

Biostatistiques : tests et choix de l'analyse

OSSIK • COURS 13-02 • LE CORRIGÉ • RÉVISION 0.5

PASS - majeure santé - classement d'après la maquette publique de Paris Cité, sans lien avec l'université. Cours original rédigé avec une IA. Le cours de ta faculté fait foi.

Sommaire du cours

Partie 1 • Tests, puissance et multiplicité	1
Partie 2 • Choisir une analyse selon le plan expérimental	7
Partie 3 • Comparer des moyennes et des proportions	12
Partie 4 • Régression multivariée et données de survie	18

Partie 1 • Tests, puissance et multiplicité

Prérequis et objectifs

Connaître hypothèse probabiliste, intervalle de confiance et erreur standard. Formuler la question avant d'observer les résultats.

À l'issue de ce cours, vous devez pouvoir :

- Interpréter une valeur p correctement.
- Relier alpha, bêta et puissance.
- Identifier multiplicité et différence entre absence de preuve et équivalence.

Fondements du thème

Formuler une hypothèse avant le calcul

L'hypothèse nulle H_0 spécifie une situation de référence, souvent absence de différence. L'alternative H_1 définit ce que l'étude cherche à distinguer de cette référence. Un test bilatéral recherche une différence dans les deux directions ; un test unilatéral ne doit pas être choisi après observation du sens des résultats. La statistique de test mesure l'écart aux prédictions de H_0 en tenant compte de sa variabilité.

La valeur p est, sous H_0 et les hypothèses du modèle, la probabilité d'obtenir une statistique au moins aussi extrême que celle observée selon la règle du test. Elle n'est ni la probabilité que H_0 soit vraie ni la probabilité que le résultat soit « dû au hasard ». Un résultat $p = 0,03$ ne signifie pas que l'effet a 97 % de chances d'exister. La taille de l'effet et son intervalle doivent accompagner la décision.

Risques alpha et bêta

Le risque alpha est la probabilité de rejeter H_0 lorsqu'elle est vraie. Le risque bêta est la probabilité de ne pas rejeter H_0 pour une alternative spécifiée ; la puissance vaut $1 - \beta$. Elle dépend de l'effet à détecter, de la variabilité, de l'effectif, du seuil et du plan. Il n'existe pas une puissance unique indépendante de l'effet réel envisagé.

Ne pas rejeter H_0 ne démontre pas l'égalité. Une étude petite peut manquer une différence importante. Démontrer une équivalence demande de définir à l'avance des marges acceptables et d'utiliser une procédure adaptée. Avec un très grand effectif, une différence minuscule peut devenir statistiquement significative sans intérêt pratique.

Multiplicité et transparence

Multiplier les tests augmente le risque d'au moins une fausse alerte. Pour m tests indépendants au seuil alpha, ce risque vaut $1 - (1 - \alpha)^m$ si toutes les hypothèses nulles sont vraies. Bonferroni utilise un seuil α/m pour contrôler le risque familial, sans exiger l'indépendance mais parfois de manière conservatrice. D'autres procédures répondent à d'autres objectifs ; il faut annoncer ce qui est contrôlé.

Sélectionner seulement les analyses favorables, modifier le critère principal après lecture des données ou arrêter au premier p inférieur à 0,05 altère l'interprétation. Un protocole, un plan d'analyse et un compte rendu des résultats négatifs rendent le raisonnement vérifiable. La significativité n'efface jamais un biais de mesure ou de sélection.

Développements et raisonnements

Le test compare une observation aux prédictions d'une hypothèse

Sous H_0 , une statistique de test possède une distribution de référence. On situe la valeur observée dans cette distribution. Plus elle est extrême selon la règle préétablie, plus les données sont difficiles à concilier avec H_0 et les autres hypothèses. Une petite valeur p peut aussi refléter une violation du modèle, par exemple des observations dépendantes traitées comme indépendantes.

La probabilité de H_0 sachant les données constitue une question inverse. Elle demanderait notamment un cadre probabiliste sur les hypothèses, absent du test fréquentiste ordinaire. Dire $p = 0,04$ signifie « 4 % de probabilité de se tromper en affirmant un effet » confond ces deux questions. Le risque alpha concerne la procédure lorsque H_0 est vraie, non la probabilité personnelle d'erreur après un résultat donné.

Puissance et taille d'effet

À effectif et dispersion constants, une différence plus grande est plus facile à détecter. La puissance doit donc être attachée à une alternative chiffrée. Un calcul de puissance avant l'étude aide à choisir l'effectif pour un effet jugé pertinent. Une puissance calculée après coup à partir de l'effet observé n'apporte pas la même information et peut simplement reformuler la valeur p .

Une étude peut fournir un résultat non significatif parce que l'effet est proche de zéro ou parce que l'intervalle est trop large pour distinguer les possibilités. Lire l'intervalle permet de séparer ces situations. L'équivalence exige des marges prédéfinies et une procédure montrant que les différences compatibles restent dans ces marges. La non-infériorité pose encore une question différente, orientée

vers une limite acceptable.

Plusieurs occasions de conclure

Tester vingt marqueurs, essayer plusieurs sous-groupes et choisir ensuite l'analyse la plus favorable multiplie les occasions de faux résultat. La famille d'hypothèses doit correspondre à la question scientifique. Bonferroni contrôle la probabilité d'au moins un faux rejet en utilisant α/m pour chaque test ; sa validité repose sur une borne, sans supposer leur indépendance.

Les analyses intermédiaires posent un problème comparable si l'on regarde les données jusqu'à obtenir p inférieur au seuil. Des plans séquentiels existent, mais leurs règles doivent être prises en compte. Une analyse exploratoire reste utile si son statut est explicite et si la confirmation future ne réutilise pas artificiellement les mêmes données comme une preuve indépendante.

Atelier de calcul et analyse de sensibilité

Une différence estimée de 3 unités possède une erreur standard de 1 dans une approximation normale. Le score z vaut 3 et la valeur p bilatérale environ 0,00270. L'intervalle à 95 % est [1,04 ; 4,96]. On dispose d'une estimation positive relativement précise, mais sa pertinence dépend de l'unité et du seuil scientifique d'intérêt.

La même différence avec une erreur standard de 3 donne $z = 1$, p environ 0,3173 et un intervalle [-2,88 ; 8,88]. Ces données restent compatibles avec un effet nul et avec des effets positifs plus grands. Les résumer par « aucun effet » éliminerait une grande partie de leur information.

Sous nullité globale et indépendance, vingt tests à 5 % donnent une probabilité d'au moins un faux positif de $1 - 0,95^{20} = 0,6415$. Avec Bonferroni, le seuil individuel devient 0,0025. Une valeur p de 0,01 serait significative dans une analyse unique préspecifiée à 5 %, mais pas selon cette correction pour vingt tests. Le calcul ne dit pas quelle famille est scientifiquement pertinente : cette définition relève du protocole.

Enfin, si la probabilité de ne pas rejeter H_0 est 0,15 lorsque la vraie différence vaut 5 unités, la puissance pour cette alternative est 85 %. Elle ne garantit pas une puissance de 85 % pour une différence de 1 unité.

Approfondissement et contrôle du modèle

Test bilatéral et choix après observation

Un test unilatéral réserve son risque de rejet à une direction choisie à l'avance. Le sélectionner après avoir vu le signe favorable du résultat utilise les données deux fois et modifie le risque réel de la procédure. Le choix doit correspondre à la question scientifique et aux conséquences d'un effet dans l'autre direction.

La frontière $p = 0,05$ ne sépare pas deux états radicalement différents de connaissance. Des valeurs de 0,049 et 0,051 peuvent provenir d'estimations presque identiques. Leur présentation doit conserver effet, intervalle et limites. Une décision opérationnelle peut exiger un seuil, mais la lecture scientifique reste continue et dépend du contexte. Cette distinction aide à éviter une surinterprétation de résultats situés juste de part et d'autre d'une convention.

Mise à l'épreuve : Non-significativité et équivalence répondent à deux questions

Problème. Une étude annonce une différence moyenne de 0,2 unité et un IC à 95 % $[-1 ; 1,4]$. Une marge d'équivalence avait été fixée avant l'étude à $[-0,5 ; 0,5]$. Peut-on conclure à l'absence de différence utile ?

Résolution. L'intervalle contient zéro, donc un test bilatéral usuel compatible avec cet intervalle ne rejette pas l'égalité au seuil de 5 %. Mais il contient aussi des différences dépassant les deux marges : l'équivalence n'est pas démontrée. Dans la procédure classique de deux tests unilatéraux à 5 %, on examine l'inclusion de l'IC à 90 % dans les marges ; l'IC à 95 % fourni ne permet pas à lui seul de réaliser cette procédure. Il faut la méthode, son intervalle et la marge préspecifiée. Choisir la marge après avoir vu les résultats transformerait la question.

Calculer une puissance dans un modèle normal explicite

Fixer la règle avant de regarder les données

Supposons une moyenne de n observations indépendantes normales, de variance connue σ^2 . On teste unilatéralement $H_0: \mu = \mu_0$ contre $\mu > \mu_0$. La règle de niveau α rejette si $\text{moyenne} > \mu_0 + z_{(1-\alpha)}\sigma/\sqrt{n}$. La puissance à une valeur alternative μ_1 est la probabilité de franchir ce seuil lorsque les données ont réellement moyenne μ_1 .

Après standardisation sous μ_1 , puissance = $1 - \Phi[z_{(1-\alpha)} - (\mu_1 - \mu_0)\sqrt{n}/\sigma]$. Cette expression montre que la puissance n'est pas une propriété du seul test : elle dépend de l'effet envisagé, de la dispersion, de n et du seuil. Une puissance élevée pour un grand effet n'assure pas une bonne détection d'un effet plus petit.

Exercice numérique

Prendre $\mu_0 = 0$, $\sigma = 10$, $n = 100$ et $\alpha = 0,05$ unilatéral, avec $z = 1,645$. Le seuil de rejet vaut 1,645. Pour $\mu_1 = 3$, la puissance vaut $1 - \Phi(-1,355) = 0,9123$ environ. Pour $\mu_1 = 1$, elle vaut $1 - \Phi(0,645) = 0,2595$. Le même protocole détecte donc ces deux alternatives avec des probabilités très différentes.

Déduire un effectif de planification

Pour viser une puissance $1 - \beta$ contre un effet δ positif, résoudre l'équation donne $n \approx [(z_{(1-\alpha)} + z_{(1-\beta)})\sigma/\delta]^2$. Avec une puissance de 80 %, $z_{(1-\beta)} = 0,842$ et $\delta = 3$, on obtient 68,72, donc au moins 69 observations dans ce modèle. L'arrondi se fait vers le haut. Le calcul ne tient pas compte ici de données manquantes, de variance estimée ni d'effet de plan.

Exercice critique. Peut-on choisir un test unilatéral après avoir observé le sens favorable ? **Non.** La règle de rejet aurait alors été choisie en utilisant les mêmes données, ce qui change son erreur de première espèce. Un test bilatéral prévu initialement ne se transforme pas après coup pour améliorer la significativité.

Cette construction sert à planifier selon un effet jugé pertinent et des hypothèses de variance explicites. Une « puissance observée » reconstruite à partir du seul effet estimé n'apporte pas une information indépendante de la statistique du test ; l'estimation et son intervalle sont plus directement interprétables après l'étude.



Schéma de raisonnement

Décision statistique

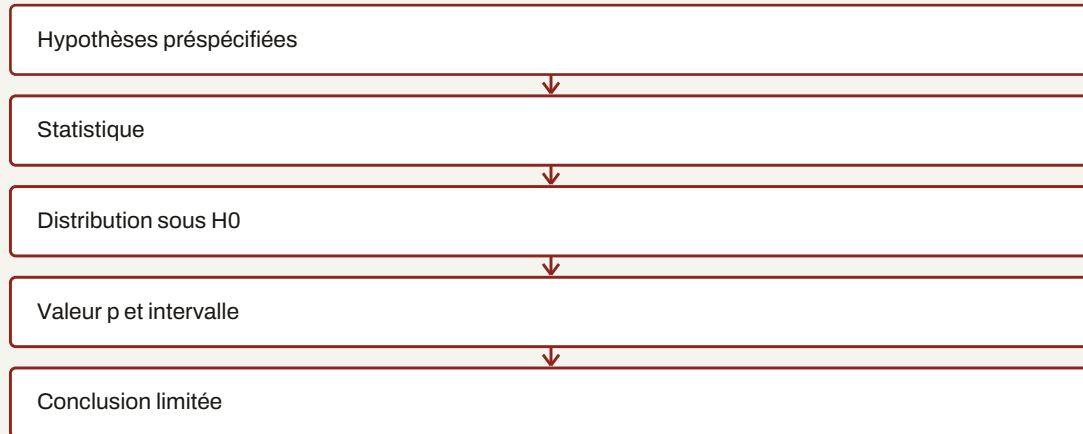


Schéma de principe original • lire les conditions et limites dans le texte

La conclusion porte sur un modèle et une procédure. Une décision statistique ne remplace ni l'évaluation du biais ni celle de l'importance clinique.

Exercices progressifs

1. Quelle est la puissance si $\beta = 0,20$ pour l'effet prévu ?
2. Un test donne $p = 0,08$ au seuil 5 %. Que conclure exactement ?
3. Pour cinq tests indépendants au seuil 5 %, estime le risque d'au moins un faux positif sous nullité globale.
4. Un seuil alpha de 1 % garantit-il 99 % de chances que chaque résultat positif soit vrai ?
5. Pour dix tests indépendants sous nullité globale au seuil 5 %, calculer le risque d'au moins un faux rejet.
6. Un intervalle $[-0,2 ; 0,3]$ est-il plus informatif qu'un intervalle $[-20 ; 30]$ pour discuter une différence proche de zéro ?

Corrections argumentées

1. 80 %. Cela décrit les répétitions de l'étude sous cette alternative, pas la probabilité que le résultat observé soit correct.
2. On ne rejette pas H_0 avec cette règle. On ne prouve ni l'absence d'effet ni l'équivalence ; il faut examiner l'intervalle et l'effectif.
3. $1 - 0,95^5 \approx 0,226$, soit 22,6 %. Une correction peut être nécessaire selon la famille de questions définie.
4. Non. Alpha décrit les faux rejets lorsque H_0 est vraie. La proportion de résultats vrais parmi les positifs dépend aussi de la fréquence des hypothèses vraies et de la puissance.



5. $1 - 0,95^{10} = 0,4013$, soit 40,13 %. La formule exacte donnée suppose l'indépendance.

6. Oui à unité identique, car il exclut davantage de grandes différences. Pour conclure formellement à l'équivalence, il faut toutefois des marges et une procédure appropriées.

Références et portée

- Penn State : cours de statistique, référence 1

Sources consultées le 1er octobre 2026 pour les notions signalées. Développements, situations et calculs pédagogiques OSSIK ; données fictives. La concordance complète avec les supports universitaires de l'année et la relecture scientifique humaine restent à réaliser. Les modèles biomédicaux ne fournissent aucune décision thérapeutique individuelle.

Partie 2 · Choisir une analyse selon le plan expérimental

Prérequis et objectifs

Savoir distinguer variable qualitative et quantitative, unités indépendantes et mesures répétées, estimation et test.

À l'issue de ce cours, vous devez pouvoir :

- Choisir l'unité et la comparaison adaptées.
- Construire des effectifs attendus.
- Reconnaître les limites des modèles de régression et de survie.

Fondements du thème

Choisir une comparaison

Pour deux groupes indépendants et une variable quantitative, un test t de Welch compare les moyennes sans supposer des variances égales, avec des hypothèses de distribution ou d'approximation à évaluer. Pour des mesures avant-après chez les mêmes personnes, l'analyse porte sur les différences individuelles : l'appariement change l'erreur standard. Un test non paramétrique n'est pas dépourvu d'hypothèses ; Mann-Whitney porte sur les distributions et ne devient un simple test de médianes que sous des conditions supplémentaires.

Pour deux variables qualitatives, un tableau de contingence et un test du chi-deux d'indépendance peuvent convenir si les effectifs attendus permettent l'approximation. Un test exact de Fisher est une option pour certains petits tableaux. Les effectifs attendus dépendent des marges : $E_{ij} = \text{total_ligne}_i \times \text{total_colonne}_j / n$. Une grande différence observé-attendu contribue au chi-deux via $(O - E)^2 / E$.

Régression et association

Une régression linéaire simple écrit $E(Y|X) = a + bX$ dans le modèle. Le coefficient b exprime la variation moyenne de Y pour une unité de X. L'interprétation suppose notamment une forme fonctionnelle adaptée ; les hypothèses d'erreurs interviennent pour l'inférence. La corrélation de Pearson mesure une liaison linéaire et peut être proche de zéro en présence d'une relation non linéaire forte.

Une régression logistique modélise le logarithme de la cote d'un événement binaire. L'exponentielle d'un coefficient donne un odds ratio conditionnel aux autres variables du modèle. Une régression ajustée ne garantit pas l'élimination de toute confusion : les variables manquantes, les erreurs de mesure et les mauvaises formes du modèle persistent. Prédire une issue et établir une cause sont deux objectifs différents.

Survie et censure

Une analyse de survie étudie le délai jusqu'à un événement défini, pas seulement le décès. La censure à droite signifie que le délai exact dépasse le temps observé : une personne encore sans événement en fin de suivi apporte de l'information. L'estimateur de Kaplan-Meier multiplie des probabilités de survie conditionnelles aux temps d'événement. Son interprétation usuelle demande une censure non informative compte tenu du cadre d'analyse.

Le risque instantané ou hazard décrit un taux conditionnel chez les personnes encore sans événement. Un hazard ratio n'est pas directement un risque relatif cumulé. Le modèle de Cox à risques proportionnels suppose un rapport de hazards constant dans sa forme standard. Un changement d'effet dans le temps peut contredire cette hypothèse.

Développements et raisonnements

Une question avant un nom de test

Comparer deux moyennes, comparer deux distributions et comparer deux probabilités ne sont pas la même cible. Pour une variable quantitative, la moyenne peut être scientifiquement pertinente même si la distribution est asymétrique. Une méthode robuste ou un modèle adapté peut alors être préférable à un changement implicite de question. Le choix ne doit pas se limiter à « normal ou non normal ».

L'indépendance dépend du plan. Deux mesures chez la même personne sont appariées ; deux groupes recrutés séparément peuvent être indépendants. Plusieurs mesures par personne ou plusieurs personnes par centre créent des structures plus complexes. Le traitement correct peut utiliser des différences, des modèles mixtes ou des erreurs standards adaptées, selon la question et le niveau attendu.

Tableau de contingence et attente sous indépendance

Sous indépendance, la proportion de la colonne j est la même dans chaque ligne. Multiplier le total de la ligne i par cette proportion donne $E_{ij} = L_i C_j / N$. La statistique de Pearson additionne $(O_{ij} - E_{ij})^2 / E_{ij}$. Les cellules de faible attente fragilisent l'approximation ; ce sont les effectifs attendus, pas seulement les observés, qu'il faut examiner.

Pour un tableau à r lignes et c colonnes, les degrés de liberté usuels valent $(r - 1)(c - 1)$, car les marges imposent des contraintes. Un résultat significatif indique une incompatibilité avec l'indépendance sous le modèle ; il ne mesure pas l'intensité de l'association. Il faut accompagner le test de proportions ou d'une mesure d'effet.

Régression et délai répondent à d'autres questions

Une corrélation quantifie une forme de liaison. Une régression précise une moyenne ou une probabilité conditionnelle. Ni l'une ni l'autre ne donne automatiquement un effet causal. Une variable de confusion peut expliquer une association ; un mauvais ajustement peut en créer une. Le rôle scientifique des variables doit guider leur inclusion.

Une variable de survie combine un temps et un statut d'événement. Supprimer les personnes censurées gaspille leur information ; les traiter comme des événements au temps de censure invente des événements. Une analyse correcte utilise leur présence dans l'ensemble à risque jusqu'à la fin de leur observation. Si la raison de censure est liée au pronostic, l'hypothèse d'une censure non informative doit être discutée.

Atelier de calcul et analyse de sensibilité

Un tableau compare deux groupes de 50 personnes. Dans le premier, 20 présentent l'événement et 30 non ; dans le second, 10 et 40. Les colonnes totalisent 30 événements et 70 non-événements. Sous indépendance, chaque ligne attend 15 événements et 35 non-événements. La statistique vaut $25/15 + 25/35 + 25/15 + 25/35 = 4,7619$. Avec un degré de liberté, la valeur p approximative est 0,0291, sans correction de continuité.

La différence de risques vaut pourtant simplement $40\% - 20\% = 20$ points. Le test et la mesure d'effet décrivent deux aspects distincts : compatibilité avec une hypothèse et amplitude observée. Un autre effectif peut changer fortement la valeur p pour les mêmes proportions.

Dans un plan avant-après, quatre différences individuelles valent 2, 4, 4 et 6 unités. Leur moyenne est 4 et leur variance corrigée $(4 + 0 + 0 + 4)/3 = 2,6667$. L'erreur standard de la moyenne des différences est $\sqrt{2,6667}/2 = 0,8165$. Le petit effectif impose une grande prudence sur les hypothèses de distribution ; il ne suffit pas de remplacer mécaniquement un test indépendant par une formule appariée.

Pour un délai, une personne sans événement après six mois apporte six mois d'observation. On sait que son délai dépasse six mois, pas qu'il vaut six mois et pas qu'elle n'aura jamais d'événement.

Approfondissement et contrôle du modèle

Pseudoréplication et unité de randomisation

Si une intervention est attribuée à des classes entières, les élèves ne sont pas tous des unités de randomisation indépendantes. Une analyse ignorant cette structure peut produire des erreurs standards trop petites. Le même problème apparaît lorsque plusieurs mesures techniques proviennent d'un seul prélèvement biologique : elles améliorent la mesure de ce prélèvement sans multiplier le nombre de personnes étudiées.

Le protocole doit donc distinguer unité de recrutement, unité d'attribution, unité de mesure et unité d'analyse. Elles peuvent coïncider, mais ce n'est pas obligatoire. Une méthode adaptée peut modéliser les niveaux ou agréger certaines mesures selon la question. Le nombre total de lignes d'un fichier n'est pas un substitut à cette description.

La réplication technique reste utile pour connaître une variabilité instrumentale. La réplication biologique informe une autre source de variation. Les confondre produit une précision apparente qui ne se transporte pas à la population cible.

Mise à l'épreuve : Des répétitions techniques ne créent pas des sujets

Problème. Douze personnes fournissent chacune un prélèvement, analysé dix fois. Un compte rendu annonce $n = 120$ pour comparer les personnes à un second groupe. Identifier l'unité et proposer une analyse de principe.

Résolution. Les 120 lectures comprennent douze unités biologiques, avec dix répétitions techniques dépendant du même prélèvement. Moyennées par personne, elles produisent douze valeurs réduisant une composante du bruit analytique. Une comparaison entre groupes peut alors utiliser ces résumés si la question le permet ; un modèle hiérarchique constitue une autre option pour séparer les variances.

Prendre 120 comme effectif indépendant sous-estimerait généralement l'incertitude. Si l'intervention était attribuée à des salles et non aux personnes, l'unité de randomisation demanderait encore un niveau supplémentaire. Le nombre de lignes du fichier ne définit jamais seul l'effectif informatif.

Schéma de raisonnement

Choisir une méthode

1. Question et effet • 2. Variable • 3. Dépendance des observations • 4. Hypothèses • 5. Estimation accompagnée

La liste ne remplace pas une analyse du plan. Certaines situations demandent une méthode hors de ces modèles élémentaires.

Exercices progressifs

1. Vingt sujets sont mesurés avant et après une intervention. Analyse indépendante ou appariée ?
2. Dans un tableau, total de ligne = 40, total de colonne = 30, $n = 100$. Effectif attendu sous indépendance ?
3. Une personne est sans événement à la fin du suivi. Faut-il la supprimer d'une analyse de survie ?
4. Un tableau 3×4 possède combien de degrés de liberté usuels pour le chi-deux d'indépendance ?
5. Deux prélèvements sur chacune de dix personnes constituent-ils vingt personnes indépendantes ?
6. Une corrélation nulle exclut-elle une relation en U ?

Corrections argumentées

1. Appariée : les deux mesures d'un sujet sont liées. On doit exploiter cette structure et examiner les différences individuelles.
2. $E = 40 \times 30/100 = 12$. C'est une attente sous le modèle nul, pas forcément un entier observé.
3. Non. Elle contribue jusqu'à sa censure. La supprimer ferait perdre de l'information et pourrait biaiser l'estimation ; le mécanisme de censure doit être discuté.
4. $(3 - 1)(4 - 1) = 6$. Les contraintes de marges expliquent ce nombre.
5. Non. Le nombre d'unités indépendantes dépend du recrutement et de la corrélation entre prélèvements.
6. Non. Pearson vise une liaison linéaire ; une relation symétrique en U peut produire une corrélation proche de zéro.

Références et portée

- Penn State : cours de statistique, référence 1

Sources consultées le 1er octobre 2026 pour les notions signalées. Développements, situations et calculs pédagogiques OSSIK ; données fictives. La concordance complète avec les supports universitaires de l'année et la relecture scientifique humaine restent à réaliser. Les modèles



biomédicaux ne fournissent aucune décision thérapeutique individuelle.

Partie 3 • Comparer des moyennes et des proportions

Prérequis et objectifs

Connaître variance d'une somme, erreur standard et Student. Fixer le sens de la différence avant tout calcul.

À l'issue de ce cours, vous devez pouvoir :

- Calculer une différence de moyennes avec Welch.
- Exploiter les différences appariées.
- Séparer erreur standard de test et d'intervalle pour les proportions.

Fondements du thème

Deux moyennes indépendantes

Deux groupes indépendants de 25 personnes ont des moyennes de 52 et 48 unités. Leurs écarts-types sont respectivement 10 et 8. On cherche d'abord une estimation de la différence $\mu_1 - \mu_2$, puis son incertitude. La différence observée est 4 unités ; ce nombre seul ne dit pas si la précision suffit à distinguer une différence de zéro.

Dans le test de Welch, l'erreur standard vaut $SE = \sqrt{s_1^2/n_1 + s_2^2/n_2}$. Ici, $SE = \sqrt{100/25 + 64/25} = \sqrt{6,56} \approx 2,56$. La statistique sous $H_0 : \mu_1 - \mu_2 = 0$ est $t = 4/2,56 \approx 1,56$. Les degrés de liberté de Welch ne sont pas automatiquement $n_1 + n_2 - 2$; leur calcul tient compte des deux variances et effectifs.

La formule est $v = (s_1^2/n_1 + s_2^2/n_2)^2 / [(s_1^2/n_1)/(n_1 - 1) + (s_2^2/n_2)/(n_2 - 1)]$. Dans cet exemple, v est proche de 45,8. Un quantile bilatéral à 95 % d'environ 2,01 donne un intervalle $4 \pm 2,01 \times 2,56$, soit approximativement $[-1,15 ; 9,15]$. Il contient zéro et des différences positives non négligeables : l'étude ne permet pas une conclusion précise d'absence d'effet.

Vérifier les conditions plutôt qu'appliquer un nom

Les observations doivent être indépendantes entre et dans les groupes pour cette formule usuelle. Si les 25 valeurs sont des prélèvements répétés sur cinq personnes, $n = 25$ ne représente pas 25 unités indépendantes. La forme des distributions et les valeurs extrêmes comptent, particulièrement avec de petits effectifs. Welch évite l'hypothèse de variances égales, mais ne résout pas tous les problèmes de données.

Une distribution très asymétrique peut nécessiter une transformation, une méthode robuste ou un autre modèle selon la question. Changer de méthode pour obtenir un résultat significatif ne constitue pas une justification. On choisit une cible : moyenne, médiane, distribution ou autre, puis une procédure cohérente. La méthode et les écarts au protocole doivent pouvoir être expliqués.

Comparaison appariée

Pour des mesures avant-après, définissons $d_i = \text{après}_i - \text{avant}_i$. Si la moyenne des différences est 4 unités, leur écart-type 5 et le nombre de sujets 25, $SE_d = 5/\sqrt{25} = 1$. Un intervalle de Student à 24 degrés de liberté utilise environ 2,064 et donne [1,94 ; 6,06]. Cette précision peut être meilleure que celle d'une analyse indépendante parce que les différences éliminent une partie de la variabilité stable entre individus.

L'avantage n'est pas automatique : il dépend de la corrélation et de la dispersion des différences. La normalité pertinente pour le test apparié exact concerne les différences, pas séparément les mesures avant et après. Une absence de groupe témoin laisse aussi des explications alternatives, comme l'évolution naturelle ou une régression vers la moyenne. Une différence avant-après significative n'établit pas seule l'efficacité causale d'une intervention.

Deux proportions

Dans deux groupes indépendants de 100 personnes, un événement concerne 30 personnes contre 15. La différence estimée est $0,30 - 0,15 = 0,15$. Un intervalle normal simple pour la différence utilise $SE = \sqrt{0,30 \times 0,70/100 + 0,15 \times 0,85/100} \approx 0,0581$. L'intervalle approximatif à 95 % est $0,15 \pm 0,114$, soit [0,036 ; 0,264], en proportion.

Pour un test d'égalité sous H_0 , une procédure classique estime la proportion commune par $p_{\text{pool}} = 45/200 = 0,225$. L'erreur standard nulle devient $\sqrt{0,225 \times 0,775 \times (1/100 + 1/100)} \approx 0,0591$. La statistique z vaut environ 2,54. Les erreurs standards d'estimation et de test ne sont donc pas nécessairement calculées de façon identique, parce qu'elles répondent à des hypothèses différentes.

Développements et raisonnements

Le dénominateur vient de la variance de l'estimateur

Si deux moyennes indépendantes estiment μ_1 et μ_2 , la variance de leur différence est la somme de leurs variances. On estime chacune par s^2/n , d'où racine($s_1^2/n_1 + s_2^2/n_2$). Le signe moins entre les moyennes ne crée pas un signe moins entre les variances, car le coefficient -1 est mis au carré. Une covariance serait nécessaire si les groupes étaient dépendants.

Welch estime les degrés de liberté à partir des contributions des deux variances. L'expression peut donner une valeur non entière, ce qui n'est pas une erreur. La méthode à variance commune utilise d'autres hypothèses et ne doit pas être substituée sans examiner leur plausibilité. Des tailles très différentes associées à des variances différentes peuvent rendre ce choix important.

L'appariement exploite une covariance

Pour $D = \text{après} - \text{avant}$, $\text{Var}(D) = \text{Var}(\text{après}) + \text{Var}(\text{avant}) - 2\text{Cov}(\text{avant}, \text{après})$. Une corrélation positive peut réduire fortement la variance des différences. Mais si l'appariement est artificiel ou si la corrélation est faible, le gain n'est pas garanti. Le test porte sur une moyenne de différences ; il ne dit pas que chaque individu évolue dans le même sens.

Une comparaison avant-après sans groupe témoin ne sépare pas l'intervention de l'évolution naturelle, de l'apprentissage du test ou de la régression vers la moyenne. La précision statistique ne résout pas cette limite de conception. Un effet temporel précis peut rester causalement ambigu.

Deux proportions et deux hypothèses de variance

Pour estimer $p_1 - p_2$, on peut utiliser séparément les proportions observées dans une approximation normale. Sous H_0 d'égalité, le test classique estime une probabilité commune en regroupant les succès et les effectifs. Le pool n'est donc pas une manipulation arbitraire ; il correspond à la contrainte de l'hypothèse nulle. Les intervalles plus performants pour petites proportions peuvent avoir une forme différente.

Comparaison	Estimation	Variabilité pertinente
Moyennes indépendantes	Moyenne 1 – moyenne 2	Deux variances de moyennes
Avant-après	Moyenne des différences	Variance des différences
Proportions indépendantes	$p_1 - p_2$	Variances binomiales
Données binaires appariées	Changement concordant ou discordant	Structure des paires

Des résultats binaires appariés ne relèvent pas du test de deux proportions indépendantes. Le nombre de paires discordantes contient l'information pertinente pour certaines comparaisons, ce qui illustre encore la priorité du plan.

Atelier de calcul et analyse de sensibilité

Deux groupes indépendants ont $n_1 = 40$, $n_2 = 60$, moyennes 15 et 12, écarts-types 6 et 5. La différence vaut 3 ; $SE = \text{racine}(36/40 + 25/60) = 1,1475$. Le score de Welch vaut 2,614. Les degrés de liberté sont environ 73,1. Un intervalle utilise le quantile correspondant, proche de 1,99, et donne environ [0,71 ; 5,29]. Les décimales exactes dépendent du quantile numérique, pas d'une convention consistant à fixer tous les tests à 1,96.

Pour 25 paires avec écart-type 10 avant et après et corrélation 0,80, la variance des différences vaut $100 + 100 - 160 = 40$. $SE_{\text{diff}} = \text{racine}(40/25) = 1,2649$. Ignorer l'appariement donnerait $\text{racine}(100/25 + 100/25) = 2,8284$. Cette comparaison montre le rôle de la corrélation, sous l'hypothèse que les paramètres annoncés décrivent bien les mêmes personnes.

Dans deux groupes de 200 personnes, 50 et 30 événements donnent $p_1 = 0,25$, $p_2 = 0,15$ et une différence de 0,10. L'erreur standard non groupée est $\text{racine}(0,25 \times 0,75/200 + 0,15 \times 0,85/200) = 0,03969$. Un intervalle normal approximatif vaut [0,0222 ; 0,1778]. Le test groupé emploie $p_{\text{pool}} = 0,20$ et $SE_0 = 0,04$, donc $z = 2,5$. Les deux dénominateurs sont proches ici, sans être identiques.

Approfondissement et contrôle du modèle

Critère principal et comparaisons secondaires

Comparer plusieurs moyennes, plusieurs temps et plusieurs sous-groupes multiplie les occasions de conclure. Une différence de deux groupes à un temps choisi après inspection ne possède pas la même interprétation qu'une comparaison prévue dans le protocole. Les estimations exploratoires peuvent être montrées, mais leur statut et la multiplicité doivent rester visibles.

Pour plusieurs mesures répétées, calculer un test indépendant à chaque temps ignore en outre la corrélation temporelle et ne répond pas directement à une question de trajectoire. Une analyse peut

viser une différence à un temps fixé, une aire sous la courbe ou une évolution globale. Ces cibles ne sont pas interchangeables. Définir celle qui importe avant le calcul évite de changer d'objectif en fonction du résultat le plus favorable.

Mise à l'épreuve : L'appariement peut changer la précision

Problème. Avant intervention, trois sujets ont 10, 20 et 30 unités. Comparer les mesures après A = (11 ; 21 ; 31) et B = (31 ; 21 ; 11), appariées dans l'ordre. Les moyennes avant/après diffèrent-elles entre scénarios ?

Résolution. La moyenne avant vaut 20 et la moyenne après 21 dans les deux scénarios. Les différences individuelles sont cependant A : (1 ; 1 ; 1), B : (21 ; 1 ; -19). Leur moyenne vaut 1 dans les deux cas, mais leur écart-type estimé est respectivement 0 et 20. Le second donne une erreur standard de $20/\sqrt{3} = 11,55$. Le premier exemple, volontairement dégénéré, ne justifie pas un test t calculé avec un dénominateur nul ; il sert à isoler l'effet des appariements. Les statistiques marginales ne suffisent pas à reconstruire une analyse appariée.

Données binaires appariées : les discordances portent le changement

Construire le tableau des mêmes personnes

Pour cent personnes mesurées avant et après, on observe quinze positives aux deux dates, vingt positives seulement avant, cinq positives seulement après et soixante négatives aux deux dates. Les marges sont 35 % de positifs avant et 20 % après, soit une diminution observée de quinze points. Les deux pourcentages ne sont pas indépendants : chaque personne contribue aux deux dates.

Les paires concordantes n'informent pas le sens du changement dans le test de symétrie des discordances. Les vingt passages positif→négatif et les cinq passages inverse constituent vingt-cinq discordances. Sous l'hypothèse nulle de probabilités de transition égales, conditionnellement au nombre de discordances, le nombre dans un sens suit une binomiale de paramètres 25 et 0,5, si les personnes sont indépendantes.

Exercice : test exact conditionnel

Calculer la probabilité bilatérale symétrique d'un déséquilibre au moins aussi extrême que 20 contre 5. On additionne $P(X \leq 5)$ et $P(X \geq 20)$, égales sous $p=0,5$. Ainsi $p_{\text{exact}} = 2 \sum_{k=0}^5 [C(25,k)/2^{25}]$, soit environ 0,00408. Ce calcul compare les deux directions du changement ; il ne teste pas l'indépendance des mesures avant et après.

L'approximation de McNemar sans correction utilise $(b-c)^2/(b+c)$, ici $225/25=9$. Une correction de continuité conduit à $(|b-c|-1)^2/(b+c)=196/25=7,84$. Les valeurs p correspondantes diffèrent du calcul exact ; il faut annoncer la méthode plutôt que mélanger une statistique approchée et une probabilité exacte.

Exercice : ce que les marges ne racontent pas

Peut-on reconstituer le nombre de discordances en connaissant seulement 35 % avant et 20 % après ?
Non. Les marges imposent une différence entre les deux types de transition, mais pas leur somme.
 Plusieurs tableaux appariés ont les mêmes marges et des incertitudes différentes. Le protocole doit conserver le lien entre les observations d'une même personne.

Le test de changement ne prouve toujours pas l'effet causal d'une intervention en l'absence de comparateur approprié. Une saison, une évolution spontanée ou une modification de mesure peuvent influencer les transitions. Les valeurs binaires doivent aussi avoir la même définition aux deux dates. Les paires incomplètes demandent une analyse du mécanisme des données manquantes ; les supprimer ne garantit pas une comparaison non biaisée.

Schéma de raisonnement

Comparer deux groupes

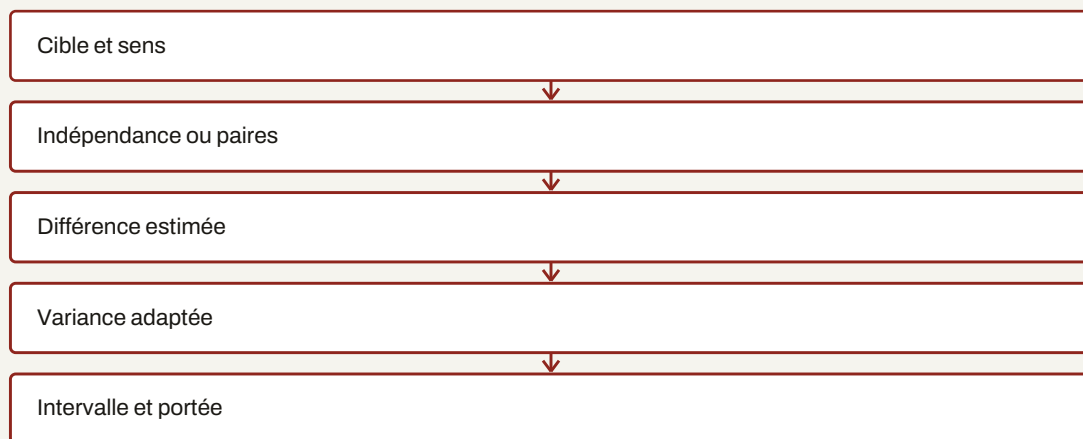


Schéma de principe original • lire les conditions et limites dans le texte

L'accord numérique de deux méthodes dans un exemple ne les rend pas interchangeables. Les conditions doivent être réévaluées pour un nouveau jeu de données.

Exercices progressifs

1. Dans une comparaison appariée, quelle variable résumer avant d'appliquer le test t ?
2. Pourquoi le dénominateur du test de Welch additionne-t-il s_1^2/n_1 et s_2^2/n_2 ?
3. Une différence de proportion de 0,15 vaut-elle 15 % ou 15 points de pourcentage ?
4. Deux moyennes indépendantes ont chacune une erreur standard de 2. Erreur standard de leur différence ?
5. Pourquoi l'écart-type des différences ne se calcule-t-il pas par $s_{\text{après}} - s_{\text{avant}}$?
6. Pour 20/100 contre 30/100, donner différence de risques et proportion groupée.

Corrections argumentées

1. Les différences individuelles calculées dans un sens constant. Les inverser change le signe de l'effet, pas la valeur absolue du test bilatéral.
2. Ce sont les variances estimées des moyennes des deux groupes ; elles s'additionnent pour leur différence sous indépendance. Avec dépendance, un terme de covariance intervient.
3. Elle vaut 15 points de pourcentage. Une variation relative demande de diviser cette différence par une proportion de référence précisée.
4. $\text{racine}(2^2 + 2^2) = 2,828$ sous indépendance.
5. La variance dépend des deux variances et de leur covariance. Soustraire les écarts-types pourrait même donner une dispersion négative.
6. Différence dans le sens premier moins second : $-0,10$, soit -10 points. La proportion groupée est $50/200 = 0,25$.

Références et portée

- Penn State, STAT 200, Lesson 9

Sources consultées le 1er octobre 2026 pour les notions signalées. Développements, situations et calculs pédagogiques OSSIK ; données fictives. La concordance complète avec les supports universitaires de l'année et la relecture scientifique humaine restent à réaliser. Les modèles biomédicaux ne fournissent aucune décision thérapeutique individuelle.

Partie 4 • Régression multivariée et données de survie

Prérequis et objectifs

Connaître fonction affine, logarithme des cotes, conditionnement et notion de censure.

À l'issue de ce cours, vous devez pouvoir :

- Interpréter coefficients et résidus.
- Transformer une prédiction logistique en probabilité.
- Calculer Kaplan-Meier en tenant compte des censures.

Fondements du thème

Régression linéaire pas à pas

Un modèle pédagogique prédit $Y = 20 + 0,5X$. Pour $X = 30$, la moyenne prédite vaut 35. Une observation réelle $Y = 38$ possède un résidu de 3, différence entre observé et prédit. Les résidus servent à examiner le modèle : courbure, dispersion variant avec X ou points extrêmes peuvent indiquer qu'une hypothèse est inadéquate.

L'ordonnée à l'origine 20 n'a pas nécessairement une interprétation pratique si $X = 0$ est en dehors du domaine observé. Extrapoler loin des données suppose que la relation demeure valable, ce qui n'est pas démontré par un bon ajustement local. R^2 décrit la part de variabilité expliquée dans un cadre précis ; il ne prouve pas la causalité et ne mesure pas directement l'erreur d'une prédiction future. La validation prédictive exige des données distinctes ou une procédure de rééchantillonnage adaptée.

Régression logistique et probabilités

Si un modèle logistique donne $\log[p/(1 - p)] = -2 + 0,7X$, alors pour $X = 0$ la cote vaut $\exp(-2) \approx 0,1353$ et $p = 0,1353/(1 + 0,1353) \approx 0,119$. Pour $X = 1$, la cote vaut $\exp(-1,3) \approx 0,2725$ et $p \approx 0,214$. L'odds ratio associé à une unité de X vaut $\exp(0,7) \approx 2,01$, tandis que le rapport des deux probabilités vaut environ 1,80.

Cette différence concrète montre pourquoi un coefficient logistique exponentié ne doit pas être annoncé comme un risque relatif. Dans un modèle avec plusieurs variables, il décrit une association conditionnelle à celles-ci dans le modèle. Une variable omise importante peut laisser de la confusion. Des données rares peuvent aussi rendre les estimations instables, même si le logiciel affiche un nombre.

Kaplan-Meier avec censure

Supposons cinq personnes observées depuis le même départ. Un événement survient au temps 1 parmi les cinq à risque : la survie estimée devient $4/5 = 0,80$. Une personne est censurée au temps 2 sans événement. Au temps 3, il reste trois personnes à risque ; un événement fait multiplier la survie par $2/3$. L'estimation devient $0,80 \times 2/3 = 0,533$.

La censure n'entraîne pas de chute directe de la courbe ; elle réduit l'ensemble à risque pour les événements suivants. Compter le second événement sur quatre au lieu de trois ignorerait la censure. Mais la méthode suppose que le mécanisme de censure soit compatible avec l'estimation : si les

personnes les plus malades quittent systématiquement le suivi, une lecture naïve peut surestimer la survie. Le nombre encore à risque aide à juger la fragilité de la fin de courbe.

Développements et raisonnements

Ajuster une relation sans changer silencieusement de question

Dans $Y = a + bX + cZ + \text{erreur}$, le coefficient b décrit la variation moyenne de Y pour une unité de X à Z fixé dans ce modèle. Il ne correspond pas nécessairement à la différence brute entre personnes de X différents. L'interprétation exige un domaine où les combinaisons de X et Z sont effectivement observées ; une extrapolation peut demander des situations absentes des données.

Un résidu est observé moins prédit. Sa moyenne peut être nulle par construction dans une régression avec constante, sans que le modèle soit bon. Une courbure des résidus, une dispersion croissante ou quelques observations influentes peuvent signaler une limite. Le R^2 résume une part de variance dans un cadre donné ; il ne fournit pas directement la qualité de prédiction sur une nouvelle population.

Logistique : les coefficients agissent sur les cotes

Le modèle $\text{logit}(p) = \eta$ implique $p = 1/(1 + \exp(-\eta))$. Ajouter b au prédicteur multiplie la cote par $\exp(b)$. La variation absolue de probabilité dépend du niveau de départ. Le même odds ratio peut donc correspondre à des changements absolus très différents. Les odds ratios ajustés possèdent aussi des propriétés de non-collapsibilité : une différence entre OR brut et ajusté ne démontre pas à elle seule une confusion.

Une interaction b_{XZ} autorise l'association de X à varier avec Z . Si elle est présente, le coefficient de X seul correspond à $Z = 0$ dans le codage adopté. Centrer Z peut rendre cette référence plus pertinente sans changer les prédictions du modèle correctement réécrit. Les unités et catégories de référence doivent accompagner les coefficients.

Construire la survie comme produit de probabilités conditionnelles

Aux temps d'événement t_j , Kaplan-Meier multiplie $(1 - d_j/n_j)$, où n_j est l'effectif encore à risque juste avant ce temps et d_j le nombre d'événements. Les censures retirent les personnes des ensembles futurs sans provoquer de chute immédiate. Une médiane de survie n'est estimable que si la courbe atteint 0,5 dans le suivi observé.

Le hazard est un taux conditionnel parmi les personnes encore sans événement. Un hazard ratio constant ne signifie pas que le risque cumulé soit multiplié par cette même valeur à toute date. Si des événements concurrents empêchent l'événement d'intérêt, les traiter comme une simple censure peut répondre à une autre question que la probabilité réelle d'observer cet événement. L'objectif de l'analyse doit rester explicite.

Atelier de calcul et analyse de sensibilité

Un modèle linéaire est $Y = 10 + 2X - 3Z$, avec Z codé 0 ou 1. À $X = 4$, les prédictions sont 18 pour $Z = 0$ et 15 pour $Z = 1$. Une observation $Y = 17$ dans le second groupe possède un résidu de +2. Passer de $X = 4$ à $X = 5$ augmente la moyenne prédite de 2 dans les deux groupes, car ce modèle ne contient pas d'interaction.

Dans un modèle logistique, une cote initiale de 0,10 correspond à $p = 9,09\%$. Un OR de 3 donne une cote de 0,30 et $p = 23,08\%$. Avec une cote initiale de 1, le même OR conduit de 50 % à 75 %. L'OR

identique ne produit donc ni un triplement uniforme de probabilité ni un gain absolu constant.

Huit personnes commencent un suivi. Au temps 1, deux événements surviennent : $S = 6/8 = 0,75$. Une personne est censurée au temps 2 ; il reste cinq personnes à risque. Au temps 3, un événement donne $S = 0,75 \times 4/5 = 0,60$. Au temps 4, deux censures laissent deux personnes à risque sans changer S . Au temps 5, un événement donne $S = 0,60 \times 1/2 = 0,30$. La médiane estimée est le temps 5, première chute à 0,5 ou moins. La fin de la courbe repose sur peu de personnes et possède une forte incertitude.

Approfondissement et contrôle du modèle

Prédiction, calibration et validation

Un modèle peut bien classer les personnes du risque le plus faible au plus élevé tout en surestimant toutes les probabilités. La discrimination et la calibration sont deux propriétés distinctes. La première concerne l'ordre ou la séparation ; la seconde l'accord entre probabilités annoncées et fréquences observées dans des groupes comparables.

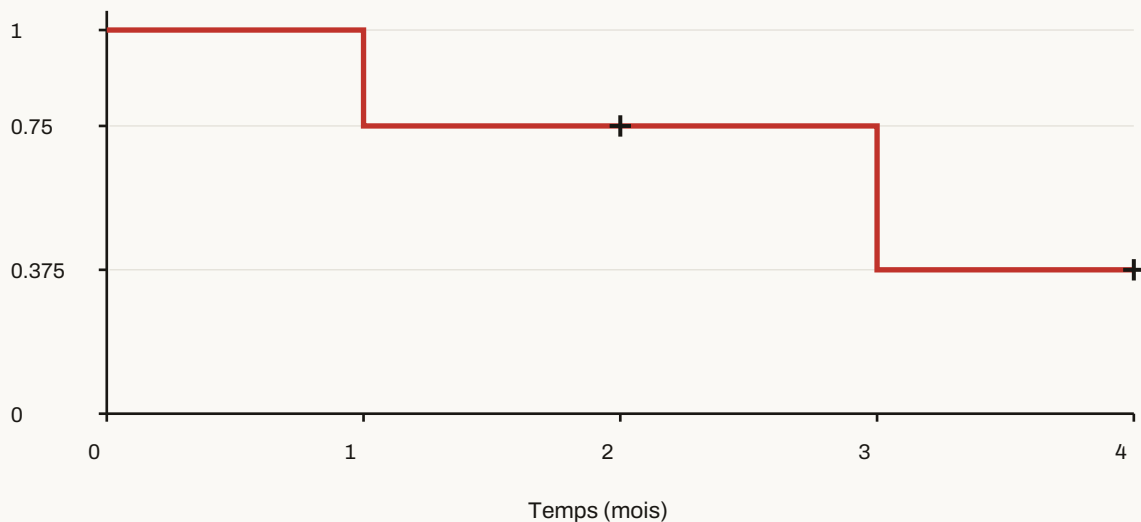
Évaluer le modèle sur les données qui ont servi à le construire peut donner un résultat optimiste, surtout si de nombreuses variables ont été essayées. Une validation séparée ou un rééchantillonnage adapté examine mieux cette performance. Un changement de population peut nécessiter une réévaluation de la calibration même lorsque les mécanismes biologiques restent plausibles.

Une bonne prédiction ne démontre pas qu'intervenir sur un prédicteur change le résultat. Un marqueur peut prédire parce qu'il reflète une cause commune ou une conséquence précoce. Le projet causal demande une autre définition des variables et d'autres hypothèses.

Lecture visuelle : Kaplan-Meier : une censure ne crée pas une chute

Kaplan-Meier : une censure ne crée pas une chute

4 personnes ; événements à 1 et 3 ; + = censures à 2 et 4 mois
Survie estimée $S(t)$



Quatre personnes sont suivies : événements à 1 et 3 mois, censures à 2 et 4 mois (croix). S passe de 1 à $3/4$ au premier événement. Après la censure, deux personnes restent à risque au second événement : $S = (3/4)(1/2) = 3/8$. Les segments sont horizontaux entre événements.

Question de lecture. Pourquoi aucune marche descendante ne se produit-elle à 2 mois ?

Correction. Une censure retire une personne du groupe encore observé ; ce n'est pas un événement. Elle modifie les dénominateurs ultérieurs, sous l'hypothèse de censure compatible avec l'estimation.

Mise à l'épreuve : Une censure n'est ni un événement ni une survie certaine

Problème. Quatre personnes sont suivies : événement à 1 mois, censure à 2 mois, événement à 3 mois, dernière personne encore suivie à 4 mois. Calculer Kaplan-Meier avant 4 mois, sans ex æquo.

Résolution. À 1 mois, quatre personnes sont à risque : $S = 1 \times (1 - 1/4) = 0,75$. La censure à 2 mois ne crée pas de chute, mais laisse deux personnes à risque à 3 mois. Le nouvel événement donne $S = 0,75 \times (1 - 1/2) = 0,375$. La proportion brute de personnes sans événement observé, $2/4$, ne donne pas cette estimation. La personne censurée contribue tant qu'elle est observée ; on ne lui impute pas une survie jusqu'au bout. L'interprétation suppose que la censure soit non informative sous les conditions du modèle, hypothèse impossible à vérifier avec ce seul tableau.



Un odds ratio constant ne donne pas une différence de risque constante

Revenir de la cote à la probabilité

La cote d'un événement de probabilité p est $p/(1-p)$. Si un modèle logistique multiplie cette cote par OR, la nouvelle probabilité est $p_1 = \text{OR} \times p_0 / [1 - p_0 + \text{OR} \times p_0]$. Cette relation est une transformation algébrique, pas une affirmation de causalité. Le coefficient logistique vaut $\ln(\text{OR})$ pour une variation d'une unité du prédicteur dans le modèle considéré.

Une probabilité et une cote n'ont pas la même borne : la première reste au plus égale à un ; la seconde peut devenir arbitrairement grande. Assimiler OR à un risque relatif peut donc produire une interprétation fortement exagérée lorsque l'événement est fréquent.

Exercice à deux risques initiaux

Pour $\text{OR}=2$, calculer p_1 lorsque $p_0=0,10$ puis lorsque $p_0=0,50$. Comparer risques relatifs et différences absolues.

Correction. Premier cas : $p_1=0,20/1,10=0,18182$; $\text{RR}=1,8182$ et différence= $0,08182$, soit 8,18 points. Second cas : $p_1=1/1,5=0,66667$; $\text{RR}=1,3333$ et différence= $0,16667$, soit 16,67 points. Un même OR ne donne donc ni le même RR ni le même écart absolu. Pour p_0 très petit, l'approximation $\text{OR} \approx \text{RR}$ devient meilleure, mais doit rester contextualisée.

Survie : risque instantané et risque cumulé

Dans un modèle de risques proportionnels, un hazard ratio compare des taux instantanés conditionnels à l'absence d'événement jusque-là. Il n'est pas un rapport constant de probabilités cumulées. Si un modèle exponentiel fictif possède des taux 0,10 et 0,20 par an, le hazard ratio vaut deux. À cinq ans, les probabilités d'événement sont $1 - e^{-0,5} = 0,39347$ et $1 - e^{-1} = 0,63212$; leur rapport vaut 1,6065 et non deux.

Exercice critique. Pourquoi le ratio des risques cumulés se rapproche-t-il de deux pour de très petites durées ? Parce que $1 - e^{(-\lambda t)} \approx \lambda t$ lorsque λt est petit. La limite vient d'une approximation locale. Censure informative, risques concurrents et variations temporelles du taux demandent d'autres modèles ; ces exemples ne constituent aucune prévision individuelle.

Schéma de raisonnement

Lire un modèle

1. Population et variables • 2. Paramètres • 3. Prédiction ou association • 4. Diagnostics • 5. Domaine de validité

Un modèle ajusté reste conditionnel à des hypothèses. Le schéma ne transforme pas une association en intervention causale.

Exercices progressifs

1. Dans le modèle $Y = 20 + 0,5X$, quel effet prédit une augmentation de X de 4 unités ?
2. Une cote de 3 correspond à quelle probabilité ?
3. Une censure fait-elle immédiatement diminuer la survie estimée de Kaplan-Meier ?
4. Un coefficient logistique vaut $\ln 2$. Que multiplie-t-il par deux ?
5. Une courbe Kaplan-Meier reste au-dessus de 0,5 jusqu'à la fin du suivi. Peut-on annoncer une médiane observée ?
6. Avec $\eta = 0$, quelle probabilité donne le modèle logistique ?

Corrections argumentées

1. La moyenne prédite augmente de 2 unités, dans le domaine où le modèle est valable. Cela n'établit pas qu'agir sur X provoquerait nécessairement cette variation.
2. $p = \text{cote}/(1 + \text{cote}) = 3/4 = 0,75$. Une cote n'est pas elle-même une probabilité et peut dépasser 1.
3. Non. Elle modifie les effectifs exposés au risque ensuite. La courbe baisse aux temps d'événements, selon les règles de l'estimateur.
4. La cote pour une unité de la variable, aux autres variables fixées dans le modèle. La probabilité n'est pas nécessairement doublée.
5. Non, la médiane n'est pas atteinte dans le suivi observé. L'extrapoler demanderait un modèle supplémentaire clairement annoncé.
6. $p = 1/(1 + \exp(0)) = 0,5$. Le logit nul correspond à des cotes égales à un.

Références et portée

- Penn State : cours de statistique, référence 1
- Penn State : cours de statistique, référence 2

Sources consultées le 1er octobre 2026 pour les notions signalées. Développements, situations et calculs pédagogiques OSSIK ; données fictives. La concordance complète avec les supports universitaires de l'année et la relecture scientifique humaine restent à réaliser. Les modèles biomédicaux ne fournissent aucune décision thérapeutique individuelle.