

REVISION (Recapitulation)
- EPREUVE PRATIQUE -

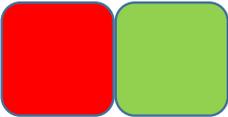


Évaluation de l'activité didactique

- les étudiants ont la possibilité d'évaluer l'activité didactique du cours et les stages/LP du premier semestre de l'année universitaire en cours (jusqu'au 14.01.2024).
- chaque étudiant a reçu à l'adresse institutionnelle (...@elearn.umfcluj.ro), un lien unique et confidentiel où il peut évaluer l'activité didactique.
- l'accès au compte de messagerie s'effectue avec les données d'authentification utilisées pour l'application MS Teams,
- le questionnaire peut également être complété depuis le téléphone mobile
- si les étudiants rencontrent des difficultés pour accéder aux formulaires ou identifient des erreurs dans les listes de matières ou dans le personnel enseignant, écrivez un email au Service d'Assurance Qualité: **dep.calitate@umfcluj.ro**.

Savoir les TYPES DES VARIABLES

- Variables **QUALITATIVES**

- ◇ dichotomiques 
- ◇ nominales 
- ◇ ordinales 

- Variables **QUANTITATIVES**

- ◇ **continues**: variables mesurées en nombres réels
- ◇ **discrètes (discontinues)**: variables mesurées en nombres entiers

Statistique descriptive: Description d'une seule variable QUALITATIVE

- Création d'un:

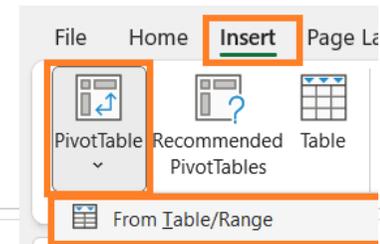
i) tableau de fréquences:

COVID-19	Fréquence absolue (nombre de patients)
positif	272
négatif	128
Total	400

- **EXCEL:**

~ à l'aide de la fonction prédéfini **COUNTIF**:
`=COUNTIF(H2:H401,"positif")`
`=COUNTIF(H2:H401,"negatif")`

~ à l'aide de l'option **PIVOTTABLE**:



Statistique descriptive: Description de la relation entre DEUX variables QUALITATIVES

- Création d'un:

ii) tableau de contingence:

	COVID-19 (positif)	COVID-19 (négatif)	Total
Tabagisme (oui)	228	104	332
Tabagisme (non)	44	24	68
Total	272	128	400

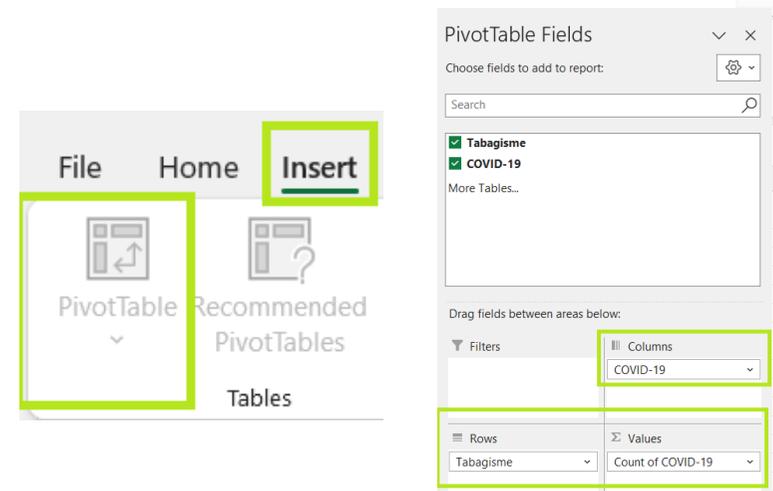
- **EXCEL:**

~ à l'aide de l'option **PIVOTTABLE** (!!!)

~ à l'aide de la fonction prédéfini COUNTIF (cela implique de trier les données selon une variable puis d'utiliser la fonction COUNTIF)

Statistique descriptive: Description de la relation entre DEUX variables

- **Tableau de contingence**
- **EXCEL:**
 - ~ à l'aide de l'option **PIVOTTABLE (!!!)**



Count of Tabagisme	Column Labels		
Row Labels	negatif	positif	Grand Total
non	24	44	68
oui	104	228	332
Grand Total	128	272	400



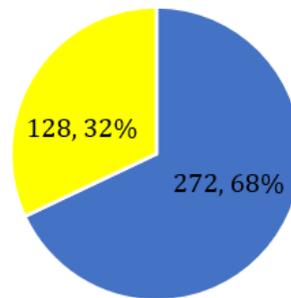
Count of Tabagisme	Column Labels		
Row Labels	positif	negatif	Grand Total
oui	228	104	332
non	44	24	68
Grand Total	272	128	400

Représentation graphique d'une seule variable QUALITATIVE

- Réalisation d'une:

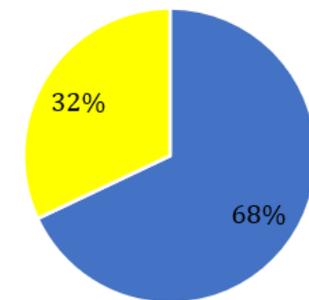
i) graphique sectoriel/camembert [engl. PIE]

Repartition des sujets par rapport au COVID-19



■ positif ■ negatif

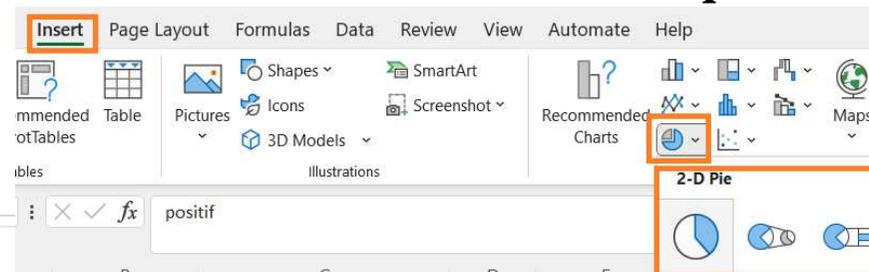
Repartition des sujets par rapport au COVID-19



■ positif ■ negatif

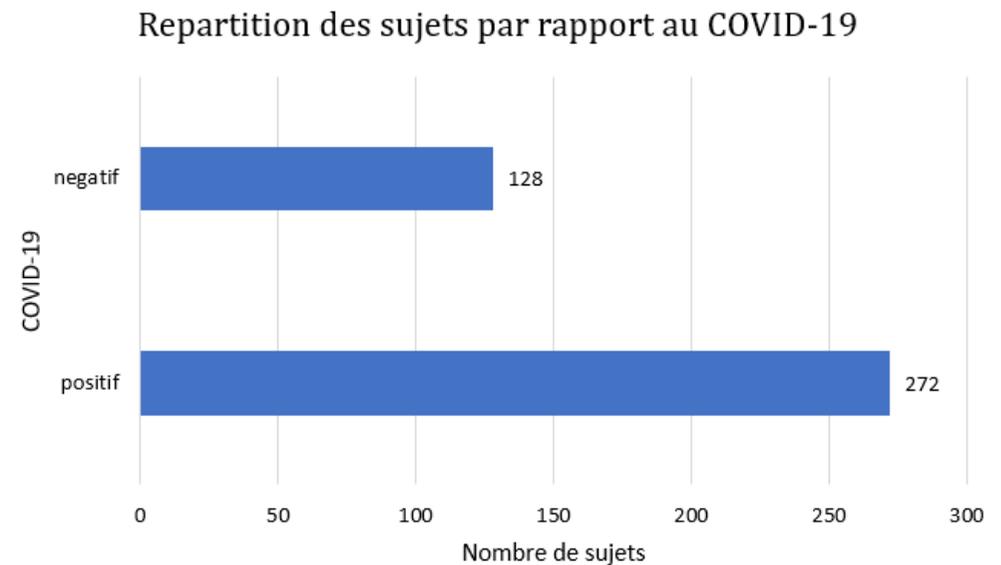
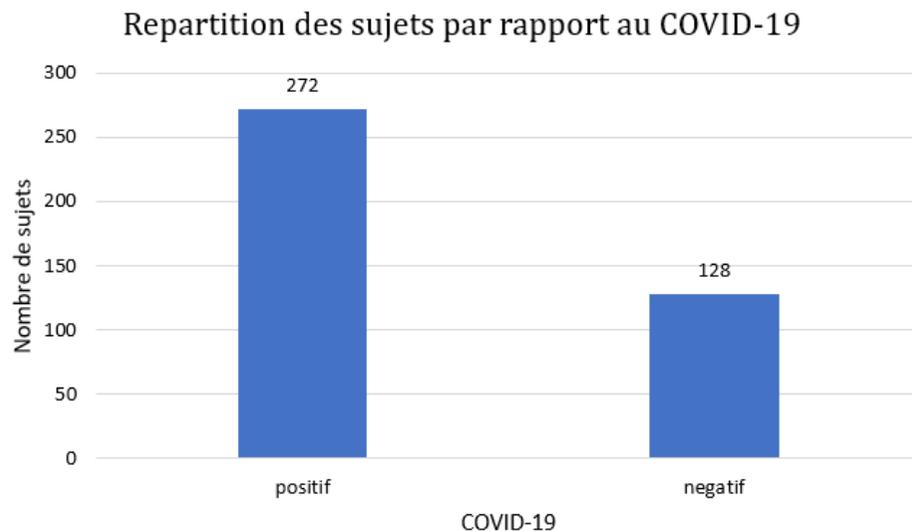
- **EXCEL:**

~ cela implique de créer préalablement un tableau de fréquences (voir les diapo antérieures) ⇒



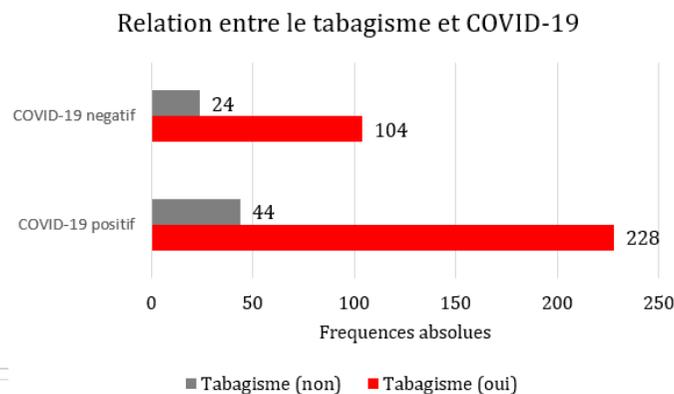
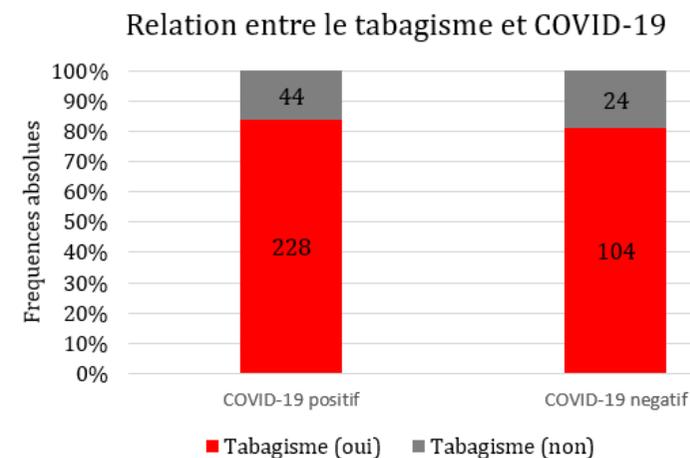
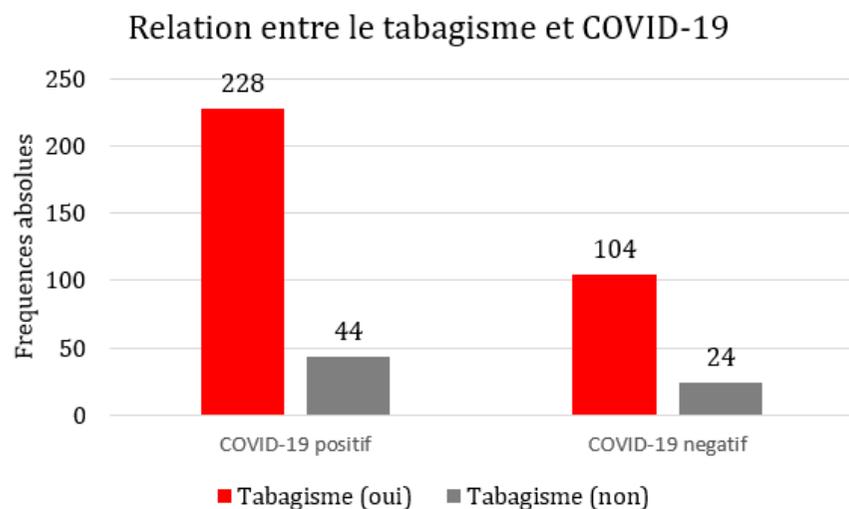
Statistique descriptive: Représentation graphique d'une seule variable QUALITATIVE

OU ii) Graphique par **des COLONNES** ou **BARRES** [engl. COLUMN/BAR]



Représentation graphique de la relation entre DEUX variables QUALITATIVES

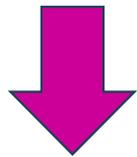
- Création d'un graphique par **des COLONNES ou BARRES** [engl. COLUMN/BAR]



Représentation graphique de la relation entre DEUX variables QUALITATIVES

- le graphique par des COLONNES (en % par rapport aux totaux des lignes) peut être obtenu à l'aide du tableau pivot (on transforme les fréquences du tableau pivot en %)

Count of Tabagisme	Column Labels		Grand Total
Row Labels	positif	negatif	
oui	228	104	332
non	44	24	68
Grand Total	272	128	400



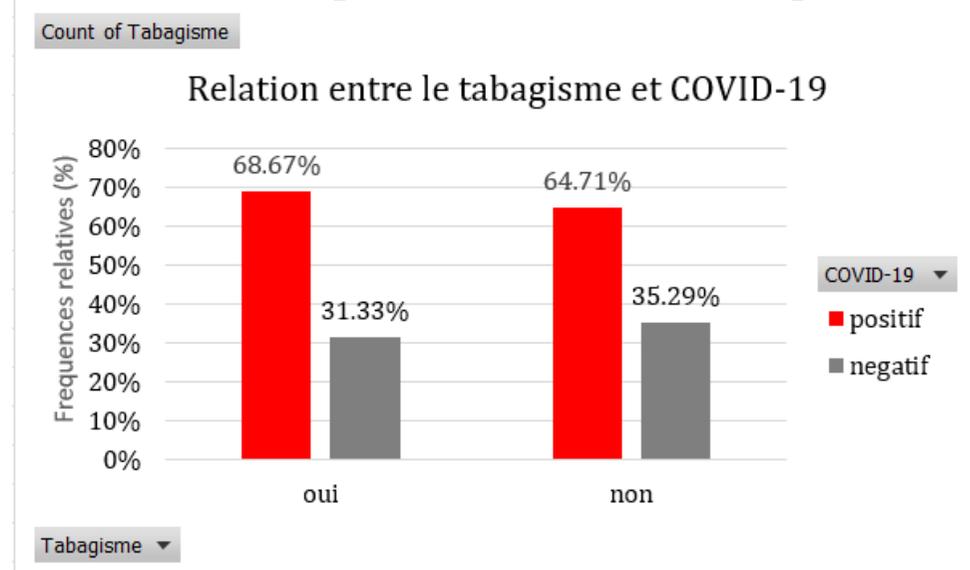
Summarize Values By > % of Grand Total

Show Values As > % of Column Total

% of Row Total

Display all the values in each row as a percentage of the total for the % of...

Count of Tabagisme	Column Labels		Grand Total
Row Labels	positif	negatif	
oui	68.67%	31.33%	100.00%
non	64.71%	35.29%	100.00%
Grand Total	68.00%	32.00%	100.00%



Représentation graphique de la relation entre DEUX variables QUALITATIVES

- le graphique par des COLONNES (en % par rapport aux **totaux des colonnes**) peut être obtenu à l'aide du tableau pivot (on transforme les fréquences du tableau pivot en %)

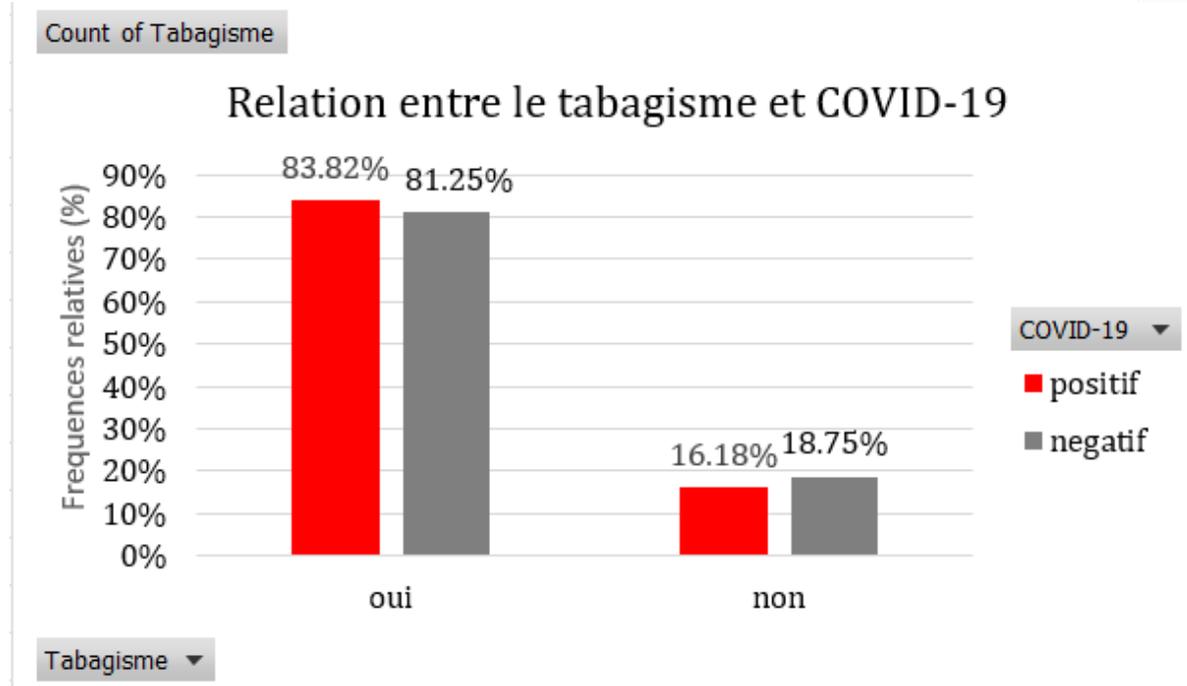
Count of Tabagisme	Column Labels			
Row Labels	positif	negatif	Grand Total	
oui	228	104	332	
non	44	24	68	
Grand Total	272	128	400	

Summarize Values By > % of Grand Total

Show Values As > % of Column Total

Value Field Settings... % of Row Total

Count of Tabagisme	Column Labels			
Row Labels	positif	negatif	Grand Total	
oui	83.82%	81.25%	83.00%	
non	16.18%	18.75%	17.00%	
Grand Total	100.00%	100.00%	100.00%	

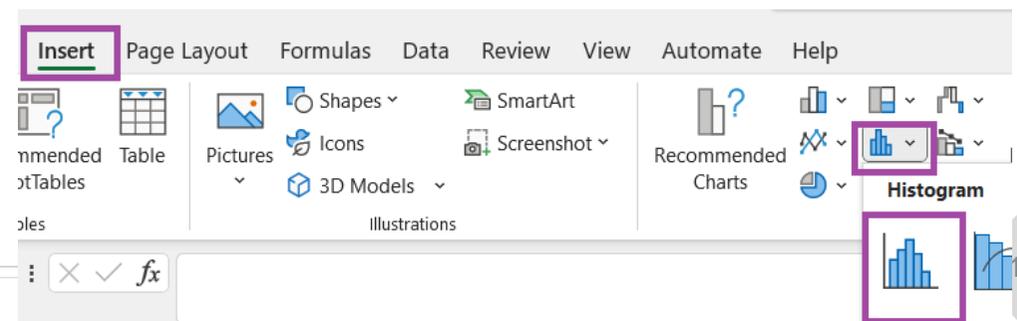
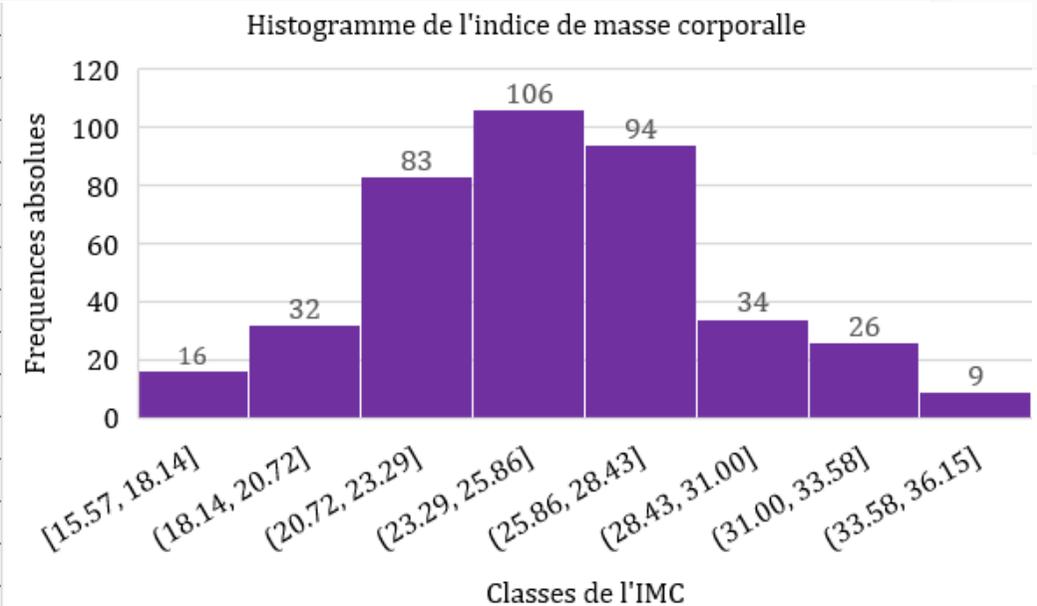


Représentation graphique d'une seule variable QUANTITATIVE

- Réalisation d'une:
 - i) **Histogramme**: utile pour interpréter la normalité des données

- **EXCEL:**

~ cela implique la sélection de données ⇒



Représentation graphique d'une seule variable QUANTITATIVE

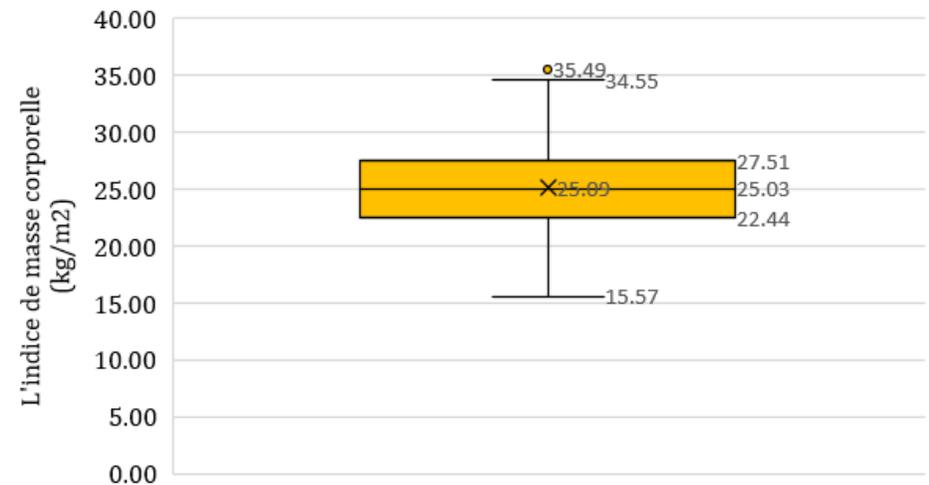
- Réalisation d'un:

ii) **Graphique boîte a moustaches** : utile pour interpréter la normalité des données (ou/et la dispersion des données)

- **EXCEL:**

~ cela implique la sélection de données (les valeurs de la variable) ⇒ **INSERT** → **CHART** → **Box & Whisker plot**

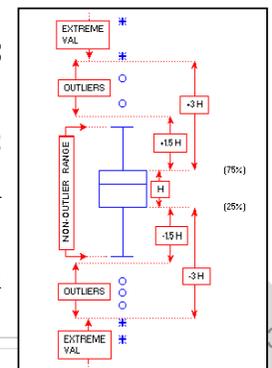
Distribution de l'indice de masse corporelle



Note: si certaines valeurs sont en dehors des moustaches, on les représente par des * ou ° :

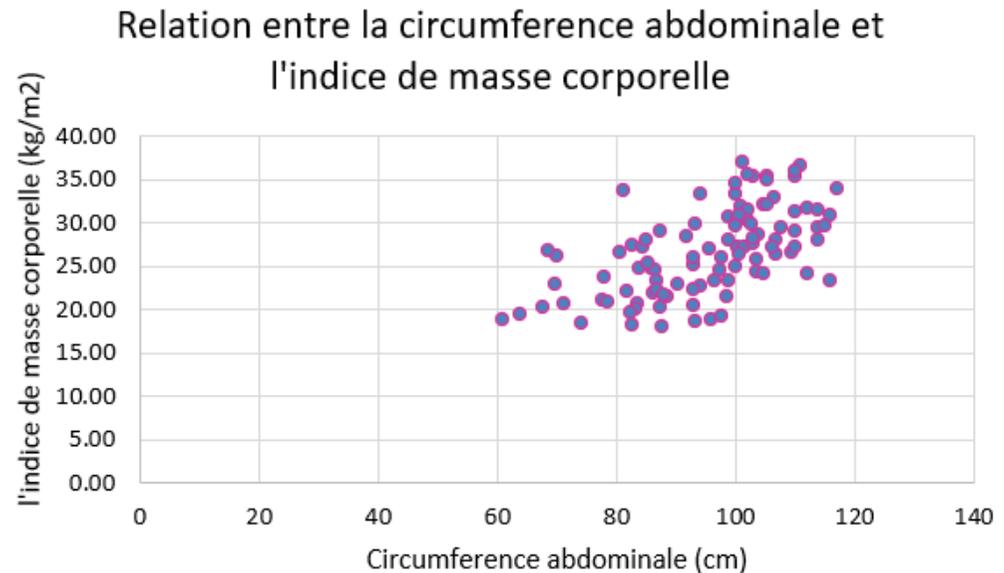
*les valeurs extrêmes = valeurs inhabituelles inférieurs à $Q_1 - 3 \times \text{EQR}$ ou supérieurs à $Q_3 + 3 \times \text{EQR}$ ou/et

° les outliers = valeurs inférieurs à $Q_1 - 1,5 \times \text{EQR}$ ou supérieurs à $Q_3 + 1,5 \times \text{EQR}$



Représentation graphique de la relation entre DEUX variables QUANTITATIVES

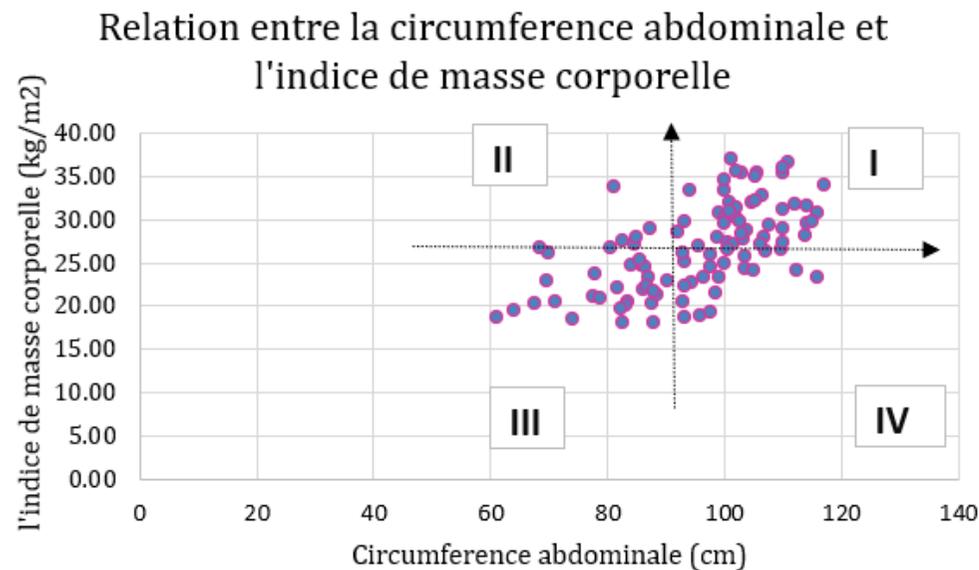
- le nouage de points
- Interprétation du graphique



- **EXCEL:** cela implique la sélection de données ⇒ **INSERT** → **CHART** → **SCATTER**

Représentation graphique de la relation entre DEUX variables QUANTITATIVES

- Les points du diagramme sont préférentiellement répartis dans les quadrants I et III.
- Dépendance entre la circonférence abdominale et l'IMC est positive: une augmentation de la circonférence abdominale implique une augmentation de l'EIM.



95 moyenne de la circonférence abd
26.68 moyenne de l'IMC



les moyennes ont été
déterminées par la fonction
AVERAGE

Représentation graphique des deux variables (QUANTITATIVE + QUALITATIVE)

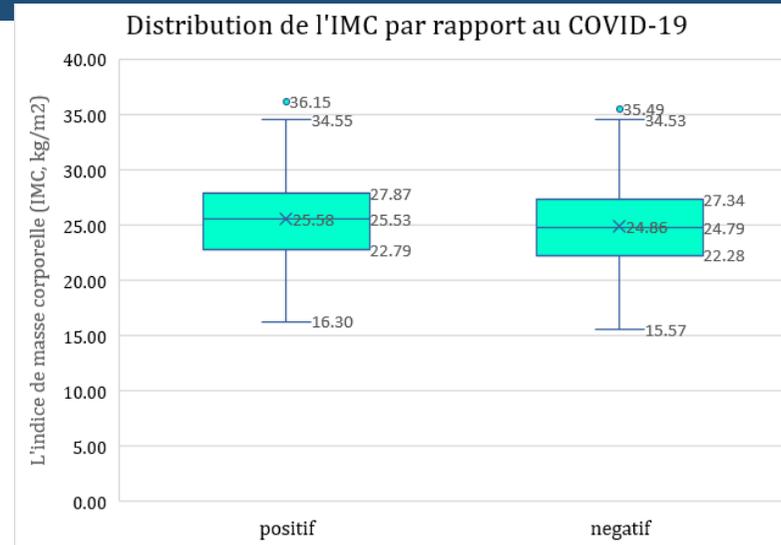
- Réalisation d'un:

i) **Graphique boîte a moustaches** : utile pour interpréter la normalité des données par rapport aux groupes (ou/et la dispersion des données)

- **EXCEL:**

~ cela implique la sélection de données (les valeurs des 2 variables) ⇒ **INSERT** → **CHART** → **Box & Whisker plot**

~ la variable qualitative qui donne les groupes doit se trouver avant la var.quantitative dans la feuille de calcul ⇒ **INSERT** → **CHART** → **Box & Whisker plot**



	A	B
COVID-19		IMC(kg/m2)
positif		22.29
positif		32.08
negatif		25.00
positif		28.13
positif		32.41

Principes pour la bonne réalisation des graphiques

Toute représentation graphique doit avoir:

- ✓ titre clair, concis et précis
- ✓ définitions des axes, sans abréviations (à l'exception des unités de mesure)
unités de mesure (ou il est le cas)
légende (s'il faut)
- ✓ tous les symboles, des abréviations ou lettres utilisés dans la figure doivent être expliqués clairement dans la légende

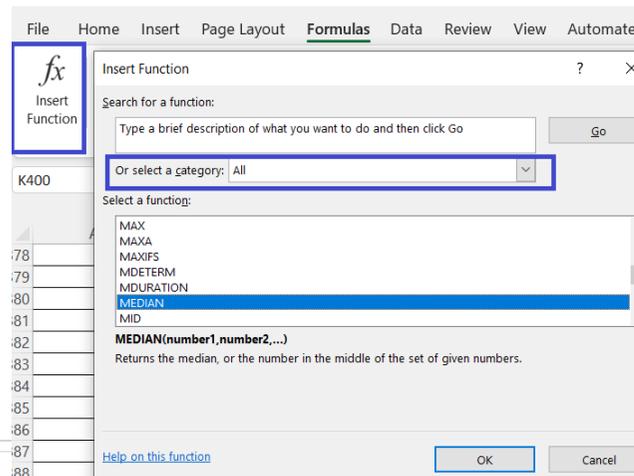
Les Statistiques descriptives

- **pour décrire une variable QUALITATIVE:**

- ◊ **dichotomique, nominale:** fréquences absolues, fréquences relatives
- ◊ **ordinales:** fréquences absolues, fréquences relatives, fréquences relatives cumulées croissantes, le mode, la médiane, les quartiles, le minimum, le maximum

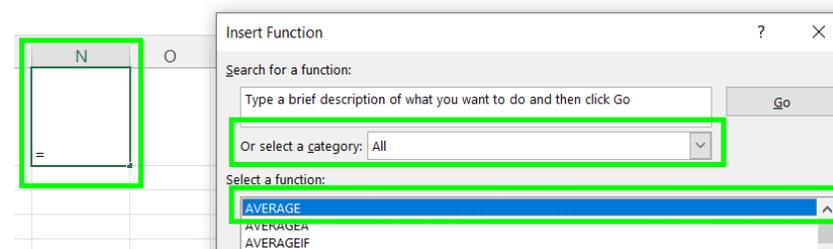
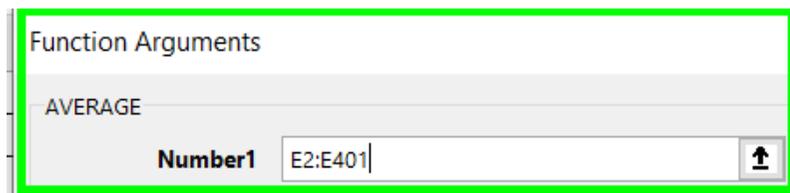
- **EXCEL:**

- ~ à l'aide de l'option PIVOTTABLE ⇒ fréquences absolues
- ~ fréquences relatives: = fréquences absolues/ taille du groupe;
- ~ fréquences relatives cumulées croissantes: = **la somme des freq. relatives** jusqu'à la valeur donnée x_i
- ~ Médiane: MEDIAN
- ~ Mode: MODE
- ~ Quartiles: QUARTILE
- ~ minimum: MIN
- ~ Maximum: MAX



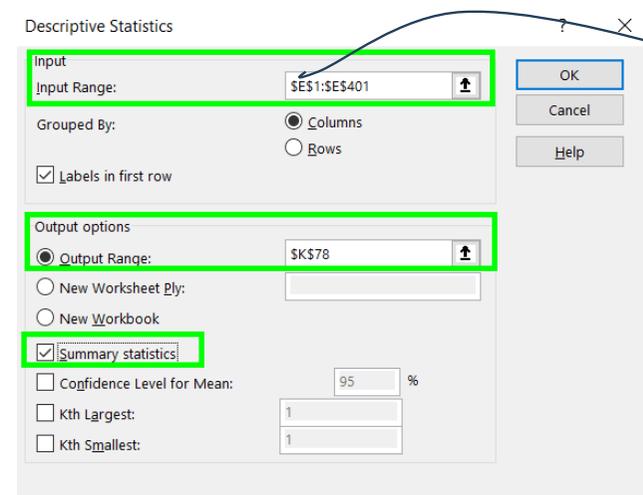
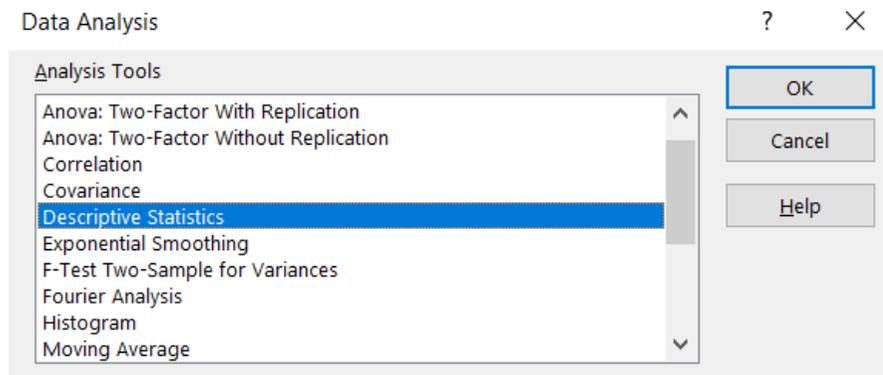
Les Statistiques descriptives

- **pour décrire une variable QUANTITATIVE:** la moyenne, médiane, mode, variance, minimum, maximum, l'amplitude, les Quartiles, les percentiles, la déviation standard, l'erreur standard, le coefficient de variation, l'asymétrie, le coefficient d'aplatissement
- **EXCEL:**
 - ◇ **en utilisant des fonctions prédéfinis:** AVERAGE, MEDIAN, MODE, VAR (ou VAR.S), MIN, MAX, QUARTILE, PERCENTILE, STDEV (ou STDEV.S), SKEW, KURT
 - ◇ ex. utilisation de la fonction AVERAGE (calcul de la moyenne)
 - ◇ choisir la cellule pour retourner le résultat de la fonction
 - ◇ sélectionner les valeurs de la variable d'intérêt



Les Statistiques descriptives

- ◇ **en utilisant des formules** (ou est le cas, par exemple: l'amplitude = max-min, le coefficient de variation, l'erreur standard, écart interquartiles $EQR = Q3 - Q1$)
- ◇ **en utilisant l'option DATA ANALYSIS → Descriptive statistics:**



la selection de données
(valeurs +/- en-tête de
colonne/s)

Savoir évaluer l'homogénéité des données

L'interprétation du coefficient de variation d'une série des données

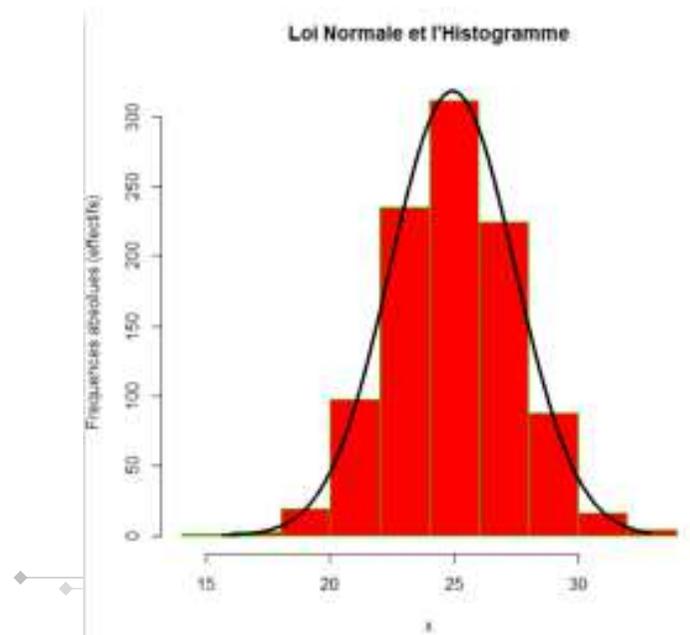
Valeur du coefficient de variation	interprétation
$CV < 10\%$	Série de données homogène
$10\% \leq CV < 20\%$	Série de données relativement homogène
$20\% \leq CV < 30\%$	Série de données relativement hétérogène
$CV \geq 30\%$	Série de données hétérogène

Savoir évaluer la normalité des données

- **en utilisant les statistiques descriptives:**

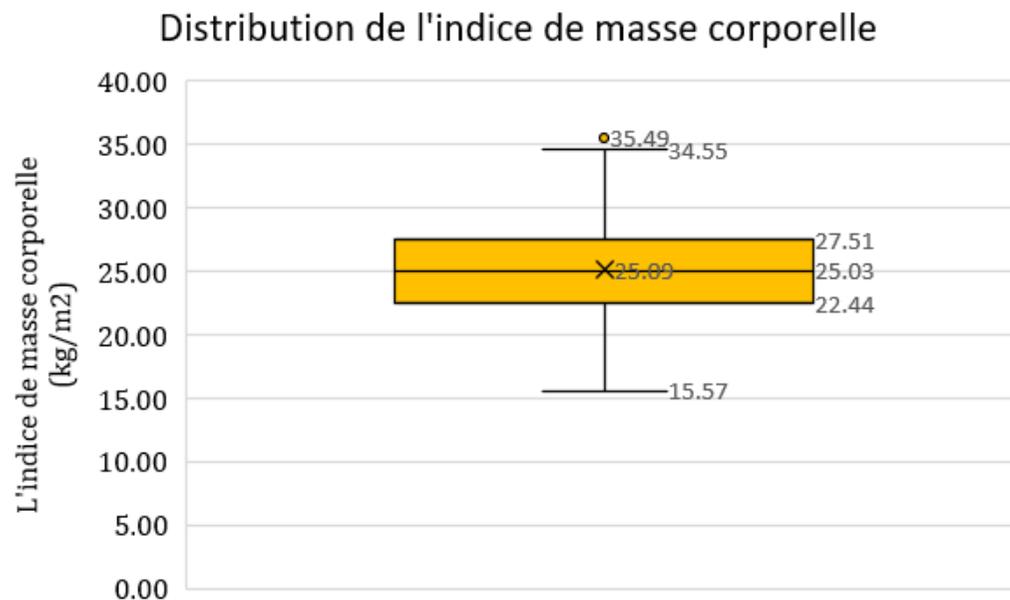
la moyenne \approx la médiane \approx le mode ET le coefficient d'aplatissement ≈ 0 (le coeff. $\in [-1,1]$) ET le coefficient d'asymétrie ≈ 0 (le coeff. $\in [-1,1]$)

- **en utilisant l'histogramme**

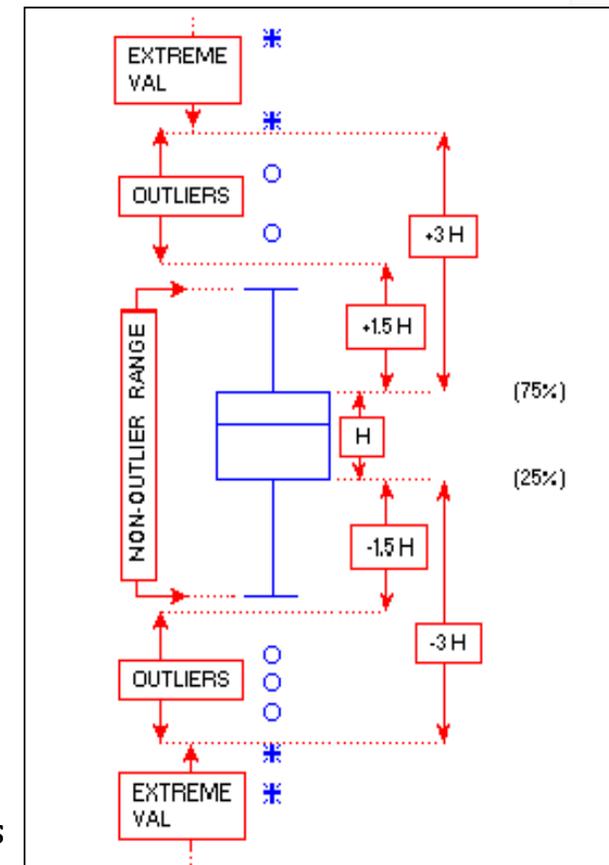


Savoir évaluer la normalité des données

- en utilisant le boite a moustaches



- Note: si certains valeurs sont en dehors des moustaches, on les représente par des
- les valeurs extrêmes = valeurs inhabituelles inférieures à $Q_1 - 3 \times \text{EQR}$ ou supérieures à $Q_3 + 3 \times \text{EQR}$ ou/et
 - ° les outliers = valeurs inférieures à $Q_1 - 1,5 \times \text{EQR}$ ou supérieures à $Q_3 + 1,5 \times \text{EQR}$



$$H = \text{EQR} = Q_3 - Q_1$$

Savoir les Calculs et les interprétations des PROBABILITES par rapport à un tableau de fréquence/contingence

- Tableau de contingence:

	COVID-19 (positif)	COVID-19 (négatif)	Total
Tabagisme (oui)	104	228	332
Tabagisme (non)	24	44	68
Total	128	272	400

Soit les évènements: $F = \{\text{être fumeur}\}$; $M = \{\text{COVID-19 positif}\}$

- savoir faire le calcul des probabilités de la forme suivante:

◊ $Pr(F)$; $Pr(M)$; $Pr(nonF)$; $Pr(nonM)$

◊ $Pr(M \cap F)$; $Pr(M \cup F)$; $Pr(M/F)$; $Pr(F/M)$; $Pr(nonM/nonF)$; $Pr(nonF/nonM)$;

- EXCEL:

	COVID-19 positif	COVID-19 negatif	Total				
Tabagisme(oui)	104	228	332	Pr(M)	=128/400	Pr(M et F)	=104/400
Tabagisme (non)	24	44	68	Pr(nonF)	=68/400	Pr(M ou F)	=N45+N46-N49
Total	128	272	400	Pr(nonM)	=272/400	Pr(M/F)	=104/332
		Pr(F)	=332/400	Pr(F)	0.83	Pr(F/M)	=104/128

Savoir le calcul et l'interprétation des probabilités conditionnées : VPP, VPN, Se, Sp

	M	nonM	Total
T	a	b	a+b
nonT	c	d	c+d
Total	a+c	b+d	N

$$Se = \Pr(T | M) \approx a / (a + c)$$

$$Sp = \Pr(\text{nonT} | \text{nonM}) \approx d / (b + d)$$

$$VPP = \Pr(M | T) \approx a / (a + b)$$

$$VPN = \Pr(\text{nonM} | \text{nonT}) \approx d / (c + d)$$

Considérons les évènements:

M = {avoir la maladie}

nonM = {sans la maladie}

T = {résultat positif au nouveau test diagnostique chez un patient choisi au hasard}

nonT = {résultat négatif au nouveau test diagnostique}

Savoir le Calcul et l'interprétation du Risque Relatif (ou l'Odds Ratio)

• calcul du RR et l'OR

	COVID-19 (positif)	COVID-19 (négatif)	Total
Tabagisme (oui)	104 =a	238=b	342=a+b
Tabagisme (non)	14=c	44=d	58=c+d
Total	118=a+c	282=b+d	400

$$RR = \frac{\Pr(M/F)}{\Pr(M/nonF)} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

$$OR = \frac{\frac{\Pr(M/F)}{\Pr(nonM/F)}}{\frac{\Pr(M/nonF)}{\Pr(nonM/nonF)}} = \frac{a \times d}{b \times c}$$

Interprétation:

- Si RR = 1, alors le risque d'avoir la maladie est le même dans les deux groupes d'exposition
- Si RR > 1, alors un sujet exposé a un risque d'être malade plus élevé qu'un sujet non exposé (F est un facteur de risque pour la maladie M)
- Si le RR < 1, un sujet exposé a un risque d'être malade plus faible qu'un sujet non exposé (F est un facteur protecteur pour la maladie M)

• EXCEL:

	COVID-19 positif	COVID-19 négatif	Total
Tabagisme(oui)	104	238	342
Tabagisme (non)	14	44	58
Total	118	282	400

RR = (104/342)/(14/58)

RR	1.259816207	OR	=(104*44)/(14*238)	OR	1.37334934
----	-------------	----	--------------------	----	------------

Corrélation et Régression Linéaires

- Calcul du coefficient de corrélation linéaire de Pearson
- Savoir l'interprétation du coefficient (la direction et l'importance de la relation - règles du Colton)
- La droite de régression [obtenue sur le graphique nouage des points].
- L'interprétation des résultats [les deux coefficients de la droite de régression : a_1 = la pente et a_0]
- Le coefficient de détermination [obtenue sur le graphique nouage des points]
- L'interprétation du coefficient de détermination
- **EXCEL: en utilisant la fonction CORREL**

=CORREL(D2:D101,E2:E101)

	C	D	E
	Genre	Circonférence abdominale(cm)	IMC(kg/m2)
5	M	112	31.73
0	F	94	22.66
0	F	100	33.39

Savoir l'analyse de Corrélation et Régression Linéaires

- Calcul du coefficient de corrélation linéaire de Pearson
- Savoir l'interprétation du coefficient (la direction et l'importance de la relation - règles du Colton)
- La droite de régression [obtenue sur le graphique nouage des points].
- L'interprétation des résultats [les deux coefficients de la droite de régression : a_1 = la pente et a_0]
- Le coefficient de détermination [obtenue sur le graphique nouage des points]
- L'interprétation du coefficient de détermination
- **EXCEL: en utilisant la fonction CORREL**

	C	D	E
	Genre	Circonférence abdominale(cm)	IMC(kg/m ²)
5	M	112	31.73
6	F	94	22.66
7	F	100	33.39

`=CORREL(D2:D101,E2:E101)`

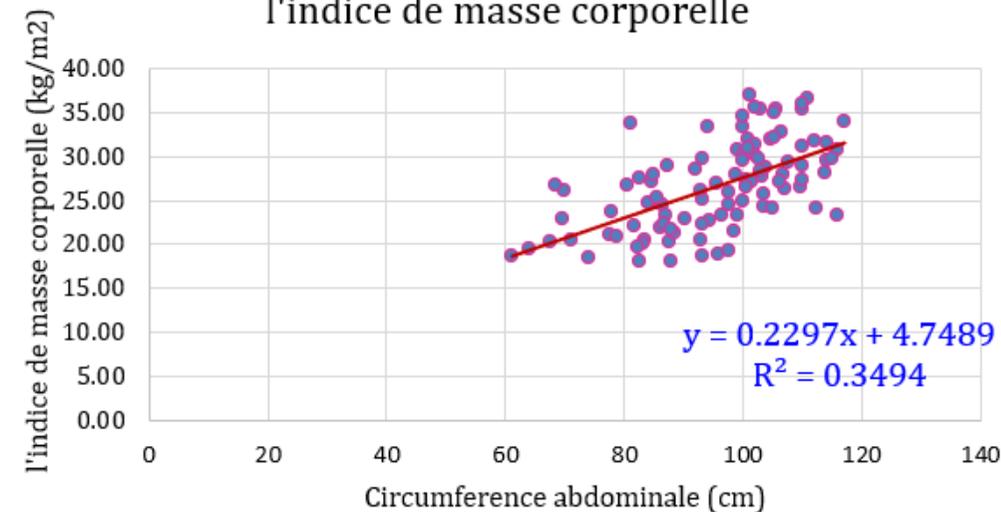
Savoir l'analyse de Corrélation et Régression Linéaires

- **Interprétation du coefficient de corrélation linéaire de Pearson:**
- Si le coefficient $r \in [-0,25; +0,25]$ → corrélation linéaire faible ou inexistante
- Si le coefficient $r \in (0,25; 0,50] \cup (-0,25; -0,50]$ → corrélation linéaire acceptable
- Si le coefficient $r \in (0,50; 0,75] \cup (-0,50; -0,75]$ → corrélation linéaire modérée à bonne
- Si le coefficient $r \in (0,75; 1] \cup (-0,75; -1]$ → forte corrélation linéaire (très bonne corrélation)

Corrélation et Régression Linéaires

La droite de régression [obtenue sur le graphique nouage des points]

Relation entre la circonférence abdominale et l'indice de masse corporelle



INTERPRÉTATION DU COEFFICIENT DE DÉTERMINATION ($d=r^2$)

- Sa valeur (en pourcentage) exprime la part de la variation de la variable dépendante (Y) qui peut être expliquée par sa relation linéaire avec la variable indépendante (X)
- Donc $r^2 = 0,35 \leftrightarrow$ dans l'échantillon d'étude 35% de la variation de l'EIM peut s'expliquer par sa relation linéaire avec la circonférence abdominale.

INTERPRÉTATION DE LA PENTE DE RÉGRESSION:

- pour chaque augmentation de la variable indépendante (X = circonférence abdominale) avec une unité de mesure (de 1 cm), la variable dépendante (Y = IMC (kg/m²)) augmente en moyenne de 0,23 unités (mm).

Savoir calculer et interpréter l'intervalle de confiance pour une moyenne

- L'INTERVALLE DE CONFIANCE à 95%:

SI $n \geq 30$:

95% IC: $[m - 1,96ES, m + 1,96ES]$ ou m = moyenne calculé sur l'échantillon; ES = l'erreur standard

SI $n < 30$:

95% IC: $[m - t_{n-1, 1-\frac{\alpha}{2}}ES, m + t_{n-1, 1-\frac{\alpha}{2}}ES]$ ou m = moyenne calculé sur l'échantillon; ES = l'erreur standard

EXCEL:

- on calcule la moyenne m et ES (par l'option Data Analysis)
- on calcule chaque extrémité de l'IC

Savoir calculer et interpréter l'intervalle de confiance pour une moyenne

- **EXCEL:**

J	K
<i>IMC(kg/m2)</i>	
Mean	25.08662531
Standard Error	0.197941026
Median	25.03243485
Mode	26.1065049
Standard Deviation	3.958820522
Sample Variance	15.67225993
Kurtosis	0.004116688
Skewness	0.12823445
Range	20.57506324
Minimum	15.57283276
Maximum	36.147896
Sum	10034.65012
Count	400
Confidence Level(95.0%)	0.389137666


 Marge d'erreur (ME) = $t_{\alpha/2, n-1} * \frac{S}{\sqrt{n}}$

M	N
Limite inferieure de l'IC:	=K3-K16
Limite superieure de l'IC:	=K3+K16

M	N
Limite inferieure de l'IC:	24.69748764
Limite superieure de l'IC:	25.47576297
95% IC	[24.70, 25.48]

Interpretation de l' IC: Nous sommes sûrs à 95% que l'intervalle de confiance [24.70, 25.48] contient la moyenne de l'IMC de la population des sujets adultes atteints d'un cancer

Savoir calculer et interpréter l'intervalle de confiance pour une différence de moyennes (échantillons appariés)

- L'INTERVALLE DE CONFIANCE à 95%:

SI $n \geq 30$:

95% IC: $[m_d - 1,96ES_d, m_d + 1,96ES_d]$ ou $m =$ la moyenne des différences calculé sur l'échantillon; ES = l'erreur standard des différences

SI $n < 30$:

95% IC: $[m_d - t_{n-1, 1-\frac{\alpha}{2}}ES_d, m_d + t_{n-1, 1-\frac{\alpha}{2}}ES_d]$ ou $m =$ la moyenne des différences calculé sur l'échantillon; ES = l'erreur standard des différences

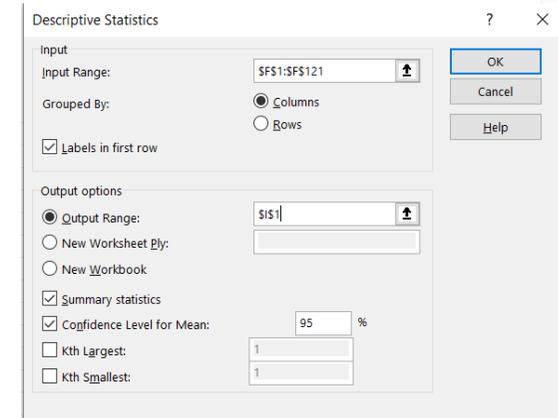
EXCEL:

- **on calcule les différences (avant-après)**
- on calcule la moyenne des différences (m_d) et l'erreur standard des différences (ES_d) par l'option Data Analysis
- on calcule chaque extrémité de l'IC

Savoir calculer et interpréter l'intervalle de confiance pour une différence de moyennes (échantillons appariés)

A	D	E
Id_patient	vitesse de sédimentation érythrocytaire (VS) à l'admission (mm)	vitesse de sédimentation érythrocytaire (VS) au moment de la sortie de l'hôpital (mm)
1	10.40	14.69
2	12.48	17.95
3	15.34	15.32
4	34.31	10.81
5	26.05	17.00
6	39.35	10.58
7	30.76	12.38

A	D	E	F
Id_patient	vitesse de sédimentation érythrocytaire (VS) à l'admission (mm)	vitesse de sédimentation érythrocytaire (VS) au moment de la sortie de l'hôpital (mm)	DifférencesVS
1	10.40	14.69	=D2-E2
2	12.48	17.95	
3	15.34	15.32	
4	34.31	10.81	
5	26.05	17.00	



I	J	K	L
DifférencesVS			
Mean	11.52359626	limite inferieure de l'IC	9.562587915
Standard Error	0.990358298	limite superieure de l'IC	13.4846046
Median	11.01469528	95% IC	[9.56, 11.48]
Mode	#N/A		
Standard Deviation	10.8488316		
Sample Variance	117.6971471		
Kurtosis	0.038631196		
Skewness	0.314473857		
Range	59.75078921		
Minimum	-14.44145604		
Maximum	45.30933316		
Sum	1382.831551		
Count	120		
Confidence Level(95.0%)	1.961008344		

Interpretation de l' IC: Nous sommes sûrs à 95% que l'intervalle de confiance [9.56, 13.48] contient la différence de moyennes de VS de la population d'intérêt → au risque 5%, il y a une différence significative entre la VS a l'admission et la moyenne de VS a la sortie de l'hôpital.

Savoir calculer et interpréter l'intervalle de confiance pour une fréquence

L'INTERVALLE DE CONFIANCE à 95%:

$$95\% \text{ IC: } \left[f - 1,96 \cdot \sqrt{\frac{f(1-f)}{n}}, f + 1,96 \cdot \sqrt{\frac{f(1-f)}{n}} \right]$$

Ou f = fréquence relative; n = taille de l'échantillon

EXCEL:

- on calcule les fréquences absolues (par l'option Pivot Table) pour la variable d'intérêt (ex. Tabagisme)

Row Labels	Count of Tabagisme
fumeur	332
non-fumeur	68
Grand Total	400

- on calcule f :

Row Labels	Count of Tabagisme	Frequence relative
fumeur	332	=332/400
non-fumeur	68	
Grand Total	400	

Savoir calculer et interpréter l'intervalle de confiance pour une fréquence

L'INTERVALLE DE CONFIANCE à 95%:

$$95\% \text{ IC: } \left[f - 1,96 \cdot \sqrt{\frac{f(1-f)}{n}}, f + 1,96 \cdot \sqrt{\frac{f(1-f)}{n}} \right]$$

Ou f = fréquence relative; n = taille de l'échantillon

EXCEL:

- on calcule chaque extrémité de l'IC:

f	0.83
limite inferieure de l'IC	=F28-1.96*sqrt(F28*(1-F28)/400)
limite supeieure de l'IC	=F28+1.96*SQRT(F28*(1-F28)/400)
95% IC	[0.79, 0.87]

Choix du test statistique

DONNÉES	NOMBRE DES ÉCHANTILLONS	TESTS PARAMÉTRIQUES	COMPARE	TESTS NON PARAMÉTRIQUES	COMPARE
QUALITATIVES	2 (ou >2) indépendants		fréquences	Chi deux <20% cellules tableau théorique/attendue <5 exact de Fisher >20% cellules tableau théorique/attendue <5	fréquences
	2 dépendants (appariés)		fréquences	Mc Nemar	
QUANTITATIVES		Données normalement distribuées		Données non normalement distribuées	
	2 indépendants	Student (t) pour échantillons indépendants avec variances ÉGALES INEGALES	moyennes	Mann Whitney U (Wilcoxon somme des rangs)	~moyennes des rangs
	2 appariés (dépendants)	Student (t) pour échantillons appariés / dépendants	moyennes	Wilcoxon pour échantillons appariés (Wilcoxon rangs signées)	~moyennes des rangs

Test pour variances (test de FISHER):
 $P > 0,05 \Rightarrow$ variances égales
 $P < 0,05 \Rightarrow$ variances inégales

le texte coloré en gris = notions qui ne sont pas nécessaires à les savoir à l'épreuve pratique (uniquement pour la théorie)

Formulations des Hypotheses des tests statistiques

- **TEST T DE STUDENT POUR 2 ECHANTILLONS INDEPENDANTES**

- ◊ **Formulation de l'hypothese nulle (H0):**

Il **n'y a pas** de **différence statistiquement significative** entre les **moyennes** de l'**IMC** dans les 2 **populations** des sujets adultes (**COVID-19 positifs** et **ceux qui n'ont pas de COVID-19**)

- ◊ **Formulation de l'hypothese alternative (H1):**

Il y a une **différence statistiquement significative** entre les **moyennes** de l'**IMC** dans les 2 **populations** des sujets adultes (**COVID-19 positifs** et **ceux qui n'ont pas de COVID-19**)

Formulations des Hypotheses des tests statistiques

- **TEST T DE STUDENT POUR 2 ECHANTILLONS DEPENDANTES:**

- ◊ **Formulation de l'hypothese nulle (H0):**

Il n'y a pas de **différence statistiquement significative** entre les **moyennes** de **vitesse de sédimentation érythrocytaire (VS)** a l'admission à l'hôpital et à la sortie dans la **population** des sujets adultes atteints d'un cancer

- ◊ **Formulation de l'hypothese alternative (H1):**

Il y a une **différence statistiquement significative** entre les **moyennes** de **vitesse de sédimentation érythrocytaire (VS)** a l'admission à l'hôpital et à la sortie dans la **population** des sujets adultes atteints d'un cancer

Formulations des Hypotheses des tests statistiques

- Test Khi-deux (χ^2):

- ◇ Formulation de l'hypothese nulle (H0):

Il n'y a pas de relation/lien/association statistiquement significative entre le tabagisme et COVID-19 dans la population des sujets adultes.

Il n'y a pas de différence statistiquement significative entre les fréquences du COVID-19 dans les 2 **populations** des sujets adultes (fumeurs et non-fumeurs)

▪

La décision d'un test statistique à l'aide de la valeur du p (p-value)

- Toutes les tests statistiques offrent une valeur p
- Si $p\text{-value} < 0,05 \Rightarrow$ on rejette $H_0 \Rightarrow$ en faveur du H_1
 - Si $p\text{-value} > 0,05 \Rightarrow$ on Ne rejette pas H_0
 - **ca ne signifie pas que H_0 est vraie!**
 - **on peut pas dire qu'il n'y a pas de différence!!!**
 - on peut dire qu'on a pas pu mettre en évidence une différence (peut être la différence existe, mais notre étude n'a pas été assez puissant pour l'observer)

Ex: $p\text{-value} = 0,003 < 0,05 \Rightarrow$ on rejette $H_0 \Rightarrow$ **en faveur du H_1**

La **conclusion** statistique: : Au risqué de 5%, **il existe une différence statistiquement significative** entre ... et ... dans

$p\text{-value} = 0,87 > 0,05 \Rightarrow$ on Ne rejette pas H_0 .

La **conclusion** statistique: : **on ne peut pas dire** qu'il existe une différence statistiquement significative entre ... et ... dans

Formulation équivalente: **on n'a pas trouvé une** différence statistiquement significative entre ... et ... dans

p-value (p-valeur)

- P-value est la probabilité d'obtenir, quand H_0 est supposée vraie, un résultat pour la statistique du test égale ou plus extrême que le résultat observé.
- Est calculée par les logiciels statistiques à partir de la valeur calculée du paramètre du test (t, Z, Khi, F...) +/- nombre d.d.l
- On peut l'identifier dans des tableaux des distributions de probabilités à partir de la valeur calculée du paramètre du test +/- nombre d.d.l

Comment choisir un test statistique

La stratégie pour faire la choix du test:

- **Quel est le TYPE DES VARIABLES?**, quel est le type de la **variable d'interet?** (malade, décès)
- **COMBIEN DES ÉCHANTILLONS (GROUPS) on a identifié** par la **variable** de regroupement (facteur de risque présent/absent; traitement A/B (aspirine/placebo); groupe A/B (avec diabète/sans diabète; fumeur/non fumeur))
- **Les échantillons** (séries des données) **sont:**
 - **DÉPENDANTES/APPARIES?**
 - **Jumeaux**
 - Le même échantillon **avant** et **après** une intervention (données répétées)
 - Etude **appariée** – pour chaque cas on trouve une témoin avec le même âge, sexe, et autres variables d'intérêt.
 - On compare la partie **gauche** avec **droite** de chaque sujet
 - On **mesure** la même variable avec **deux techniques** différentes sur **le même sujet**
 - **INDÉPENDANTES?** (on exclue les situations pour les échantillons dépendants)

Comment choisir un test statistique

La stratégie pour faire la chois du test:

- **Quelle est la nature des donnes pour les données QUANTITATIVES?**
 - Distribution normale?
 - Variances égales/ inégales?

EXCEL: Tests statistiques

- **Test t de Student pour deux échantillons indépendantes:**

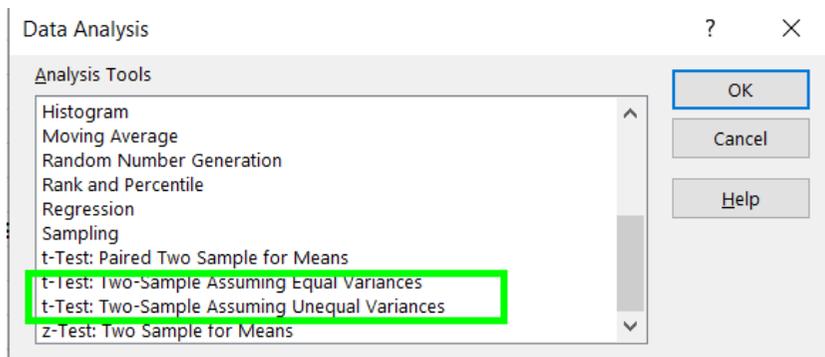
- ◊ trier les données par rapport a la variable de regroupement par l'option **Home → Sort & Filter → Custom Sort**
 - sélection de la base de données entiere
 - **Home → Sort & Filter → Custom Sort** → choisir la variable de regroupement (dans notre ex. COVID-19)

A	B	C	D	E	F	G	H	
id_pacient	Comté du patient	Tranche d'age (1: age<=44 ans; 2=45-54 ans; 3=55-74 ans; 4: age>=75 ans)	Sexe	IMC(kg/m2)	Tabagisme (1=fumeur; 0=non-fumeur)	Hypertension	COVID-19	
1	CJ						tif	
2	CJ						tif	
3	BH						tif	
4	CJ						tif	
5	BH						tif	
6	BH						tif	
7	BH						tif	
8	CJ						tif	
9	SM						tif	
10	BH						tif	
11	BH						tif	
12	CJ						tif	
13	SM						tif	
14	CJ						tif	
15	CJ		3	F	25.39	0	non	negatif
16	BH		2	M	27.48	0	oui	positif

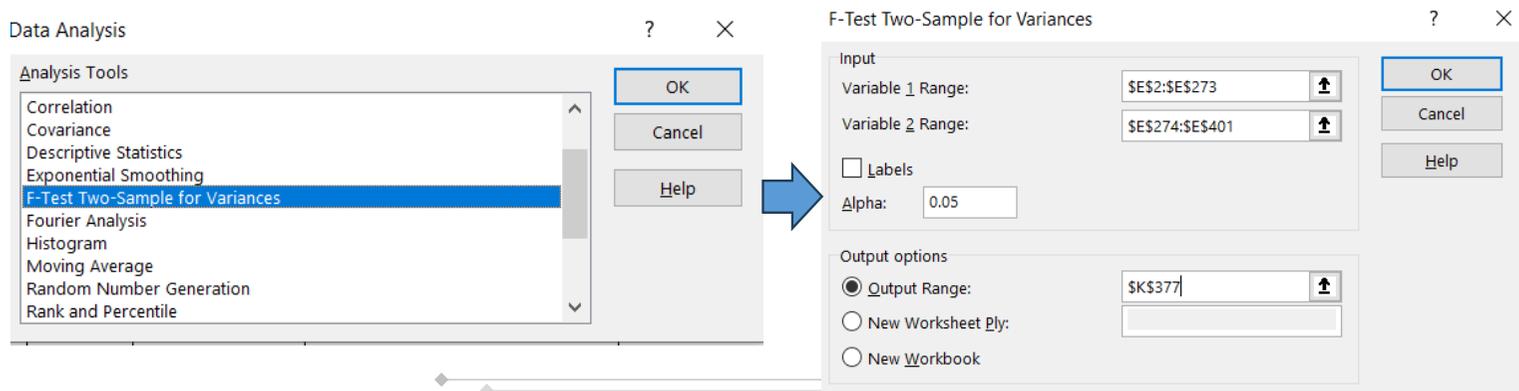
- ◊ evaluer si la variable quantitative d'intérêt (dans notre cas, l'IMC) suit la loi normale sur les deux groupes (calcul des statistiques descriptives/histogramme)

EXCEL: Tests statistiques

- utiliser l'option **Data Analysis**:



- tester l'égalité des variances à l'aide du test de Fisher



Note: il n' était pas nécessaire de cocher l'option *Labels* car nous n'avons pas sélectionné l'en-tête de colonne

EXCEL: Tests statistiques

	K	L	M
F-Test Two-Sample for Variances			
		<i>Variable 1</i>	<i>Variable 2</i>
Mean		24.85582	25.5771
Variance		15.5729	15.65114
Observations		272	128
df		271	127
F		0.995001	
P(F<=f) one-tail		0.479553	
F Critical one-tail		0.783241	

t-Test: Two-Sample Assuming Equal Variances

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

Labels

Alpha:

Output options

Output Range:

New Worksheet Ply:

New Workbook

OK Cancel Help

	t-Test: Two-Sample Assuming Equal Variances	
	<i>Variable 1</i>	<i>Variable 2</i>
Mean	24.85581511	25.5771
Variance	15.57289716	15.65114
Observations	272	128
Pooled Variance	15.59786273	
Hypothesized Mean	0	
df	398	
t Stat	-1.703852153	
P(T<=t) one-tail	0.04459463	
t Critical one-tail	1.648691174	
P(T<=t) two-tail	0.089189259	
t Critical two-tail	1.965942324	

p-valeur **du test F** $> 0,05 \Rightarrow$ NE rejette pas $H_0 \Rightarrow$ **on n'a pas trouvé une différence statistiquement** significative entre les **variances de l'IMC** dans les 2 **populations** \Rightarrow pour tester la différence entre les moyennes on va utiliser t-test avec **les variances égales**

SI p-valeur $< 0,05 \Rightarrow$ on rejette $H_0 \Rightarrow$ **il y a une différence statistiquement** significative entre les **variances de l'IMC** dans les 2 **populations** \Rightarrow pour tester la différence entre les moyennes on va utiliser t-test avec **les variances inégales**

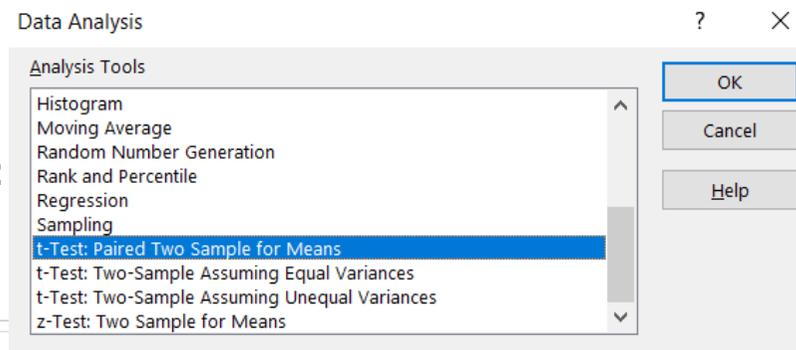
EXCEL: Tests statistiques

- **Test t de Student pour deux échantillons dépendantes:**

- ◊ évaluer si la variable quantitative d'intérêt (dans notre cas, VS) mesuré avant et après, suit la loi normale (calcul des statistiques descriptives/histogramme)

D	E
vitesse de sédimentation érythrocytaire (VS) à l'admission (mm)	vitesse de sédimentation érythrocytaire (VS) au moment de la sortie de l'hôpital (mm)
10.40	14.69
12.48	17.95
15.34	15.32
34.31	10.81
26.05	17.00
39.35	10.58
30.76	12.38
24.87	8.74
21.80	15.86

- ◊ utiliser l'option **Data Analysis:**



EXCEL: Tests statistiques

• Test t de Student pour deux échantillons dépendantes:

t-Test: Paired Two Sample for Means

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

Labels

Alpha:

Output options

Output Range:

New Worksheet Ply:

New Workbook

OK
Cancel
Help

t-Test: Paired Two Sample for Means

	<i>vitesse de sédimentation érythrocytaire (VS) à l'admission (mm)</i>	<i>vitesse de sédimentation érythrocytaire (VS) au moment de la sortie de l'hôpital (mm)</i>
Mean	23.65964918	12.13605292
Variance	80.18755137	29.71084937
Observations	120	120
Pearson Correlation	-0.079888412	
Hypothesized Mean Difference	0	
df	119	
t Stat	11.63578503	
P(T<=t) one-tail	1.15803E-21	
t Critical one-tail	1.657759285	
P(T<=t) two-tail	2.31606E-21	
t Critical two-tail	1.980099876	

EXCEL: Tests statistiques

- **Test Khi-deux (χ^2):**

- ◇ faire le **tableau de contingence observé** par l'option **PIVOT TABLE**
- ◇ faire le **tableau de contingence théorique**
- ◇ calculer la **p-Valeur** à l'aide de la fonction

- **Exemple: tester l'association entre le tabagisme et COVID-19**

Tabagisme	COVID-19
fumeur	negatif
non-fumeur	negatif
fumeur	negatif

PivotTable from table or range

Select a table or range

Table/Range: Sheet5!\$A\$1:\$B\$401

Choose where you want the PivotTable to be placed

New Worksheet

Existing Worksheet

Location:

Choose whether you want to analyze multiple tables

Add this data to the Data Model

OK Cancel

EXCEL: Tests statistiques

- Test Khi-deux (χ^2):

◊ le tableau de contingence observé par l'option PIVOT TABLE

Count of Tabagisme	Column Labels	negatif	positif	Grand Total
fumeur		228	104	332
non-fumeur		44	24	68
Grand Total		272	128	400



Count of Tabagisme	Column Labels	positif	negatif	Grand Total
fumeur		104	228	332
non-fumeur		24	44	68
Total		128	272	400

◊ le tableau de contingence théorique

	positif	negatif	Total
fumeur	=F10*H8/H10		332
non-fumeur			68
Total	128	272	400



	positif	negatif	Total
fumeur	106.24	225.76	332
non-fumeur	21.76	46.24	68
Total	128	272	400

EXCEL: Tests statistiques

- **Test Khi-deux (χ^2):**

- ◊ la **p-Valeur** a l'aide de la fonction la fonction **CHISQ.TEST (ou CHITEST)**

Function Arguments

CHISQ.TEST

Actual_range F3:G4 = {104,228;24,44}

Expected_range F8:G9 = {106.24,225.76;21.76,46.24}

= 0.522704029

Returns the test for independence: the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

Expected_range is the range of data that contains the ratio of the product of row totals and column totals to the grand total.

Formula result = 0.522704029

[Help on this function](#) OK Cancel

Travail sur EXCEL

- Fusionner des multiples cellules dans une seule cellule
- Formatage du texte dans une cellule
- Ecrire de formules données dans Excel en utilisant des opérateurs mathématiques (incluant la puissance, racine carrée), l'ordre des opérations.
- utiliser des références relative (A2) et absolues (ex. A\$2, \$A2, \$A\$2 - pour fixer les cellules sur les colonnes, les lignes, et les lignes et colonnes).
- Autres Fonctions prédéfinies: **IF (si) , COUNT, COUNTIF, NORM.DIST**
- Multiplier les formules

Travail sur le system d'opération [WINDOWS]

- La création des dossiers avec le nom indiqué, dans la place indiquée
- L'enregistrement des fichiers avec le nom indiqué, dans le dossier indiqué
- Copier coller des objets (textes, tableaux, graphiques) entre des différents applications (WORD, PowerPoint)
- faire le formatage des documents WORD, PowerPoint

Merci pour votre attention!

