

Coller ici votre troisième étiquette.

Année universitaire 2017–2018
Session 1 - Semestre 1

Licence 1 mention Economie, parcours Economie – Mathématiques et
Informatique Appliquées

Epreuve : Statistique

Enseignant : E. Leconte

Date de l'épreuve : 11/12/2017

Durée de l'épreuve : 1h30

Liste des documents autorisés : aucun.

Liste des matériels autorisés : calculatrice de type Casio FX 92.

Nombre de pages : 9.

Répondre directement sur ce sujet aux endroits indiqués sans dégrafer les feuilles.
Justifier et détailler les calculs. La note tiendra compte de la qualité de la rédaction.
Arrondir tous les calculs à deux décimales.

Barème indicatif : A : 1 – B : 7 – C : 1 – D : 6 – E : 5.

Description des données pour tout le devoir

On dispose de données sur des appartements en location situés dans une ville américaine.
Les variables sont les suivantes :

- loyer : loyer mensuel (en dollars),
- surface : surface de l'appartement (en m^2),
- statio : nature du stationnement fourni avec l'appartement (aucun, place de parking, garage fermé),
- balcon : oui si présence d'un balcon, non sinon.

On fournit les sorties de R suivantes (arrondies à deux décimales), **qui vous seront nécessaires pour toute la suite du devoir** :

```
> table(statio)
statio
  aucun parking  garage
     2         10      19
> summary(loyer)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
540.0  664.0   798.0   814.5  925.0  1220.0
```

```

> summary(surface)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 30.00  33.00   36.00   37.74   41.00   55.00
> var(loyer) # attention, le dénominateur est (n-1)
[1] 33960.19
> var(surface) # attention, le dénominateur est (n-1)
[1] 35.93
> cor(loyer,surface)
[1] 0.90
> range(reg$residuals) # on note reg la sortie de lm(loyer~surface)
[1] -171.44  153.11
> table(balcon,statio)
      statio
balcon aucun parking garage
   non      2         5      10
   oui      0         5       9
> tapply(loyer,statio,mean)
 aucun parking  garage
549.50  720.80  891.74
> tapply(loyer^2,statio,mean) # moyenne des carrés des loyers
# selon la nature du stationnement
 aucun parking  garage
302040.5 532640.2 823939.7

```

A Questions générales

1. Quelle est la population étudiée? Quelle est sa taille (on la notera n)?
2. Donner le type de chacune des quatre variables étudiées.

B Etude de la liaison entre le loyer (noté Y) et la surface (notée X)

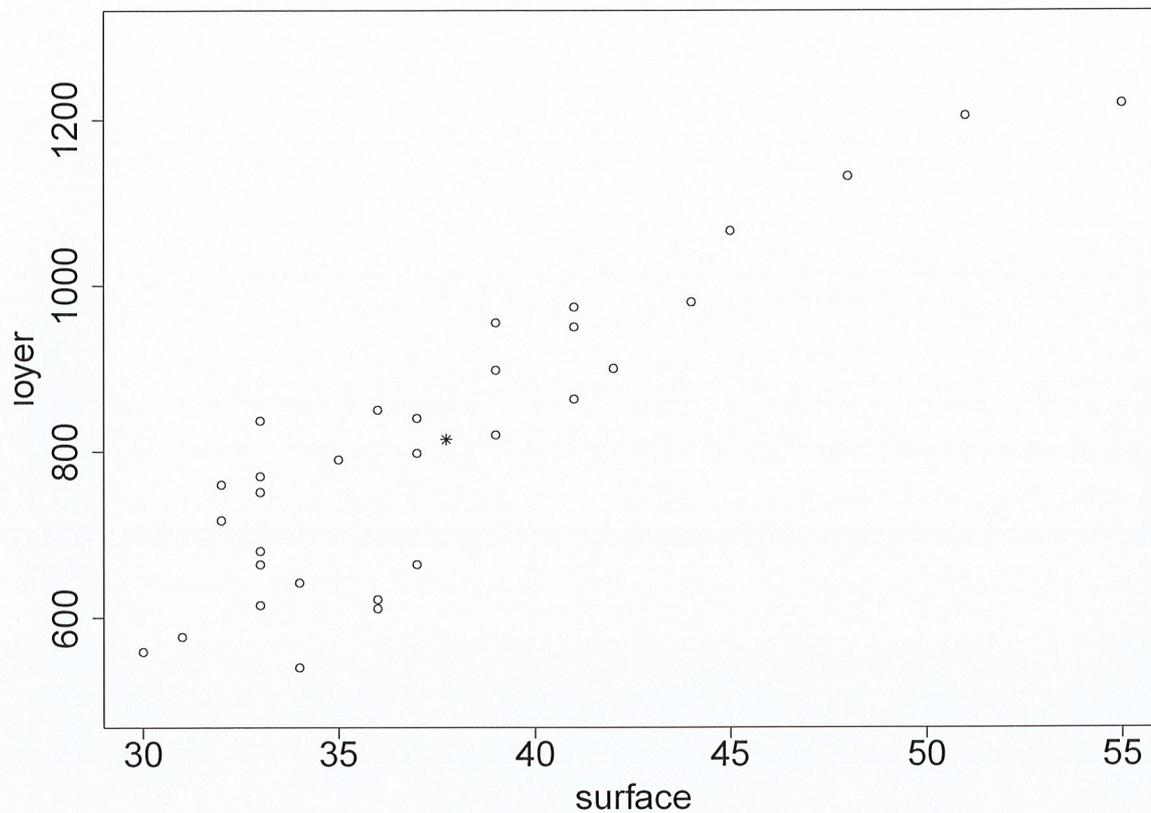
1. Calculer les variances puis la covariance des variables surface et loyer.

Var X =

Var Y =

Cov(X,Y)=

2. Commenter l'allure du nuage de points ci-dessous et dire quel modèle il vous inspire. Dire à quoi correspond le point *.



Modèle :

Point * :

3. Déterminer l'équation de la droite de régression du montant du loyer sur la surface. On notera \hat{a} et \hat{b} respectivement la pente et l'ordonnée à l'origine de cette droite.

Formule théorique pour \hat{a} :

Calcul de \hat{a} :

Formule théorique pour \hat{b} :

Calcul de \hat{b} :

Equation de la droite :

4. Tracer cette droite sur le nuage de point de la question 2.
5. Calculer le coefficient de détermination R^2 du modèle et interpréter concrètement la valeur de ce coefficient. Conclure sur la qualité du modèle.

Calcul de R^2 :

Interprétation de la valeur :

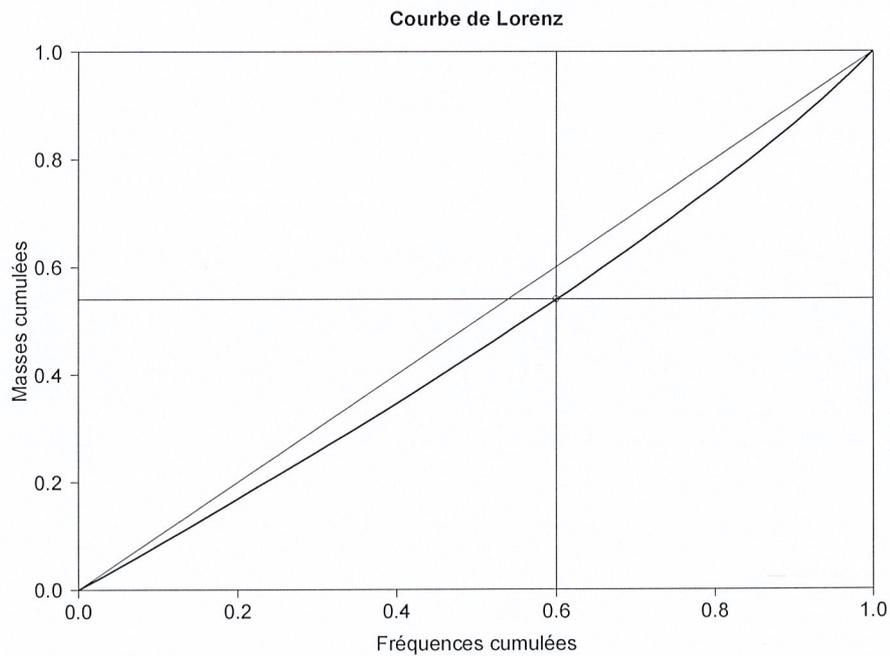
Conclusion :

6. D'après ce modèle, combien faut-il déboursier de plus chaque mois si on souhaite disposer de $5 m^2$ supplémentaires ?
7. Quel montant du loyer prédit ce modèle pour un appartement de $40 m^2$?
8. Rappeler la définition des résidus de la régression (on les notera \hat{e}_i) et montrer que la somme des résidus est toujours nulle (démonstration faite en TD).

9. Entourer sur le nuage de points l'appartement dont le loyer est le moins bien prédit par ce modèle et expliquer pourquoi. Que vaut son résidu ?

C Etude de la concentration des surfaces des appartements

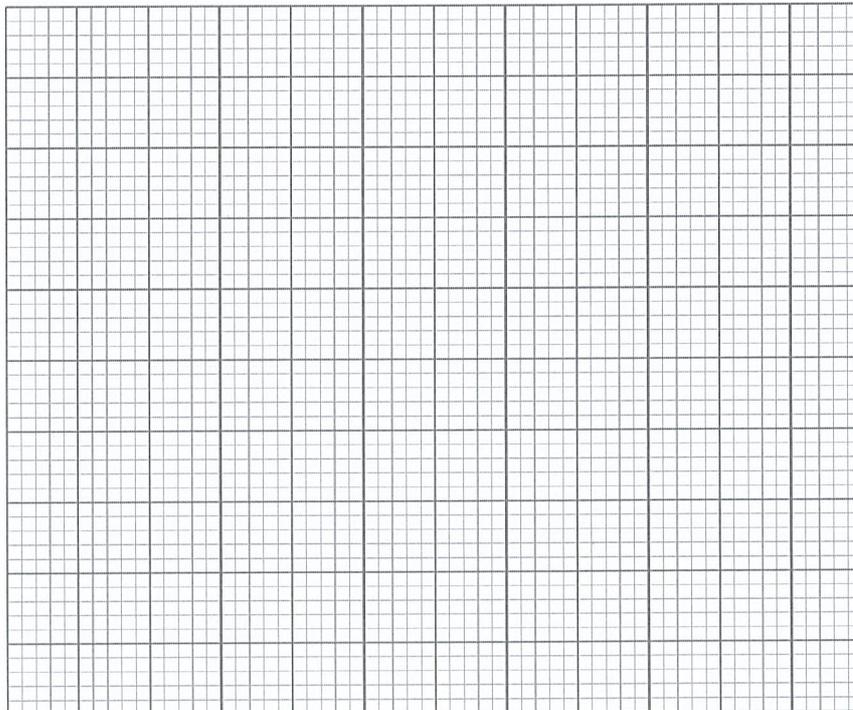
Le graphique ci-dessous a été tracé pour la variable surface à partir des données brutes.



1. Faire une phrase pour interpréter le point entouré (de coordonnées $(0,6 ; 0,54)$).
2. Sachant que l'aire sous la courbe de Lorenz vaut 0,46, calculer l'indice de Gini correspondant et commenter sa valeur.

D Etude de la liaison entre la nature du stationnement et la présence d'un balcon

1. Donner le tableau de contingence du couple (balcon, statio) (avec balcon en ligne et statio en colonne).
2. En déduire la distribution marginale de la variable balcon (en effectif et en fréquence relative).
3. Etablir le tableau des profils-colonnes et tracer les graphiques en colonnes juxtaposés associés à ces distributions.



4. Compléter le tableau suivant qui contient les effectifs théoriques t_{ij} attendus en cas d'indépendance parfaite entre les deux variables (donner la formule théorique et ne détailler le calcul que pour la cellule non/parking).

$t_{ij} =$

$t_{12} =$

	aucun	parking	garage
non	1,10		
oui			

5. On donne le tableau suivant qui contient pour chaque cellule sa contribution c_{ij} au calcul de la distance du χ^2 :

	aucun	parking	garage
non		0,04	0,02
oui	0,90	0,05	0,02

- (a) Rappeler la formule de c_{ij} et calculer la valeur manquante du tableau.

$c_{ij} =$

$c_{11} =$

- (b) Rappeler la formule théorique de la distance du χ^2 et la calculer.

Formule théorique : $\chi^2 =$

Application numérique : $\chi^2 =$

- (c) Quel est le couple de modalités qui apporte la plus grande contribution à la distance du χ^2 (ce n'est pas la valeur manquante) ? Qualifier l'association entre ces modalités.

6. Calculer l'indice C de Cramer et conclure sur la liaison entre les deux variables.

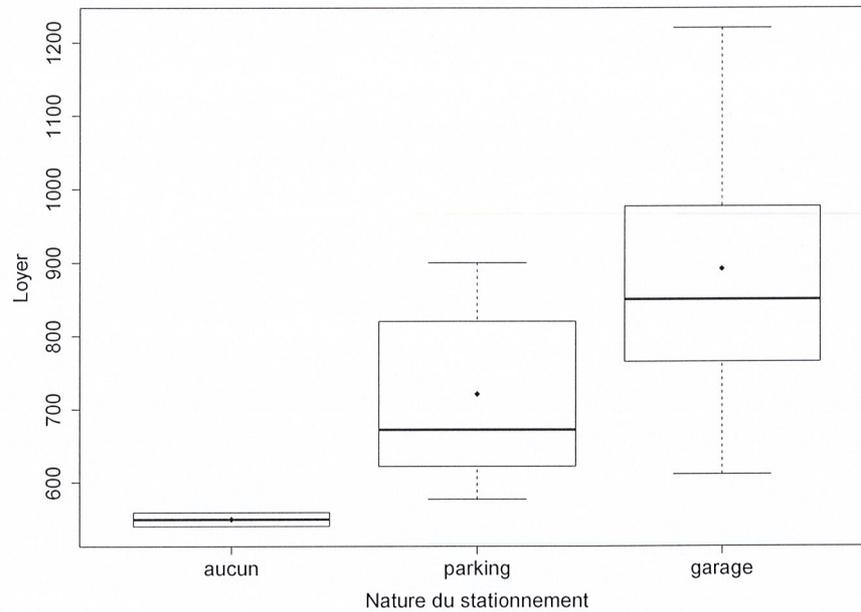
Formule théorique :

Application numérique :

Commentaire :

E Etude de la liaison entre le montant du loyer et la nature du stationnement

1. Commenter le graphique ci-dessous.



2. Quelle relation relie la moyenne marginale de la variable loyer aux moyennes conditionnelles de la variable loyer sachant la variable statio ?

Formule théorique :

Vérification sur les données :

3. Calculer les variances conditionnelles de la variable `loyer` sachant la variable `statio`.

4. En déduire la variance intra-groupes.

Formule théorique :

Application numérique :

5. Rappeler la formule d'analyse de la variance et en déduire la valeur de la variance inter-groupes.

Formule :

Variance inter-groupes :

6. Calculer le rapport de corrélation η^2 et conclure sur l'intensité de la liaison entre le montant du loyer et la nature du stationnement.

Formule :

Application numérique :

Commentaire :