

RAPPEL

- ✗ Examen : Lundi 09 janvier de 10H30 à 12H30 en Amphi 4.
- ✗ Autorisation : notes de cours, ouvrages, dictionnaire langue française
- ✗ Pas le droit à un ordinateur
- ✗ Etudiants ayant 1/3 de temps supplémentaire : début examen à **09H50** dans la salle **C213**

Question 1 (8 points)

Voici des démarches de recueil de données accompagnées de questions de recherche ou d'hypothèses. A quel type de recueil correspondent-ils ? Justifiez votre réponse en utilisant le cours et notez bien sur la feuille si vous répondez à 1a, 1b, 1c, etc.

1A

Un sociologue veut vérifier l'hypothèse que les jeunes non diplômés qui réussissent à trouver un emploi sont ceux qui sont soutenus par leur réseau social (famille, amis, autres connaissances) ou qui réussissent à le mobiliser (pour contacter un employeur, etc.). Pour vérifier ces hypothèses, lui et son équipe suivent 100 jeunes sortis du système éducatif sans avoir obtenu de diplôme (ni bac, ni CAP, ni BEP, etc.). Ils les revoient chaque année pendant 4 ans (entre 18 et 21 ans) pour un entretien long lors duquel les jeunes sont invités à raconter librement leurs tentatives pour trouver du travail, leurs réussites et leurs échecs.

1A

Méthode : Interview des individus lors d'entretiens

Structuration : faible = entretien non directif ou libre

Taille : Etude comprenant une 100^{aine} d'individus

Temps : Etude en diachronie = chaque individu revient tous les 4 ans

Quantitatif/qualitatif : les deux analyses sont possibles

1B

Dans son mémoire, un étudiant de master *Métiers de l'Enseignement Scolaire* s'intéresse à la façon dont l'enseignement de l'orthographe à l'école primaire est présenté dans le *Bulletin Officiel du Ministère de l'Education National* (le « BO », qui fixe les programmes, les horaires, etc.). L'étudiant suppose que l'importance accordée à l'apprentissage de l'orthographe dans le BO a diminué depuis l'extension de l'informatique et du web (vers 1995). Il dépouille donc systématiquement tous les BO entre 1995 et 2010. Il compte les apparitions du mot *orthographe* et recense les contenus à acquérir à la fin du CM2 (ex : distinguer les principaux homophones grammaticaux : à-a, où-ou, etc. ; connaître les particularités des marques du pluriel de certains noms en -al, - eau, - eu, - ou ; etc.). Son hypothèse est que la fréquence du mot *orthographe* diminue dans le BO entre 1995 et 2010 et que le niveau à atteindre à la fin du CM2 est de moins en moins exigeant.

1B

Méthode : Analyser des documents constitués par d'autres (étude de corpus)

Structuration : plutôt faible car concentration sur deux points : orthographe et contenus disciplinaires à acquérir

Taille : corpus de taille importante (tous les BO sur 15 ans)

Temps : Etude en diachronie

Quantitatif/qualitatif : quantitatif (fréquence du mot *orthographe*)

1c

On veut savoir si des enfants de 2 ans sont déjà capables de comprendre certaines distinctions syntaxiques, par exemple la différence entre la voix active (*le chat blanc poursuit le chat noir*) et la voix passive (*le chat blanc est poursuivi par le chat noir*). Dans ce but, on utilise la méthode d'orientation du regard auprès d'un échantillon de 50 enfants de 2 ans. Chaque enfant est assis sur les genoux de sa mère et entend des phrases passives, par exemple, *le chat blanc est poursuivi par le chat noir*. Sur deux écrans situés à gauche et à droite de l'enfant, une image apparaît : un chat noir poursuivant un chat blanc d'un côté, un chat blanc poursuivant un chat noir de l'autre côté. On enregistre la direction du regard de l'enfant et on vérifie que le regard se stabilise plus vite sur l'image correspondant à la phrase entendue.

1E

Méthode : Mise en place d'une expérimentation + Observation d'individus

Structuration : forte pour l'expérimentation et faible pour l'observation des enfants

Taille : importante (50 enfants âgés de 2 ans)

Temps : synchronie (à un moment T)

Quantitatif/qualitatif : quantitatif

1D

Pour vérifier si le QI est lié au développement du langage, on enregistre 90 enfants de même milieu social répartis dans trois groupes d'âge (3 ans, 4 ans et 5 ans) pendant dix heures d'interactions dans le cadre familial. Parallèlement, on leur fait passer un test de QI. On transcrit les énoncés et on calcule les LME (longueur moyenne des énoncés) individuelles. On calcule ensuite la corrélation entre QI et LME.

1D

Méthode : Recueillir des productions langagières orales

Structuration : faible (enregistrement audio → transcription → comptage LME)

Taille : importante (90 enfants)

Temps : synchronie

Quantitatif/qualitatif : Quantitatif (calcul de la corrélation entre QI et LME)

Question 2

(6 points)

Voici des données (fictives) sur des notes obtenues par 17 étudiants de L à un examen de phonologie. Parallèlement, leur professeur a noté les absences et les présences pendant le semestre. Il a fait deux catégories d'étudiants:

- Les « assidus » qui sont venus à 8 séances sur 12 au moins.
- Les « **non assidus** » qui sont venus à moins de 8 séances sur 12.

SUJET	NOTE	ASSIDUITE	SUJET	NOTE	ASSIDUITE
Sujet 1	19	Assidu	Sujet 10	9	Non assidu
Sujet 2	17	Non assidu	Sujet 11	9	Assidu
Sujet 3	18	Assidu	Sujet 12	9	Non assidu
Sujet 4	14	Non assidu	Sujet 13	9	Assidu
Sujet 5	14	Assidu	Sujet 14	8	Non assidu
Sujet 6	12	Non assidu	Sujet 15	8	Assidu
Sujet 7	11	Assidu	Sujet 16	3	Non assidu
Sujet 8	11	Non assidu	Sujet 17	7	assidu
Sujet 9	10	Assidu			

Le professeur de phonologie calcule les moyennes et les écarts types des notes. Il obtient les résultats listés dans le tableau suivant et il en conclut que l'assiduité est un facteur important de la réussite aux examens, puisque la moyenne des assidus est supérieure de plus d'un point à la moyenne des non assidus.

	Les 17 étudiants	Les 9 étudiants assidus	Les 8 étudiants non assidus
Moyenne	11.05	11.67	10.37
Écart type	4.2	4.3	4.2

Un collègue à lui qui est calé en statistiques fait un test de significativité de comparaison des moyennes entre les assidus et les non assidus : il trouve $p = 0.54$.

Répondez aux cinq questions suivantes. Justifiez brièvement votre réponse en utilisant le cours et notez bien sur la feuille si vous répondez à 2a, 2b, 2c, etc.

2a – Calculez la médiane et le mode de l'échantillon des 17 notes.

2b – Il y a deux variables impliquées dans ces données : l'assiduité (assidu / non assidu) et la note. De quel type de variable s'agit-il ?

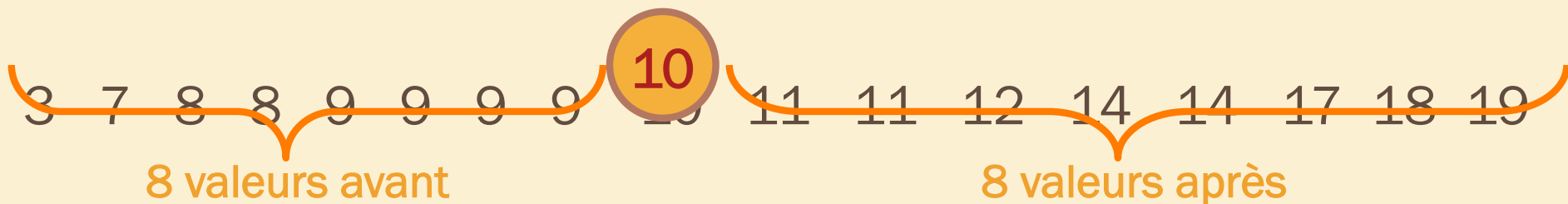
2c – Quelle type de comparaison de moyennes est utilisé pour comparer les assidus et les non assidus ?

2d – Comment s'appelle le test de significativité fait par le collègue calé en statistiques ?

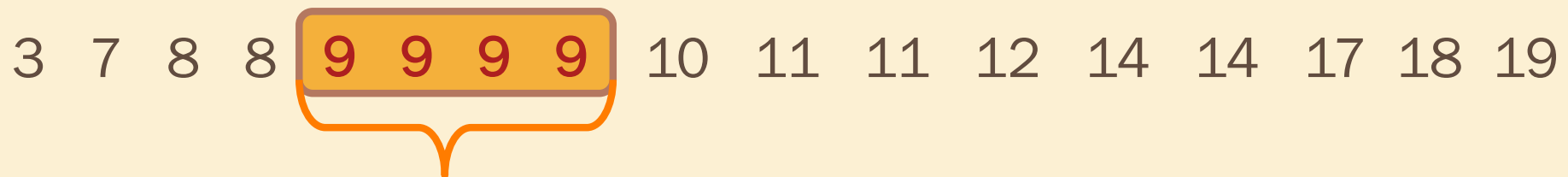
2e – Quelle conclusion peut-on tirer du fait que $p = 0.54$?

2a – Calculez la médiane et le mode de l'échantillon des 17 notes.

Médiane : désigne la valeur qui divise un ensemble en deux moitiés → la médiane est le point milieu de l'ensemble



Mode : désigne la valeur la plus représentée d'une variable quelconque dans une population d'objets, de personnes, de choses



X 4 → Le mode est donc 4

2b – Il y a deux variables impliquées dans ces données : **l'assiduité** (assidu / non assidu) et **la note**. De quel type de variable s'agit-il ?

L'assiduité :

possibilité entre assidu ou non-assidu (= étiquette)

➔ **Variable nominale** (dit aussi catégorielle)

La note :

Possibilité entre 0 et 20 (= valeur numérique)

➔ **Variable quantitative**

2c – Quel type de comparaison de moyennes est utilisé pour comparer les **assidus** et les **non assidus** ?

× Comparaison non-appariée



car on compare deux moyennes de deux groupes différents (assidus vs. non-assidus)

Comparaison appariée : On compare deux moyennes issues d'un même groupe (groupe des *assidus* / moyenne des assidus fille vs. Moyenne des assidus garçons)

2d – Comment s'appelle le test de significativité fait par le collègue calé en statistiques ?

✗ Le test de Student OU test T

2e – Quelle conclusion peut-on tirer du fait que $p = 0.54$?

✕ Rappel, en SHS :

Si $p < 0,05$ → résultat significatif

Si $p > 0,05$ → résultat non significatif

$P = 0,54$ → Résultat non significatif

Interprétation : J'ai 54% de chance de généraliser à tort en affirmant que les moyennes sont différentes OU 46% de chance de généraliser à tort en affirmant que les moyennes sont égales.

Par conséquent, le résultat n'est pas significatif → On ne peut donc pas tirer de conclusion sur le lien entre l'assiduité en cours et la note à l'examen

Question 3

(6 points)

Sur le site de l'Observatoire français des drogues et des toxicomanies, on trouve le compte rendu suivant :

« Les liens entre consommation de cannabis et accidents de la route ont été étudiés dans le cadre de l'enquête SAM (Stupéfiants et Accidents Mortels de la circulation routière). Entre octobre 2001 et septembre 2003, tous les conducteurs impliqués dans un accident mortel de la route ont fait l'objet d'un dépistage en vue de savoir s'ils avaient consommé des stupéfiants. Au cours de la période d'enquête, 7 % de conducteurs impliqués dans un accident mortel se sont révélés positifs au cannabis (...). La comparaison des conducteurs responsables (de l'accident) à un groupe de conducteurs non responsables a permis de montrer que les conducteurs sous l'influence du cannabis ont 1,8 fois plus de risque d'être responsables d'un accident mortel que les conducteurs négatifs. Au total, les résultats conduisent à évaluer à 230 les décès annuels par accidents de la route attribuables à la consommation de cannabis au début des années 2000. »

<http://www.ofdt.fr/ofdtdev/live/produits/cannabis/consequ.html>

L'interprétation qui est habituellement donnée à ces faits est de considérer que l'accroissement du risque d'accident est une conséquence des effets du cannabis sur les états mentaux : euphorie, troubles de la mémoire immédiate et de la perception, difficultés à se concentrer, diminution des réflexes, troubles de la coordination motrice. **En vous basant sur la partie du cours qui traite un problème similaire, proposez une autre interprétation (ou plusieurs) à cette relation statistique.**

Ce que l'enquête nous dit...

- ✗ L'enquête montre que la consommation de cannabis multiplie par 1,8 le risque d'avoir un accident mortel ;
- ✗ Autrement dit, les conducteurs qui ne consomment pas de cannabis causent moins d'accidents mortels sur la route.
- ✗ 230 décès/an suite à un accident de la route sont dus à des conducteurs consommateurs de cannabis

Ce que l'enquête ne nous dit pas...

- ✗ Combien de sujets ont été testés ? Une dizaine, une centaine, un millier ?
- ✗ Comment les enquêteurs ont-ils trouvé les 1,8 fois plus de risque de causer un accident mortel sous l'influence du cannabis ?
- ✗ 230 décès liés au cannabis sur combien en tout ?
- ✗ Variables non prises en compte : est-ce que les consommateurs de cannabis avaient également bu, étaient fatigués, conduisaient à une heure nocturne, etc.

POUR CONCLURE...

- ✗ En l'absence de données (taille de l'échantillon, protocole expérimental, autres variables), ces résultats n'ont pas de valeur scientifique !
- ➔ Simplifier pour mieux convaincre ! Les politiques publiques ont-elles simplifié les données scientifiques ?