**Université Panthéon-Assas Paris 2**

**Sujet d’examen du cours de « Etudes et recherche en gestion (analyse des données) »**

**M24078AC U.E.C. 1**

**Année universitaire 2022-2023**

**Master 1 Economie et Gestion**

**Professeurs Patrick Hetzel et Simon Porcher**

**Durée de l’examen : 3h (parties 1 et 2).**

**Aucun document n’est autorisé. Calculatrice non autorisée.**

**Merci de répondre aux questions des parties 1 et 2 sur deux copies différentes.**

*Ce sujet comporte 7 pages. Avant de composer, assurez-vous que votre sujet est complet.*

*\*\*\**

**Partie 1**

Veuillez répondre à la question suivante :

**Question (5 points) :** Quelles sont les grandes différences entre les approches qualitatives et les approches quantitatives en gestion ?

**Partie 2**

**Questions à choix multiples (1 point par question)**

Choisissez la bonne réponse.

1. La corrélation entre deux variables est faible si le coefficient de corrélation de

Pearson est :

a. supérieur a 0,6.

b. entre –0,3 et 0.

c. entre 0 et 0,3.

d. entre –0,3 et 0,3.

2. Le *F* d’une anova est le rapport entre :

a. la variance intergroupe et la variance intragroupe.

b. la variance intragroupe et la variance intergroupe.

c. la variance totale et l’écart-type.

d. la variance totale et la variance intergroupe.

3. Le modèle de régression linéaire multiple permet de :

a. calculer la corrélation de différentes variables indépendantes.

b. faire le lien entre un groupe unique de variables indépendantes et différentes variables dépendantes.

c. réaliser une série de régressions linéaires entre deux variables, une variable dépendante et différents groupes de variables indépendantes.

d. mesurer l’impact de différentes variables indépendantes sur la variable dépendante.

4. Une analyse typologique permet :

a. d’établir l’existence d’une relation significative entre deux variables.

b. de segmenter une population d’objets.

c. d’analyser la corrélation entre deux facteurs.

5. Un modèle de régression linéaire permet de prédire efficacement les résultats d’une variable dépendante binaire.

a. Oui.

b. Non.

c. Cela dépend du type de variables indépendantes.

d. Cela dépend du niveau de corrélation entre les variables indépendantes.

**Etude de cas (10 points)**

La base *bank* contient des informations sur 43 193 clients d’une banque portugaise. L’unité d’observation est le client. Cette base de données est mise à disposition d’une équipe de télémarketeurs chargée d’appeler les clients de la banque afin de leur proposer de souscrire à un nouveau produit d’épargne. Les variables présentes dans la base sont:

* l’âge du client (*age*) ;
* l’argent sur son compte (*balance*) en euros ;
* une série de variables dichotomiques égales à 1 si ‘oui’ et 0 si ‘non’ :
	+ le fait qu’il ait souscrit au produit d’épargne (*souscrit*);
	+ le fait qu’il ait un prêt à la consommation (*prêt*), immobilier (*immo*) ou fait défaut dans le passé (*défaut*) ;
	+ le fait qu’il soit marié (*married*) ou divorcé (*divorcé*). S’il n’est ni marié, ni divorcé, il est célibataire non divorcé. Notez bien qu’il est soit marié, soit divorcé, soit célibataire (il ne peut pas être divorcé et célibataire par exemple) ;
	+ le niveau d’éducation qui peut être soit primaire (*primaire*), soit secondaire (*secondaire*), soit supérieur (*tertiaire*) ;
	+ la catégorie socio-professionnelle : cadre supérieur (*management*), entrepreneur (*entrepreneur*), employé (*admin*), technicien (*technician*), ouvrier (*blue-collar*), services à la personne (*services*), travailleur indépendant (*self-employed*), servant (*housemaid*), retraité (*retired*), étudiant (*student*) ou chômeur (*unemployed*).

Nous allons tenter de comprendre ce qui peut amener un client de la banque à souscrire au produit d’épargne. *souscrit* est la variable dépendante et les autres variables sont les variables indépendantes.

**Partie I**

La table 1 reporte les statistiques descriptives pour la base de données.

|  |
| --- |
| **Table 1 : Statistiques descriptives** |
|  | N | Minimum | Maximum | Moyenne | Ecart type |
| age | 43193 | 18 | 95 | 40,76 | 10,513 |
| balance | 43193 | -8019 | 102127 | 1354,03 | 3042,104 |
| souscrit | 43193 | 0 | 1 | ,12 | ,321 |
| prêt | 43193 | 0 | 1 | ,16 | ,371 |
| immo | 43193 | 0 | 1 | ,56 | ,496 |
| défaut | 43193 | 0 | 1 | ,02 | ,133 |
| married | 43193 | 0 | 1 | ,60 | ,490 |
| divorcé | 43193 | 0 | 1 | ,12 | ,321 |
| tertiaire | 43193 | 0 | 1 | ,31 | ,461 |
| secondaire | 43193 | 0 | 1 | ,54 | ,499 |
| primaire | 43193 | 0 | 1 | ,16 | ,364 |
| management | 43193 | 0 | 1 | ,21 | ,410 |
| entrepreneur | 43193 | 0 | 1 | ,03 | ,178 |
| admin | 43193 | 0 | 1 | ,12 | ,320 |
| retired | 43193 | 0 | 1 | ,05 | ,217 |
| student | 43193 | 0 | 1 | ,02 | ,133 |
| technician | 43193 | 0 | 1 | ,17 | ,376 |
| blue-collar | 43193 | 0 | 1 | ,21 | ,411 |
| services | 43193 | 0 | 1 | ,09 | ,290 |
| self-employed | 43193 | 0 | 1 | ,04 | ,185 |
| housemaid | 43193 | 0 | 1 | ,03 | ,164 |
| unemployed | 43193 | 0 | 1 | ,03 | ,169 |
| N valide (liste) | 43193 |  |  |  |  |

La table 2 reporte les corrélations de Pearson entre les différentes variables indépendantes de la base (hors catégories socio-professionnelles).

1. Pourquoi réalise-t-on la table des corrélations entre les variables indépendantes du modèle ? (1 point)
2. Identifiez les corrélations moyennes dans la table 2 (1 point).

|  |
| --- |
| **Table 2 : Corrélations** |
|  | age | balance | prêt | immo | défaut | married | divorcé | tertiaire | secondaire | primaire |
| agebalanceprêtimmodéfautmarrieddivorcétertiairesecondaireprimaire | 1 |  |  |  |  |  |  |  |  |  |
| ,098\*\* | 1 |  |  |  |  |  |  |  |  |
| -,010\* | -,084\*\* | 1 |  |  |  |  |  |  |  |
| -,185\*\* | -,067\*\* | ,037\*\* | 1 |  |  |  |  |  |  |
| -,016\*\* | -,066\*\* | ,076\*\* | -,009 | 1 |  |  |  |  |  |
| ,280\*\* | ,026\*\* | ,038\*\* | ,018\*\* | -,014\*\* | 1 |  |  |  |  |
| ,169\*\* | -,022\*\* | ,016\*\* | -,001 | ,017\*\* | -,445\*\* | 1 |  |  |  |
| -,076\*\* | ,088\*\* | -,054\*\* | -,110\*\* | -,016\*\* | -,098\*\* | -,012\* | 1 |  |  |
| -,083\*\* | -,070\*\* | ,062\*\* | ,096\*\* | ,013\*\* | -,016\*\* | ,017\*\* | -,715\*\* | 1 |  |
| ,209\*\* | -,015\*\* | -,017\*\* | ,008 | ,002 | ,146\*\* | -,009 | -,288\*\* | -,464\*\* | 1 |
|  |

**Partie II**

Nous allons maintenant utiliser un modèle de régression linéaire avec la variable *souscrit* comme variable dépendante. Voici les résultats du modèle :

|  |
| --- |
| **Table 3 : Coefficients de la régression linéairea** |
| Modèle | Coefficients non standardisés | Coefficients standardisés | t | Sig. |
| B | Erreur standard | Bêta |
| 1 | (Constante) | ,133 | ,009 |  | 14,786 | ,000 |
| age | ,000 | ,000 | ,011 | 1,837 | ,066 |
| balance | 0,000003 | ,000 | ,031 | 6,423 | ,000 |
| prêt | -,044 | ,004 | -,051 | -10,763 | ,000 |
| immo | -,072 | ,003 | -,111 | -22,395 | ,000 |
| défaut | -,037 | ,011 | -,016 | -3,280 | ,001 |
| divorcé | -,024 | ,006 | -,024 | -4,284 | ,000 |
| tertiaire | ,054 | ,006 | ,078 | 8,989 | ,000 |
| secondaire | ,019 | ,005 | ,029 | 3,940 | ,000 |
| management | ,001 | ,006 | ,002 | ,201 | ,841 |
| entrepreneur | -,020 | ,009 | -,011 | -2,161 | ,031 |
| admin | ,027 | ,006 | ,027 | 4,619 | ,000 |
| retired | ,104 | ,009 | ,070 | 12,165 | ,000 |
| student | ,142 | ,012 | ,059 | 11,522 | ,000 |
| technician | ,002 | ,005 | ,003 | ,455 | ,649 |
| services | ,001 | ,006 | ,001 | ,232 | ,816 |
| self-employed | -,003 | ,009 | -,002 | -,293 | ,769 |
| housemaid | -,019 | ,010 | -,010 | -1,913 | ,056 |
| unemployed | ,039 | ,010 | ,021 | 4,078 | ,000 |
| married | -,037 | ,004 | -,057 | -9,646 | ,000 |
| a. Variable dépendante : souscrit |
| **Table 4 : Récapitulatif des modèles** |
| Modèle | R | R-deux | R-deux ajusté | Erreur standard de l'estimation |
| 1 | ,200a | ,040 | ,039 | ,314 |
|  |

1. Pourquoi ne pas avoir intégré *primaire* et *blue-collar* dans la régression Table 3 ? Comment peut-on interpréter le coefficient de *management* ? (2 points)
2. Pourquoi le R² est-il aussi bas à votre avis (Table 4) ? (1 point)

Nous allons maintenant utiliser un modèle logit avec *souscrit* en variable dépendante. Voici les résultats (Table 5):

|  |
| --- |
| **Table 5 : Résultat du Logit** |
|  | B | E.S | Wald | ddl | Sig. | Exp(B) |
| Pas 1a | age | ,004 | ,002 | 5,183 | 1 | ,023 | 1,004 |
| balance | ,000 | ,000 | 29,882 | 1 | ,000 | 1,000 |
| prêt | -,566 | ,051 | 122,158 | 1 | ,000 | ,568 |
| immo | -,725 | ,033 | 494,456 | 1 | ,000 | ,485 |
| défaut | -,508 | ,152 | 11,241 | 1 | ,001 | ,602 |
| divorcé | -,231 | ,056 | 16,845 | 1 | ,000 | ,794 |
| married | -,383 | ,039 | 98,013 | 1 | ,000 | ,682 |
| tertiaire | ,584 | ,054 | 118,399 | 1 | ,000 | 1,793 |
| secondaire | ,244 | ,053 | 21,290 | 1 | ,000 | 1,277 |
| entrepreneur | -,234 | ,101 | 5,338 | 1 | ,021 | ,791 |
| admin | ,286 | ,053 | 28,663 | 1 | ,000 | 1,331 |
| retired | ,786 | ,070 | 127,079 | 1 | ,000 | 2,194 |
| student | ,872 | ,089 | 95,007 | 1 | ,000 | 2,392 |
| technician | ,032 | ,047 | ,482 | 1 | ,488 | 1,033 |
| services | ,007 | ,065 | ,011 | 1 | ,917 | 1,007 |
| self-employed | -,022 | ,084 | ,068 | 1 | ,795 | ,978 |
| housemaid | -,184 | ,108 | 2,918 | 1 | ,088 | ,832 |
| unemployed | ,351 | ,083 | 17,782 | 1 | ,000 | 1,421 |
| Constante | -2,009 | ,092 | 474,502 | 1 | ,000 | ,134 |
| 1. A partir de la Table 5, identifiez les variables ayant un coefficient significatif et classez-les selon que l’impact soit positif ou négatif (1 point).
2. Que peut-on en déduire quant au profil-type du client qu’il faut appeler ? (1 point)
3. Comment pourrait-on calculer la probabilité qu’a un client moyen de souscrire au produit d’épargne ? Il n’est pas nécessaire de calculer la probabilité mais il faut poser l’équation qui permet de la calculer. Un client moyen est celui qui prend une valeur moyenne pour chaque variable, vous pouvez donc vous référer aux statistiques descriptives plus hautes (1 point).
 |

**Partie III**

Nous allons enfin réaliser un test du Khi-deux entre le fait de souscrire (*souscrit*) et les catégories socio-professionnelles. Voici les résultats :

1. Commentez les résultats de la Table 6 (1 point).
2. Qu’est-ce qu’un effectif théorique (Table 7) ? (1 point)

|  |
| --- |
| **Table 6: Tests du khi-deux** |
|  | Valeur | ddl | Signification asymptotique (bilatérale) |
| khi-deux de Pearson | 772,489a | 10 | ,000 |
| Rapport de vraisemblance | 690,364 | 10 | ,000 |
| N d'observations valides | 43193 |  |  |
| a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 90,09. |

|  |
| --- |
| **Table 7 : Tableau croisé** |
|  | souscrit | Total |
| 0 | 1 |
| job | admin. | Effectif | 4387 | 613 | 5000 |
| Effectif théorique | 4418,8 | 581,2 | 5000,0 |
| blue-collar | Effectif | 8603 | 675 | 9278 |
| Effectif théorique | 8199,5 | 1078,5 | 9278,0 |
| entrepreneur | Effectif | 1295 | 116 | 1411 |
| Effectif théorique | 1247,0 | 164,0 | 1411,0 |
| housemaid | Effectif | 1090 | 105 | 1195 |
| Effectif théorique | 1056,1 | 138,9 | 1195,0 |
| management | Effectif | 7963 | 1253 | 9216 |
| Effectif théorique | 8144,7 | 1071,3 | 9216,0 |
| retired | Effectif | 1659 | 486 | 2145 |
| Effectif théorique | 1895,7 | 249,3 | 2145,0 |
| self-employed | Effectif | 1358 | 182 | 1540 |
| Effectif théorique | 1361,0 | 179,0 | 1540,0 |
| services | Effectif | 3654 | 350 | 4004 |
| Effectif théorique | 3538,6 | 465,4 | 4004,0 |
| student | Effectif | 549 | 226 | 775 |
| Effectif théorique | 684,9 | 90,1 | 775,0 |
| technician | Effectif | 6538 | 817 | 7355 |
| Effectif théorique | 6500,0 | 855,0 | 7355,0 |
| unemployed | Effectif | 1076 | 198 | 1274 |
| Effectif théorique | 1125,9 | 148,1 | 1274,0 |
| Total | Effectif | 38172 | 5021 | 43193 |
| Effectif théorique | 38172,0 | 5021,0 | 43193,0 |