Traitement et Analyse des Données

Exercice 1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| xi | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| ni | **7500** | **3700** | **2300** | **2200** | **1100** | **530** | **200** | **120** | **30** |
| fi | **0.42** | **0.21** | **0.13** | **0.12** | **0.06** | **0.03** | **0.01** | **0.006** | **0.002** |
| Ni | **7500** | **11200** | **13500** | **15700** | **16800** | **17330** | **17530** | **17650** | **17680** |
| Fi | **0.42** | **0.633** | **0.764** | **0.888** | **0.95** | **0.98** | **0.992** | **0.998** | **1.00** |

* **Total ni : 17680**
* **Ni = ∑ ni (Effectif cumulé)**
* **fi = ni / 17680 (Fréquence)**
* **Fi = Prendre Ni et divisé par Tot(ni) = 17680 (Fréquence cumulé)**

Exercice 2

**Énoncé : On cherche à étudier la relation entre le nombre d’heure de travail dans un cours et la note finale (sur 20) a l’examen de ce cours.**

**Données : Les données obtenues sur les 20 étudiants de la classe sont les suivantes :**

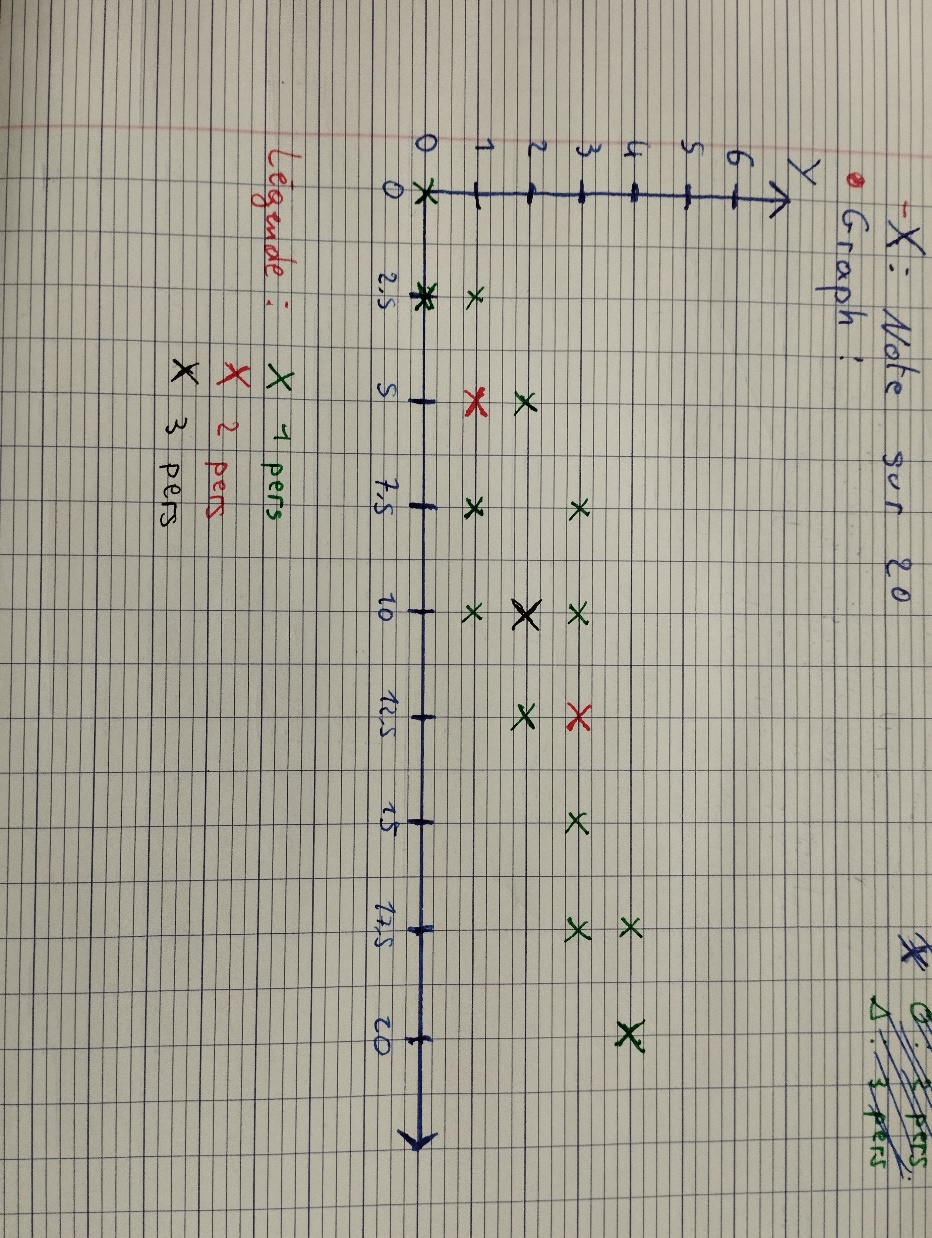
* **2 étudiants n’ont pas révisé et ont obtenue l’un 0/20 et l’autre 2.5/20**
* **Pour les 5 étudiants qui ont étudié pendant 1 heure, il y en a 1 qui a obtenu 2.5/20, 2 qui ont obtenue 5/20, 1 qui a obtenu 7.5/20 et 1 qui a eu 10/20**
* **Pour les 5 étudiants qui ont étudié pendant 2 heures, il y en a 1 qui a obtenu 5/20, 3 qui ont obtenu 10/20, et 1 qui a obtenu 12.5/20**
* **Pour les 6 étudiants qui ont étudié pendant 3 heures, 1 a obtenu 7.5/20, 1 a obtenu 10/20, 2 ont obtenu 12.5/20, 1 a obtenu 15/20 et 1 a obtenu 17.5/20**
* **2 étudiants ont révisé pendant 4 heures et ont obtenu pour l’un 17.5/20 et l’autre 20/20**

**Consigne :**

* **1] Réaliser une représentation graphique approprié (voir feuille)**
  + **Choix de X et Y ?**
  + **Qu’elle stratégie adopter pour certains points ?**
* **2] Calculer le coefficient de corrélation PX,Y a partir du tableau de contingence**
* **3] Calculer :** 
  + **Px**
  + **Px \* X**
  + **Py**
  + **Py \* Y**
  + **Px X^2**
  + **Py Y^2**
  + **∑ Pxy X\*Y**
  + **Cov(X,Y) = E(XY) – E(X)E(Y)**

**Réponses :**

**1] Pour le choix des axes, je choisis :**

* + **X : Note sur 20**
  + **Y : Nb heure de travail**

**Certains points sont superposés, j’utilise donc un code couleur, voici le graph :**

**Faut inverser le graph en fait, faut mettre nbHeure en X et note en Y**

**2 et 3] Voici le tableau de contingence (nbHeure en X et note en Y)**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ****X/Y**** | ****0**** | ****2.5**** | ****5**** | ****7.5**** | ****10**** | ****12.5**** | ****15**** | ****17.5**** | ****20**** |
| ****0**** | **1** | **1** | **0** | **0** | **0** | **0** | **0** | **0** | **0** |
| ****1**** | **0** | **1** | **2** | **1** | **1** | **0** | **0** | **0** | **0** |
| ****2**** | **0** | **0** | **1** | **0** | **3** | **1** | **0** | **0** | **0** |
| ****3**** | **0** | **0** | **0** | **1** | **1** | **2** | **1** | **1** | **0** |
| ****4**** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **1** | **1** |

**Voici le même tableau mais avec des fréquences et tous les trucs :**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ****X/Y**** | ****0**** | ****2.5**** | ****5**** | ****7.5**** | ****10**** | ****12.5**** | ****15**** | ****17.5**** | ****20**** | ****Px**** | ****Px\*X**** | ****Px\*X²**** |
| ****0**** | **0.05** | **0.05** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0.1** | **0** | **0** |
| ****1**** | **0** | **0.05** | **0.10** | **0.05** | **0.05** | **0** | **0** | **0** | **0** | **0.25** | **0.25** | **0.25** |
| ****2**** | **0** | **0** | **0.05** | **0** | **0.15** | **0.05** | **0** | **0** | **0** | **0.25** | **0.5** | **1** |
| ****3**** | **0** | **0** | **0** | **0.05** | **0.05** | **0.10** | **0.05** | **0.05** | **0** | **0.25** | **0.9** | **2.7** |
| ****4**** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0.05** | **0.05** | **0.1** | **0.4** | **1.6** |
| ****Py**** | **0.05** | **0.10** | **0.15** | **0.10** | **0.25** | **0.15** | **0.05** | **0.10** | **0.05** |  | **∑=**  **2.05 E(X)** | **∑=**  **5.55 E(X²)** |
| ****Py\*Y**** | **0** | **0.25** | **0.75** | **0.75** | **2.5** | **1.075** | **0.75** | **1.75** | **1** | **∑= 9.625 E(Y)** |  |  |
| ****Py\*Y²**** | **0** | **0.625** | **3.75** | **5.625** | **25** | **23.4375** | **11.25** | **30.625** | **20** | **∑= 120.3125 E(Y²)** |  |  |
| **∑PXY X\*Y** | **0** | **0.125** | **1** | **1.5** | **5** | **5** | **2.25** | **6.125** | **4** | **∑= 25 E(XY)** |  |  |

**Cov(X,Y) = E(XY) – E(X)E(Y)**

**= 25 – (2.05\*9.625)**

**= 5.26875**

**σ²X = E(X²) – E²(X)**

**= 5.55 – (2.05)²**

**= 1.3475**

**σ²Y = E(Y²) – E²(Y)**

**= 120.3125 – (9.625)²**

**= 27.67188**

**PXY = (cov(X,Y) / (σX \* σY))**

**= (5.26875 / (sqrt(1.3475) \* sqrt(27.67188)))**

**= 0.86 = 86%**

Exercice 3

**Énoncé : Pile ou face, chaque élève fais 10 lancé 10 fois, on note le nbr de pile qu’ils ont fait pour chaque série**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ****0**** | ****1**** | ****2**** | ****3**** | ****4**** | ****5**** | ****6**** | ****7**** | ****8**** | ****9**** | ****10**** |
| ****0**** | **3** | **14** | **22** | **39** | **45** | **44** | **17** | **12** | **4** | **0** |

**Moyenne = 4.965**

**Trouvé sigma (écart type σ)**

Traitement et Analyse des Données

Page 36

Données :

* La durée de vie :
  + Xbarre = 1/50 \* SOMME(i a 50 de Xi) = 3040.3 heure
* Mono : muM = 3000 heures et sigmaM = 150 heures
* Test
  + Echantillon : N = 50 moteurs et XbarreM = 3040.3 heures
  + Test au seuil alpha = 5%
  + Statistique du test : sqrt(N)((Xbarre – mu) / sigma)

**Q : On fait un test d’hypothèse :**

**Référence : muM = 3000 h et sigmaM = 150h**

**Echantillon : 50 nouveaux moteur 🡪 XbarreM = 3040.3 h**

**Hypothèse H0 : Le nouveau procédé ne change pas la durée de vie**

**Hypothèse H1 : Le nouveau procédé augmente la durré de vie**

* **Comment décidé ?**
  + **Q : Est-ce que l’écart observé XbarreM – muM = 40.3h est « suffisamment grand » pour rejeter H0 ?**
    - **D’apres le théorèmeSi H0 est vrai, l’écart normalizé**

**Z = (XbarreM - muM)/(sigmaM / sqrt(N)) environ loi normal N(0,1)**

**R : 1.65**

**Q : Même question avec un échantillon de 30 moteurs**

**R : On fait un test d’hypothèse :**

**Référence : muM = 3000 h et sigmaM = 150h**

**Echantillon : 30 nouveaux moteur 🡪 XbarreM = ?**

**Hypothèse H0 : Le nouveau procédé ne change pas la durée de vie**

**Hypothèse H1 : Le nouveau procédé augmente la durré de vie**

* **Comment décidé ?**
  + **Q : Est-ce que l’écart observé XbarreM – muM = ? est « suffisamment grand » pour rejeter H0 ?**
    - **D’apres le théorèmeSi H0 est vrai, l’écart normalizé**

**Z = (XbarreM - muM)/(sigmaM / sqrt(N)) environ loi normal N(0,1)**

**Q : Même question avec un ecart type sigmaM = 200 heures**

**R : On fait un test d’hypothèse :**

**Référence : muM = 3000 h et sigmaM = 200h**

**Echantillon : 50 nouveaux moteur 🡪 XbarreM = ?**

**Hypothèse H0 : Le nouveau procédé ne change pas la durée de vie**

**Hypothèse H1 : Le nouveau procédé augmente la durré de vie**

* **Comment décidé ?**
  + **Q : Est-ce que l’écart observé XbarreM – muM = ? est « suffisamment grand » pour rejeter H0 ?**
    - **D’apres le théorèmeSi H0 est vrai, l’écart normalizé**

**Z = (XbarreM - muM)/(sigmaM / sqrt(N)) environ loi normal N(0,1)**

**Q : Même question avec un seuil alpha = 1%**

**R : On fait un test d’hypothèse :**

**Référence : muM = 3000 h et sigmaM = 200h**

**Echantillon : 50 nouveaux moteur 🡪 XbarreM = 3040.3 h**

**Hypothèse H0 : Le nouveau procédé ne change pas la durée de vie**

**Hypothèse H1 : Le nouveau procédé augmente la durré de vie**

* **Comment décidé ?**
  + **Q : Est-ce que l’écart observé XbarreM – muM = 40.3 est « suffisamment grand » pour rejeter H0 ?**
    - **D’apres le théorèmeSi H0 est vrai, l’écart normalizé**

**Z = (XbarreM - muM)/(sigmaM / sqrt(N)) = 1.90 < 2.33 peut pas rejeter l’hypothèse nul. P-value d’environ 3%**

Traitement et Analyse des Données