

Nom de naissance :

Premier  
prénom :Numéro  
candidature :

15 / 20

(Remplir cette partie à l'aide de la notice)

Concours / Examen : Inspecteur externeSession : 2024Epreuve n° : 2Matière : Econometrie et statistiques

## CONSIGNES

- Remplir soigneusement, sur CHAQUE feuille officielle, la zone d'identification en MAJUSCULES.
- Ne pas signer la composition et ne pas y apporter de signe distinctif pouvant indiquer sa provenance.
- Numéroter chaque feuille A3 dans le cadre à droite et placer les feuilles dans le bon sens et dans l'ordre.
- Rédiger avec un stylo à encre foncée (bleue ou noire) et ne pas utiliser de stylo plume à encre claire.
- N'effectuer aucun collage ou découpage de sujets ou de feuille officielle. Ne joindre aucun brouillon.

Feuille :

0	1	/	0	2
---	---	---	---	---

Exercice 1.

1. Un caractère statistique est un caractère que l'on peut accorder à un individu dans une population. Ce caractère peut être continu, comme l'âge d'une personne (tout réel positif), ou qualitatif, comme son sexe (homme ou femme). Dans ce cas on dit aussi qu'il s'agit d'un booléen et on peut par exemple décider d'accorder une valeur à une qualité (par exemple, "femme" = 1 et "homme" = 0). Un caractère statistique ne peut donc pas être à la fois qualitatif et continu.

FAUX

2. Le premier quantile, généralement noté  $Q_1$ , est la valeur étudiée de l'individu tel qu'un quart des individus de l'échantillon aient une valeur supérieure à  $Q_1$ . De la même façon, la médiane est la valeur qui sépare en deux moitiés l'échantillon. Les individus de la 1<sup>ère</sup> moitié ont une valeur inférieure à la médiane, la seconde une valeur supérieure. Le premier quantile est donc toujours inférieur à la médiane, mais il peut aussi lui être égal. Il n'est donc pas toujours strictement inférieur à la médiane.

FAUX

3. Il n'est pas précisé de quel intervalle de confiance on parle. Admettons qu'on cherche par exemple à estimer un caractère qualitatif moyen, par exemple la taille moyenne d'un groupe de  $N$  personnes. Dans ce cas on cherche un intervalle de confiance autour de la vraie valeur de la moyenne, en calculant un estimateur de cette moyenne. Plus  $N$  sera grand, plus la variance de l'estimateur sera faible, et la longueur de l'intervalle de confiance aussi, rapprochant le résultat de l'enquête de la réalité. FAUX

4. La question est: Si on a  $(X_n)$  suite de v.a.n. t.q.  
 $(X_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X = A$  loi constante ;

alors a-t-on,  $\forall a \in X(\Omega)$ ,  
 $\lim_{n \rightarrow +\infty} P(X_n = a) = P(X = a)$  ?

Comme on a la convergence en loi, on peut déduire que les  $X_i$ ,  $i \in \llbracket 1; n \rrbracket$  sont indépendantes, indépendamment distribués (i.i.d.).

Le plus, par définition de la convergence en loi, on a la convergence en probabilités pour tout réel  $a$  appartenant à  $X(\Omega)$ . Ici  $X$  étant constante,  $X(\Omega)$  est réduite au singleton d'un réel, que j'ai noté  $a$ .

VRAI

~~Soit  $S$  la variable aléatoire réelle (v.a.r.) égale au rang du premier agent homme affecté dans le nouveau service de VILLOUVEF.~~

~~Par construction,  $S \sim \mathcal{G}(p = 1 - 0,58 = 0,42)$  c'est-à-dire (càd.) que~~

### Exercice 2

Le nouveau service comporte  $n$  agents. On cherche  $n$  tel que la probabilité de l'événement  $H =$  "il y a au moins un homme parmi les  $n$  agents" soit supérieure à 0,9. Notons que  $\bar{H} =$  "il n'y a que des femmes parmi les  $n$  agents".

Les arrivées des agents étant indépendantes, on a que  $P(\bar{H}) = 0,58^n$ .

Donc  $P(H) = 1 - 0,58^n$ .

On cherche  $n$  tel que  $P(H) \geq 0,9$

$$\Leftrightarrow 1 - 0,58^n \geq 0,9 \Leftrightarrow -0,58^n \geq -0,1 \Leftrightarrow 0,58^n \leq 0,1$$

(car on multiplie de part et d'autre par un négatif),

$$\Leftrightarrow n \ln 0,58 \leq \ln 0,1 \quad \text{car } \ln \text{ est une fonction croissante,}$$

$$\Leftrightarrow n \geq \frac{\ln 0,1}{\ln 0,58} \quad \text{car } \ln 0,58 \text{ est négatif.}$$

$$\Leftrightarrow n \geq 4,23 \Leftrightarrow n \geq 5 \quad \text{car } n \text{ est un entier naturel.}$$



Conclusion : il faut au moins mettre 5 agents dans le service pour que la probabilité qu'au moins l'un d'entre eux soit un fumeur soit supérieure à 90%.

### Exercice 3

1. On note A l'événement "les 11 premières personnes prennent 5 croissants parmi les 6" et B l'événement "la 12<sup>e</sup> personne prend un croissant". Ainsi l'événement décrit est  $(B \cap A)$ , la 12<sup>e</sup> personne prend le 6<sup>e</sup> et dernier croissant.

On cherche donc  $P(A \cap B)$  ou d'après les probabilités conditionnelles,

$$P(A \cap B) = P(A) \times P_A(B)$$

loi hypergéométrique de tirage sans remise  
tirer le croissant dans un panier composé de 1 croissant et 6 pains au chocolat.

$$P(A) = \frac{\text{nb de cas favorables}}{\text{nb de cas possibles}} = \frac{\binom{6}{5} \binom{12}{6}}{\binom{18}{11}} = \frac{6!}{5!1!} \times \frac{12!}{6!6!} \times \frac{11!7!}{18!}$$

$$P(A) = \frac{12!11!7!}{5!6!18!} = \frac{12!11!6! \times 7}{6!6!18!} = \frac{12!11!}{5!18!} \times 7$$

$$P_A(B) = \frac{1}{18-11} = \frac{1}{7}$$

$$\text{donc } P(A \cap B) = \frac{12!11!}{5!18!} \times 7 \times \frac{1}{7} = \frac{12!11!}{5!18!} = 0,0249$$

la probabilité que le douzième collègue prenne le dernier croissant est 0,0249

2.

#### Exercice 4.

On a deux séries de données, les points  $(x_i; y_i)$   $\forall i \in \llbracket 1; 10 \rrbracket$ .

On cherche s'il existe une corrélation entre les des  $x_i$  et celle des  $y_i$ .

On commence par calculer le coefficient de corrélation:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = r. \text{ La calculatrice fournit la valeur } r = 0,8766.$$

La valeur de  $r$  étant assez proche de 1, on peut penser qu'il y a une certaine corrélation entre  $X$  et  $Y$ . Mais pour aller plus loin, il faut tester la significativité de la pente de la régression linéaire de  $Y$  sur  $X$ .

On note que la statistique:

$$T = \sqrt{n-2} \frac{r}{\sqrt{1-r^2}} \rightarrow T(n-2)$$

suit une loi de Student à  $n-2=8$  degrés de liberté.

Calculons la réalisation  $t$  de notre statistique.

$$\text{On a : } \sqrt{10-2} \frac{0,87661}{\sqrt{1-0,8766^2}} \approx 5,152$$

On se reporte à la table de la loi de Student, à la ligne DDL=8.

Notre hypothèse  $H_0$  est que la pente de la régression linéaire est nulle. Soit  $H_0: \beta_1=0$  avec  $Y = \beta_0 + \beta_1 X + \varepsilon$ .

Notre réalisation  $t = 5,15$  dépasse la valeur seuil de signification pour le test bilatéral, y compris à un seuil de  $\alpha = 1\% = 0,001$ .

On peut donc rejeter  $H_0$  au seuil de certitude de 1% et conclure que la régression linéaire est significative, donc que les données  $x_i$  et  $y_i$  sont bien corrélées.



Nom de naissance :



Premier prénom :

Numéro candidature :

15 / 20

(Remplir cette partie à l'aide de la notice)

Concours / Examen : Inspéteur externe

Session : 2024

Epreuve n° : 2

Matière : Économétrie et statistiques

**CONSIGNES**

- Remplir soigneusement, sur CHAQUE feuille officielle, la zone d'identification en MAJUSCULES.
- Ne pas signer la composition et ne pas y apporter de signe distinctif pouvant indiquer sa provenance.
- Numéroté chaque feuille A3 dans le cadre à droite et placer les feuilles dans le bon sens et dans l'ordre.
- Rédiger avec un stylo à encre foncée (bleue ou noire) et ne pas utiliser de stylo plume à encre claire.
- N'effectuer aucun collage ou découpage de sujets ou de feuille officielle. Ne joindre aucun brouillon.

Feuille :

02 / 02

Exercice 5 1.

Soit  $X_1, \dots, X_m$  avec  $m = 90$  le résultat de chacune des  $m$  réservations.

Si la réservation  $i$  est confirmée,  $X_i = 1$ . Sinon,  $X_i = 0$ .

On sait que  $\forall i \in \llbracket 1; m \rrbracket; X_i \hookrightarrow \mathcal{B}(p)$  avec  $p = 0,8$ .

Càd que les  $X_i$  sont i.i.d. et suivent une loi de Bernoulli de paramètre  $0,8$ .

Donc  $E(X_i) = 0,8$  et  $V(X_i) = 0,8 \times 0,2 = 0,16$ ,  $\forall i$ .

$N$  est le nombre de confirmations total. Donc  $N = \sum_{i=1}^m X_i$  avec  $m = 90$ .

Donc  $E(N) = E\left(\sum_{i=1}^m X_i\right) = \sum_{i=1}^m E(X_i) = m E(X_1) = mp = 90 \times 0,8 = 72 = E(N)$

↓  
par linéarité de l'espérance  
car les  $X_i$  sont de même loi

De même,  $V(N) = V\left(\sum_{i=1}^m X_i\right) = \sum_{i=1}^m V(X_i) = m V(X_1) = mpq = 90 \times 0,16 = 14,4 = V(N)$

et  $\sigma_N = \sqrt{14,4} = 3,79$ .

↓  
car indépendance deux à deux des  $X_i$

2. D'après le théorème central limite, toutes les conditions ont été vérifiées dans la question précédente, on a :

$$\frac{N - E(N)}{\sqrt{m} \sigma_N} \hookrightarrow N(0,1)$$

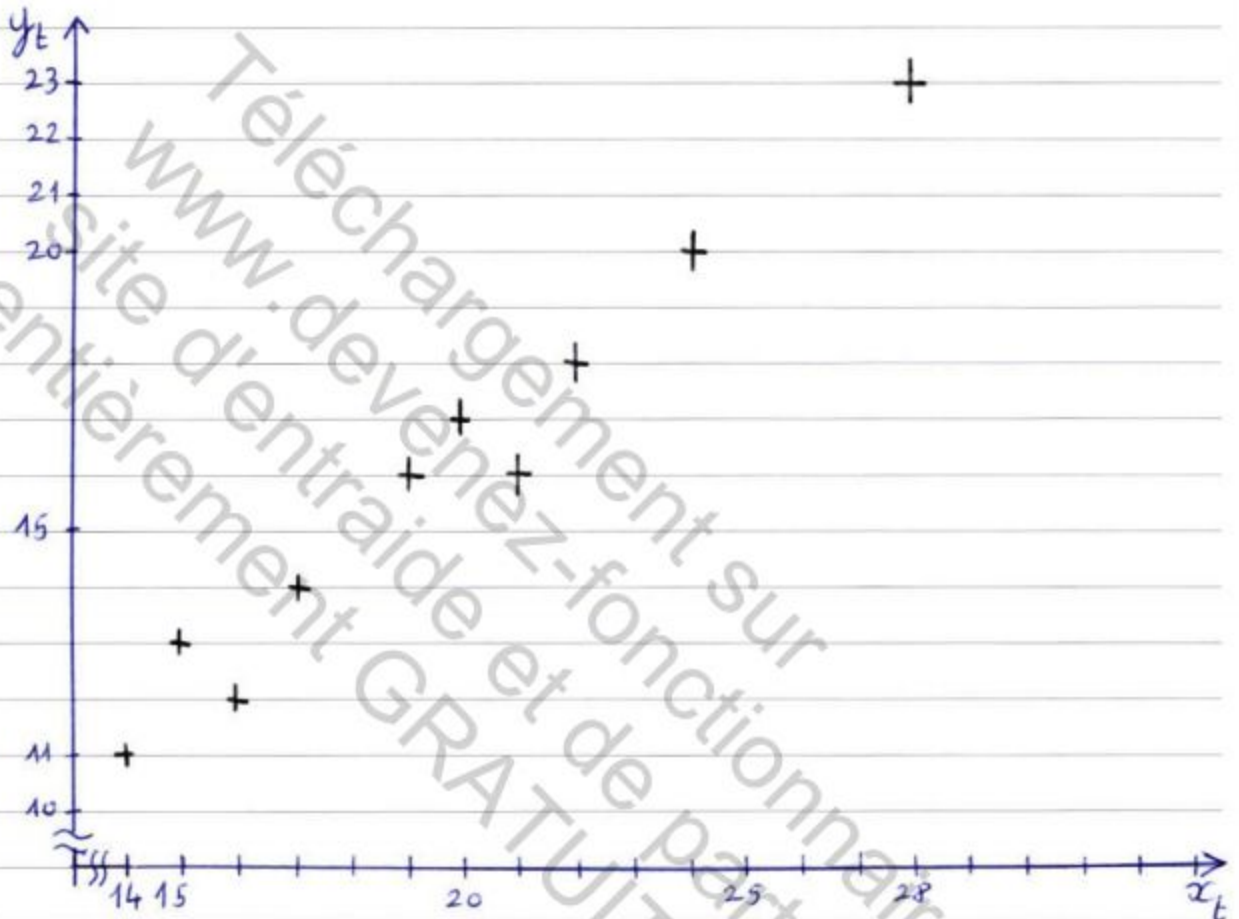
$$\text{On cherche } P(N \geq 70) = 1 - P(N < 70) = 1 - P\left(\frac{N - E(N)}{\sqrt{m} \sigma_N} < \frac{70 - E(N)}{\sqrt{m} \sigma_N}\right)$$

$$= 1 - P\left(\frac{N - E(N)}{\sqrt{m} \sigma_N} < \frac{70 - 72}{\sqrt{90} \sqrt{14,4}}\right) = 1 - \Phi(-0,055) = \Phi(0,055)$$

avec  $\Phi$  fonction de répartition d'une loi normale centrée réduite.

On trouve dans la table  $\Phi(0,055) \approx 0,5239$

Donc  $P(N \geq 70) = 0,5239$

Exercice 6

1. Le nuage de points ci-dessus suggère une corrélation linéaire entre  $y_t$  et  $x_t$ .

2. On calcule les coefficients de la régression linéaire suivante :

$$y_t = b + ax_t + \varepsilon_t$$

On sait que  $a = \frac{\text{cov}(Y, X)}{V(X)}$ . On calcule  $a$  à la calculatrice.

$$a \approx 0,839$$

On sait aussi que la droite de régression passe par le point moyen  $(\bar{x}, \bar{y})$ .

$$\text{Donc } \bar{y} = b + a\bar{x} \Leftrightarrow b = -(\bar{x} - \bar{y})$$

On calcule  $b$  à la calculatrice.

$$b \approx -0,448$$

On a aussi  $r \approx 0,984$ ; ce qui suggère en effet une bonne corrélation.



Années	t	$y_t$	$x_t$	$\hat{y}_t = \hat{\beta} + \hat{\alpha}x_t$ ②	$e_t = y_t - \hat{y}_t$
2012	1	11	14	11,298	-0,298
2013	2	13	15	12,137	0,863
2014	3	14	17	13,815	0,185
2015	4	16	19	15,493	0,507
2016	5	12	16	12,976	-0,976
2017	6	17	20	16,332	0,668
2018	7	20	24	19,688	0,312
2019	8	16	21	17,171	-1,171
2020	9	18	22	18,010	-0,010
2021	10	23	28	23,044	-0,044

2. les régresses se situent en colonne ②.

3. les résidus sont les  $e_t = y_t - \hat{y}_t$ . Ils sont en colonne ③.

$\bar{e}_t \approx 0$  au millième près, on peut donc affirmer que la moyenne des résidus est bien nulle.

4. L'erreur  $E_t = e_t + er_t$  avec  $er_t$  l'erreur due à la régression, et  $E_t$  l'erreur totale.

On a  $er_t = \hat{y}_t - \bar{y}$ ; pour rappel  $e_t = y_t - \hat{y}_t$  est l'erreur due aux résidus, donc

$$E_t = y_t - \bar{y}. \text{ On cherche à estimer } \sum_{t=1}^{10} (y_t - \bar{y})^2.$$

Calculons son espérance.

$$E(E_t) = \sum_{t=1}^{10} E((y_t - \bar{y})^2) = \sum E(y_t^2) + E(\bar{y}^2) - 2E(y_t \bar{y})$$

$$= 10(E(y_t^2)) \dots = MV(E_t)$$

$$\text{Donc soit } S_T = \frac{1}{\sqrt{n}} \sum_{t=1}^{10} (y_t - \bar{y})^2. \quad E(S_T) = \left(\frac{1}{\sqrt{n}}\right)^2 MV(E_t) = V(E_t).$$

$S_T$  est l'estimateur de la variance de l'erreur.

5. On teste la significativité de la pente, c'est l'hypothèse  $H_0: "a=0"$ .

Sous  $H_0$  on sait que  $\frac{1}{\sqrt{10-2}} \frac{1}{\sqrt{1-\alpha^2}} \hookrightarrow T(10-2) \text{ DOL}$ .

$$\text{Notre réalisation } t \text{ est } = \sqrt{8} \frac{10,8861}{\sqrt{1-0,994^2}} \approx 15,62$$

la valeur seuil de la loi de Student à 8 DOL et à 5% de risque de première espèce, pour le test bilatéral, est d'après la table de 2,306.

Notre réalisation étant bien supérieure, on peut rejeter  $H_0$  à 5%, et conclure que notre régression est bien significative à 5%.

6. On sait que  $\sqrt{p} \frac{\hat{a} - a}{\sigma_a} \hookrightarrow T(8) \text{ DDL}$ .

$\Rightarrow$  au seuil bilatéral de 5%,  $T(8) \text{ DDL}$  a pour valeur seuil 2,306.  
Donc un intervalle de confiance au niveau 95% approchant la vraie valeur de  $a$  sera :

$$[0,61 - 0,839]$$

$$= [2,$$

$$= [2, \dots \times 0,$$

Téléchargement sur  
[www.devenez-fonctionnaire.fr](http://www.devenez-fonctionnaire.fr)  
Site d'entraide et de partage  
entièrement GRATUIT



Question 4. On demande de calculer "l'estimateur de la variance de l'erreur". Le nom peut changer d'un énoncé à l'autre, mais dit tel quel et surtout après une question qui demandait de calculer les résidus il s'agit de calculer l'estimateur  $s$  avec:

$$s^2 = \frac{1}{n-2} \sum_i (y_i - \hat{y}_i)^2 = \frac{1}{n-2} \sum_i e_i^2 = \frac{1}{8} \times 3,994 = 0,499 = s^2$$

Ici, les  $e_i$ ,  $\forall i \in [1; 10]$  ayant été calculés, on obtient:  $s = 0,707$  (arrondi au millième).

Question 5: voir ci-dessus.

Question 6. L'intervalle de confiance pour  $a$  à 95% s'écrit:

$$IC = [\hat{a} - t \hat{\sigma}_a; \hat{a} + t \hat{\sigma}_a]$$

avec:  $\hat{a}$  la valeur estimée de  $a$ ,  $= 0,839$  ici;

$t$  réalisation d'une loi de Student à  $n-2 = 8$  D.D.L, ici  $t = 2,306$  au risque de 5%;

$\hat{\sigma}_a$  l'écart-type estimé de  $a$ .

$$\hat{\sigma}_a^2 = \frac{s^2}{\sum_i (x_i - \bar{x})^2} = \frac{s^2}{V(x) (n-1)}$$

$$\text{Donc } \hat{\sigma}_a = \frac{s}{s_x \sqrt{n}} = 0,05413$$

$s_x^2$  dans la calculatrice sachant que  $(n-1)s_x = n\sigma_x$

$$\text{Donc } IC = [0,714; 0,964]$$

Question 7 la variance totale est égale à celle due aux résidus + celle due à la régression. On note les sommes des carrés ainsi:

$$SST (\text{total}) = SSR (\text{résidus}) + SSE (\text{régression}).$$

$$\text{Avec } SSR = \sum_i e_i^2 = 3,994 \text{ et } SSE = \sum_i (\hat{y}_i - \bar{y})^2$$

$$\text{On a } r^2 = \frac{SSE}{SST} = \frac{SSE}{SSE + SSR}$$

Ici  $SSE \approx 120$ , elle est donc en effet bien supérieure à  $SSR$ .

→ plus  $r^2$  est proche de 1, et plus la variance totale est surtout composée de la variance due à la régression, ce qui confirme que la régression est de bonne qualité (on explique une grande partie de la variance observée).