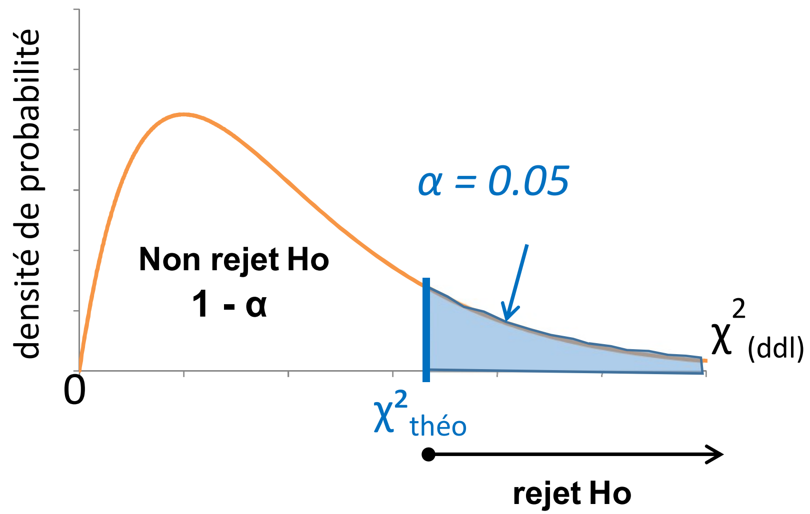
Tests relevant de la statistique du χ^2

Règle de décision :

- Déterminer la valeur critique de la statistique de test $\chi^2_{théo}$ délimitant les zones d'acceptation et rejet de H_0 (pour un risque α fixé ,avec ddl= nb modalités -1)

→ Table loi χ^2 (**unilatéral**)

Zone de rejet (bleu) : 5% des valeurs les plus extrêmes de la statistique de test qui n'ont presque aucune chance d'être observée si H0 est vraie. *(χ^2 = unilatéral)*



Tests relevant de la statistique du χ^2

Règle de décision :

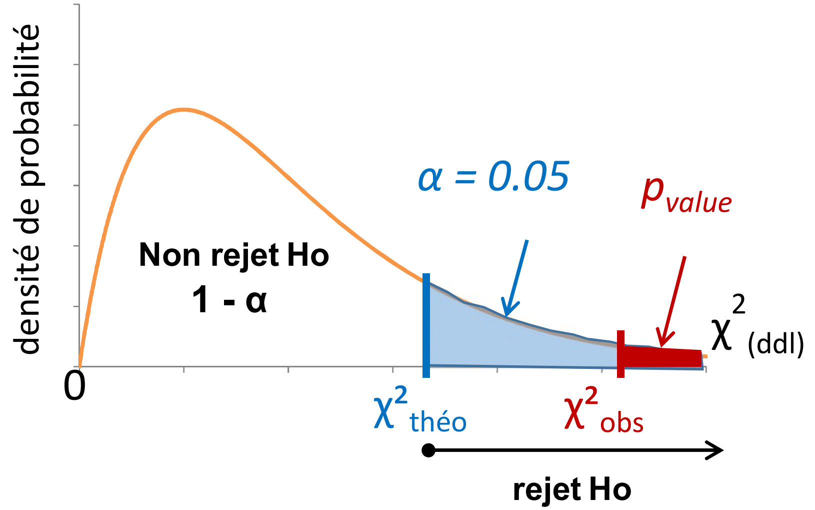
- Calcul du χ^2_{obs} à partir des observations sur l'échantillon

→ Comparaison à distribution sous H_0 : si $\chi^2_{obs} \geq \chi^2_{\text{théo}}$ → rejet H_0 (accepte H_1)

- On peut aussi utiliser la **p-value (aire rouge):**

« probabilité d'obtenir sous H_0 une valeur de la statistique de test au moins aussi extrême que celle obtenue à partir des observations »

→ Si p-value **<** α, on rejette H_0 et on accepte H_1



Tests relevant de la statistique du χ^2

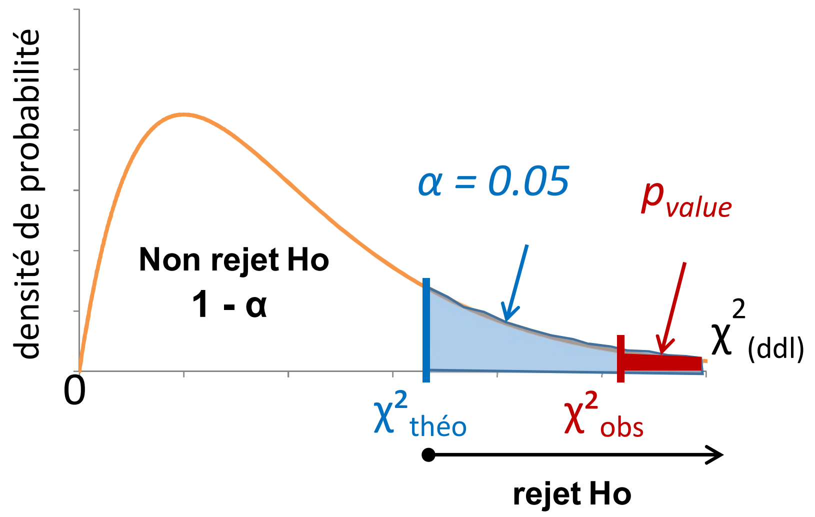
Règle de décision :

Quelques remarques :

- Si χ^2_{obs} très faible = bonne adéquation à la distribution théorique → non rejet de H0.

- Plus χ^2_{obs} (écart effectif observés et attendus) sera grand, moins H0 sera probable.

- Pour un même écart, χ^2_{obs} sera d'autant plus grand que l'effectif total est grand.



Tests relevant de la statistique du χ^2

Application :

On a effectué le croisement de balsamines blanches avec des balsamines pourpres. En première génération (F1) les fleurs sont toutes pourpres.

On obtient en seconde génération (F2) 4 catégories avec les effectifs suivants :

Couleur	Pourpre	Rose	Blanc lavande	blanc
Effectifs	1790	547	548	213

Peut on accepter l'hypothèse de répartition mendélienne (9/16, 3/16, 3/16, 1/16) pour $\alpha = 0.05$?

→ Revoir la réalisation du test « à la main » en vidéo sur Moodle ←

Tests relevant de la statistique du χ^2

Solution :

H_0 : distribution observée **conforme** à la théorique

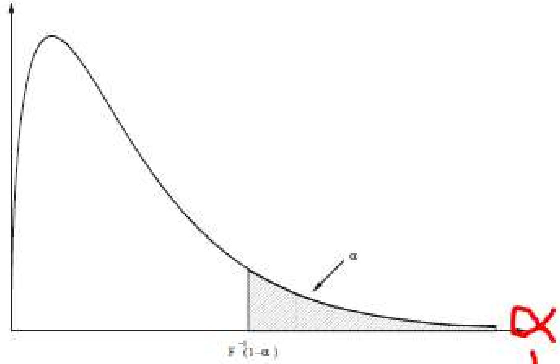
H_1 : distribution observée **non conforme** à la théorique

Trouver $\chi^2_{théo}$ dans la table de la loi du χ^2 :

pour α = 0.05

et degré de liberté (**ddl**) = nb modalités-1 = 4 – 1 = 3

➔ $\chi^2_{théo} = 7.81$



ddl →

n \ α	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.001
1	0.0002	0.001	0.0039	0.0158	2.71	3.84	5.02	6.63	10.83
2	0.02	0.05	0.1	0.21	4.61	5.99	7.38	9.21	13.82
3	0.11	0.22	0.35	0.58	6.25	$\boxed{7.81}$	9.35	11.34	16.27
4	0.3	0.48	0.71	1.06	7.78	9.49	11.14	13.28	18.47

Tests relevant de la statistique du χ^2

Solution :

H_0 : distribution observée **conforme** à la théorique

H_1 : distribution observée **non conforme** à la théorique

$\chi^2_{théo} = 7.81$

	Couleurs	Pourpre	Rose	B-Lavande	Blanc	Totaux
Effectifs observés	O_i	1790	547	548	213	**3098 = n**
Proportions attendues	p_i	9/16	3/16	3/16	1/16	**1**
Effectifs attendus						

Tests relevant de la statistique du χ^2

Solution :

H_0 : distribution observée **conforme** à la théorique

H_1 : distribution observée **non conforme** à la théorique

$\chi^2_{théo} = 7.81$

Couleurs	Pourpre	Rose	B-Lavande	Blanc	Totaux
Effectifs observés O_i	1790	547	548	213	**3098 = n**
Proportions attendues p_i	9/16	3/16	3/16	1/16	**1**
Effectifs attendus $c_i = n^* p_i$	1742.6	580.8	580.8	193.8	**3098 = n**

$$\chi^2 = \sum \frac{(O_i - C_i)^2}{C_i}$$

$$\chi^2_{obs} = \frac{(1790 - 1742.6)^2}{1742.6} + \frac{(547 - 580.8)^2}{580.8} + \frac{(548 - 580.8)^2}{580.8} + \frac{(213 - 193.8)^2}{193.8} = 7.06$$

$\chi^2_{obs} < \chi^2_{théo}$ **non rejet de H_0** → Hypothèse de répartition mendélienne acceptée

Effectuer ce test sous R

Code : 1) On crée les vecteurs de données :

```
eff_obs<-c(1790,547,548,213)
prop_theo<-c(9/16,3/16,3/16,1/16)

eff_theo<-sum(eff_obs)*prop_theo
```

$C_i = n*p_i$

```
ifelse(sum(eff_theo)==sum(eff_obs),
       "c'est bon!:)", "y'a un pb! :(")
```

Doit être == effectif total

2) On utilisera la fonction « **chisq.test** » :

```
chisq.test(eff_obs,p=prop_theo)
```

OU

```
chisq.test(eff_obs,p=eff_theo,rescale.p = T)
```

Quand on donne 1 vecteur d'effectifs (R doit recalculer les probabilités théoriques)

Effectuer ce test sous R

Sortie : A vous d'interpréter la sortie R!

```
	Chi-squared test for given probabilities

data:  c(1790, 547, 548, 213)
X-squared = 7.0628, df = 3, p-value = 0.06992
```

χ^2_{obs} ddl p-value > 0.05, non rejet de H0

Tests relevant de la statistique du χ^2

3) Test d'homogénéité /d'égalité

But : Comparer la distribution d'1 VA qualitative V1 observée dans J populations

➔ vérifier que la répartition des effectifs dans les différentes modalités sont équivalentes entre les J échantillons (suivent la même loi de probabilité).

Ex : Est-ce que la distribution des groupes sanguins est homogènes entre les localités A et B ?

2 variables qualitatives:

V1: VA d'intérêt (VD) à I modalités mesurée sur les individus statistiques

V2 : Variable (VI) à J modalités, **fixées** par l'expérimentateur, **déterminant** les **échantillons** (groupes) à comparer

Tests relevant de la statistique du χ^2

Les données :

Tableau de contingence (TDC) = croise les modalités de 2 variables qualitatives (**V1** en colonnes et **V2** en lignes) et contient des **effectifs observés O_{ij}** (N individus distribués dans le **I** x **J** cases du tableau).

Σ effectifs obs. par ligne

Structure TDC:

Modalités de V1

Modalités de V2

V2/V1	1	i	I	Marge colonne
1 (échantillon A)	O_{11}	O_{1i}	O_{1I}	$n_{1.}$
J (échantillon B)	O_{J1}	O_{Ji}	O_{JI}	$n_{J.}$
Marge ligne	$n_{.1}$	$n_{.i}$	$n_{.I}$	N

Σ effectifs obs. par colonne

Effectif total

Tests relevant de la statistique du χ^2

Les hypothèses :

H_0 : la distribution de V1 est identique dans toutes les populations d'où sont tirés les échantillons (V2) → homogénéité/égalité des effectifs entre échs.

Les échantillons proviennent de la même population et les différences d'effectifs observées sont dues aux fluctuations d'échantillonnage.

H_1 : la distribution de V1 n'est pas identique dans toutes les populations d'où sont tirés les échs (V2) → hétérogénéité/différence des effectifs entre échs.

Ils proviennent de populations différentes et les variations observées sont liées aux caractéristiques des groupes comparés (= modalités du facteur principal V2).

Tests relevant de la statistique du χ^2

La statistique de test :

$$\chi^2 = \sum \frac{(O_{ij}-C_{ij})^2}{C_{ij}}$$

Avec : O_{ij}= effectif observé et C_{ij}= effectif attendu

Comment calculer l'effectif théorique attendu C_{ij} si inconnu?

$$C_{ij} = \frac{Total\ au\ bout\ de\ la\ ligne\ \times total\ au\ bout\ de\ la\ colonne}{Effectif\ total}$$

Ex : $C_{11} = \frac{n1. \times n.1}{N}$

V2/V1	1	i	I	Marge colonne
1 (échantillon A)	O_{11}	O_{1i}	O_{1I}	$n_{1.}$
	C_{11}	C_{1i}	C_{1I}	
J (échantillon B)	O_{J1}	O_{Ji}	O_{JI}	$n_{J.}$
	C_{J1}	C_{Ji}	C_{JI}	
Marge ligne	$n_{.1}$	$n_{.i}$	$n_{.I}$	N

Tests relevant de la statistique du χ^2

Règle de décision :

- Déterminer la valeur de la statistique de test $X^2_{théo}$ au seuil α et avec **ddl** = (nb modalités V1 -1)*(nb modalités V2 -1) → Table loi X^2 (**unilatéral**)

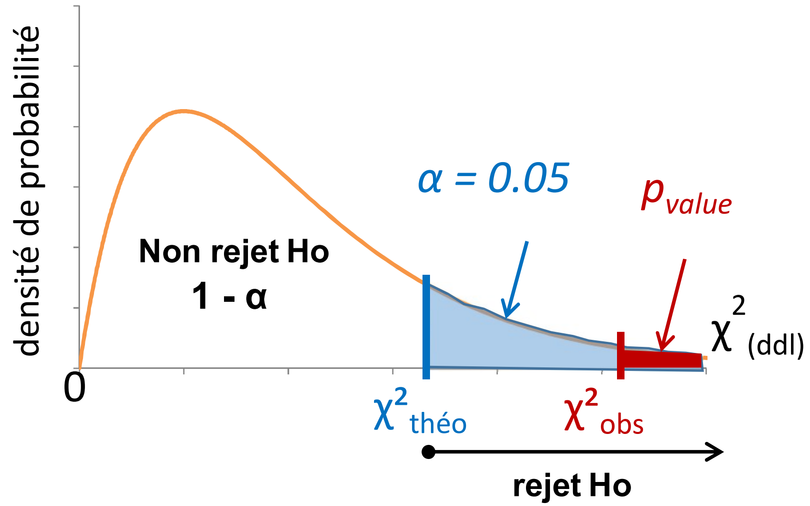
- Calcul du χ^2_{obs} à partir des observations sur les échantillons

→ Si $\chi^2_{obs} \geq \chi^2_{théo}$ ou p-value < α, on rejette H_0 et on accepte H_1

Les différences d'effectifs observées entre les échantillons sont :

- soit dues aux fluctuations d'échantillonnage (H0),

- soit liées aux modalités des groupes comparés = effet du facteur principal testé (H1).



Tests relevant de la statistique du χ^2

Application :

On veut comparer le salaire des femmes et des hommes dans une entreprise. On prend un échantillon de 120 hommes et un échantillon de 150 femmes puis on note le salaire : "faible", "moyen" ou "élevé".

On observe les effectifs suivants :

	Faible	Moyen	Elevé	Marge colonne
Homme	10	70	40	120
Femme	30	60	60	150
Marge ligne	40	130	100	270

Peut-on dire au seuil α=5% que les hommes et les femmes ont un niveau de salaire équivalent?

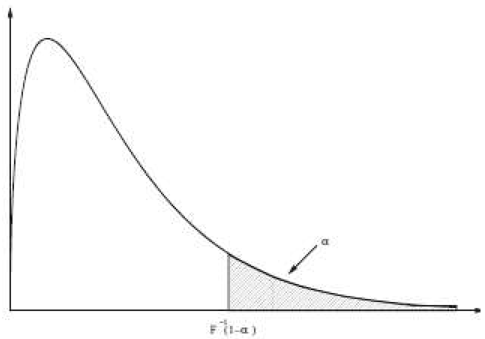
→ Revoir la réalisation du test « à la main » en vidéo sur Moodle ←

Tests relevant de la statistique du χ^2

Solution :

H_0 : les distributions observées sont équivalentes
H_1 : les distributions observées sont différentes

Trouver $\chi^2_{théo}$ dans la table de la loi du χ^2 :



pour **α** = 0.05

et **ddl** = (nb de modalités V1 -1)*(nb de modalités V2 -1)
= (3-1)*(2-1) = 2

➔ $\chi^2_{théo} = 5.99$

n \ α	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.001
1	0.0002	0.001	0.0039	0.0158	2.71	3.84	5.02	6.63	10.83
2	0.02	0.05	0.1	0.21	4.61	$\boxed{5.99}$	7.38	9.21	13.82
3	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	16.27

Tests relevant de la statistique du χ^2

Solution :

H_0 : les distributions observées sont équivalentes
H_1 : les distributions observées sont différentes

$\chi^2_{théo} = 5.99$

	Faible	Moyen	Elevé	Marge colonne
Homme	10 **17.8**	70 **57.8**	40 **44.4**	120
Femme	30 **22.2**	60 **72.2**	60 **55.6**	150
Marge ligne	40	130	100	270

$$C_{ij} = \frac{Total\ au\ bout\ de\ la\ ligne\ \times total\ au\ bout\ de\ la\ colonne}{Effectif\ total}$$

Ex : $C_{11} = \dfrac{120 \times 40}{270}$ = 17.8

Tests relevant de la statistique du χ^2

Solution :

H_0 : les distributions observées sont équivalentes
H_1 : les distributions observées sont différentes

$$\chi^2_{théo} = 5.99$$

	Faible	Moyen	Elevé	Marge colonne
Homme	10 **17.8**	70 **57.8**	40 **44.4**	120
Femme	30 **22.2**	60 **72.2**	60 **55.6**	150
Marge ligne	40	130	100	270

$$\chi^2 = \sum \frac{(O_{ij}-C_{ij})^2}{C_{ij}}$$

$$\chi^2_{obs} = \frac{(10-17.8)^2}{17.8} + \frac{(70-57.8)^2}{57.8} + \cdots + \frac{(60-72.2)^2}{72.2} + \frac{(60-55.6)^2}{55.6} = 11.58$$

χ^2_{obs} **>** $\chi^2_{théo}$ **rejet de H_0** → hommes et femmes sont traités différemment au niveau du salaire

Effectuer ce test sous R

Code :

```
### Rentrer les données
# vecteurs des 2 echs (effectifs obs)
Hom<-c(10,70,40)
Fem<-c(30,60,60)

# en faire une matrice
table<- matrix(c(Hom, Fem),2,3,byrow=T)

# ajouter le noms de lignes et colonnes
colnames(table)<-c("Faible","Moyen","Elevé")
row.names(table)<-c("Homme","Femme")

# pour ajouter les marges L et C
addmargins(table)
```

Matrice au format : 2 L * 3 C

Remplissage par lignes (T=TRUE)

Sortie :

```
      Faible Moyen Elevé Sum
Homme     10    70    40 120
Femme     30    60    60 150
Sum       40   130   100 270
```

Tableau de contingence

Effectuer ce test sous R

Code : On utilise « **chisq.test** », directement sur la matrice qu'on vient de créer :

```
khi_test2 <- chisq.test(table)
khi_test2
```

La matrice 2 L * 3 C

Sortie :

```
        Pearson's Chi-squared test

data:  table
X-squared = 11.579, df = 2, p-value = 0.00306
```

χ^2_{obs}

ddl

p-value très significative (< 0.01)

➔ rejet de H0

Effectuer ce test sous R

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P-value	Code	Niveau de significativité
1 à 0.1		Non Significatif (NS)
0.1 à 0.05	.	« presque » significatif
0.05 à 0.01	*	Significatif
0.01 à 0.001	**	Très significatif
0.001 à 0	***	Hautement significatif

Effectuer ce test sous R

Code :

```
## Récupérer les valeurs compilées dans le test :
```

```
#effectifs attendus (théoriques)
round(khi_test2$expected,2)
```

Arrondir à 2 chiffres après la virgule

NB: l'objet khi_test2 est une liste!

khi_test2	list [9] (S3: htest)	List of length 9
statistic	double [1]	11.57885
parameter	integer [1]	2
p.value	double [1]	0.003059747
method	character [1]	'Pearson\'s Chi-squared test'
data.name	character [1]	'table'
observed	double [2 x 3]	10 30 70 60 40 60
expected	double [2 x 3]	17.8 22.2 57.8 72.2 44.4 55.6
residuals	double [2 x 3]	-1.845 1.650 1.608 -1.438 -0.667
stdres	double [2 x 3]	-2.68 2.68 3.00 -3.00 -1.13 1.13

Sortie :

```
       Faible Moyen Elevé
Homme   17.78 57.78 44.44
Femme   22.22 72.22 55.56
```

Tests relevant de la statistique du χ^2

4) Test d'indépendance

1 seul échantillon de n individus statistiques (issus d'1 seule population), sur lesquels on mesure simultanément 2 caractéristiques (2 VA quali. V1 et V2).

But : Tester l'existence d'un lien statistique entre ces 2 variables qui serait susceptibles de créer une différence de répartition d'effectif par rapport à ceux attendus sous l'hypothèse d'indépendance.

Qu'est ce que veut dire l'indépendance ? 2 variables sont dites indépendantes lorsque les variations de l'une n'a aucune influence sur les variations de l'autre.

Ex : est-ce que connaître le sexe de quelqu'un permet de supposer la couleur de ses cheveux? Réponse : non.

est-ce que connaître l'origine ethnique de quelqu'un permet de supposer celle de ses yeux? Réponse : peut-être bien...