

# ATELIER SAS – TESTS D’HYPOTHESES

## PARTIE EXERCICE

---

→ Chaque question doit être organisée selon le schéma présenté en cours.

### EXERCICE 1 : LES DEPENSES A NOËL

1. Importez le jeu de données « *depense.txt* »

Ce jeu de données décrit les dépenses à Noël de plusieurs familles selon leur taille, la CSP du chef de famille, ainsi que de la pratique de la religion. (Instruction DATA)

2. Décrivez les variables (nombre d’individus, nombre de modalités...) (PROC DATASETS + PROC FREQ)

3. La dépense moyenne à Noël chez les familles pratiquantes est-elle significativement plus élevée que chez les familles non pratiquantes ?

- Représentez graphiquement les variables à l’aide d’un boxplot (pensez à trier le jeu de données par la variable adéquate). Que pouvez-vous constater ? (PROC SORT + PROC BOXPLOT)
- Quel test semble le plus adapté pour répondre à la question ?
- Quelles sont les hypothèses de ce test (H0 et H1) ?
- Vérifiez les conditions (échantillons gaussiens, homoscedasticité des variances) (PROC UNIVARIATE + PROC TABULATE)
- Réalisez le test de Student pour échantillons indépendants. (PROC TTEST)
- Quelle est la p-value ? Qu’en conclure ?

4. Existe-t-il un lien entre la taille de la famille et les dépenses à Noël ?

- Représentez graphiquement les variables à l’aide d’un nuage de points. Que pouvez-vous constater ? (PROC GPLOT)
- Quel test semble le plus adapté pour répondre à la question ?
- Quelles sont les hypothèses de ce test (H0 et H1) ?
- Réalisez le test de corrélation. (PROC CORR)
- Quelle est la p-value ? Qu’en conclure ?

5. Y-a-t-il un lien entre la CSP et le fait d’être pratiquant ou non ?

- Représentez graphiquement les variables à l’aide d’un graphique. Que pouvez-vous constater ? (PROC GCHART)
- Réalisez un tableau de contingence. (PROC FREQ)
- Réalisez un tableau des effectifs théoriques. (PROC FREQ avec expected)
- Quel test semble le plus adapté pour répondre à la question ?
- Les conditions d’application du test choisi sont-elles respectées ?
- Quelles sont les hypothèses de ce test (H0 et H1) ?
- Réalisez le test du khi-deux. (PROC FREQ avec chisq)
- Quelle est la p-value ? Qu’en conclure ?

## EXERCICE 2 : LA VO2MAX CHEZ LES CYCLISTES

Une équipe cycliste suit la progression de la VO2max de ses coureurs au fil des années. Pour 10 coureurs cyclistes, les VO2max suivantes ont été notées en 2015 et en 2016.

<b>Coureur</b>	<b>Annee2015</b>	<b>Annee2016</b>
C1	68	69
C2	60	65
C3	74	70
C4	68	69
C5	67	73
C6	74	72
C7	70	69
C8	63	72
C9	74	65
C10	71	74

1. Rentrez ce jeu de données via une étape data (par exemple : instruction DATALINES ou CARDS).
2. Créez une nouvelle variable qui définit la différence de VO2max entre 2016 et 2015.
3. Faites un histogramme sur cette nouvelle variable. (PROC GCHART)
4. Ici, nous voulons tester si les VO2max ont augmenté d'une année à l'autre. Faites un test de conformité de la moyenne sur cette nouvelle variable. (PROC TTEST side = U)
5. Représentez graphiquement les variables Annee2015 et Annee2016. (PROC GCHART)
6. Quel test est utilisé pour comparer la moyenne de deux échantillons appariés ?
7. Quelles sont les hypothèses de ce test (H0 et H1) ?
8. Vérifiez les conditions d'application de ce test. (PROC UNIVARIATE)
9. Faites un test de Student pour échantillons appariés. Qu'en conclure ? (PROC TTEST side = U)

## EXERCICE 3 : LE POIDS DES POULETS

1. Importer le jeu de données chickwts.txt (PROC IMPORT par exemple)  
Ce jeu de données décrit le poids des poulets selon leurs compléments alimentaires.

2. Existe-t-il un lien entre le poids et le complément alimentaire donné ?
  - a) Combien y a-t-il d'individus par type de compléments alimentaires ? (Pensez à trier les données) (PROC SORT + PROC FREQ)
  - b) Représentez graphiquement les données. (PROC BOXPLOT)  
Que peut-on en déduire ?
  - c) Quel test semble le plus adapté pour répondre à la question ?
  - d) Quelles sont les hypothèses de ce test (H0 et H1) ?

- e) Vérifiez les conditions (échantillons gaussiens, homoscedasticité des variances, normalité des résidus). (PROC UNIVARIATE NORMAL PLOT sur les échantillons + PROC GLM + PROC UNIVARIATE NORMAL PLOT sur les résidus)
- f) Réalisez le test de l'Anova. Qu'en conclure ? (PROC GLM)

#### EXERCICE 4 : LE TEMPS PASSE DEVANT LA TELE

1. Importez le jeu de données temps.csv (PROC IMPORT)

Ce jeu de données représente le temps journalier passé devant la télévision en minutes, par genre et par tranche d'âge (variable « periode »).

2. Recodez la variable « genre » de la façon suivante : (DATA + SET + SELECT... WHEN)

<i>Sexe</i>	<i>Recode_sexe</i>
1	Homme
2	Femme

3. Recodez la variable « periode » de la façon suivante : (DATA + SET + SELECT... WHEN)

<i>Période</i>	<i>Recode_période</i>
A	A - moins de 20 ans
B	B - 20-30 ans
C	C - 30-40 ans
D	D - 40-60 ans
E	E - plus de 60 ans

4. Le genre et l'âge influent-ils sur le temps passé devant la télévision ?

- a) Décrivez les données (nombre d'observations, moyenne et écart-type pour chaque croisement de modalités). (PROC SORT + PROC MEANS : dans l'output, vous pouvez créer une table qui contient la moyenne de chaque croisement de modalités qui sera utile pour la représentation graphique).
- b) Représentez graphiquement les données. (PROC PLOT : ici, on peut utiliser les variables « moyenne », « sexe » et « période »). Que peut-on en déduire ?
- c) Quel test vous semble le plus adapté pour répondre à la question ?
- d) Vérifiez les conditions (normalité des résidus). (PROC GLM + PROC UNIVARIATE NORMAL PLOT sur les résidus)
- e) Réalisez une ANOVA à deux facteurs (sexe et période). (PROC ANOVA)  
Qu'en conclure ?