

EISTI ING1 : Examen de rattrapage de Statistiques Descriptives

15 Février 2011 : documents papiers autorisés, durée 1heure

1. Nature de caractères

Question 1) On est souvent tenté de transformer un caractère qualitatif ordinal en un caractère quantitatif discret en attribuant une note (qui respecte l'ordre) à chaque modalité. Expliquer pourquoi cela est abusif sauf si une hypothèse (à préciser) est vérifiée à priori.

Question 2) A propos de nature de caractère, on classe l'ensemble des types possibles sur une échelle (pauvreté ... richesse).

(a) Expliquer le sens de cette échelle.

(b) Quelle est la conséquence principale de cette échelle quant à l'utilisation des telles ou telles méthodes statistiques ?

2. Questions de réflexion

Question 1) Régulièrement, on publie dans les journaux télévisés, le salaire médian et le salaire moyen des français. On constate souvent que le 1^{er} est nettement inférieur au second. Que faut-il en tirer comme conclusion ?

Question 2) Dans une régression linéaire (simple ou multiple) de Y en fonction X_1, \dots, X_n , on calcule R^2 comme le rapport entre la variance de Y expliquée par la droite de régression et la variance totale de Y. Donner l'équivalent R^2 dans le cas d'un croisement d'un caractère quantitatif Y et d'un caractère qualitatif, On explicitera sa réponse.

Question 3) Dans le calcul de quantiles d'un caractère quantitatif représenté par intervalles, on est amené à faire des interpolations linéaires. Quelle est l'hypothèse sous-jacente qui justifie cette interpolation ?

3. Analyse bivariée : une petite régression

On considère les 2 caractères quantitatifs x et y suivants :

X	1	2	7	4	6
Y	5	4	1	3	2

On donne $\sum x_i = 20$, $\sum y_i = 15$, $\sum x_i^2 = 106$, $\sum y_i^2 = 55$ et $\sum x_i y_i = 44$

Question 1) Calculer les moyennes et les variances de chaque caractère ainsi que la covariance.

Question 2) Déterminer l'équation de la droite de régression de y en fonction de x.

Question 3) Pouvez-vous prévoir la valeur de y lorsque x = 8.

Question 4) Calculer le coefficient de régression linéaire et évaluer le degré de corrélation linéaire entre les deux caractères.

4. Analyse bivariée : étude générale

Dans une entreprise de 130 salariés, on souhaite étudier le lien entre le salaire annuel (en milliers d'euros), la catégorie (cadre supérieur, cadre moyen et ouvrier) et le sexe (homme, femme).

EISTI ING1 : Examen de rattrapage de Statistiques Descriptives
15 Février 2011 : documents papiers autorisés, durée 1heure

Sexe	femme	homme	Tous sexes
Catégorie			
Cadre sup.	44,00	40,94	41,40
Cadre moy.	116,50	116,38	116,40
Ouvrier	20,65	22,29	21,73
Toutes catégories	27,36	34,04	32,04

Tab 1 : Salaire moyen annuel par sexe et par catégorie

Sexe	femme	homme	Total général
Catégorie			
Cadre sup.	3	17	20
Cadre moy.	2	8	10
Ouvrier	34	66	100
Total général	39	91	130

Tab 2 : Tableau de contingence des effectifs

Sexe	femme	homme	Freq. marginales
Catégorie			
Cadre sup.	2,31	13,08	15,38
Cadre moy.	1,54	6,15	7,69
Ouvrier	26,15	50,77	76,92
Freq. Marginales	30	70	100

Tab 3 : Tableau de contingence des fréquences (%)

1) Etude du lien entre le salaire et les deux autres caractères

- Expliquer comment est calculée la cellule grisée du tableau 1.
- A la lecture du tableau 1, quelles conclusions pouvez-vous faire sur la corrélation entre, le salaire et le sexe d'une part, et entre le salaire et la catégorie d'autre part.
- Que faudrait-il calculer pour confirmer ces conclusions ?

2) Etude du lien entre la catégorie et le sexe

Sexe	femme	homme
Catégorie		
Cadre sup.	15	85
Cadre moy.	20	80
Ouvrier	34	66

Tab 4 : Profils lignes (%)

Sexe	femme	homme
Catégorie		
Cadre sup.	7,69	18,68
Cadre moy.	5,13	8,79
Ouvrier	87,18	72,53

Tab 5 : Profils colonnes (%)

- Dans les tableaux 2, 3 et 4, expliquer ce que signifient les cases grisées.
- Avec quels chiffres peut-on comparer les tableaux 4 et 5 ?
- Que pouvez-vous en déduire sur la corrélation entre ces deux caractères ?
- Peut-on utiliser le coefficient de contingence pour tester l'indépendance de ces caractères ?