

Fiche 3 - TISD - Master Pro

Variables qualitatives

Tran Viet Chi, chi.tran@math.univ-lille1.fr, bureau 316 (bâtiment M3).

1 Etude d'une variable qualitative avec SAS

Exercice 1 (Travail des femmes)

Les données étudiées dans cet exercice sont téléchargeables à partir de

<http://math.univ-lille1.fr/~tran/enseignements.html>

Les données de la table `travf.sas7bdat` sont issues de l'"Enquête Emploi en Continu" 2005 de l'INSEE. Pour constituer le fichier, nous nous sommes restreints aux femmes de 20 à 65 ans, vivant dans des ménages "classiques" (personnes seules, familles monoparentales, couples avec ou sans enfants) et dont la place est celle de chef de ménage ou conjoint du chef de ménage. Nous disposons de 99 959 individus et des variables suivantes (voir le descriptif) :

- identifiant (IDENT),
- activité (ACT6),
- âge (AGE)
- niveau d'étude (CITE97),
- type de ménage dans laquelle vit la personne (TYPMEN5), indicatrice de vie en couple (COHAB)
- de la CSP de la personne (CSTOT), du chef de ménage ou du conjoint (CSTOTCJ), et de la personne de référence dans le ménage (CSTOTPR),
- du nombre d'enfants (NBENF18)
- de la tranche de salaire (SALREDTR),
- du temps de travail (TPPRED) et du nombre d'heures travaillées pendant une semaine de référence (EMPNBH).

1. Tracer un diagramme en tuyaux d'orgue pour la variable ACT6 avec la PROC GCHART et l'option `hbar` ou `vbar`.
2. Toujours en utilisant la PROC GCHART et avec l'option `pie`, dessiner un diagramme à secteurs pour la variable CITE97.
3. A l'aide de la PROC FREQ, obtenir la liste des modalités et les fréquences pour les variables ACT6, CITE97.
4. Croiser les variables ACT6 et COHAB et donner la valeur des indicateurs de liaison suivants : χ^2 , Φ^2 et coefficient de Cramer. Nous verrons à la fin du semestre une façon de *tester* la liaison de ces variables.

2 Etude d'une variable qualitative avec R

Exercice 2 (Salaires)

Les données de la table `cps85.txt` proviennent d'une enquête auprès des ménages effectuée par le U.S. Census Bureau en mai 1985. Le fichier contient 534 individus. Nous nous intéresserons aux variables suivantes :

- NONWH qui vaut 1 si l'individu interrogé n'est ni Blanc ni Hispanique,
- HISP qui vaut 1 si l'individu interrogé est Hispanique,
- FE qui vaut 1 si l'individu interrogé est une femme,
- UNION qui vaut 1 si l'individu est syndiqué,

Pour chacune de ces variables, tracer le diagramme en secteur et établir pour chacune de leurs modalités l'effectif et la fréquence. Croiser les variables NONWH et UNION avec la commande `table`.

Exercice 3 (Passagers du Titanic)

1. Les données sur les passagers du Titanic, fournies par **R**, peuvent-être chargées par la commande `data(Titanic)`. Demander la description de cette table par

```
help("Titanic")
```

2. Explorer la table (en particulier sa structure, demander `dim(Titanic)`) et décrire les différentes variables. Calculer les fréquences marginales et tracer leurs diagrammes en secteur.
3. Obtenir la table de contingence du croisement de la variable `Survived` avec la variable `Class`. Calculer les indicateurs de liaison χ^2 , Φ^2 et coefficient de Cramer avec les commandes `chisq.test`.
4. Faire de même avec le croisement de `Survived` et `Sex` et `Survived` et `Age`. Avec quelle variable `Survived` est-elle le plus liée?
5. Recommencer pour le croisement de `Survived` avec `Class` par sexe. La liaison de ces deux variables est-elle plus forte pour les hommes ou pour les femmes?