**Exemples des exercices pour l’examen pratique**

***Daniel Leucuța, Mihaela Iancu***

**Objectives**

* Révision des toutes les activités des travaux pratiques

**Ce n’est pas obligatoire de faire ces exercices, mais nous vous conseillons de faire tous les exercices pour vous a maison pouvez observer les types des exercices que vous pouvez avoir pendant l’examen pratique**.

**Scénario:**

Dans le service de médecine interne d’un l'hôpital, la situation des patients a été vérifiée et les données ont été collectées dans le fichier Base\_de\_donnees-Exemples.xlsx.

**Demandes:**

1. **Le travail avec le system d’opération**
2. **Eliminatoire:** Créez **dans votre compte** (la partition disque avec votre nom d’utilisateur) un **dossier** avec le nom ExemplesInfo.
3. **Eliminatoire :** **Sauvegardez** le fichier Excel (voir en haut), dans le dossier crée, et **renommez** le ExemplesExcel
4. **Eliminatoire:** dans le dossier crée réalisez un fichier Word appelée ExemplesWord

Pour chaque point vous devez copier les résultats dans Word !

Écrivez le nombre de la question avant chaque réponse copiée dans Word !

1. **Microsoft Word ֍**
2. Insérez et formatez le tableau suivant.Ecrivez le texte avant et après le tableau avec formatage (le tableau, la fusion des cellules, le formatage en gras et italique, le symbole, les accents, le subscript (Q1‒Q3), et le superscript n**0** [voir le fichier des conseils pour Microsoft Word]

--------------------------

# Résultats

***Tableau 1***. Les caractéristiques des patients en fonction du traitement

|  |  |  |  |
| --- | --- | --- | --- |
|  | Group | | Valeur P |
| Caractéristiques | Traitement | Contrôle |  |
| L’adiponectine (**µ**g/dL) – moyenne (DS) | 132 (24) | 176(32) | 0,005 |
| sCTX - médiane (Q1‒Q3) | 255 (172–287) | 170 (143–202) | 0,86 |
| Les obeses n**0** (%) | 26 (26) | 74 (74) | 0,06 |

# Discussion

L’adiponectine moyenne trouvée par autres chercheurs a été de 253 **µ**g/dL [Collet B. 2015]

------------------------

1. Formatez le texte avant avec la police (font) Garamond, couleur verte, dimension 14, alignement simultané à gauche et à droite (justified), espace entre les lignes 1,5, espace avant le paragraphe de 6 points. Modifiez les bordures du page pour avoir 2 cm ou 0,8 " inch (pouces). Changez le format de la page dans - A4 [voir le fichier des conseils pour Microsoft Word]
2. Insérez le numéro de page [voir le fichier des conseils pour Microsoft Word]
3. Créez une table de matière pour les titres Résultats et Discussion [voir le fichier des conseils pour Microsoft Word]
4. Editez le texte Collet B. 2015 en haut comme une référence. [voir le fichier des conseils pour Microsoft Word]
5. Mettez dans le Header de la page le texte suivant : L’adiponectine et l’obésité[voir le fichier des conseils pour Microsoft Word]
6. **Eliminatoire:** Ecrivez l’hypothèse nulle, l’hypothèse alternative, et l’interprétation du p, du test statistique présentée dans le tableau 1, en ce qui concerne l’adiponectine, sachant que les données sont normalement distribuées, et les variances égales. Ecrivez le nom du test statistique a été plus probablement utilise. [voir le fichier des conseils pour Microsoft Excel] [Ici il y a une demande pour un **test statistique** depuis la demande, mais aussi dans le tableau on observe la valeur p qui indique qu’il y a les résultats d’un test statistiques (en effet ici il y a trois tests statistiques, parmi lesquels il y a un test pour l’adiponectine). Pour la chois d’un test statistique on doit **vérifier** dans l‘énonce **s’il s'agit d’une** **comparaison**. Ici on observe dans le tableau deux colonnes qui indique des groupes: ceux qui ont reçu un traitement et le groupe de contrôle. **Si on observe une comparaison** des groupes (n’importe leur nombre**) on doit se poser la question si les observations sont dépendantes (appariées) ou non**. Ici on observe qu’ils comparent ceux qui ont reçu un traitement avec le groupe de contrôle. On cherche donc dans l’ énonce les 5 situation types des observations dépendantes: mesures répétées sur les mêmes sujets (ex. on mesure la température d’un patient avant et après la prise d’un médicament); comparaison des jumeaux identiques; comparaison des échantillons appariées (pour chaque sujet dans un group on trouve un sujet dans l’autre groupe avec caractéristiques similaires); comparaison des plusieurs observations faites sur certaines parties du corps du même individu, qui conceptuellement ne sont pas différentes – ex: partie gauche avec la partie droite (ex. un test dermatologique avec une traitement pour diminuer la perte de cheveux); comparaison des deux (ou plusieurs) méthodes pour mesurer une caractéristique (ou deux techniques diagnostiques [ex tension artérielle]) sur les mêmes sujets. Ici on observe qu’ils comparent ceux qui ont reçu un traitement avec le groupe de contrôle, donc aucune des 5 situations types pour des échantillons dépendantes => il s’agit des observations indépendantes. Pour la chois d’un test statistique, s’il s’agit d’une comparaison, on se pose la question, **qu’est qu’il comparent (quelle est le type de la variable/ caractéristique d’intérêt comparée)?** Le test concerne une variable quantitative (parce que l’adiponectine a des valeurs numériques, mais si on ne sait pas ça, on observe que c’est écrit que la variable a une distribution normale, donc cette affirmation on peut le faire pour les variables quantitatives). D’une manière implicite, quand on compare des caractéristiques quantitatives, notre but est de comparer leur centre (la moyenne – si les données sont normale distribuées, et parfois les médianes quand les données ne sont pas normale distribuées). Pour la chois d’un test statistique, si la variable d’intérêt est quantitative, on doit se poser la question: **est que cette variable quantitative est normale distribuée ou non?** On observe dans le scenario/question que les données **suivent une distribution normale**. Pour la chois d’un test statistique, si la variable d’intérêt est quantitative, normale distribuées, et les échantillons sont indépendants, **c’est important de savoir si les variances sont égales** ou non. Ici on est informée que les variances sont égales. On tenant compte de toutes ces informations (comparaison des deux échantillons indépendants concernant la moyenne d’une variable quantitative, normale distribuée, avec variances égales) on peut déduire => le nom du test est : le test t (Student) pour échantillons indépendants avec variances égales. L’hypothèse nulle : il n’y a pas de différence statistiquement significative entre les sujets qui ont reçu un traitement et les sujets qui ont reçu l’intervention de contrôle concernant la moyenne de l’adiponectine. L’hypothèse alternative : il y a une différence statistiquement significative entre les sujets qui ont reçu un traitement et les sujets qui ont reçu l’intervention de contrôle concernant la moyenne de l’adiponectine. Interprétation du p : parce que p < 0.05 rejeté l’hypothèse nulle, et on accepte l’hypothèse alternative, donc il y a une différence statistiquement significative entre les sujets qui ont reçu un traitement et les sujets qui ont reçu l’intervention de contrôle concernant la moyenne de l’adiponectine.
7. **Eliminatoire:** Ecrivez l’hypothèse nulle, l’hypothèse alternative, et l’interprétation du p, du test statistique présentée dans le tableau 1, en ce qui concerne l’obésité. Ecrivez le nom du test statistique a été plus probablement utilise. [voir le fichier des conseils pour Microsoft Excel] [Ici il y a une demande pour un **test statistique** depuis la demande, mais aussi dans le tableau on observe la valeur p qui indique qu’il y a les résultats d’un test statistiques (en effet ici il y a trois tests statistiques, parmi lesquels il y a un test pour l’obésité). Pour la chois d’un test statistique on doit **vérifier** dans l‘énonce **s’il s'agit d’une** **comparaison**. Ici on observe dans le tableau deux colonnes qui indique des groupes: ceux qui ont reçu un traitement et le groupe de contrôle. **Si on observe une comparaison** des groupes (n’importe leur nombre**) on doit se poser la question si les observations sont dépendantes (appariées) ou non**. Ici on observe qu’ils comparent ceux qui ont reçu un traitement avec le groupe de contrôle. On cherche donc dans l’ énonce les 5 situation types des observations dépendantes: mesures répétées sur les mêmes sujets (ex. on mesure la température d’un patient avant et après la prise d’un médicament); comparaison des jumeaux identiques; comparaison des échantillons appariées (pour chaque sujet dans un group on trouve un sujet dans l’autre groupe avec caractéristiques similaires); comparaison des plusieurs observations faites sur certaines parties du corps du même individu, qui conceptuellement ne sont pas différentes – ex: partie gauche avec la partie droite (ex. un test dermatologique avec une traitement pour diminuer la perte de cheveux); comparaison des deux (ou plusieurs) méthodes pour mesurer une caractéristique (ou deux techniques diagnostiques [ex tension artérielle]) sur les mêmes sujets. Ici on observe qu’ils comparent ceux qui ont reçu un traitement avec le groupe de contrôle, donc aucune des 5 situations types pour des échantillons dépendantes => il s’agit des observations indépendantes. Pour la chois d’un test statistique, s’il s’agit d’une comparaison, on se pose la question, **qu’est qu’il comparent (quelle est le type de la variable/ caractéristique d’intérêt comparée)?** Le test concerne une variable qualitative (parce que les valeurs de ces variables sont représentées par texte : oui/non – avec et sans obésité. Pour concision les tableaux ne montrent que la présence de l’obésité, les fréquences absolues et relatives des obeses). D’une manière implicite, quand on compare des caractéristiques qualitatives, notre but est de comparer leurs fréquences. On tenant compte de toutes ces informations (comparaison des deux échantillons indépendants concernant les fréquences d’une variable qualitative) on peut déduire => le nom du test est : soit le test Khi deux, soit le test exact Fisher. La chois entre les deux sera faite en regardant le tableau de contingence théorique. L’hypothèse nulle : il n’y a pas de différence statistiquement significative entre les sujets qui ont reçu un traitement et les sujets qui ont reçu l’intervention de contrôle concernant les fréquences d’obésité OU il n’y a pas de relation/association/lien statistiquement significative entre le traitement et l’obésité. L’hypothèse alternative : il y a une différence statistiquement significative entre les sujets qui ont reçu un traitement et les sujets qui ont reçu l’intervention de contrôle concernant les fréquences d’obésité OU il y a une relation/association/lien statistiquement significative entre le traitement et l’obésité. Interprétation du p : parce que p > 0.05 on ne peut pas rejeter l’hypothèse nulle, donc on ne peut pas dire qu’il y a une différence statistiquement significative entre les sujets qui ont reçu l’intervention de contrôle concernant les fréquences d’obésité OU on ne peut pas dire qu’il y a une relation/association/lien statistiquement significative entre le traitement et l’obésité. Attention !!! si p>0.05 s’il vous plait de ne pas énoncer des affirmations avec des erreurs comme les suivantes : « on ne peut pas dire qu’il y a une **différence** (- ici le mot différence n'est pas correct mais on peut dire relation/association…) statistiquement significative entre le traitement et l’obésité » ou de dire qu’« on rejette l’hypothèse alternative », ou de dire qu’« on accepte l’hypothèse nulle », ou on dire « qu’il n’y a pas des différences … »]]
8. **Microsoft Excel**
9. **Eliminatoire:** Représentez graphiquement la répartition par hypertension. [graphique une variable qualitative (parce que les valeurs de cette variables sont représentées par du texte : oui/non) => graphique colonne/ bar / camembert (pour le faire, sélectionnez toutes les données, réalisez un tableau de fréquences avec Insert + Insert + PivotTable, puis Insert + Charts + – Bar)]
10. **Eliminatoire:** Réalisez le tableau de contingence pour la distribution des paresthésies par hypertension. [Tableau de pour la relation entre deux variables qualitatives (parce que les valeurs de cette variables sont représentées par du texte : oui/non) => tableau de contingence (pour le faire, sélectionnez toutes les données, réalisez un tableau de contingence avec Insert + PivotTable : variable hypertension sur les lignes, variable paresthésies sur les colonnes, variable hypertension dans Values)]
11. **Eliminatoire:** Représentez graphiquement la distribution de l’hypertension par Activité physique> 30 minutes par jour. [graphique pour la relation entre deux variables qualitatives (parce que les valeurs de cette variables sont représentées par du texte : oui/non) => graphique colonne/ bar – idéale graphique des colonnes empilées (pour le faire, sélectionnez toutes les données, réalisez un tableau de contingence avec Insert + PivotTable: variable hypertension sur les lignes, variable Activité physique> 30 minutes par jour sur les colonnes, variable hypertension dans Values, puis Insert + Charts + Column puis 100% stacked column)]
12. **Eliminatoire:** Créez la table de fréquence pour la variable Hypertension. (avec COUNTIF) [voir le fichier des conseils pour Microsoft Excel]
13. **Eliminatoire:** Faites un graphique pour la variable Glucose initiale. Interprétez la normalité. [voir le fichier des conseils pour Microsoft Excel] [graphique une variable quantitative (parce que les valeurs de cette variables sont représentées par des valeurs numériques : 160, 176 …) => graphique histogramme (pour le faire, sélectionnez les valeurs de la variable, avec le titre, insérez le graphique avec Insert + Charts + Histogram. Une autre possibilité est d’utiliser la fonction Histogram du Data Analysis (avec l’option Chart Output).). L’histogramme ne semble pas symétrique, donc il y a une suggestion que la variable glucose initiale ne semble pas suivre une distribution normale]
14. **Eliminatoire:** Calculez pour la variable Poids avec des fonctions les statistiques suivantes: moyenne arithmétique, médiane, mode, variance, maximum, minimum, l’amplitude, écart type, coefficient d’asymétrie, coefficient d’aplatissement, l’erreur standard, le coefficient de variation, le quartile 1 et 3. [voir le fichier des conseils pour Microsoft Excel]

Moyenne: La valeur au milieu des données si la distribution des données est normale est 75.34 4

Médiane: 50% des valeurs sont ≤ 77, 50% sont ≥77. L'observation qui divise la distribution sur deux partis égaux est 77.00

Mode: La valeur de l'observation associée à la fréquence la plus élevée est 84.00

Variance: La racine carrée de la déviation standard est 117.48

Maximum: La plus grande valeur de la série des données est 96.00

Minimum: La plus petite valeur de la série des données est 54.00

Amplitude: La différence entre la plus grande et la plus petite valeur dans la série des données est 42.00

Ecart type: Est la racine carrée de la variance. Montre la dispersion des données autour de la Moyenne. La racine carrée de la moyenne des carrées des écarts vers la moyenne est: 10.84

Déviation standard: Identique a l’écart type, voir en haut 10.84

Coefficient asymétrie: *il est proche a 0 (*appartient à l'intervalle **[-1, 1]**) dont il y a une distribution symétrique ; on observe qu’il est négative, donc la distribution est un peu étalée vers la vers la gauche -0.11

Coefficient aplatissement: *il est proche a 0 (*appartient à l'intervalle **[-1, 1]**) dont il y a une distribution mesokurtique ; on observe qu’il est négative, donc il est un peu aplati -0.91

Coefficient variation: 0.14 est comprise entre **≥10% - <20%** la série des donnes peut être considérée comme relativement homogène

Quartile 1: 25% des valeurs sont ≤ 66, 75% sont ≥ 66.00

Quartile 3: 75% des valeurs sont ≤ 84, 25% sont ≥ 84.00

L'erreur standard Est utilisée dans le calcul de l'intervalle de confiance de 95% autour de la moyenne (pour une distribution normale) - 1.22

1. **Eliminatoire:** **֍** Vérifiez si la glycémie finale est statistiquement significatif diffèrent de la glycémie initiale pour les sujets qui ont eu activité physique, sachant que les données sont normalement distribuées. Ecrivez le nom du test statistique, la valeur du p, la valeur critique du test, la région du rejet, la valeur du paramètre du test statistique. **Eliminatoire:** Ecrivez l’hypothèse nulle, l’hypothèse alternative, et l’interprétation du p. [voir le fichier des conseils pour Microsoft Excel] [Ici il y a une demande pour un **test statistique** (parce que il demande de vérifier la présence d’un différence statistiquement significative). Pour la chois d’un test statistique on doit **vérifier** dans l‘énonce **s’il s'agit d’une comparaison**. Ici on observe qu’ils comparent la glycémie initiale avec la glycémie finale. **Si on observe une comparaison** des groupes (n’importe leur nombre**) on doit se poser la question si les observations sont dépendantes (appariées) ou non**. Ici on observe qu’ils comparent la glycémie initiale avec la glycémie finale. On cherche donc dans l’ énonce les 5 situation types des observations dépendantes: mesures répétées sur les mêmes sujets (ex. on mesure la température d’un patient avant et après la prise d’un médicament); comparaison des jumeaux identiques; comparaison des échantillons appariées (pour chaque sujet dans un group on trouve un sujet dans l’autre groupe avec caractéristiques similaires); comparaison des plusieurs observations faites sur certaines parties du corps du même individu, qui conceptuellement ne sont pas différentes – ex: partie gauche avec la partie droite (ex. un test dermatologique avec une traitement pour diminuer la perte de cheveux); comparaison des deux (ou plusieurs) méthodes pour mesurer une caractéristique (ou deux techniques diagnostiques [ex tension artérielle]) sur les mêmes sujets. Ici ils comparent la glycémie initiale avec la glycémie finale, donc il s’agit des observations dépendantes (appariées). Pour la chois d’un test statistique, s’il s’agit d’une comparaison, on se pose la question, **qu’est qu’il comparent (quelle est le type de la variable/ caractéristique d’intérêt comparée)?** Le test concerne une variable quantitative (parce que les valeurs de ces variables sont représentées par des valeurs numériques : 160, 176 …). D’une manière implicite, quand on compare des caractéristiques quantitatives, notre but est de comparer leur centre (la moyenne – si les données sont normale distribuées). Pour la chois d’un test statistique, si la variable d’intérêt est quantitative, on doit se poser la question: **est que cette variable quantitative est normale distribuée ou non?** On observe dans le scenario/question que les données **suivent une distribution normale** (si ce n’était pas le cas on doit le vérifier, avec les coefficients d’asymétrie et d’aplatissement). On tenant compte de toutes ces informations (comparaison des deux séries d’observations dépendants concernant la moyenne d’une variable quantitative, normale distribuée) on peut déduire => le nom du test est : le test t (Student) pour échantillons dépendants (pour le faire on doit trier les donnes en fonction de la variable Activité physique afin de pouvoir sélectionner des données concernant des sujets qui n’ont pas eu activité physique, puis dans l’onglet Data, section Analysis – outil Data Analysis, on fait clic sur t-Test Paired Two Sample For Means. La on doit choisir dans les champs Variable 1 Range et Variable 2 Range les valeurs de la Glycémie initiale et finale qui correspond aux sujets avec Activité physique = non: $G$2:$G$42 $H$2:$H$42. Dans la cellule P(T<=t) two-tail c’est la valeur du p du test=1.85E-55 (donc 1.85 \* 10-52), dans la cellule t Critical two-tail c’est la valeur critique du test=2.02 => donc la région du rejet est (-infini, -2,02] U [2.02, +infini), dans la cellule t Stat c’est la valeur du paramètre du test statistique=117.8). L’hypothèse nulle : il n’y a pas de différence statistiquement significative entre la moyenne de la glycémie initiale et celle finale pour les sujets qui ont eu activité physique. L’hypothèse alternative : il y a une différence statistiquement significative entre la moyenne de la glycémie initiale et celle finale pour les sujets qui ont eu activité physique. Interprétation du p : parce que p < 0.05 on rejet l’hypothèse nulle, et on accepte l’hypothèse alternative, donc il y a une différence statistiquement significative entre la moyenne de la glycémie initiale et celle finale pour les sujets qui ont eu activité physique]
2. **Microsoft PowerPoint**
3. Créez une présentation avec le titre Exemples examen et votre nom
4. Changer le design des diapositives
5. Insérer le numéro de la diapositive
6. Insérer la date sur chaque diapositive
7. Copiez un graphique et un tableau que vous avez réalisé avant.
8. **Microsoft Excel - continuation**
9. **Eliminatoire:** **֍** Vérifiez si la glycémie finale est statistiquement significatif diffèrent entre ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique, sachant que les données sont normalement distribuées, et que les variances sont inégales. Ecrivez le nom du test statistique, la valeur du p, la valeur critique du test, la région du rejet, la valeur du paramètre du test statistique. Interprétez le résultat du test statistique en utilisant la région du rejet. **Eliminatoire:** Ecrivez l’hypothèse nulle, l’hypothèse alternative, et l’interprétation du p. [voir le fichier des conseils pour Microsoft Excel] [Ici il y a une demande pour un **test statistique** (parce que il demande de vérifier la présence d’un différence statistiquement significative). Pour la chois d’un test statistique on doit **vérifier** dans l‘énonce **s’il s'agit d’une** **comparaison**. Ici on observe qu’ils comparent ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique. **Si on observe une comparaison** des groupes (n’importe leur nombre**) on doit se poser la question si les observations sont dépendantes (appariées) ou non**. Ici on observe qu’ils comparent ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique. On cherche donc dans l’ énonce les 5 situation types des observations dépendantes: mesures répétées sur les mêmes sujets (ex. on mesure la température d’un patient avant et après la prise d’un médicament); comparaison des jumeaux identiques; comparaison des échantillons appariées (pour chaque sujet dans un group on trouve un sujet dans l’autre groupe avec caractéristiques similaires); comparaison des plusieurs observations faites sur certaines parties du corps du même individu, qui conceptuellement ne sont pas différentes – ex: partie gauche avec la partie droite (ex. un test dermatologique avec une traitement pour diminuer la perte de cheveux); comparaison des deux (ou plusieurs) méthodes pour mesurer une caractéristique (ou deux techniques diagnostiques [ex tension artérielle]) sur les mêmes sujets. Ici ils comparent ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique, donc aucune des 5 situations types pour des échantillons dépendantes => il s’agit des observations indépendantes. Pour la chois d’un test statistique, s’il s’agit d’une comparaison, on se pose la question, **qu’est qu’il comparent (quelle est le type de la variable/ caractéristique d’intérêt comparée)?** Le test concerne une variable quantitative (parce que les valeurs de ces variables sont représentées par des valeurs numériques : 160, 176 …). D’une manière implicite, quand on compare des caractéristiques quantitatives, notre but est de comparer leur centre (la moyenne – si les données sont normale distribuées). Pour la chois d’un test statistique, si la variable d’intérêt est quantitative, on doit se poser la question: **est que cette variable quantitative est normale distribuée ou non?** On observe dans le scenario/question que les données **suivent une distribution normale** (si ce n’était pas le cas on doit le vérifier, avec les coefficients d’asymétrie et d’aplatissement). Pour la chois d’un test statistique, si la variable d’intérêt est quantitative, normale distribuées, et les échantillons sont indépendants, **c’est important de savoir si les variances sont égales** ou non. Ici on est informée que les variances sont inégales (si ce n’était pas le cas on doit le vérifier avec un test F pour les variances). On tenant compte de toutes ces informations (comparaison des deux échantillons indépendants concernant la moyenne d’une variable quantitative, normale distribuée, avec variances inégales) on peut déduire => le nom du test est : le test t (Student) pour échantillons indépendants avec variances inégales (pour le faire on doit trier les donnes en fonction de la variable Activité physique afin de pouvoir sélectionner des données concernant des sujets qui ont eu et qui n’ont pas eu activité physique, puis dans l’onglet Data, section Analysis – outil Data Analysis, on fait clic sur t-Test: Two-Sample Assuming Unequal Variances. La on doit choisir dans les champs Variable 1 Range et Variable 2 Range les valeurs de la Glycémie finale qui correspond aux sujets avec Activité physique = oui et non: $H$2:$H$42 $H$43:$H$80. Dans la cellule P(T<=t) two-tail c’est la valeur du p du test=4.53E-18 (donc 4.53 \* 10-18), dans la cellule t Critical two-tail c’est la valeur critique du test=1.99 => donc la région du rejet est (-infini, -1.99] U [1.99, +infini), dans la cellule t Stat c’est la valeur du paramètre du test statistique=11.37). L’hypothèse nulle : il n’y a pas de différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la moyenne de la glycémie finale. L’hypothèse alternative : il y a une différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la moyenne de la glycémie finale. Interprétez le résultat du test statistique en utilisant la région du rejet : parce que la valeur du paramètre du test statistique se trouve à l’intérieur de la région du rejet on rejet l’hypothèse nulle, et on accepte l’hypothèse alternative, donc il y a une différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la moyenne de la glycémie finale. Interprétation du p : parce que p < 0.05 on rejet l’hypothèse nulle, et on accepte l’hypothèse alternative, donc il y a une différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la moyenne de la glycémie finale]
10. **Eliminatoire:** **֍** Vérifiez si la glycémie initiale est statistiquement significatif diffèrent entre ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique, sachant que les données sont normalement distribuées. Ecrivez le nom du test statistique, la valeur du p, la valeur critique du test, la région du rejet, la valeur du paramètre du test statistique. **Eliminatoire:** Ecrivez l’hypothèse nulle, l’hypothèse alternative, et l’interprétation du p.

[voir le fichier des conseils pour Microsoft Excel] [Ici il y a une demande pour un **test statistique** (parce que il demande de vérifier la présence d’un différence statistiquement significative). Pour la chois d’un test statistique on doit **vérifier** dans l‘énonce **s’il s'agit d’une** **comparaison**. Ici on observe qu’ils comparent ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique. **Si on observe une comparaison** des groupes (n’importe leur nombre**) on doit se poser la question si les observations sont dépendantes (appariées) ou non**. Ici on observe qu’ils comparent ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique. On cherche donc dans l’ énonce les 5 situation types des observations dépendantes: mesures répétées sur les mêmes sujets (ex. on mesure la température d’un patient avant et après la prise d’un médicament); comparaison des jumeaux identiques; comparaison des échantillons appariées (pour chaque sujet dans un group on trouve un sujet dans l’autre groupe avec caractéristiques similaires); comparaison des plusieurs observations faites sur certaines parties du corps du même individu, qui conceptuellement ne sont pas différentes – ex: partie gauche avec la partie droite (ex. un test dermatologique avec une traitement pour diminuer la perte de cheveux); comparaison des deux (ou plusieurs) méthodes pour mesurer une caractéristique (ou deux techniques diagnostiques [ex tension artérielle]) sur les mêmes sujets. Ici ils comparent ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique, donc aucune des 5 situations types pour des échantillons dépendantes => il s’agit des observations indépendantes. Pour la chois d’un test statistique, s’il s’agit d’une comparaison, on se pose la question, **qu’est qu’il comparent (quelle est le type de la variable/ caractéristique d’intérêt comparée)?** Le test concerne une variable quantitative (parce que les valeurs de ces variables sont représentées par des valeurs numériques : 160, 176 …). D’une manière implicite, quand on compare des caractéristiques quantitatives, notre but est de comparer leur centre (la moyenne – si les données sont normale distribuées). Pour la chois d’un test statistique, si la variable d’intérêt est quantitative, on doit se poser la question: **est que cette variable quantitative est normale distribuée ou non?** On observe dans le scenario/question que les données **suivent une distribution normale** (si ce n’était pas le cas on doit le vérifier, avec les coefficients d’asymétrie et d’aplatissement). Pour la chois d’un test statistique, si la variable d’intérêt est quantitative, normale distribuées, et les échantillons sont indépendants, **c’est important de savoir si les variances sont égales** ou non. Ici on n’est pas informée si les variances sont égales ou inégales, donc on doit le vérifier avec un test F pour les variances (pour le faire on doit trier les donnes en fonction de la variable Activité physique afin de pouvoir sélectionner des données concernant des sujets qui ont eu et qui n’ont pas eu activité physique, puis dans l’onglet Data, section Analysis – outil Data Analysis, on fait clic sur F-Test Two-Sample for Variances. La on doit choisir dans les champs Variable 1 Range et Variable 2 Range les valeurs de la Glycémie initiale qui correspond aux sujets avec Activité physique = oui et non : $G$2:$G$42 $G$43:$G$80. Dans la cellule P(F<=f) one-tail, on trouve la valeur p du test F pour les variances =0.11, qui est plus grande que 0.05, donc on ne peut pas rejeter l’hypothese nulle du test F pour les variances, donc on ne peut pas dire qu’il y a une différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la variance de la glycémie => on va choisir un test supposant des variances égales). On tenant compte de toutes ces informations (comparaison des deux échantillons indépendants concernant la moyenne d’une variable quantitative, normale distribuée, avec variances égales) on peut déduire => le nom du test est : le test t (Student) pour échantillons indépendants avec variances égales (pour le faire on doit trier les donnes en fonction de la variable Activité physique afin de pouvoir sélectionner des données concernant des sujets qui ont eu et qui n’ont pas eu activité physique, puis dans l’onglet Data, section Analysis – outil Data Analysis, on fait clic sur t-Test: Two-Sample Assuming Equal Variances. La on doit choisir dans les champs Variable 1 Range et Variable 2 Range les valeurs de la Glycémie finale qui correspond aux sujets avec Activité physique = oui et non: $G$2:$G$42 $G$43:$G$80. Dans la cellule P(T<=t) two-tail c’est la valeur du p du test=0.16, dans la cellule t Critical two-tail c’est la valeur critique du test=1.99 => donc la région du rejet est (-infini, -1.99] U [1.99, +infini), dans la cellule t Stat c’est la valeur du paramètre du test statistique=1.41). L’hypothèse nulle : il n’y a pas de différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la moyenne de la glycémie initiale. L’hypothèse alternative : il y a une différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la moyenne de la glycémie initiale. Interprétation du p : parce que p > 0.05 on ne peut pas rejeter l’hypothèse nulle, donc on ne peut pas dire qu’il y a une différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la moyenne de la glycémie initiale. Attention !!! si p>0.05 s’il vous plait de ne pas énoncer des affirmations avec des erreurs comme les suivantes : « on rejette l’hypothèse alternative », ou de dire qu’« on accepte l’hypothèse nulle », ou on dire « qu’il n’y a pas des différences … »]]

1. **Eliminatoire:** **֍** Vérifiez si la variance de la glycémie initiale est statistiquement significatif diffèrent entre ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique, sachant que les données sont normalement distribuées. Ecrivez le nom du test statistique, la valeur du p. **Eliminatoire:** Ecrivez l’hypothèse nulle, l’hypothèse alternative, et l’interprétation du p. [voir le fichier des conseils pour Microsoft Excel] [Ici il s’agit de comparer deux échantillons indépendants concernant des variables quantitatives normale distribuées, pour voir si les variances sont égales ou non => on doit le vérifier avec un test F pour les variances (pour le faire on doit trier les donnes en fonction de la variable Activité physique afin de pouvoir sélectionner des données concernant des sujets qui ont eu et qui n’ont pas eu activité physique, puis dans l’onglet Data, section Analysis – outil Data Analysis, on fait clic sur F-Test Two-Sample for Variances. Là on doit choisir dans les champs Variable 1 Range et Variable 2 Range les valeurs de la Glycémie initiale qui correspond aux sujets avec Activité physique = oui et non : $G$2:$G$42 $G$43:$G$80. Dans la cellule P(F<=f) one-tail, on trouve la valeur p du test F pour les variances =0.11. L’hypothèse nulle : il n’y a pas de différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la variance de la glycémie initiale. L’hypothèse alternative : il y a une différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la variance de la glycémie initiale. Interprétation du p : parce que p est plus grande que 0.05, donc on ne peut pas rejeter l’hypothèse nulle du test F pour les variances, donc on ne peut pas dire qu’il y a une différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant la variance de la glycémie]
2. **Eliminatoire:** **֍** Vérifiez s’il y a une relation statistiquement significative entre l’activité physique et les paresthésies. Calculez la valeur du p du test statistique et le paramètre du test statistique. Ecrivez le nom du test statistique, la valeur du p, la région du rejet, sachant que la valeur critique du test est 3,84. Interprétez le résultat du test statistique en utilisant la région du rejet. **Eliminatoire:** Ecrivez l’hypothèse nulle, l’hypothèse alternative, et l’interprétation du p. [voir le fichier des conseils pour Microsoft Excel] [ Ici il y a une demande pour un **test statistique** (parce que il demande de vérifier la relation statistiquement significative). Pour la chois d’un test statistique on doit **vérifier** dans l‘énonce **s’il s'agit d’une** **comparaison**. Ici on observe qu’ils demande la relation entre deux variables, et les deux variables sont qualitatives dichotomiques, donc on peut déduire qu’on peut regarder la situation comme une comparaison entre les ceux sans activité physique et ceux avec activité physique (on considère ici la variable activité physique comme une variable de regroupement), et la variable paresthésies est la variable comparée. **Si on observe une comparaison** des groupes (n’importe leur nombre**) on doit se poser la question si les observations sont dépendantes (appariées) ou non**. Ici on observe qu’ils comparent ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique. On cherche donc dans l’ énonce les 5 situation types des observations dépendantes: mesures répétées sur les mêmes sujets (ex. on mesure la température d’un patient avant et après la prise d’un médicament); comparaison des jumeaux identiques; comparaison des échantillons appariées (pour chaque sujet dans un group on trouve un sujet dans l’autre groupe avec caractéristiques similaires); comparaison des plusieurs observations faites sur certaines parties du corps du même individu, qui conceptuellement ne sont pas différentes – ex: partie gauche avec la partie droite (ex. un test dermatologique avec une traitement pour diminuer la perte de cheveux); comparaison des deux (ou plusieurs) méthodes pour mesurer une caractéristique (ou deux techniques diagnostiques [ex tension artérielle]) sur les mêmes sujets. Ici ils comparent ceux qui ont eu activité physique, et ceux qui ont n’ont pas eu activité physique, donc aucune des 5 situations types pour des échantillons dépendantes => il s’agit des observations indépendantes. Pour la chois d’un test statistique, s’il s’agit d’une comparaison, on se pose la question, **qu’est qu’il comparent (quelle est le type de la variable/ caractéristique d’intérêt comparée)?** Le test concerne une variable qualitative (parce que les valeurs de ces variables sont représentées par texte : oui/non). D’une manière implicite, quand on compare des caractéristiques qualitatives, notre but est de comparer leurs fréquences. On tenant compte de toutes ces informations (comparaison des deux échantillons indépendants concernant les fréquences d’une variable qualitative) on peut déduire => le nom du test est : soit le test Khi deux, soit le test exact Fisher. La chois entre les deux sera faite en regardant le tableau de contingence théorique. tableau de contingence. (pour faire le test Khi2 : Réalisez le tableau observée : sélectionnez toutes les données, réalisez un tableau de contingence avec Insert + PivotTable : variable activité physique sur les lignes, variable paresthésies sur les colonnes, variable activité physique dans Values. Réalisez le tableau théorique : Recopiez le tableau sans le formatage, en éliminant le contenu, et en gardant les totaux. Calculez les valeurs des cellules = total sur la ligne \* total sur la colonne / le grand total. Calculez la statistique Khi deux avec la formule somme i=1..4 de (f. observée - f. théorique)^2/f. théorique. Calculez la valeur du p avec la fonction CHITEST). La valeur du p du test=0.08, la valeur critique du test=3.84 => donc la région du rejet est [3.84, +infini), la valeur du paramètre du test statistique=3.0). L’hypothèse nulle : il n’y a pas de différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant les fréquences des paresthésies OU il n’y a pas de relation statistiquement significative entre l’activité physique et les paresthésies. L’hypothèse alternative : il y a une différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant les fréquences des paresthésies OU il y a une relation statistiquement significative entre l’activité physique et les paresthésies. Interprétation du p : parce que p > 0.05 on ne peut pas rejeter l’hypothèse nulle, donc on ne peut pas dire qu’il y a une différence statistiquement significative entre les sujets qui ont eu activité physique et les sujets qui n’ont pas eu activité physique concernant les fréquences des paresthésies OU on ne peut pas dire qu’il y a une relation statistiquement significative entre l’activité physique et les paresthésies. Attention !!! si p>0.05 s’il vous plait de ne pas énoncer des affirmations avec des erreurs comme les suivantes : « on ne peut pas dire qu’il y a une **différence** (- ici le mot différence n'est pas correct mais on peut dire relation/association…) significative entre l’activité physique et les paresthésies» ou de dire qu’« on rejette l’hypothèse alternative », ou de dire qu’« on accepte l’hypothèse nulle », ou on dire « qu’il n’y a pas des différences … »]]
3. **Eliminatoire:** Créez la table de fréquence pour la variable Paresthésie. (avec PivotTable) [voir le fichier des conseils pour Microsoft Excel]
4. **Eliminatoire:** Calculer pour les variables suivantes: glucose initiale et finale les statistiques suivants: moyenne arithmétique, médiane, mode, maximum, minimum, écart type, coefficient d’asymétrie et coefficient d’aplatissement avec Data Analysis. [voir le fichier des conseils pour Microsoft Excel]
5. Interprétez le niveau d'homogénéité de la variable Poids. [voir le fichier des conseils pour Microsoft Excel] [On doit calculer le coefficient de variance = déviation standard(STDEV)/moyenne(AVERAGE) = 0.14 – le résultat est un pourcentage : 14%, donc le poids est une variable relative homogène – elle est compris entre 10% et 20% - voir le fichier des conseils Excel]
6. **Eliminatoire:** Interprétez normalité de la distribution de la variable Poids en fonction du coefficient d’asymétrie et le coefficient d’aplatissement [voir le fichier des conseils pour Microsoft Excel] [On doit calculer le coefficient d’asymétrie (SKEW)=-0.11 et le coefficient d’aplatissement (KURT)= - 0.9 – les deux sont a l’intérieur de l’intervalle [-1, 1], donc il y a une suggestion que les données sont normale distribuées - voir le fichier des conseils Excel]
7. Interprétez normalité de la distribution de la variable Poids en fonction de la moyenne et de la médiane [voir le fichier des conseils pour Microsoft Excel] [si on regarde seulement la moyenne (AVERAGE) et la médiane (MEDIAN), on doit penser s’ils sont similaires, ici 75, et 77 sont proches, et ca est une suggestion de normalité]
8. Interprétez pour la variable Poids : moyenne arithmétique, médiane, mode, maximum, minimum, l’amplitude, écart type, coefficient d’asymétrie, coefficient d’aplatissement, le quartile 1, 2 et 3. [voir le fichier des conseils pour Microsoft Excel]
9. Triez (ordonnez les données en ascendant) les données en fonction de l’hypertension. [voir le fichier des conseils pour Microsoft Excel]
10. Pour chaque patient, calculer l'indice de masse corporelle (IMC) en utilisant la formule:

IMC = poids (kg) / taille2 (m2)

1. Quel est le pourcentage de patients ayant le glucose initial dans l’intervalle moyenne ± 1 déviation standard? [voir le fichier des conseils pour Microsoft Excel] [~68% - connu depuis la théorie si la variable est normale distribuée]
2. Quel est le pourcentage de patients ayant le glucose final dans l’intervalle moyen ± 2 écart types? [voir le fichier des conseils pour Microsoft Excel] [~95% - connu depuis la théorie si la variable est normale distribuée]
3. **Eliminatoire:** Pour chaque patient, vérifiez si la valeur du IMC est >= 30 et créez une colonne nommée Obésité ou surpoids, avec oui (si IMC >=30) et non comme des valeurs. [=IF(K2>=30,"oui","non")] [voir le fichier des conseils pour Microsoft Excel]
4. Calculer l'estimateur ponctuel pour la moyenne du glucose initial **chez les sujets avec activité physique**! Calculez l'intervalle de confiance à 95% associée, sachant que la valeur critique de 1,97, et que les données sont normales distribuées. Interprétez l’intervalle de confiance. [voir le fichier des conseils pour Microsoft Excel] [ici il s’agit de l’intervalle de confiance pour une moyenne d’un seul group. Pour le faire on doit trier les sujets en fonction de l’activité physique, pour réussir appliquer les formules sur les sujets avec activité physique. Puis on calcule la moyenne arithmétique, la déviation standard, et le nombre des sujets, avec des formules. La moyenne représente aussi l’estimateur ponctuel de la moyenne. Après on applique les formules (voir le fichier des conseils) pour estimer la borne inférieure et supérieure de l’intervalle de confiance pour une moyenne. Enfin on écrit l’intervalle. La vraie valeur de la moyenne du glucose initial dans la population cible se trouve entre 157.1 - 163.1 avec une probabilité de 95%.]
5. Calculer l'estimateur ponctuel pour la fréquence des sujets avec Hypertension. Calculez l'intervalle de confiance à 95% associée, sachant que la valeur critique de 1,96. Interprétez l’intervalle de confiance. [voir le fichier des conseils pour Microsoft Excel] [ici il s’agit de l’intervalle de confiance pour une fréquence d’un seul group. Pour le on calcule la fréquence absolue (nombre des sujets avec hypertension), le nombre total des sujets, et la fréquence relative avec des formules. La fréquence relative représente aussi l’estimateur ponctuel de la fréquence. Après on applique les formules (voir le fichier des conseils) pour estimer la borne inférieure et supérieure de l’intervalle de confiance pour une fréquence. Enfin on écrit l’intervalle. La vraie valeur de la fréquence de l’hypertension dans la population cible se trouve entre 0.21 - 0.42 avec une probabilité de 95%.]
6. Calculer l'estimateur ponctuel pour la moyenne de la différence du glucose finale entre les sujets avec activité physique et ceux sans avec activité physique. Calculez l'intervalle de confiance à 95% associée, sachant que l’erreur standard est de 3,4, et la valeur critique de 1,98, et que les données sont normales distribuées. Interprétez l’intervalle de confiance. [voir le fichier des conseils pour Microsoft Excel] [ici il s’agit de l’intervalle de confiance pour la différence des deux moyennes (des deux groups). Pour le faire on doit trier les sujets en fonction de l’activité physique, pour réussir appliquer les formules sur les sujets avec et sans activité physique. Puis on calcule la moyenne arithmétique, la pour les deux groupes, avec des formules, et puis on fait la différence entre les deux moyennes. La différence moyenne représente aussi l’estimateur ponctuel de la moyenne de la différence. Après on applique les formules (voir le fichier des conseils) pour estimer la borne inférieure et supérieure de l’intervalle de confiance pour la moyenne de la différence. Enfin on écrit l’intervalle. La vraie valeur de la moyenne de la différence du glucose finale entre ceux avec activité physique et ceux sans activité physique dans la population cible se trouve entre 20.9 et 34.3 avec une probabilité de 95%.]
7. Calculer l'estimateur ponctuel pour la différence des pourcentages des hypertensives entre les sujets avec activité physique et ceux sans avec activité physique. Calculez l'intervalle de confiance à 95% associée, sachant que l’erreur standard est de 1.6, et la valeur critique de 1,96. Interprétez l’intervalle de confiance. [voir le fichier des conseils pour Microsoft Excel] [ici il s’agit de l’intervalle de confiance pour la différence des deux fréquences relatives (des deux groups). Pour le faire on doit faire le tableau de contingence pour la relation entre les deux variables avec PivotTable. Puis on calcule la fréquence absolue et relative, la pour les deux groupes, avec des formules, et puis on fait la différence entre les deux fréquences. La différence des fréquences représente aussi l’estimateur ponctuel de la moyenne des fréquences. Après on applique les formules (voir le fichier des conseils) pour estimer la borne inférieure et supérieure de l’intervalle de confiance pour la différence des fréquences relatives. Enfin on écrit l’intervalle. La vraie valeur de la différence des fréquences de l’hypertension entre ceux avec activité physique et ceux sans activité physique dans la population cible se trouve entre -3.1 - 3.2 avec une probabilité de 95%.]
8. Précisez le type des variables suivantes: Age, Hypertension [voir le fichier des conseils pour Microsoft Excel] [Age – quantitative continue ; hypertension – qualitative dichotomique]
9. Calculer le risque relative d’avoir des paresthésies pour ceux qui ont eu activité physique par rapport à ceux qui n’ont pas eu activité physique. Interprétez le résultat [voir le fichier des conseils pour Microsoft Excel] [on doit réaliser le tableau de contingence pour la relation entre les deux variables. Puis on calcule RR=[a/(a+b)]/[c/(c+d)]=0.35. Il est inferieur a 1, dont l’activité physique est un facteur protectif pour l’apparition des paresthésies – il diminue le risque.]
10. **Eliminatoire:** **֍** Sélectionnez la colonne avec L’activité physique et la colonne avec la glycémie finale (avec leurs titres), en utilisant la touche **CTRL** (pour la sélection des colonnes multiples). Dans le tab **Insert**, section **Charts**, image de l’histogramme, **Box and Wisker**.

**Les exercices suivants peuvent être réalisées après le Travail Pratique de corrélation et régression**

1. **Eliminatoire:** Représentez graphiquement la distribution du poids en fonction de l’âge. Ecrivez l’interprétation : graphique linéaire ou non, relation (directe) proportionnelle, inverse proportionnelle. Interprétez la pente de la droite. Interprétez le coefficient de détermination. [voir le fichier des conseils pour Microsoft Excel] [ici il s’agit d’un graphique pour la relation entre deux variables quantitatives – donc un graphique nouage des points (pour le faire on sélectionne les deux colonnes, avec la touche CTRL foncée. Puis on insert le graphique, avec un clic dans l’onglet Insert, Charts, Scatter. Puis on ajoute la droite de régression, l’équation de régression, et le coefficient de détermination, avec clic droit – Add trendline.) On observe une pente ascendante – qui indique une relation directe proportionnelle (plus les valeurs de l’âge sont importantes, plus les valeurs du poids sont importants), et une relation linéaire (les points semble suivre une droite). Le coefficient du x, la pante est = 0.53 => pour chaque année en plus le poids augmente en moyenne avec 0.53 kg. Le coefficient de détermination est R2 = 0.08. On considère ça comme un pourcentage => 8%. 8% de la variance du poids est expliquées par la relation linéaire avec l’âge]
2. **Eliminatoire:** Calculez le coefficient de corrélation pour la relation entre le poids en l’âge. **Eliminatoire:** Interprétez le coefficient (direction et importance de la relation). [voir le fichier des conseils pour Microsoft Excel] [=CORREL(C2:C80,D2:D80)=-0.21. Parce qu’il est négatif, ça indique une relation inverse proportionnelle. Parce que la valeur est entre (-0.25, et 0) l’importance de la relation est négligeable ou aucune corrélation linéaire entre les variables.]
3. **Optionnel.** Représentez graphiquement le graphique des moyennes (colonnes avec barrés des erreurs) pour la Glucose initiale par hypertension, en supposant une distribution des données normale distribuées. [voir le fichier des conseils pour Microsoft Excel]