SOLUTION TP n° 8

Solution 1. On lance 100 fois un dé à 6 faces numérotées de 1 à 6. On obtient les résultats suivants :

Numéro	1	2	3	4	5	6
Nombre de fois	18	23	19	12	11	15

Peut-on affirmer que le dé est truqué ? *(on fera une analyse graphique convenable, puis un test statistique adapté au risque 5%)* :

```
nb = c(18, 23, 19, 12, 11, 15)
bar = barplot(nb / 100, col = "white")
points(bar, rep(1 / 6, 6), type = "h")
proba = rep(1 / 6, 6)
chisq.test(nb, p = proba)$p.value
```

Cela renvoie : `[1] 0.2757299`. Comme p-valeur > 0.05, on ne peut pas affirmer que le dé est truqué.

Solution 2. Une enquête effectuée auprès du comptoir de 116 coopératives agricoles a permis d'étudier l'arrivée dans le temps des usagers de ces coopératives. Pendant une heure, on a les résultats suivants :

Nombre d'usagers arrivés	0	1	2	3	4	5
Nombre de coopératives	28	31	27	19	7	4

Soit X la *var* égale au nombre d'usager arrivés en une heure. Peut-on affirmer, au risque 5%, que X suit la loi de Poisson $\mathcal{P}(2)$?

```
proba = c(dpois(0:4, 2), 1 - ppois(4, 2))
chisq.test(c(28, 31, 27, 19, 7, 4), p = proba)$p.value
```

Cela renvoie : `[1] 0.03073932`. Comme p-valeur < 0.05, on peut affirmer, au risque 5%, que X ne suit pas la loi de Poisson $\mathcal{P}(2)$.

Solution 3. Le fabriquant d'un certain type d'appareil affirme que :

- la durée de vie moyenne d'un appareil est de 1500 heures,
- la durée de vie d'un appareil peut être modélisée par une *var* suivant une loi exponentielle.

Afin de tester cette affirmation, on mesure la durée de vie en heures de 10 de ces appareils pris au hasard. Les résultats sont :

576	617	1718	698	3335	512	2024	3259	3146	754

Peut-on dire, au risque 5%, que le fabriquant a tort ?

```
x = c(576, 617, 1718, 698, 3335, 512, 2024, 3259, 3146, 754)
ks.test(x, "pexp", 1/1500)$p.value
```

Cela renvoie : `[1] 0.3098538`. Comme p-valeur > 0.05, on ne peut rien dire sur le fait que le fraiquant ait tort.

Solution 4. Les données considérées sont les observations d'un caractère que l'on modélise comme une *var* X. Elles sont disponibles ici :

`https://chesneau.users.lmno.cnrs.fr/observations.txt`

1. Mettre les données dans un vecteur `x` :

```
x = scan("https://chesneau.users.lmno.cnrs.fr/observations.txt")
```

2. Représenter l'histogramme des fréquences, le boxplot et le QQ plot associés à `x`. Est-que X semble suivre une loi normale ?

```
hist(x)
boxplot(x)
qqnorm(scale(x))
```

3. Apporter une réponse plus tranchée avec un test statistique adapté :

```
shapiro.test(x)$p.value
```

Cela renvoie : `[1] 0.3426605`. Comme p-valeur > 0.05, l'hypothèse que X suit une loi normale n'est pas rejetée.

Solution 5. La pression artérielle systolique est la pression maximale du sang dans les artères au moment de la contraction du cœur. Celle-ci a été mesurée pour 29 individus de différents âges. Ainsi, pour chacun d'entre eux, on dispose :

- de leur pression systolique en mmHg,
- de leur âge en années.

On modélise ces deux variables comme des *var* Y et X_1. Le jeu de données "pression" est disponible ici :

`https://chesneau.users.lmno.cnrs.fr/pression.txt`

1. Mettre le jeu de données sous la forme d'une data frame `w`, puis attacher les noms des colonnes :

```
w = read.table("https://chesneau.users.lmno.cnrs.fr/pression.txt", header = T)
attach(w)
```

2. Peut-on affirmer, au risque 5%, que Y ne suit pas une loi normale ?

   ```
   shapiro.test(Y)$p.value
   ```

   Cela renvoie : `[1] 0.6421197`. Comme p-valeur > 0.05, l'hypothèse que Y suit une loi normale n'est pas rejetée.

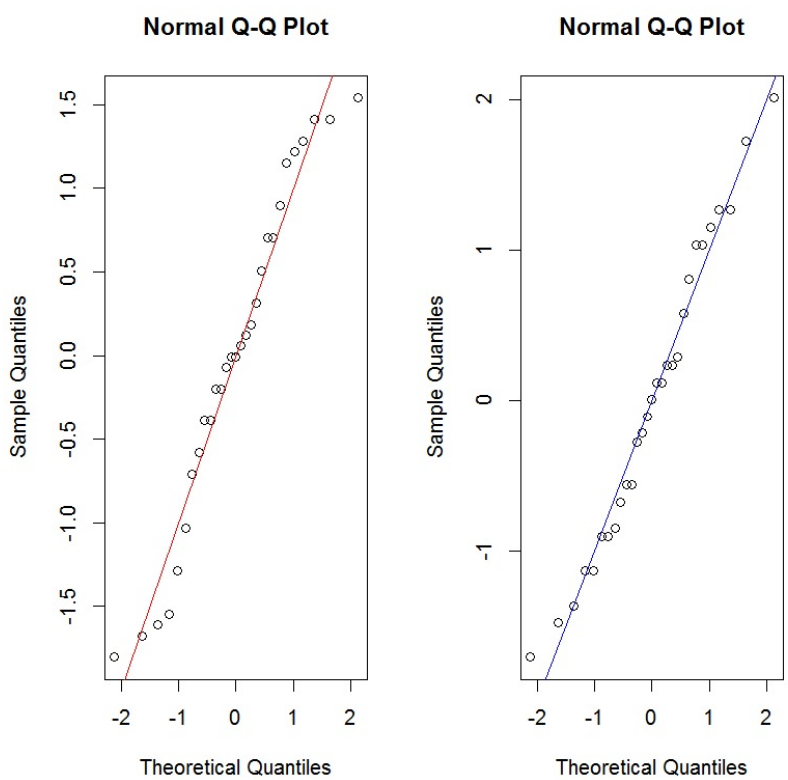
3. Peut-on affirmer, au risque 5%, que X_1 ne suit pas une loi normale ?

   ```
   shapiro.test(X1)$p.value
   ```

   Cela renvoie : `[1] 0.2083349`. Comme p-valeur > 0.05, l'hypothèse que X_1 suit une loi normale n'est pas rejetée.

4. Reproduire et comprendre l'enjeu des commandes suivantes :

   ```
   par(mfrow = c(1, 2))
   qqnorm(scale(X1))
   abline(0, 1, col = "red")
   qqnorm(scale(Y))
   abline(0, 1, col = "blue")
   ```



On analyse les QQ plot associés aux `var` X_1 et Y. On peut alors remarquer que les points sont presque alignés sur la droite diagonale d'équation $y = x$. On peut alors admettre que X_1 et Y suivent des lois normale.

Solution 6. Soient X et Y deux *var* indépendantes. Illustrer les résultats ci-dessous avec la commande `qqplot` :

Pour les QQ plot à venir, on peut remarquer que les points sont presque alignés sur la droite diagonale d'équation $y = x$, ce qui traduit l'égalité en loi étudiée.

- Caractérisation de la loi du chi-deux $\chi^2(2)$: Si $X \sim \mathcal{N}(0,1)$ et $Y \sim \mathcal{N}(0,1)$, alors

$$X^2 + Y^2 \sim \chi^2(2) :$$

```
a = rnorm(1000)
b = rnorm(1000)
c = rchisq(1000, 2)
qqplot(a^2 + b^2, c)
```

- Caractérisation de la loi de Student $\mathcal{T}(\nu)$: Si $X \sim \mathcal{N}(0,1)$ et $Y \sim \chi^2(\nu)$, alors

$$\frac{X}{\sqrt{\frac{Y}{\nu}}} \sim \mathcal{T}(\nu) :$$

Prendre $\nu = 3.9$:

```
a = rnorm(1000)
b = rchisq(1000, 3.9)
c = rt(1000, 3.9)
qqplot(a / sqrt(b / 3.9), c)
```

- Caractérisation de la loi de Fisher $\mathcal{F}(\nu_1, \nu_2)$: Si $X \sim \chi^2(\nu_1)$ et $Y \sim \chi^2(\nu_2)$, alors

$$\frac{\frac{X}{\nu_1}}{\frac{Y}{\nu_2}} = \frac{\nu_2 X}{\nu_1 Y} \sim \mathcal{F}(\nu_1, \nu_2) :$$

Prendre $(\nu_1, \nu_2) = (2.1, 8.3)$:

```
a = rchisq(1000, 2.1)
b = rchisq(1000, 8.3)
c = rf(1000, 2.1, 8.3)
qqplot((a / 2.1) / (b / 8.3), c)
```