

Seance 9.1: Test de chi-carré

Visseho Adjivanou, PhD.

15 March 2023

Au programme aujourd'hui

- 1 Rappel
- 2 Logique du test statistique
- 3 Test de chi-carré
- 4 Labo
- 5 A faire pour la semaine prochaine

Rappel des objectifs de l'analyse descriptive

- 1 Y a-t-il une relation entre deux variables ?
- 2 Quelle est l'intensité de cette relation?
- 3 Quelles sont la direction et la forme de cette relation?

Rappel : association entre deux variables

Type de la variable indépendante			Type de la variable dépendante		
			Catégorielle		Continue
			Nominale	Ordinale	Intervalle/ratio
	Catégorielle	Nominale	Tableau croisé	Tableau croisé	Analyse de la variance
			Test du chi-carré		Test t
			V de Cramer		État carré
		Ordinale	Diagramme de barre empilée		Diagramme boxplot
	Continue	Intervalle/ratio	Transformer la VI en catégorielle		idem
					Corrélation / régression
					Test t
					R carré
					Diagramme de nuage de points

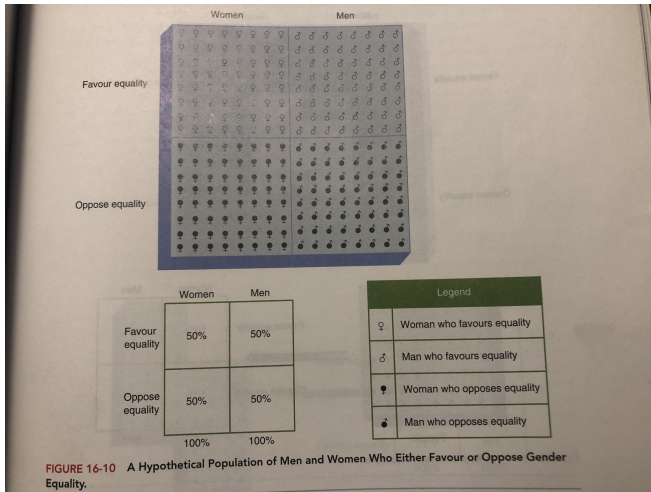
Figure 1: Tableau récapitulatif de l'association entre deux variables

Logique de test statistique

Question

- 4. Pouvons-nous **généraliser** la relation à la population de laquelle est tiré l'échantillon?

Logique de test statistique



Logique de test statistique

On voit que dans cette population, il y a un même pourcentage d'hommes et de femmes qui sont favorables aux droits des femmes. Il n'y a donc pas de relation entre le sexe et l'attitude par rapport à l'égalité.

Logique de test statistique

- Tirage d'un échantillon aléatoire du quart de cette population.

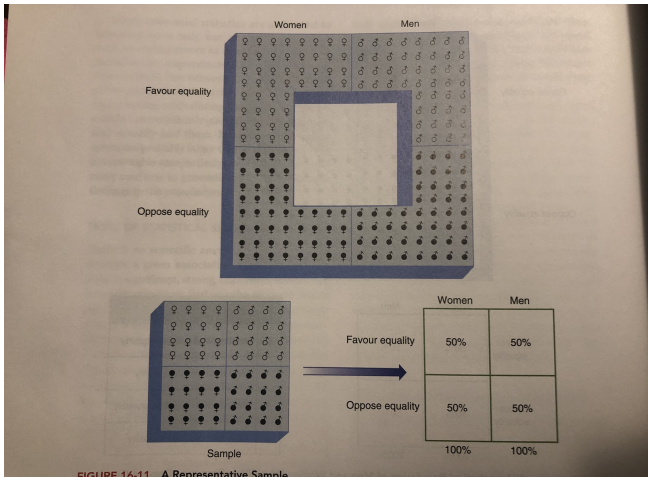


FIGURE 16-11 A Representative Sample.

Logique de test statistique

- On peut voir clairement que cet échantillon est bien représentatif de la population.

Logique de test statistique

- On peut voir clairement que cet échantillon est bien représentatif de la population.
- Il n'existe pas de relation entre les deux variables.

Logique de test statistique

- On peut voir clairement que cet échantillon est bien représentatif de la population.
- Il n'existe pas de relation entre les deux variables.
- Cet échantillon nous aurait ainsi permis de tirer une conclusion correcte sur la population.

Logique de test statistique

- On peut voir clairement que cet échantillon est bien représentatif de la population.
- Il n'existe pas de relation entre les deux variables.
- Cet échantillon nous aurait ainsi permis de tirer une conclusion correcte sur la population.
- A travers la logique de l'échantillonnage, nous concluerons qu'il n'y a pas de relation au sein de la population d'où est tiré l'échantillon.

Logique de test statistique

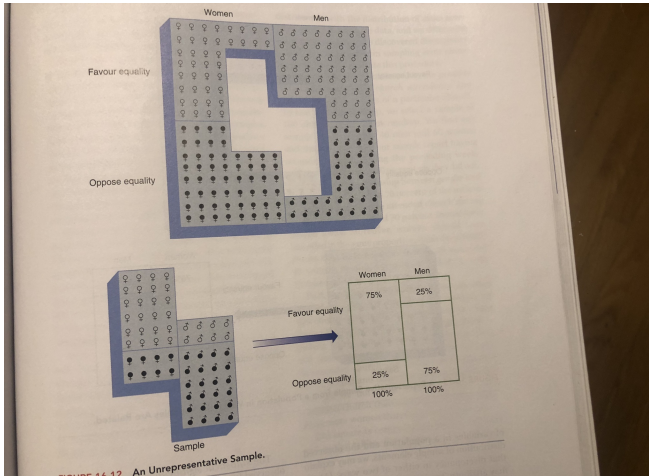
- On peut voir clairement que cet échantillon est bien représentatif de la population.
- Il n'existe pas de relation entre les deux variables.
- Cet échantillon nous aurait ainsi permis de tirer une conclusion correcte sur la population.
- A travers la logique de l'échantillonnage, nous concluerons qu'il n'y a pas de relation au sein de la population d'où est tiré l'échantillon.
- Il est évident aussi qu'on ne serait pas toujours tombé sur un échantillon aussi parfait de la population.

Logique de test statistique

- On peut voir clairement que cet échantillon est bien représentatif de la population.
- Il n'existe pas de relation entre les deux variables.
- Cet échantillon nous aurait ainsi permis de tirer une conclusion correcte sur la population.
- A travers la logique de l'échantillonnage, nous concluerons qu'il n'y a pas de relation au sein de la population d'où est tiré l'échantillon.
- Il est évident aussi qu'on ne serait pas toujours tombé sur un échantillon aussi parfait de la population.
- Il ne serait pas inhabituel de trouver 2 hommes de plus qui sont favorables aux droits des femmes ou inversement.

Logique de test statistique

- Maintenant, regardons un autre échantillon:



Logique de test statistique

- On voit que cet échantillon n'est pas représentatif de la population précédente.

Logique de test statistique

- On voit que cet échantillon n'est pas représentatif de la population précédente.
- Nous avons 75% des femmes qui sont favorables aux droits des femmes contre 25% des hommes.

Logique de test statistique

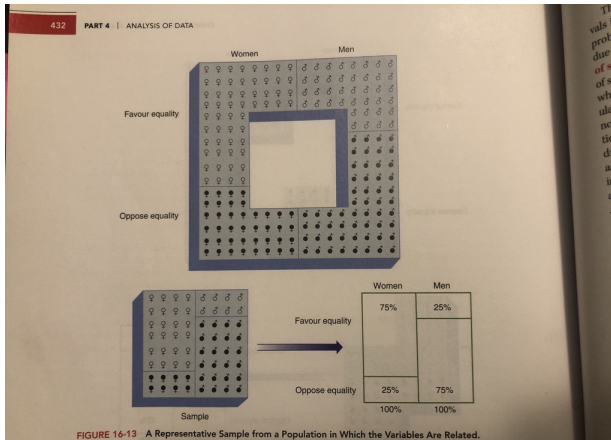
- On voit que cet échantillon n'est pas représentatif de la population précédente.
- Nous avons 75% des femmes qui sont favorables aux droits des femmes contre 25% des hommes.
- Si nous avons tiré un tel échantillon dans une population où il n'y a pas de relation entre les deux variables, nous serions fortement induits en erreur par l'analyse de notre échantillon.

Logique de test statistique

- On voit que cet échantillon n'est pas représentatif de la population précédente.
- Nous avons 75% des femmes qui sont favorables aux droits des femmes contre 25% des hommes.
- Si nous avons tiré un tel échantillon dans une population où il n'y a pas de relation entre les deux variables, nous serions fortement induits en erreur par l'analyse de notre échantillon.
- Selon la logique de l'échantillonnage, il est très peu probable que nous tirions un tel échantillon de notre population où il n'y a pas d'association entre les deux variables.

Logique de test statistique

- Plus probable que cet échantillon provienne plutôt d'une population comme celle-ci:



Logique de test statistique

- L'échantillon tiré de cette population nous indique aussi une relation forte entre le sexe et l'opinion.
- Mais, cette fois-ci, cet échantillon est représentatif de la population.

Logique de test statistique

- Dans les faits:
 - 1 Nous ne connaissons pas la distribution dans la population

Logique de test statistique

- Dans les faits:
 - 1 Nous ne connaissons pas la distribution dans la population
 - 2 Nous ne tirons qu'un seul échantillon.

Logique de test statistique

- Dans les faits:
 - 1 Nous ne connaissons pas la distribution dans la population
 - 2 Nous ne tirons qu'un seul échantillon.
- Si nous faisons face à une association au sein de notre échantillon alors que nous pensons qu'il n'y en a pas dans la population, nous pouvons dire deux choses:

Logique de test statistique

- Dans les faits:
 - ① Nous ne connaissons pas la distribution dans la population
 - ② Nous ne tirons qu'un seul échantillon.
- Si nous faisons face à une association au sein de notre échantillon alors que nous pensons qu'il n'y en a pas dans la population, nous pouvons dire deux choses:
 - i. Soit que notre échantillon n'est pas représentatif ou que

Logique de test statistique

- Dans les faits:
 - ① Nous ne connaissons pas la distribution dans la population
 - ② Nous ne tirons qu'un seul échantillon.
- Si nous faisons face à une association au sein de notre échantillon alors que nous pensons qu'il n'y en a pas dans la population, nous pouvons dire deux choses:
 - i. Soit que notre échantillon n'est pas représentatif ou que
 - ii. Soit qu'il existe bel et bien une association entre les variables dans la population d'où est tirée notre échantillon.

Logique des tests statistiques

- Déterminer la probabilité de découvrir une relation dans notre échantillon quand il y en a au sein de la population.

Logique des tests statistiques

- Déterminer la probabilité de découvrir une relation dans notre échantillon quand il y en a au sein de la population.
- Si cette probabilité est petite ($1/20$ ou 5%, d'où l'idée du seuil de 5%), et si nous découvrons une relation au sein de l'échantillon, nous pourrions conclure qu'il existe probablement une relation dans la population.

Logique des tests statistiques

- Déterminer la probabilité de découvrir une relation dans notre échantillon quand il y en a au sein de la population.
- Si cette probabilité est petite ($1/20$ ou 5%, d'où l'idée du seuil de 5%), et si nous découvrons une relation au sein de l'échantillon, nous pourrions conclure qu'il existe probablement une relation dans la population.
- Alors il s'agit de tester la supposition qu'il **n'existe pas de relation** dans la population.

Logique des tests statistiques

- Déterminer la probabilité de découvrir une relation dans notre échantillon quand il y en a au sein de la population.
- Si cette probabilité est petite (1/20 ou 5%, d'où l'idée du seuil de 5%), et si nous découvrons une relation au sein de l'échantillon, nous pourrions conclure qu'il existe probablement une relation dans la population.
- Alors il s'agit de tester la supposition qu'il **n'existe pas de relation** dans la population.
- On appelle cela l'**hypothèse nulle**, notée H_0 .

Logique des tests statistiques

- Déterminer la probabilité de découvrir une relation dans notre échantillon quand il y en a au sein de la population.
- Si cette probabilité est petite (1/20 ou 5%, d'où l'idée du seuil de 5%), et si nous découvrons une relation au sein de l'échantillon, nous pourrions conclure qu'il existe probablement une relation dans la population.
- Alors il s'agit de tester la supposition qu'il **n'existe pas de relation** dans la population.
- On appelle cela l'**hypothèse nulle**, notée H_0 .
- Ainsi, rejeter l'hypothèse nulle, revient à dire qu'il existe une relation dans la population.

Logique des tests statistiques

- Déterminer la probabilité de découvrir une relation dans notre échantillon quand il y en a au sein de la population.
- Si cette probabilité est petite (1/20 ou 5%, d'où l'idée du seuil de 5%), et si nous découvrons une relation au sein de l'échantillon, nous pourrions conclure qu'il existe probablement une relation dans la population.
- Alors il s'agit de tester la supposition qu'il **n'existe pas de relation** dans la population.
- On appelle cela l'**hypothèse nulle**, notée H_0 .
- Ainsi, rejeter l'hypothèse nulle, revient à dire qu'il existe une relation dans la population.
- On parle de relation **statistiquement significative**

Logique des tests statistiques

- A l'inverse, si les chances de trouver une relation dans l'échantillon alors qu'il n'y en a pas dans la population sont élevées (supérieures à 1 sur 20), nous NE pouvons croire en toute confiance à l'existence d'une relation dans la population.

Logique des tests statistiques

- A l'inverse, si les chances de trouver une relation dans l'échantillon alors qu'il n'y en a pas dans la population sont élevées (supérieures à 1 sur 20), nous NE pouvons croire en toute confiance à l'existence d'une relation dans la population.
- La relation trouvée au sein de l'échantillon est probablement due au hasard seul.

Logique des tests statistiques

- A l'inverse, si les chances de trouver une relation dans l'échantillon alors qu'il n'y en a pas dans la population sont élevées (supérieures à 1 sur 20), nous NE pouvons croire en toute confiance à l'existence d'une relation dans la population.
- La relation trouvée au sein de l'échantillon est probablement due au hasard seul.
- Dans ce cas, nous disons que nous NE rejetons PAS l'hypothèse nulle

Logique des tests statistiques

- A l'inverse, si les chances de trouver une relation dans l'échantillon alors qu'il n'y en a pas dans la population sont élevées (supérieures à 1 sur 20), nous NE pouvons croire en toute confiance à l'existence d'une relation dans la population.
- La relation trouvée au sein de l'échantillon est probablement due au hasard seul.
- Dans ce cas, nous disons que nous NE rejetons PAS l'hypothèse nulle
- La relation que nous obtenons dans l'échantillon est probablement factice

Logique des tests statistiques

- A l'inverse, si les chances de trouver une relation dans l'échantillon alors qu'il n'y en a pas dans la population sont élevées (supérieures à 1 sur 20), nous NE pouvons croire en toute confiance à l'existence d'une relation dans la population.
- La relation trouvée au sein de l'échantillon est probablement due au hasard seul.
- Dans ce cas, nous disons que nous NE rejetons PAS l'hypothèse nulle
- La relation que nous obtenons dans l'échantillon est probablement factice
- Remarque: On dit **ne pas rejeter l'hypothèse nulle** et non **accepter l'hypothèse nulle**.

Logique des tests statistiques

- **Niveau de significativité** = probabilité de retrouver grâce au hasard une relation au sein d'un échantillon, en dépit de l'absence de relation dans la population. Il est noté α .
- On n'utilisera souvent les niveaux de significativité de 5%, 1% et 0.1%.

Logique des tests statistiques

- **Niveau de significativité** = probabilité de retrouver grâce au hasard une relation au sein d'un échantillon, en dépit de l'absence de relation dans la population. Il est noté α .
- On n'utilisera souvent les niveaux de significativité de 5%, 1% et 0.1%.
- Le fait que la signification statistique repose sur une probabilité implique que nous ne pouvons jamais être absolument certains de faire le bon choix.

Logique des tests statistiques

- **Niveau de significativité** = probabilité de retrouver grâce au hasard une relation au sein d'un échantillon, en dépit de l'absence de relation dans la population. Il est noté α .
- On n'utilisera souvent les niveaux de significativité de 5%, 1% et 0.1%.
- Le fait que la signification statistique repose sur une probabilité implique que nous ne pouvons jamais être absolument certains de faire le bon choix.
- Pour le faire, il faut les données de la population directement.

Logique des tests statistiques

- **Niveau de significativité** = probabilité de retrouver grâce au hasard une relation au sein d'un échantillon, en dépit de l'absence de relation dans la population. Il est noté α .
- On n'utilisera souvent les niveaux de significativité de 5%, 1% et 0.1%.
- Le fait que la signification statistique repose sur une probabilité implique que nous ne pouvons jamais être absolument certains de faire le bon choix.
- Pour le faire, il faut les données de la population directement.
- Ainsi, on peut commettre deux types d'erreurs.

Logique des tests statistiques

- Erreur de type I ou erreur **alpha** : rejet de l'hypothèse nulle alors qu'elle est vraie. La probabilité d'une erreur de type 1 est α

Décision concernant H_0	Si H_0 est vraie	Si H_0 est fausse
Rejeter H_0	Erreur de type I	Pas d'erreur
Conserver H_0	Pas d'erreur	Erreur de type II

Logique des tests statistiques

- Erreur de type I ou erreur **alpha** : rejet de l'hypothèse nulle alors qu'elle est vraie. La probabilité d'une erreur de type 1 est α
- Erreur de type II ou erreur **bêta** : Non rejet de l'hypothèse nulle alors qu'elle est fausse.

Décision concernant H_0	Si H_0 est vraie	Si H_0 est fausse
Rejeter H_0	Erreur de type I	Pas d'erreur
Conserver H_0	Pas d'erreur	Erreur de type II

Logique des tests statistiques

- Erreur de type I ou erreur **alpha** : rejet de l'hypothèse nulle alors qu'elle est vraie. La probabilité d'une erreur de type 1 est α
- Erreur de type II ou erreur **bêta** : Non rejet de l'hypothèse nulle alors qu'elle est fausse.
- Si les chances de l'erreur de type I augmentent, les chances de l'erreur de type II diminuent et vice versa

Décision concernant H_0	Si H_0 est vraie	Si H_0 est fausse
Rejeter H_0	Erreur de type I	Pas d'erreur
Conserver H_0	Pas d'erreur	Erreur de type II

Test de chi-carré

Le test du chi-carré

- Le chi-carré, χ^2 , est un nombre qui compare les fréquences observées dans un tableau bivarié aux fréquences auxquelles on devrait s'attendre s'il n'y avait pas du tout de relation entre les deux variables dans la population (les fréquences anticipées).
- Sa formule est:

$$\chi^2 = \sum \frac{(f_o - f_a)^2}{f_a}$$

où :

$$f_a = \left(\frac{\text{Total de la colonne}}{N} \right) * (\text{Total de la rangée})$$

- fréquence anticipée d'une cellule: f_a ;
- fréquence observée d'une cellule: f_o et
- N est le nombre total de cas

Tests de signification statistique de chi-carré

- La distribution d'échantillonnage et la méthode pour déterminer la signification statistique dépendent de la nature des données que nous analysons.

Tests de signification statistique de chi-carré

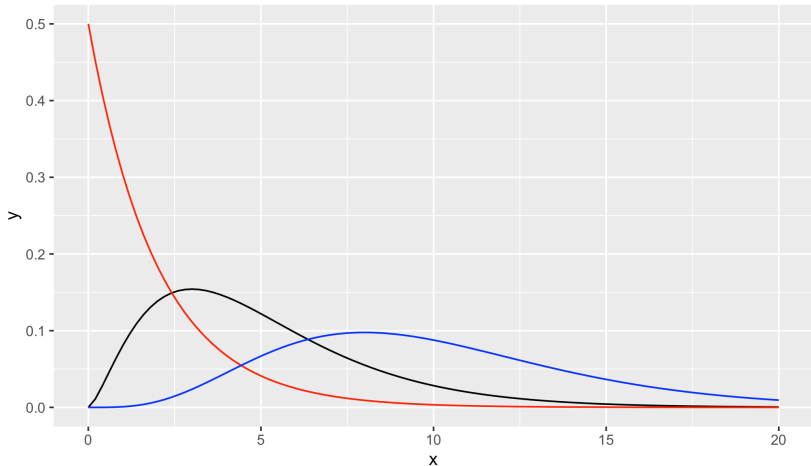
- La distribution d'échantillonnage et la méthode pour déterminer la signification statistique dépendent de la nature des données que nous analysons.
- Pour les données disposées en tableau, le test du **chi-carré** est utilisé.

Tests de signification statistique de chi-carré

- La distribution d'échantillonnage et la méthode pour déterminer la signification statistique dépendent de la nature des données que nous analysons.
- Pour les données disposées en tableau, le test du **chi-carré** est utilisé.
- Il repose sur la distribution d'échantillonnage du chi-carré.

Distribution du chi-carré

Distribution avec du chi carré avec 2(rouge), 5(noir) et 10(bleu) degrés de liberté



Test de signification de chi-carré

Le test de chi-carré

- Il existe une table de la distribution d'échantillonnage du χ^2 , c'est-à-dire une table qui donne les probabilités d'obtenir un χ^2 au moins aussi grand qu'une certaine valeur si, dans la population de laquelle fut tiré l'échantillon, il n'y a pas de relation entre les deux variables.

Le test de chi-carré

- Il existe une table de la distribution d'échantillonnage du χ^2 , c'est-à-dire une table qui donne les probabilités d'obtenir un χ^2 au moins aussi grand qu'une certaine valeur si, dans la population de laquelle fut tiré l'échantillon, il n'y a pas de relation entre les deux variables.
- Cette probabilité dépend de ce qu'on appelle les **degrés de liberté (dl ou ddl)**.

Degré de liberté

- Supposons que je vous demande de trouver 3 chiffres dont la moyenne vaut 11.
- Quelles sont les réponses possibles?

Degré de liberté

- Supposons que je vous demande de trouver 3 chiffres dont la moyenne vaut 11.
- Quelles sont les réponses possibles?
- On peut avoir (11, 11, 11), (11, 12, 10), ...

Degré de liberté

- Supposons que je vous demande de trouver 3 chiffres dont la moyenne vaut 11.
- Quelles sont les réponses possibles?
- On peut avoir (11, 11, 11), (11, 12, 10), ...
- Vous vous rendez compte qu'il y a une infinité de solution

Degré de liberté

- Maintenant, supposez que je vous dise qu'un de ces chiffres vaut 10.
- Quelles sont les réponses possibles?

Degré de liberté

- Maintenant, supposez que je vous dise qu'un de ces chiffres vaut 10.
- Quelles sont les réponses possibles?
- On peut avoir (10, 11, 12), (10, 13, 10), ...

Degré de liberté

- Maintenant, supposez que je vous dise qu'un de ces chiffres vaut 10.
- Quelles sont les réponses possibles?
- On peut avoir (10, 11, 12), (10, 13, 10), ...
- Vous vous rendez compte à nouveau que vous avez une infinité de solutions.

Degré de liberté

- Finalement, supposez que je vous donne deux chiffres, soit (9 et 12)
- Quelles sont les réponses possibles?

Degré de liberté

- Finalement, supposez que je vous donne deux chiffres, soit (9 et 12)
- Quelles sont les réponses possibles?
- On peut avoir (9, 12, 12), ou ...

Degré de liberté

- Finalement, supposez que je vous donne deux chiffres, soit (9 et 12)
- Quelles sont les réponses possibles?
- On peut avoir (9, 12, 12), ou ...
- Vous vous rendez compte que c'est la seule réponse possible

Degré de liberté

- Finalement, supposez que je vous donne deux chiffres, soit (9 et 12)
- Quelles sont les réponses possibles?
- On peut avoir (9, 12, 12), ou ...
- Vous vous rendez compte que c'est la seule réponse possible
- Ainsi, à travers cet exemple, on se rend compte que seuls deux chiffres sont "libres". Une fois qu'ils sont fixés, la dernière réponse est déterminée.

Degré de liberté

- Finalement, supposez que je vous donne deux chiffres, soit (9 et 12)
- Quelles sont les réponses possibles?
- On peut avoir (9, 12, 12), ou ...
- Vous vous rendez compte que c'est la seule réponse possible
- Ainsi, à travers cet exemple, on se rend compte que seuls deux chiffres sont "libres". Une fois qu'ils sont fixés, la dernière réponse est déterminée.
- Dans cette situation, nous disons qu'il y a 2 degrés de liberté.

Degré de liberté

- Finalement, supposez que je vous donne deux chiffres, soit (9 et 12)
- Quelles sont les réponses possibles?
- On peut avoir (9, 12, 12), ou ...
- Vous vous rendez compte que c'est la seule réponse possible
- Ainsi, à travers cet exemple, on se rend compte que seuls deux chiffres sont "libres". Une fois qu'ils sont fixés, la dernière réponse est déterminée.
- Dans cette situation, nous disons qu'il y a 2 degrés de liberté.
- De manière générale, chaque fois que nous regardons la moyenne de N valeurs, le degré de liberté vaut toujours $(N-1)$

Degré de liberté

Les degrés de liberté font référence au nombre maximum de valeurs logiquement indépendantes, qui sont des valeurs qui ont la liberté de varier, dans l'échantillon de données.

Degré de liberté dans un tableau bivarié

- On peut aussi déterminer le degré de liberté dans un tableau
- Maintenant, supposer que je vous donne ce tableau
- Quelles sont les valeurs possibles de A, B, C, et D?

Attitude envers l'avortement	Homme	Femme	Total
-----	-----	-----	-----
Approuve	A	C	300
Désapprouve	B	D	100
Total	200	200	400
-----	-----	-----	-----

Figure 5: ddl

- Ici aussi, vous avez plusieurs réponses possibles

Degré de liberté dans un tableau bivarié

- On peut aussi déterminer le degré de liberté dans un tableau
- Maintenant, supposer que je vous donne ce tableau
- Quelles sont les valeurs possibles de A, B, C, et D?

Attitude envers l'avortement	Homme	Femme	Total
-----	-----	-----	-----
Approuve	A	C	300
Désapprouve	B	D	100
Total	200	200	400
-----	-----	-----	-----

Figure 5: ddl

- Ici aussi, vous avez plusieurs réponses possibles
- On peut avoir (100, 100, 200, 0), (110, 90, 190, 10)

Degré de liberté dans un tableau bivarié

- Maintenant, supposer que je vous donne $A = 150$
- Quelles sont les autres valeurs possibles?

Attitude envers l'avortement	Homme	Femme	Total
-----	-----	-----	-----
Approuve	150	C	300
Désapprouve	B	D	100
Total	200	200	400
-----	-----	-----	-----

Figure 6: ddl2

- On se rend compte que les valeurs de (B, C, D) sont fixes

Degré de liberté dans un tableau bivarié

- Maintenant, supposer que je vous donne $A = 150$
- Quelles sont les autres valeurs possibles?

Attitude envers l'avortement	Homme	Femme	Total
-----	-----	-----	-----
Approuve	150	C	300
Désapprouve	B	D	100
Total	200	200	400
-----	-----	-----	-----

Figure 6: ddl2

- On se rend compte que les valeurs de (B, C, D) sont fixes
- $(B = 50, C = 150, D = 50)$ est la seule réponse possible

Degré de liberté dans un tableau bivarié

- Maintenant, supposer que je vous donne $A = 150$
- Quelles sont les autres valeurs possibles?

Attitude envers l'avortement	Homme	Femme	Total
-----	-----	-----	-----
Approuve	150	C	300
Désapprouve	B	D	100
Total	200	200	400
-----	-----	-----	-----

Figure 6: ddl2

- On se rend compte que les valeurs de (B, C, D) sont fixes
- $(B = 50, C = 150, D = 50)$ est la seule réponse possible
- Ainsi, nous disons que le tableau a 1 degré de liberté

Degré de liberté dans un tableau bivarié

- Maintenant, réfléchis à un tableau de deux colonnes et trois rangés
- Trouve le degré de liberté dans ce tableau si tu connais les informations à la marge?

Degré de liberté dans un tableau bivarié

- En règle générale, le degré de liberté dans les tableaux bivariés vaut $(r-1)(c-1)$

Degré de liberté dans un tableau bivarié

- En règle générale, le degré de liberté dans les tableaux bivariés vaut $(r-1)(c-1)$
- r étant le nombre de rangées et

Degré de liberté dans un tableau bivarié

- En règle générale, le degré de liberté dans les tableaux bivariés vaut $(r-1)(c-1)$
- r étant le nombre de rangées et
- et c le nombre de colonnes dans le tableau.

Degré de liberté dans un tableau bivarié

- En règle générale, le degré de liberté dans les tableaux bivariés vaut $(r-1)(c-1)$
- r étant le nombre de rangées et
- et c le nombre de colonnes dans le tableau.
- A partir de cette valeur, on peut lire dans le tableau de distribution du chi-carré :

Degré de liberté dans un tableau bivarié

- En règle générale, le degré de liberté dans les tableaux bivariés vaut $(r-1)(c-1)$
- r étant le nombre de rangées et
- et c le nombre de colonnes dans le tableau.
- A partir de cette valeur, on peut lire dans le tableau de distribution du chi-carré :
- la valeur minimale du chi-carré nécessaire pour obtenir un résultat statistiquement significatif au seuil (ou niveau de significativité) voulu (0.05, 0.02, 0.01 ou 0.001).

Tableau de distribution du chi-deux

- page 343 de votre livre de cours

	<i>p</i>					
<i>v</i>	0.100	0.050	0.025	0.010	0.005	0.001
1	2.7055	3.8415	5.0239	6.6349	7.8794	10.8276
2	4.6052	5.9915	7.3778	9.2103	10.5966	13.8155
3	6.2514	7.8147	9.3484	11.3449	12.8382	16.2662
4	7.7794	9.4877	11.1433	13.2767	14.8603	18.4668
5	9.2364	11.0705	12.8325	15.0863	16.7496	20.5150
6	10.6446	12.5916	14.4494	16.8119	18.5476	22.4577
7	12.0170	14.0671	16.0128	18.4753	20.2777	24.3219
8	13.3616	15.5073	17.5345	20.0902	21.9550	26.1245
9	14.6837	16.9190	19.0228	21.6660	23.5893	27.8772
10	15.9872	18.3070	20.4832	23.2093	25.1882	29.5883
11	17.2750	19.6751	21.9200	24.7250	26.7568	31.2641
12	18.5493	21.0261	23.3367	26.2170	28.2995	32.9095
13	19.8119	22.3620	24.7356	27.6882	29.8195	34.5282
14	21.0641	23.6848	26.1189	29.1412	31.3193	36.1233
15	22.3071	24.9958	27.4884	30.5779	32.8013	37.6973
16	23.5418	26.2962	28.8454	31.9999	34.2672	39.2524
17	24.7690	27.5871	30.1910	33.4087	35.7185	40.7902
18	25.9894	28.8693	31.5264	34.8053	37.1564	42.3124
19	27.2036	30.1435	32.8523	36.1909	38.5823	43.8202
20	28.4120	31.4104	34.1696	37.5662	39.9968	45.3147
21	29.6151	32.6706	35.4789	38.9322	41.4011	46.7970
22	30.8133	33.9244	36.7807	40.2894	42.7957	48.2679
23	32.0069	35.1725	38.0756	41.6384	44.1813	49.7282

Tableau de distribution du chi-deux

Ce tableau nous donne :

- Sur la première ligne le niveau de significativité souhaité (par exemple 5%)

Tableau de distribution du chi-deux

Ce tableau nous donne :

- Sur la première ligne le niveau de significativité souhaité (par exemple 5%)
- Sur la première colonne le degré de liberté calculé (par exemple 9)

Tableau de distribution du chi-deux

Ce tableau nous donne :

- Sur la première ligne le niveau de significativité souhaité (par exemple 5%)
- Sur la première colonne le degré de liberté calculé (par exemple 9)
- L'intersection de la ligne et de cette colonne vous donne la **valeur minimale** du chi-deux pour obtenir un résultat significatif au seuil voulu.

Tableau de distribution du chi-deux

Ce tableau nous donne :

- Sur la première ligne le niveau de significativité souhaité (par exemple 5%)
- Sur la première colonne le degré de liberté calculé (par exemple 9)
- L'intersection de la ligne et de cette colonne vous donne la **valeur minimale** du chi-deux pour obtenir un résultat significatif au seuil voulu.
- Dans l'exemple ici, la valeur du chi-deux vaut : 16.9190

Tableau de distribution du chi-deux

Ce tableau nous donne :

- Sur la première ligne le niveau de significativité souhaité (par exemple 5%)
- Sur la première colonne le degré de liberté calculé (par exemple 9)
- L'intersection de la ligne et de cette colonne vous donne la **valeur minimale** du chi-deux pour obtenir un résultat significatif au seuil voulu.
- Dans l'exemple ici, la valeur du chi-deux vaut : 16.9190
- Cela signifie que si nous devons prendre 100 échantillons (aléatoires) différents dans une population où il n'y a pas d'association entre les deux variables, dans 5% des cas, nous obtiendrons une valeur égale ou supérieure à 16.9190.

Tableau de distribution du chi-deux

Ce tableau nous donne :

- Sur la première ligne le niveau de significativité souhaité (par exemple 5%)
- Sur la première colonne le degré de liberté calculé (par exemple 9)
- L'intersection de la ligne et de cette colonne vous donne la **valeur minimale** du chi-deux pour obtenir un résultat significatif au seuil voulu.
- Dans l'exemple ici, la valeur du chi-deux vaut : 16.9190
- Cela signifie que si nous devons prendre 100 échantillons (aléatoires) différents dans une population où il n'y a pas d'association entre les deux variables, dans 5% des cas, nous obtiendrons une valeur égale ou supérieure à 16.9190.
- Au seuil de 1%, on trouve la valeur du chi-carré égale à 21.66.

En conclusion

- Je me donne 5% de chance de me tromper (niveau de significativité ou erreur alpha)
- Avec ce niveau et le degré de liberté, je lis la valeur du chi-carré qui leur est associée

En conclusion

- Je me donne 5% de chance de me tromper (niveau de significativité ou erreur alpha)
- Avec ce niveau et le degré de liberté, je lis la valeur du chi-carré qui leur est associée
- Si le chi-carré que je calcule avec mon échantillon est **supérieur** à cette valeur, je suis bien en phase avec le risque que je me suis donné: je dis qu'il y a une association au sein de la population

En conclusion

- Je me donne 5% de chance de me tromper (niveau de significativité ou erreur alpha)
- Avec ce niveau et le degré de liberté, je lis la valeur du chi-carré qui leur est associée
- Si le chi-carré que je calcule avec mon échantillon est **supérieur** à cette valeur, je suis bien en phase avec le risque que je me suis donné: je dis qu'il y a une association au sein de la population
- Autrement dit, il est peu probable que mon échantillon provienne d'une population dans laquelle il n'y a pas d'association

Décision

- Si votre **chi-deux calculé** est supérieur ou égal au **chi-deux lu**
==> Rejeter l'hypothèse nulle. Cela revient à dire qu'il existe une association significative (n'est pas du au hasard) entre vos deux variables (catégorielles)

Décision

- Si votre **chi-deux calculé** est supérieur ou égal au **chi-deux lu** \implies Rejeter l'hypothèse nulle. Cela revient à dire qu'il existe une association significative (n'est pas du au hasard) entre vos deux variables (catégorielles)
- Si votre **chi-deux calculé** est inférieur au **chi-deux lu** \implies Vous ne pouvez pas rejeter l'hypothèse nulle. Vous ne pouvez pas conclure à l'existence d'une association non nulle entre vos deux variables.

En résumé

Pour faire un test d'association entre deux variables (catégorielles):

- ➊ Posez votre hypothèse nulle (et alternative)
 - H_0 : Il **n'existe pas** d'association entre les deux variables
 - H_1 : Il **existe** une association non nulle entre les deux variables
- ➋ Choisissez votre niveau de significativité (5%)

➌ *Prendre une décision:*

- ***chi-deux calculé \geq chi-deux lu \implies Rejeter l'hypothèse nulle. \implies Relation est **statistiquement*****

En résumé

Pour faire un test d'association entre deux variables (catégorielles):

- ➊ Posez votre hypothèse nulle (et alternative)
 - H_0 : Il **n'existe pas** d'association entre les deux variables
 - H_1 : Il **existe** une association non nulle entre les deux variables
- ➋ Choisissez votre niveau de significativité (5%)
- ➌ Trouvez votre degré de liberté
- ➍
- ➎ *Prendre une décision:*
 - ***chi-deux calculé \geq chi-deux lu \implies Rejeter l'hypothèse nulle. \implies Relation est **statistiquement*****

En résumé

Pour faire un test d'association entre deux variables (catégorielles):

- ➊ Posez votre hypothèse nulle (et alternative)
 - H_0 : Il **n'existe pas** d'association entre les deux variables
 - H_1 : Il **existe** une association non nulle entre les deux variables
- ➋ Choisissez votre niveau de significativité (5%)
- ➌ Trouvez votre degré de liberté
- ➍ Trouvez la valeur du **chi-deux** pour rejeter l'hypothèse nulle
- ➎ *Prendre une décision:*
 - ***chi-deux calculé \geq chi-deux lu \implies Rejeter l'hypothèse nulle. \implies Relation est **statistiquement*****

En résumé

Pour faire un test d'association entre deux variables (catégorielles):

- ➊ Posez votre hypothèse nulle (et alternative)
 - H_0 : Il **n'existe pas** d'association entre les deux variables
 - H_1 : Il **existe** une association non nulle entre les deux variables
- ➋ Choisissez votre niveau de significativité (5%)
- ➌ Trouvez votre degré de liberté
- ➍ Trouvez la valeur du **chi-deux** pour rejeter l'hypothèse nulle
- ➎ Calculez votre chi-deux à partir de votre échantillon
- ➏ *Prendre une décision:*
 - ***chi-deux calculé \geq chi-deux lu \implies Rejeter l'hypothèse nulle. \implies Relation est **statistiquement*****

Remarques

- 1 Une relation qui est significative au seuil de 5% est **aussi significative** à tous les seuils supérieurs à 5%.
- 2 Une relation qui est significative au seuil de 5% n'est pas automatiquement significative au seuil plus faible (par exemple 1%)

Remarques

- 1 Une relation qui est significative au seuil de 5% est **aussi significative** à tous les seuils supérieurs à 5%.
- 2 Une relation qui est significative au seuil de 5% n'est pas automatiquement significative au seuil plus faible (par exemple 1%)
- 3 Plus grande est la valeur du chi-carré calculée, peu probable est que cette valeur soit attribuée à une erreur d'échantillonnage

Remarques

- 1 Une relation qui est significative au seuil de 5% est **aussi significative** à tous les seuils supérieurs à 5%.
- 2 Une relation qui est significative au seuil de 5% n'est pas automatiquement significative au seuil plus faible (par exemple 1%)
- 3 Plus grande est la valeur du chi-carré calculée, peu probable est que cette valeur soit attribuée à une erreur d'échantillonnage
- 4 Une relation statistiquement significative n'est pas toujours une **signification substantielle**. Avec une grande taille d'échantillon, les très petites différences peuvent être significatives

Remarques

- 1 Une relation qui est significative au seuil de 5% est **aussi significative** à tous les seuils supérieurs à 5%.
- 2 Une relation qui est significative au seuil de 5% n'est pas automatiquement significative au seuil plus faible (par exemple 1%)
- 3 Plus grande est la valeur du chi-carré calculée, peu probable est que cette valeur soit attribuée à une erreur d'échantillonnage
- 4 Une relation statistiquement significative n'est pas toujours une **signification substantielle**. Avec une grande taille d'échantillon, les très petites différences peuvent être significatives
- 5 Le test de chi-deux est basé sur un échantillon issu d'une procédure d'échantillonnage probabilistique

Remarques

- ⑥ Pour les autres tests que nous serons amenés à faire (test de Student de comparaison des moyennes, test de Fisher pour l'analyse de la variance), la seule différence serait le changement de la distribution de la statistique à consider.
- Pour tous les autres tests, on suivra la même logique
- Étapes 1 à 3 sont pareilles
- Étape 4, trouver la valeur de la statistique appropriée
- Étape 5, calculer votre statistique
- Étape 6: prendre une décision

Labo

Pour la semaine prochaine

- Lire les chapitres 6 (à nouveau) et le chapitre 7
- Refaire le labo chez vous.