

Master 2 SISE

EXAMEN : INFERENCE STATISTIQUE & ANOVA - ANCOVA

Année Universitaire 2022-2023 - février 2023 Durée 1h 30

Recommandations : Soigner la rédaction. La note prendra largement en compte la qualité des explications. La copie-brouillon et la copie qui ne comporte que des résultats sont mal perçues par le correcteur. Aucun document n'est permis. Les machines à calculer sont autorisées. Les trois exercices sont indépendants. Des extraits de tables statistiques sont donnés en annexe 1 ainsi que les principaux résultats SAS de l'exercice 3, en annexe 2.

Exercice 1 : (Barème de notation : a) 1.5 pt b) 1.5 pt c) 1.5 pt d) 1.5 pt e) 1.5 pt = 7.5 pts)

On a relevé deux échantillons de personnes testées positives à la Covid-19. La répartition des personnes testées positives à la Covid-19 parmi les personnes vaccinées et non-vaccinées est présentée dans le tableau suivant :

| Statut vaccinal | Non-vacciné | Vacciné |
|---------------------------------------|-------------|---------|
| Nombre de personnes testées positives | 225 | 50 |
| Nombre total de personnes testées | 300 | 200 |

- Peut-on considérer avec un risque d'erreur $\alpha = 5\%$, que la proportion de personnes non-vaccinées testées positives est supérieure à 70% ?
- Etablir un intervalle de confiance bilatéral symétrique de niveau $1 - \alpha = 95\%$ de la proportion de personnes non-vaccinées testées positives.
- Déterminer le niveau de confiance $1 - \alpha$ que l'on attribue à l'intervalle bilatéral symétrique $[69.86\%; 80.14\%]$ de la proportion de personnes non-vaccinées testées positives, obtenu à partir de d'un échantillon de 300 personnes non-vaccinées ?
- Quel doit-être la taille de l'échantillon de personnes non-vaccinées à tester pour établir un intervalle bilatéral symétrique de la proportion de personnes non-vaccinées testées positives, d'un niveau de 95% avec une marge d'erreur inférieure ou égale à 2.45% ?
- Peut-on considérer avec un risque d'erreur $\alpha = 5\%$, qu'il y a plus de personnes testées positives chez les personnes non-vaccinées que chez les personnes vaccinées ?

Exercice 2 : (Barème de notation : a) 1.5 pt b) 2.5 pts c) 1.5 pt d) 2.5 pts = 8 pts)

On a relevé les durées de séjours en réanimation des patients contaminés par la Covid-19 selon le type de variant Delta ou Omicron. Les durées de séjour des patients en réanimation sont supposées suivre une loi normale. Les résultats sont présentés dans le tableau suivant :

| Variant Covid-19 | Delta | Omicron |
|---|-------|---------|
| Nombre de patients positifs | 50 | 25 |
| Durée moyenne de séjour observé (jours) | 13.6 | 9.9 |
| Variance observée de la durée de séjour | 2.64 | 1.36 |

a) Peut-on affirmer avec un risque d'erreur $\alpha = 5\%$, que la durée moyenne de séjour en réanimation des patients positifs au variant Omicron est significativement inférieure 10 jours ?

b) Déterminer la puissance du test précédent où sous $H_0 : m_0 = 9.9$ jours, alors qu'en réalité le temps moyen de séjour en réanimation des patients contaminés par le variant Omicron est sous $H_1 : m_1 = 9.3$ jours.

c) Peut-on considérer avec un risque d'erreur $\alpha = 5\%$, que l'écart-type de la durée de séjour en réanimation des patients positifs au variant Delta est significativement supérieur à 2 jour ?

Le chef de service de l'hôpital a affirmé que les patients atteints du variant Omicron ont des durées de séjour à l'hôpital moins longues que celles des patients atteints du variant Delta.

d) Peut-on conclure avec un risque d'erreur $\alpha = 5\%$, que l'affirmation émise par le chef de service est vraie ?

Exercice 3 : (Barème de notation : a) 1.5 pt b) 1.5 pts c) 1.5 pt = 4.5 pts)

L'objectif de cette étude est d'analyser les effets de certaines caractéristiques des voitures sur leur prix. On a extrait onze caractéristiques d'une base de données constituée de 27 petites voitures du marché belge.

DICTIONNAIRE DES VARIABLES

| Iden | Variable | Type |
|-----------|---|-----------------|
| 1 . PRIX | Prix de la voiture (MF Belge) | (CONTINUE) |
| 2 . MARQ | Marque du constructeur : Française Etrangère | (2 MODALITES) |
| 3 . PFIS | Puissance Fiscale - Nombre de chevaux : 4CV 5CV 6CV | (3 MODALITES) |
| 4 . CARB | Type de carburant : Diesel SP95 SP98 | (3 MODALITES) |
| 5 . CYLI | Cylindrée de la voiture (cm3) | (CONTINUE) |
| 6 . VITE | Vitesse (km/h) | (CONTINUE) |
| 7 . VOLU | Volume du coffre | (CONTINUE) |
| 8 . POID | Poids du véhicule | (CONTINUE) |
| 9 . REPR | Reprise | (CONTINUE) |
| 10 . FREI | Freinage | (CONTINUE) |
| 11 . BRUI | Bruit | (CONTINUE) |

Pour l'interprétation des résultats des modèles, justifier vos réponses en indiquant le numéro des résultats correspondants de l'annexe 2.

a) Peut-on considérer avec un risque d'erreur $\alpha = 5\%$, qu'il y a un effet de la marque du constructeur sur le prix de la voiture ? Etablir les hypothèses statistiques et interpréter les principaux résultats de ce modèle.

b) Peut-on conclure avec un risque d'erreur $\alpha = 5\%$, qu'il y a un effet de la marque du constructeur et de la puissance fiscale sur le prix de la voiture ? Donner une interprétation résumée des principaux résultats de ce modèle explicatif.

c) On a ajouté une nouvelle co-variable explicative cylindrée de la voiture - au modèle précédent. Interpréter les principaux résultats de ce nouveau modèle.



ANNEXE 1

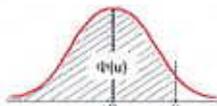


Table de la loi Normale Centrée Réduite : $U \rightarrow N(0; 1)$

Φ : Fonction de répartition : $\Phi(u) = P(U \leq u)$; $\Phi(-u) = P(U \leq -u) = 1 - \Phi(u)$

| u | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 0.0 | 0.50000 | 0.50399 | 0.50798 | 0.51197 | 0.51595 | 0.51994 | 0.52392 | 0.52790 | 0.53188 | 0.53586 |
| 0.1 | 0.53983 | 0.54380 | 0.54776 | 0.55172 | 0.55567 | 0.55962 | 0.56356 | 0.56749 | 0.57142 | 0.57535 |
| 0.2 | 0.57926 | 0.58317 | 0.58706 | 0.59095 | 0.59483 | 0.59871 | 0.60257 | 0.60642 | 0.61026 | 0.61409 |
| 0.3 | 0.61791 | 0.62172 | 0.62552 | 0.62930 | 0.63307 | 0.63683 | 0.64058 | 0.64431 | 0.64803 | 0.65173 |
| 0.4 | 0.65542 | 0.65910 | 0.66276 | 0.66640 | 0.67003 | 0.67364 | 0.67724 | 0.68082 | 0.68439 | 0.68793 |
| 0.5 | 0.69146 | 0.69497 | 0.69847 | 0.70194 | 0.70540 | 0.70884 | 0.71226 | 0.71566 | 0.71904 | 0.72240 |
| 0.6 | 0.72575 | 0.72907 | 0.73237 | 0.73565 | 0.73891 | 0.74215 | 0.74537 | 0.74857 | 0.75175 | 0.75490 |
| 0.7 | 0.75804 | 0.76115 | 0.76424 | 0.76730 | 0.77035 | 0.77337 | 0.77637 | 0.77935 | 0.78230 | 0.78524 |
| 0.8 | 0.78814 | 0.79103 | 0.79389 | 0.79673 | 0.79955 | 0.80234 | 0.80511 | 0.80785 | 0.81057 | 0.81327 |
| 0.9 | 0.81594 | 0.81859 | 0.82121 | 0.82381 | 0.82639 | 0.82894 | 0.83147 | 0.83398 | 0.83646 | 0.83891 |
| 1.0 | 0.84134 | 0.84375 | 0.84614 | 0.84849 | 0.85083 | 0.85314 | 0.85543 | 0.85769 | 0.85993 | 0.86214 |
| 1.1 | 0.86433 | 0.86650 | 0.86864 | 0.87076 | 0.87286 | 0.87493 | 0.87698 | 0.87900 | 0.88100 | 0.88298 |
| 1.2 | 0.88493 | 0.88686 | 0.88877 | 0.89065 | 0.89251 | 0.89435 | 0.89617 | 0.89796 | 0.89973 | 0.90147 |
| 1.3 | 0.90320 | 0.90490 | 0.90658 | 0.90824 | 0.90988 | 0.91149 | 0.91309 | 0.91466 | 0.91621 | 0.91774 |
| 1.4 | 0.91924 | 0.92073 | 0.92220 | 0.92364 | 0.92507 | 0.92647 | 0.92785 | 0.92922 | 0.93056 | 0.93189 |
| 1.5 | 0.93319 | 0.93448 | 0.93574 | 0.93699 | 0.93822 | 0.93943 | 0.94062 | 0.94179 | 0.94295 | 0.94408 |
| 1.6 | 0.94520 | 0.94630 | 0.94738 | 0.94845 | 0.94950 | 0.95053 | 0.95154 | 0.95254 | 0.95352 | 0.95449 |
| 1.7 | 0.95543 | 0.95637 | 0.95728 | 0.95818 | 0.95907 | 0.95994 | 0.96080 | 0.96164 | 0.96246 | 0.96327 |
| 1.8 | 0.96407 | 0.96485 | 0.96562 | 0.96638 | 0.96712 | 0.96784 | 0.96856 | 0.96926 | 0.96995 | 0.97062 |
| 1.9 | 0.97128 | 0.97193 | 0.97257 | 0.97320 | 0.97381 | 0.97441 | 0.97500 | 0.97558 | 0.97615 | 0.97670 |
| 2.0 | 0.97725 | 0.97778 | 0.97831 | 0.97882 | 0.97932 | 0.97982 | 0.98030 | 0.98077 | 0.98124 | 0.98169 |
| 2.1 | 0.98214 | 0.98257 | 0.98300 | 0.98341 | 0.98382 | 0.98422 | 0.98461 | 0.98500 | 0.98537 | 0.98574 |
| 2.2 | 0.98610 | 0.98645 | 0.98679 | 0.98713 | 0.98745 | 0.98778 | 0.98809 | 0.98840 | 0.98870 | 0.98899 |
| 2.3 | 0.98928 | 0.98956 | 0.98983 | 0.99010 | 0.99036 | 0.99061 | 0.99086 | 0.99111 | 0.99134 | 0.99158 |

Exemples : $\Phi(1.26) = P(U \leq 1.26) = 0.89617$; $\Phi(u) = P(U \leq u) = 97.50\% \Rightarrow u = 1.96$

ANNEXE 1

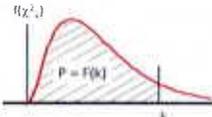


Table de la loi du $\chi^2_{v,d.d.l.}$ Fonction de répartition : $F(k) = P(\chi^2_v \leq k)$

| v P | 0,010 | 0,020 | 0,025 | 0,050 | 0,10 | 0,15 | 0,20 | 0,80 | 0,90 | 0,95 | 0,975 | 0,98 | 0,99 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 4 | 0,297 | 0,429 | 0,484 | 0,711 | 1,064 | 1,366 | 1,649 | 5,989 | 7,779 | 9,488 | 11,143 | 11,668 | 13,277 |
| 24 | 10,856 | 11,992 | 12,401 | 13,848 | 15,659 | 16,969 | 18,062 | 29,553 | 33,196 | 36,415 | 39,364 | 40,270 | 42,980 |
| 25 | 11,524 | 12,697 | 13,120 | 14,611 | 16,473 | 17,818 | 18,940 | 30,675 | 34,382 | 37,652 | 40,646 | 41,566 | 44,314 |
| 40 | 22,164 | 23,838 | 24,433 | 26,509 | 29,051 | 30,856 | 32,345 | 47,269 | 51,805 | 55,758 | 59,342 | 60,436 | 63,691 |
| 48 | 28,177 | 30,080 | 30,755 | 33,098 | 35,949 | 37,965 | 39,621 | 55,993 | 60,907 | 65,171 | 69,023 | 70,197 | 73,683 |
| 49 | 28,941 | 30,871 | 31,555 | 33,930 | 36,818 | 38,859 | 40,534 | 57,079 | 62,038 | 66,339 | 70,222 | 71,406 | 74,919 |
| 50 | 29,707 | 31,664 | 32,357 | 34,764 | 37,689 | 39,754 | 41,449 | 58,164 | 63,167 | 67,505 | 71,420 | 72,613 | 76,154 |

Exemples : $v = 4$ d.d.l. $F(k) = P(\chi^2_4 \leq k) = 0,95 \Rightarrow k = 9,488$ $k = 7,779 \Rightarrow F(7,779) = P(\chi^2_4 \leq 7,779) = 0,90$

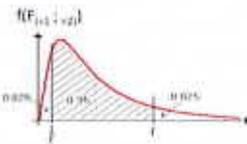


Table de la loi de Fisher-Snedecor

Valeur f de la variable de Fisher-Snedecor $F(v_1; v_2)$ ayant la probabilité 2,5% d'être dépassée.

v_1 : degrés de liberté du numérateur v_2 : degrés de liberté du dénominateur

Fonction de répartition : $F(f) = P(F(v_1, v_2) \leq f) = 97,50\%$

| $v_2 \ v_1$ | 2 | 4 | 20 | 24 | 25 | 40 | 45 | 49 | 50 |
|-------------|------|------|------|------|------|------|------|------|------|
| 24 | 4,32 | 3,38 | 2,33 | 2,27 | 2,26 | 2,15 | 2,12 | 2,11 | 2,11 |
| 25 | 4,29 | 3,35 | 2,30 | 2,24 | 2,23 | 2,12 | 2,10 | 2,08 | 2,08 |
| 40 | 4,05 | 3,13 | 2,07 | 2,01 | 1,99 | 1,88 | 1,85 | 1,84 | 1,83 |
| 45 | 4,01 | 3,09 | 2,03 | 1,96 | 1,95 | 1,83 | 1,81 | 1,79 | 1,79 |
| 49 | 3,98 | 3,06 | 2,00 | 1,94 | 1,92 | 1,80 | 1,78 | 1,76 | 1,76 |
| 50 | 3,97 | 3,05 | 1,99 | 1,93 | 1,92 | 1,80 | 1,77 | 1,76 | 1,75 |

Exemples : $v_1 = 25$ et $v_2 = 40$: $P(F_{97,5\% : 25 : 40} \leq f) = 0,975 \Rightarrow f = 1,99$

$P(F_{2,5\% : 25 : 40} \leq f') = 0,025$ $P(F_{97,5\% : 40 : 25} \leq f) = 0,975 \Rightarrow f = 2,12 \Rightarrow f' = \frac{1}{2,12} = 0,04717$

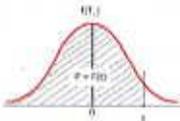


Table de la loi de Student - Fractiles de la loi de Student à v degrés de liberté

Valeur du fractile t - Fonction de répartition : $P = F(t) = P(T_v \leq t)$.

| v P | 0,7 | 0,75 | 0,8 | 0,85 | 0,9 | 0,95 | 0,955 | 0,96 | 0,965 | 0,97 | 0,975 | 0,99 | 0,995 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 10 | 0,542 | 0,700 | 0,879 | 1,093 | 1,372 | 1,812 | 1,877 | 1,948 | 2,028 | 2,120 | 2,228 | 2,764 | 3,169 |
| 24 | 0,531 | 0,685 | 0,857 | 1,059 | 1,318 | 1,711 | 1,767 | 1,828 | 1,896 | 1,974 | 2,064 | 2,492 | 2,797 |
| 25 | 0,531 | 0,684 | 0,856 | 1,058 | 1,316 | 1,708 | 1,764 | 1,825 | 1,893 | 1,970 | 2,060 | 2,485 | 2,787 |
| 40 | 0,529 | 0,681 | 0,851 | 1,050 | 1,303 | 1,684 | 1,737 | 1,796 | 1,862 | 1,936 | 2,021 | 2,423 | 2,704 |
| 49 | 0,528 | 0,680 | 0,849 | 1,048 | 1,299 | 1,677 | 1,730 | 1,788 | 1,852 | 1,925 | 2,010 | 2,405 | 2,680 |
| 50 | 0,528 | 0,679 | 0,849 | 1,047 | 1,299 | 1,676 | 1,729 | 1,787 | 1,852 | 1,924 | 2,009 | 2,403 | 2,678 |

Exemple : $v = 10$ d.d.l. $P(T_{10} \leq t) = 0,975 \Rightarrow t = +2,2281$ et $P(T_{10} \leq -t) = 0,025 \Rightarrow t = -2,228$

ANNEXE 2

N°: SAS1 ----- The GLM Procedure
 Levene's Test for Homogeneity of PRIX Variance
 ANOVA of Squared Deviations from Group Means

| Source | DDL | Somme | | Valeur | |
|--------|-----|------------|---------------------|--------|--------|
| | | des carrés | Moyenne quadratique | F | Pr > F |
| MARQ | 1 | 1.0041E8 | 1.0041E8 | 1.24 | 0.2758 |
| Error | 25 | 2.0217E9 | 80869829 | | |

N°: SAS2 ----- Marque du constructeur = Etrangère -----

Tests de normalité PRIX

| Test | -Statistique-- | -----p-Value----- |
|--------------|----------------|-------------------|
| Shapiro-Wilk | W 0.904815 | Pr < W 0.3190 |

N°: SAS3 ----- Puissance fiscale = 4CV -----

Tests de normalité PRIX

| Test | -Statistique-- | -----p-Value----- |
|--------------|----------------|-------------------|
| Shapiro-Wilk | W 0.973052 | Pr < W 0.9278 |

N°: SAS4 ----- Puissance fiscale = 6CV -----

Mesures statistiques de base
Tests de normalité PRIX

| Test | -Statistique-- | -----p-Value----- |
|--------------|----------------|-------------------|
| Shapiro-Wilk | W 0.931328 | Pr < W 0.7026 |

N°: SAS5 ----- Dependent Variable: PRIX Prix de la voiture -----

| Source | DDL | Somme des carrés | Moyenne quadratique | Valeur F | Pr > F |
|-----------------|-----|------------------|---------------------|----------|--------|
| Model | 6 | 160775.6414 | 26795.9402 | 22.00 | <.0001 |
| Error | 20 | 24360.5704 | 1218.0285 | | |
| Corrected Total | 26 | 185136.2119 | | | |

| Source | DDL | Type III SS | Moyenne quadratique | Valeur F | Pr > F |
|-----------|-----|-------------|---------------------|----------|--------|
| MARQ | 1 | 4253.79053 | 4253.79053 | 3.49 | 0.0764 |
| PFIS | 2 | 21225.81142 | 10612.90571 | 8.71 | 0.0019 |
| MARQ*PFIS | 2 | 6753.92827 | 3376.96413 | 2.77 | 0.0865 |
| CYLIN | 1 | 300.93682 | 300.93682 | 0.25 | 0.6246 |

N°: SAS6 ----- Variable : CYLI (Cylindrée cm3) -----

Tests de normalité CYLIN

| Test | -Statistique-- | -----p-Value----- |
|--------------|----------------|-------------------|
| Shapiro-Wilk | W 0.958902 | Pr < W 0.4168 |

The GLM Procedure

N°: SAS7 ----- Marque du constructeur = Française -----

Tests de normalité PRIX

| Test | -Statistique-- | -----p-Value----- |
|--------------|----------------|-------------------|
| Shapiro-Wilk | W 0.897271 | Pr < W 0.2730 |

N°: SAS8 ----- Dependent Variable: PRIX Prix de la voiture

| Source | DDL | Somme des carrés | Moyenne quadratique | Valeur F | Pr > F |
|-----------------|-----|------------------|---------------------|----------|--------|
| Model | 1 | 6901.7182 | 6901.7182 | 0.97 | 0.3346 |
| Error | 25 | 178234.4936 | 7129.3797 | | |
| Corrected Total | 26 | 185136.2119 | | | |

| Source | DDL | Type III SS | Moyenne quadratique | Valeur F | Pr > F |
|--------|-----|-------------|---------------------|----------|--------|
| MARQ | 1 | 6901.718205 | 6901.718205 | 0.97 | 0.3346 |

N°: SAS9 ----- The GLM Procedure

| Tukey Groupement | Moyenne | N | MARQ |
|------------------|---------|----|-----------|
| A | 340.22 | 10 | Française |
| A | 307.11 | 17 | Etrangère |

N°: SAS10 ----- Puissance fiscale = 5CV -----

| Test | -Statistique- | -----p-Value----- |
|--------------|---------------|-------------------|
| Shapiro-Wilk | W 0.9295 | Pr < W 0.5930 |

N°: SAS11 ----- Dependent Variable: PRIX Prix de la voiture

Informations sur le niveau de classe

| Classe | Niveaux | Valeurs |
|-----------------------------|---------|---------------------|
| MARQ | 2 | Etrangère Française |
| PFIS | 3 | 4CV 5CV 6CV |
| Number of Observations Read | | 27 |
| Number of Observations Used | | 27 |

| Source | DDL | Somme des carrés | Moyenne quadratique | Valeur F | Pr > F |
|-----------------|-----|------------------|---------------------|----------|--------|
| Model | 5 | 160474.7046 | 32094.9409 | 27.33 | <.0001 |
| Error | 21 | 24661.5072 | 1174.3575 | | |
| Corrected Total | 26 | 185136.2119 | | | |

| Source | DDL | Type III SS | Moyenne quadratique | Valeur F | Pr > F |
|-----------|-----|-------------|---------------------|----------|--------|
| MARQ | 1 | 4713.8719 | 4713.8719 | 4.01 | 0.0582 |
| PFIS | 2 | 143702.7666 | 71851.3833 | 61.18 | <.0001 |
| MARQ*PFIS | 2 | 11644.3207 | 5822.1603 | 4.96 | 0.0172 |

N°: SAS12 ----- The GLM Procedure

Levene's Test for Homogeneity of PRIX Variance
ANOVA of Squared Deviations from Group Means

| Source | DDL | Somme des carrés | Moyenne quadratique | Valeur F | Pr > F |
|--------|-----|------------------|---------------------|----------|--------|
| PFIS | 2 | 83493631 | 41746815 | 15.79 | <.0001 |
| Error | 24 | 63446427 | 2643601 | | |