Travail de synthèse

Présentation orale d'un article et implantation de la méthode

Pondération	30 % de la note finale (corrigé sur 100 points)
Mode	Travail individuel
Choix de l'article	au plus tard le **vendredi 19 février 2027** (séance 6)
Présentations	**vendredi 16 avril** ou **vendredi 23 avril 2027**, pendant la séance (8 h 30 à 11 h 20)
Dépôt du code	au plus tard **la veille de votre présentation, 23 h 55** (jeudi 15 avril ou jeudi 22 avril 2027), boîte de dépôt « Travail de synthèse » sur Brio

1 Consigne officielle et objectifs

Consigne officielle

L'étudiant doit donner une présentation orale d'un article et remettre le code documenté de son implantation ou de son application de la méthode biostatistique décrite dans l'article. L'utilisation d'une fiche de lecture et la préparation d'exercices en lien avec l'article présenté sont encouragées.

Le travail de synthèse vous fait parcourir, sur un article de votre choix, le cycle complet pratiqué dans les études de cas : lire un article méthodologique, en dégager les choix statistiques à l'aide de la grille méthodologique du cours, comprendre la théorie qui les justifie, puis implanter ou appliquer la méthode en R de façon reproductible. Au terme du travail, vous devriez être en mesure de :

1. expliquer une méthode biostatistique récente à un auditoire de statisticiens, en distinguant clairement l'objet d'intérêt, la nuisance et l'inférence ;
2. situer la méthode par rapport aux méthodes vues en classe ;
3. implanter ou appliquer la méthode en R avec un code documenté et reproductible ;
4. porter un regard critique sur les hypothèses, les résultats et les limites de l'article.

2 Choix de l'article

1. Choisissez un article dans la liste du tableau 1 (les PDF sont déposés sur le site du cours) *ou* proposez un autre article méthodologique en biostatistique. Un article hors liste doit

être approuvé par l'enseignant : il doit proposer ou évaluer une méthode statistique (et non seulement l'appliquer), avoir été publié dans une revue avec comité de lecture ou être une prépublication reconnue, et sa méthode doit pouvoir être implantée ou appliquée en R dans un délai raisonnable.

2. Transmettez votre choix par courriel à l'enseignant au plus tard le **vendredi 19 février 2027**, avec un deuxième choix. Pour un article hors liste, joignez le PDF et deux ou trois phrases sur ce que vous comptez implanter.

3. Un même article ne peut être présenté que par une seule personne ; les articles sont attribués selon l'ordre de réception des choix. L'enseignant confirme l'attribution et la date de présentation (16 ou 23 avril) au plus tard le vendredi 26 février 2027.

Tableau 1: Articles suggérés et séances du cours auxquelles ils se rattachent.

Référence	**Thème**	**Séances**
Aktas Samur, A., Coskunfirat, N. et Saka, O. (2014). *PeerJ*, 2, e648.	Comparaison des GEE et des modèles linéaires mixtes généralisés pour des réponses binaires longitudinales.	3, 6
Pedroza, C. et Truong, V. T. T. (2017). *Trials*, 18, 512.	Estimation de risques relatifs dans les études multicentriques comportant peu de centres : comparaison de méthodes par simulation.	2, 3, 6
Talbot, D., Mésidor, M., Chiu, Y., Simard, M. et Sirois, C. (2022). arXiv:2112.00547.	Perspective alternative sur la régression de Poisson robuste pour estimer des rapports de risques ou de prévalences.	1, 6
Zivich, P. N., Ross, R. K., Shook-Sa, B. E., Cole, S. R. et Edwards, J. K. (2024). *Statistics in Medicine*, 43, 5562–5572.	Estimateur sandwich empirique de la variance pour la g-computation par espérances conditionnelles itérées (ICE).	6
Ristl, R., Hothorn, L., Ritz, C. et Posch, M. (2020). *Statistical Methods in Medical Research*, 29(6), 1746–1762.	Inférence simultanée pour plusieurs modèles GEE marginaux.	6, 11
Joo, J. W. J., Hormozdiari, F., Han, B. et Eskin, E. (2016). *Genome Biology*, 17, 62.	Correction pour tests multiples dans les modèles linéaires mixtes (études d'association génétique).	3, 11
Finak, G. et coll. (2015). *Genome Biology*, 16, 278.	MAST : modèle en deux parties pour l'expression génique de cellules uniques (ARN), tests d'expression différentielle.	3, 11
Miller, Roe, Hu et Bell (2020). *BMC Medical Research Methodology*, 20, 50.	Puissance du test du khi-deux comparée à celle d'un modèle linéaire mixte généralisé en présence de données manquantes.	2, 3, 8
Ségalas, C., Helmer, C., Genuer, R. et Proust-Lima, C. (2024). *Statistics in Medicine*.	Analyse en composantes principales fonctionnelle comparée aux modèles mixtes pour des mesures répétées avec données manquantes.	3–4, 8

3 Présentation orale

Durée. 20 minutes de présentation, suivies de 10 minutes de questions de l'enseignant et de la classe. La présentation est interrompue à 22 minutes. Les présentations sont données en français.

Structure suggérée. Le minutage est indicatif ; adaptez-le à votre article.

Partie	Contenu attendu	Durée
1. Contexte	Problème biomédical, question de recherche, données ou devis ; pourquoi les méthodes usuelles ne suffisent pas.	3 min
2. Grille méthodologique	Les six rubriques de la grille pour la méthode de l'article (une diapositive de synthèse est recommandée).	3 min
3. Théorie	Modèle, critère d'estimation ou de test, résultats clés (propriétés, variance, conditions) ; lien avec les méthodes vues en classe.	5 min
4. Implantation	Ce que vous avez implanté ou appliqué, choix de programmation, difficultés rencontrées, vérifications (p. ex. comparaison avec un module existant ou simulation).	4 min
5. Résultats	Vos résultats, comparés à ceux de l'article ; interprétation.	3 min
6. Limites et regard critique	Hypothèses fortes, situations où la méthode échoue, alternatives, questions ouvertes.	2 min

Implantation ou application ? Vous pouvez soit *implanter* la méthode (écrire vous-même l'estimateur, la variance, l'algorithme), soit *appliquer* la méthode à des données (celles de l'article si elles sont disponibles, un jeu de données du cours ou des données simulées) au moyen d'un module existant. Dans le second cas, on s'attend à une application qui va au-delà d'un simple appel de fonction : reproduction d'un résultat de l'article, petite étude de simulation qui vérifie une propriété annoncée, ou comparaison avec une méthode vue en classe.

Attentes supplémentaires au doctorat

Au doctorat, la partie théorique est plus approfondie : on s'attend à ce que vous présentiez au moins un argument théorique complet ou esquissé avec soin, par exemple la dérivation de la fonction d'influence de l'estimateur, l'obtention de sa loi asymptotique ou de sa variance sandwich, ou la démonstration d'une propriété d'efficacité ou de double robustesse. La durée de la présentation est la même ; condensez au besoin le contexte. Les descripteurs du critère « Compréhension et exposé de la méthode » de la grille sont appliqués en conséquence.

4 Livrables

1. **Diapositives** de la présentation, en PDF, nommées `STT7360_TS_Nom_Prenom_diapos.pdf`, déposées avec le code (la veille de la présentation, 23 h 55). Vous présenterez à partir de ce fichier.
2. **Code R documenté et reproductible**, dans une archive `STT7360_TS_Nom_Prenom.zip` qui contient :
   — un fichier `README` (texte ou Markdown) : objectif du code, lien avec l'article (quelles sections, quels tableaux ou figures sont reproduits), instructions d'exécution, modules

requis et leur version (sortie de `sessionInfo()`), temps d'exécution approximatif ;

- le code : script `.R` commenté, ou document R Markdown ou Quarto avec son rendu (HTML ou PDF) ; les fonctions que vous écrivez sont documentées (arguments, valeur retournée) ;
- les données utilisées, ou le code qui les génère (simulation) avec le germe aléatoire ; si les données ne peuvent être redistribuées, expliquez comment les obtenir ;
- les sorties principales (tableaux, figures) produites par le code.

Le code doit s'exécuter du début à la fin sur un autre ordinateur, sans modification autre que le répertoire de travail (utilisez des chemins relatifs).

3. **Fiche de lecture** (facultative, encouragée) : gabarit en annexe A, à joindre en PDF à l'archive.
4. **Exercices** (facultatifs, encouragés) : deux à quatre exercices en lien avec l'article, du type de ceux des études de cas (dérivation, interprétation, petite analyse en R), avec leur solution, en PDF dans l'archive. Les meilleurs exercices pourront, avec votre accord, être proposés à la classe comme préparation à l'examen.

Retards, intégrité. Le dépôt tardif de l'archive est soumis à la politique de retard du plan de cours ; une absence à la présentation sans motif valable entraîne la note zéro pour la partie orale. Le code d'autrui (y compris celui des auteurs de l'article) peut être réutilisé s'il est clairement cité dans le `README` et dans le code ; c'est alors votre contribution propre qui est évaluée. L'usage d'outils d'intelligence artificielle est déclaré selon les règles du plan de cours.

5 Grille de correction (100 points)

Pour chaque critère, la note est attribuée dans la plage de points indiquée pour le niveau atteint : *excellent* (environ 90 % des points et plus), *satisfaisant* (de 60 à 89 % environ), *insuffisant* (moins de 60 %).

Critère	Excellent	Satisfaisant	Insuffisant
A. Compréhension et exposé de la méthode (20 pts)	Méthode exposée avec exactitude et rigueur ; modèle, critère et résultats clés correctement énoncés et justifiés ; lien clair avec le cours. Au doctorat : argument théorique (fonction d'influence, loi asymptotique, etc.) présenté correctement. *18 à 20 pts*	Méthode globalement juste, mais certains résultats énoncés sans justification ou avec des imprécisions ; lien avec le cours partiel. Au doctorat : argument théorique incomplet. *12 à 17 pts*	Erreurs de fond, confusion entre objet d'intérêt et nuisance, ou exposé limité à une paraphrase de l'article. *0 à 11 pts*
B. Grille méthodologique (10 pts)	Les six rubriques sont identifiées et justifiées avec précision ; la grille éclaire la comparaison avec les méthodes vues en classe. *9 à 10 pts*	Rubriques présentes mais certaines vagues ou mal justifiées. *6 à 8 pts*	Grille absente, incomplète ou erronée. *0 à 5 pts*
C. Implantation, code et reproductibilité (25 pts)	Implantation correcte et vérifiée (comparaison avec un module, simulation ou résultats publiés) ; code clair, structuré, commenté ; `README` complet ; s'exécute du début à la fin sans modification. *23 à 25 pts*	Implantation essentiellement correcte mais peu vérifiée, ou code qui exige des corrections mineures pour s'exécuter ; documentation partielle. *15 à 22 pts*	Code absent, incorrect, non reproductible ou réduit à un appel de fonction sans documentation. *0 à 14 pts*
D. Résultats et interprétation (15 pts)	Résultats présentés clairement (tableaux, figures soignés), comparés à ceux de l'article, écarts expliqués ; interprétation statistique et biomédicale juste. *14 à 15 pts*	Résultats présentés mais comparaison ou interprétation sommaire. *9 à 13 pts*	Résultats absents, illisibles ou mal interprétés. *0 à 8 pts*
E. Regard critique (10 pts)	Hypothèses et limites discutées avec pertinence ; situations d'échec, alternatives et pistes d'amélioration proposées. *9 à 10 pts*	Quelques limites mentionnées, surtout celles que les auteurs signalent eux-mêmes. *6 à 8 pts*	Aucun recul critique. *0 à 5 pts*

Critère	Excellent	Satisfaisant	Insuffisant
F. Qualité de la présentation orale (10 pts)	Exposé structuré, clair et vivant ; diapositives lisibles (notation cohérente, figures lisibles) ; temps respecté. *9 à 10 pts*	Structure correcte mais quelques passages confus, diapositives surchargées ou dépassement léger du temps. *6 à 8 pts*	Exposé désorganisé, lu, ou dépassement important du temps. *0 à 5 pts*
G. Réponses aux questions (10 pts)	Réponses exactes et précises, qui montrent une maîtrise au-delà du contenu présenté. *9 à 10 pts*	Réponses généralement justes mais parfois vagues ou incomplètes. *6 à 8 pts*	Incapacité à répondre aux questions de base sur la méthode ou le code. *0 à 5 pts*
Total : A + B + C + D + E + F + G = 20 + 10 + 25 + 15 + 10 + 10 + 10 = 100 points			

Points de reconnaissance (facultatifs). Une fiche de lecture complète et des exercices pertinents avec solutions peuvent valoir chacun jusqu'à **2,5 points** supplémentaires (jusqu'à 5 points au total), sans que la note du travail puisse dépasser 100. Leur absence n'entraîne aucune pénalité.

Calendrier récapitulatif

- Vendredi 19 février 2027 : choix de l'article (avec un deuxième choix).
- Vendredi 26 février 2027 : confirmation de l'article et de la date de présentation.
- Veille de la présentation, 23 h 55 : dépôt des diapositives et de l'archive de code (fiche de lecture et exercices facultatifs).
- Vendredi 16 ou 23 avril 2027 : présentation (20 min + 10 min de questions).

A Gabarit de fiche de lecture (facultatif)

Référence complète de l'article

Question de recherche (méthodologique et biomédicale)

Données ou devis (source, taille, structure, variables clés ; simulation)

Grille méthodologique

Élément	Choix	Justification
1. Objet statistique d'intérêt		
2. Paramètres nuisibles, données manquantes, sources de biais		
3. Approche pour traiter la nuisance ou éviter les biais		
4. Critère d'estimation ou de test		
5. Méthode d'optimisation ou de calcul		
6. Obtention de la distribution des statistiques		

Résultats principaux (théoriques, simulation, application)

Limites (signalées par les auteurs et relevées par moi)

Ce que j'implante ou j'applique (quoi, avec quelles données, comment je vérifie)