

TESTS STATISTIQUES DE FRÉQUENCES

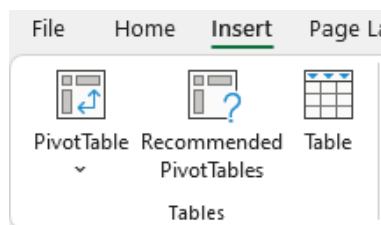
Daniel Leucuța, Mihaela Iancu

Solutions

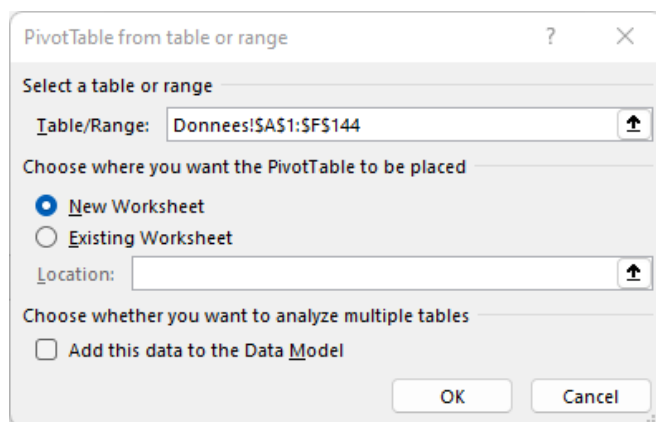
Réalisation du **Tableau de contingence observée**.

Sélectionnez le tableau de données, y compris les en-têtes de colonne, de A1 à F144.

Dans l'onglet Insertion, section Tableaux (Tables), cliquez sur Tableau croisé dynamique (**PivotTable**).



Dans la fenêtre ouverte, on vérifie si toutes les données ont été correctement sélectionnées dans le champ **Table/Range**, sinon resélectionnées. Nous quittons la sélection **NewWorksheet** (pour insérer le tableau de contingence sur une nouvelle feuille de calcul).



Dans la nouvelle feuille de données ouverte, la variable Traitement sera déplacée vers le champ Lignes (**Rows**), la variable Hyperglycémie vers le champ Colonnes

(**Columns**) et l'une ou l'autre des deux variables (Traitement ou Hyperglycémie) vers le champ Valeurs (**Values**).

Ainsi on obtiendra:

Count of Hyperglycémie	Column Labels			
Row Labels	Non	Oui	Grand Total	
Dexaméthasone	43	27	70	
Méthylprednisolone	20	53	73	
Grand Total	63	80	143	

Nous voulons changer l'ordre des lignes afin que sur la première ligne nous ayons la méthylprednisolone. Pour cela, nous faisons un clic droit sur Méthylprednisolone, et choisissons Move Methylprednisolone Up. Aussi, on veut changer l'ordre des colonnes afin que sur la première colonne nous ayons Oui. Pour cela, nous faisons un clic droit sur Oui, et choisissons Move Oui Up. On écrit en haut du tableau: **Tableau de contingence observée**

2	Tableau de contingence observée			
3	Count of Hyperglycémie	Column Labels		
4	Row Labels	Oui	Non	Grand Total
5	Méthylprednisolone	53	20	73
6	Dexaméthasone	27	43	70
7	Grand Total	80	63	143

1. Réalisation du **Tableau de contingence théorique**.

În dreapta tabelului de contingență realizați un tabel în formatul următor: copiați denumirile titlurilor liniilor și coloanelor, precum și valorile totalurilor din tabelul observat, lăsând liber conținutul tabelului:

	A	B	C	D	E	F	G	H	I
1									
2	Tableau de contingence observée					Tableau de contingence théorique			
3	Count of Hyperglycémie	Column Labels							
4	Row Labels	Oui	Non	Grand Total		Traitement/Hyperglycémie	Oui	Non	Total
5	Méthylprednisolone	53	20	73		Méthylprednisolone			73
6	Dexaméthasone	27	43	70		Dexaméthasone			70
7	Grand Total	80	63	143		Total	80	63	143

A l'intérieur de chacune des quatre cellules du contenu du tableau, saisissez la formule de calcul des fréquences théoriques : le total sur la ligne qui correspond à la cellule * le total sur la colonne qui correspond à la cellule / le total de tout le tableau :

	F	G	H	I
1				
2	Tableau de contingence théorique			
3				
4	Traitement/Hyperglycémie	Oui	Non	Total
5	Méthylprednisolone	=I5*G7/I7	=I5*H7/I7	73
6	Dexaméthasone	=I6*G7/I7	=I6*H7/I7	70
7	Total	80	63	143

Rezultatul calculelor este următorul:

	F	G	H	I
1				
2	Tableau de contingence théorique			
3				
4	Traitement/Hyperglycémie	Oui	Non	Total
5	Méthylprednisolone	40.8392	32.1608	73
6	Dexaméthasone	39.1608	30.8392	70
7	Total	80	63	143

On observe que les deux variables sont de type qualitatif, leurs catégories s'excluent mutuellement, on ne nous dit pas dans l'énoncé qu'il s'agit d'observations dépendantes - on en déduit donc qu'elles sont indépendantes et toutes les cellules du contenu du tableau de contingence théorique ont des valeurs supérieures à 5 (et aucune n'est à 0), le test χ^2 peut donc être appliqué à ces données.

2. Calcul de la statistique de test χ^2 et de la valeur p

La statistique de test χ^2 a la formule suivante : $\chi^2 = \sum_{i=1}^{l*c} \frac{(f_i^o - f_i^t)^2}{f_i^t}$

Où χ^2 est la statistique de test, l = nombre de lignes dans le tableau de contingence, c = nombre de colonnes dans le tableau de contingence, f_i^t = fréquence théorique de la cellule i, f_i^o = fréquence observée de la cellule i.

Sous le tableau de contingence théorique entrez le tableau suivant :

	F	G	H	I
1				
2	Tabel de contingență teoretic			
3				
4	Tratament/hiperglicemie	Da	Nu	Total
5	Metilprednisolon	40.83916	32.16084	73
6	Dexametazona	39.16084	30.83916	70
7	Total	80	63	143
8				
9				
10	Nivelul de semnificație statistică (alfa)		0,05	
11	Numărul de grade de libertate		1	
12	Valoarea critică a testului Hi pătrat		3,84	
13	Statistica testului			
14	Valoarea lui p			

Dans la cellule vide correspondant à la statistique du test, saisissez la formule suivante :

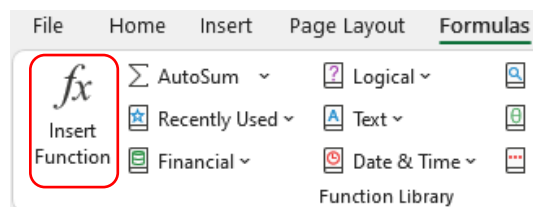
$$=(B5-G5)^2/G5+(C5-H5)^2/H5+(B6-G6)^2/G6+(C6-H6)^2/H6$$

Après le calcul, vous obtiendrez pour la statistique de test la valeur de 16.79.

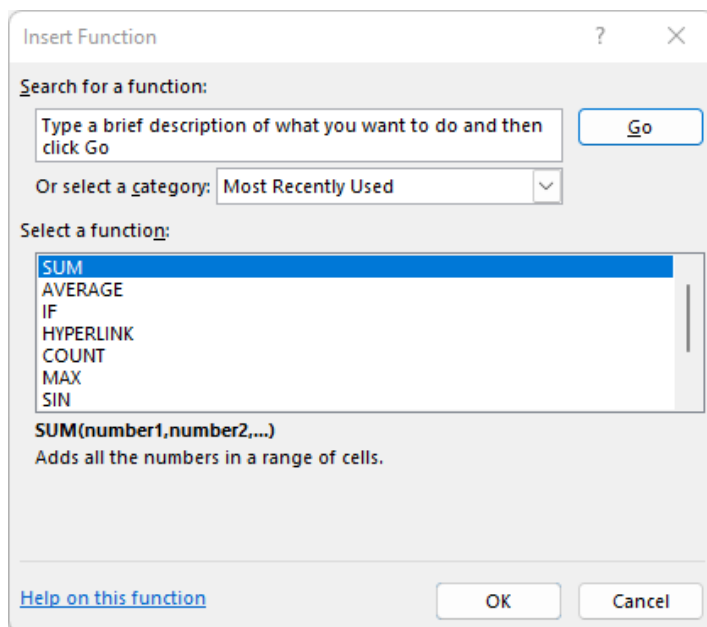
Pour calculer la valeur p du test χ^2 , saisissez une formule prédéfinie dans Excel en sélectionnant la cellule vide à droite de la valeur p et en appuyant sur le bouton Fx au-dessus de la colonne A :



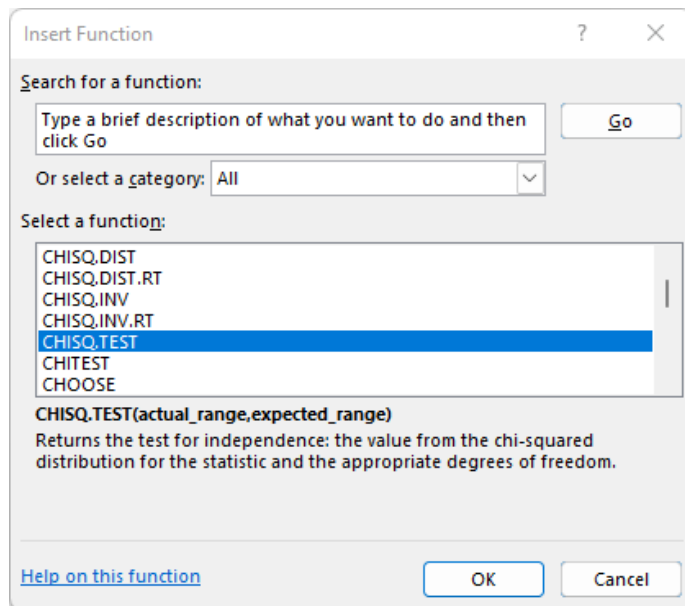
Une autre option consiste à appuyer sur le bouton **Fx** dans l'onglet Formules, section Bibliothèque de fonctions (**Function Library**).



La fenêtre suivante s'ouvrira :



Pour trouver la fonction nécessaire au calcul de la valeur de p, dans le deuxième champ (**Or select a category**), sélectionnez **All**, pour avoir accès à toutes les fonctions, puis identifiez dans la section Sélectionner une fonction (**Select a function**), la fonction **CHISQ.TEST** (ou **CHITEST** - l'ancienne version de la formule) nous le sélectionnons et appuyons sur le bouton **OK**.

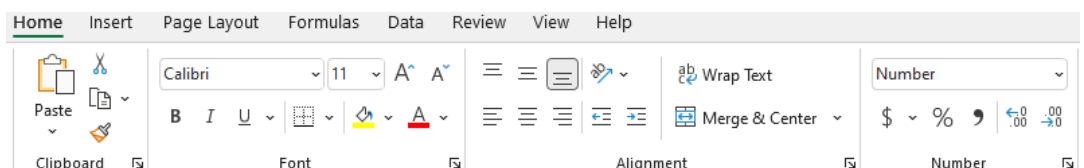


Dans la fenêtre ouverte, nous sélectionnons les quatre cellules à l'intérieur du tableau de contingence observé dans le champ **Actual_range**, et nous sélectionnons les quatre cellules à l'intérieur du tableau de contingence théorique dans le champ **Expected_range**, après quoi nous appuyons sur le bouton **OK**.

La valeur suivante sera obtenue :

La valeur de p	4.2E-05
----------------	---------

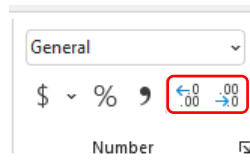
Cette valeur est le format scientifique pour représenter les valeurs numériques et signifie $4,2 * 10^{-5}$. Nous pouvons sélectionner cette cellule et changer son format en Nombre dans l'onglet Accueil (**Home**), section Nombre (**Number**):



Nous obtiendrons la valeur suivante :

La valeur de p	0.00
----------------	------

Pour voir plus de décimales, nous pouvons appuyer sur les flèches dans l'onglet Accueil (**Home**), section Nombre (**Number**):



Ainsi, nous pouvons obtenir la valeur au format suivant :

La valeur de p	0.000042
----------------	----------

- Interprétation du résultat du test statistique, basé sur la statistique χ^2 , puis basé sur la valeur de p.

Entrez ce qui suit sous le tableau des statistiques de test χ^2 précédemment calculées :

Interprétations :						
Hypothèse nulle:						
Hypothèse alternative:						
Niveau de signification statistique (alpha) : 0,05						
Valeur critique du test du Khi carré : 3,84						
Région de rejet (critique) :						
Interprétez le résultat du test en fonction de la statistique du chi carré et de la région de rejet :						
Justifier la réponse précédente :						
Interprétez le résultat du test en fonction de la statistique de la valeur p :						
Justifier la réponse précédente :						

Les bonnes réponses sont les suivantes :

Hypothèse nulle : Il n'y a pas d'association statistiquement significative entre le traitement et l'hyperglycémie chez les sujets atteints de COVID-19, ou il n'y a pas de différences statistiquement significatives entre les sujets atteints de COVID-19 qui ont reçu de la méthylprednisolone et ceux qui ont reçu de la dexaméthasone concernant la fréquence de l'hyperglycémie.

Hypothèse alternative : Il existe une association statistiquement significative entre le traitement et l'hyperglycémie chez les sujets atteints de COVID-19, ou il existe des différences statistiquement significatives entre les sujets atteints de

COVID-19 qui ont reçu de la méthylprednisolone et ceux qui ont reçu de la dexaméthasone concernant la fréquence de l'hyperglycémie.

Niveau de signification statistique (alpha) : 0,05

La valeur critique du test du Chi carré : 3,84 – pour le cas de deux variables qualitatives dichotomiques.

Région de rejet (critique) : [3,84 ; + infini), la région est construite à partir de la valeur critique [critical value ; + infini).

Interprétez le résultat du test en fonction de la statistique du chi carré et de la région de rejet : Il existe une association statistiquement significative entre le traitement et l'hyperglycémie chez les sujets atteints de COVID-19, ou Il existe des différences statistiquement significatives entre les sujets atteints de COVID-19 qui ont reçu de la méthylprednisolone et ceux qui ont reçu de la dexaméthasone concernant la fréquence de l'hyperglycémie.

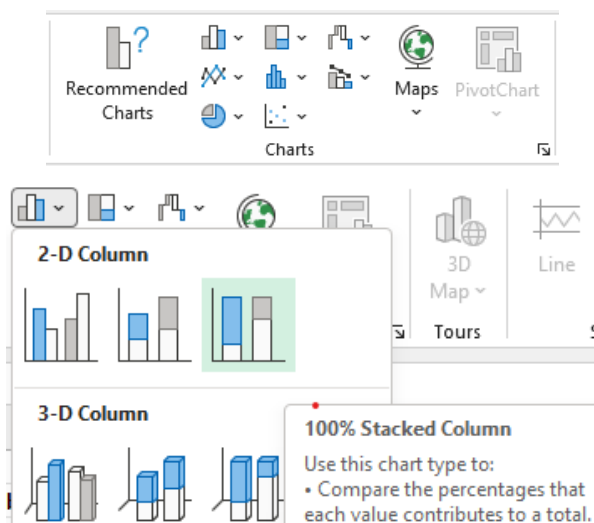
Justifiez votre réponse : étant donné que la statistique de test de 16,79 se situe dans la région de rejet (critique) - [3,84 ; + infini), on rejette l'hypothèse nulle et on accepte l'hypothèse alternative.

Interpréter le résultat du test en fonction de la statistique de la valeur p : il existe une association statistiquement significative entre le traitement et l'hyperglycémie chez les sujets atteints de COVID-19, ou il existe des différences statistiquement significatives entre les sujets atteints de COVID-19 qui ont reçu de la méthylprednisolone et ceux qui ont reçu de la dexaméthasone, en termes de fréquence de l'hyperglycémie.

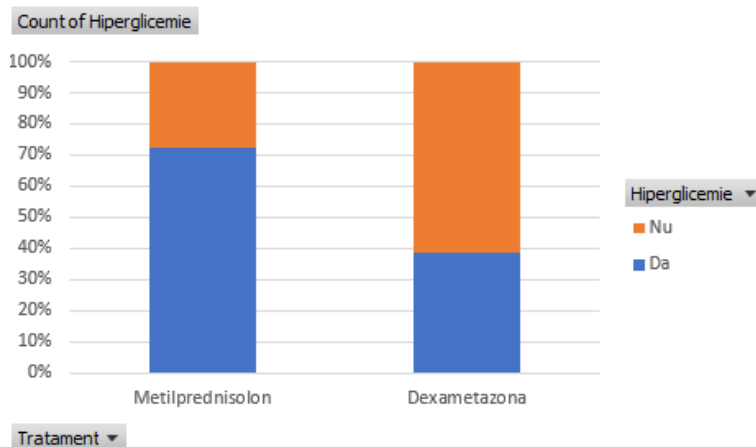
Justifiez votre réponse : étant donné que la valeur de p (0,000042) est inférieure à 0,05 (le niveau de signification statistique, alpha), nous rejetons l'hypothèse nulle et acceptons l'hypothèse alternative.

4. Pour choisir le graphique, nous pensons au nombre de variables impliquées : deux. Ensuite, nous pensons à leur type. Les deux sont qualitativement dichotomiques. Pour cette situation, d'un graphique qui identifie le lien entre deux variables qualitatives, des graphiques à colonnes ou à barres conviennent, de préférence en pourcentage.

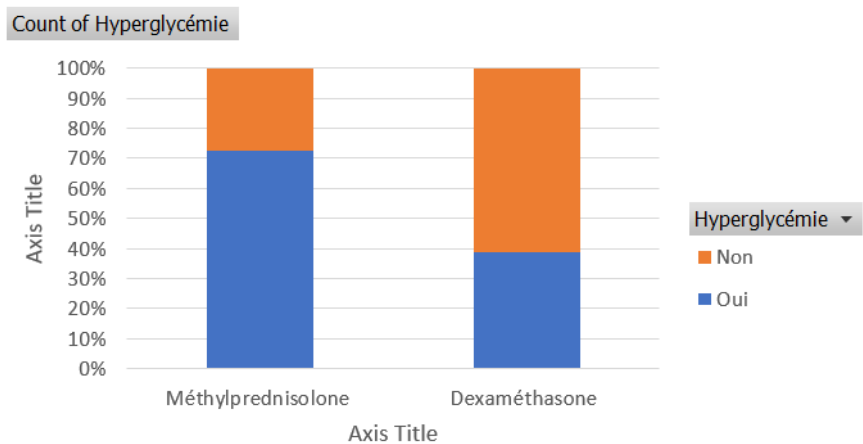
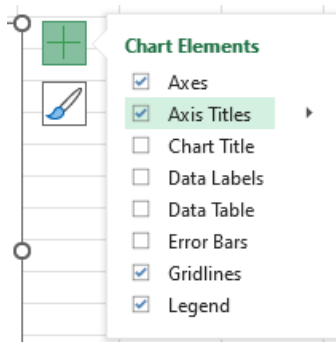
Nous sélectionnons une cellule dans le tableau de contingence observé et, dans l'onglet Insertion (**Insert**), section Graphiques (**Charts**), nous choisissons Colonne empilée à 100 % (**100% Stacked Column**).



On obtient le graphe suivant :

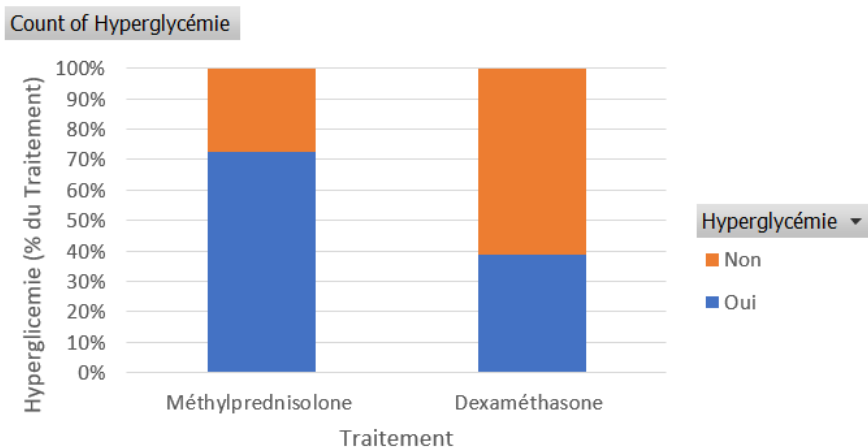


Întrucât nu avem titlul axelor trecute pe grafic, dăm un clic pe grafic într-o zonă fără elemente și apoi apăsăm pe butonul + din colțul dreapta sus al graficului și bifăm **Axis Titles**:



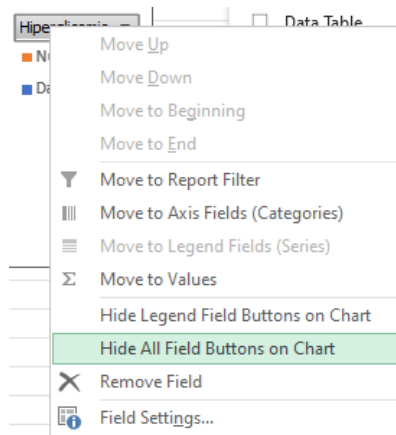
Traitement ▼

Nous nommons les axes selon l'image ci-dessous :

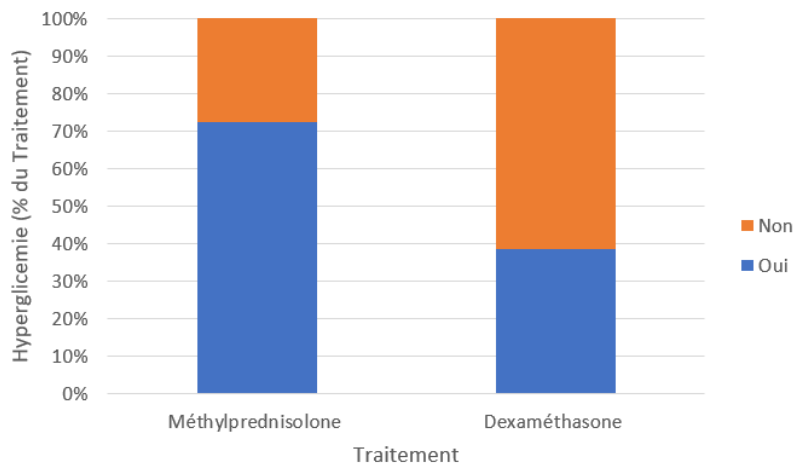


Traitement ▼

Enfin, faites un clic droit sur le bouton dans la légende et choisissez Masquer tous les boutons de champ du graphique (**Hide all field buttons on chart**):



L'image suivante est obtenue :



Sur l'image, on peut voir que dans le groupe qui a reçu de la méthylprednisolone, il y a un pourcentage plus élevé de sujets souffrant d'hyperglycémie (sur le graphique, en bleu), par rapport à ceux du groupe qui a reçu de la dexaméthasone. Les différences entre les deux groupes semblent relativement importantes.

- Pour trouver le pourcentage de sujets souffrant d'hyperglycémie par traitement, nous copierons le tableau des contingences observées ailleurs sur la feuille de calcul. Nous le sélectionnons initialement :

2	Tableau de contingence observée			
3	Count of Hyperglycémie	Column Labels		
4	Row Labels	Oui	Non	Grand Total
5	Méthylprednisolone	53	20	73
6	Dexaméthasone	27	43	70
7	Grand Total	80	63	143

Sélectionnez ensuite une cellule vide et faites un clic droit pour coller le contenu :

2	Tableau de contingence observée			
3	Count of Hyperglycémie	Column Labels		
4	Row Labels	Oui	Non	Grand Total
5	Méthylprednisolone	53	20	73
6	Dexaméthasone	27	43	70
7	Grand Total	80	63	143
8				
9				
10	Count of Hyperglycémie	Column Labels		
11	Row Labels	Oui	Non	Grand Total
12	Méthylprednisolone	53	20	73
13	Dexaméthasone	27	43	70
14	Grand Total	80	63	143

Nous sélectionnons n'importe quelle cellule du contenu du tableau observé, ni du total, ni des catégories de variables, nous faisons un clic droit et dans le menu, choisissez Show Value As - % of Row Total (pour présenter les valeurs en pourcentage sur les lignes):

The screenshot shows an Excel interface with a PivotTable. A right-click context menu is open over the 'Grand Total' row. The 'Show Value As' option is expanded, and '% of Row Total' is selected. The background shows the same contingency table as above, but with the 'Show Value As' menu open.

Enfin on obtient le tableau souhaité :

10	Count of Hyperglycémie	Column Labels		
11	Row Labels	Oui	Non	Grand Total
12	Méthylprednisolone	72.60%	27.40%	100.00%
13	Dexaméthasone	38.57%	61.43%	100.00%
14	Grand Total	55.94%	44.06%	100.00%

Dans ce tableau on observe que dans le groupe recevant la méthylprednisolone il y avait 72,6% de sujets qui avaient une hyperglycémie, et dans le groupe recevant la dexaméthasone 38,57% avaient une hyperglycémie. Ainsi, ceux qui ont reçu de la méthylprednisolone ont présenté une hyperglycémie 34 % plus fréquente que ceux qui ont reçu de la dexaméthasone.

le Risque relatif d'avoir **hyperglycémie** pour ceux qui ont reçu de la **méthylprednisolone** par rapport a ceux qui ont reçu de la **dexaméthasone** se calcule avec la formule $[a / (a + b)] / [c / (c + d)] = 1.88$, a partir du tableau de contingence observée, avec les fréquences absolues (pas avec les fréquences relatifs – pourcentages !)

Nommez la feuille de calcul : Traitement Hyperglycémie.

6. Le tableau théorique observé et la valeur de la statistique χ^2 , ainsi que la valeur p, pour tester l'association entre le traitement et les décès sont présentés ci-dessous :

2	Tabel de contingență observat					Tabel de contingență teoretic			
3	Count of Deces	Column Labels							
4	Row Labels	Da	Nu	Grand Total		Tratament/Deces	Da	Nu	Total
5	Metilprednisolon	9	64	73		Metilprednisolon	10.72027972	62.27972	73
6	Dexametazona	12	58	70		Dexametazona	10.27972028	59.72028	70
7	Grand Total	21	122	143		Total	21	122	143
8									
9									
10						Nivelul de semnificație statistică (alfa)		0,05	
11						Numărul de grade de libertate		1	
12						Valoarea critică a testului Hi pătrat		3,84	
13						Statistica testului		0.661007	
14						Valoarea lui p		0.416205	

L'interprétation de ces résultats est la suivante :

Hypothèse nulle : Il n'y a pas d'association statistiquement significative entre le traitement et les décès chez les sujets atteints de COVID-19, ou il n'y a pas de différences statistiquement significatives entre les sujets atteints de COVID-19 qui ont reçu de la méthylprednisolone et ceux qui ont reçu de la dexaméthasone, en termes de fréquence des décès.

Hypothèse alternative : Il existe une association statistiquement significative entre le traitement et les décès chez les sujets atteints de COVID-19, ou il existe des différences statistiquement significatives entre les sujets atteints de COVID-19 qui

ont reçu de la méthylprednisolone et ceux qui ont reçu de la dexaméthasone concernant la fréquence des décès.

Niveau de signification statistique (alpha) : 0,05

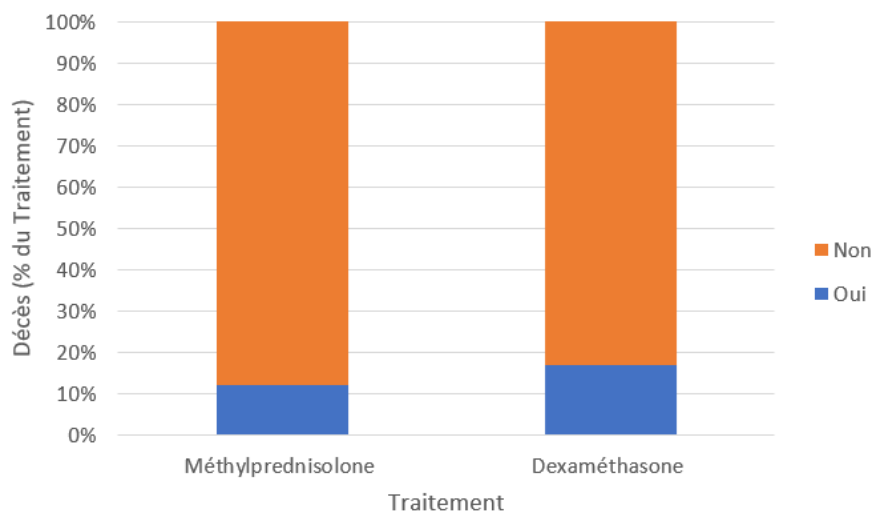
La valeur critique du test du Khi carré : 3,84 – pour le cas de deux variables qualitatives dichotomiques.

Région de rejet (critique) : [3,84 ; + infini), la région est construite à partir de la valeur critique [critical value ; + infini).

Interpréter le résultat du test en fonction de la statistique du chi carré et de la région de rejet : on ne peut pas dire qu'il existe une association statistiquement significative entre le traitement et les décès chez les sujets atteints de COVID-19, ou on ne peut pas dire qu'il existe des différences statistiquement significatives entre les sujets atteints COVID-19 qui ont reçu de la méthylprednisolone et ceux qui ont reçu de la dexaméthasone, en termes de fréquence de décès.

Justifiez votre réponse : Puisque la statistique de test, 0,661, n'est pas dans la région de rejet (critique) - [3,84 ; + infini), nous ne pouvons pas rejeter l'hypothèse nulle.

7. Un graphique comme celui-ci sera obtenu :



Sur l'image, on peut voir que dans le groupe qui a reçu de la méthylprednisolone, il y a un pourcentage inférieur de sujets décédés (sur le graphique en bleu), par rapport à ceux du groupe qui a reçu de la dexaméthasone. Cependant, les différences entre les deux groupes semblent relativement faibles.

8. Un tableau comme celui-ci sera obtenu :

10	Count of Décès	Column Labels <input type="button" value="v"/>		
11	Row Labels <input type="button" value="v"/>	Oui	Non	Grand Total
12	Méthylprednisolone	12.33%	87.67%	100.00%
13	Dexaméthasone	17.14%	82.86%	100.00%
14	Grand Total	14.69%	85.31%	100.00%

Ce tableau montre que dans le groupe méthylprednisolone, 12,33 % des sujets sont décédés et dans le groupe dexaméthasone, 17,14 % sont décédés. Ainsi, ceux qui ont reçu de la méthylprednisolone sont décédés 4,81 % moins souvent que ceux qui ont reçu de la dexaméthasone.

Nommez la feuille de calcul : Traitement Décès.

Enregistrez le fichier.

A retenir

Dans le cas où l'on s'intéresse à l'association entre deux variables qualitatives, avec des catégories mutuellement exclusives, des observations indépendantes et des fréquences attendues suffisamment élevées, l'un des tests que l'on peut utiliser est le test χ^2 .

Pour obtenir la p-value du test χ^2 dans Excel, un tableau de contingence observé est inséré à l'aide d'un tableau croisé dynamique, puis le tableau théorique est calculé (chaque cellule de contenu du tableau étant le produit du total ligne et du total colonne correspondant à la cellule, divisé par le total sur toute la table), et enfin la fonction CHISQ.TEST est utilisée, en sélectionnant le contenu des deux tables.