

Toutes les réponses doivent être justifiées. Le barème est donné à titre indicatif.

Ex. 1 — Des TDs [2 point]. Au debut de chaque séance de TD, les étudiants doivent résoudre un petit exercice, qui est différent par les deux groupes.

1. Si les exercices sont assignés aléatoirement et équitablement aux groupes, calculer la probabilité que le deuxième groupe reçoive l'exercice plus difficile dans 5 ou plus des 6 séances de TDs.
2. Concevoir un test d'hypothèse pour décider si l'asignation des exercices est effectivement équitable.

Ex. 2 — Estimateurs [2 points]. Considerons une variable aléatoire X avec densité de probabilité $f_X(x) = \lambda/2e^{-\lambda|x|}$ pour $x \in \mathbf{R}$.

1. Vérifier que $f_X(x)$ satisfait toutes les propriétés d'une densité de probabilité.
2. Soit $\{x_1, x_2, \dots, x_n\}$ un échantillon de valeurs de la variable X tirées indépendamment. Déterminer l'estimateur du maximum de vraisemblance de λ .

Ex. 3 — Ventes [4 points]. Pendant 10 jours un magasins d'électroménagers vend le nombre suivant de réfrigérateurs. La variance des ventes est connue et vaut $\sigma^2 = 4$.

5	7	10	3	5	8	11	8	7	6
---	---	----	---	---	---	----	---	---	---

1. Estimer un intervalle de confiance à 95% pour la moyenne journalière des ventes de réfrigérateurs.
2. Pour combien de jours faudrait-il collecter les données sur les ventes pour avoir un intervalle à 95% avec une incertitude relative de 1% ?
3. Historiquement, la moyenne des ventes est de 9 réfrigérateurs par jour. Que vaut la valeur-p associée à l'hypothèse que la moyenne des ventes de réfrigérateurs soit 9 (hypothèse H_0) ?
4. Rejetez-vous avec une seuil de signification de 5% l'hypothèse H_0 que la moyenne des ventes de réfrigérateurs soit 9 ?
5. Utilisez le tableau dessous, obtenu avec la méthode du bootstrap, pour donner un interval de confiance à 50% pour la moyenne des ventes de réfrigérateurs. Expliquez comme le tableau a été obtenu.

11	8	11	5	7	10	3	3	8	6
5	3	5	6	7	8	11	6	7	8
8	11	10	7	10	3	5	8	7	7
7	3	5	11	6	8	7	6	8	11

Ex. 4 — Risques [1 point]. Définir en general le risque de première espèce et celui de deuxième espèce.

Ex. 5 — Régression [4 points]. Le jeu des données *Credit.csv* (disponible ici <http://www-bcf.usc.edu/~gareth/ISL/Credit.csv>) montre plusieurs variables quantitatives pour 400 personnes : *Balance* (dette moyenne par carte de crédit pour un certain nombre de personnes), *Age* (l'âge de la personne), *Cards* (le nombre de cartes de crédit), *Education* (nombre d'années d'études), *Income* (le revenu en milliers de dollars), *Limit* (la limite de crédit), et *Rating* (cote de crédit).

Une régression linéaire avec *Limit* comme réponse et les autres données comme variables explicatives donne le résultat suivant avec *R* :

```
> summary(lm(Limit ~ Balance+Income+Rating
+Cards+Age+Education, data=credit))
Call:
lm(formula = Limit ~ Balance + Income + Rating
+ Cards + Age + Education, data = credit)
```

Residuals :

Min	1Q	Median	3Q	Max
-389.18	-101.34	13.88	104.51	344.44

Coefficients :

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-322.05951	54.39320	-5.921	6.99e-09	***
Balance	0.11224	0.04730	2.373	0.0181	*
Income	0.96502	0.50759	1.901	0.0580	.
Rating	14.44512	0.20324	71.074	< 2e-16	***
Cards	-72.01842	5.62521	-12.803	< 2e-16	***
Age	0.02575	0.45348	0.057	0.9547	
Education	2.97900	2.45238	1.215	0.2252	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 152.7 on 393 degrees of freedom

Multiple R-squared: 0.9957, Adjusted R-squared: 0.9956

F-statistic: 1.514e+04 on 6 and 393 DF, p-value: < 2.2e-16

- 1.Existe-t-il une relation entre les variables explicatives et la limite de crédit ?
- 2.Quantifier la dépendance entre les variables explicatives et la limite de crédit.
- 3.Quelles variables contribuent à déterminer la limite de crédit ?

4. Estimer les intervalles de confiance à 90% pour le paramètre Balance (la dette moyenne).

5. Dire si la dette moyenne explique la limite de crédit avec un risque de première espèce de 1%.

On passe à une régression linéaire avec seulement Rating and Cards comme variables explicatives et on obtient :

```
> summary(lm(Limit ~ Rating+Cards, data=credit))
```

Call:

```
lm(formula = Limit ~ Rating + Cards, data = credit)
```

Residuals:

Min	1Q	Median	3Q	Max
-360.67	-97.63	13.09	108.56	354.49

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-341.15903	24.76708	-13.78	<2e-16 ***
Rating	14.90573	0.04968	300.04	<2e-16 ***
Cards	-72.31808	5.60545	-12.90	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 153.3 on 397 degrees of freedom

Multiple R-squared: 0.9956, Adjusted R-squared: 0.9956

F-statistic: 4.502e+04 on 2 and 397 DF, p-value: < 2.2e-16

6. Commenter les changements des valeurs de R^2 et adjusted- R^2 .

7. Dire quel est le meilleur modèle.