

Teste statistice pentru frecvențe

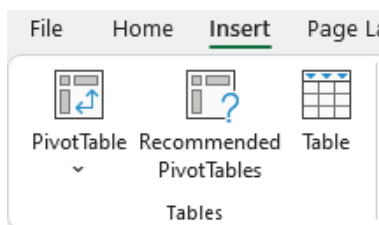
Daniel-Corneliu Leucuța

SOLUȚII

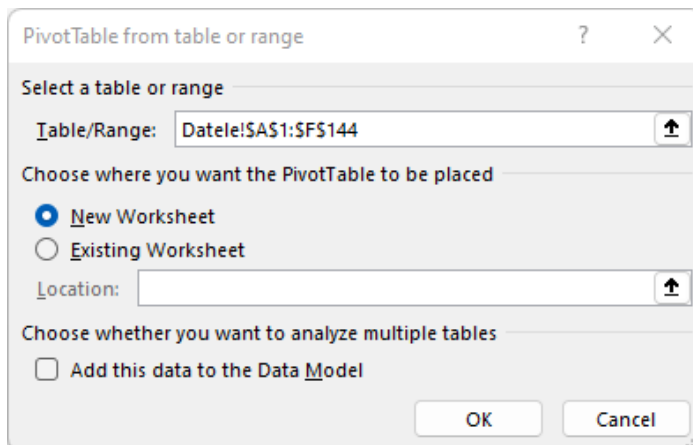
1. Realizarea tabelului de contingență observat.

Se selectează tabelul cu datele, incluzând titlurile coloanelor, de la A1 până la F144.

În tabul **Insert**, secțiunea **Tables**, se dă clic pe **PivotTable**.



În fereastra deschisă, se verifică dacă toate datele au fost corect selectate în câmpul **Table/Range**, dacă nu se reselectează. Lăsăm selecția **New Worksheet** (să insere tabelul de contingență pe o foaie de calcul nouă).



În noua foaie de date deschisă se va deplasa variabila Tratament în câmpul **Rows**, variabila Hiperglicemie în câmpul **Columns** și oricare din cele două variabile (Tratament sau Hiperglicemie) în câmpul **Values**.

PivotTable Fields

Choose fields to add to report:

Search

- ☐ Id
- ☐ Traitement
- ☐ Deces
- ☐ Therapie intensive
- ☐ Hipopotasemie
- ☐ Hiperglicemie

Drag fields between areas below:

Filters	Columns
Rows	Values

☐ Defer Layout Update Update

PivotTable1

To build a report, choose fields from the PivotTable Field List

Astfel vom obține:

PivotTable Fields

Choose fields to add to report:

Search

- ☐ Id
- ☒ Traitement
- ☐ Deces
- ☐ Therapie intensive
- ☐ Hipopotasemie
- ☒ Hiperglicemie

Drag fields between areas below:

Filters	Columns
	Hiperglicemie
Rows	Values
Traitement	Count of Hiperglicemie

☐ Defer Layout Update Update

Count of Hiperglicemie Column Labels

Row Labels Da Nu Grand Total

Dexametazona	27	43	70
Metilprednisolon	53	20	73
Grand Total	80	63	143

Dorim să modificăm ordinea liniilor, astfel încât pe prima linie să avem Metilprednisolon. Pentru aceasta, dăm clic dreapta pe Metilprednisolon, și alegem Move Metilprednisolon Up.

Tastăm deasupra denumirea tabelului: **Tabel de contingență observat**

2	Tabel de contingență observat		
3	Count of Hiperglicemie	Column Labels	
4	Row Labels	Da	Nu Grand Total
5	Metilprednisolon	53	20 73
6	Dexametazona	27	43 70
7	Grand Total	80	63 143

2. Realizarea tabelului de contingență teoretic.

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	Tabel de contingență observat					Tabel de contingență teoretic			
3	Count of Hiperglicemie	Column Labels							
4	Row Labels	Da	Nu	Grand Total		Tratament/hiperglicemie	Da	Nu	Total
5	Metilprednisolon	53	20	73		Metilprednisolon			73
6	Dexametazona	27	43	70		Dexametazona			70
7	Grand Total	80	63	143		Total	80	63	143

În interiorul fiecărei din cele patru celule din conținutul tabelului se introduce formula pentru calculul frecvențelor teoretice: totalul pe linie care corespunde celei * totalul pe coloană care corespunde celei / totalul întregului tabel:

	F	G	H	I
1				
2	Tabel de contingență teoretic			
3				
4	Tratament/hiperglicemie	Da	Nu	Total
5	Metilprednisolon	=I5*G7/I7	=I5*H7/I7	73
6	Dexametazona	=I6*G7/I7	=I6*H7/I7	70
7	Total	80	63	143

Rezultatul calculelor este următorul:

	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

Se observă că cele două variabile sunt de tip calitativ, categoriile lor sunt mutual exclusive, nu ni se spune în enunț că ar fi observații dependente – astfel deducem că ar fi independente și toate celulele din conținutul tabelului de contingență teoretic au valori mai mari decât 5 (și niciuna nu este 0), astfel testul χ^2 poate fi aplicat pe aceste date.

3. Calculul statisticii χ^2 a testului și a valorii lui p

Statistica testului χ^2 are formula următoare: $\chi^2 = \sum_{i=1}^{l*c} \frac{(f_i^o - f_i^t)^2}{f_i^t}$

Unde χ^2 este statistica testului, l = numărul de linii în tabelul de contingență, c = numărul de coloane în tabelul de contingență, f_i^t = frecvența teoretică a celulei i, f_i^o = frecvența observată a celulei i.

Sub tabelul de contingență teoretic introduceți tabelul următor:

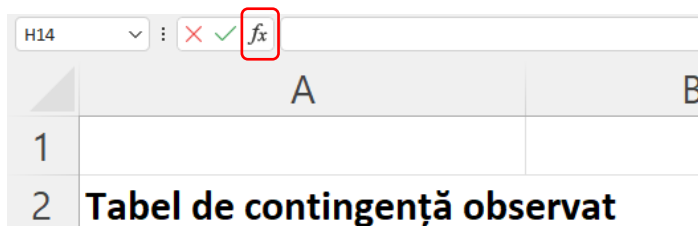
	F	G	H	I
1				
2	Tabel de contingență teoretic			
3				
4	Tratament/hiperglicemie	Da	Nu	Total
5	Metilprednisolon	40.83916	32.16084	73
6	Dexametazona	39.16084	30.83916	70
7	Total	80	63	143
8				
9				
10	Nivelul de semnificație statistică (alfa)		0,05	
11	Numărul de grade de libertate		1	
12	Valoarea critică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

În celula goală corespunzătoare statisticii testului introduceți formula următoare:

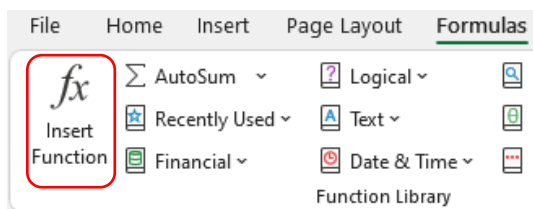
$$=(B5-G5)^2/G5+(C5-H5)^2/H5+(B6-G6)^2/G6+(C6-H6)^2/H6$$

În urma calculului veți obține pentru statistica testului valoarea de 16.79.

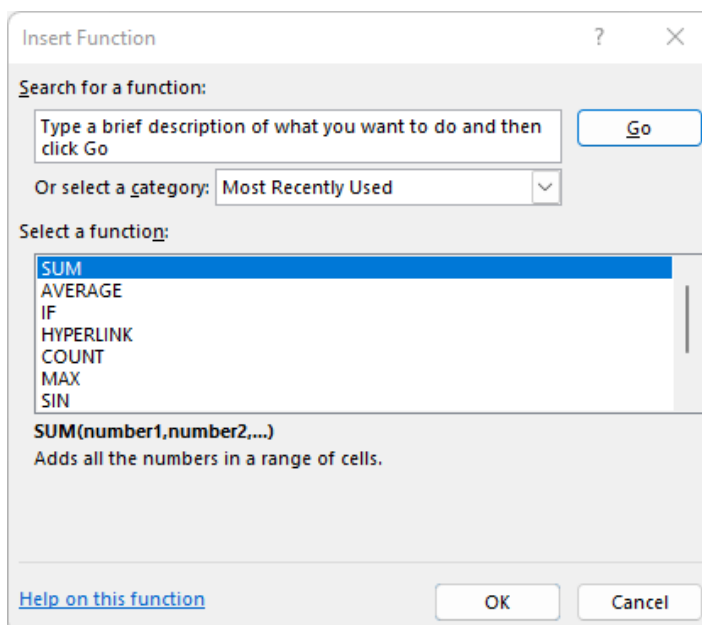
Pentru calculul valorii p a testului χ^2 , se va introduce o formulă predefinită din Excel, selectând celula goală din dreapta Valorii lui p, și apăsând butonul fx de deasupra coloanei A:



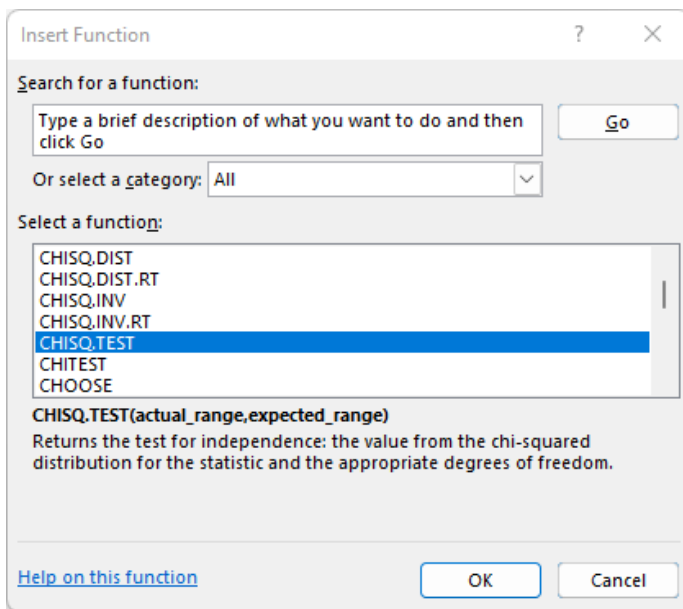
O altă variantă este să apăsăm pe butonul **Fx** din tabul **Formulas**, secțiunea **Function Library**.



Se va deschide fereastra următoare:



Pentru a găsi funcția necesară pentru calcularea valorii lui p, în al doilea câmp (**Or select a category**), selectăm **All**, pentru a avea acces la toate funcțiile, iar apoi identificăm în secțiunea **Select a function**, funcția **CHISQ.TEST** (sau CHITEST – versiunea veche formulei) o selectăm și apăsăm butonul **OK**.

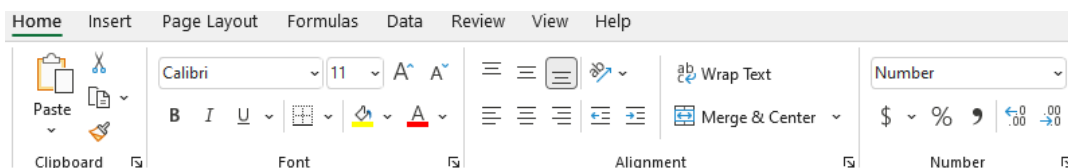


În fereastra deschisă, selectăm cele patru celule din interiorul tabelului de contingență observat în câmpul **Actual_range**, și selectăm cele patru celule din interiorul tabelului de contingență teoretic în câmpul **Expected_range**, după care apăsăm butonul **OK**.

Se va obține următoarea valoare:

Valoarea lui p	4.17246E-05
----------------	-------------

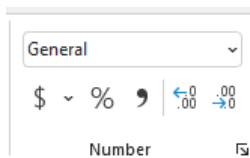
Aceasta valoare este formatul științific de reprezentare a valorilor numerice și înseamnă $4,17246 \cdot 10^{-5}$. Putem selecta această celulă și să schimbăm formatul ei în **Number**, în tabul **Home**, secțiunea **Number**:



Vom obține valoarea următoare:

Valoarea lui p	0.00
----------------	------

Pentru a vedea mai multe zecimale, putem apăsa pe săgețile din tabul Home, secțiunea Number:



Astfel putem obține valoarea în formatul următor:

Valoarea lui p	0.000042
----------------	----------

4. Interpretarea rezultatului testului statistic, pe baza statisticii χ^2 , iar apoi pe baza valorii lui p.

Introduceți sub tabelul cu statisticile testului χ^2 , calculate anterior, următoarele:

Interpretări:					
Ipoteza nulă:					
Ipoteza alternativă:					
Nivelul de semnificație statistică (alfa): 0,05					
Valoarea critică a testului Hi pătrat: 3,84					
Regiunea de respingere (critică):					
Interpretați rezultatul testului pe baza statisticii hi pătrat și a regiunii de respingere:					
Argumentați răspunsul anterior:					
Interpretați rezultatul testului pe baza statisticii valorii lui p:					
Argumentați răspunsul anterior:					

Răspunsurile corecte sunt următoarele:

Ipoteza nulă: Nu există asociere statistic semnificativă între tratament și hiperglicemie la subiecții cu COVID-19, sau nu există diferențe statistic semnificative între subiecții cu COVID-19 care au primit metilprednisolon și cei care au primit dexametazonă, în ce privește frecvența hiperglicemiei.

Ipoteza alternativă: Există asociere statistic semnificativă între tratament și hiperglicemie la subiecții cu COVID-19, sau există diferențe statistic semnificative între subiecții cu COVID-19 care au primit metilprednisolon și cei care au primit dexametazonă, în ce privește frecvența hiperglicemiei.

Nivelul de semnificație statistică (alfa): 0,05

Valoarea critică a testului Hi pătrat: 3,84 – pentru cazul a două variabile calitative dihotomiale.

Regiunea de respingere (critică): [3,84; + infinit), regiunea se construiește pe baza valorii critice [valoare critică; + infinit).

Interpretați rezultatul testului pe baza statisticii hi pătrat și a regiunii de respingere: Există asociere statistic semnificativă între tratament și hiperglicemie la subiecții cu COVID-19, sau Există diferențe statistic semnificative între subiecții cu COVID-19.

Argumentați răspunsul: întrucât statistica testului, de 16,79 se află în regiunea de respingere (critică) - [3,84; + infinit), respingem ipoteza nulă și acceptăm ipoteza alternativă.

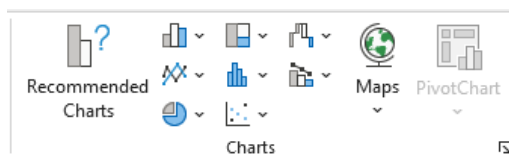
Interpretați rezultatul testului pe baza statisticii valorii lui p: Există asociere statistic semnificativă între tratament și hiperglicemie la subiecții cu COVID-19, sau există diferențe statistic semnificative între subiecții cu COVID-19 care au primit metilprednisolon și cei care au primit dexametazonă, în ce privește frecvența hiperglicemiei.

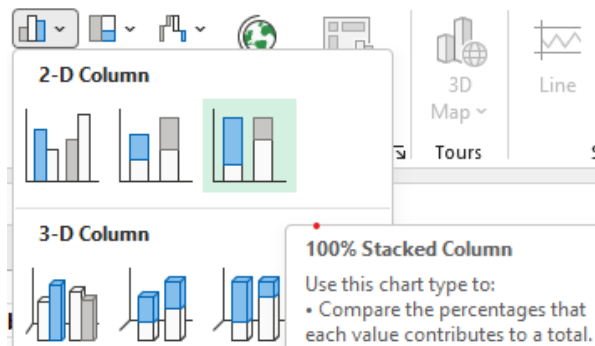
Argumentați răspunsul: întrucât valoarea lui p (0,000042) este mai mică decât 0,05 (nivelul de semnificație statistică, alfa), respingem ipoteza nulă și acceptăm ipoteza alternativă.

5. Pentru a alege graficul ne gândim câte variabile sunt implicate: două. Apoi ne gândim la tipul acestora. Ambele sunt de tip calitativ dihotomial. Pentru această situație, a unui grafic ce identifică legătura dintre două variabile de tip calitativ, sunt potrivite grafice de tip coloană sau bară, preferabil procentual.

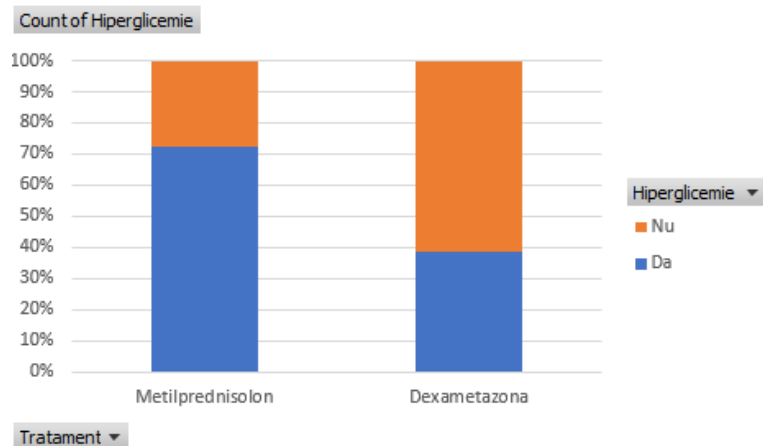
Selecționăm o celulă din tabelul de contingență observat, și în tabul **Insert**,

Insert, secțiunea **Charts**, alegem **100% Stacked Column**.

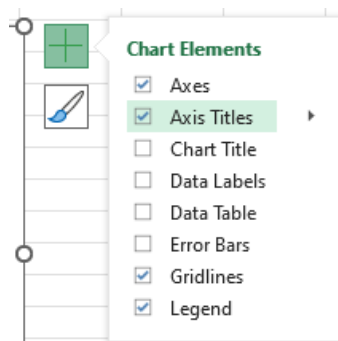


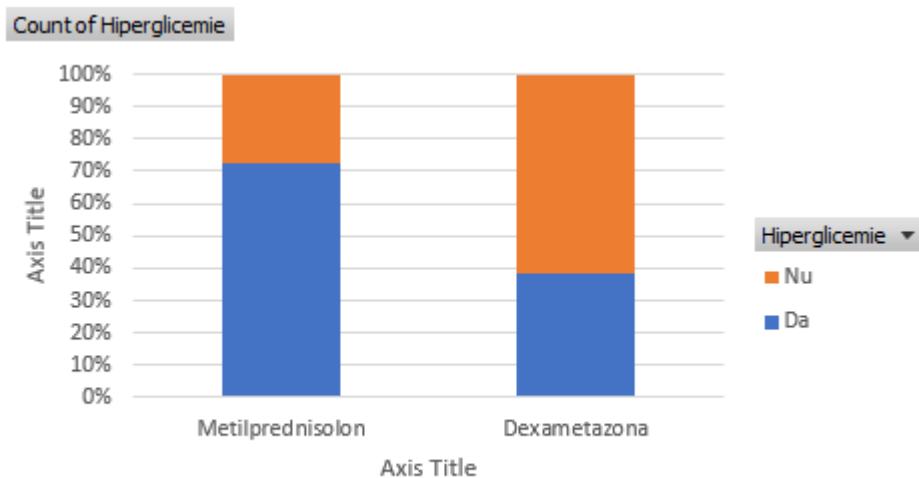


Obținem următorul grafic:



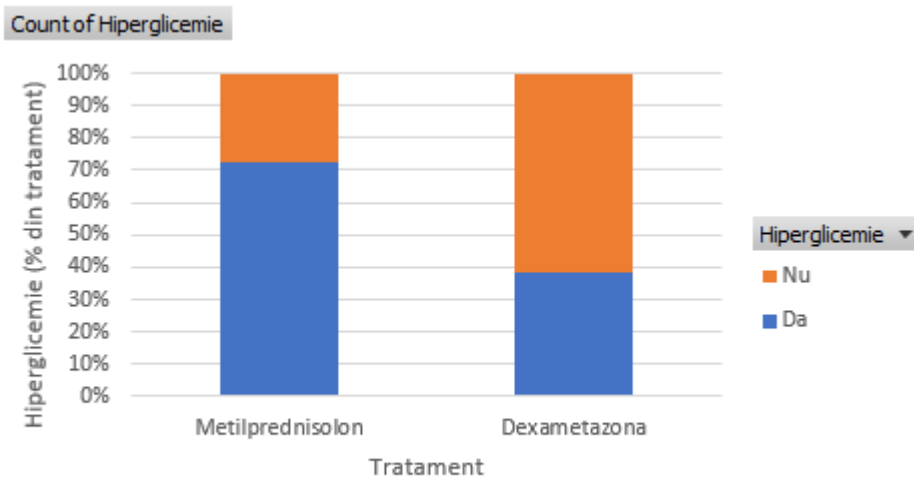
Întrucât nu avem titlul axelor trecute pe grafic, dăm un clic pe grafic într-o zonă fără elemente și apoi apăsăm pe butonul + din colțul dreapta sus al graficului și bifăm **Axis Titles**:





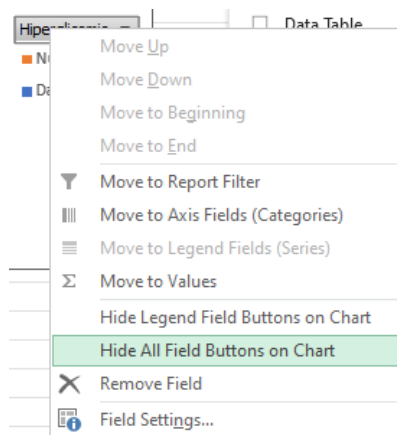
Tratament ▼

Denumim axele conform imaginii de mai jos:

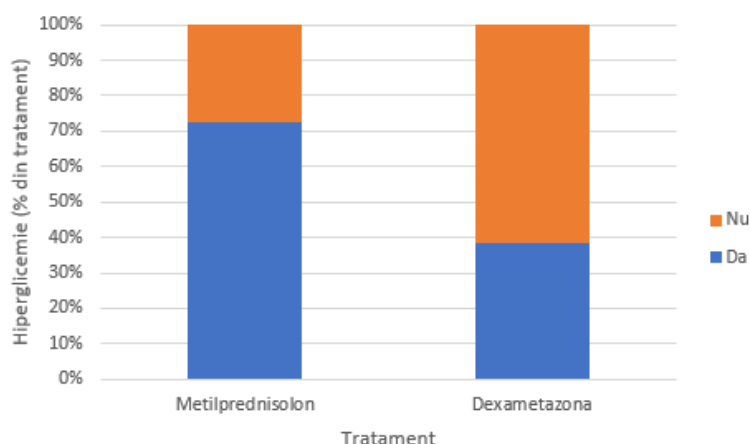


Tratament ▼

La final, dăm clic dreapta pe butonul de la legendă și alegem **Hide all field buttons on chart**:



Se obține imaginea următoare:



În imagine se observă că în grupul care a primit metilprednisolon există un procent mai mare de subiecți cu hiperglicemie (pe grafic colorat cu albastru), față de cei din grupul care au primit dexametazonă. Diferențele dintre cele două grupuri par relativ mari.

- Pentru a afla procentul subiecților cu hiperglicemie în funcție de tratament, vom copia tabelul de contingență observat în altă poziție pe foaia de calcul. Îl selectăm inițial:

2	Tabel de contingență observat			
3	Count of Hiperglicemie	Column Labels		
4	Row Labels	Da	Nu	Grand Total
5	Metilprednisolon	53	20	73
6	Dexametazona	27	43	70
7	Grand Total	80	63	143

Apoi selectăm o celulă goală și cu clic dreapta lipim conținutul:

	A	B	C	D
1				
2	Tabel de contingență observat			
3	Count of Hiperglicemie	Column Labels		
4	Row Labels	Da	Nu	Grand Total
5	Metilprednisolon	53	20	73
6	Dexametazona	27	43	70
7	Grand Total	80	63	143
8				
9				
10	Count of Hiperglicemie	Column Labels		
11	Row Labels	Da	Nu	Grand Total
12	Metilprednisolon	53	20	73
13	Dexametazona	27	43	70
14	Grand Total	80	63	143

Selectăm oricare celulă din conținutul tabelului observat, nu din total, nici din categoriile variabilelor, facem clic dreapta și din meniu alegem Show Value As - % of Row Total (să prezinte valorile sub formă de procente pe linii):

În final obținem tabelul dorit:

	A	B	C	D
10	Count of Hiperglicemie	Column Labels		
11	Row Labels	Da	Nu	Grand Total
12	Metilprednisolon	72.60%	27.40%	100.00%
13	Dexametazona	38.57%	61.43%	100.00%
14	Grand Total	55.94%	44.06%	100.00%

În acest tabel se observă că în grupul care a primit metilprednisolon au fost 72,6% subiecți care au avut hiperglicemie, iar în grupul care a primit dexametazonă

38,57% au avut hiperglicemie. Astfel cei care au primit metilprednisolon au avut cu 34% mai frecvent hiperglicemie față de cei care au primit dexametazonă.

Denumiți foaia de calcul: Tratament Hiperglicemie.

7. Tabelul observat, teoretic și valoarea statisticii χ^2 , precum și valoarea lui p, pentru testarea asocierii dintre tratament și decese sunt prezentate mai jos:

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		

Interpretarea acestor rezultate se află în continuare:

Ipoteza nulă: Nu există asociere statistic semnificativă între tratament și decese la subiecții cu COVID-19, sau nu există diferențe statistic semnificative între subiecții cu COVID-19 care au primit metilprednisolon și cei care au primit dexametazonă, în ce privește frecvența deceselor.

Ipoteza alternativă: Există asociere statistic semnificativă între tratament și decese la subiecții cu COVID-19, sau există diferențe statistic semnificative între subiecții cu COVID-19 care au primit metilprednisolon și cei care au primit dexametazonă, în ce privește frecvența deceselor.

Nivelul de semnificație statistică (alfa): 0,05

Valoarea critică a testului Hi pătrat: 3,84 – pentru cazul a două variabile calitative dihotomiale.

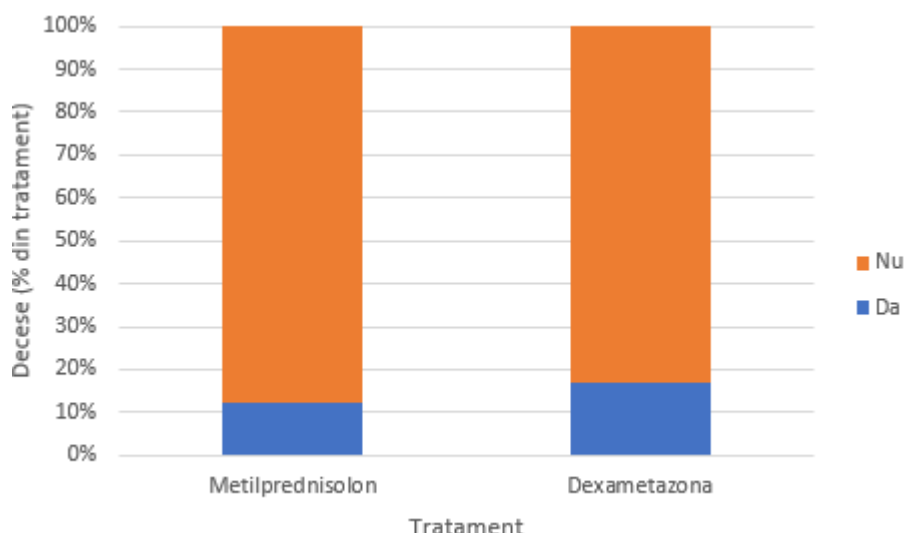
Regiunea de respingere (critică): [3,84; + infinit), regiunea se construiește pe baza valorii critice [valoare critică; + infinit).

Interpretați rezultatul testului pe baza statisticii hi pătrat și a regiunii de respingere: nu putem spune că există asociere statistic semnificativă între tratament și decese la subiecții cu COVID-19, sau nu putem spune că există diferențe statistic semnificative între subiecții cu COVID-19 care au primit

metilprednisolon și cei care au primit dexametazonă, în ce privește frecvența deceselor.

Argumentați răspunsul: întrucât statistica testului, de 0,661, nu se află în regiunea de respingere (critică) - $[3,84; +\infty)$, nu putem respinge ipoteza nulă.

8. Se va obține un grafic precum următorul:



În imagine se observă că în grupul care a primit metilprednisolon există un procent mai mic de subiecți cu decese (pe grafic colorat cu albastru), față de cei din grupul care au primit dexametazonă. Totuși diferențele dintre cele două grupuri par relativ mici.

9. Se va obține un tabel precum următorul:

10	Count of Deces	Column Labels ▼		
11	Row Labels ▼	Da	Nu	Grand Total
12	Metilprednisolon	12.33%	87.67%	100.00%
13	Dexametazona	17.14%	82.86%	100.00%
14	Grand Total	14.69%	85.31%	100.00%

În acest tabel se observă că în grupul care a primit metilprednisolon au fost 12,33% subiecți care au decedat, iar în grupul care a primit dexametazonă 17,14% au decedat. Astfel cei care au primit metilprednisolon au decedat cu 4,81% mai rar față de cei care au primit dexametazonă.

Denumiți foaia de calcul: Tratament Decese.

Salvați fișierul.

DE REȚINUT!

- În situația în care ne interesează asocierea dintre două variabile calitative, cu categorii mutual exclusive, observații independente, și frecvențe așteptate suficient de mari, unul din testele pe care le putem folosi este testul χ^2 .
- Pentru a obține valoarea lui p a testului χ^2 în Excel, se inseră un tabel de contingență observat cu ajutorul PivotTable, apoi se calculează tabelul teoretic (fiecare celulă de conținut de tabel fiind produsul dintre totalul pe linie și cel pe coloană corespunzător celulei, împărțit la totalul pe întregul tabel), iar în final se folosește funcția CHISQ.TEST, selectând conținutul celor două tabele.