**Université de Lille** **L3 MIASHS**

# Statistiques décisionnelles : TP 1

Nom : Belladjo

Prénom: Keroudine

Prof : R. Butez

Exercice1

*On dispose de l'échantillon de 50 variables indépendantes suivant, les données ont été classées par ordre croissant pour en faciliter le traitement. Les ex-aequo proviennent d'arrondis.*

$X_1$=0.02, $X_2$=0.02, $X_3$=0.07 , $X_4$=0.07, $X_5$=0.07, $X_6$=0.09, $X_7$=0.11, $X_8$ =0.13, $X_9$=0.16, $X_{10}$=0.16, $X_{11}$=0.2, $X_{12}$=0.2, $X_{13}$=0.21, $X_{14}$=0.25, $X_{15}$ =0.29, $X_{16}$=0.32, $X_{17}$=0.35, $X_{18}$=0.35, $X_{19}$=0.37, $X_{20}$=0.37, $X_{21}$=0.38, $X_{22}$ 0.4, $X_{23}$=0.4, $X_{24}$=0.41, $X_{25}$=0.47, $X_{26}$=0.49 ,$X_{27}$=0.54 , $X_{28}$=0.55, $X_{29}$ =0.59, $X_{30}$=0.6, $X_{31}$=0.68, $X_{32}$=0.68 , $X_{33}$=0.7 , $X_{34}$=0.73, $X_{35}$=0.75, $X_{36}$ =0.79, $X_{37}$=0.79, $X_{38}$=0.8, $X_{39}$=0.81, $X_{40}$=0.81, $X_{41}$=0.86, $X_{42}$=0.89, $X_{43}$=1.03, $X_{44}$=1.04, $X_{45}$=1.04, $X_{46}$=1.09, $X_{47}$=1.12, $X_{48}$=1.21, $X_{49}$=1.33, $X_{50}$=1.42

*1)* Testons l'adéquation de ces données à une loi Unif[0,1.5] à l'aide d'un test du $X^2$ au niveau 90% . On proposera un découpage en 5 intervalles de même longueur. Calculer la p-valeur.

Fx(t)= $\int_0^t 1/1.5dx$

= $\int_0^t 2/3dx$

Fx(t) = 2/3t

n=50

| Valeurs (Xi) | Effectifs observés (Ni) | Pref = (P(Xi<=t) | Effectifs théoriques (n*Prefi) | Proportion (Pi) |
|---|---|---|---|---|
| [0,0.3] | 15 | 0,2 | 10 | 15/50=0.3 |
| [0.3,0.6] | 15 | 0,2 | 10 | 15/50= 0.3 |
| [0.6,0.9] | 12 | 0,2 | 10 | 12/50=0.24 |
| [0.9,1.2] | 5 | 0,2 | 10 | 5/50=0.1 |
| [1.2,1.5] | 3 | 0,2 | 10 | 3/50=0.06 |

$$D = \frac{(15-10)^2}{10} + \frac{(15-10)^2}{10} \frac{(12-10)^2}{10} + \frac{(5-10)^2}{10} + \frac{(3-10)^2}{10}$$

$$D = 50 * (\frac{(0.3-0.2)^2}{0.2} + \frac{(0.3-0.2)^2}{0.2} + \frac{(0.24-0.2)^2}{0.2} + \frac{(0.1-0.2)^2}{0.2} + \frac{(0.06-0.2)^2}{0.2})$$

D= 50 * (0.256)
D=12.8
$D \sim \chi^2(4)$
$h_{1-a}$= 7,7794

Code R :

```
3:1    (Top Level) ÷

Console   Terminal ×   Jobs ×

R 4.1.2 · ~/Documents/statdecision/tp/
R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML
Tapez 'q()' pour quitter R.

> X=c(15,15,12,5,3)
> pref=c(0.2,0.2,0.2,0.2,0.2)
> chisq.test(X,p=pref)

        Chi-squared test for given probabilities

data:  X
X-squared = 12.8, df = 4, p-value = 0.0123

>
```

D>$h_{1-a}$ alors on rejette $H_0$

P-valeurs = 0.0123

2) Tester l'adéquation de ces données à la loi dont la densité est

f(x) = cos(x)1x∊[0,π/2]
à l'aide d'un test du χ2, au niveau 90%. On pourra calculer la fonction de répartition de cette fonction et déterminer les probabilités des intervalles construits précédemment, ou bien en construire de nouveaux à l'aide des quantiles. Calculer la p-valeur.

$n = 50$

$\propto = 0{,}90$

Calcule de la fonction de répartition:

$$P(X \leq t) = \int_0^t cos(x)dx \ \text{ avec } \ t \in [0, \pi/2]$$

$$= [sin(x)]_0^t$$

$$= sin(t) - sin(0)$$

$$P(X \leq t) = sint(t)1_{0 \leq t \leq \pi/2}$$

| Valeurs (Xi) | Effectifs observés (Ni) | $P_i^{ref}$ | Effectifs théoriques $(n * P_i^{ref})$ | Proportion (Pi) |
|---|---|---|---|---|
| [0,0.3] | 15 | 0.295 | 14.5 | 15/50=0.3 |
| [0.3,0.6] | 15 | 0.269 | 13.45 | 15/50= 0.3 |
| [0.6,0.9] | 12 | 0,218 | 10.9 | 12/50=0.24 |
| [0.9,1.2] | 5 | 0,148 | 7,4 | 5/50=0.1 |
| [1.2,1.5] | 3 | 0,065 | 3,27 | 3/50=0.06 |

Calcule de la fonction des $P_i^{ref}$:

$$P_i^{ref} = \int_{t-0.3}^{t} cos(x)dx = sin(t) - sin(t - 0.3)$$

*Pour tout intervalle de la forme [t-0.3 , t] avec t-(t-0.3) = 0.3 pour t* $\in [0, 1.5]$

$n * P_5^{ref}$ =50*0.065 =3,27 < 5 , c'et la valeurs de colonne 4 et de la ligne 6. On regroupe alors les lignes 5 et plus , on a donc ;

| Valeurs (Xi) | Effectifs observés (Ni) | $P_i^{ref}$ | Effectifs théoriques $(n * P_i^{ref})$ | Proportion (Pi) |
|---|---|---|---|---|
| [0,0.3] | 15 | 0.295 | 14.5 | 15/50=0.3 |
| [0.3,0.6] | 15 | 0.269 | 13.45 | 15/50= 0.3 |
| [0.6,0.9] | 12 | 0,218 | 10.9 | 12/50=0.24 |
| [0.9,+∞[ | 8 | 0.218 | 10.9 | 8/50=0.16 |

$$P_4^{ref} = 1 - \sum_{i=1}^{3} P_i^{ref}$$

$$= 1 - 0{,}782$$

$$P_4^{ref} = 0.218$$

$$D = \frac{(15 - 14.5)^2}{14.5} + \frac{(15 - 13.45)^2}{13.45} \frac{(12 - 10.9)^2}{10.9} + \frac{(8 - 10.9)^2}{10.9}$$

$D = 1.07$

CODE R:

```
> X=c(15,15,12,8)
> pref=c(0.295,0.269,0.218,0.218)
> chisq.test(X,p=pref)

        Chi-squared test for given probabilities

data:  X
X-squared = 1.0654, df = 3, p-value = 0.7854

>
```

$D \sim \chi^2(3)$

$h_{1-a}$= 6.251

D < $h_{1-a}$

On conserve alors $H_0$

P-valeurs = 0.7854

3) Reprendre les questions précédentes avec un test de Kolmogorov-Smirnov en modifiant les valeurs pour ne pas avoir d'ex-æquo (il n'y en avait aucun avant l'arrondi). Calculer la p-valeur dans chacun des cas et comparez aux p-valeurs obtenues précédemment.

$X_1$=0.02, $X_2$=0.021, $X_3$=0.07 , $X_4$=0.071, $X_5$=0.0711, $X_6$=0.09, $X_7$=0.11, $X_8$=0.13, $X_9$=0.16, $X_{10}$=0.161, $X_{11}$=0.2, $X_{12}$=0.201, $X_{13}$=0.21, $X_{14}$=0.25, $X_{15}$=0.29, $X_{16}$=0.32, $X_{17}$=0.35, $X_{18}$=0.351, $X_{19}$=0.37, $X_{20}$=0.371, $X_{21}$ =0.38, $X_{22}$=0.4, $X_{23}$=0.401, $X_{24}$=0.41, $X_{25}$=0.47, $X_{26}$=0.49 ,$X_{27}$=0.54 , $X_{28}$=0.55, $X_{29}$=0.59, $X_{30}$=0.6, $X_{31}$=0.68, $X_{32}$=0.681 , $X_{33}$=0.7 , $X_{34}$=0.73, $X_{35}$=0.75, $X_{36}$=0.79, $X_{37}$=0.791, $X_{38}$=0.8, $X_{39}$=0.81, $X_{40}$=0.811, $X_{41}$ =0.86, $X_{42}$=0.89, $X_{43}$=1.03, $X_{44}$=1.04, $X_{45}$=1.041, $X_{46}$=1.09, $X_{47}$=1.12, $X_{48}$=1.21, $X_{49}$=1.33, $X_{50}$=1.42

Testons l'adéquation de ces données à une loi Unif[0,1.5] à l'aide de Kolmogorov-Smirnov au niveau 90% . On proposera un découpage en 5 intervalles de même longueur. Calculer la p-valeur.

La fonction de répartition:

$$F_X(t) = P(X \leq t)$$

$$= \int_0^{1.5} 1/1.5 dx$$

$$= [\frac{2}{3}x]_0^t$$

$$F_X(t) = \frac{2}{3}t1_{0\leq t\leq 1.5}$$

La fonction de répartition empirique:

$$F_n(t) = \sum_{i=1}^{3} \frac{1}{50} 1_{X\leq t}$$

Le test statistique de Kolmogorov-Smirnov:

| $X_i$ | $F_X(t)$ | $F_n(t)$ | $\lvert\frac{i}{50} - F_X(t)\rvert$ | $\lvert\frac{i-1}{50} - F_X(t)\rvert$ |
|---|---|---|---|---|
| 1.42 | 0.946 | 1.0 | 0.053 | 0.034 |
| 1.33 | 0.886 | 0.98 | 0.093 | 0.073 |

| $X_i$ | $F_X(t)$ | $F_n(t)$ | $\lvert\frac{i}{50} - F_X(t)\rvert$ | $\lvert\frac{i-1}{50} - F_X(t)\rvert$ |
|---|---|---|---|---|
| 1.21 | 0.806 | 0.96 | 0.153 | 0.133 |
| 1.12 | 0.746 | 0.94 | 0.193 | 0.173 |
| 1.041 | 0.694 | 0.9 | 0.206 | 0.186 |
| 1.04 | 0.693 | 0.88 | 0.187 | 0.167 |
| 1.03 | 0.686 | 0.86 | 0.173 | 0.153 |
| 0.89 | 0.593 | 0.84 | 0.247 | 0.227 |
| 0.86 | 0.573 | 0.82 | 0.247 | 0.227 |
| 0.811 | 0.541 | 0.8 | 0.259 | 0.239 |
| 0.81 | 0.54 | 0.78 | 0.24 | 0.220 |
| 0.8 | 0.53 | 0.76 | 0.227 | 0.21 |
| 0.79 | 0.52 | 0.72 | 0.193 | 0.173 |
| 0.75 | 0.5 | 0.7 | 0.220 | 0.180 |
| 0.73 | 0.486 | 0.68, | 0.193 | 0.174 |
| 0.7 | 0.466 | 0.66 | 0.193 | 0.174 |
| 0.681 | 0.454 | 0.64 | 0.186 | 0.166 |
| 0.68 | 0.453 | 0.62 | 0.167 | 0.147 |
| 0.6 | 0.399 | 0.6 | 0.2 | 0.18 |
| 0.59 | 0.393 | 0.58 | 0.186 | 0.166 |
| 0.55 | 0.366 | 0.56 | 0.193 | 0.173 |
| 0.54 | 0.36 | 0.54 | 0.180 | 0.160 |
| 0.49 | 0.326 | 0.52 | 0.193 | 0.173 |
| 0.47 | 0.313 | 0.5 | 0.187 | 0.167 |
| 0.41 | 0.273 | 0.48 | 0.207 | 0.187 |
| 0.401 | 0.267 | 0.46 | 0.193 | 0.173 |
| 0.4 | 0.266 | 0.44 | 0.173 | 0.153 |

| $X_i$ | $F_X(t)$ | $F_n(t)$ | $\lvert\frac{i}{50} - F_X(t)\rvert$ | $\lvert\frac{i-1}{50} - F_X(t)\rvert$ |
|---|---|---|---|---|
| *0.38* | 0.253 | 0.42 | 0.167 | 0.147 |
| *0.371* | 0.247 | 0.4 | 0.153 | 0.133 |
| *0.351* | 0.234 | 0.36 | 0.126 | 0.106 |
| *0.35* | 0.233 | 0.34 | 0.107 | 0.087 |
| *0.29* | 0.193 | 0.3 | 0.107 | 0.087 |
| *0.25* | 0.167 | 0.28 | 0.113 | 0.093 |
| *0.201* | 0.134 | 0.24 | 0.105 | 0.086 |
| *0.161* | 0.107 | 0.2 | 0.093 | 0.073 |
| *0.16* | 0.106 | 0.18 | 0.073 | 0.053 |
| *0.13* | 0.086 | 0.16 | 0.073 | 0.053 |
| *0.11* | 0.073 | 0.14 | 0.067 | 0.047 |
| *0.09* | 0.06 | 0.12 | 0.06 | 0.040 |
| *0.0711* | 0.047 | 0.1 | 0.053 | 0.033 |
| *0.071* | 0.047 | 0.08 | 0.33 | 0.013 |
| *0.07* | 0.047 | 0.06 | 0.013 | 0.007 |
| *0.021* | 0.014 | 0.04 | 0.026 | 0.014 |
| *0.02* | 0.014 | 0.02 | 0.006 | |
| *1.09* | 0.726 | 0.92 | 0.193 | 0.173 |
| *0.791* | 0.52 | 0.74 | 0.212 | 0.193 |
| *0.37* | 0.246 | 0.38 | 0.133 | 0.113 |
| *0.32* | 0.213 | 0.32 | 0.107 | 0.087 |
| *0.21* | 0.140 | 0.26 | 0.120 | 0.1 |
| *0.2* | 0.133 | 0.22 | 0.086 | 0.067 |

$$h(F_X(t), F_n(t)) = \max(\max(\lvert\frac{i}{50} - F_X(t)\rvert, \lvert\frac{i-1}{50} - F_X(t)\rvert))$$

$$= \max(0.259; 0.239)$$

$$h(F_X(t), F_n(t)) = 0.259$$

CODE R :

```
Console   Terminal ×   Jobs ×
R 4.1.2 · ~/Documents/statdecision/tp/
> X=c(0.021,0.021,0.07,0.071,0.0711, 0.09, 0.11, 0.13, 0.16, 0.161, 0.2, 0.201, 0.21, 0.25, 0.29, 0.32,
 0.35, 0.351, 0.37, 0.371, 0.38,0.4, 0.401, 0.41, 0.47, 0.49, 0.54, 0.55, 0.59, 0.6, 0.68, 0.681, 0.7, 0.
73, 0.75, 0.79, 0.791,0.8, 0.81, 0.811, 0.86, 0.89, 1.03, 1.04, 1.041, 1.09, 1.12, 1.21, 1.33, 1.42)
> ks.test(X,"punif",0,1.5)

	One-sample Kolmogorov-Smirnov test

data:  X
D = 0.25933, p-value = 0.0024
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test(X, "punif", 0, 1.5) :
  ties should not be present for the Kolmogorov-Smirnov test
```

P-valeurs = 0.0024

$h_{1-a} = 0{,}1696$

$h(F_X(t), F_n(t)) > h_{1-a}$ on rejette alors $H_0$

2) Tester l'adéquation de ces données à la loi dont la densité est (x) = cos(x)1x∈[0,π/2] à l'aide d'un test de Kolmogorov-Smirnov, au niveau 90%. On pourra calculer la fonction de répartition de cette fonction et déterminer les probabilités des intervalles construits précédemment, ou bien en construire de nouveaux à l'aide des quan- tiles. Calculer la p-valeur.

$n = 50$

$\propto = 0{,}90$

Calcule de la fonction de répartition:

$$P(X \leq t) = \int_0^t cos(x)dx \text{ avec } t \in [0, \pi/2]$$

$$= [sin(x)]_0^t$$

$$= sin(t) - sin(0)$$

$$P(X \leq t) = sint(t)1_{0 \leq t \leq \pi/2}$$

| $X_i$ | $F_n(t)$ | $F_X(t)$ | $\lvert\frac{i}{50} - F_X(t)\rvert$ | $\lvert\frac{i-1}{50} - F_X(t)\rvert$ |
|---|---|---|---|---|
| 0.02 | 0.02 | 0.210 | 0.00099 | 0.210 |
| 0.021 | 0.04 | 0.021 | 0.019 | 0.050 |
| 0.07 | 0.06 | 0.069 | 0.009 | 0.031 |
| 0.071 | 0.08 | 0.070 | 0.009 | 0.011 |
| 0.0711 | 0.10 | 0.071 | 0.029 | 0.010 |
| 0.09 | 0.12 | 0.090 | 0.030 | 0.010 |
| 0.11 | 0.14 | 0.1097 | 0.030 | 0.019 |
| 0.13 | 0.16 | 0.1296 | 0.030 | 0.0003 |
| 0.16 | 0.18 | 0.159 | 0.021 | 0.019 |
| 0.161 | 0.20 | 0.160 | 0.039 | 0.0004 |
| 0.2 | 0.22 | 0.198 | 0.021 | 0.012 |
| 0.201 | 0.24 | 0.1996 | 0.040 | 0.007 |
| 0.21 | 0.26 | 0.208 | 0.051 | 0.026 |
| 0.25 | 0.28 | 0.247 | 0.032 | 0.035 |
| 0.29 | 0.30 | 0.285 | 0.014 | 0.043 |
| 0.32 | 0.32 | 0.314 | 0.005 | 0.024 |
| 0.35 | 0.34 | 0.343 | 0.003 | 0.022 |
| 0.351 | 0.36 | 0.344 | 0.016 | 0.003 |
| 0.37 | 0.38 | 0.361 | 0.018 | 0.009 |
| 0.371 | 0.40 | 0.363 | 0.037 | 0.011 |
| 0.38 | 0.42 | 0.371 | 0.049 | 0.0296 |
| 0.4 | 0.44 | 0.389 | 0.050 | 0.041 |
| 0.401 | 0.46 | 0.390 | 0.070 | 0.007 |
| 0.41 | 0.48 | 0.3986 | 0.081 | 0.009 |
| 0.47 | 0.50 | 0.453 | 0.047 | 0.014 |

| $X_i$ | $F_n(t)$ | $F_X(t)$ | $\lvert \frac{i}{50} - F_X(t) \rvert$ | $\lvert \frac{i-1}{50} - F_X(t) \rvert$ |
|---|---|---|---|---|
| 0.49 | 0.52 | 0.471 | 0.049 | 0.003 |
| 0.54 | 0.54 | 0.514 | 0.026 | 0.016 |
| 0.55 | 0.56 | 0.523 | 0.037 | 0.005 |
| 0.59 | 0.58 | 0.556 | 0.024 | 0.04994 |
| 0.6 | 0.60 | 0.565 | 0.035 | 0.0295 |
| 0.68 | 0.62 | 0.629 | 0.009 | 0.024 |
| 0.681 | 0.64 | 0.629 | 0.010 | 0.027 |
| 0.7 | 0.66 | 0.644 | 0.016 | 0.022 |
| 0.73 | 0.68 | 0.667 | 0.013 | 0.030 |
| 0.75 | 0.70 | 0.682 | 0.018 | 0.011 |
| 0.79 | 0.72 | 0.710 | 0.010 | 0.003 |
| 0.791 | 0.74 | 0.711 | 0.029 | 0.016 |
| 0.8 | 0.76 | 0.717 | 0.043 | 0.035 |
| 0.81 | 0.78 | 0.724 | 0.056 | 0.022 |
| 0.811 | 0.80 | 0.724 | 0.075 | 0.023 |
| 0.86 | 0.82 | 0.758 | 0.062 | 0.037 |
| 0.89 | 0.84 | 0.777 | 0.063 | 0.022 |
| 1.03 | 0.86 | 0.857 | 0.003 | 0.010 |
| 1.04 | 0.88 | 0.862 | 0.018 | 0.007 |
| 1.041 | 0.90 | 0.862 | 0.037 | 0.0001 |
| 1.09 | 0.92 | 0.886 | 0.033 | 0.016 |
| 1.12 | 0.94 | 0.900 | 0.039 | 0.0311 |
| 1.21 | 0.96 | 0.936 | 0.024 | 0.029 |
| 1.33 | 0.98 | 0.971 | 0.009 | 0.002 |
| 1.42 | 1.00 | 0.988 | 0.011 | 0.008 |

$h(F_X(t), F_n(t)) = \max(0.081{,}0.04994)$

$h(F_X(t), F_n(t)) = 0.081$

P-valeurs = 0.8949

```
20:307   (Top Level) ÷                                             R Markdown ÷
Console   Terminal ×   Jobs ×
R 4.1.2 · ~/Documents/statdecision/tp/
> X=c(0.021,0.021,0.07,0.071,0.0711, 0.09, 0.11, 0.13, 0.16, 0.161, 0.2, 0.201, 0.21, 0.25, 0.29, 0.32,
 0.35, 0.351, 0.37, 0.371, 0.38,0.4, 0.401, 0.41, 0.47, 0.49, 0.54, 0.55, 0.59, 0.6, 0.68, 0.681, 0.7, 0.
73, 0.75, 0.79, 0.791,0.8, 0.81, 0.811, 0.86, 0.89, 1.03, 1.04, 1.041, 1.09, 1.12, 1.21, 1.33, 1.42)
> ks.test(X,sin)

        One-sample Kolmogorov-Smirnov test

data:  X
D = 0.081391, p-value = 0.8949
alternative hypothesis: two-sided
```

$h_{1-a} = 0{,}1696$

$h(F_X(t), F_n(t)) < h_{1-a}$ , on conserve alors $H_0$

Exercice 2

1) L'énoncé

Un calculateur a simulé un échantillon de n = 10 valeurs distribuées selon une loi normale.

14.63427, 12.62687, 14.63303, 14.25101, 16.59716, 13.34633, 10.59761, 16.91010, 22.15402 ,22.09362

Les valeurs Xi produites sont rangées par ordre croissant :

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $Xi$ | 10.59761 | 12.62687 | 13.34633 | 14.25101 | 14.63303 | 14.63427 | 16.59716 | 16.91010 | 22.09362 | 22.15402 |

a) Donner une estimation de la moyenne et l'écart-type de l'échantillon.

b) Exécuter un test de Kolmogorov-Smirnov au seuil de 5% pour décider si la distribution de l'échantillon est en adéquation avec la loi normale.

2) On a utilisiser la fonction R rnorm pour générer nos données:

Ci-dessous, on genèere n = 10 observations de la loi normale N(μ = 15, σ2 = 3).

```
> rnorm(1:10,15,3)
 [1] 14.63427 12.62687 14.63303 14.25101 16.59716 13.34633 10.59761 16.91010 22.15402 22.09362
> 
```

3) La solution:

a) - une estimation de la moyenne l'échantillon :

```
>
> X=c(14.63427, 12.62687, 14.63303, 14.25101, 16.59716, 13.34633, 10.5976
1, 16.91010, 22.15402 ,22.09362)
> moyenne=mean(X)
> moyenne
[1] 15.7844
> 
```

$\bar{x} = 15.7844$

- une estimation de l'écart-type de l'échantillon :

```
> X=c(14.63427, 12.62687, 14.63303, 14.25101, 16.59716, 13.34633, 10.5976
1, 16.91010, 22.15402 ,22.09362)
> ecartype=sqrt(var(X))
> ecartype
[1] 3.802379
>
```

$\sigma = 3.802379$

b) Exécuter un test de Kolmogorov-Smirnov au seuil de 5% pour décider si la distribution de l'échantillon est en adéquation avec la loi normale.

Voici la sortie du logiciel R pour le test:

```
> X=c(14.63427, 12.62687, 14.63303, 14.25101, 16.59716, 13.34633, 10.5976
1, 16.91010, 22.15402 ,22.09362)
> moyenne=mean(X)
> ecartype=sqrt(var(X))
> ks.test(X,"pnorm",moyenne,ecartype)

        One-sample Kolmogorov-Smirnov test

data:  X
D = 0.21886, p-value = 0.6487
alternative hypothesis: two-sided

>
```

On trouve $D = 0.21886$. La $p - valeur = 0.6487$ est nettement supérieure à 0.05, donc on accepte l'hypothèse $H_0$ selon laquelle la distribution de l'échantillon est en adéquation avec la loi normale

Exercice 3

1) Avec cette définition, si n = 1000, quelle variable faut-il regarder si on veut le quantile d'ordre 0.95 ? Quel le variable faut-il regarder si on veut le quantile d'ordre 0.975 ?

On sait que :

$$Q_n(p) = X_{(np)} \quad \text{si np est entier}$$

$$Q_n(p) = X_{\lfloor (np) \rfloor + 1} \text{ sinon.}$$

On a donc pour n=1000 et p=0.95 :

$n * p = 1000 * 0.95 = 950$

Donc si on veut le quantile d'ordre 0.95 on regarde le variable $X_{950}$

De la même manière si on veut le quantile d'ordre 0.975 on regardera variable $X_{975}$

2) Quelle valeur de n faut-il choisir pour garantir que le quantile estimé a ses deux premiers chiffres après la virgule corrects ? Dans toute la suite, on utilisera au minimum cette valeur de n.

D'après l'énoncé l'estimation du quantile est d'ordre $\frac{1}{\sqrt{n}}$ , on choisira donc la valeurs n telle que :

$$\frac{1}{\sqrt{n}} = 0.01$$

$$\Leftrightarrow n = 10000$$

3) Utilisez la fonction rnorm pour recréer la table des quantiles de la loi normale entre 0.8 et 0.95, échantillonnée tous les 0.1. Comparez aux vrais quantiles de la loi normale. Quelle valeur de n faut-il prendre pour obtenir une précision de 10−2 ?

```
> Q_0_1=rnorm(10000)
> Q=sort(Q_0_1)
> Q=c(Q)
> Q_estime=Q[c(8000,  8100,  8200,  8300,  8400,  8500,  8600,  8700,  8800,  8900,  9000,  9100,  9200,
 9300,  9400,  9500)]
> Q_estime
 [1] 0.8492806 0.8856229 0.9213354 0.9632237 1.0008201 1.0353325 1.0815584 1.1302207 1.1782310
[10] 1.2151721 1.2686954 1.3225429 1.4032546 1.4705290 1.5536896 1.6474880
> Q_reel=c(qnorm(0.8),qnorm(0.81),qnorm(0.82),qnorm(0.83),qnorm(0.84),qnorm(0.85),qnorm(0.86),qnorm(0.8
7),qnorm(0.88),qnorm(0.89),qnorm(0.90),qnorm(0.91),qnorm(0.92),qnorm(0.93),qnorm(0.94),qnorm(0.95))
> Q_reel
 [1] 0.8416212 0.8778963 0.9153651 0.9541653 0.9944579 1.0364334 1.0803193 1.1263911 1.1749868
[10] 1.2265281 1.2815516 1.3407550 1.4050716 1.4757910 1.5547736 1.6448536
> Erreur=c(Q_estime-Q_reel)
> Erreur=abs(Erreur)
> Erreur_moyen=mean(Erreur)
> Erreur_moyen
[1] 0.006213265
>
```

On a utiliser la commande rnorm(10000) pour générer nos 10000 données . On les a ensuite ordonné avec la commande sort() et on les nommé Q

Et comme on sait pour obtenir le quantile d'ordre 0.8 par exemple on regarde la variable $X_{8000}$. Ainsi en suivant le même principe nous savons que nos quantiles estimer sont a la possition (8000, 8100, 8200, 8300, 8400, 8500, 8600, 8700, 8800, 8900, 9000, 9100, 9200, 9300, 9400, 9500) d'ordre 0.8 à 0.95 échantillonné tous les O.1. pour le faire on a effectuer la commansde suivant:

Q_estime=Q[c(8000, 8100, 8200, 8300, 8400, 8500, 8600, 8700, 8800, 8900, 9000, 9100, 9200, 9300, 9400, 9500)]

Nous Avon ensuite calculer les vraies quantiles Q_reel et nous l'avons comparé avec les quantiles estimés Q_estime. Pour faire la comparaison nous avant calculer l'erreur moyenne. Et nous constatons une erreur moyenne de 0.006213265

Pour obtenir une précision de 10−2 on choisira donc la valeurs de n telle que:

$$\frac{c}{\sqrt{10000}} = 0.006$$

$$\Leftrightarrow c = 0.006 * \sqrt{10000}$$

$$\Leftrightarrow c = 0.6$$

On a donc :

$$\frac{0.6}{\sqrt{n}} = 0.01$$

$$n = 3600$$