



DEUX MESURES RÉPÉTÉES SUR UNE VARIABLE BINAIRE

ANALYSER CHAQUE MESURE

ÉVOLUTION DES RÉPONSES ENTRE LES DEUX MESURES RÉPÉTÉES

Mots-clés : Variable binaire, Mesures répétées, Tableau de contingence, Tri croisé, Test du Khi^2 de McNémar ; Profils de réponses.

UN EXEMPLE : LE DOSSIER DEBAT¹

On s'intéresse sur l'effet d'un débat sur la popularité d'un candidat à une élection. On tire au hasard un échantillon de 280 électeurs. On les interroge, avant et après le débat, sur leur opinion sur ce candidat, favorable (Favor.) ou défavorable (Défav.).

On considérera qu'un écart de 10 points de %, dans un sens ou dans l'autre, est un écart important

Le tableau de données pourrait être présenté sous la forme d'un tableau Individus x Variables. Ce tableau comporterait

- 280 lignes : une par électeur,
- 2 colonnes : la réponse AVANT le débat (Favor./Défav.) et la réponse après (Favor./Défav.).

Tableau 1 : Données DEBAT (données individuelles)

PROFILS	AVANT	APRES	POIDS
s001	Favor.	Défav.	1
s002	Favor.	Défav.	1
s003	Défav.	Favor.	1
s004	Favor.	Favor.	1
(...)	(...)	(...)	(...)
s279	Défav.	Favor.	1
s280	Défav.	Défav.	1

Il y a donc, potentiellement, quatre *profils de réponses* possibles :

- Favorable avant et après (Favor., Favor),
- Favorable avant et Défavorable après (Favor., Défav.),
- Défavorable avant et Favorable après (Défav., Favor.),
- Défavorable avant et après (Défav., Défav.).

Saisie informatique des données

Lorsque les données individuelles comportent d'autres variables (Sexe, Ville, Age...) le tableau de données sera saisi sous la forme du Tableau 1 (cf. ci-dessus). Mais lorsque les données ne comportent que ces deux variables, il sera plus commode de les saisir sous la forme du Tableau 2.

Ce tableau ne comprend que 4 lignes. Il regroupe dans une même ligne tous les individus ayant le même profil de réponse. On y indique les effectifs de chacun des quatre profils observés.

¹ Il s'agit de données fictives.



Deux mesures répétées sur une variable binaire
Le dossier DEBAT
Comparer les deux mesures répétées

Tableau 2 : Données DEBAT (liste des profils et effectifs)

PROFILS	AVANT	APRES	EFFECTIF
p01	Favor.	Favor.	102
p02	Favor.	Défav.	10
p03	Défav.	Favor.	80
p04	Favor.	Défav.	88

Question

On se demande si les opinions ont évolué favorablement pour ce candidat après le débat, et si oui dans quelle proportion. On répondra à ces questions au niveau de l'échantillon et on s'interrogera également sur ce qu'il peut en être chez l'ensemble des électeurs.

Type et statut des variables

On parle de *mesures répétées* dans la mesure où, même si le tableau des données individuelles comporte deux colonnes, c'est une même variable (une variable binaire à deux modalités Favor./Défav.) qui est observée à deux reprises avant et après le débat, et non pas deux variables différentes.

Ouverture du fichier

- Lancer SESAnalyse
- Menu Fichier - Ouvrir un dossier SES (*.SES)
- Sélectionner le dossier DEBAT.SES.

AFFICHAGE DES DONNÉES

Il est agréable de visualiser les données sous une autre forme : un Tableau de contingence ou Tri croisé.

- SES-Pégase
- Menu **Analyse**
- Voir la distribution
- Tri croisé

Tableau 3 : Données DEBAT (tableau de contingence)

n_jk	Favor.	Défav.	Total
Favor.	102	10	112
Défav.	80	88	168
Total	182	98	280

Ce tableau présente les données sous une autre forme que le tableau de données initial. Ainsi, on a 10 électeurs qui étaient favorables avant le débat et qui sont défavorables après le débat.

Les totaux marginaux nous donnent d'autres informations :

- le nombre d'électeurs favorables aux candidats avant le débat (indépendamment de leur opinion après le débat). Il y en a donc 112.
- le nombre d'électeurs favorables après le débat (indépendamment de leur opinion avant le débat). Il y en a 182.



ANALYSE DE CHAQUE MESURE RÉPÉTÉES

Opinion avant le débat

- Menu Nouvelle analyse ou cliquer sur l'icône :
- Sélectionner la « variable » AVANT comme variable à analyser

AVANT (Opinion avant le débat)

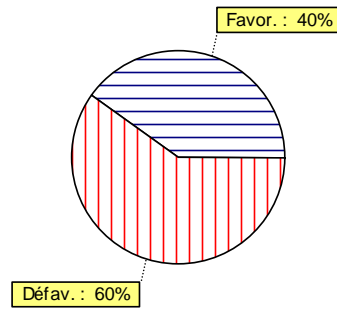


Figure 1 : Répartition en % des réponses favorables avant le débat

Opinion après le débat

APRES (Opinion après le débat)

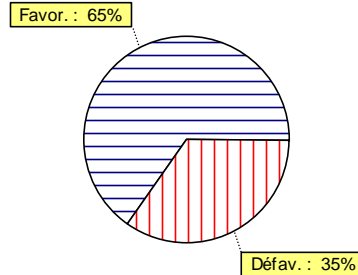


Figure 2 : Répartition en % des réponses favorables après le débat

ANALYSER L'ÉVOLUTION DES RÉPONSES

Sélectionner les variables à analyser

SES-Pégase

- Menu Nouvelle analyse ou cliquer sur l'icône :
- Sélectionner les deux variables, AVANT et APRES, en tant que variables à analyser.

Sélectionner la case « Mesures répétées »

Mesures répétées ?

Valider :

En dehors de l'examen des effectifs et pourcentages conjoints, les procédures d'analyse de mesures répétées sont radicalement différentes de celles mises en œuvre pour l'analyse d'un tableau de contingence classique.

En particulier, du point de vue des procédures inférentielles traditionnelles, le test approprié est le test du Kh^2 de McNémar (et non pas le test du Kh^2 d'indépendance).



Sens de l'évolution des pourcentages

Dans l'échantillon

- Menu Analyse
- Comparaison des mesures répétées
- Sens de l'écart entre les pourcentages

Tableau 4 : Évolution des réponses favorables, en %

Favor.	
AVANT	40.00%
APRES	65.00%

On constate que :

- le pourcentage de FAV est plus important APRES (65%) qu'AVANT (40%).

Le débat a eu pour effet d'augmenter le pourcentage d'opinions favorables pour ce candidat (et par conséquent, puisque la réponse « Sans opinion » était exclue, de diminuer le nombre d'opinions défavorables).

Dans la population

Même si les opinions ont évolué favorablement dans cet échantillon de 280 personnes, il n'est pas certain que l'évolution aille dans le même sens dans la population (l'ensemble des électeurs). Qu'en est-il ?

- Menu Analyse
- Comparaison des mesures répétées
- Sens de l'écart entre les pourcentages

Tableau 5 : Test du χ^2 de McNemar

Ddl	1
McNemar	54,44
p	<0.0001
McNemarC	52,90
p	<0.0001

On considère plutôt le résultat du test corrigé et la valeur p associée :

- si $p > .05$, on dit que « le test est significatif ». Cela permet de conclure à l'existence d'une différence du % de réponses favorables, dans le même sens que dans l'échantillon. On peut conclure à un effet du débat, dans un sens favorable au candidat.

- si $p > .05$ on ne peut pas conclure à l'existence d'une différence dans la population. On ne peut pas conclure à un effet du débat.

Ici, le test nous indique $p < .0001$, donc très petit

On peut conclure que, dans la population parente, il existe une différence de % de répondants favorables, entre avant et après le débat, soit ici dans le sens d'une augmentation du pourcentage d'opinions favorables (χ^2 de McNemar = 52.90, $p < 0.0001$).

Ampleur de l'écart entre les pourcentages

Dans l'échantillon

- SES-Pegase
- Menu Analyse
- Comparaison des mesures répétées
- Ampleur de l'écart entre les pourcentages

Tableau 6 : Différence des % entre les deux mesures répétées

Favor.	
AVANT	40.00%
APRES	65.00%
diff %	+25 pts %



*Deux mesures répétées sur une variable binaire
Le dossier DEBAT
Comparer les deux mesures répétées*

La différence de pourcentages de réponses favorables (Favor.) entre la première mesure, avant le débat (AVANT) et la seconde mesure, après le débat (APRES) est égale à +25 pts %.

On peut conclure, pour cet échantillon de 280 électeurs, à une augmentation importante du pourcentage d'opinions favorables au candidat après le débat (d = 25 points de %)

Dans la population

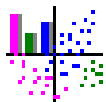
Pour répondre à la question de l'ampleur de l'écart dans la population, on calcule un intervalle de confiance sur la différence entre les deux pourcentages.

Tableau 7 : Intervalle de confiance (IC) sur la différence des % entre les enquêtes

d% Favor.	+25 pts %
d% Défav.	-25 pts %
p	0.05
Confiance	95.00%
IC_inf	+19 pts %
IC_sup	+31 pts %

L'intervalle de confiance (au seuil .05 ou au niveau de garantie de 95%) est ici [+19pts% ; 31 Pts%]. Il nous indique que, dans la population d'où a été tiré cet échantillon, la différence de % en faveur du candidat, est comprise entre 19 et 31 pts de %.

Sachant qu'on considère une différence de 10 pts de % comme importante, et que toutes les valeurs de l'intervalle sont supérieures à 10, on peut conclure à une différence importante dans la population.



RÉFÉRENCE

Corroyer, D., & Wolff, M. (2003). *L'Analyse Statistique des Données en Psychologie; Concepts et Méthodes de base*. Paris: Armand Colin (Cursus). p.131-160.

Rouanet, H., Bernard, J.-M., & Le Roux, B. (1990). *Statistique en Sciences Humaines. Analyse Inductive des Données*. Paris: Dunod. p.115-170.

Agresti, A. (2002). *Categorical Data Analysis (2nde ed.)*. Hoboken (New Jersey): John Wiley & Sons.

Liste des Tableaux

Tableau 1 : Données DEBAT (données individuelles).....1

Tableau 2 : Données DEBAT (liste des profils et effectifs)2

Tableau 3 : Données DEBAT (tableau de contingence).....2

Tableau 4 : Évolution des réponses favorables, en %4

Tableau 5 : Test du χ^2 de McNemar4

Tableau 6 : Différence des % entre les deux mesures répétées4

Tableau 7 : Intervalle de confiance (IC) sur la différence des % entre les enquêtes5

Liste des Figures

Figure 1 : Répartition en % des réponses favorables avant le débat3

Figure 2 : Répartition en % des réponses favorables après le débat3

SOMMAIRE

Un exemple : le dossier DEBAT1

Saisie informatique des données 1

Question..... 2

Type et statut des variables 2

Ouverture du fichier 2

Affichage des données2

Analyse de chaque mesure répétées3

Opinion avant le débat..... 3

Opinion après le débat 3

Analyser l'évolution des réponses3

Sélectionner les variables à analyser 3

Sens de l'évolution des pourcentages..... 4

 Dans l'échantillon4

 Dans la population4

Ampleur de l'écart entre les pourcentages4

 Dans l'échantillon4

 Dans la population5

Référence6

Liste des tableaux.....6

Liste des figures.....6