,749	19.5%
Principale	fcs_pr_fish	1,749	19.5%
Principale	fcs_pr_egg	1,749	19.5%
Principale	hdds_pulse	1,683	18.8%
Principale	fcs_sugar_s_rf	1,150	12.8%
Principale	hdds_sugar	778	8.7%
Principale	fcs_cond_s_rf	609	6.8%

Base de données	Variable	Nb de NA	% de NA
Principale	fcs_fat_s_rf	569	6.4%
Principale	hdds_cond	348	3.9%
Principale	hdds_fat	311	3.5%
Principale	sers_avertissement_even	68	0.8%
Principale	fcs_stap_s_rf	54	0.6%
Principale	sers_preparer_future	52	0.6%
Principale	hdds_stap_cer	29	0.3%
Principale	hdds_stap_root	29	0.3%
MAD	ever_breast_f	7	0.3%

Après diagnostic, nous avons constaté que la base mad ne contient qu’une variable avec des valeurs manquantes (c’est la variable `ever_breast_f`). Par contre, ce n’est pas le cas pour la base principale. En effet, les variables `hhh_main_activity` “activité principale du chef de ménage”, `hhh_matrimonial` “situation matrimoniale du chef de ménage” et `fcs_pr_s_rf` “source principale d’approvisionnement pour la consommation de viande, poisson et œufs” sont entièrement vides. Elles sont suivies par `fcs_fruit_s_rf` ((source principale d’approvisionnement pour la consommation de fruits)) : 96,9 % de NA; `fcs_veg_s_rf` (source principale d’approvisionnement pour la consommation de légumes) : 88,7 % de NA; `fcs_fruit_org` (nombre de jours de consommation de fruits riches en vitamine A) : 85,9 % de NA; `hh_source_income` (source principale de revenu du ménage) : 80,9 % de NA. Pour un début, nous allons procéder uniquement à la suppression des variables qui sont complètement vides et après, suivant les questions des prochaines parties du devoir, nous nous ajusterons aux nécessités de corrections des valeurs manquantes.

Nous allons passer maintenant à la partie relative à l’Analyse des données et calcul d’indicateurs

II. Analyse des données et calcul d’indicateurs

III- II.1. Analyse socio-démographique des ménages

L’objectif ici c’était de bien identifier les noms des colonnes et de voir celles que nous allons utiliser pour la synthèse socio-démographique.

Table 2: Synthèse socio-démographique des ménages

Variable	Statistique
Taille ménage (moyenne \pm SD)	7.5 \pm 12.9
Âge chef de ménage (moyenne \pm SD)	42.9 \pm 12.6
Proportion d'hommes (chef de ménage)	44 % (hommes)
Niveau d'éducation primaire (chef de ménage)	30.8 % primaire
Région la plus fréquente	Batha
District le plus fréquent	Mamdi
Source de revenu la plus fréquente	1

Pour bien analyser la sortie, il me faut d'abord aller chercher la source de revenu qui correspond à la valeur 1.

```
## [1] "Agriculture vivrière/ou de rente (hors maraîchage)"
```

Maintenant, je vais passer à l'a partie l'Analyse de la sortie du tableau de synthèse socio-démographique des ménages :

- Taille du ménage : en moyenne 7,5 membres (écart-type 12,9) – la dispersion élevée reflète la présence de quelques ménages très nombreux.
- Âge du chef de ménage : moyenne 42,9 ans (écart-type 12,6), avec une distribution classique autour de la quarantaine.
- Sexe du chef de ménage : 44 % sont des hommes (donc 56 % de femmes), ce qui suggère une proportion relativement élevée de ménages à tête féminine.
- Niveau d'éducation (primaire) : seuls 30,8 % des chefs de ménage ont un niveau d'études primaires, indiquant un faible taux de scolarisation formelle.
- Répartition géographique : la région de Batha est la plus représentée, et au sein de celle-ci, le district de Mamdi concentre le plus grand nombre de ménages
- Source de revenu : Agriculture vivrière/ou de rente (hors maraîchage (1) est la source de revenu la plus fréquente.

IV- II.2. Score de consommation des ménages

IV-1. II.2.1. Analyse descriptive des variables qui composent le SCA.

Nous remarquons que : Céréales (6,7 \pm 1,0 jours) Les céréales sont consommées presque quotidiennement (en moyenne 6,7 jours sur 7) avec très peu de dispersion (SD = 1), ce qui

confirme qu'elles constituent l'aliment de base de la majorité des ménages. La consommation de légumes est modérée et très hétérogène : certains ménages les prennent presque tous les jours, d'autres quasiment jamais (SD élevée). En moyenne, les familles consomment ces protéines 2 à 3 jours par semaine. Là encore, la variabilité est importante, signe d'inégalités d'accès ou de prix. Légèrement moins fréquentes que les protéines animales, les légumineuses sont prises en moyenne 2 jours/semaine, avec une dispersion comparable à celle des protéines. Les produits laitiers sont consommés très occasionnellement et de façon très inégale (SD > moyenne), ce qui peut refléter un coût élevé ou une moindre tradition de consommation. La presque absence de fruits (0,3 jour/semaine) indique un risque potentiel de carence en vitamines, notamment Vitamine C et fibres.

IV-2. II.2.2. Calcul du score de consommation

IV-3. II.3. Tableau des poids par groupe alimentaire

Le poids des condiments est mis à 0 car ils sont considérés comme des additifs et n'ont pas d'impact significatif sur la sécurité alimentaire. En revanche, les autres groupes alimentaires ont des poids qui reflètent leur importance dans l'alimentation quotidienne (pour le cas spécifique du Tchad).

SCA moyen par région (Tchad)

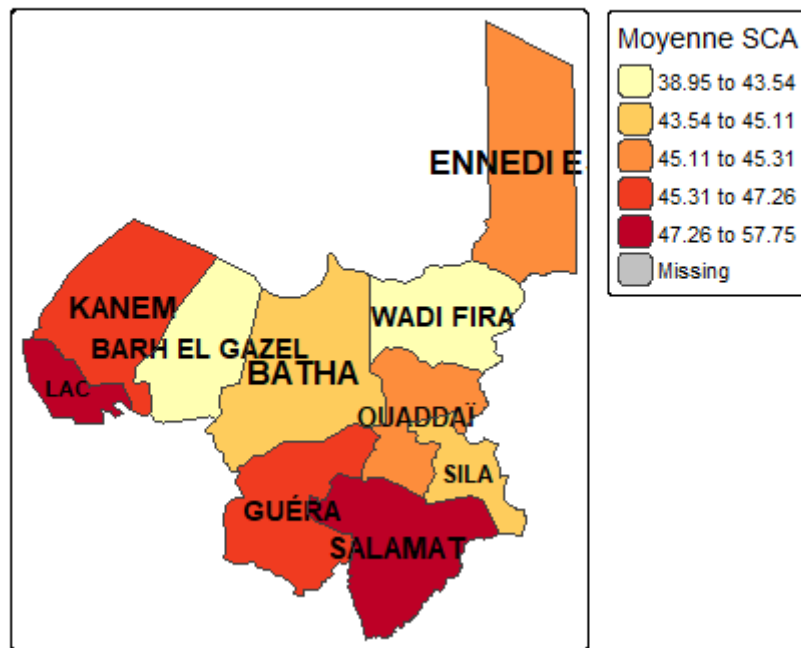


Figure 1: Carte du SCA moyen par région (Tchad)

Analyse de la carte – SCA moyen par région (Tchad)

Écarts marqués entre les régions

Salamat se distingue avec le score le plus élevé (entre 47,3 et 57,8), signe d'une diversité alimentaire et d'une fréquence de consommation maximales.

Guéra et Lac suivent dans la classe supérieure (45,3–47,3), indiquant eux aussi un accès relativement aisé à la plupart des groupes alimentaires.

Zones à consommation modérée

Kanem, Sila, Ennedi Est et Ouaddaï présentent des scores intermédiaires (45,1–45,3), traduisant une consommation régulière mais un peu moins systématique de certains aliments (notamment légumes et produits laitiers).

Régions les plus vulnérables

Batha et Barh el Gazel, avec des scores dans la catégorie la plus basse (38,9–43,5), montrent une moindre fréquence de consommation de groupes clés (fruits, protéines, produits laitiers), ce qui peut refléter des difficultés d'accès ou de pouvoir d'achat.

Tendances géographiques

Un gradient sud–nord apparaît, le sud (Salamat, Guéra) bénéficiant de meilleures ressources ou d'une production locale plus diversifiée, tandis que le centre–nord (Batha, Barh el Gazel) reste en retrait.

Les régions périphériques (Lac à l'ouest, Ennedi Est à l'est) se situent autour de la moyenne nationale.

Conclusion : La carte met en évidence de fortes disparités régionales : les ménages du sud du Tchad jouissent d'un régime plus diversifié et fréquent, alors que ceux du centre-nord semblent plus exposés aux contraintes qui limitent leur consommation alimentaire. Ces résultats peuvent guider des interventions ciblées (amélioration des marchés, renforcement des filières locales) pour réduire ces écarts.

V- II.3. Indice réduit des stratégies de survie

Là nous allons faire une inspection des variables qui composent l'indice réduit des stratégies de survie (IRSS). Nous allons aussi faire une analyse descriptive de ces variables.

V-1.II.3.1. Inspection des variables

Table 1 : Synthèse descriptive des composantes du rCSI

Stratégie comportementale	Moyenne \pm SD (jours)
Réduction de la qualité des repas	1.1 \pm 1.5
Emprunt de nourriture/argent	1.1 \pm 1.4
Réduction de la taille des repas	0.7 \pm 1.1
Adultes réduisent les repas	0.4 \pm 0.9
Réduction du nombre de repas	0.6 \pm 1.1

Stratégies les plus fréquentes : -Réduction de la qualité des repas et Emprunt de nourriture/argent sont les deux stratégies en moyenne les plus utilisées : 1,1 jour sur 7 chacune. Cela signifie qu'un ménage typique a dû baisser la qualité de ses repas ou emprunter au moins une fois par semaine.

-Variabilité importante Les écarts-types sont presque aussi élevés que les moyennes (1,4–1,5 jour), ce qui indique une très grande hétérogénéité : certains ménages recourent fréquemment à ces stratégies, d'autres jamais. Cela reflète sans doute des différences d'accès aux ressources ou de capacité d'emprunt.

-Stratégies moins fréquentes Les actes plus « extrêmes » – réduire la taille des repas (0,7 \pm 1,1 jour), réduire le nombre de repas (0,6 \pm 1,1 jour) ou faire sauter le repas des adultes (0,4 \pm 0,9 jour) – sont en moyenne moins utilisés. Ces stratégies sont donc considérées comme ultimes,

prises en œuvre par moins de ménages ou moins régulièrement. Par conséquent, le rCSI pondère ces stratégies ; ces moyennes suggèrent un stress alimentaire modéré à élevé selon les ménages.

V-2.II.3.2. Calcul de l'indice réduit des stratégies de survie (rCSI)

min_rCSI	max_rCSI	mean_rCSI	sd_rCSI
0	56	5.8	7.5

V-3.II.3.3. Tableau des poids du rCSI

Table 2 : Poids attribués à chaque stratégie pour le calcul du rCSI

Stratégie	Poids
Réduction qualité des repas	1
Emprunt nourriture/argent	2
Réduction taille des repas	1
Adultes sautent repas	3
Réduction du nombre de repas	1

V-4.II.3.4. Visualisation de la distribution du rCSI suivant les niveaux administratifs

Nous allons commencer par les régions

rCSI moyen par région (Tchad)

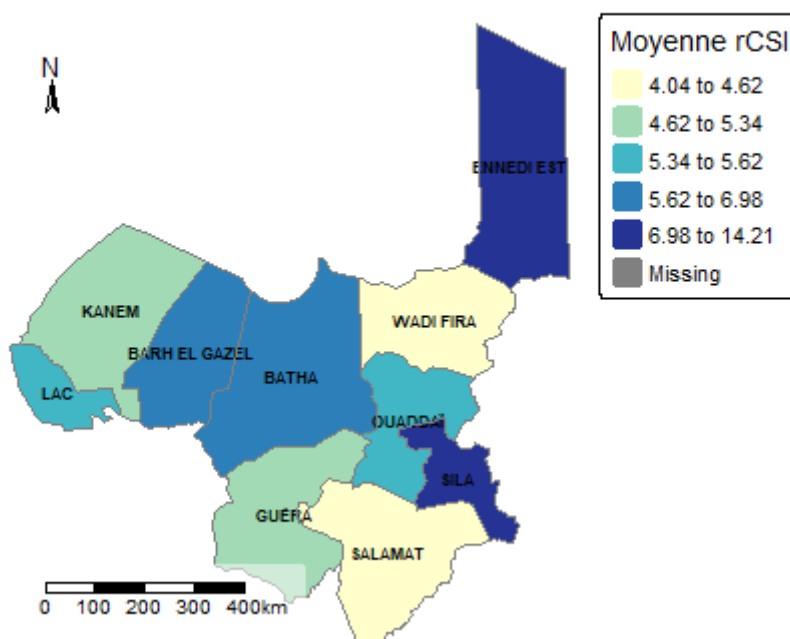


Figure 2: Carte du rCSI moyen par région (Tchad)

Analyse de la carte – rCSI moyen par région (Tchad)

Régions les plus exposées (rCSI élevé)

Ennedi Est se détache nettement avec le score moyen le plus élevé (entre 6,98 et 14,21), signe que certains ménages y recourent fréquemment à des stratégies de survie consommatrices de capital (stress, crise et urgence).

Sila figure également dans la classe supérieure (6,98–14,21), suggérant là aussi un usage intensif de stratégies d'adaptation sévères.

Niveaux intermédiaires

Batha, Barh el Gazel et Ouaddaï se situent dans la tranche 5,62–6,98. Ces régions montrent un recours régulier aux stratégies de survie, mais pas aussi systématique qu'à Ennedi Est ou Sila. Lac et Guéra appartiennent à la classe 5,34–5,62, indiquant un niveau modéré de rCSI, donc un recours ponctuel aux stratagèmes de survie.

Régions moins affectées

Kanem (4,62–5,34) et Wadi Fira (4,04–4,62) présentent les scores moyens les plus bas. Les ménages y recourent moins souvent ou moins intensément aux stratégies d'adaptation, témoignant d'une meilleure résilience ou de chocs moins fréquents.

Implications géographiques

Un gradient est-ouest/nord-sud se dessine : l'extrême nord-est et le sud-est subissent le stress le plus fort, tandis que l'ouest et le centre-est restent relativement épargnés.

Les zones à rCSI élevé (Ennedi Est, Sila) devraient être prioritaires pour un renforcement des filets de sécurité et un soutien au capital productif, afin d'éviter le basculement en stratégies de crise ou d'urgence.

Conclusion : La carte souligne de vives disparités régionales dans l'intensité des stratégies de survie basées sur la consommation. Les régions d'Ennedi Est et de Sila semblent particulièrement vulnérables, nécessitant des interventions ciblées pour restaurer le pouvoir d'achat et limiter l'érosion des actifs ménagers. Nous allons passer à la représentation suivant les départements du Tchad

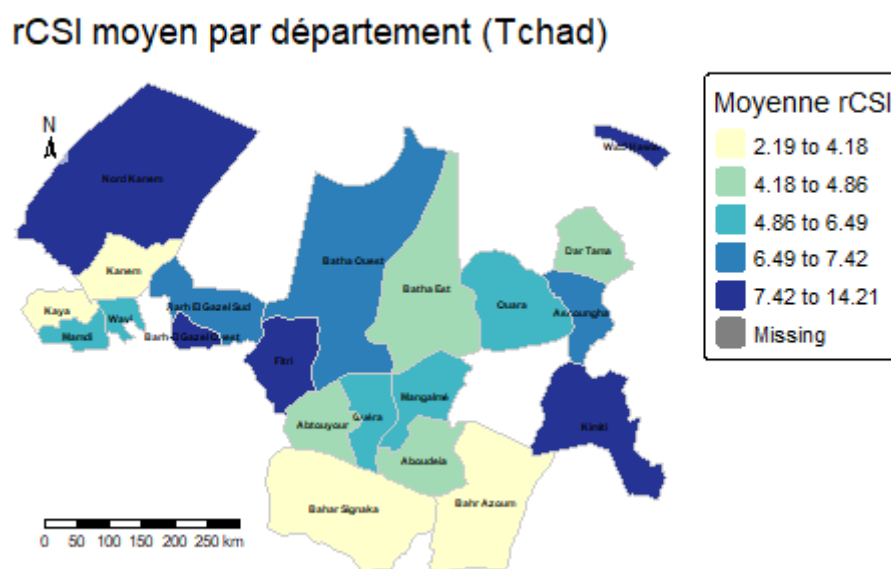


Figure 3: Carte du rCSI moyen par département (Tchad)

Nous allons passer aux stratégies d'adaptation aux moyens d'existence

VI- II.4. Stratégies d'adaptation aux moyens d'existence

VI-1. II.4.1. Inspection des variables

Remarque : on affiche ici le nombre et pourcentage de réponses pour chaque modalité de chaque stratégie (1=Non confronté; 2=Stratégie épuisée; 3=Oui; 4=Non applicable)

Maintenant, nous allons passer à la représentation graphique de la distribution des stratégies d'adaptation aux moyens d'existence suivant les niveaux administratifs.

VI-2. II.4.2. Visualisation de la distribution du LhCSI suivant les niveaux administratifs

LhCSI majoritaire par région (Tchad)

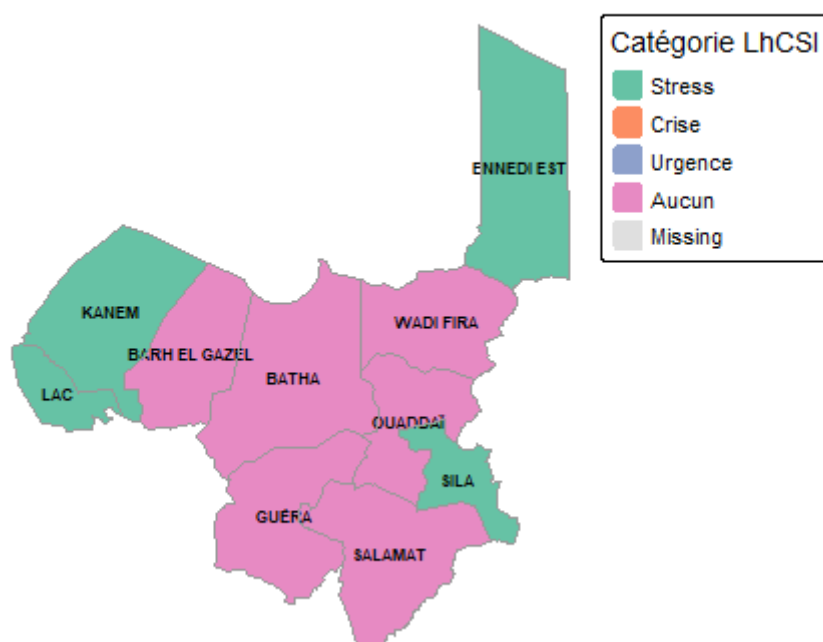


Figure 4: Carte de la catégorie LhCSI majoritaire par région (Tchad)

Analyse de la carte – LhCSI majoritaire par région (Tchad)

Prépondérance de « Aucun » Dans la majorité des régions (Barh el Gazel, Batha, Guéra, Ouaddaï, Wadi Fira, Salamat), la catégorie « Aucun » domine. Cela signifie que la plupart des ménages n'ont pas eu à recourir à des stratégies d'adaptation, ni en mode stress, ni en crise, ni en urgence au cours des 30 derniers jours ou 12 derniers mois.

Zones sous « Stress » Quatre régions sont classées en stress : Kanem, Lac, Ennedi Est et Sila. Ces ménages se sont appuyés principalement sur des stratégies initiales (épuisement d'épargne, vente d'actifs simples, emprunts) sans basculer vers des mesures plus drastiques.

Absence de « Crise » et d'« Urgence » Aucune région n'est passée aux stratégies de crise (vente d'actifs productifs, réduction des dépenses essentielles) ni d'urgence (vente de biens ultimes, mendicité, pillage). Cela suggère qu'aucun ménage n'a été confronté à un effondrement total de ses moyens d'existence.

Implications géographiques – Les régions périphériques (ouest et est) sont plus fréquemment sous stress, ce qui indique une vulnérabilité accrue aux chocs climatiques ou économiques. – Le centre du pays paraît plus résilient, avec très peu ou pas de recours aux stratégies d'adaptation.

Conclusion Le comportement des ménages tchadiens montre une capacité de résistance notable : ils limitent pour l'instant leur réponse aux seules stratégies de stress. Toutefois, les régions identifiées en stress nécessitent une attention et un soutien ciblés pour éviter qu'elles ne glissent vers des stratégies de crise ou d'urgence, plus difficiles à inverser.