Série de tp n°=3

Exercice1 : Un directeur d'agence bancaire réalise une étude relative à la durée de traitement des dossiers, supposée suivre une distribution normale. Sur 30 dossiers les données récoltées sont les suivants :

Durée mn	0-10	10-20	20-30	30-40	40-50	50-60
Effectif	3	6	10	7	3	1

1) Calculer la moyenne des durées de traitement des dossiers de cet échantillon
2) En déduire les estimations ponctuelles de la moyenne μ(sans biais) et de l'écart typeσ de la population des dossiers.
3) Donner une estimation de μ par intervalle de confiance au seuil du risque5%

Exercice 2 : soient 200 élèves de lycée interrogés sur le type d'études supérieures qu'ils désirent entreprendre. Les résultats figurent ci-dessous :

Sexe / Type étude	garçon	fille
Littérature	60	60
Scientifique	42	18
technique	18	2

a)étudier s'il y a une relation entre les études et le sexe avec un risqueα=0.001 ?
b) Effectuer cette étude sous R

Exercice3 : Un agriculteur suppose que ces récoltes sont chaque année aussi bonnes (les mêmes) (H0).soit le relevé des récoltes, suivant :

Année	1iere	2ieme	3ieme	4ieme	5ieme
Récolte en tonnes	20	30	18	16	10

Réaliser un test de chi deux qui vérifie cette hypothèse (manuellement puis à l'aide de commandes R), en prenant α=0.05

Exercice4 : On veut déterminer de quelle façon la force de tension d'un certain alliage dépend du pourcentage du zinc qu'il contient .on a les données suivantes :

%de Zinc	4.5	4.6	4.7	4.8	4.9	5.0	5.1
Force de tension	1.0	1.1	1.2	1.4	1.5	1.5	1.7

1) dessiner le graphe représentant ces données.
2) calculer le modèle de régression linéaire correspondant, en prenant x : le pourcentage du zinc et y : la force de tension
3) prédire la valeur de Y avec X=6 .
3) effectuer sous R, les questions 2, tout en dessinant la droite correspondante sur le graphe précédent.

Solution

Exercice1 :

1-moyenne : $\mu=\overline{x}=\Sigma(ni*xci)/n=26.3$

$Sn=\sqrt{ni/n * \Sigma(xci-\mu)2}=\sqrt{151.55}=12.31$

2- $\mu=\overline{x}=\hat{\mu}$ (est estimateur efficace=>sans biais)

Sn=12.31 (un estimateur asymptotiquement sans biais)

$\hat{\sigma}=\sqrt{\frac{n}{n-1}}$ Sn =1.01709*12.31=12.5 ……………………………. Estimateur sans biais.

3- $\bar{x} \sim N(\mu, \sigma/\sqrt{n})$ nous estimons ICμ,

Avec σ inconnue=>$U_{1-\alpha/2}$ est le fractile d'ordre 1-α /2 =0.975 OU α /2(c'est le même) de la loi de Student t_α (n-1)=$t_{0.025}$ (29)= t0.05(29)=2.04525 (n-1 est degré de liberté)

Avec $\hat{\sigma}$=>ICμ= [26.3-(2.0452*($\hat{\sigma}/\sqrt{n}$)), 26.3+(2.0452*($\hat{\sigma}/\sqrt{n}$))]

ICμ= [26.3-(2.0452*(12.5/ $\sqrt{30}$)), 26.3+ (2.0452*(12.5/ $\sqrt{30}$))]=[21.632,30.967]

Remarque :

n=30≥30=> on peut utiliser le fractile de la loi normale U1- α /2=1.96=>ICμ=[21.826, 30.773]

Exercice 2 :

a)Si on considère que :

X : le caractère (variable) qualitatif « sexe », ayant comme modalité :'garçon' et 'fille'

Y : le caractère (variable) qualitatif « type d'étude » comme modalité :'littéraire', 'scientifique' et 'technique'.

1-définir les hypothèses

H0 : les caractères X et Y sont indépendants

H1 : les caractères X et Y ne sont pas sont indépendants

α=0.001

On va utiliser le test d'indépendance de chi2 :

2-tester condition d'application :

Les effectifs théorique ≥5 : pour une matrice de (3,2)

nij=Total*pij avec pij=$\sum_{k=1}^{3} nik$ /Total * $\sum_{l=1}^{2} nlj$ /Total, sachant que Total =200 est l'effectif total. P11=120/200*120 /200, est la probabilité que les deux événements se produisent sachant qu'ils sont indépendants=> p(littéraire$\cap$ garçon)=p(littéraire)*p(garçon) chaque probabilité est le rapport entre l'effectif de la modalité et effectif total.

n11=200*p11=120*120 /200=72

Etudes	Garçon	fille	Total1
Littéraire	72	48	120
Scientifique	36	24	60
Technique	12	8	20
Total	120	80	200

=>les effectifs théoriques sont >5=> on peut appliquer le test de chi2

3-calculer la statistique (ici la distance quadratique)

D=Σ(nij-$n_{ij\text{-théorique}}$)^2/ $n_{ij\text{-théorique}}$

D=$(60-72)^2/72+(42-36)^2/36+(18-12)^2/12+(60-48)^2/48+(18-24)^2/24+(2-8)^2/8$

D= $\chi2_{(\text{observé})}$ =15

$\chi2_{(1-\alpha)}$(ddl)= $\chi2_{(1-\alpha)}$ (l-1)*(c-1) / ddl=degré de liberté =(l-1)*(c-1)=(3-1)*(2-1)

L : nombre de modalités de 1 lère variable, c : nombre de modalités de la 1 lère variable

$\chi2_{(1-\alpha)}$(ddl)= $\chi2_{(0.999)}$(2)=13.82

4-appliquer la règle de décision :

Il s'agit de comparer D et $\chi2_{(1-\alpha)}$(ddl)=>D=15>13.82=> H0 est rejeté au seuil du risque 0.001

=>H1 est acceptée : On peut donc considérer que le type d'étude et le sexe des étudiants sont liés significativement.

l'erreur de rejeter H0 à tort est non significative => on rejette H0 => il y a une relation significative entre X et Y.

Exercice3 :

0-variable considérée, X : poids de récolte en tonnes par année, c'est une variable discrète.

1-définir H0 : les récoltes sont chaque année aussi bonnes (les mêmes)

H1 : les récoltes de chaque année sont différentes

On va utiliser le test d'ajustement de chi2 à une distribution de 20/94=0.2 sur les Cinq années de récolte avec n=94

2-tester la condition d'application :

Année	1iere	2ieme	3ieme	4ieme	5ieme
Pi(théorique)	0.2	0.2	0.2	0.2	0.2
n*pi	18.8	18.8	18.8	18.8	18.8

Les n*pi, qui constituent les effectives théoriques sont de 18.8>5=> la condition d'application du test est respectée

3-calculer la statistique :

$D=\Sigma(nij-n_{i\text{-théorique}})^2/\ n_{i\text{-théorique}}$

$D=(20-18.8)^2/18.8+(30-18.8)^2/18.8+(18-18.8)^2/18.8+(16-18.8)^2/18.8+(10-18.8)^2/18.8$

$D= \chi 2_{(\text{observé})} =11.3191$

$\chi 2_{(1-\alpha)}(ddl)= \chi 2_{(1-\alpha)}\ (k-r-1)$ / K: nombre de modalités (ou classes) de la variable=5, r : nombre de paramètres estimés de la loi à laquelle on va ajuster l'échantillon=0.

ddl=degré de liberté= 5-0-1=4

$\chi 2_{(1-\alpha)}(ddl)= \chi 2_{(0.95)}(4)=9.49$

4-appliquer la règle de décision :

Il s'agit de comparer D et $\chi 2_{(1-\alpha)}(ddl)$=>D=11.3191>9.49= $\chi 2_{(0.95)}(4)$=> H0 est rejetée au seuil du risque 0.05

=>H1 est acceptée **:** On peut donc considérer que les récoltes de chaque année sont différentes

Exercice 4 :

1)x=c(4.5,4.6,4.7,4.8,4.9,5.0,5.1) ;y=c(1.0,1.1,1.2,1.4,1.5,1.5,1.7)

plot(x,y,main= " ",xlab= "pourcentage en zinc ",ylab= "force de tension ")

2) calcule des coefficients :

La forme du modèle de régression linéaire simple : $Y=\beta_0+\beta_1X+\varepsilon$

Selon la méthode des moindres carrés on peut estimer β_0 et $\beta_1$somme suit :

$\widehat{\beta 1}=\frac{\sum_{i=1}^{n}(xi-\bar{x})*(yi-\bar{y})}{\sum_{i=1}^{n}(xi-\bar{x})^{2}}$, $\widehat{\beta 0}=\bar{y}-\widehat{\beta 1}\bar{x}$

$\bar{y}$=1.34

$\bar{x}$=4.8

xi	xi-$\bar{x}$	(xi-$\bar{x}$)^2	yi	Yi-$\bar{y}$	(xi-$\bar{x}$)* (Yi-$\bar{y}$)
4.5	-0.3	0.09	1.0	-0.34	0.102
4.6	-0.2	0.04	1.1	-0.24	0.048
4.7	-0.1	0.01	1.2	-0.14	0.014
4.8	0	0	1.4	0.06	0
4.9	0.1	0.01	1.5	0.16	0.016
5.0	0.2	0.04	1.5	0.16	0.032
5.1	0.3	0.09	1.7	0. 36	0.108
TOTAL	---	0.28	---	---	0.320

=>$\widehat{\beta 1}$=0.320 /0.28 =1.1428 d'où $\widehat{\beta 0}$=1.34-1.1428*4.8=-4.145 => Y=-4.145+ 1.1428 *X

3) Y=-4.145+ 1.1428 *(6)=**2.7118**