SOLUTION TP n° 6

Solution 1. Une usine fabrique des billes métalliques. L'usine s'est engagée à fournir à un client des billes dont le diamètre moyen est de 25.21 millimètres. Le client réceptionne sa commande. Dans le lot reçu, il prélève un échantillon de 20 billes choisies au hasard et mesure les diamètres en millimètres suivants :

24.72	24.94	25.03	25.17	25.28	25.12	25.15	25.28	25.67	25.46
24.84	24.99	25.15	25.07	25.61	25.11	25.25	25.36	25.46	25.51

On suppose que le diamètre en millimètres d'une bille peut être modélisé par une *var* X suivant une loi normale.

À partir de ces données, peut-on affirmer que l'usine ne respecte pas ses engagements ? Faire un test statistique adapté au risque 5% :

```
x = c(24.72, 24.94, 25.03, 25.17, 25.28, 25.12, 25.15, 25.28, 25.67, 25.46,
24.84, 24.99, 25.15, 25.07, 25.61, 25.11, 25.25, 25.36, 25.46, 25.51)
t.test(x, mu = 25.21)$p.value
```

Cela renvoie : `[1] 0.9789626`. Comme p-valeur > 0.05, on ne peut rien affirmer.

Solution 2. On s'intéresse à la contenance des bouteilles de vin Château Beaulieu 2014. On mesure le contenu de 12 bouteilles extraites au hasard dans cette production. Les résultats, en centilitres, sont :

75.15	74.32	74.96	73.64	74.41	75.22	73.78	74.56	74.82	74.12	74.92	75.34

On suppose que le contenu en centilitres d'une bouteille cette production peut être modélisé par une *var* X suivant une loi normale.

Peut-on affirmer, au risque 5%, que le contenu moyen d'une bouteille de cette production est inférieur à 75 centilitres ?

```
x = c(75.15, 74.32, 74.96, 73.64, 74.41, 75.22, 73.78, 74.56, 74.82,
74.12, 74.92, 75.34)
t.test(x, mu = 75, alternative = "less")$p.value
```

Cela renvoie : `[1] 0.01607362`. Comme p-valeur < 0.05, la réponse est oui.

Solution 3. Un producteur de pommes a dans son cahier des charges la clause suivante : "le diamètre d'une pomme dans chaque lot fourni devra avoir un écart-type inférieur ou égal à 0.65 centimètres". Sur 8 pommes extraites au hasard dans un lot, on obtient les diamètres en centimètres suivants :

9.35	10.14	11.31	8.97	8.51	9.92	9.43	10.81

On suppose que le diamètre en centimètres d'une pomme de la production peut être modélisé par une *var* X suivant une loi normale.

Peut-on affirmer, au risque 1%, que le cahier des charges n'est pas respecté ?

```
x = c(9.35, 10.14, 11.31, 8.97, 8.51, 9.92, 9.43, 10.81)
library(OneTwoSamples)
var_test1(x, 0.65^2, side = 1)$P_value
```

Cela renvoie : `[1] 0.04310831`. Comme p-valeur > 0.01, on ne peut rien affirmer.

Solution 4. Un producteur d'œufs de poules a constaté que sur toute sa production de l'année 2008, la proportion d'œufs non-propres à la consommation (cassures, malformation. . .) était égale à 3.8%. En 2009, la proportion p d'œufs non-conformes n'est pas connue et le producteur ne peut travailler que sur un échantillon. Il souhaite n'avoir que 5 chances sur 100 de se tromper en disant à tort que la proportion d'œufs non-conformes pour 2009 est différente de celle de 2008. Il examine au hasard 300 œufs dans la production 2009. Parmi ceux-ci, 20 sont défectueux.

Effectuer le test statistique qui s'impose et énoncer clairement votre conclusion.

```
prop.test(20, 300, p = 0.038)$p.value
```

Cela renvoie : `[1] 0.01444767`. Comme p-valeur < 0.05, au risque 5%, on peut affirmer que la proportion d'œufs non-conformes pour 2009 est différente de celle de 2008.

On aurait aussu pu faire un test statistique exact :

```
binom.test(20, 300, p = 0.038)$p.value
```

Cela renvoie : `[1] 0.01479992`. Comme p-valeur < 0.05, au risque 5%, on peut affirmer que la proportion d'œufs non-conformes pour 2009 est différente de celle de 2008.

Solution 5. On cherche à savoir si le rendement moyen en blé dans une région A est strictement supérieur à celui dans une région B. On considère alors 16 parcelles différentes réparties sur les deux régions. Les résultats, en quintaux par hectare, sont :

- pour la région A :

48.12	48.24	50.41	53.59	54.62	56.38	57.77	58.65	60.52

- pour la région B :

44.27	46.31	48.29	48.47	50.58	51.23	55.44

Le rendement en blé en quintaux par hectare dans la région A peut être modélisé par une *var* X_1, et celui dans la région B peut être modélisé par une *var* X_2. On suppose que X_1 et X_2 suivent des lois normales.

Proposer un test statistique adapté au problème et conclure.

```
x1 = c(48.12, 48.24, 50.41, 53.59, 54.62, 56.38, 57.77, 58.65, 60.52)
x2 = c(44.27, 46.31, 48.29, 48.47, 50.58, 51.23, 55.44)
t.test(x1, x2, alternative = "greater")$p.value
```

Cela renvoie : `[1] 0.01368057`. Comme p-valeur < 0.05, au risque 5%, on peut affirmer que le rendement moyen en blé dans une région A est strictement supérieur à celui dans une région B.

Solution 6. On dispose de deux lots de boîtes de sauce italienne conditionnées de la même manière mais provenant de producteurs différents. On s'intéresse à la teneur en grammes de viande dans celles-ci.

- On extrait 7 boîtes provenant du premier producteur et on mesure leur teneur de viande. Les résultats, en grammes, sont :

12.12	12.03	13.58	13.38	11.81	15.92	13.65

- On extrait 6 boîtes provenant du deuxième producteur et on mesure leur teneur de viande. Les résultats, en grammes, sont :

14.81	13.93	14.91	15.87	15.62	15.39

La teneur en grammes de viande dans une boîte provenant du premier producteur peut être modélisée par une *var* X_1, et celle dans une boîte provenant du deuxième producteur peut être modélisée par une *var* X_2. On suppose que X_1 et X_2 suivent des lois normales.

1. Peut-on admettre que la dispersion de la teneur de viande dans une boîte ne diffère pas selon les producteurs ?

```
x1 = c(12.12, 12.03, 13.58, 13.38, 11.81, 15.92, 13.65)
x2 = c(14.81, 13.93, 14.91, 15.87, 15.62, 15.39)
var.test(x1, x2)$p.value
```

   Cela renvoie : `[1] 0.1374583`. Comme p-valeur > 0.05, on peut admettre cela (égalité des variances).

2. Peut-on affirmer qu'il y a une différence entre les producteurs quant à la teneur moyenne en viande dans les boîtes ?

```
t.test(x1, x2, var.equal = TRUE)$p.value
```

   Cela renvoie : `[1] 0.01398439`. Comme p-valeur < 0.05, la réponse est oui (la différence est significative $\star$).

Solution 7. Un industriel veut commercialiser une nouvelle production de rillette de saumon "Saumonmiam". Sur le marché existe déjà cette production mais sous la marque "Bonsaumon". L'industriel ne réalisera l'investissement que si un sondage où les réponses possibles sont "Mauvais","Moyen", "Bon" et "Très bon", permettra de conclure que, dans toute la population des consommateurs potentiels, le pourcentage de consommateurs qui donnent une appréciation "Bon" ou "Très bon" à "Saumonmiam" est supérieur au pourcentage de consommateurs qui donnent une appréciation "Bon ou Très bon" à "Bonsaumon". Il ne veut avoir que 5 chances sur 100 de se tromper dans son investissement. Un institut de dégustation réalise le sondage.

Dans la population des consommateurs potentiels, il interroge au hasard 130 personnes sur leur appréciation de "Bonsaumon". Parmi ceux-ci 98 donnent une appréciation "Bon ou Très bon" à "Bonsaumon". Dans la population des consommateurs potentiels, il interroge au hasard 150 autres personnes sur leur appréciation de "Saumonmiam". Parmi ceux-ci 122 donnent une bonne appréciation à "Bon ou Très bon" à "Saumonmiam". L'industriel investira-t-il dans la production de "Saumonmiam" ?

```
prop.test(c(122, 98), c(150, 130), alternative = "greater")$p.value
```

Cela renvoie : `[1] 0.1437014`. Comme p-valeur > 0.05, on ne peut pas conclure.

Solution 8. On souhaite tester si deux goûteurs donnent en moyenne la même appréciation à des produits différents. Chacun des deux goûteurs donne une note de 1 à 20. On présente 10 produits différents aux 2 goûteurs G1 et G2. Les résultats sont :

G1	G2
8	8.5
12	11
15.5	18
14	16
10	9
10	12
5	6.5
9	9
18.5	20
13	16

La note donnée par le goûteur G1 peut être modélisée par une *var* X_1 et celle donnée par le goûteur G2 peut être modélisée par une *var* X_2. On suppose que $X_1 - X_2$ suit une loi normale.

Peut-on affirmer, au risque 5%, que les jugements ne sont pas en moyenne identiques ?

```
x1 = c(8, 12, 15.5, 14, 10, 10, 5, 9, 18.5, 13)
x2 = c(8.5, 11, 18, 16, 9, 12, 6.5, 9, 20, 16)
t.test(x1, x2, paired = TRUE)
```

Solution 9. Un expérimentateur cherche à savoir si en moyenne la concentration d'α-lactalbumine dans le collostrum de la vache est significativement différente de la concentration d'α-lactalbumine dans le lait de la vache. Il examine 12 vaches et pour chacune d'elles, il mesure en mg/mL la concentration d'α-lactalbumine. Les résultats sont :

Collostrom	Lait
1.50	1.21
2.39	2.18
1.82	1.43
2.45	2.24
2.45	2.34
2.26	2.17
1.95	1.82
1.66	1.36
2.33	2.15
2.08	1.80
2.00	1.54
1.78	1.73

La concentration d'α-lactalbumine dans le collostrum d'une vache peut être modélisée par une *var* X_1 et la concentration d'α-lactalbumine dans le lait d'une vache peut être modélisée par une *var* X_2. On suppose que $X_1 - X_2$ suit une loi normale.

Proposer un test statistique adapté au problème et conclure.

```
x1 = c(1.50, 2.39, 1.82, 2.45, 2.45, 2.26, 1.95, 1.66, 2.33, 2.08, 2.00, 1.78)
x2 = c(1.21, 2.18, 1.43, 2.24, 2.34, 2.17, 1.82, 1.36, 2.15, 1.80, 1.54, 1.73)
t.test(x1, x2, paired = TRUE)$p.value
```

Cela renvoie : `[1] 5.919004e-05`. Comme p-valeur < 0.001, la différence est hautement significative $\star\star\star$.

Solution 10. Un sondage a été réalisé auprès de plusieurs employés de l'entreprise Normandie Dream afin de connaître leur niveau de satisfaction vis-à-vis de leur travail. Les résultats sont :

Salaire annuel \ Niveau de satisfaction	faible	moyen	élevé
moins de 20 000 €	10	27	56
plus de 20 000 €	15	44	38

Peut-on affirmer, au risque 5%, que le niveau de satisfaction d'un employé dépend de son salaire annuel ?

```
A = matrix(c(10, 27, 56, 15, 44, 38), nrow = 2, byrow = TRUE)
chisq.test(A)$p.value
```

Cela renvoie : `[1] 0.01472248`. Comme p-valeur < 0.05, la réponse est oui.

Solution 11. On a demandé à 257 personnes choisies au hasard dans la rue d'indiquer la langue étrangère qu'il connait le mieux. Les résultats sont :

Langue \ Sexe	homme	femme
anglais	65	55
espagnol	33	21
allemand	15	27
aucune	17	24

Peut-on affirmer, au risque 5%, que les connaissances en langues étrangères dépendent du sexe ?

```
A = matrix(c(65, 55, 33, 21, 15, 27, 17, 24), nrow = 4, byrow = TRUE)
chisq.test(A)$p.value
```

Cela renvoie : `[1] 0.04419228`. Comme p-valeur < 0.05, la réponse est oui.

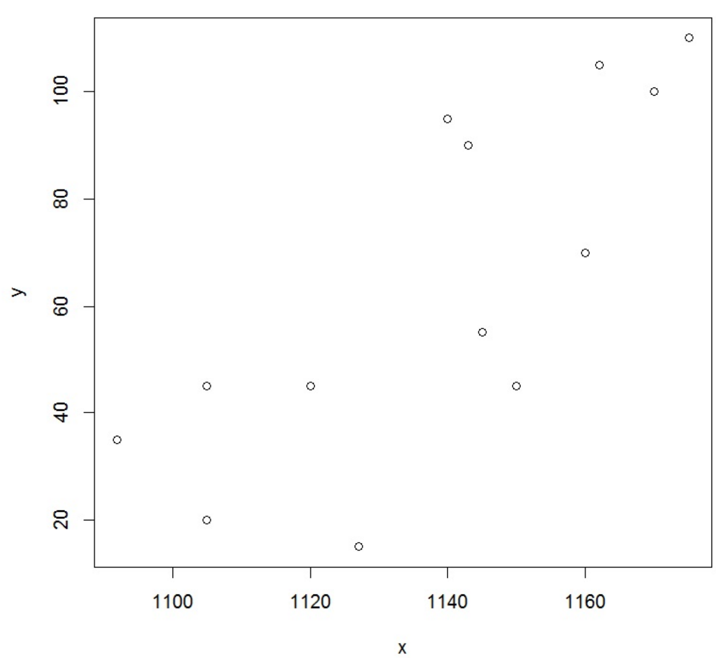
Solution 12. On étudie la vitesse de dissolution d'un type de streptomycine en poudre. La vitesse de dissolution est un caractère Y et la densité de cette streptomycine dans un lot est un caractère X. Sur un échantillon aléatoire de 13 lots, des expérimentateurs observent les valeurs $(x_i, y_i)_{i\in\{1,\ldots,13\}}$ de (X, Y) suivantes :

x_i	1140	1092	1127	1175	1162	1105	1160	1143	1170	1105	1150	1145	1120
y_i	95	35	15	110	105	20	70	90	100	45	45	55	45

1. Représenter graphiquement le nuage de points $(x_i, y_i)_{i\in\{1,\ldots,13\}}$. Est-il raisonnable de modéliser (X, Y) comme un vecteur de *var* suivant une loi normale bidimensionnelle ?

   ```
   x = c(1140, 1092, 1127, 1175, 1162, 1105, 1160, 1143, 1170, 1105, 1150, 1145, 1120)
   y = c(95, 35, 15, 110, 105, 20, 70, 90, 100, 45, 45, 55, 45)
   plot(x, y)
   ```

   Cela renvoie :

   La silhouette du nuage de points est de forme ellipsoïdale, on peut admettre l'hypothèse de normalité sur (X, Y). Il est donc raisonable de modéliser (X, Y) comme un vecteur de *var* suivant une loi normale bidimensionnelle.

2. Peut-on affirmer, au risque 5%, que Y et X sont liées ?

```
cor.test(x,y)$p.value
```

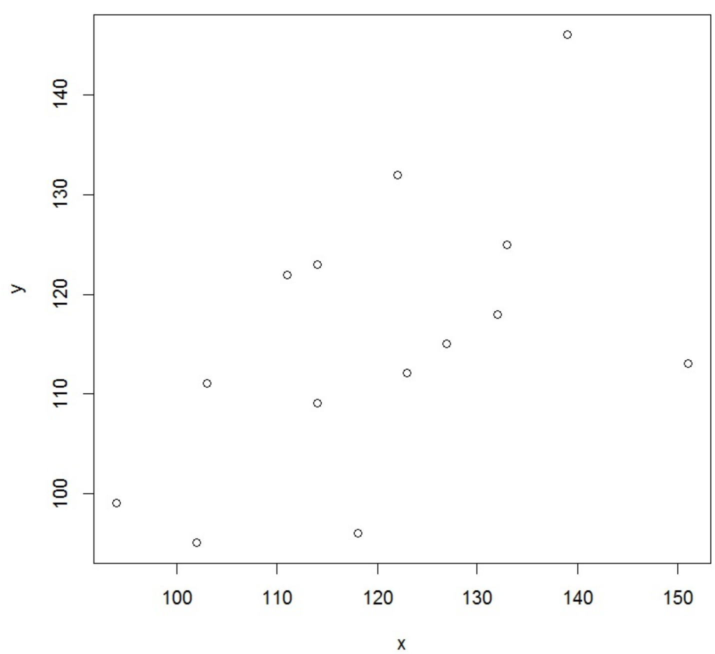
Cela renvoie : `[1] 0.00175879`. Comme p-valeur < 0.05, la réponse est oui. La dépendance est même très significative $\star\star$ car p-valeur $\in [0.001, 0.01[$.

Solution 13. Sur 14 familles composées d'une mère et d'une fille, on examine le QI de la mère et le QI de la fille. Les résultats sont :

Mère	123	132	118	114	102	139	133	94	111	127	122	103	114	151
Fille	112	118	96	123	95	146	125	99	122	115	132	111	109	113

Peut-on affirmer qu'il y a une liaison significative entre le QI de la mère et le QI de la fille ?

```
x = c(123, 132,  118, 114, 102, 139, 133, 94, 111, 127, 122, 103, 114, 151)
y = c(112, 118, 96, 123, 95, 146, 125, 99, 122, 115, 132, 111, 109, 113)
plot(x, y)
cor.test(x,y)$p.value
```



La silhouette du nuage de points est de forme ellipsoïdale, on peut admettre l'hypothèse de normalité sur (X, Y).

On a aussi : `[1] 0.04714025`. Comme p-valeur < 0.05, la réponse est oui.