



IUT STID, 1^{ère} année & APPC

Statistique descriptive

Interrogation 2 : à rendre le
Jeudi 18 novembre

Avant propos : Cette interrogation est à effectuer en binôme (les binômes sont les binômes de TP). Les réponses sont à compléter directement sur le fichier **Interro2-etud.odt**, téléchargeable sur mon site web (<http://www.nathalievilla.org>, en cliquant sur Enseignements / IUT STID Carcassonne / Statistique descriptive), et le fichier, renommé « Interro2-nom1-nom2.odt » (où *nom1* et *nom2* sont les noms des deux membres du binômes), est à envoyer à nathalie.villa@univ-perp.fr, **avant le jeudi 18 novembre à 20h**. Un accusé de réception de la bonne réception de votre fichier doit vous parvenir.

Noms : _____

Pour effectuer cette interrogation, vous utiliserez le fichier de données R **TP1.RData** et correspondant à l'étude tirée de l'article :

Desbois, D. (2008) Introduction to scoring methods: financial problems of farm holdings. *Case Studies in Business, Industry and Government Statistics*, 2(1), 56-76.

Ce fichier est téléchargeable sur mon site web, <http://www.nathalievilla.org>, en cliquant sur Enseignements / IUT STID Carcassonne / Statistique descriptive, sous le nom **TP1.RData**. En particulier, l'objet de cette interrogation est l'étude de deux variables particulières de ce jeu de données, « OWNLAND » (qualitative) et « r1 » (quantitative). L'étude comprend une étude globale puis une comparaison selon les modalités de la variable d'intérêt principal du jeu de données qui est la variable « DIFF » indiquant un incident (ou non) de paiement dans le remboursement d'un crédit.

1 Étude de la variable « OWNLAND ».....	2
2 Étude de la variable « r1 ».....	3
3 Comparaison entre entreprises ayant eu un problème de remboursement et entreprises n'ayant pas eu de problème.....	5
3.1 Création de deux jeux de données.....	5
3.2 Comparaison de la variable « OWNLAND ».....	5
3.3 Comparaison de la variable « r1 ».....	6

1 Étude de la variable « OWNLAND »

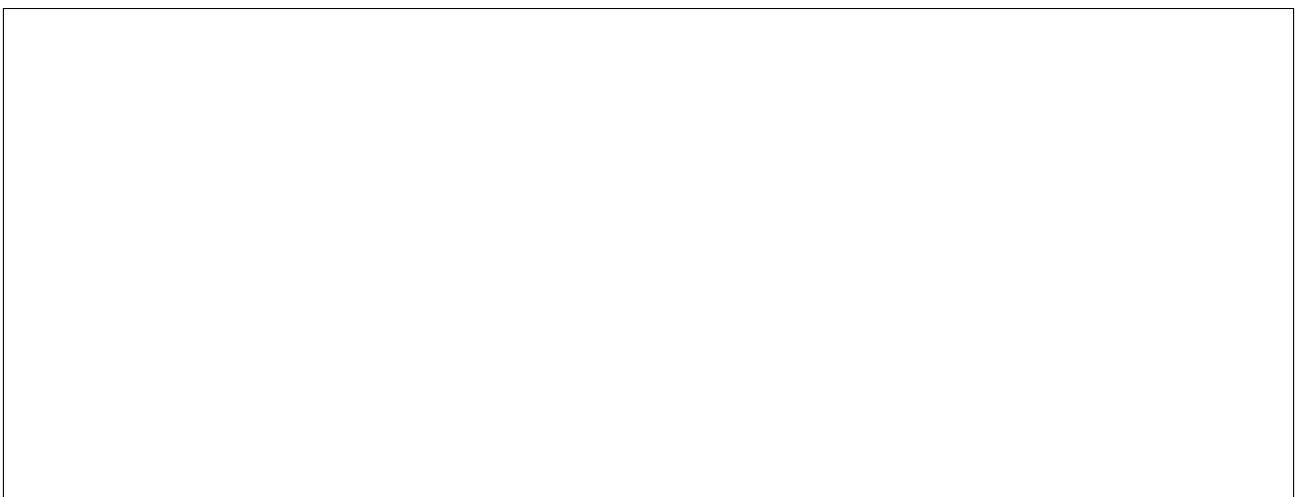
Dans cette partie, on se propose d'étudier la variable « OWNLAND » : donnez le tableau d'effectifs de cette variable ainsi qu'un graphique permettant de représenter sa distribution de manière adapté. Commentez les résultats obtenus.

On insèrera dans le cadre Code 1 le code généré pour effectuer ces deux éléments.

Tableau d'effectifs

Représentation graphique de la distribution de « OWNLAND »

Commentaires



Code 1: Code généré pour effectuer le tableau d'effectifs et le graphique

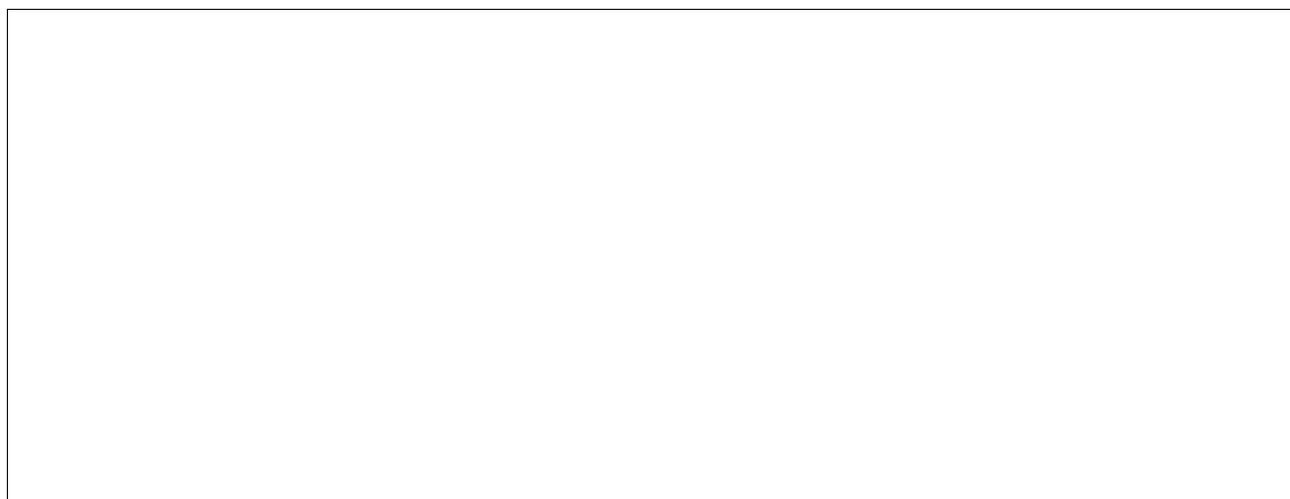
2 Étude de la variable « r1 »

Dans cette seconde partie, on étudie la variable « r1 » qui est le ratio entre la dette totale et les avoirs de l'entreprise.

1. Déterminez la moyenne, la médiane, le minimum, le maximum et l'écart type de « r1 » : vous donnerez les valeurs de ces trois statistiques puis vous commenterez la différence entre moyenne et médiane. De plus, vous préciserez dans le cadre Code 2 le code généré pour déterminer ces trois valeurs.

Statistiques

Commentaires



Code 2 : Code généré pour déterminer les statistiques principales de la variable « r1 »

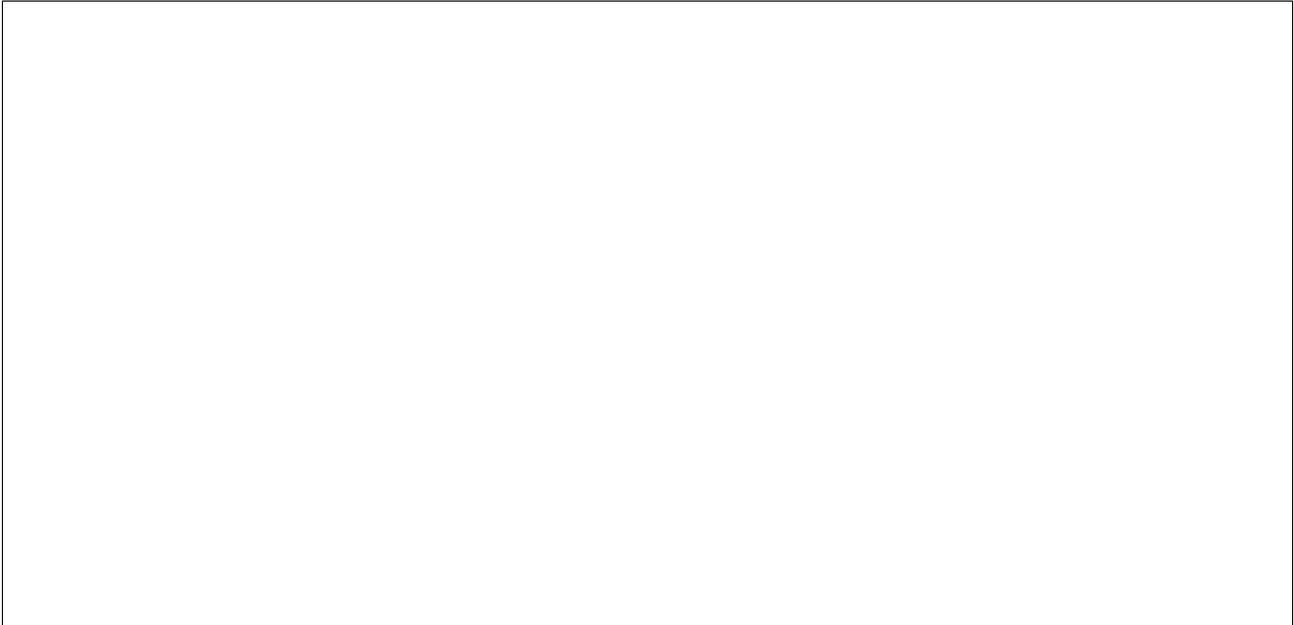
2. Que signifie concrètement, pour une entreprise, d'avoir une valeur de « r1 » supérieure à 1 ? Que peut-on attendre quant à l'influence de cette variable sur les incidents de paiement (variable DIFF) ?
3. Effectuez un regroupement des valeurs de « r1 » en 10 classes pertinentes. Donnez le tableau d'effectifs de ces dix classes puis l'histogramme associé que vous commenterez (notamment en relation avec ce que vous avez dit à la question 1 ci-dessus).

Enfin, vous préciserez dans le cadre Code 3 le code généré pour effectuer le regroupement en classes, déterminer le tableau d'effectifs et construire l'histogramme.

Tableau d'effectifs du regroupement en classes

Histogramme

Commentaires



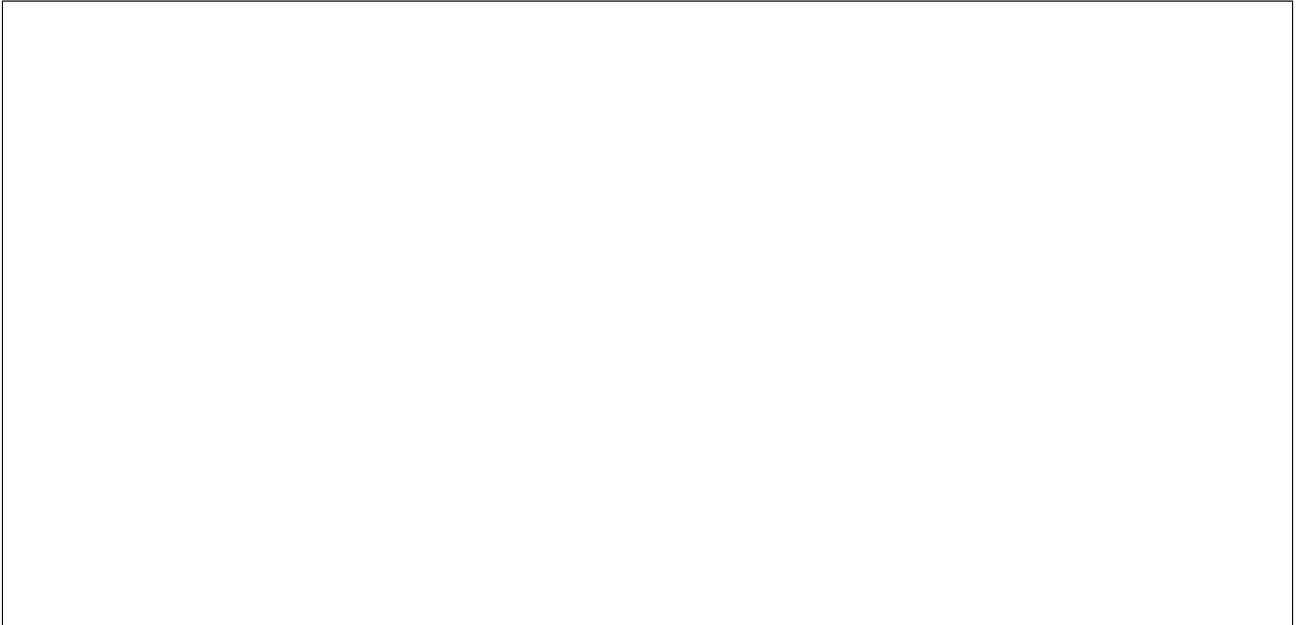
Code 3: Code généré pour effectuer le regroupement en classes, le tableau d'effectifs et l'histogramme

3 Comparaison entre entreprises ayant eu un problème de remboursement et entreprises n'ayant pas eu de problème

Le but de cette partie est de comparer les valeurs des variables « OWNLAND » et « r1 » entre les entreprises ayant eu un incident de paiement et celle n'en ayant pas eu.

3.1 Création de deux jeux de données

À partir du jeu de données initial, créez deux jeux de données : un nommé « donnees.sain » contenant les valeurs de toutes les variables pour les entreprises n'ayant pas eu d'incident de paiement (variable « DIFF » égale à « sain ») et un autre, nommé « donnees.incident » contenant les valeurs de toutes les variables pour les entreprises ayant eu un incident de paiement (variable « DIFF » égale à « incident »). Précisez dans le cadre Code 4 le code généré pour créer ces deux jeux de données.



Code 4: Code généré pour créer deux jeux de données à partir des données initiales

3.2 Comparaison de la variable « OWNLAND »

Effectuez le diagramme circulaire de la distribution de la variable « OWNLAND » pour les entreprises ayant eu un incident de paiement et pour les entreprises n'en ayant pas eu. Commentez les différences : les résultats étaient-ils prévisibles ? Si oui, pourquoi ?

Diagrammes circulaires

Commentaires

3.3 Comparaison de la variable « r1 »

1. Déterminez la moyenne et l'écart type de la variable « r1 » pour les entreprises ayant eu un incident de paiement et pour celles n'en ayant pas eu. Commentez les différences ? Celles-ci étaient-elles prévisibles et si oui, pourquoi ?

Statistiques

Commentaires

2. Déterminez la variable centrée réduite issue de « r1 » pour les entreprises ayant eu un incident de paiement et pour celles n'en ayant pas eu. Vous copierez le code permettant d'obtenir ces deux variables dans le cadre Code 5. Entre l'entreprise saine numéro 1 et l'entreprise ayant eu un incident de paiement numéro 78, laquelle a la situation financière la plus favorable de son groupe ? Justifier.

Code 5: Code généré pour créer les deux variables centrées réduites

Comparaison entreprise 1 et entreprise 78