Analyse statistique Bivariée

2. Deux variables qualitatives

- Pour tester l’existence d’un lien entre les modalités de deux variable qualitatives:
  - On utilise de test de **CHI-2.**
  - Une représentation graphique.

Tableau de contingence

- Pour étudier la relation entre deux variables qualitatives, on construit d'abord un tableau appelé table de contingence ou tableau croisé.

		Variable Y						
		y_1	y_2	$\dots$	y_j	$\dots$	y_m	Total
Variable X	x_1	n_{11}	n_{12}	$\dots$	n_{1j}	$\dots$	n_{1m}	$n_{1.}$
	x_2	n_{21}	n_{22}	$\dots$	n_{2j}	$\dots$	n_{2m}	$n_{2.}$
	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
	x_i	n_{i1}	n_{i2}	$\dots$	n_{ij}	$\dots$	n_{im}	$n_{i.}$
	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
	x_l	n_{l1}	n_{l2}	$\dots$	n_{lj}	$\dots$	n_{lm}	$n_{l.}$
	Total	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.j}$	$\dots$	$n_{.m}$	n

Tableau de contingence

n_{ij} : effectif joint de la modalité x_i et de la modalité y_i
n_i : effectif marginal de la modalité x_i
n_j : effectif marginal de la modalité y_i
n : taille de l'échantillon

Tableau d’effectifs corrigés

- L’effectif corrigé se calcul ainsi:

$$n_{ij}^{*} = \frac{n_i n_j}{n}$$

		Variable Y						
		y_1	y_2	$\dots$	y_j	$\dots$	y_m	Total
Variable X	x_1	n_{11}^*	n_{12}^*	$\dots$	n_{1j}^*	$\dots$	n_{1m}^*	$n_{1.}$
	x_2	n_{21}^*	n_{22}^*	$\dots$	n_{2j}^*	$\dots$	n_{2m}^*	$n_{2.}$
	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
	x_i	n_{i1}^*	n_{i2}^*	$\dots$	n_{ij}^*	$\dots$	n_{im}^*	$n_{i.}$
	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
	x_l	n_{l1}^*	n_{l2}^*	$\dots$	n_{lj}^*	$\dots$	n_{lm}^*	$n_{l.}$
	Total	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.j}$	$\dots$	$n_{.m}$	n

Tableau d’effectifs corrigés

Test de Khi-deux

Hypothèses du test:

H_0:Les variables X et Y sont indépendantes.

H_1:Il existe une liaison entre Les variables X et Y

Statistique du test:

Sous H_0: $\chi^2 = \sum_{i=1}^{l} \sum_{j=1}^{m} \frac{(n_{ij}-n_{ij}^*)2}{n_{ij}^*} \sim \chi^2_{(l-1)(m-1)}$

Règle de décision:

Si la valeur de la statistique de test χ^2 est **inférieure** à la valeur seuil $\chi^2_{(l-1)(m-1)}$ alors on accepte l'hypothèse nulle.

Si la valeur de la statistique de test χ^2 est **supérieure** à la valeur seuil $\chi^2_{(l-1)(m-1)}$ alors on rejette l'hypothèse nulle.

Exemple:

Exposition	Etat du système immunitaire		Total
	I^-	I^+	
Exposé	135	120	255
Non exposé	150	195	345
Total	285	315	600

Croisement entre l'exposition au produit chimique et l'état du système immunitaire

Exposition	Etat du système immunitaire		Total
	I^-	I^+	
Exposé	121.1	133.9	255
Non exposé	163.9	181.1	345
Total	285	315	600

Tableau de contingence th´eorique

Sous H_0: $\chi^2 = \sum_{i=1}^{l} \sum_{j=1}^{m} \frac{(n_{ij}-n_{ij}^*)2}{n_{ij}^*} \sim \chi^2_{(2-1)(2-1)}$

Pour un risque α = 0.05, la région critique conduisant au rejet de H_0 pour une loi du χ_1^2 est définit par : [3,841,+∞[

Application numérique:

$$\chi^2 = \frac{(135-121.1)^2}{121.1} + \frac{(120-133.9)^2}{133.9} + \frac{(150-63.9)^2}{163.9} + \frac{(195-181.1)^2}{181.1} = 5.36$$

Au risque $\alpha = 0{,}05$, la valeur 5,36 appartient à la zone critique. On rejette alors H_0

Illustration graphique

- Graphiquement, nous pouvons aussi déduire l'existence de relation entre deux variables qualitatives en utilisant:

le diagramme en barre regroupé

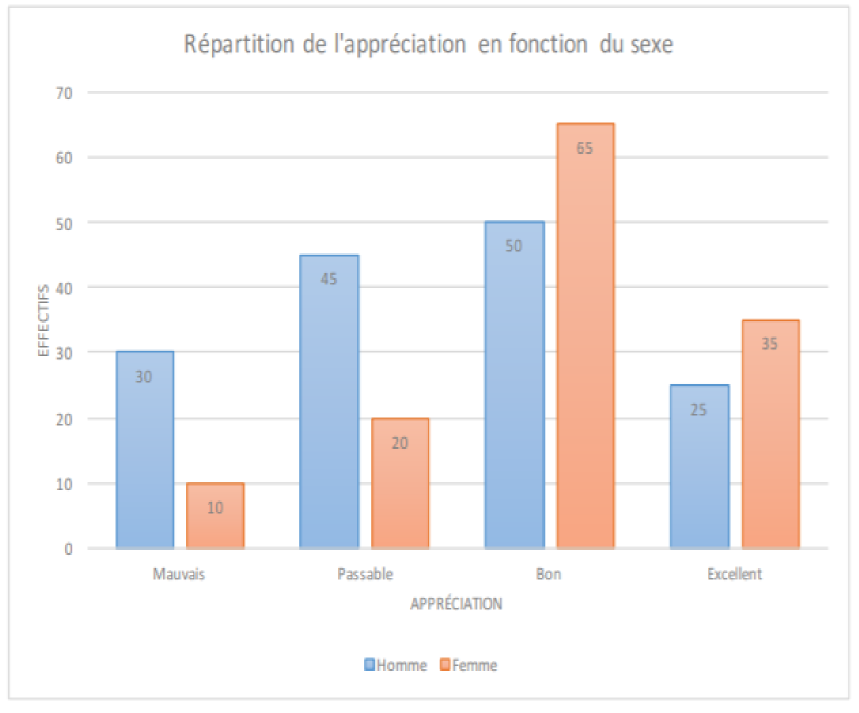

Diagramme en mosaïque