9. Encore des tests d'hypothèses...

t-tests, testing $\mu_1 = \mu_2$, and randomized experiments

Sam Gyetvay

ECO 2273 – Économetrie I

21 novembre 2025

Aujourd'hui

Tests de t

Tests d'égalité des moyennes

Expériences randomisées

Vers

Test de z

Jusqu'à présent, nous avons utilisé la distribution normale standard $z \sim N(0, 1)$ pour sélectionner les valeurs critiques pour les tests d'hypothèses

- Pour un test bilatéral à un niveau de signification de 95%, nous utilisons $z_{0.975} = .196 \approx 2$
- Sous l'hypothèse nulle, la statistique de test $\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$
- Nous rejetons $H_0 : \mu = \mu_0$ si $\left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{0.975}$

Cela est parfois appelé un test de z

Lorsque nous effectuons un test de z, nous faisons implicitement l'hypothèse que σ, l'écart-type de X, est un nombre fixe et connu

Estimation de σ

En réalité, σ est presque toujours inconnu

- La moyenne de la population μ est la valeur que nous obtiendrions si nous échantillonnions 100% de la population et calculions la moyenne
- L'écart-type de la population σ est la valeur que nous obtiendrions si nous échantillonnions 100% de la population et calculions l'écart-type

Lorsque σ n'est pas connu, il doit être estimé à partir des données

- Tout comme nous avons utilisé $\bar{X}$ comme notre estimation de μ...

Notre estimation est

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Statistique de test lorsque σ est inconnu

Lorsque nous utilisons s au lieu de σ, notre statistique de test devient

$$\frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Maintenant, le numérateur *et* le dénominateur contiennent des nombres que nous avons estimés à partir des données. Cela signifie qu'ils dépendent de l'échantillon, et donc sont des variables aléatoires

Il s'avère que nous connaissons la distribution de la statistique de test dans ce cas. C'est la distribution **student-t** avec $n-1$ **degrés de liberté**

Rappel : Variables aléatoires chi-carré

Une variable aléatoire chi-carré $X \sim \chi^2_k$ provient de la somme des carrés de k normales standard indépendantes :

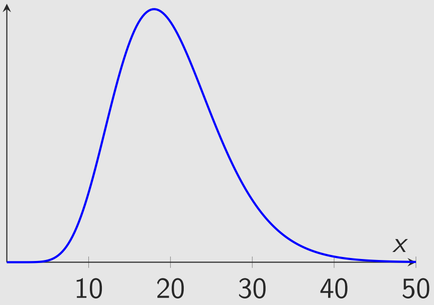
$$X = Z_1^2 + Z_2^2 + \cdots + Z_k^2, \quad Z_i \sim N(0, 1)$$

- Toujours non-négative : $X \geq 0$
- Asymétrie positive, surtout pour les petits k
- Moyenne $E[X] = k$, variance $Var(X) = 2k$

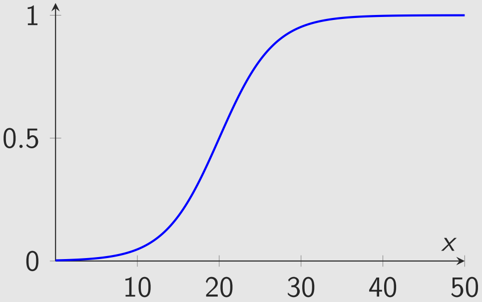
Nous appelons χ^2_k le “chi-carré avec k degrés de liberté”

PDF et CDF de χ^2_{20} (lisse)

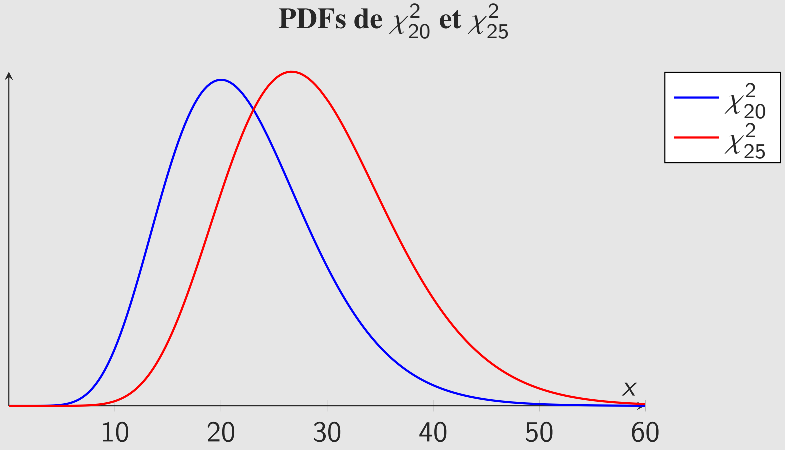
PDF de $X \sim \chi^2_{20}$



CDF de $X \sim \chi^2_{20}$



PDFs de χ^2_{20} vs χ^2_{25} (visuel)



Sommes de normales au carré décentrées

Lorsque nous prenons la somme de k normales indépendantes au carré décentrées

$$X = (Z_1 - \bar{Z})^2 + (Z_2 - \bar{Z})^2 + \cdots + (Z_k - \bar{Z})^2, \quad Z_i \sim N(0,1)$$

où $\bar{Z} = \frac{1}{n}\sum_i Z_i$, la variable résultante X est χ^2_{k-1}

Pourquoi ? Parce qu'une fois que nous connaissons les valeurs de Z_1, Z_2, ..., Z_{k-1}, et $\bar{Z}$, nous pouvons parfaitement déterminer Z_k

$$Z_k = n\bar{Z} - \sum_{i=1}^{k-1} Z_i = \sum_{i=1}^{k} Z_i - \sum_{i=1}^{k-1} Z_i$$

Par conséquent, il s'agit seulement de la somme des carrés de $k - 1$ normales standard indépendantes. La k-ième n'est pas indépendante des autres.

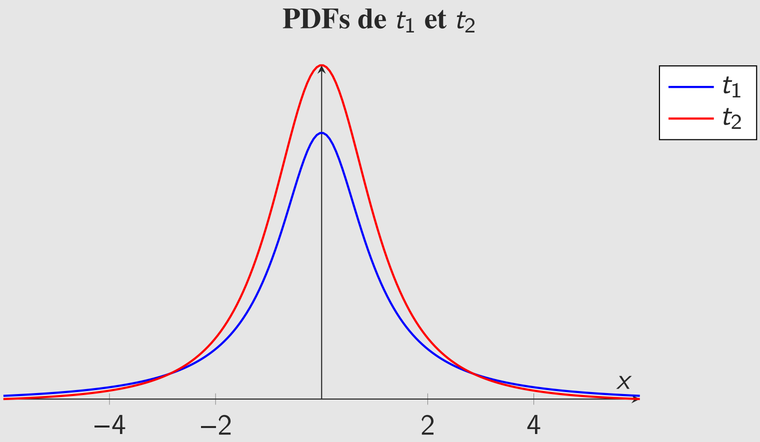
Rappel : Variables aléatoires Student-t

Si $X \sim N(0,1)$ et $Y \sim \chi^2_d$, alors

$$W = \frac{X}{\sqrt{Y/d}} \sim t_d$$

Nous appelons t_d la "distribution t de Student"

PDFs de Student-t t_1 vs t_2



Pourquoi la statistique de test suit-elle une distribution de Student-t ?

Affirmation : $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} \sim t_{n-1}$

Preuve :

$$t = \frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{\bar{X}-\mu}{s/\sqrt{n}} \cdot \frac{\sigma}{\sigma} = \frac{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)}{\sqrt{\frac{(n-1)s^2}{\sigma^2} \cdot \frac{1}{n-1}}}$$

Puisque $s^2 = \frac{1}{n-1}\sum_i (X_i - \bar{X})^2$, $\frac{(n-1)s^2}{\sigma^2} = \sum_i \frac{(X_i-\bar{X})}{\sigma}\frac{(X_i-\bar{X})}{\sigma}$ est une somme de normales standard au carré avec une contrainte linéaire ($\bar{X} = \frac{1}{n}\sum X_i$), ce qui donne χ^2_{n-1}. Par conséquent, le numérateur est $N(0,1)$ et le dénominateur est $\sqrt{\chi^2_{n-1}/(n-1)}$, comme requis. $\square$

Quantiles de t_{n-1}

La seule différence entre un test-t et un test-z est que nous utilisons les quantiles de la distribution de Student-t, au lieu des quantiles de la distribution normale standard. Cependant, pour de grands échantillons n, ces derniers sont pratiquement identiques

n	$\mathbf{t_{0.975,n-1}}$	$\mathbf{t_{0.95,n-1}}$
10	2.228	1.812
100	1.984	1.660
500	1.965	1.648
1000	1.962	1.646
5000	1.9605	1.6455
10000	1.9603	1.6453

Rappelez-vous, $z_{0.975} = 1.96$, $z_{0.95} = 1.645$. Ainsi, les tests sont presque identiques pour de grands échantillons.

Lors de l'examen, je vous fournirai tous les quantiles pertinents. Pas besoin de tables

Tests t à deux côtés : la recette

Pour réaliser un test t **à deux côtés** au niveau de **significativité de 5%**

Étape 1 : Énoncer l'**hypothèse nulle**

$$H_0 : \mu = \mu_0$$

Étape 2 : former la **statistique de test**

$$t = \frac{\bar{X} - mu_0}{s/\sqrt{n}}$$

Étape 3 : comparer la statistique de test avec la **valeur critique** $t_{0.975,n-1}$

- Si $|t| > t_{0.975,n-1}$, **rejeter** l'hypothèse nulle au niveau de significativité de 5%
- Si $|t| \leq t_{0.975,n-1}$, **ne pas rejeter** l'hypothèse nulle au niveau de significativité de 5%

Tests t à un côté : la recette

Pour réaliser un test d'**hypothèse à un côté** au niveau de **significativité de 5%**

Étape 1 : Énoncer l'**hypothèse nulle**

$$H_0 : \mu \leq \mu_0$$

Étape 2 : former la **statistique de test**

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Étape 3 : comparer la statistique de test avec la **valeur critique** $t_{0.95,n-1}$

- Si $t > t_{0.95,n-1}$, **rejeter** l'hypothèse nulle au niveau de significativité de 5%
- Si $t \leq t_{0.95,n-1}$, **ne pas rejeter** l'hypothèse nulle au niveau de significativité de 5%

Problèmes pratiques

1. Jacques a interrogé un échantillon aléatoire de $n = 100$ travailleurs à Montréal sur leur revenu familial. Dans l'échantillon de Jacques, le revenu moyen était de 36 000$ et l'écart-type était de 4 500$. Tester l'hypothèse que le revenu moyen à Montréal est d'au moins 40 000$ au niveau de significativité de 5%.

2. Jean a interrogé un échantillon aléatoire de $n = 10$ étudiants à l'UQAM sur leurs notes. Dans l'échantillon de Jean, la note moyenne était de 72% et l'écart-type était de 8%. Tester l'hypothèse que la note moyenne à l'UQAM est de 70.

3. Jermaine a interrogé un échantillon aléatoire de $n = 1000$ résidents de Montréal sur leur loyer. Dans l'échantillon de Jermaine, le loyer moyen payé était de 1210$ et l'écart-type était de 250$. Tester l'hypothèse que le loyer moyen est au plus de 1200$.

Tests de l'(in)égalité de deux moyennes

Parfois, nous voulons utiliser notre ensemble de données pour calculer deux moyennes différentes, $\bar{X}_1$ et $\bar{X}_2$, représentant généralement différents sous-groupes ou périodes de temps, puis tester des hypothèses sur la relation entre elles

Par exemple :

1. Supposons que nous utilisions un échantillon de travailleurs canadiens pour calculer les salaires moyens des travailleurs à Toronto et à Montréal. Sont-ils payés le même salaire ?
2. Supposons que nous utilisions un échantillon de travailleurs canadiens pour calculer les salaires moyens des travailleurs en 2024 et 2025. Les salaires ont-ils augmenté ?

Dans ces cas, l'hypothèse nulle implique deux moyennes de population inconnues

1. $H_0 : \mu_{TOR} = \mu_{MTL}$, où μ_{TOR} est le salaire moyen réel à Toronto et μ_{MTL} est le salaire moyen réel à Montréal
2. $H_0 : \mu_{2025} > \mu_{2024}$, où μ_{2025} est le salaire moyen réel en 2025 et μ_{2024} est le salaire moyen réel en 2024

Statistique de test pour un test bilatéral d'égalité

Pour un test de l'hypothèse nulle $H_0 : \mu_1 = \mu_2$, la statistique de test que nous utiliserons dans ce cours est

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

où

- $\bar{X}_1$ et $\bar{X}_2$ sont les moyennes des échantillons dans le groupe 1 et le groupe 2
- s_1^2 et s_2^2 sont les variances des échantillons dans le groupe 1 et le groupe 2
- n_1 et n_2 sont les nombres d'observations dans le groupe 1 et le groupe 2

Degrés de liberté ?

Puisque nous utilisons le test t, nous devons connaître les degrés de liberté pour savoir quelle valeur critique utiliser

Malheureusement, les formules pour les degrés de liberté sont laides et complexes. Par exemple, la formule de Welch-Satterthwaite, à la p.562 du manuel de Anderson et al., est

$$n_{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1} + \frac{\left(s_2^2/n_2\right)^2}{n_2-1}}.$$

Je n'ai jamais utilisé cette formule dans ma vie, donc je ne vous obligerai pas à l'apprendre

Pour nous simplifier la vie, dans ce cours nous supposerons que n_1 et n_2 sont suffisamment grands pour que t soit normalement distribué. Ainsi, nous pouvons utiliser $z_{0.975} \approx 2$ et $z_{0.095} \approx 1.645$.

Tests bilatéraux de l'égalité des moyennes : la méthode

Pour réaliser un test **bilatéral** au niveau de **significativité de 5%**

Étape 1 : Énoncer l'**hypothèse nulle**

$$H_0 : \mu_1 = \mu_2$$

Étape 2 : former la **statistique de test**

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Étape 3 : comparer la statistique de test avec la **valeur critique** $t_{0.975, n_{df}} \approx z_{0.975} \approx 2$

- Si $|t| > 2$, **rejeter** l'hypothèse nulle au niveau de significativité de 5%
- Si $|t| \leq 2$, **ne pas rejeter** l'hypothèse nulle au niveau de significativité de 5%

Tests unilatéraux de l'inégalité des moyennes : la méthode

Pour réaliser un test **unilatéral** au niveau de **significativité de 5%**

Étape 1 : Énoncer l'**hypothèse nulle**

$$H_0 : \mu_1 < \mu_2$$

Étape 2 : former la **statistique de test**

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Étape 3 : comparer la statistique de test avec la **valeur critique** $t_{0.95,n_{df}} \approx z_{0.95} \approx 1.645$

- Si $t > 1.645$, **rejeter** l'hypothèse nulle au niveau de significativité de 5%
- Si $t \leq 1.645$, **ne pas rejeter** l'hypothèse nulle au niveau de significativité de 5%

Tests de l'égalité des proportions

Rappel : lorsque X_i est binaire de sorte que $X_i \in \{0, 1\}$, $\bar{X} = p$ où p est une proportion égale à la part de fois où $X_i = 1$. La variance et l'écart-type de p sont donnés par les formules :

$$\sigma^2 = \frac{p(1-p)}{n} \quad \sigma = \sqrt{\frac{p(1-p)}{n}}$$

La statistique de test pour un test de l'(in)égalité des proportions est donc

$$t = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Expériences randomisées

Une application très importante du cadre de test d'hypothèses que nous avons développé est l'analyse des expériences randomisées

Vous êtes probablement familier avec les expériences randomisées en médecine

Lors du test de leur vaccin contre la COVID-19, Pfizer a pris un groupe d'individus et les a divisés aléatoirement en deux groupes : le **groupe de traitement** et le **groupe témoin**

Les individus dans le groupe de traitement ont reçu le vaccin, tandis que ceux dans le groupe témoin ont reçu un placebo

Ils ont ensuite mesuré si chaque individu contractait la COVID-19 dans le mois suivant. La différence dans les taux de COVID-19 entre les deux groupes mesure l'efficacité du vaccin

Randomisation

Nous appelons ces expériences *randomisées* car chaque individu est *aléatoirement* assigné aux groupes de traitement et témoin

Pourquoi randomisons-nous ?

La randomisation assure que la différence entre le **groupe de traitement** et le **groupe témoin** représente l'**effet causal** du traitement

Pourquoi cela fonctionne-t-il ?

La randomisation garantit que le traitement est **indépendant** de toutes autres caractéristiques. Que vous soyez grand, petit, jeune, vieux, homme, femme, votre probabilité d'être dans le groupe de traitement est la même

Cela signifie que les seules différences entre le groupe de traitement et le groupe témoin, en moyenne, sont dues au traitement !

Évaluation de la randomisation à l'aide de tests d'équilibre

Lorsque nous menons une expérience, nous voulons souvent vérifier que la randomisation a été mise en œuvre avec succès

Pour ce faire, nous réalisons des **tests d'équilibre**

Dans un test d'équilibre, nous comparons les valeurs moyennes de différentes caractéristiques **prédéterminées** dans le groupe de traitement et le groupe témoin et testons si elles sont égales

Chaque test d'équilibre est un test d'égalité des moyennes : $H_0 : \mu_T = \mu_C$, où μ_T est la moyenne dans le groupe de traitement et μ_C est la moyenne dans le groupe témoin. Si nous rejetons l'hypothèse nulle, cela indique un échec de la randomisation, et les groupes ne sont pas comparables

Il est très important que nous incluions uniquement les caractéristiques qui ont été mesurées **avant** que l'expérience ait lieu dans notre test d'équilibre. Sinon, elles pourraient être influencées par le traitement

Tableaux d’équilibre

Nous rassemblons souvent les tests d’équilibre dans un **tableau d’équilibre**

Dans un tableau d’équilibre, chaque ligne correspond à une caractéristique prédéterminée différente. La première ligne est la moyenne parmi le groupe traité, la deuxième ligne est la moyenne parmi le groupe témoin, et la troisième ligne est la différence, généralement accompagnée d’une erreur standard, d’une valeur de p ou d’une statistique t

Dans une étude expérimentale, le tableau d’équilibre est typiquement le Tableau 1

Caractéristique	Moyenne Traitée	Moyenne Témoin	Diff	p-valeur
Taille (cm)	175.2	174.8	0.6	0.72
Salaire ($ / h)	24.5	24.8	-0.3	0.58
Âge (années)	38.1	37.6	0.5	0.64
Genre (femme)	0.48	0.49	-0.01	0.89
Race (non blanc)	0.32	0.31	0.01	0.77

Vers

Maintenant, nous allons parler de quelque chose de différent : les vers

Vers

Maintenant, nous allons parler de quelque chose de différent : les vers

Plus précisément, nous lirons et discuterons de l'article

> *"Worms : Identifying impacts on education and health in the present of treatment externalities"*

par Ted Miguel et Michael Kremer, publié dans la revue Econometrica en 2004

Cet article était l'une des premières et des plus innovantes applications des expériences randomisées par les économistes du développement

Kremer, avec Abhijit Banerjee et Esther Duflo, a remporté le prix Nobel d'Économie en 2019, pour avoir développé une "approche expérimentale pour alléger la pauvreté mondiale"

Econometrica, Vol. 72, No. 1 (January, 2004), 159–217

WORMS: IDENTIFYING IMPACTS ON EDUCATION AND HEALTH IN THE PRESENCE OF TREATMENT EXTERNALITIES

BY EDWARD MIGUEL AND MICHAEL KREMER[1]

Intestinal helminths—including hookworm, roundworm, whipworm, and schistosomiasis—infect more than one-quarter of the world's population. Studies in which medical treatment is randomized at the individual level potentially doubly underestimate the benefits of treatment, missing externality benefits to the comparison group from reduced disease transmission, and therefore also underestimating benefits for the treatment group. We evaluate a Kenyan project in which school-based mass treatment with deworming drugs was randomly phased into schools, rather than to individuals, allowing estimation of overall program effects. The program reduced school absenteeism in treatment schools by one-quarter, and was far cheaper than alternative ways of boosting school participation. Deworming substantially improved health and school participation among untreated children in both treatment schools and neighboring schools, and these externalities are large enough to justify fully subsidizing treatment. Yet we do not find evidence that deworming improved academic test scores.

KEYWORDS: Health, education, Africa, externalities, randomized evaluation, worms.

L'ancylostome, l'ascaris, le trichocéphale et la schistosomiase infectent un quart de la population mondiale. Ils sont particulièrement répandus parmi les enfants d'âge scolaire dans les pays en développement. ***Nous examinons l'impact d'un programme dans lequel soixante-quinze écoles primaires rurales kényanes ont été intégrées dans un traitement de déparasitage dans un ordre randomisé****. Nous constatons que* ***le programme a réduit l'absentéisme scolaire d'au moins un quart****, avec des gains de participation particulièrement importants parmi les enfants les plus jeunes, faisant du déparasitage un moyen très efficace d'augmenter la participation scolaire chez les jeunes enfants.*

…

[L]e coût par année supplémentaire de participation n'est que de 3,50 $, rendant le déparasitage considérablement plus rentable que d'autres méthodes pour augmenter la participation scolaire, telles que les subventions scolaires. …Nous ne trouvons aucune preuve que le déparasitage ait augmenté les scores aux tests académiques.

…

Il existe une vaste littérature documentant des corrélations positives entre la santé et les résultats économiques. ***Nos résultats suggèrent un lien de causalité allant de la santé à l'éducation***

L'ancylostome et l'ascaris infectent chacun environ 1,3 milliard de personnes dans le monde, tandis que le trichocéphale affecte 900 millions et 200 millions sont infectés par la schistosomiase (Bundy (1994)). Bien que la plupart aient des infections légères qui peuvent être asymptomatiques, une minorité a des infections graves, qui peuvent entraîner une anémie ferriprive, une malnutrition protéino-énergétique, des douleurs abdominales et une léthargie.

Des thérapies orales à dose unique à faible coût peuvent tuer les vers, réduisant les infections par l'ancylostome, l'ascaris et la schistosomiase de 99 pour cent*. …La réinfection est rapide, cependant. …****L'Organisation mondiale de la santé a approuvé des programmes de déparasitage scolaire de masse dans les zones à forte infection, car cela élimine le besoin de dépistage parasitologique individuel coûteux, réduisant les coûts à aussi peu que 49 cents par personne par an en Afrique.***

Il existe seulement des preuves empiriques limitées sur les externalités du traitement de déparasitage, mais celles qui existent suggèrent que ***le déparasitage basé à l'école peut créer des externalités substantielles****. Cependant,* ***ces études se basent sur des comparaisons avant-après dans les mêmes villages pour estimer les externalités pour les individus non traités. Cela les laisse sans un groupe de comparaison plausible.***

…

La phase de randomisation dans les écoles de l'intervention de déparasitage que nous examinons nous permet de capturer l'effet global du déparasitage même en présence d'externalités entre individus au sein des écoles. *…Notre échantillon de 75 écoles est également beaucoup plus grand que les études existantes, qui étaient généralement menées dans cinq villages ou moins.*

Nous évaluons le Projet de Déparasitage des Écoles Primaires (PDEP) …Le projet a eu lieu dans le sud de Busia, une région agricole pauvre et densément peuplée dans l'ouest du Kenya, dans une zone ayant les taux d'infection par les helminthes les plus élevés du district de Busia. Les 75 écoles du projet comprennent presque toutes les écoles primaires rurales de cette région, et avaient un total d'inscriptions de plus de 30 000 élèves âgés de six à dix-huit ans.

En janvier 1998, les soixante-quinze écoles du PDEP ont été divisées au hasard en trois groupes de vingt-cinq écoles chacun *…En raison des contraintes administratives et financières de l'ICS, l'intervention sanitaire a été mise en place progressivement sur plusieurs années.* ***Les écoles du groupe 1 ont reçu un traitement de déparasitage gratuit en 1998 et 1999, les écoles du groupe 2 en 1999, tandis que les écoles du groupe 3 ont commencé à recevoir le traitement en 2001. Ainsi, en 1998, les écoles du groupe 1 étaient des écoles de traitement, les groupes 2 et 3 étaient des écoles de comparaison, et en 1999, les écoles du groupe 2 étaient des écoles de traitement et les écoles du groupe 3 étaient des écoles de comparaison.***

TABLE I

1998 AVERAGE PUPIL AND SCHOOL CHARACTERISTICS, PRE-TREATMENT[a]

	Group 1 (25 schools)	Group 2 (25 schools)	Group 3 (25 schools)	Group 1 − Group 3	Group 2 − Group 3
Panel A: Pre-school to Grade 8					
Male	0.53	0.51	0.52	0.01	−0.01
				(0.02)	(0.02)
Proportion girls <13 years, and all boys	0.89	0.89	0.88	0.00	0.01
				(0.01)	(0.01)
Grade progression (= Grade − (Age − 6))	−2.1	−1.9	−2.1	−0.0	0.1
				(0.1)	(0.1)
Year of birth	1986.2	1986.5	1985.8	0.4**	0.8***
				(0.2)	(0.2)
Panel B: Grades 3 to 8					
Attendance recorded in school registers (during the four weeks prior to the pupil survey)	0.973	0.963	0.969	0.003	−0.006
				(0.004)	(0.004)
Access to latrine at home	0.82	0.81	0.82	0.00	−0.01
				(0.03)	(0.03)
Have livestock (cows, goats, pigs, sheep) at home	0.66	0.67	0.66	−0.00	0.01
				(0.03)	(0.03)
Weight-for-age Z-score (low scores denote undernutrition)	−1.39	−1.40	−1.44	0.05	0.04
				(0.05)	(0.05)
Blood in stool (self-reported)	0.26	0.22	0.19	0.07**	0.03
				(0.03)	(0.03)
Sick often (self-reported)	0.10	0.10	0.08	0.02**	0.02**
				(0.01)	(0.01)

TABLE V

JANUARY TO MARCH 1999, HEALTH AND HEALTH BEHAVIOR DIFFERENCES BETWEEN GROUP 1 (1998 TREATMENT) AND GROUP 2 (1998 COMPARISON) SCHOOLS[a]

	Group 1	Group 2	Group 1 – Group 2
Panel A: Helminth Infection Rates			
Any moderate-heavy infection, January–March 1998	0.38	–	–
Any moderate-heavy infection, 1999	0.27	0.52	−0.25***
			(0.06)
Hookworm moderate-heavy infection, 1999	0.06	0.22	−0.16***
			(0.03)
Roundworm moderate-heavy infection, 1999	0.09	0.24	−0.15***
			(0.04)
Schistosomiasis moderate-heavy infection, 1999	0.08	0.18	−0.10*
			(0.06)
Whipworm moderate-heavy infection, 1999	0.13	0.17	−0.04
			(0.05)

Problème pratique n° 4

Un groupe d'économistes de la santé a mené une expérience randomisée pour étudier l'effet de l'assurance santé sur la santé. Ils ont recruté 500 personnes non assurées pour leur étude, et ont attribué aléatoirement à 250 d'entre elles une assurance santé gratuite. Ceux du groupe témoin n'ont rien reçu. Un an plus tard, ils ont mesuré la santé de chaque individu selon plusieurs dimensions. Ils ont combiné leurs mesures en un score de santé, une variable numérique allant de 0 à 5, où 0 est une santé médiocre et 5 est une excellente santé.

Le score de santé moyen des individus dans le groupe de traitement était de 2,5 (écart-type 1,2). Le score de santé moyen des individus dans le groupe témoin était de 2,25 (écart-type 1,3).

Testez si l'assurance santé a eu un effet sur le score de santé au niveau de signification de 5

Problème pratique n° 5

Un groupe d'économistes de la santé a mené une expérience randomisée pour étudier l'effet de l'assurance santé sur la santé. Ils ont recruté 500 personnes non assurées pour leur étude, et ont alloué aléatoirement 250 d'entre elles pour recevoir une assurance santé gratuite. Ceux du groupe témoin n'ont rien reçu. Un an plus tard, ils ont mesuré la santé de chaque individu selon plusieurs dimensions. Une des mesures était de savoir si les individus avaient subi une crise cardiaque au cours de l'année précédente.

Dans le groupe de traitement, 5 individus ont eu une crise cardiaque, tandis que dans le groupe témoin, 8 individus ont eu une crise cardiaque.

Testez si l'assurance santé a eu un effet sur le risque de crise cardiaque au niveau de signification de 5%.

Problème pratique n° 6

Un groupe d'économistes de la santé a mené une expérience randomisée pour étudier l'effet de l'assurance santé sur la santé. Ils ont recruté 500 personnes non assurées pour leur étude, et ont alloué aléatoirement 250 d'entre elles pour recevoir une assurance santé gratuite. Ceux du groupe témoin n'ont rien reçu. Avant de distribuer l'assurance santé, les économistes ont mesuré un ensemble de caractéristiques pour chaque individu.

En utilisant le tableau ci-dessous, déterminez si les différences entre le groupe de traitement et le groupe témoin sont significatives au niveau de 5%. La randomisation a-t-elle été réussie ?

Caractéristique	Traités	Témoin	Diff	t	p
Âge (années)	38.2 (10.0)	37.9 (9.8)	0.30		
Poids (kg)	78.5 (14.0)	77.0 (13.5)	1.50		0.22
Femme (1=oui)	0.48	0.45	0.03	0.90	
Fumeur (1=oui)	0.15	0.18	-0.03		