

# Plan. Chapitre IV : Exploiter

- I. Analyses de bases
- II. Tests statistiques



# Avant de partir

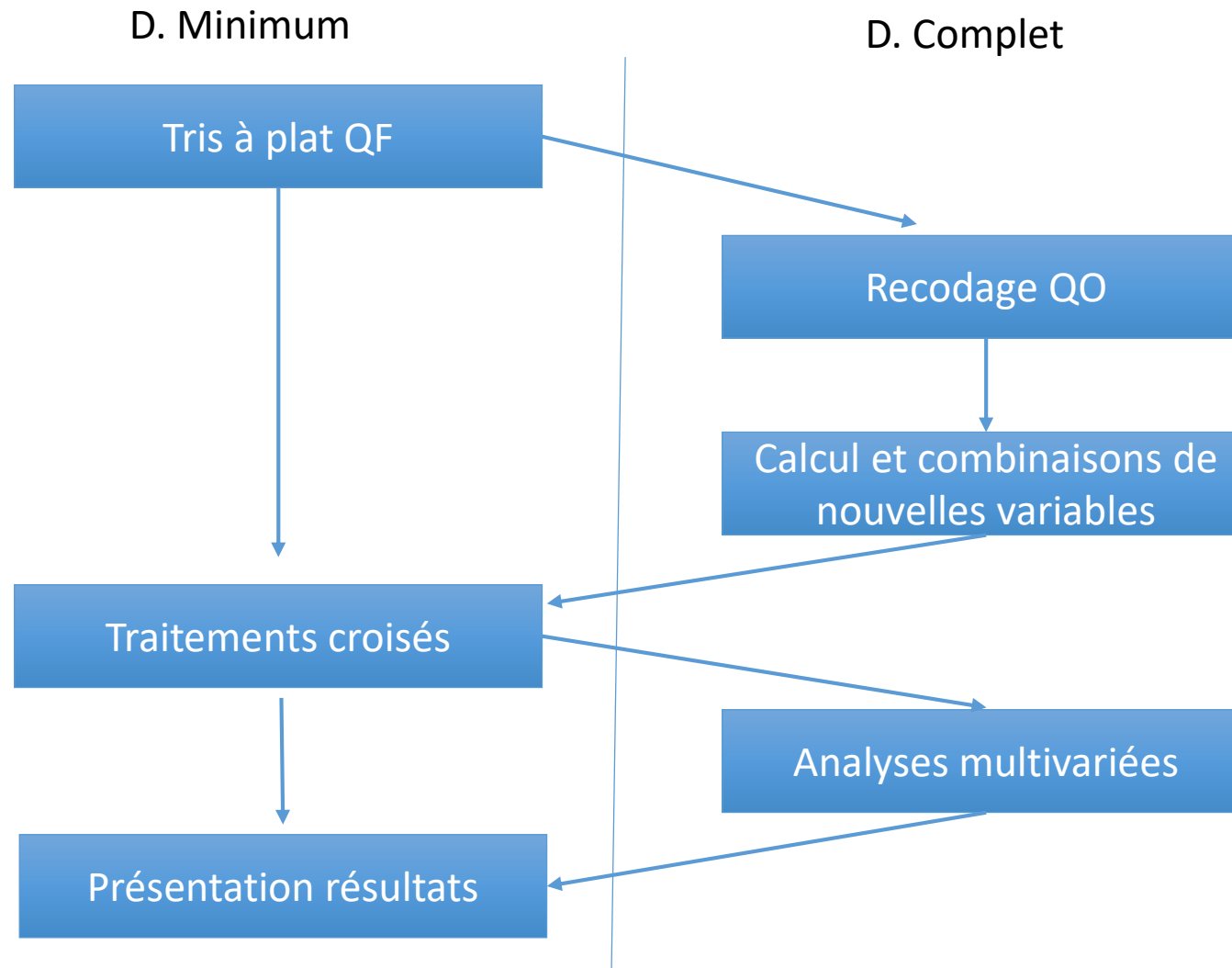
**Contrôler** votre saisie et nettoyer les données :

- Exclure de l'analyse les individus dont les réponses sont trop incomplètes (50 % à 60 % de non réponses)
- Supprimer les réponses « systématiques » : toujours la même modalité de réponse choisie par la répondant.
- Exclure les individus qui ont répondu mais ne font pas partie de l'échantillon.

# Avant de partir

- **Traiter les valeurs manquantes :**
  - . Conserver non-réponses
  - . Remplacer par moyenne ou le mode.
  - . Remplacer par la valeur du « plus proche voisin » : identifier les réponses d'un autre répondant qui « lui ressemble le plus ».

# Schéma général dépouillement d'une enquête.



# L'analyse des réponses.

- **Tris à plat** : calcul des effectifs et des pourcentages effectués question par question.

On parle d'analyse univariée.

# Analyse à plat : 1° regard sur les résultats de l'enquête

La présentation des résultats se fera à l'aide :

- tableaux d'effectifs et de fréquence
- représentations graphiques adéquates (*camemberts/secteurs, diagramme, histogrammes...*)
- indicateurs de tendance centrale (moyenne...) et de dispersion (écart-type...) pour les variables métriques

Ces **résultats seront brièvement commentés** de manière à faire ressortir les points importants.

### Avez-vous déjà acheté des produits locaux ?

(n = 294, échantillon redressé)



**Figure 6 : Achats de produits locaux dans le Pays de la Jeune Loire**

**95 % des répondants déclarent avoir déjà acheté des produits locaux\*.**

\*Un produit est dit local s'il est produit ou transformé en Haute-Loire ou à 80 km de son lieu d'achat.

Ce résultat global peut être nuancé en analysant les comportements d'achats relatifs aux produits locaux.



BILAN ET POTENTIEL DE CONSOMMATION DE PRODUITS  
LOCAUX DANS LE PAYS DE LA JEUNE LOIRE ET SES RIVIERES



4 réponses

RÉSUMÉ

INDIVIDUEL

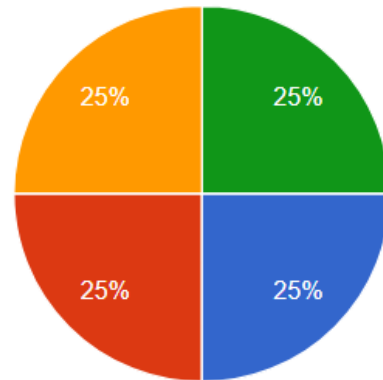


Réponses acceptées



Comment qualifiez-vous votre satisfaction concernant l'achat que vous venez d'effectuer sur notre site ?

4 réponses



- Tout à fait satisfait
- Très satisfait
- Moyennement satisfait
- Pas du tout



QUESTIONS

RÉPONSES

4

## 4 réponses

Message pour les personnes interrogées

Le formulaire "Questionnaire satisfact

RÉSUMÉ

INDIVIDUEL



Recevoir une notification lorsqu'une réponse est ajoutée

Sélectionner la destination des réponses

Supprimer le lien vers le formulaire



Télécharger les réponses (.csv)



Imprimer toutes les réponses

Supprimer toutes les réponses

# Questionnaire découverte Forms

thierry.paradinas@gmail.com

Fichier Édition Affichage Insertion Format Données Outils Formulaire Modules complémentaires Aide Toutes les modifications ont été enregistrées dans Drive

Commentaires

Partager

100% € % ,0 .00 123 Arial 10 B I S A

Horodateur

A	B	C	D	E	F	G	H	I
Horodateur	Comment qualifiez-vous votre satisfaction concernant l'achat que vous venez d'effectuer sur notre site ?	Notez votre degré de satisfaction sur les points suivants [Trouver produit]	Notez votre degré de satisfaction sur les points suivants [Choix]	Notez votre degré de satisfaction sur les points suivants [Description]	Notez votre degré de satisfaction sur les points suivants [Passage de commande]	Notez votre degré de satisfaction sur les points suivants [Choix modes de paiement]	Notez votre degré de satisfaction sur les points suivants [Respect des délais de livraison]	Envisagez-vous un nouvel achat sur notre site ?
14/09/2017 14:09:55	Tout à fait satisfait	6	5		7	7	7	Oui probablement
14/09/2017 14:11:05	Pas du tout	7	7	7	6	5	7	Non
14/09/2017 14:11:41	Moyennement satisfait	5	5	5	5	5	5	Oui probablement
14/09/2017 14:25:45	Très satisfait	2	3	4	5	6	6	Oui probablement



Réponses au formulaire 1 Tableau croisé dynamique 3 Tableau croisé dynamique 1

Alonso.xls

Explorer

Tout afficher X

Questionnaire découverte Forms

thierry.paradinas@gmail.com

Fichier Édition Affichage Insertion Format Données Outils Formulaire Modules complémentaires Aide Toutes les modifications ont été enregistrées dans Drive

Commentaires

Partager

100% € % .0 .00 123 Arial 10 B I S A

A	B	C	D	E	F	G
Horodateur	Comment qualifiez-vous votre satisfaction concernant l'achat que vous venez d'effectuer sur notre site ?	Notez votre degré de satisfaction sur les points suivants [Trouver produit]	Notez votre degré de satisfaction sur les points suivants [Choix]	Notez votre degré de satisfaction sur les points suivants [Description]	Notez votre degré de satisfaction sur les points suivants [Passage de commande]	Notez votre degré de satisfaction sur les points suivants [Choix modes de paiement]
14/09/2017 14:09:55	Tout à fait satisfait	6	5			7
14/09/2017 14:11:05	Pas du tout	7	7	7		6
14/09/2017 14:11:41	Moyennement satisfait	5	5	5	5	5
14/09/2017 14:25:45	Très satisfait	2	3	4		5

Explorer

Pour A1:P5 MODIFIER

ANALYSE

Count of Envisagez-vous un nouvel achat sur notre site ?

Count of Aviez-vous déjà acheté sur sur notre site avant aujourd'hui ?

Non

Explorer

Nombre de Aviez-vous déjà acheté sur sur notre site avant aujourd'hui ?

Nombre de Envisagez-vous un nou

Réponses au formulaire 1 Tableau croisé dynamique 3 Tableau croisé dynamique 1

# Que faire des « non-réponses » ?

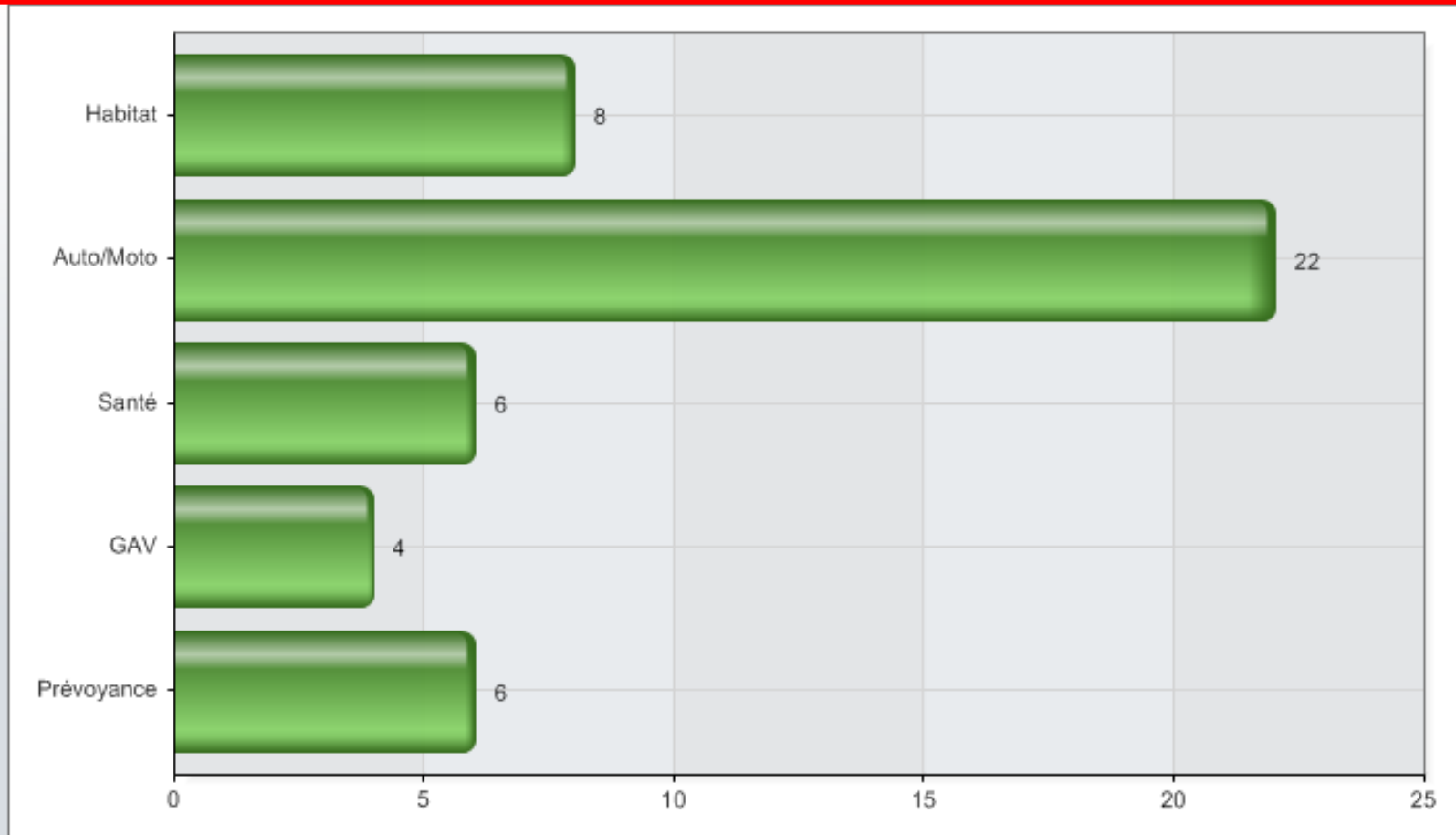
- Dans la plupart des cas on exclura les NR de la restitution des résultats.
- Eventuellement on pourra préciser le taux de réponse de la question ou le nombre de répondants.

# Types de variables

- **Nominales**
- **Ordinales**
- **Métriques**

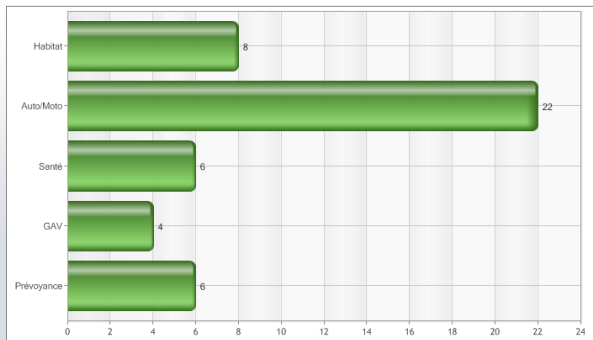
# Types de variables

- **Nominales** : mode, fréquences...



**Quels types d'assurance aviez-vous?**

Habitat	8
Auto/Moto	22
Santé	6
GAV	4
Prévoyance	6
autre à préciser	0
<b>TOTAL</b>	<b>46</b>



Quels types d'assurance aviez-vous?	
Habitat	8
Auto/Moto	22
Santé	6
GAV	4
Prévoyance	6
autre à préciser	0
TOTAL	46
Non-réponse(s)	93

Quels types d'assurance aviez-vous?	
Habitat	17,39%   25%
Auto/Moto	47,83%   68,75%
Santé	13,04%   18,75%
GAV	8,70%   12,5%
Prévoyance	13,04%   18,75%
autre à préciser	00,00% (0%)
TOTAL	100%



Pourcentage citations

Pourcentage observations

Quels types d'assurance aviez-vous?	
Habitat	17,39%   25%
Auto/Moto	47,83%   68,75%
Santé	13,04%   18,75%
GAV	8,70%   12,5%
Prévoyance	13,04%   18,75%
autre à préciser	00,00% (0%)
TOTAL	100%



# Types de variables

- **Ordinales** : variable qualitative dont les modalités sont ordonnées.

## Choix ordonné

Ordonnez  
par ordre  
de  
préférence

---

L'accueil

3

---

Le niveau des intervenants

1

---

Le discours d'introduction

2

---

Le repas

4

# Types de variables

- **Métriques**

# Variables métriques dans e-questionnaire

- QO Entier / numérique
- Q échelles **avec notation**
- Questions groupées Entier / Numérique

# QO Entier / numérique

► Libellé de la question (Numérotation automatique = 8)

Quel est votre âge ?

Afficher le titre dans le questionnaire

► Type de la question

- Choix multiple (case à cocher)
- Choix unique (bouton radio)
- Choix unique (liste déroulante)
- Ouverte (texte libre)
- Cachée

► Type de la question ouverte

- Texte
- Entier
- Numérique
- Pourcentage
- Date(jj/mm/aaaa)

*Vous devez dépublier l'enquête pour pouvoir la modifier.*

Population	3
Somme	82,0
Moyenne	27,3
Variance	276,3
Ecart type	16,6
Minimum	12,0
Maximum	45,0
Données	#
12	1
25	1
45	1

# Q échelles avec notation

Libellé de la question (Numérotation automatique = 3)

Etes-vous satisfait du conseil fait par votre conseiller lors de la souscription de votre assurance?

Afficher le titre dans le questionnaire

Type de la question

- Choix multiple (case à cocher)
- Choix unique (bouton radio)
- Choix unique (liste déroulante)
- Ouverte (texte libre)
- Cachée

Propriétés complémentaires

- Cette question est obligatoire
- L'ordre des modalités est aléatoire
- Ajouter du texte libre après la dernière modalité
- Attribuer une valeur par réponse (pondération)
- Colonnes

Réponses possibles en colonne

Titre

1	Satisfait	✗
2	Plutôt satisfait	✗
3	Plutôt insatisfait	✗
4	Insatisfait	✗

Attribuer une valeur par réponse (pondération)

5	▼	🎨
3	▼	🎨
2	▼	🎨
1	▼	🎨

Couleur graphique

Ajouter Importer des choix pré-définis

Saisissez un titre pour une nouvelle réponse ou plusieurs titres séparés par des point-virgule ;

✗ Supprimer

Insatisfait: 15,62%(5)	Satisfait: 12,5%(4)
Plutôt insatisfait: 12,5%(4)	Plutôt satisfait: 59,38%(19)

Etes-vous satisfait du conseil fait par votre conseiller lors de la souscription de votre assurance?

Satisfait	4
Plutôt satisfait	19
Plutôt insatisfait	4
Insatisfait	5
TOTAL	32
Non-réponse(s)	93

Etes-vous satisfait du conseil fait par votre conseiller lors de la souscription de votre assurance?

Satisfait	12,50%
Plutôt satisfait	59,38%
Plutôt insatisfait	12,50%
Insatisfait	15,63%
TOTAL	100%

Etes-vous satisfait du conseil fait par votre conseiller lors de la souscription de votre assurance?

MOYENNE	2,81/5
---------	--------

# Questions groupées Entier / Numérique

Notez votre degré de satisfaction pour chaque point

texte visible sous le titre de l'élément

Afficher le titre dans le questionnaire

## Type de la question

- Choix multiple (case à cocher)
- Choix unique (bouton radio)
- Choix ordonné
- Ouverte (texte libre)

## Type de la question ouverte

- Texte
- Entier
- Numérique
- Pourcentage
- Date(jj/mm/aaaa)

## Propriétés complémentaires

- Cette question est obligatoire
- Précision
- Minimum (Intervalle)
- Maximum (Intervalle)

## Questions en ligne

Titre

- 1 Trouver un produit
- 2 Choix des produits
- 3 Description / informations sur les produits
- 4 Facilité de passer les commandes
- 5 Choix des modes de paiement
- 6 Respect des délais de livraison



+

Ajouter

Saisissez un titre pour une nouvelle question ou plusieurs titres séparés par des point-virgule ;'

## Réponses possibles en colonne

Titre

- 1 Note



Couleur graphique



+

Ajouter

Importer des choix pré-définis

Saisissez un titre pour une nouvelle réponse

Description / informations s...	
Facilité de passer les comm...	
Choix des modes de paiement	
Respect des délais de livrai...	
	0,0

Note

Trouver un produit							
Modalité	Population	Somme	Moyenne	Variance	Ecart type	Minimum	Maximum
Note	8	41,0	5,1	11,0	3,3	2,0	10,0
Choix des produits							
Modalité	Population	Somme	Moyenne	Variance	Ecart type	Minimum	Maximum
Note	8	39,0	4,9	8,7	2,9	2,0	11,0
Description / informations sur les produits							
Modalité	Population	Somme	Moyenne	Variance	Ecart type	Minimum	Maximum
Note	7	68,0	9,7	87,9	9,4	2,0	30,0
Facilité de passer les commandes							
Modalité	Population	Somme	Moyenne	Variance	Ecart type	Minimum	Maximum
Note	8	38,0	4,8	4,2	2,1	2,0	8,0
Choix des modes de paiement							
Modalité	Population	Somme	Moyenne	Variance	Ecart type	Minimum	Maximum
Note	8	45,0	5,6	16,0	4,0	1,0	12,0
Respect des délais de livraison							
Modalité	Population	Somme	Moyenne	Variance	Ecart type	Minimum	Maximum
Note	8	36,0	4,5	14,0	3,7	1,0	10,0

8&rnd=0,7747401



## Paramètres de position

- Moyenne
- Médiane
- Mode

## Paramètres de dispersion

- Etendue
- Ecart-type

# Paramètre de position : moyenne

- Mesure de tendance centrale la plus courante : centre de gravité d'une série.
- Problème de la sensibilité aux observations extrêmes.
- Doit être complétée par notion de **DISPERSION** : variance, écart-type, étendue...

# Variance et écart-type

- **Variance** : donne une connaissance de la distance entre chaque point et la moyenne.
- **Ecart-type** : plus exploitable que la variance car élimine l'effet du carré.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Pris isolément = pe

**fficient de variation.**

## Coefficient de variation :

$$CV = \frac{\sigma}{\text{Moyenne}} \times 100$$

On estime qu'une série est dispersée si **CV > 25%**

Dans ce cas, **creuser par tris croisés.**

# Allez plus loin...

- Appliquer des **filtres**
- Intérêt ?  
Analyser les résultats de manière plus fine.

**Tableau de bord**  
Suivez le déroulement de votre enquête

Suivez le déroulement de votre enquête

Statut	Total	%		
Initialisé	40	10,08%	1	0
Email lu	0	0%	0	0
Ouvert	26	6,55%	19	0
En cours	0	0%	0	0
Terminé	331	83,37%	0	0
<b>Total</b>	<b>397</b>	<b>100%</b>	<b>20</b>	<b>0</b>

Rapport

Options

**Filtres** Codification

Filtres - Google Chrome

manager.e-questionnaire.com/DA\_Filter.asp?new=1&t=0

Définition des filtres

<< Liste des filtres

Description

Filtres - Google Chrome

manager.e-questionnaire.com/DA\_Filter.asp?numQY=40153&new=C&t=0

Définition des filtres

<< Liste des filtres

Description

essai

Liste des critères

[Ajout d'un critère](#)

Aucun critère disponible

Ajout d'un critère

Champ de la question	Opérateur	Valeur
...		

Filtres - Google Chrome

manager.e-questionnaire.com/DA\_Filter.asp?numQY=40153&new=C&t=0

Définition des filtres

<< Liste des filtres

Description

essai

Liste des critères

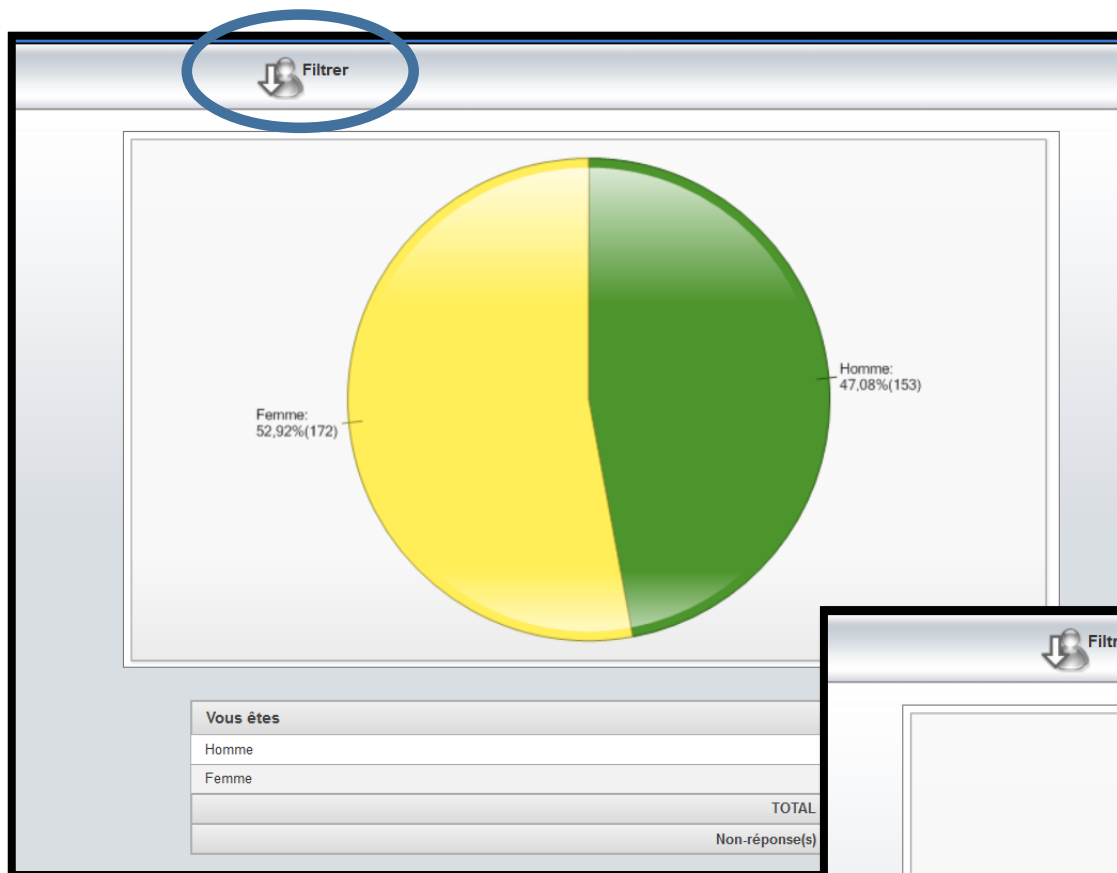
[Ajout d'un critère](#)

Aucun critère disponible

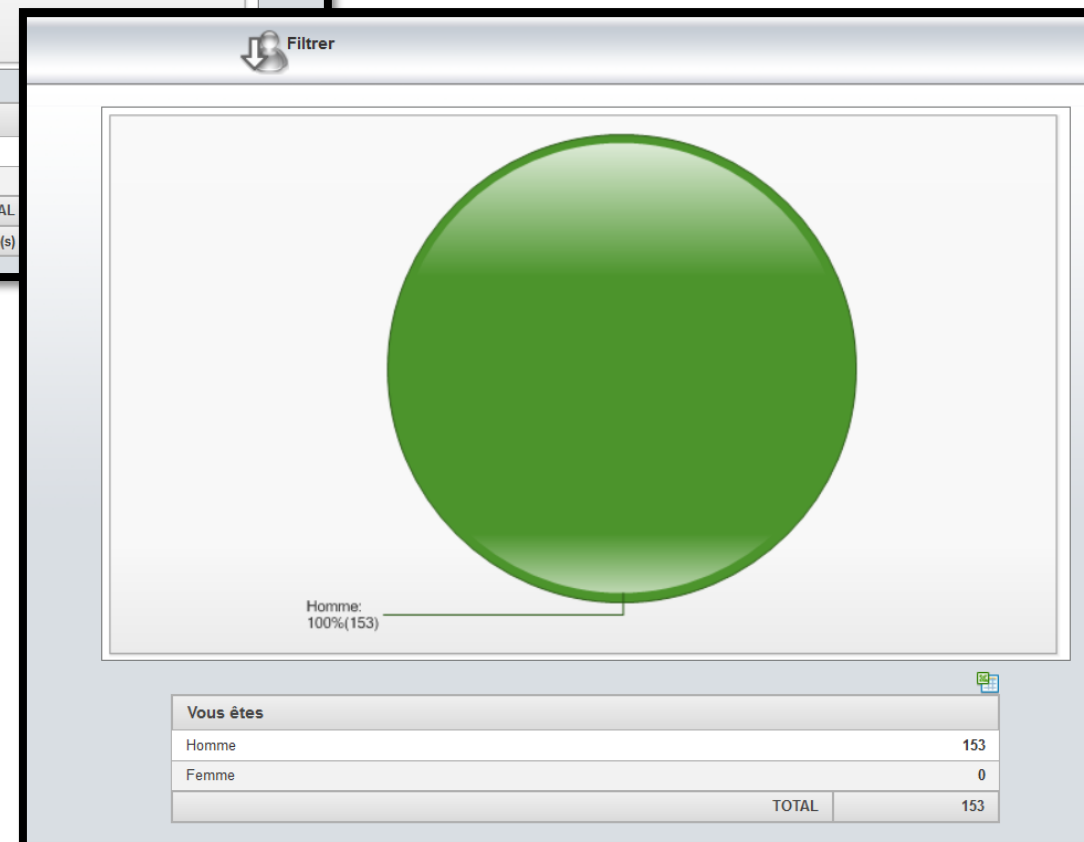
Ajout d'un critère

Champ de la question	Opérateur	Valeur
Quelle est votre banque principale ?	Est égal à	Banque populaire

Titre complet du champ de la question:  
Quelle est votre banque principale ?

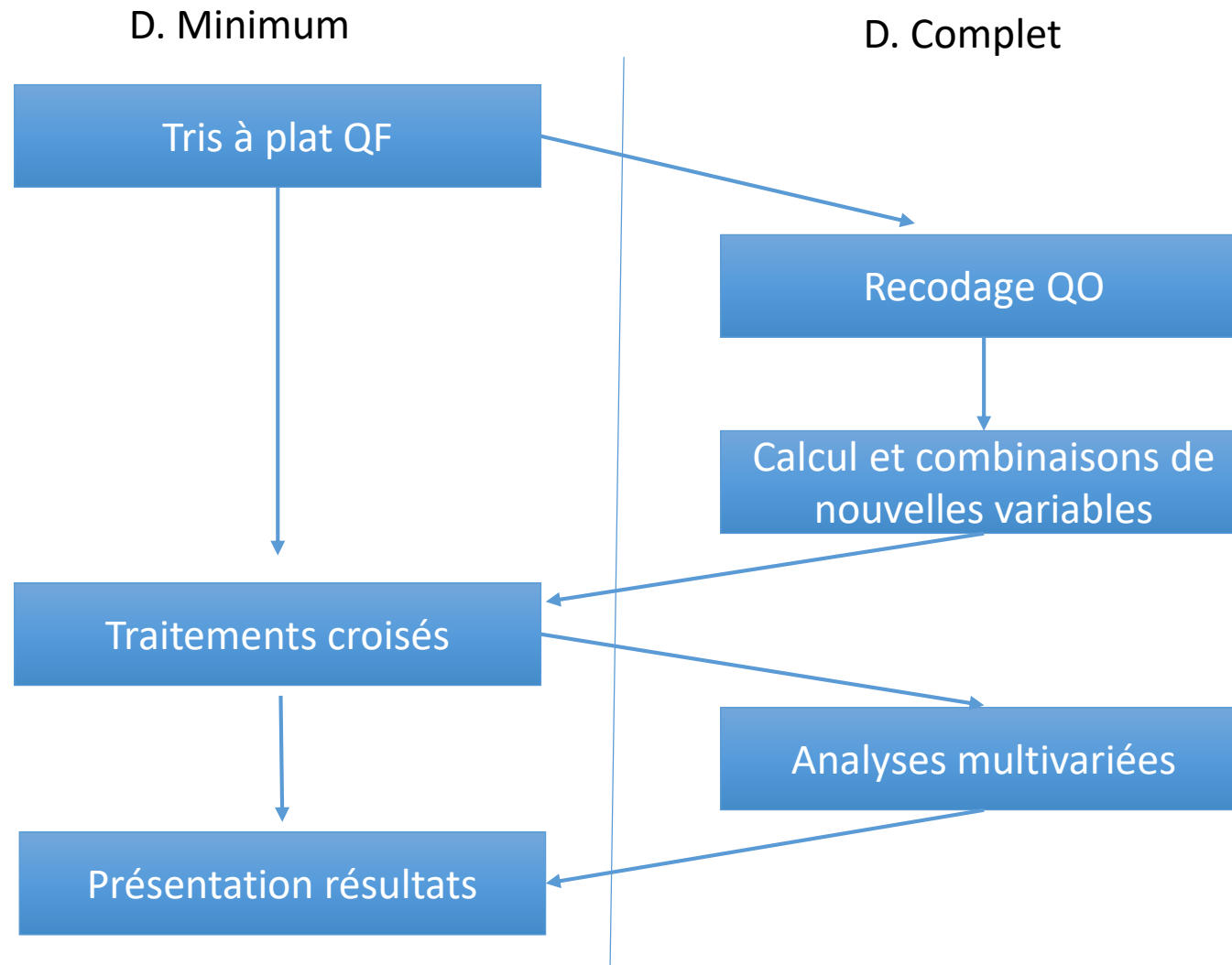


Ici on filtre pour n'avoir que les réponses des hommes.....





# Schéma général dépouillement d'une enquête.

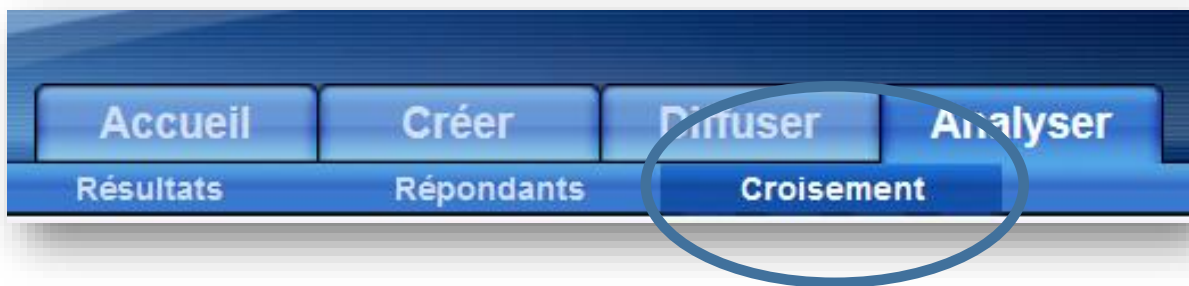


# L'analyse des réponses.

- **Tris croisés** : les effectifs et pourcentages sont effectués en croisant 2 questions.

On parle d'**analyse bivariée**

Savoir si les résultats globaux observés varient en fonction des caractéristiques des répondants.



Accueil Résultats  
Créer Répondants  
Diffuser Croisement  
Analyser

**Analyse Croisée**  
Selectionnez les deux questions dont vous souhaitez croiser les résultats

Question en ligne

Merci de répondre aux questions suivantes....

Comment qualifiez-vous.....     Comment qualifiez-vous.....

Notez votre degré de satisfaction pour chaque point

Trouver un produit	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Trouver un produit
Choix des produits	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Choix des produits
Description / informations sur les produits	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Description / informations sur les produits
Facilité de passer les commandes	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Facilité de passer les commandes
Choix des modes de paiement	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Choix des modes de paiement
Respect des délais de livraison	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Respect des délais de livraison

Merci de répondre aux questions suivantes....

Quels sont les éléments qui vous ont décidé à acheter ?     Quels sont les éléments qui vous ont décidé à acheter ?

Envisagez-vous un nouvel achat sur notre site ?     Envisagez-vous un nouvel achat sur notre site ?

# Tris croisés.

		Etes-vous satisfait du conseil fait par votre conseiller lors de la souscription de votre assurance?			
		Satisfait	Plutôt satisfait	Plutôt insatisfait	Insatisfait
Quels types d'assurance aviez-vous?	Habitat	1 12,5%	3 37,5%	1 12,5%	3 37,5%
	Auto/Moto	2 9,09%	13 59,09%	3 13,64%	4 18,18%
	Santé	0 0%	4 66,67%	0 0%	2 33,33%
	GAV	0 0%	4 100%	0 0%	0 0%
	Prévoyance	1 16,67%	3 50%	2 33,33%	0 0%
	autre à préciser	0	0	0	0

Format Données Outils Formulaire Modules complémentaires Aide Toutes les modifications ont été enregistrées dans Drive

.0 .00 123 Arial 10 B I S A

Éditeur de rapports

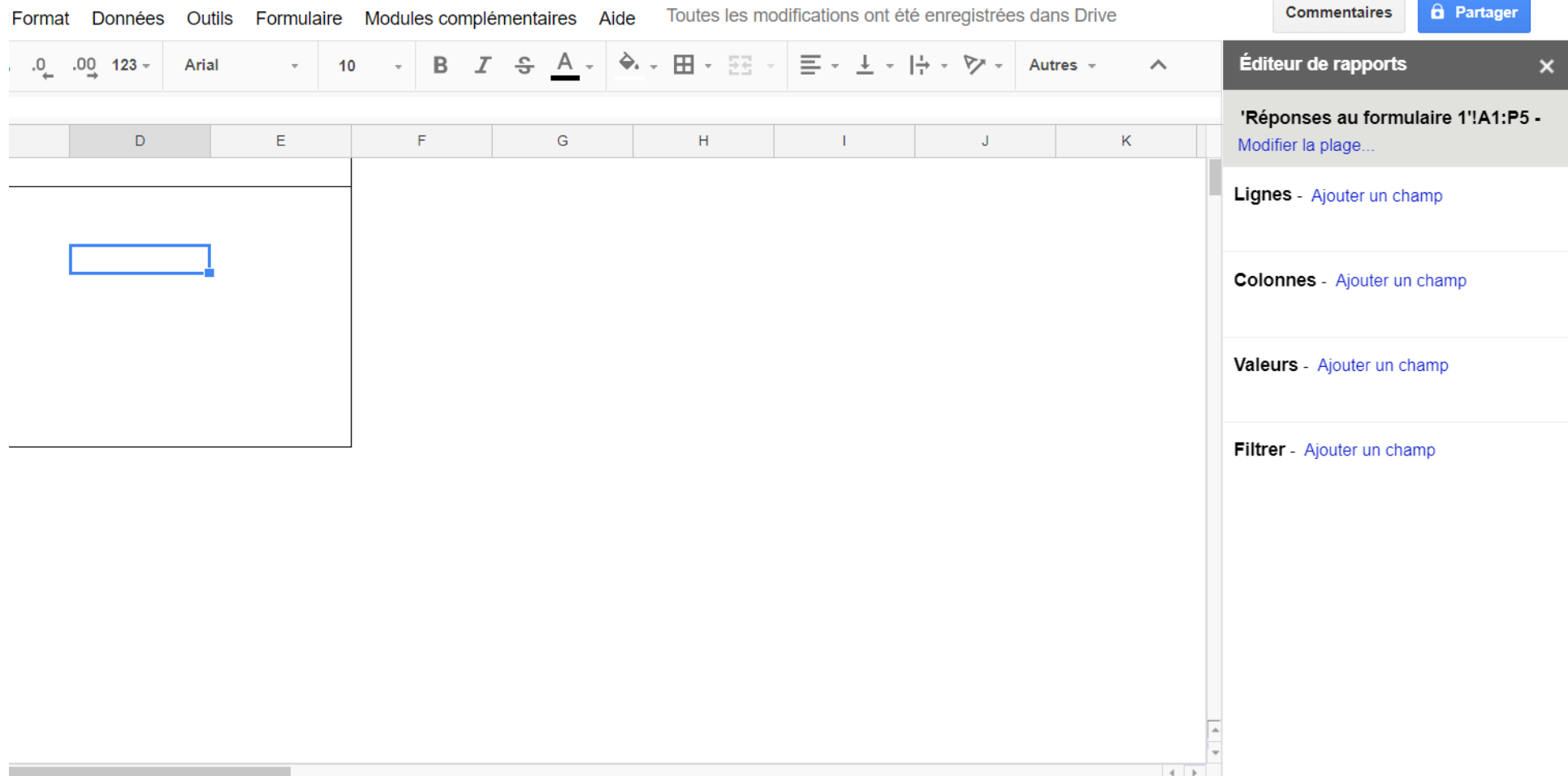
'Réponses au formulaire 1'!A1:P5 -  
[Modifier la page...](#)

Lignes - [Ajouter un champ](#)

Colonnes - [Ajouter un champ](#)

Valeurs - [Ajouter un champ](#)

Filtrer - [Ajouter un champ](#)



# Tris croisés avec Google Forms

Utilisation des TCD.

# Plan. Chapitre IV : Exploiter

## I. Analyses de bases

## II. Tests statistiques de base



**Analyses bivariées -> test statistique.**



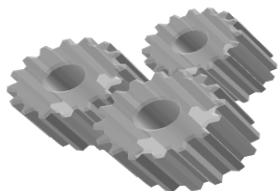
**Méthodes selon les types de questions qui sont croisées.**

<b>Objectifs</b>	<b>Analyse appropriée</b>
<p>Tester un <u>lien</u> <b>d'association</b> entre deux <u>variables nominales</u> (fermée et fermée)  <i>Existe-t-il un lien entre le fait que l'ouvrier se déclare <b>satisfait ou insatisfait de ses conditions de travail</b> et / le <b>type de formation</b> qu'il déclare avoir préalablement suivi ? – Le fait d'être un H/F - selon son âge...</i></p>	<p><b>Test du Chi-deux</b></p>
<p>Tester un lien d'association entre deux <u>variables ordinales</u>  <i>Existe un lien entre <b>l'attractivité de certaines TIC</b> et le <b>type de loisirs préférés</b> (dans les deux cas , l'individu devait classer des propositions).</i></p>	<p><b>Coefficient de corrélation de Spearman-Kendall</b></p>
<p>Tester un lien d'association entre deux <u>variables métriques</u> (<u>Numériques</u>)  <i>Le <b>score d'épanouissement professionnel</b> (de 1 à 8) est-il associée à <b>l'âge des salariés</b> (à l'ancienneté " ? Au salaire)</i></p>	<p><b>Coefficient de corrélation linéaire de Pearson</b></p>



## Calcul du Chi-2 .

**Chi 2** : compare les résultats observés ce qui ce passerait en théorie s'il n'y avait aucun lien entre les variables.



# Les questions posées....

- Existe-t-il une **association** (et pas une causalité) entre les variables ?
- Quelle est la **force** de cette relation ?
- Dans l'affirmative, **comment se traduit-elle dans les faits** ?

# Contraintes

- **Variables nominales**

*NB : une variable métrique ou ordinale peut être « dégradée » en nominale*

- **Tableau de contingence** -> tri croisé

- **Effectif mini dans les cellules du tableau : 5**

# Exemple de départ...

Ex: Conso de pâtes au chocolat / situation de famille

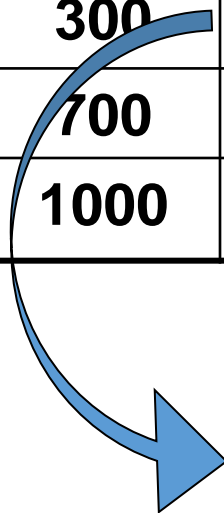
Existe-t-il une association entre la situation familiale et la consommation de pâtes au chocolat ?

**Résultats du tri croisé :**

	TOTAL	CEL	MARIE Sans enfants	Mariés avec enfants	Veufs	Divorcés
OUI	<b>300</b>	<b>40</b>	<b>60</b>	<b>150</b>	<b>20</b>	<b>30</b>
NON	<b>700</b>	<b>70</b>	<b>120</b>	<b>200</b>	<b>110</b>	<b>200</b>
	<b>1000</b>	<b>110</b>	<b>180</b>	<b>350</b>	<b>130</b>	<b>230</b>

**Tableau de contingence = tableau des fréquences théoriques** : répond à la question, quels seraient les effectifs si aucun rapport entre conso et situation familiale ?

	TOTAL	CEL	MARIE Sans enf	Mariés avec enfants	Veufs	Divorcés
OUI	<b>300</b>					
NON	<b>700</b>					
	<b>1000</b>	<b>110</b>	<b>180</b>	<b>350</b>	<b>130</b>	<b>230</b>



Si 30% des personnes consomment, on devrait retrouver ce pourcentage chez les célibataires =  $110 \times 300/1000$

	TOTAL	CEL	MARIE Sans enf	Mariés avec enfants	Veufs	Divorcés
OUI	<b>300</b>	<b>33</b>	<b>54</b>	<b>105</b>	<b>39</b>	<b>69</b>
NON	<b>700</b>	<b>77</b>	<b>126</b>	<b>245</b>	<b>91</b>	<b>161</b>
	<b>1000</b>	<b>110</b>	<b>180</b>	<b>350</b>	<b>130</b>	<b>230</b>

Les résultats diffèrent des **fréquences observées** :

	TOTAL	CEL	MARIE Sans enf	Mariés avec enfants	Veufs	Divorcés
OUI	<b>300</b>	<b>40</b>	<b>60</b>	<b>150</b>	<b>20</b>	<b>30</b>
NON	<b>700</b>	<b>70</b>	<b>120</b>	<b>200</b>	<b>110</b>	<b>200</b>
	<b>1000</b>	<b>110</b>	<b>180</b>	<b>350</b>	<b>130</b>	<b>230</b>

Est-ce que cette différence est « pertinente » pour pouvoir dire qu'il y a un lien ?

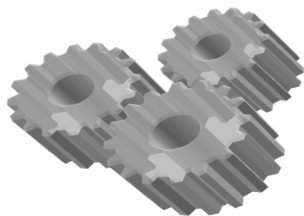
# Calcul du Chi-2.

$$\frac{\text{somme } [(fréquences théoriques - f observées)]^2}{$$

$$[f \text{ théoriques}]$$

khi 2 qui mesure la **distance** entre **ce qui est observé** (le premier tableau) et ce **qui ce passerait en théorie s'il n'y avait aucun lien entre les variables**.

Compare le résultat à un « **seuil de dépendance** »





	TOTAL	CEL	MARIE	Mariés		
			Sans enf	avec	Veufs	Divorcés
				enfants		
OUI	300	1,48	0,67	19,29	9,26	22,04
NON	700	0,64	0,29	8,27	3,97	9,45
	1000					



Dans notre exemple, **Chi-2 = 75,5.**

Seuil de Fiabilité = 95% ←

$\chi^2$	X°.005	X°.01	X°.025	X°.05	X°.10	X°.25	X°.50	X°.75	X°.90	X°.95	X°.975	X°.99	X°.995	X°.999
1										3,84				
2										5,99				
3										7,81				
4										9,49				
5										11,1				
6										12,6				
7										14,1				
8										15,5				
9										16,9				
10										18,3				
11										19,7				
12										21,0				
13										22,4				
14										23,7				
15										25,0				
16										26,3				
17										27,6				
18										28,9				
19										30,1				
20										31,4				
21										32,7				
22										33,9				
23										35,2				
24										36,4				
25										37,7				
26										38,9				
27										40,1				
28										41,3				
29										42,6				
30										43,8				
40										55,8				
50										67,5				
60										79,1				
70										90,5				
80										102				
90										113				
100										124				

Degré de liberté  
=  
 $(c-1) \times (l-1)$   
C : nombre de colonnes  
L : nombre de lignes  
Du tableau de contingence  
**Attention : ne pas  
considérer lignes et  
colonnes TOTAUX**



Ici, ddl = 4

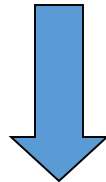
# Calcul du Chi-2.

La valeur trouvée s'appelle le SEUIL D'INDEPENDANCE

**Khi-2 > seuil**

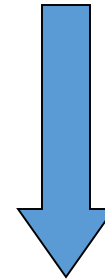


Liaison significative  
Entre les variables



Le tri croisé est intéressant.  
Association/ dépendance entre les variables

**Khi-2 < seuil**



Pas de liaison

	Oui	Non	Total
Femme	12	41	53
Homme	23	33	56
<b>Total</b>	<b>35</b>	<b>74</b>	<b>109</b>
	32,1%	67,9%	
Calcul fréquences théoriques			
	Oui	Non	Total
Femme	17,01834862	35,98165138	53
Homme	17,98165138	38,01834862	56
<b>Total</b>	<b>35</b>	<b>74</b>	<b>109</b>

Probabilité KHI DEUX =	=TEST.KHIDEUX(F7:G8;F14:G15)
	TEST.KHIDEUX(plage_réelle; plage_attendue)

La formule est :

=TEST.KHIDEUX(valeurs\_tableau\_frequence\_observees;valeurs\_tableau\_frequencies\_theoriques)

# Calcul sur Excel

Probabilité du Khi2	1,6895E-15	<code>=TEST.KHIDEUX(L6:P7;D6:H7)</code>
---------------------	------------	---

Valeur du Khi2	75,34	<code>=KHIDEUX.INVERSE(G16;4)</code>
Valeur du khi 2 à 5%	9,49	<code>=KHIDEUX.INVERSE(5%;4)</code>

DDL

Valeur du Chi-2 théorique

# Exercice

Existe-t-il un lien entre le niveau d'études et le choix d'une destination de vacances ?

		Corrèze	New York	Berlin	Total
Primaire		<b>300</b>	<b>50</b>	<b>100</b>	
Secondaire		<b>250</b>	<b>80</b>	<b>20</b>	
Supérieur		<b>50</b>	<b>120</b>	<b>30</b>	
Total					

Seuil de Fiabilité = 95% ←

$\chi^2$	X°.005	X°.01	X°.025	X°.05	X°.10	X°.25	X°.50	X°.75	X°.90	X°.95	X°.975	X°.99	X°.995	X°.999
1										3,84				
2										5,99				
3										7,81				
4										9,49				
5										11,1				
6										12,6				
7										14,1				
8										15,5				
9										16,9				
10										18,3				
11										19,7				
12										21,0				
13										22,4				
14										23,7				
15										25,0				
16										26,3				
17										27,6				
18										28,9				
19										30,1				
20										31,4				
21										32,7				
22										33,9				
23										35,2				
24										36,4				
25										37,7				
26										38,9				
27										40,1				
28										41,3				
29										42,6				
30										43,8				
40										55,8				
50										67,5				
60										79,1				
70										90,5				
80										102				
90										113				
100										124				

Degré de liberté

=

$$(c-1) \times (l-1)$$

C : nombre de colonnes

L : nombre de lignes

Du tableau de contingence

Attention : ne pas

considérer lignes et

colonnes **TOTAUX**

Ici, ddl = 4

### 1 Effectif réel

	1	1	1		3	
	Corrèze	New York	Berlin		TOTAL	
Primaire	300	50	100		450	1
Secondaire	250	80	20		350	1
Supérieur	50	120	30		200	1
					0	
					0	
					0	
					0	
					0	
					0	
					0	

0 TOTAL

### 2 Effectif théorique

	Corrèze	New York	Berlin	0	0	TOTAL
Primaire	270,0	112,5	67,5			450
Secondaire	210,0	87,5	52,5			350
Supérieur	120,0	50,0	30,0			200
0						0
0						0
0						0
0						0
0						0
0						0
0						0
0						0
<b>TOTAL</b>	<b>600</b>	<b>250</b>	<b>150</b>	<b>0</b>	<b>0</b>	<b>1000</b>



### 3 Calcul du Khi deux

	Corrèze	New York	Berlin	0	0
Primaire	3,3	34,7	15,6		
Secondaire	7,6	0,6	20,1		
Supérieur	40,8	98,0	0,0		
0					
0					
0					
0					
0					
0					
0					
0					

Khi deux = 220,92

#### 4 Calcul du ddl et conclusion

Nombre de lignes	3
Nombre de colonnes	3
DDL (Degré de liberté)	4
Valeur critique au seuil de 5%	9,49
Khi deux observé	220,92

Conclusion

Dépendance entre les variables

<b>Objectifs</b>	<b>Analyse appropriée</b>
<p>Tester un <u>lien</u> d'association entre deux <b><u>variables nominales</u></b> (fermée et fermée) <i>Existe-t-il un lien entre le fait que l'ouvrier se déclare <b>satisfait ou insatisfait de ses conditions de travail</b> et / le <b>type de formation</b> qu'il déclare avoir préalablement suivi ? – Le fait d'être un H/F - selon son âge...</i></p>	<p><b>Test du Chi-deux</b></p>
<p>Tester un lien d'association entre deux <b><u>variables ordinales</u></b> <i>Existe un lien entre <b>l'attractivité de certaines TIC</b> et le <b>type de loisirs préférés</b> (dans les deux cas , l'individu devait classer des propositions).</i></p>	<p><b>Coefficient de corrélation de Spearman-Kendall</b></p>
<p>Tester un lien d'association entre deux <b><u>variables métriques</u></b> <b>(Numériques)</b> <i>Le <b>score d'épanouissement professionnel</b> (de 1 à 8) est-il associée à <b>l'âge des salariés</b> (à l'ancienneté " ? Au salaire)</i></p>	<p><b>Coefficient de corrélation linéaire de Pearson</b></p>

# Le principe

- Indice : **Rhô de Spearman**
- Mesurer intensité du lien et le sens de cette relation sur les rangs des observations (et non sur les observations en elles-mêmes).  
*Ex : marques classées sur 2 critères par les répondants*

# L'exemple

Pré-test publicitaire sur 5 films. Question : convergence entre classement des cibles et pros.

	Clt conso	Clt Pros	Écart di	Ecart di <sup>2</sup>
F1	3	2		
F2	1	3		
F3	4	5		
F4	2	1		
F5	5	4		

	Clt conso	Clt Pros	Écart di	Ecart di <sup>2</sup>
F1	3	2	1	1
F2	1	3	-2	4
F3	4	5	-1	1
F4	2	1	1	1
F5	5	4	1	1
				8

# Le calcul

$$r = 1 - \frac{6 \sum d^2}{N(N^2 - 1)}$$

Nombre de couples  
d'observations

Différence de rangs  
entre les couples  
d'observations

# Le traitement

- On compare la valeur du **Rhô calculé** à celle du **Rhô théorique** (valeur critique) lue dans une table (attention n compris entre 4 et 25)
- **$R_s = 0,6$** . Valeur critique au seuil de 5% pour  $n = 5$  est de **1**
- Pas d'inférence possible.



n	$\alpha = 0.05$	$\alpha = 0.01$	n	$\alpha = 0.05$	$\alpha = 0.01$
5	1.000		31	.356	.459
6	.886	1.000	32	.350	.452
7	.786	.929	33	.345	.446
8	.738	.881	34	.340	.439
9	.700	.833	35	.335	.433
10	.648	.794	36	.330	.427
11	.618	.755	37	.325	.421
12	.587	.727	38	.321	.415
13	.560	.703	39	.317	.410
14	.538	.675	40	.313	.405
15	.521	.654	41	.309	.400
16	.503	.635	42	.305	.395
17	.485	.615	43	.301	.391
18	.472	.600	44	.298	.386
19	.460	.584	45	.294	.382
20	.447	.570	46	.291	.378
21	.435	.556	47	.288	.374
22	.425	.544	48	.285	.370
23	.415	.532	49	.282	.366
24	.406	.521	50	.279	.363
25	.398	.511	60	.255	.331
26	.390	.501	70	.235	.307
27	.382	.491	80	.220	.287
28	.375	.483	90	.207	.271
29	.368	.475	100	.197	.257
30	.362	.467			

Source : <http://calamar.univ-ag.fr/uag/staps/cours/stat/stat.htm>

<b>Objectifs</b>	<b>Analyse appropriée</b>
<p>Tester un <u>lien</u> d'association entre deux <b><u>variables nominales</u></b> (fermée et fermée)  <i>Existe-t-il un lien entre le fait que l'ouvrier se déclare <b>satisfait ou insatisfait de ses conditions de travail</b> et / le <b>type de formation</b> qu'il déclare avoir préalablement suivi ? – Le fait d'être un H/F - selon son âge...</i></p>	<p><b>Test du Chi-deux</b></p>
<p>Tester un lien d'association entre deux <b><u>variables ordinales</u></b>  <i>Existe un lien entre <b>l'attractivité de certaines TIC</b> et le <b>type de loisirs préférés</b> (dans les deux cas , l'individu devait classer des propositions).</i></p>	<p><b>Coefficient de corrélation de Spearman-Kendall</b></p>
<p>Tester un lien d'association entre deux <b><u>variables métriques</u></b> (<b><u>Numériques</u></b>)  <i>Le <b>score d'épanouissement professionnel</b> (de 1 à 8) est-il associée à <b>l'âge des salariés</b> (à l'ancienneté " ? Au salaire)</i></p>	<p><b>Coefficient de corrélation linéaire de Pearson</b></p>

# Coefficient de corrélation

- Compris entre 1 et -1
- Si sa valeur absolue est supérieure à 0,95, il y a une liaison forte entre les deux séries.
- Si la valeur est proche de 1, les séries évoluent dans le même sens, si elle est proche de - 1, elles évoluent en sens opposé.
- La présence d'une corrélation n'est pas forcément la preuve d'une liaison de cause à effet, car les deux séries de données peuvent être influencées conjointement par une troisième.

## Le calcul

$$r(x, y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \times \sqrt{\sum(y_i - \bar{y})^2}}$$

$$\begin{aligned} r_{xy} &= \frac{COV(X, Y)}{\sqrt{V(X) \times V(Y)}} \\ &= \frac{COV(X, Y)}{\sigma_x \times \sigma_y} \end{aligned}$$

Fin de séance....

