UNIVERSITÉ DE NANTES

Master 1 Ingénierie Mathématique

Statistique Inférentielle

Exercices

Ex 1. Soit $(S_n)_{n\in\mathbb{N}}$ et $(T_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires qui convergent en loi quand $n \to \infty$ vers S et T respectivement.

1. Si S_n et T_n sont indépendantes pour tout n, montrer que (S_n, T_n) converge en loi vers (S, T) avec S et T indépendantes.
2. Soit $X_1, X_2, ...$ une suite de variables aléatoires réelles iid de carré intégrable et telles que $\mathbb{E}(X_i) = 0$. On pose
$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \quad \text{et} \quad T_n = \frac{1}{\sqrt{n}} \sum_{i=n+1}^{2n} X_i.$$
Calculer la limite en loi de (S_n, T_n), puis de $Z_n = S_{2n} - S_n$.
3. En déduire que S_n ne converge pas en probabilité.

Ex 2. Soit $X_1, ..., X_n$ des v.a. réelles iid de loi uniforme sur $[0, \theta]$ avec $\theta > 0$.

1. Montrer que $\widehat{\theta} = \max\{X_1, ..., X_n\} = X_{(n)}$ est un estimateur consistant de θ.
2. Déterminer sa vitesse de convergence.
3. Montrer que $\widetilde{\theta} = \frac{2}{n} \sum_{i=1}^{n} X_i$ est également un estimateur consistant de θ et donner sa vitesse de convergence.

Ex 3. Soit $X_1, ..., X_n$ iid de fonction de répartition F. On veut estimer $F(x)^2$ pour $x \in \mathbb{R}$.

1. Calculer l'espérance de $\widehat{F}(x)^2$.
2. Déterminer la limite en loi de $\sqrt{n}(\widehat{F}(x)^2 - F(x)^2)$ quand $n \to \infty$ (indication: écrire $\widehat{F}^2 = F^2 + 2F(\widehat{F} - F) + (\widehat{F} - F)^2$).
3. Illustrer numériquement la convergence en loi sur $N = 1000$ échantillons iid $X_1, ..., X_{50}$ simulés, de loi exponentielle de paramètre 1 pour $x = 1$.

Ex 4. Soient $X_1, ..., X_n$ iid de densité f et fonction de répartition F. On note q_α le quantile d'ordre α de X_1 et on suppose la densité f continue et strictement positive sur un voisinage de q_α (on a donc unicité du quantile d'ordre α avec $F(q_\alpha) = \alpha$).

1. Justifier que pour tout $\epsilon > 0$,
$$\delta_\epsilon := \inf_{x:|x-q_\alpha|\geq\epsilon} |F(x) - \alpha|$$
est strictement positif.
2. Soit $\lceil n\alpha \rceil$ le plus petit entier supérieur ou égal à $n\alpha$. En utilisant le théorème de Glivenko-Cantelli, déduire que le quantile empirique $X_{(\lceil n\alpha \rceil)}$ est un estimateur consistant de q_α.

Ex 5. Soit $U_1, ..., U_n$ un échantillon de variables aléatoires iid de loi uniforme sur $]0,1[$. On s'intéresse au comportement asymptotique du quantile empirique d'ordre $\alpha \in]0,1[$, $U_{(n_\alpha)}$ où on note $n_\alpha = \lceil n\alpha \rceil$. On veut montrer que

$$Z_n := \sqrt{n}\big(U_{(n_\alpha)} - \alpha\big) \xrightarrow[n\to\infty]{loi} \mathcal{N}(0, \alpha(1-\alpha)).$$

1. Vérifier le résultat numériquement.
2. Rappeler la densité de la k-ième statistique d'ordre $U_{(k)}$.
3. Soit $X_1, ..., X_{n+1}$ des variables aléatoires indépendantes de loi exponentielle de paramètre 1, montrer que pour tout $k = 1, ..., n+1$, $S_k = \sum_{i=1}^k X_i$ suit une loi Gamma $\Gamma(k,1)$ de densité
$$\gamma_k(x) = \frac{x^{k-1}}{(k-1)!} e^{-x} \ , \ x > 0.$$
4. On admet le résultat suivant: Soient S, T deux v.a. positives indépendantes de densités f_S, f_T respectivement, la variable $X = T/S$ a pour densité
$$f_X(x) = \int_0^{+\infty} f_S(y) f_T(xy) y dy \ , \ x > 0.$$
Montrer que S_k/S_{n+1} a même loi que $U_{(k)}$ (indication: poser $S = S_k$ et $T = S_{n+1} - S_k$).
5. Déterminer la limite en loi de $\sqrt{n_\alpha}\big(S_{n_\alpha}/n_\alpha - 1\big)$.
6. En remarquant que
$$\lim_{n\to\infty} \frac{n_\alpha}{n} = \alpha \quad \text{et} \quad \lim_{n\to\infty} \sqrt{n}\Big(\frac{n_\alpha}{n} - \alpha\Big) = 0,$$
montrer d'après la question précédente que
$$\sqrt{n}\Big(\frac{S_{n_\alpha}}{n} - \alpha\Big) \xrightarrow[n\to\infty]{loi} \mathcal{N}(0,\alpha) \ \text{ et } \ \sqrt{n}\Big(\frac{S_{n+1} - S_{n_\alpha}}{n} - (1-\alpha)\Big) \xrightarrow[n\to\infty]{loi} \mathcal{N}(0, 1-\alpha).$$
7. En déduire que
$$\sqrt{n}\Big(\alpha \ \frac{S_{n+1} - S_{n_\alpha}}{n} - (1-\alpha) \ \frac{S_{n_\alpha}}{n}\Big) \xrightarrow[n\to\infty]{loi} \mathcal{N}\big(0, \alpha(1-\alpha)\big),$$
puis que
$$Y_n := \sqrt{n}\Big(\alpha \ \frac{S_{n+1} - S_{n_\alpha}}{S_{n_\alpha}} - (1-\alpha)\Big) \xrightarrow[n\to\infty]{loi} \mathcal{N}\Big(0, \frac{1-\alpha}{\alpha}\Big).$$
8. En remarquant que $U_{(n_\alpha)} \overset{loi}{=} \alpha/(1 + Y_n/\sqrt{n})$ (cf. question 4), montrer qu'il existe ξ_n compris entre 0 et $Y_n/\sqrt{n}$ tel que
$$U_{(n_\alpha)} = \alpha - \frac{\alpha}{(1+\xi_n)^2} \frac{Y_n}{\sqrt{n}}.$$
9. En conclure que
$$Z_n := \sqrt{n}\big(U_{(n_\alpha)} - \alpha\big) \xrightarrow[n\to\infty]{loi} \mathcal{N}(0, \alpha(1-\alpha)).$$

Ex 6. Soit un modèle paramétrique $\mathcal{M} = \{\mu_\theta : \theta \in \Theta \subset \mathbb{R}^4\}$ où $\theta = (m, \sigma^2, \kappa, \eta)$ avec

$$m = \mathbb{E}_\theta(X_1) \ , \ \sigma^2 = \mathrm{var}_\theta(X_1) \ , \ \kappa = \mathbb{E}_\theta(X_1 - m)^3 \ \text{ et } \ \eta = \mathbb{E}_\theta(X_1 - m)^4.$$

On observe $X_1, ..., X_n$ iid dans ce modèle et on note $\widehat{m}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \widehat{m})^2$.

1. Montrer que
$$\widehat{\sigma}_n^2 = \frac{1}{2n(n-1)} \sum_{i,j=1}^n (X_i - X_j)^2.$$
2. En déduire par récurrence que pour $n > 3$,
$$\mathrm{var}_\theta(\widehat{\sigma}_n^2) = \frac{1}{n}\Big(\eta - \frac{(n-3)\sigma^4}{n-1}\Big).$$
3. Déterminer $\mathrm{cov}_\theta(\widehat{m}_n, \widehat{\sigma}_n^2)$ en fonction de θ. Sous quelles conditions la covariance est-elle nulle?

Ex 7. Déterminer si les modèles suivants sont des familles exponentielles, en précisant le cas échéant les statistiques et paramètres naturels.

1. Le modèle Uniforme $\mathcal{M} = \{\mathcal{U}[a,b] : -\infty < a < b < +\infty\}$.
2. Le modèle Binomial $\mathcal{M} = \{\mathcal{B}(n,p) : n \in \mathbb{N}^*, p \in]0,1[\}$.
3. La famille des lois Gamma, de densités $f_{a,b}(x) = x^{a-1}e^{-x}/b^a, x > 0 : a, b > 0$.
4. Le modèle de Cauchy de densités $f_\theta(x) = \left[\pi(1 + (\theta - x)^2)\right]^{-1} : \theta \in \mathbb{R}$.

Ex 8. Montrer que le modèle Gaussien multivarié centré et non dégénéré (restreint aux matrices de covariance inversibles) est une famille exponentielle en précisant des paramètres et statistiques naturels.

Ex 9. Soit $X \sim \mathcal{N}(0,1)$, on rappelle que la fonction caractéristique de X est donnée par

$$\varphi_X(t) = \exp(-t^2/2) \ , \ t \in \mathbb{R}.$$

1. Calculer la fonction caractéristique d'un vecteur Gaussien standard $X = (X_1, ..., X_k)^\top$.
2. En déduire la fonction caractéristique d'un vecteur Gaussien quelconque $Y \sim \mathcal{N}(\mathsf{m}, \Sigma)$. On pourra utiliser que $Y \overset{loi}{=} \mathsf{m} + \Sigma^{1/2}X$.

Ex 10. Soit $-1 < \rho < 1$, on pose

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

1. Montrer qu'il existe un vecteur Gaussien $X = (X_1, X_2)^\top$ de moyenne nulle et de variance Σ.
2. On pose $Y_1 = (X_1 + X_2)$ et $Y_2 = (X_1 - X_2)$. Donner la loi de $Y = (Y_1, Y_2)^\top$.
3. Les variables Y_1 et Y_2 sont-elles indépendantes?
4. Justifier que Y admet une densité sur $\mathbb{R}^2$ et la calculer.

Ex 11. Soit $X \sim \mathcal{N}(0,1)$ et $a > 0$. on pose

$$Y_a := X\mathbb{1}\{|X| < a\} - X\mathbb{1}\{|X| \geq a\}$$

1. Montrer que Y_a est une variable aléatoire Gaussienne.
2. Montrer qu'il existe $b > 0$ tel que $\int_0^b x^2 e^{-x^2/2} dx = \sqrt{2\pi}/4$.
3. Calculer $\mathrm{cov}(X, Y_b)$. Le vecteur $(X, Y_b)^\top$ est-il un vecteur Gaussien?

Ex 12. On rappelle les trois caractérisations importantes de l'espérance conditionnelle:

i) Si la loi conditionnelle de Y sachant X est de densité $f_{Y|X}$ par rapport à une mesure ν qui admet un moment d'ordre 1 presque-sûrement, alors l'espérance conditionnelle de Y sachant X est définie comme l'espérance sous la densité conditionnelle

$$\mathbb{E}(Y|X) = \int y f_{Y|X}(y) d\nu(y).$$

ii) Si Y est intégrable, alors $\mathbb{E}(Y|X)$ est l'unique variable aléatoire (p.s.) fonction de X qui vérifie pour toute fonction h telle que $\mathbb{E}|h(X)Y| < \infty$,

$$\mathbb{E}\big(h(X)Y\big) = \mathbb{E}\big(h(X)\mathbb{E}(Y|X)\big).$$

iii) Si Y est de carré intégrable, alors $\mathbb{E}(Y|X) = \rho^*(X)$ où ρ^* minimise $\rho \mapsto \mathbb{E}(Y - \rho(X))^2$ parmi les fonctions mesurables de X.

1. Montrer que la caractérisation *iii)* de l'espérance conditionnelle implique que pour toute fonction h telle que $\mathbb{E}|h(X)Y| < \infty$, $\mathrm{cov}(Y - \mathbb{E}(Y|X), h(X)) = 0$.
2. En déduire que *iii)* et *ii)* sont équivalentes si Y est de carré intégrable.
3. Montrer que *i)* et *iii)* sont équivalentes si (X, Y), X et Y sont de densités $f_{X,Y}$, f_X et f_Y par rapport à ν et Y est de carré intégrable.

Ex 13. Soit X une variable aléatoire de loi exponentielle de paramètre $\lambda > 0$ (d'espérance $1/\lambda$) et Y dont la loi conditionnelle sachant X est une loi normale d'espérance mX, $m \in \mathbb{R}$ et de variance 1. On pose $\theta = (\lambda, m)$.

1. Déterminer la densité jointe $f^{\theta}_{X,Y}$ de (X, Y).
2. Montrer que $\mathbb{E}_\theta\big(Y/X|X\big) = \mathbb{E}_\theta(Y/X)$. Les variables X et Y/X sont-elles indépendantes?
3. Soit $(X_1, Y_1), ..., (X_n, Y_n)$ un échantillon iid de même loi que (X, Y). Montrer que
$$\widehat{m} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i} \quad \text{et} \quad \widetilde{m} = \frac{1}{n}\sum_{i=1}^n \frac{Y_i}{X_i}$$
sont des estimateurs consistants de m. Sont-ils sans biais?
4. Comparer numériquement les erreurs quadratiques moyennes de $\widehat{m}$ et $\widetilde{m}$ pour $n = 100$ et $\lambda = 1$.
5. En travaillant conditionnellement à $X_1, ..., X_n$ dans un premier temps, déterminer le meilleur estimateur de m de la forme
$$\sum_{i=1}^n \phi_i(X_1, ..., X_n)Y_i$$
où les ϕ_i sont des fonctions (mesurables) quelconque.
6. Illustrer numériquement les résultats des deux questions précédentes.

Ex 14. Montrer que $\widehat{\theta} = \max\{X_1, ..., X_n\}$ est une statistique exhaustive et totale dans le modèle d'échantillonage iid uniforme sur $[0, \theta]$, $\theta > 0$.

Ex 15. Calculer l'information de Fisher dans le modèle d'échantillonage iid Gaussien $\mathcal{M} = \{\mathcal{N}(m, \sigma^2) : m \in \mathbb{R}, \sigma^2 > 0\}$. Les estimateurs empiriques
$$\widehat{m} = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{et} \quad \widehat{\sigma}^2 \frac{1}{n-1}\sum_{i=1}^n (X_i - \widehat{m})^2$$
sont-ils efficaces?

Ex 16. Soit le modèle d'échantillonage iid de Poisson de paramètre $\theta > 0$.

1. Montrer que $S = S(X) = \sum_{i=1}^n X_i$ est une statistique exhaustive complète.
2. En déduire l'estimateur sans biais optimal de θ et montrer qu'il est efficace.
3. Soit $T = T(X) = \sum_{i=1}^n X_i^2$. Montrer que $\widetilde{\theta} = (T - S^2/n)/(n-1)$ est également un estimateur sans biais de θ. Est-il efficace?
4. En déduire $\mathbb{E}_\theta\big(T|S\big)$.
5. On veut maintenant estimer $\psi = \psi(\theta) = e^{-\theta}$. Montrer que $\widehat{\psi} = \mathbb{1}\{X_1 = 0\}$ est un estimateur sans biais de ψ. Est-il consistant?
6. En déduire l'estimateur sans biais optimal de ψ.

Ex 17. Soient $X_1, ..., X_n$ iid à valeurs dans $\{-1, 1\}$ avec θ la probabilité que X_1 soit égal à 1.

1. Décrire le modèle statistique et montrer que c'est une famille exponentielle.
2. Calculer l'estimateur du maximum de vraisemblance de θ et montrer qu'il est efficace.
3. Pour $t \in \mathbb{R}$, on note $\phi_\theta(t) = \mathbb{E}_\theta(e^{-tX_1})$. Proposer un estimateur $\widehat{\phi}_\theta(t)$ de $\phi_\theta(t)$ par la méthode des moments.
4. Calculer l'estimateur du maximum de vraisemblance de $\phi_\theta(t)$. Montrer qu'il est asymptotiquement efficace.

Ex 18. Soit $X_1, ..., X_n$ iid de loi de Cauchy de densité $f_\theta(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}, x \in \mathbb{R}$ avec $\theta \in \mathbb{R}$.

1. Calculer la log-vraisemblance $v(\theta, X)$, puis l'information de Fisher (on admettra que le modèle est régulier).
2. On note $\widehat{\theta}$ l'estimateur du maximum de vraisemblance. Justifier qu'il existe, est unique et rappeler son comportement en loi quand $n \to \infty$.
3. Rappeler le comportement asymptotique de $\widetilde{\theta} = X_{(\lceil n/2 \rceil)}$, la médiane de l'échantillon. Cet estimateur est-il asymptotiquement efficace?
4. En notant $S_\theta(X) = \partial_\theta v(\theta, X)$ la fonction de score et $H_\theta(X) = \partial_\theta S_\theta(X)$, montrer que
$$\widehat{\theta} - \widetilde{\theta} = -\frac{S_{\widetilde{\theta}}(X)}{H_{\widetilde{\theta}}(X)} + o_P(\sqrt{n}).$$
5. En déduire un estimateur asymptotiquement efficace de θ qui soit fonction de $\widetilde{\theta}$.
6. Illustrer numériquement les performances (convergence, biais, variance, EQM, ...) de ces trois estimateurs calculés sur des échantillons de tailles variant de $n = 20$ à $n = 2000$ (prendre une valeur quelconque de θ).

Ex 19. Soit un modèle paramétrique régulier avec l'estimateur du maximum de vraisemblance asymptotiquement Gaussien de variance $\mathcal{I}(\theta)^{-1}$. On considère un paramètre $\psi = \psi(\theta) \in \mathbb{R}^q$ tel que ψ est différentiable et la borne de Cramer-Rao $\theta \mapsto \Sigma_\psi(\theta) := \nabla\psi(\theta)^\top \mathcal{I}(\theta)^{-1} \nabla\psi(\theta)$ est continue est inversible en tout point.

1. Construire une région de confiance asymptotique pour ψ de niveau asymptotique $1-\alpha$ basée sur l'estimateur du maximum de vraisemblance.
2. Coder une fonction qui à une log-vraisemblance (donnée sous forme de fonction de θ) supposée issue d'un modèle régulier, associe une région de confiance asymptotique sur θ de niveau de confiance $1-\alpha$. On pourra se restreindre au cas $\Theta \subseteq \mathbb{R}$.
3. Tester les performances de cet intervalle de confiance dans un modèle logistique de densité $f_\theta(x) = e^{-(x-\theta)}/(1+e^{-(x-\theta)})^2, x \in \mathbb{R}$ sur différentes tailles d'échantillons variant de $n = 5$ à $n = 500$.

Ex 20. Soit $X_1, ..., X_n$ iid de loi $\Gamma(2, \theta), \theta > 0$.

1. Générer un échantillon iid $X_1, ..., X_n$ de taille $n = 5$ avec pour vraie valeur du paramètre $\theta = 2$.
2. On veut tester (maladroitement) l'hypothèse $H_0 : \theta = 2$ contre $H_1 : \theta \neq 2$ en supposant les données issues d'une loi normale de moyenne 2θ et de variance inconnue. Construire (à la main) la statistique de test et la p-value associée en précisant la région de rejet.
3. Comparer avec le résultat de la fonction `t.test`.
4. On veut évaluer la probabilité de rejeter H_0 si elle est vraie (l'erreur de première espèce) lorsqu'on se fixe un niveau asymptotique $\alpha = 0.05$. Reproduire l'expérience 10000 fois et calculer la fréquence de rejet de H_0. Commenter.
5. Reprendre la question précédente pour $n = 50$. Commenter.
6. On veut évaluer la probabilité de rejeter H_0 si elle est fausse (la puissance du test) au niveau $\alpha = 0.05$. Générer 10000 échantillons de taille $n = 5$ de loi $\Gamma(2, 1.5)$ et calculer la fréquence de rejet de H_0. Recommencer pour $n = 50$ et commenter.
7. Représenter graphiquement la puissance du test pour $n = 50$ et θ variant dans l'intervalle $[1, 3]$.
8. En remarquant que sous H_0, $S := \sum_{i=1}^n X_i$ suit une loi Gamma $\Gamma(2n, \theta)$, construire un test exact de H_0 de niveau $\alpha = 0.05$.
9. Comparer la puissance avec celle du test précédent sur l'intervalle $[1, 3]$.

Ex 21. Soient $X_1, ..., X_n$ iid de loi de Pareto de densité $f_{a,b}(x) = ab^a x^{-(a+1)} \mathbb{1}\{x > b\}, x \in \mathbb{R}$ avec $a, b > 0$ les paramètres du modèles.

1. Calculer les estimateurs du maximum de vraisemblance $\widehat{a}$ et $\widehat{b}$ de a et b.
2. En remarquant que λX_i est de loi de Pareto de paramètres $a, \lambda b$ pour tout $\lambda > 0$, montrer que la loi de $\widehat{a}$ ne dépend pas de b.
3. Montrer que $g(a, X) = \widehat{a}/a$ est une fonction pivotale.
4. Estimer les quantiles de $g(a, X)$ par Monte-Carlo pour $n = 50$.
5. Générer un échantillon de taille $n = 50$ avec $a = 2$ et $b = 1$ et construire un intervalle de confiance le plus petit possible sur a de niveau de confiance $1 - \alpha = 0.95$, en utilisant les quantiles estimés dans la question précédente.
6. Proposer un test de niveau $\alpha = 0.05$ de $H_0 : a = 2$ contre $H_1 : a \neq 2$.
7. Estimer par Monte-Carlo la puissance du test sous des valeurs de a variant entre 0.1 et 3, et la représenter graphiquement.
8. Reprendre les questions précédentes pour tester $H_0 : \theta = 2$ contre $H_1 : \theta > 2$.
9. Comparer avec le test du rapport de vraisemblance sur $H_0 : \theta = 2$ contre $H_1 : \theta = 3$.

Ex 22. Soit le modèle iid Gaussien $\mathcal{N}(\theta, 1), \theta \in \mathbb{R}$.

1. Calculer la statistique du rapport de vraisemblance pour $H_0 : \theta = 0$ contre $H_1 : \theta = 1$.
2. Montrer que la région de rejet du test du rapport de vraisemblance est identique pour toute hypothèse alternative $H_1 : \theta = \theta_1$ avec $\theta_1 > 0$.
3. En déduire que le test du rapport de vraisemblance pour $H_0 : \theta = 0$ contre $H_1 : \theta > 0$ est UPP.
4. Montrer que ce test est sans biais si on prend comme hypothèse alternative $H_1 : \theta \neq 0$ mais qu'il n'est plus UPP.
5. Montrer qu'il est UPPSB.
6. Montrer que le théorème de Wilks dans ce cas.

Ex 23. Soit $(X_1, Y_1), ..., (X_n, Y_n)$ un échantillon iid d'un couple de variables aléatoires (X, Y) à valeurs dans des ensembles finis $\{x_1, ..., x_K\}$ et $\{y_1, ..., y_L\}$. Pour chaque paire (k, ℓ), on note $N_{k\ell}$ l'effectif observé

$$N_{k\ell} = \sum_{i=1}^{n} \mathbb{1}\{X_i = x_k, Y_i = y_\ell\},$$

et $\theta_{k\ell} = \mathbb{P}(X = x_k, Y = y_\ell)$. On considère le modèle paramétré par le vecteur $\theta = (\theta_{k\ell})_{k=1,...,K,\ell=1,...,L}$ sur les observations $N_{k\ell}, k = 1, ..., K, \ell = 1, ..., L$.

1. Décrire l'espace Θ des paramètres. Quelle est sa dimension?
2. Ecrire la vraisemblance du modèle et calculer les estimateurs du maximum de vraisemblance des $\theta_{k\ell}$.
3. On veut tester si les variables X et Y sont indépendantes. Décrire le sous-ensemble de paramètres Θ_0 correspondant à cette hypothèse et préciser sa dimension.
4. Calculer la statistique du test de rapport de vraisemblance dans ce cas. Rappeler sa loi limite sous H_0 (théorème de Wilks).
5. On veut tester s'il y a des différences significatives de salaire entre hommes et femmes dans une populations. On relève la répartition suivante sur un échantillon de 575 personnes:

Salaires	entre 1000 et 2000 €	entre 2000 et 3000€	entre 3000 et 4000€	> 4000€
Femmes	80	75	100	30
Hommes	50	70	110	60

   Appliquer le résultat précédent pour tester l'absence d'inégalités sur les salaires.
6. Comparer le résultat avec celui obtenu par la fonction `chisq.test`.