Centre d'Immunologie et des Maladies Infectieuses

4V001 - Ateliers de biologie moleculaire et cellulaire

Biostatistiques appliquées

Martin LARSEN

www.immulab.fr/cms/index.php/teaching

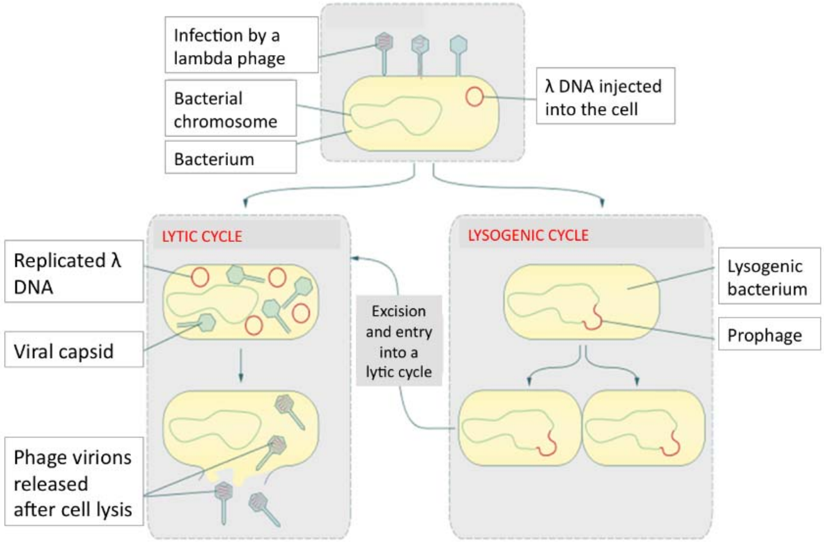
INSERM U1135, CHU Pitié-Salpetrière, Paris, France

Hypothèse: Etape par étape

1. Définir le problème
2. Indiquer les objectifs
3. Indiquer l'hypothèse nulle (H0)
4. Indiquer l'hypothèse alternative (H1)
5. Sélectionner le test statistique approprié
6. Décider si le test de l'hypothèse sera testé par test unilatéral ou bilatéral
7. Indiquer le niveau de risque alpha (α)
8. Indiquer le niveau de risque bêta (1-β)
9. Indiquer ou établisser (nécessite des connaissances antérieures) la taille d'effet.
10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon
11. Rassembler les échantillons
12. Collecter et pré-analyser les données
13. Calculer le test statistique
14. Déterminer la valeur de test critique et la valeur p
15. Si p-value < α, rejeter H0
16. Si p-value > α, ne parvient pas à rejeter H0
17. Analyse *post hoc*

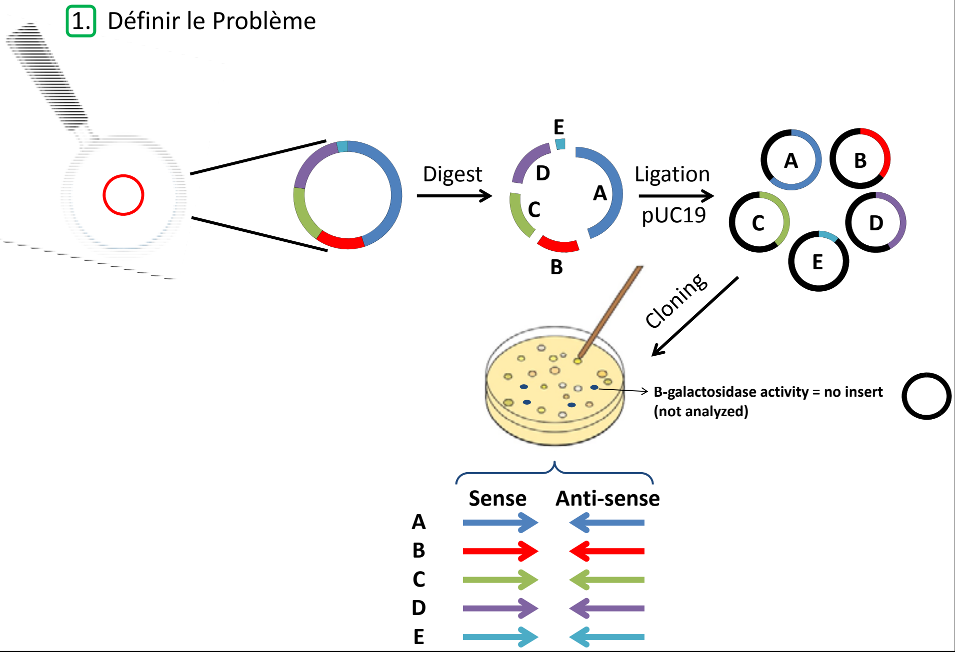
Tester l'hypothèse- Quel est le problème?

1. Définir le Problème



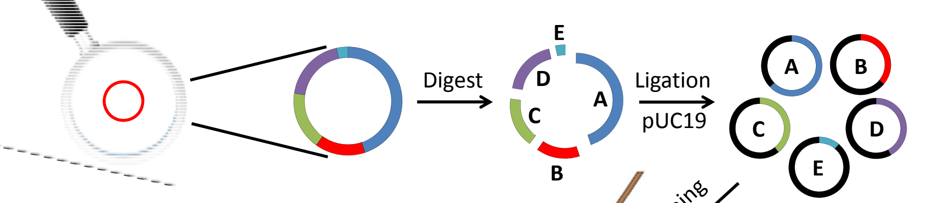
Tester l'hypothèse- Quel est le problème?

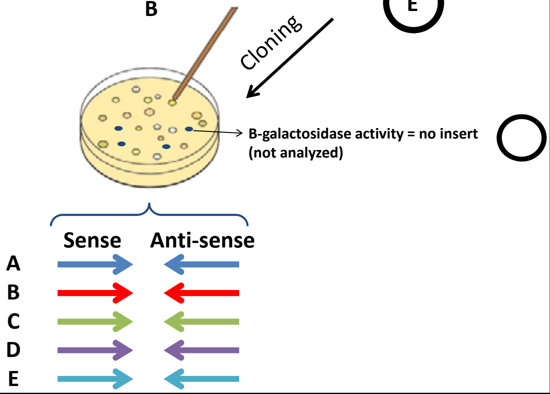
1. Définir le Problème



Tester l'hypothèse- Objectifs

1. Définir le Problème
2. Etablir les Objectifs





Objectifs:

1)

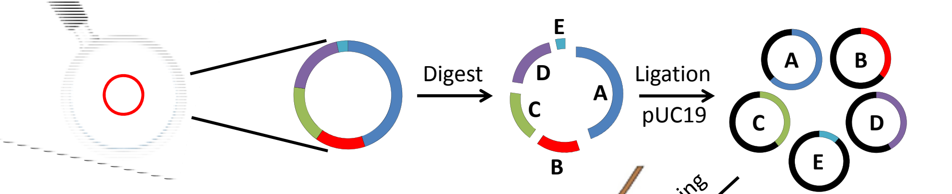
?

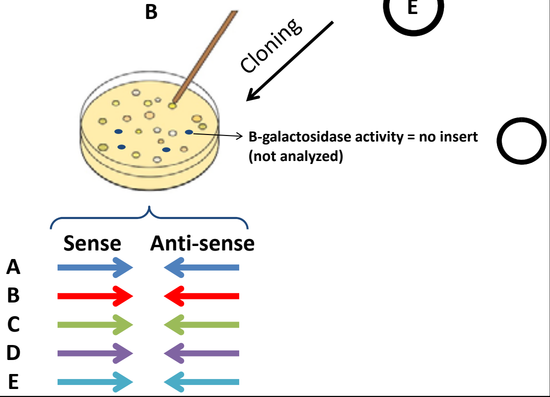
2)

?

Tester l'hypothèse- Objectifs

1. Définir le Problème
2. Etablir les Objectifs





Objectifs:

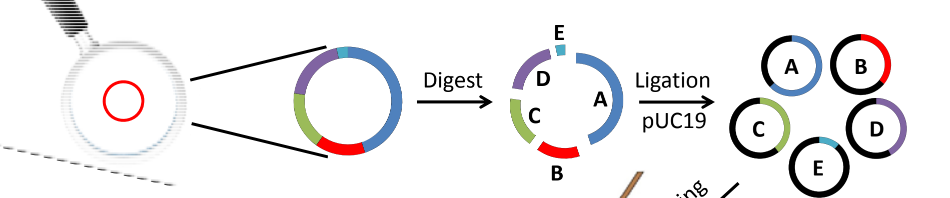
1) La ligation génère-t-elle une répartition uniforme de chaque fragment (A,B,C,D et E)?

2)

?

Tester l’hypothèse- Objectifs

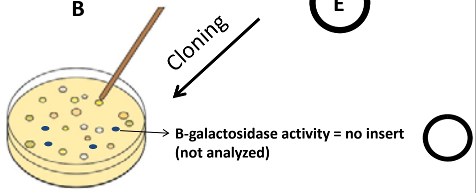
1. Définir le Problème
2. Etablir les Objectifs

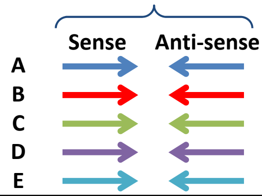


Objectifs:

1) La ligation génère-t-elle une répartition uniforme de chaque fragment (A,B,C,D et E)?

2) Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?





Tester l’hypothèse- Formuler l’hypothèse

1. Définir le Problème
2. Etablir les objectifs
   - A) La ligation génère-t-elle une répartition uniforme de chaque fragment (A,B,C,D et E)?
   - B) Les sens et les anti-sens sont-ils également représentés de façon égale pour chaque fragment?
3. Formuler l’hypothèse nulle (H_0)
   - H_0: les fragments sont distribués de façon
     - Obj. 1: ?
     - Obj. 2: ?
4. Formuler l’hypothèse alternative (H_1)
   - H_1: les fragments ne sont pas distribués de façon
     - Obj. 1: ?
     - Obj. 2: ?

Tester l'hypothèse- Formuler l'hypothèse

1. Définir le Problème
2. Etablir les objectifs
   - A) La ligation génère-t-elle une répartition uniforme de chaque fragment (A,B,C,D et E)?
   - B) Les sens et les anti-sens sont-ils également représentés de façon égale pour chaque fragment?
3. Formuler l'hypothèse nulle (H_0)
   - H_0: les fragments sont distribués de façon
     - Obj. 1: uniforme
     - Obj. 2: ?
4. Formuler l'hypothèse alternative (H_1)
   - H_1: les fragments ne sont pas distribués de façon
     - Obj. 1: uniforme
     - Obj. 2: ?

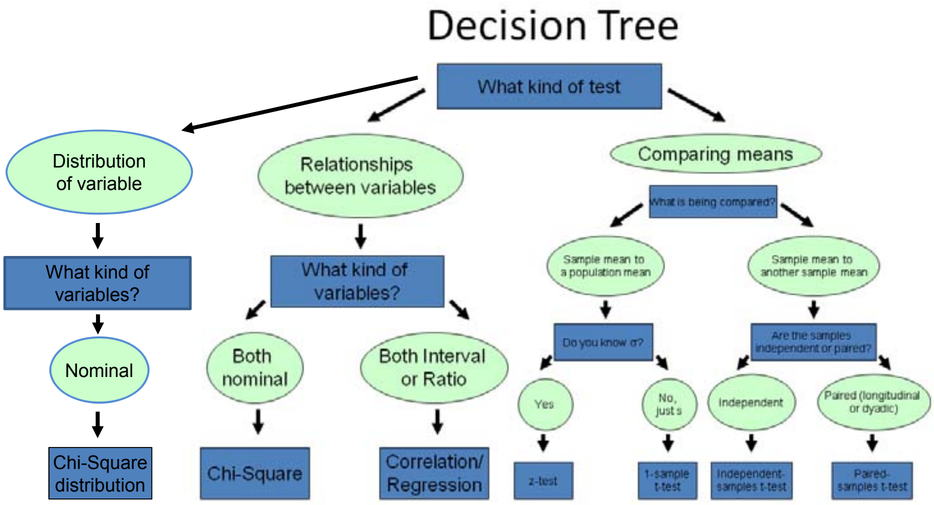
Tester l'hypothèse- Formuler l'hypothèse

1. Définir le Problème
2. Etablir les objectifs
   - A) La ligation génère-t-elle une répartition uniforme de chaque fragment (A,B,C,D et E)?
   - B) Les sens et les anti-sens sont-ils également représentés de façon égale pour chaque fragment?
3. Formuler l'hypothèse nulle (H_0)
   - H_0: les fragments sont distribués de façon
     - Obj. 1: uniforme
     - Obj. 2: proportionnelle
4. Formuler l'hypothèse alternative (H_1)
   - H_1: les fragments ne sont pas distribués de façon
     - Obj. 1: uniforme
     - Obj. 2: proportionnelle

Tester l'hypothèse - Quel test?

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

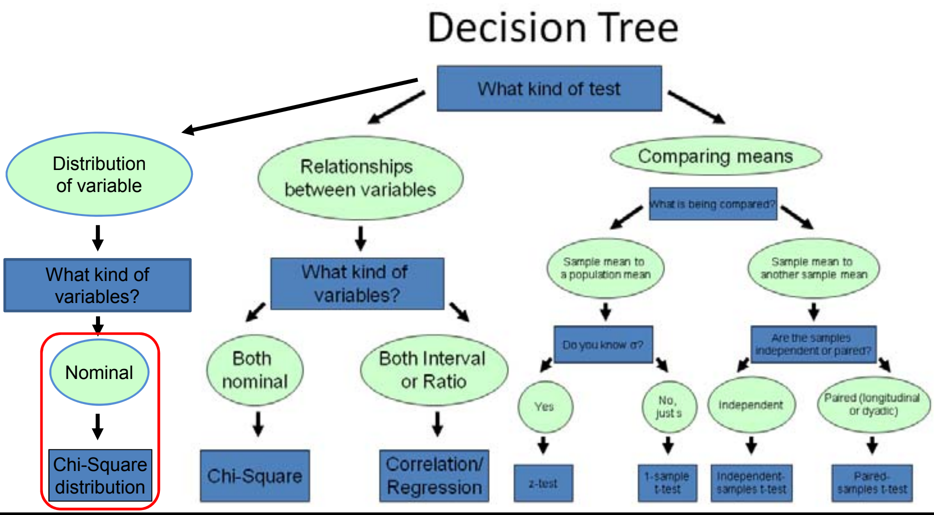
5. Sélectionner le test statistique approprié



Tester l'hypothèse - Quel test?

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

5. Sélectionner le test statistique approprié



Tester l'hypothèse - Quel test χ^2 ?

χ^2 – conformité

Variable mesurée

O_1	O_2	O_3	O_4

O_i est la ième valeur observée

Les valeurs attendues sont connues à priori

χ^2 - homogeneity

Variable mesurée

Variable de stratification de groupe	O_1	O_2
	O_3	O_4

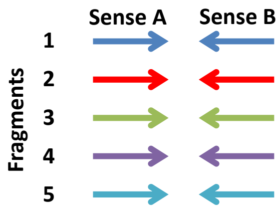
Les valeurs attendues sont définies en fonction des H0, souvent totaux des colonnes et des lignes (distribution proportionnelle)

χ^2 - independence

Variable 1 mesurée

Variable 2 mesurée	O_1	O_2
	O_3	O_4

Data structure



	A	B
1	O_{1A}	O_{1B}
2	O_{2A}	O_{2B}
3	O_{3A}	O_{3B}
4	O_{4A}	O_{4B}
5	O_{5A}	O_{5B}

Objectifs:

1) La ligation génère-t-elle une répartition uniforme de chaque fragment (A,B,C,D et E)?
2) Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Tester l'hypothèse - Quel test χ^2 ?

χ^2 - conformité

Variable mesurée

O_1	O_2	O_3	O_4

O_i est la ième valeur observée

Les valeurs attendues sont connues à priori

χ^2 - homogeneity

Variable mesurée

Variable de stratification de groupe	O_1	O_2
	O_3	O_4

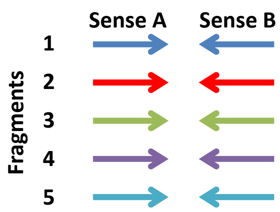
Les valeurs attendues sont définies en fonction des H0, souvent totaux des colonnes et des lignes (distribution proportionnelle)

χ^2 - independence

Variable 1 mesurée

Variable 2 mesurée	O_1	O_2
	O_3	O_4

Data structure



	A+B
1	$O_{1A}+O_{1B}$
2	$O_{2A}+O_{2B}$
3	$O_{3A}+O_{3B}$
4	$O_{4A}+O_{4B}$
5	$O_{5A}+O_{5B}$

Objectifs:

1) La ligation génère-t-elle une répartition uniforme de chaque fragment (A,B,C,D et E)?
2) Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Tester l'hypothèse - Quel test χ^2 ?

χ^2 - conformité

Variable mesurée

O_1	O_2	O_3	O_4

O_i est la ième valeur observée

Les valeurs attendues sont connues a priori

χ^2 - homogeneity

Variable mesurée

Variable de stratification de groupe		
	O_1	O_2
	O_3	O_4

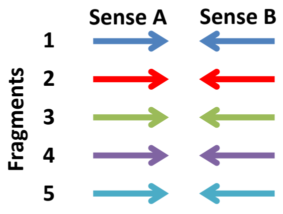
Les valeurs attendues sont définies en fonction des H0, souvent totaux des colonnes et des lignes (distribution proportionnelle)

χ^2 - independence

Variable 1 mesurée

Variable 2 mesurée		
	O_1	O_2
	O_3	O_4

Data structure



	A+B	E(x)
1	$O_{1A}+O_{1B}$	?
2	$O_{2A}+O_{2B}$	?
3	$O_{3A}+O_{3B}$	?
4	$O_{4A}+O_{4B}$	?
5	$O_{5A}+O_{5B}$	?

Objectifs:

1) La ligation génère-t-elle une répartition uniforme de chaque fragment (A,B,C,D et E)?
2) Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Tester l'hypothèse - Quel test χ^2 ?

χ^2 - conformité

Variable mesurée

O_1	O_2	O_3	O_4

O_i est la ième valeur observée

Les valeurs attendues sont connues à priori. En effet, elles sont uniformèment distribuées (1/5 chacun)

χ^2 - homogeneity

Variable mesurée

Variable de stratification de groupe		
	O_1	O_2
	O_3	O_4

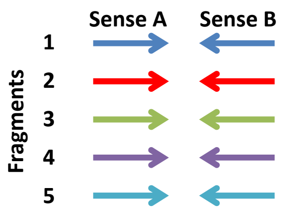
Les valeurs attendues sont définies en fonction des H0, souvent totaux des colonnes et des lignes (distribution proportionnelle)

χ^2 - independence

Variable 1 mesurée

Variable 2 mesurée		
	O_1	O_2
	O_3	O_4

Data structure



	A+B	E(x)
1	$O_{1A}+O_{1B}$	1/5
2	$O_{2A}+O_{2B}$	1/5
3	$O_{3A}+O_{3B}$	1/5
4	$O_{4A}+O_{4B}$	1/5
5	$O_{5A}+O_{5B}$	1/5

Objectifs: χ^2 - conformité

1) La ligation génère-t-elle une répartition uniforme de chaque fragment (A,B,C,D et E)?
2) Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Tester l'hypothèse - Quel test χ^2 ?

χ^2 - conformité

Variable mesurée

O_1	O_2	O_3	O_4

O_i est la ième valeur observée

Les valeurs attendues sont connues à priori

χ^2 - homogeneity

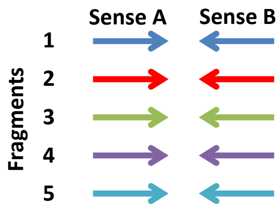
	Variable mesurée	
Variable de stratification de groupe	O_1	O_2
	O_3	O_4

Les valeurs attendues sont définies en fonction des H0, souvent totaux des colonnes et des lignes (distribution proportionnelle)

χ^2 - independence

	Variable 1 mesurée	
Variable 2 mesurée	O_1	O_2
	O_3	O_4

Data structure



	A	B
1	O_{1A}	O_{1B}
2	O_{2A}	O_{2B}
3	O_{3A}	O_{3B}
4	O_{4A}	O_{4B}
5	O_{5A}	O_{5B}

Objectifs:

1) La ligation génère-t-elle une répartition uniforme de chaque fragment (A,B,C,D et E)?
2) Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Tester l'hypothèse - Quel test χ^2 ?

χ^2 - conformité

Variable mesurée

O_1	O_2	O_3	O_4

O_i est la ième valeur observée

Les valeurs attendues sont connues à priori

χ^2 - homogeneity

	Variable mesurée	
Variable de stratification de groupe	O_1	O_2
	O_3	O_4

Les valeurs attendues sont définies en fonction des H0, souvent totaux des colonnes et des lignes (distribution proportionnelle)

χ^2 - independence

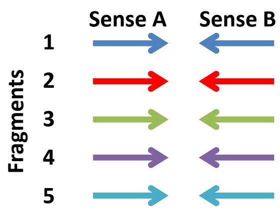
	Variable 1 mesurée	
Variable 2 mesurée	O_1	O_2
	O_3	O_4

$$E_{ij} = \frac{n_i \cdot n_j}{n}$$

e.g.

$$E_{1A} = \frac{n_1 \times n_A}{n}$$

Data structure



	A	B	Total
1	O_{1A}	O_{1B}	n_1
2	O_{2A}	O_{2B}	n_2
3	O_{3A}	O_{3B}	n_3
4	O_{4A}	O_{4B}	n_4
5	O_{5A}	O_{5B}	n_5
Total	n_A	n_B	n

Objectifs:

1) La ligation génère-t-elle une répartition uniforme de chaque fragment?

χ^2 - independence

2) Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Tester l'hypothèse - Quel test χ^2 ?

χ^2 - conformité

Variable mesurée

O_1	O_2	O_3	O_4

O_i est la ième valeur observée

Les valeurs attendues sont connues à priori

χ^2 - homogeneity

Variable mesurée

Variable de stratification de groupe		
	O_1	O_2
	O_3	O_4

Les valeurs attendues sont définies en fonction des H0, souvent totaux des colonnes et des lignes (distribution proportionnelle)

χ^2 - independence

Variable 1 mesurée

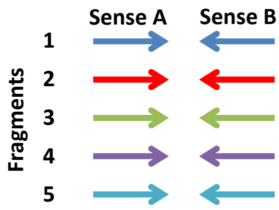
Variable 2 mesurée		
	O_1	O_2
	O_3	O_4

$E_{ij} = n_i \cdot 0.5$

e.g.

$E_{1A} = E_{1B} = n_1 \cdot 0.5$

Data structure



	A	B	Total
1	O_{1A}	O_{1B}	n_1
2	O_{2A}	O_{2B}	n_2
3	O_{3A}	O_{3B}	n_3
4	O_{4A}	O_{4B}	n_4
5	O_{5A}	O_{5B}	n_5
Total	n_A	n_B	n

Objectifs:

1) La ligation génère-t-elle une répartition uniforme de chaque fragment?

χ^2 - independence

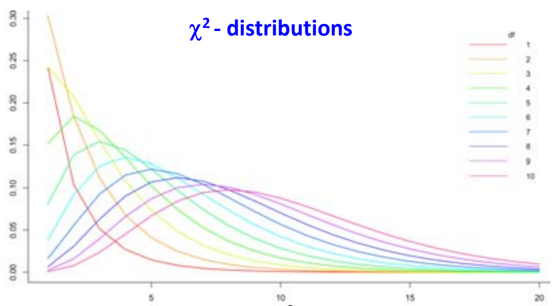
2) Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Tester l'hypothèse - distributions χ^2

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionnelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

5. Sélectionner le test statistique approprié
   - Chi-square conformité (objectif 1)
   - Chi-square test for independence (objectif 2)

χ^2 est calculé de la même manière pour les deux tests.



χ^2

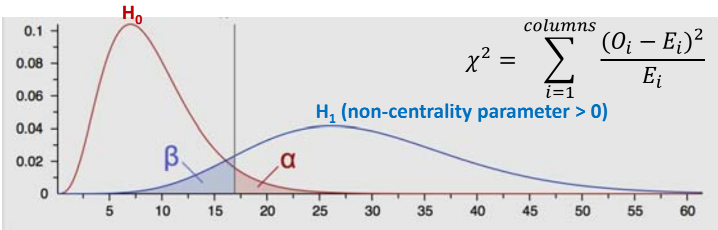
Notez que df est basé sur le nombre de catégories associées à la variable mesurée. PAS au nombre total d'observations (Contrairement à un test t).

$$\chi^2 = \sum_{i=1}^{columns} \frac{(O_i - E_i)^2}{E_i}, df = columns - 1$$

Tester l'hypothèse - Analyse de puissance

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

5. Sélectionner le test statistique approprié
   - Chi-square test de conformité.
6. Décider si le test de l'hypothèse est uni ou bilatéral.
   - par definition unilatéral droit (χ^2 augmente pour chaque erreur)



$$\chi^2 = \sum_{i=1}^{columns} \frac{(O_i - E_i)^2}{E_i}$$

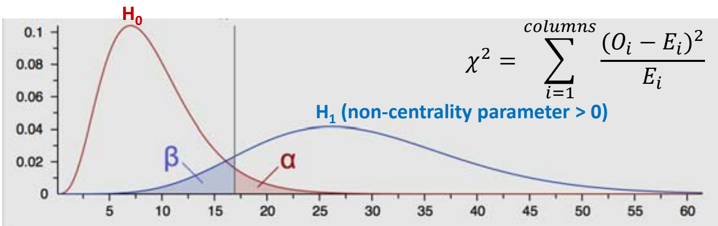
χ^2

Tester l'hypothèse - Analyse de puissance

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

5. Sélectionner le test statistique approprié
   - Chi-square test de conformité.
6. Décider si le test de l'hypothèse est uni ou bilatéral.
   - par definition unilatéral droit (χ^2 augmente pour chaque erreur)

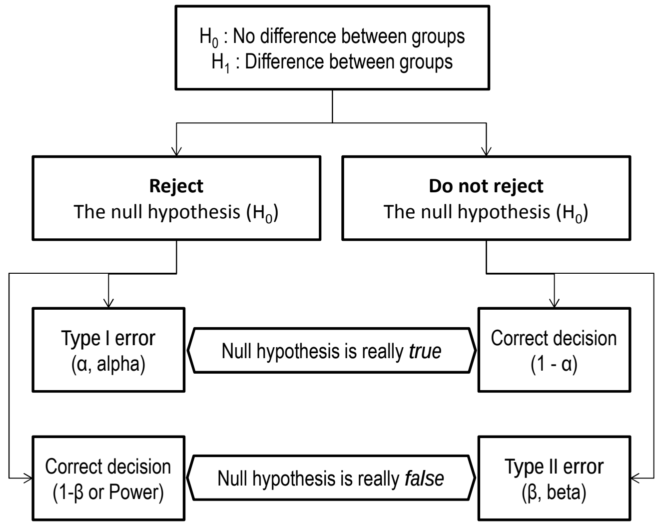
7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**



$$\chi^2 = \sum_{i=1}^{columns} \frac{(O_i - E_i)^2}{E_i}$$

χ^2

Erreur de Type I, de Type II et Puissance

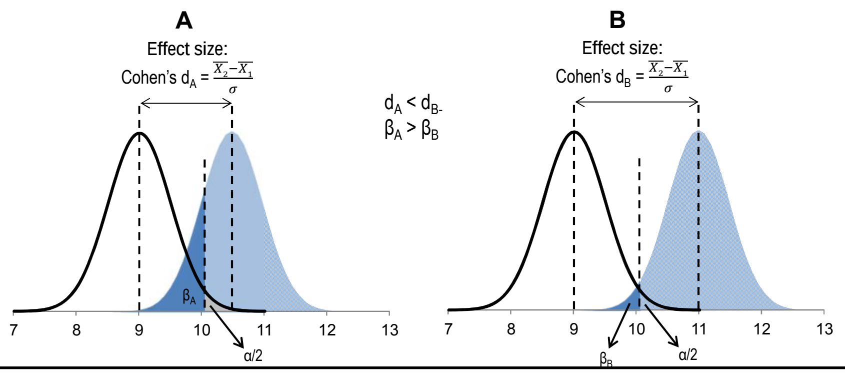


Tester l'hypothèse - Analyse de puissance

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir (nécessite des connaissances antérieures) taille d'effet
    - Interpretation de Cohen sur la taille d'effet :

    Low: 0.2 Medium: 0.5 High: 0.8



Tester l'hypothèse - Analyse de puissance

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

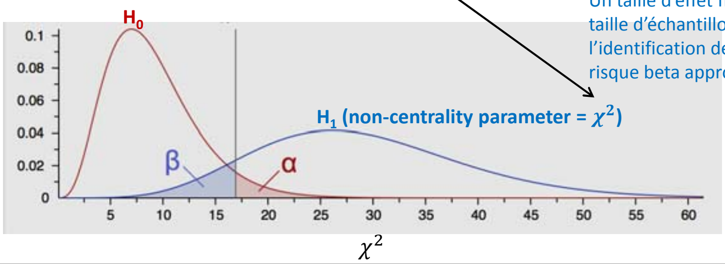
7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir la taille d'effet
   - Interpretation de la taille d'effet (dépend de df):

   Low: 0.1 Medium: 0.3 **High: 0.5**

- Calculer la taille d'effet estimé par phi(ϕ):

$$\phi = \sqrt{\frac{\chi^2}{N}} \leftrightarrow \chi^2 = \phi^2 \times N \text{ (identifie risque beta)}$$

Un taille d'effet fixé et une taille d'échantillon permettent l'identification de χ^2 avec un risque beta approprié .



Tester l'hypothèse - Analyse de puissance

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

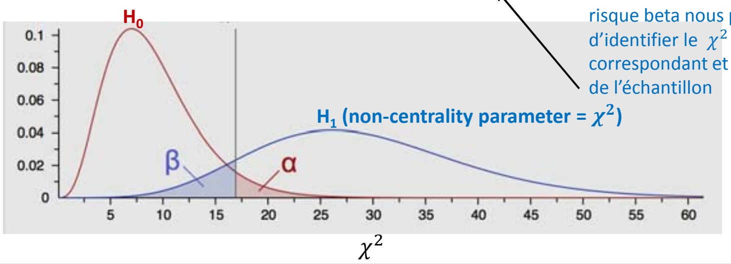
7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir la taille d'effet
   - Interpretation de la taille d'effet (dépend de df):

   Low: 0.1 Medium: 0.3 **High: 0.5**

- Calculer la taille d'effet estimé par phi(ϕ):

$$\phi = \sqrt{\frac{\chi^2}{N}} \leftrightarrow N = \frac{\chi^2}{\phi^2} \text{(identifie taille echantillon)}$$

Une taille d'effet fixé et un risque beta nous permettent d'identifier le χ^2 correspondant et ainsi la taille de l'échantillon



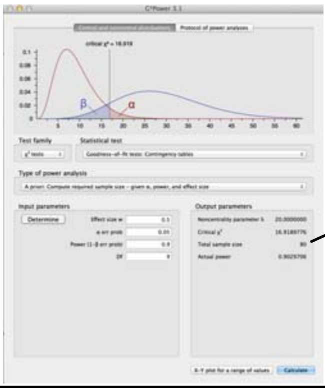
Tester l'hypothèse - Analyse de puissance

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir la taille d'effet : ϕ **= 0.5**

10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon

- Utiliser G*Power et/ou Real Statistics (logiciel gratuit)



Taille de l'échantillon

ϕ	N
0.5	80

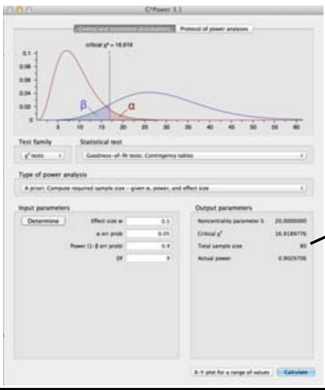
Tester l'hypothèse - Analyse de puissance

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir la taille d'effet : ϕ **= 0.5**

10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon

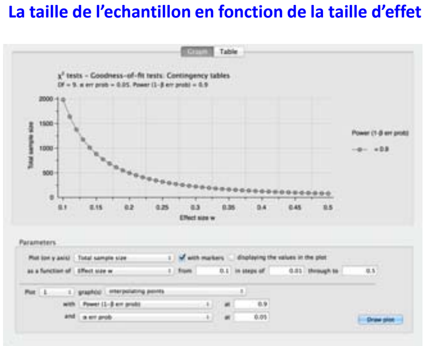
- Utiliser G*Power et/ou Real Statistics (logiciel gratuit)



Taille de l'échantillon

ϕ	N
0.5	80

La taille de l'echantillon en fonction de la taille d'effet



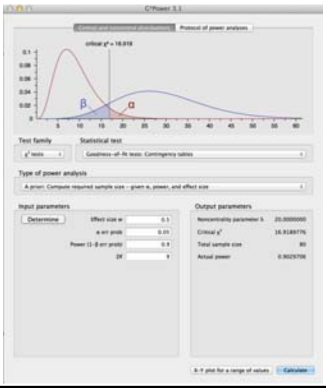
Tester l'hypothèse - Analyse de puissance

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir la taille d'effet : **ϕ = 0.5**

10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon

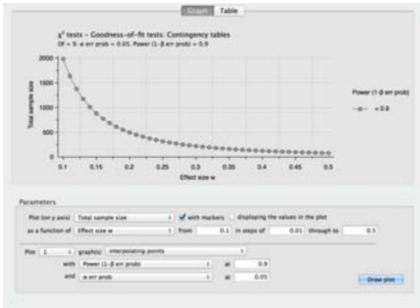
- Utiliser G*Power et/ou Real Statistics (logiciel gratuit)



Taille de l'échantillon

ϕ	N
0.1	1983
0.3	221
0.5	80
0.7	41

La taille de l'echantillon en fonction de la taille d'effet



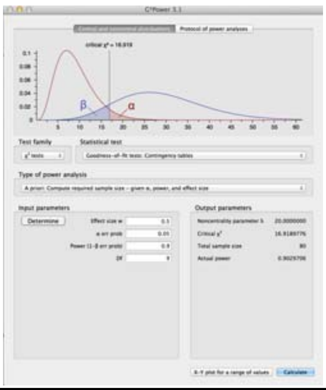
Tester l'hypothèse - Analyse de puissance

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir la taille d'effet : **ϕ = 0.5**

10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon

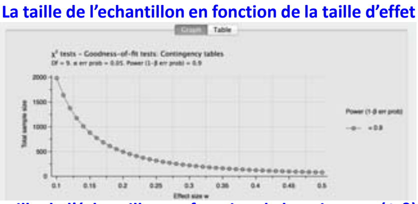
- Utiliser G*Power et/ou Real Statistics (logiciel gratuit)



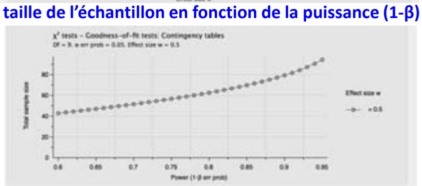
Taille de l'échantillon

ϕ	N
0.1	1983
0.3	221
0.5	80
0.7	41

La taille de l'echantillon en fonction de la taille d'effet



La taille de l'échantillon en fonction de la puissance (1-β)



Tester l'hypothèse - Cohorte et collection de données

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir la taille d'effet : ***ϕ* = 0.5**
10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon **n=80**
11. Rassembler les échantillons
12. Collecter et pré-analyser les données

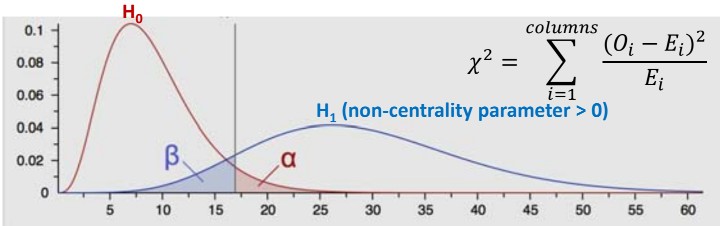
Tester l'hypothèse - Cohorte et collection de données

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir la taille d'effet : ***ϕ* = 0.5**
10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon **n=80**
11. Rassembler les échantillons
12. Collecter et pré-analyser les données
13. Calculer le test statistique
    - Obj. 1: La ligation génére-t-elle une distribution égale de chaque fragment?

Data de 2016

	A	B
1	3	0
2	12	6
3	6	9
4	15	0
5	0	2



$$\chi^2 = \sum_{i=1}^{columns} \frac{(O_i - E_i)^2}{E_i}$$

χ^2

Tester l'hypothèse- χ^2 tester l'objectif1

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir la taille d'effet : ϕ **= 0.5**
10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon **n=80**
11. Rassembler les échantillons
12. Collecter et pré-analyser les données
13. Calculer le test statistique
    - Obj. 1: La ligation génére-t-elle une distribution égale de chaque fragment?

Data de 2016

	A	B
1	3	0
2	12	6
3	6	9
4	15	0
5	0	2

Somme des lignes →

Observé

	O
1	3
2	18
3	15
4	15
5	2

N=53

Attendu

	E
1	53/5
2	53/5
3	53/5
4	53/5
5	53/5

$$\chi^2 = \sum_{i=1}^{columns} \frac{(O_i - E_i)^2}{E_i} =$$

$$\frac{(3-10.6)^2}{10.6} + \frac{(18-10.6)^2}{10.6} + \frac{(15-10.6)^2}{10.6} + \frac{(15-10.6)^2}{10.6} + \frac{(2-10.6)^2}{10.6} = 21.25$$

Degré de liberté (df) = 5-1 = 4

Tester l'hypothèse- χ^2 tester l'objectif1

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir la taille d'effet : ϕ **= 0.5**
10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon **n=80**
11. Rassembler les échantillons
12. Collecter et pré-analyser les données
13. Calculer le test statistique
    - Obj. 1: La ligation génére-t-elle une distribution égale de chaque fragment?

	alpha	
df	0.05	0.01
1	3.84	6.64
2	5.99	9.21
3	7.82	11.34
4	9.49	13.28
5	11.07	15.09
6	12.59	16.81
7	14.07	18.48
8	15.51	20.09
9	16.92	21.67
10	18.31	23.21

Observé

	O
1	3
2	18
3	15
4	15
5	2

N=53

Attendu

	E
1	53/5
2	53/5
3	53/5
4	53/5
5	53/5

$\chi^2 = 21.25$

Degrés de liberté (df) = 5-1 = 4

$\chi^2_{Critical} = 9.49 < 21.25$

Conclusion statistique:
Rejeter l'hypothèse H_0. Les échantillons ne dérivent pas d'une distribution uniforme

Conclusion biologique:
Les fragments ADN λ Phage ne sont pas clonés dans pUC19 en proportions égales.

Tester l'hypothèse - objectif 1 *post hoc*

- **Question complémentaire**: les fragments ne sont pas clonés à des proportions égales, mais quels fragments sont significativement différents de la répartition uniforme?

Hypothesis testing - objective 1 *post hoc*

- **Add-on question:** The fragments are not cloned at equal proportions, but which fragments are significantly different from the uniform distribution?
- This question can be answered with *post-hoc* (Latin, meaning "after this") methods.
- Many *post-hoc* methods are available. The most simple is the repetitive analysis of all possible combinations corrected for Type-I error using **Bonferroni** correction (dividing alpha with the number of combinations tested).
- For objective 1 we could compare each of the fragments with all the other fragments combined (using multiple Chi-square tests).
- An alternative is to consider the observations as a normal distributed random variable and transform them to adjusted residuals following the standardized normal distribution (N(0,1)).
- The latter approach will be scrutinized here.

χ^2-squared - analyse *post hoc*

- Supposons que X soit une variable aléatoire normale standard (moyenne = 0 et variance = 1). X ~ N (0,1)
- Un échantillon tiré au hasard de X est normalement distribué.
- Les résidus de l'échantillon seront également répartis normalement cette fois avec une moyenne de 0. A noter, la variance sera différente pour chaque O_{ij}.
  - La variance est inversement correlée avec p_i et p_j. Plus une observation est fréquente pour une variable mesurée, moins il y aura de variance. Nous pouvons donc corriger en divisant le résidu avec $\sqrt{(1-p_i)}$. Pour les observations fréquentes, le résidu ajusté augmentera plus que pour les observations peu fréquentes.
  - Les résidus ajustés peuvent être comparés au score Z critique
    - Dans excel pour l'analyse bilatérale = NORM.INV(α /2, 0, 1) = NORM.INV(0.025, 0, 1) = 1.96
    - Faire une correction pour comparaisons multiples. Bonferroni: $\alpha_{corr} = \alpha / n_{comparisons}$

$E_i = p_i \cdot n$ (espérance) and $p_i = \frac{1}{5}$

Residu: $r_i = \frac{O_i - E_i}{\sqrt{E_i}}$

Résidu ajusté $= \frac{O_i - E_i}{\sqrt{E_i(1-p_i)}}$

Tester l'hypothèse - objectif 1 *post hoc*

- **Question complémentaire**: les fragments ne sont pas clonés à des proportions égales, mais quels fragments sont significativement différents de la répartition uniforme?

$E_i = p_i \cdot n$ (espérance) and $p_i = \frac{1}{5}$

Residu: $r_i = \frac{O_i - E_i}{\sqrt{E_i}}$

Residu ajusté $= \frac{O_i - E_i}{\sqrt{E_i(1-p_i)}}$

Conclusion:

Observé

	O_i
1	3
2	18
3	15
4	15
5	2

n=53

Résidus ajustés

	r_{adj}	Size (bp)
1	-2.61	16841
2	2.54	5626
3	1.51	6527
4	1.51	7234
5	-2.95	1275

α = 0.05
Tests = 5
$\alpha_{bonferroni} = 0.05/5 = 0.01$
N(0.01/2,0,1)=2.58

Tester l'hypothèse - objectif 1 *post hoc*

- **Question complémentaire**: les fragments ne sont pas clonés à des proportions égales, mais quels fragments sont significativement différents de la répartition uniforme?

$E_i = p_i \cdot n$ (espérance) and $p_i = \frac{1}{5}$

Residu: $r_i = \frac{O_i - E_i}{\sqrt{E_i}}$

Residu ajusté $= \frac{O_i - E_i}{\sqrt{E_i(1-p_i)}}$

Observé

	O_i
1	3
2	18
3	15
4	15
5	2

n=53

Résidus ajustés

	r_{adj}	Size (bp)
1	-2.61	16841
2	2.54	5626
3	1.51	6527
4	1.51	7234
5	-2.95	1275

$\alpha = 0.05$
Tests = 5
$\alpha_{bonferroni} = 0.05/5=0.01$
N(0.01/2,0,1)=2.58

Conclusion:
Le fragment 2 est près d'être significativement surreprésenté et le fragment 1 et 5 sont significativement sous-représenté.

Pourquoi? Identifie cause potentielle.
1) Large fragment: ?
2) Petit fragment: ?

Tester l'hypothèse - objectif 1 *post hoc*

- **Question complémentaire**: les fragments ne sont pas clonés à des proportions égales, mais quels fragments sont significativement différents de la répartition uniforme?

$E_i = p_i \cdot n$ (espérance) and $p_i = \frac{1}{5}$

Residu: $r_i = \frac{O_i - E_i}{\sqrt{E_i}}$

Residu ajusté $= \frac{O_i - E_i}{\sqrt{E_i(1-p_i)}}$

Observé

	O_i
1	3
2	18
3	15
4	15
5	2

n=53

Résidus ajustés

	r_{adj}	Size (bp)
1	-2.61	16841
2	2.54	5626
3	1.51	6527
4	1.51	7234
5	-2.95	1275

$\alpha = 0.05$
Tests = 5
$\alpha_{bonferroni} = 0.05/5=0.01$
N(0.01/2,0,1)=2.58

Conclusion:
Le fragment 2 est près d'être significativement surreprésenté et le fragment 1 et 5 sont significativement sous-représenté.

Pourquoi? Identifie cause potentielle.
1) Large fragment: Efficacité de transfection
2) Petit fragment: Purification ADN

Tester l'hypothèse - objectif 1

- **Appliquez à vos propres échantillons!**

$$\chi^2 = \sum_{i=1}^{columns} \frac{(O_i - E_i)^2}{E_i}$$

$E_i = p_i \cdot n$ (espérance) and $p_i = \frac{1}{5}$

Residu: $r_i = \frac{O_i - E_i}{\sqrt{E_i}}$

Résidu ajusté $= \frac{O_i - E_i}{\sqrt{E_i(1-p_i)}}$

Observé

	O
1	O_1
2	O_2
3	O_3
4	O_4
5	O_5

Attendu

	E
1	E_1
2	E_2
3	E_3
4	E_4
5	E_5

Tester l'hypothèse- Test χ^2 objectif 2

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir l'effet de la taille: ϕ **= 0.5**
10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon **n=80**
11. Rassembler les échantillons
12. Collecter et pré-analyser les données
13. Calculer le test statistique

Obj 2: Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Observations

	A	B	Total
1	3	0	3
2	12	6	18
3	6	9	15
4	15	0	15
5	0	2	2
Total	36	17	53

50:50?

Attendu

	A	B
1	E_{1A}	E_{1B}
2	E_{2A}	E_{2B}
3	E_{3A}	E_{3B}
4	E_{4A}	E_{4B}
5	E_{5A}	E_{5B}

50:50

$E_{ij} = n_i \cdot 0.5$

$$\chi^2 = \sum_{j=1}^{rows} \sum_{i=1}^{columns} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Tester l'hypothèse- Test χ^2 objectif 2

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir l'effet de la taille: **ϕ = 0.5**
10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon **n=80**
11. Rassembler les échantillons
12. Collecter et pré-analyser les données
13. Calculer le test statistique

Obj 2: Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Observations

	A	B	Total
1	3	0	3
2	12	6	18
3	6	9	15
4	15	0	15
5	0	2	2
Total	36	17	53

50:50?

Attendu

	A	B
1	1.5	1.5
2	9	9
3	7.5	7.5
4	7.5	7.5
5	1	1

50:50

40% E_{ij}<5

$$E_{ij} = n_i \cdot 0.5$$

$$\chi^2 = \sum_{j=1}^{rows} \sum_{i=1}^{columns} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 22.6$$

Tester l'hypothèse- Test χ^2 objectif 2

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir l'effet de la taille: **ϕ = 0.5**
10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon **n=80**
11. Rassembler les échantillons
12. Collecter et pré-analyser les données
13. Calculer le test statistique

Obj 2: Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Observations

	A	B	Total
1+5	3	2	5
2	12	6	18
3	6	9	15
4	15	0	15
Total	36	17	53

50:50?

Attendu

	A	B
1+5	2.5	2.5
2	9	9
3	7.5	7.5
4	7.5	7.5

50:50

25% E_{ij}<5

$$E_{ij} = n_i \cdot 0.5$$

$$\chi^2 = \sum_{j=1}^{rows} \sum_{i=1}^{columns} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 17.8$$

Tester l'hypothèse- Test χ^2 objectif 2

- H_0: les fragments sont distribués de façon uniforme (O1) ou proportionelle (O2)
- H_1: les fragments ne sont pas distribués de façon uniforme ou proportionelle

7. Déterminer le niveau de risque alpha (α) : **5% (0.05) (erreur de Type 1)**
8. Déterminer le niveau de risque beta (1-β) : **90% (0.90) (erreur de Type 2)**
9. Définir ou Etablir l'effet de la taille: **ϕ = 0.5**
10. Créer un plan d'échantillonnage, déterminer la taille de l'échantillon **n=80**
11. Rassembler les échantillons
12. Collecter et pré-analyser les données
13. Calculer le test statistique

Obj 2: Les sens et les anti-sens sont-ils représentés de façon égale pour chaque fragment?

Observations

	A	B	Total
1+5	3	2	5
2	12	6	18
3	6	9	15
4	15	0	15
Total	36	17	53

50:50?

Attendu

	A	B
1+5	2.5	2.5
2	9	9
3	7.5	7.5
4	7.5	7.5

50:50

25% E_{ij}<5

$E_{ij} = n_i \cdot 0.5$

$$\chi^2 = \sum_{j=1}^{rows} \sum_{i=1}^{columns} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 17.8$$

Degré de libérté (df) = (ligne-1)*(colonne-1) = (2-1)*(4-1) = 3

$\chi^2_{Critique} = 7.82 < 17.8$ => Rejeter H_0

Hypothesis testing - objective 2 *post hoc*

- **Question complémentaire :** Quel fragment a une distribution sens/anti-sens biaisée?

χ^2-squared - analyse *post hoc*

- Supposons que X soit une variable aléatoire normale standard (moyenne = 0 et variance = 1). X ~ N (0,1)

- Un échantillon tiré au hasard de X est normalement distribué.
- Les résidus de l'échantillon seront également répartis normalement cette fois avec une moyenne de 0. A noter, la variance sera différente pour chaque O_{ij}.
  - La variance est inversement correlée avec p_i et p_j. Plus une observation est fréquente pour une variable mesurée, moins il y aura de variance. Nous pouvons donc corriger en divisant le résidu avec $\sqrt{(1-p_i)(1-p_j)}$. Pour les observations fréquentes, le résidu ajusté augmentera plus que pour les observations peu fréquentes.
  - Les résidus ajustés peuvent être comparés au score Z critique
    - Dans excel pour l'analyse bilatérale = NORM.INV(α /2, 0, 1) = NORM.INV(0.025, 0, 1) = 1.96
    - Faire une correction pour omparaisons multiples. Bonferroni: $\alpha_{corr} = \alpha/n_{comparisons}$

$E_{ij} = n_i \cdot pj$ (espérance) and $p_i = \frac{n_i}{n}$, p_j=0.5

Résidu: $r_{ij} = \frac{O_{ij}-E_{ij}}{\sqrt{E_{ij}}}$

Résidu ajusté = $\frac{O_{ij}-E_{ij}}{\sqrt{E_{ij}(1-p_i)(1-p_j)}}$

		M_i		
		j=1	j=2	
M_j	i=1	O_{ij}	O_{ij}	$n_{i=1}$
	i=2	O_{ij}	O_{ij}	$n_{i=2}$
		$n_{j=1}$	$n_{j=2}$	n

M = Mesure (variable)
O = Observation
n = Ligne, colonne ou somme totale

Hypothesis testing - objective 2 *post hoc*

- **Question complémentaire :** Quel fragment a une distribution sens/anti-sens biaisée?

$E_{ij} = n_i \cdot pj$ (espérance) and $p_i = \frac{n_i}{n}$, p_j=0.5

Résidu: $r_{ij} = \frac{O_{ij}-E_{ij}}{\sqrt{E_{ij}}}$

Résidu ajusté = $\frac{O_{ij}-E_{ij}}{\sqrt{E_{ij}(1-p_i)(1-p_j)}}$

Observé

	A	B	Total
1	3	0	3
2	12	6	18
3	6	9	15
4	15	0	15
5	0	2	2
Total	36	17	53

Résidus ajustés

	A	B
1	1.78	-1.78
2	1.74	-1.74
3	-0.91	0.91
4	4.57	-4.57
5	-1.44	1.44

α = 0.05
Tests = 10
$\alpha_{bonferroni}$ = 0.05/10=0.005
N(0.005/2,0,1)=2.81

Conclusion:
Le fragment 4 est nettement plus représenté dans le sens sens (sens A) par rapport à l'anti-sens.

Pourquoi? Identifie cause potentielle.
1) Toxicité

Tester l'hypothèse - objectif 2

- **Appliquez à vos propres échantillons!**

$$\chi^2 = \sum_{j=1}^{rows} \sum_{i=1}^{columns} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

E_{ij} = $n_i \cdot pj$ (espérance) and $p_i = \frac{n_i}{n}$, p_j=0.5

Résidu: $r_{ij} = \frac{O_{ij}-E_{ij}}{\sqrt{E_{ij}}}$

Résidu ajusté = $\frac{O_{ij}-E_{ij}}{\sqrt{E_{ij}(1-p_i)(1-p_j)}}$

Observé

	A	B	Total
1	O_{1A}	O_{1B}	n_1
2	O_{2A}	O_{2B}	n_2
3	O_{3A}	O_{3B}	n_3
4	O_{4A}	O_{4B}	n_4
5	O_{5A}	O_{5B}	n_5
Total	n_A	n_B	n

Estimé

	A	B
1	E_{1A}	E_{1B}
2	E_{2A}	E_{2B}
3	E_{3A}	E_{3B}
4	E_{4A}	E_{4B}
5	E_{5A}	E_{5B}