

Corrigé type examen final

GADM

Master 1

Exercice 1:

1 - Les conditions de validité d'un test ANOVA sont :

- 1 - L'indépendance des échantillons. 0,5 pt
- 2 - La normalité de la distribution. 0,5 pt
- 3 - L'homogénéité des variances. 0,5 pt

2 - Oui il existe d'autres tests à effectuer 1 pt

Comme le test de Kruskal et Wallis pour la normalité, il ya d'autres tests aussi. 1 pt.

3 - La différence entre ANOVA et ANCOVA c'est que ANOVA est utilisé si on explique une variable quantitative avec des variables qualitatives et ANCOVA est utilisé si on explique une variable quantitative avec des variables qualitatives et quantitatives. Dans ce cas, la variable quantitative explicative est dite : covariable. 1,5 pts

4 - On effectue un test ANOVA 1 pt.

Par il ya un seul facteur qui est la catégorie du désordre 1 pt

- $F_{obs} = 9,37$.

On pose un test d'hypothèse.

H_0 : il n'y a pas une différence significative entre les 3 catégories

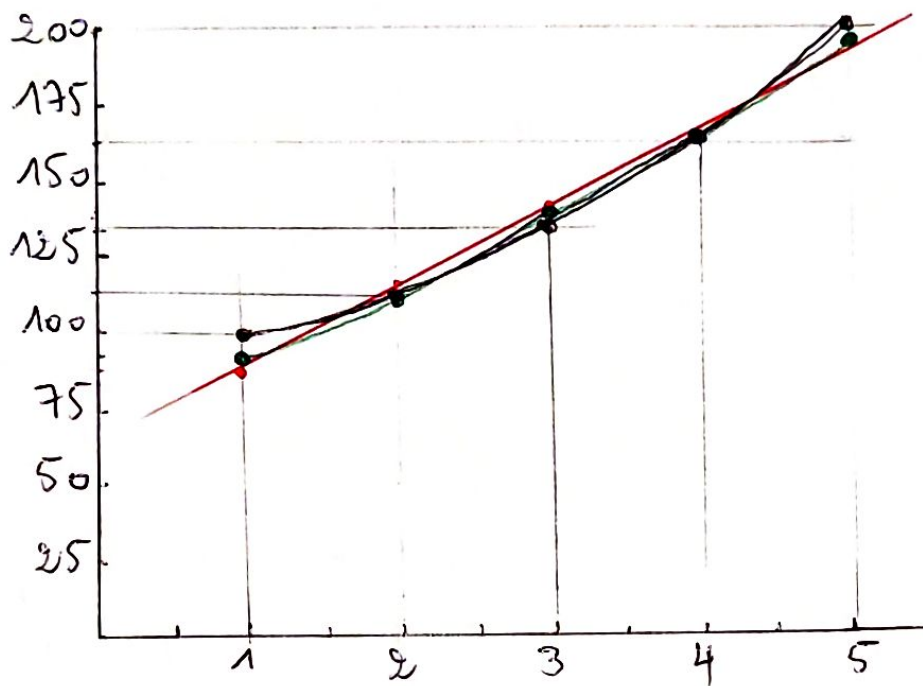
H_1 : il y a une différence significative entre les 3 catégories.

2 pt.

$$F_{0,05,2,9} = 4,26.$$

$F_{obs} > F_{0,05,2,9}$ donc on rejette H_0

Exercice 2:



- Selon le tableau, le modèle le plus approprié est le modèle exponentiel, car les x_i sont en progression arithmétique et les y_i en progression géométrique. L'équation est $y = b \times e^{ax}$.
 - le modèle exponentiel est calculé comme suit :
- selon le tableau, on calcule $Y = \ln y$.

Année	x_i	y_i	$Y = \ln y_i$	x_i^2	$x_i Y_i$
2018	1	100	4,60	1	4,60
2019	2	112	4,71	4	9,42
2020	3	130	4,96	9	14,58
2021	4	160	5,07	16	20,28
2022	5	200	5,29	25	26,45

$$\sum x_i = 15, \sum Y_i = 24,53, \sum x_i^2 = 55, \sum x_i Y_i = 75,33$$

$$\bar{n} = 3, \bar{Y} = 4,90, \sigma_n^2 = \frac{55}{5} - 3^2 = 2, \sigma_{nY} = \frac{75,33}{5} - 3 \times 4,90 = 0,36$$

$$a = \frac{0,36}{2} = 0,18$$

$$\text{On a } Y - \bar{Y} = a(n - \bar{n})$$

$$\Rightarrow Y - 4,90 = 0,18(n - 3)$$

$$\Rightarrow Y = 0,18n + 4,36$$

2pts

$$\Rightarrow \ln Y = 0,18n + 4,36$$

$$\Rightarrow Y = e^{0,18n + 4,36}$$

$$\Rightarrow Y = e^{0,18n} \times e^{4,36}$$

$$\Rightarrow Y = 78,25 \times e^{0,18n}$$

- Les y estimés sont:

$$\hat{y}_1 = 78,25 \times e^{0,18} = 93,68$$

$$\hat{y}_2 = 78,25 \times e^{0,36} = 112,15$$

1,5pts

$$\hat{y}_3 = 134,27$$

$$\hat{y}_4 = 160,75$$

$$\hat{y}_5 = 192,46$$

La courbe des $\hat{y}_i$ est représentée en vert sur le premier

plan.

- pour le modèle $y = 24,8n + 66$ (est un modèle linéaire) (1pt)
 $\hat{y}_1 = 90,8, \hat{y}_2 = 115,6, \hat{y}_3 = 140,4, \hat{y}_4 = 165,2, \hat{y}_5 = 190$

La droite de regression est en rouge.

$$- \bar{y} = 140,4$$

Pour le modèle $y = 78,25 \times e^{0,18n}$

$$SCE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 6142,76$$

$$SCT = \sum_{i=1}^n (y_i - \bar{y})^2 = 6483,2$$

$$\text{le } R^2 = \frac{6142,76}{6483,2} \approx 0,94$$

Pour le modèle $y = 24,8n + 66$.

La $SCE = 6150,4$ et la SCT est la même, c'est à dire
 $SCT = 6483,2$ donc $R^2 = \frac{6150,4}{6483,2} \approx 0,94$.

Pour $y = 78,2 \times e^{0,18n}$.

Les ε_i sont $\varepsilon_1 = 6,32, \varepsilon_2 = -0,15, \varepsilon_3 = -4,27, \varepsilon_4 = -0,71$
 $\varepsilon_5 = 7,54$ donc $\sum \varepsilon_i^2 = 115,61$. (2 pts)

Pour $y = 24,8n + 66$

Les ε_i sont $\varepsilon_1 = 9,2, \varepsilon_2 = -3,6, \varepsilon_3 = -10,4, \varepsilon_4 = -5,2, \varepsilon_5 = 10$
donc $\sum \varepsilon_i^2 = 332,8$.

- Selon les R^2 des deux modèles, on peut dire que c'est pratiquement la même valeur donc on ne peut pas choisir entre les deux modèles. Par en réalité, on a un tableau constitué de 5 valeurs et ce nombre est insuffisant. (2 pts)
- La $\sum \varepsilon_i^2$ du premier modèle est nettement inférieure à celle du 2^{eu} modèle, cela montre que le premier modèle est meilleur que le deuxième. On peut aussi voir ça en regardant les 3 courbes. La noire et la verte vont vers la même direction (montent d'une manière exponentielle) alors que la droite en rouge se détache des deux courbes en $n = 5$.
- Le nombre d'abonnés en 2023 ≈ 230 abonnés. (0,5 pt)