Université Grenoble Alpes **Frédérique Leblanc**
L2/STA351 –

Examen Terminal 2018-2019

Durée : 2h.

Documents autorisés : 2 pages manuscrites recto-verso, tables statistiques (8 pages), calculatrice programmable (celle du Bac).

Recommandations : soin dans la rédaction et l'orthographe ; inutile de recopier les notes de cours au hasard ; indiquer lorsque c'est explicitement demandé l'expression formelle des statistiques utilisées pour produire les valeurs numériques obtenues ; répondre dans les champs grisés et directement sur le sujet.

Barême : Il est seulement indicatif et sur 50 points

Les quatre premiers exercices portent sur les données utilisées en TP du fichier `mtcars` où sont observés 11 variables sur 32 individus (modèles de véhicules). Seront utilisés les échantillons suivants :

1. `mpg` : éch. de la variable qu'on notera M " miles p. galon" et qui donne le nombre de miles parcourus pour 1 galon de carburant (donc plus cette quantité est grande plus le véhicule est économique).
2. `am` : ech. de la variable A, "manuelle ou automatique" indicatrice d'une boite de vitesse manuelle (la valeur 1 correspond à "manuelle")
3. `cyl` : ech. de la variable C nombres de cylindres dans le moteur (à valeurs dans $\{4, 6, 8\}$)
4. `disp` : ech. de la variable D : volume déplacé dans les cylindres

Exercice 1 (13 points) :

On s'intérresse ici à la variable M, μ son espérance et σ son écart-type. On dispose d'un échantillon observé de M pour $n = 32$ individus et on a obtenu :

$$\sum_{i=1}^{32} m_i = 642.9 \quad \text{et} \quad \sum_{i=1}^{32} m_i^2 = 14042.31$$

Les résultats sur μ et σ seront donnés à 10^{-2}.

1. (3 points) Donner les estimations sans biais de μ et σ^2 et en déduire l'estimation de σ :

Solution:

$$\begin{aligned} \hat{\mu} &= \\ \hat{\sigma^2} &= \\ \hat{\sigma} &= \end{aligned}$$

2. (3 points) On supposera la variable M normalement distribuée. Calculer un intervalle de confiance pour la distance moyenne parcourue avec un galon d'essence pour un véhicule de 1973, au niveau de confiance 0.95.

Solution:

— Donner d'abord l'expression formelle de l'intervalle de confiance de niveau $1 - \alpha$ utilisé puis le calculer :

$$I(\quad ;\alpha) = \left[\qquad\qquad\qquad ; \qquad\qquad\qquad \right]$$

— **A.N.** : pour $\bar{m} = \quad ; n = \quad ; \alpha = \quad$ .... on obtient :

$$I_{calc}(\quad ; \quad) = [\qquad\qquad ; \qquad\qquad]$$

3. (3 points) Pour cet échantillon, quel est le niveau de confiance de l'intervalle $[18.09; 22.09]$?

Solution:

4. (4 points) On se propose ici de calculer l'intervalle de confiance de niveau 80% pour l'écart-type σ de M. On calculera d'abord l'intervalle de confiance pour σ^2 puis celui sur σ.

Solution:

L'intervalle de confiance de niveau $1 - \alpha$ pour .... est donné par :

$$I(\quad ;\alpha) = \left[\qquad\qquad ; \qquad\qquad \right]$$

et pour l'échantillon observé on a :
A.N. : $\alpha = \quad , n = \quad ,$ ....
d'où :

$$I_{calc}(\quad ; \quad) = [\qquad\qquad ; \qquad\qquad]$$

et on en déduit l'intervalle de confiance sur σ : ...

$$I_{calc}(\quad ; \quad) = [\qquad\qquad ; \qquad\qquad]$$

Exercice 2 (7 points) :

On s'intéresse à la proportion p inconnue de véhicules à boite manuelle dans le parc automobile des années 73-74. Dans l'échantillon observé 13 sont des véhicules à boite manuelle sur les 32 véhicules considérés.

1. (1 point) Quelle estimation de p proposez-vous ?

Solution:

$$\hat{p} = \qquad\qquad = $$

2. (4 points) En utilisant l'approximation normale, donner un intervalle de confiance symétrique pour p de niveau de confiance approximatif 0.95 :

Solution:
A condition que : ....
on peut utiliser l'intervalle dont l'expression formelle est donné par :

$$I(\quad ; \alpha) = \left[\qquad\qquad\qquad ; \qquad\qquad\qquad \right]$$

A.N. : pour $\hat{p} =$; $n =$; $\alpha =$; ... on obtient :

$$I_{calc}(\quad ; \quad) = [\qquad ; \qquad]$$

3. (2 points) Combien de véhicules devraient être observés pour pouvoir calculer un intervalle de confiance de niveau 0.95 avec une précision de plus ou moins 1% (on supposera que $\hat{p}$ est celui obtenu précédemment) ?

Solution:

Exercice 3 (10 points) :

On cherche à présent à comparer les moyennes des deux variables suivantes qui seront ici notées X et Y (d'esp. μ_X et μ_Y) sur la sous-population des véhicules ayant un moteur à 4 cylindres :

— X : la variable M pour "les véhicules 4 cylindres"

— Y : la variable $-0.14D + 41$ pour "les véhicules 4 cylindres"

Les valeurs observées de M et D sont données par :

`mpg`	22.8	24.4	22.8	32.4	30.4	33.9	21.5	27.3	26.0	30.4	21.4
`disp`	108	147	141	79	76	71	120	79	120	95	121

1. (2 points) Calculer les valeurs des échantillons de Y et de $D' = X - Y$

`y`											
`d'`											

2. (1 point) Ces échantillons sont-ils appariés ou indépendants ?

Solution:

3. (2 points) *On veut savoir si la consommation d'un véhicule a la même moyenne que la nouvelle variable Y construite par transformation linéaire de la variable D.*

Solution:
Préciser le nom du test mis en oeuvre et ses conditions d'application. On fait le test statistique de : ....
sous la condition (hypothèse de modélisation) :

4. (5 points) Mettre en oeuvre le test au niveau 5% :

Solution:
— Poser les hypothèses formelles du test :

$\mathcal{H}_0$: $\mathcal{H}_1$:

— La statistique de test et la région de rejet sont données par :

$T =$ et $W_\alpha = \{ \quad \}$

— **A.N. :** $n =$, $\alpha =$, ...

$T_{calc} =$ et $W_{0.05} = \{ \quad \}$

— **Conclusion littérale :**

Exercice 4 (12 points)

On souhaite ici répondre à la question :

En 1973, le type de boite de vitesse a-t-il un effet en moyenne sur la consommation d'un véhicule ?

Pour y répondre on dispose des deux échantillons de tailles respectives 13 (manuelle) et 19 (automatique) :

— X consommation en miles par galon pour les véhicules à boite manuelle
— Y consommation en miles par galon pour les véhicules à boite automatique

On a les résumés numériques suivants :

$$\bar{x} = 24.39; \quad \bar{y} = 17.15; \quad s_x^2 = 35.1001; \quad s_y^2 = 13.9257.$$

On notera μ_x, μ_y les espérances, σ_x^2, σ_y^2 les variances et n_x, n_y les tailles des échantillons.

1. (1 point) Préciser les hypothèses de modélisation nécessaires pour comparer les variances :

Solution:

2. (5 points) Tester l'égalité des variances pour n'importe quel niveau α :

Solution:
— Poser les hypothèses du test :

$\mathcal{H}_0$: $\mathcal{H}_1$:

— La statistique de test et la région de rejet au niveau α sont données par :

$$T = \qquad \text{et} \quad W_\alpha = \{T ... \qquad \text{ou} \quad T ... \qquad \}$$

— **A.N. :** $n_x = ...$, $n_y = ...$, ...

$$T_{calc} = \qquad \text{et} \qquad p - valeur =$$

Indiquer la fonction de la calculatrice utilisée pour le calcul de la p-val ainsi que ses arguments (ou si vous n'avez pas de calculatrice donner l'expression formelle de cette p-valeur en notant F la fonction de répartition de la statistique de test sous l'hypothèse nulle) :

— Conclusion littérale :

3. (6 points) Compte tenu du résultat précédent obtenu pour $\alpha = 5\%$, indiquer l'hypothèse supplémentaire à faire sur le modèle pour pouvoir comparer les espérances de X et Y et mettre en oeuvre le test permettant de répondre au problème posé au début de l'exercice :

Solution:

— Hypothèse à rajouter sur le modèle :

— Hypothèses du test :

$$\mathcal{H}_0 : \qquad \mathcal{H}_1 :$$

— Statistique de test et région de rejet :

$$T = \qquad \text{avec} \quad \Sigma^2 = \qquad \text{et} \quad W_\alpha = \{ \qquad \}$$

— Calcul de Σ^2, T et de la p-valeur :

$$\Sigma^2_{calc} = \qquad T_{calc} = \qquad ; \, p{-}valeur = \qquad =$$

— **Conclusion littérale :**

Exercice 5 (8 points) :

Un traitement est administré à trois doses différentes (D1, D2 et D3) à trois groupes de sujets atteints d'une même maladie. L'expérimentation est faite en double aveugle. On compte le nombre de guérisons pour chaque dose. Les résultats sont les suivants :

Etat patient / Dose	guéri	non guéri
D1	30	30
D2	42	35
D3	58	31

On souhaite déterminer si l'efficacité du traitement est liée à la dose utilisée.

1. (4 points) Analyse descriptive : calculer les trois répartitions conditionnelles de l'état du patient selon la dose prescrite et les représenter sur un même graphique en bleu pour la répartition conditionnelle au dosage D1, rouge pour la répartition conditionnelle au dosage D2 et vert pour la répartition conditionnelle au dosage D3. Interpréter ce graphique.

   Solution:

2. (4 points) Quel test statistique doit-on mettre en oeuvre pour valider (ou non) la conjecture précédente ? Préciser le nom du test réalisé, les hypothèses du test, les éventuelles conditions requises pour sa mise en oeuvre, la statistique de test calculée, la p-valeur et la conclusion de l'étude.

   Solution: