Statistiques et Applications - TD2
Tests d'hypothèses

1 Exercice 1

Vous êtes un inspecteur de la répression des fraudes et vous avez pour mission de vérifier qu'un producteur de vin remplit honnêtement ses bouteilles, c'est-à-dire qu'il fait en sorte qu'il y ait au moins 75 cl de vin dans chaque bouteille vendue. Cependant, une machine remplit automatiquement les bouteilles et le volume de chaque bouteille est aléatoire. Vous voudrez alors savoir si le producteur a réglé sa machine de telle sorte qu'il y ait en moyenne au moins 75 cl dans les bouteilles vendues.

Pour mesurer le volume d'une bouteille mise en vente, il faut l'ouvrir et donc détruire la bouteille. Vous ne pouvez donc pas contrôler toutes les bouteilles, et vous devrez prendre une décision à partir d'un échantillon réduit de taille n.

On note Y_i la variable aléatoire égale au volume de vin dans la i^{e} bouteille de l'échantillon. On suppose que les variables Y_i sont indépendantes et suivent toute la même loi

$$Y_i \sim N(\mu, \sigma^2)$$

1. Télécharger depuis moodle le fichier `Volumes.csv` donnant pour chacune des bouteilles de l'échantillon son volume et enregistrez-le sur votre répertoire de travail.
2. Copier l'adresse de votre répertoire de travail. Pour la trouver, cliquer droit sur un fichier, propriétés, et copier l'adresse.
3. Démarrer Rstudio. Dans l'éditeur de texte, tapez votre première commande `setwd('AdresseDuRépertoire')`. Envoyer cette ligne de code dans la console en cliquant sur le bouton RUN.
4. Charger ce fichier `Volumes.csv` sur **R**, pour cela recopier dans l'éditeur de texte la commande `Vol=read.table('Volumes.csv', header=TRUE, sep=',', dec='.')`. Envoyez la ligne de code dans la console avec RUN. Une nouvelle variable apparait dans l'environnement de travail. Il s'agit de la variable `Vol` qui appartient à la classe `data.frame`. Vous pouvez vérifier que la table est correctement chargée avec les commandes `head(Vol)` et `tail(Vol)`.
5. Quelle est la classe de `Vol$bouteille` et de `Vol$volume` ? A quoi sert le symbol `$` ?
6. A l'aide des fonctions `mean` et `var`, donnez une estimation de la moyenne et de l'écart-type du volume des bouteilles. Pourquoi ne pouvez-vous pas prendre une décision quant à l'honnêteté du producteur à partir de la simple information de la moyenne de l'échantillon ?
7. On suppose tout d'abord que le producteur est honnête, et donc qu'il a réglé sa machine de telle sorte qu'il y ait en moyenne exactement 75 cl dans les bouteilles. Pour comparer graphiquement la loi théorique sous cette hypothèse, avec les volumes effectivement observés dans l'échantillon, on va représenter les données sous forme d'histogramme et l'on va rajouter la densité de la théorique par-dessus. Pour cela recopier et comprener les instructions suivantes

```
  hist(Vol$volume,freq=FALSE)
x=seq(72,78,0.01)
```

```
y=dnorm(x,mean=75,sd=sd(Vol$volume))
lines(x,y,col='red')
```

Qu'en pensez-vous ? Sous cette hypothèse, donnez la loi de $\overline{X}$ la moyenne de l'échantillon.

8. Proposer un test d'hypothèse pour conclure sur l'honnêteté du producteur.
   (a) Rappeler le modèle sur les données.
   (b) Donner l'hypothèse nulle et l'hypothèse alternative.
   (c) Donner la statistique de test ainsi que sa loi sous $\mathcal{H}_0$. Calculer la valeur de la statistique de test.
   (d) Donner la règle de décision associée à ce test, et conclure sur le test. Vous pourrez trouver les quantiles de la loi normale avec la fonction `qnorm`.
   (e) Calculer la probabilité critique, vous pourrez utiliser la fonction `pnorm()`.

   Un tel test peut être réalisé avec **R**, grâce à la commande `t.test(Vol$volume, alternative='less',mu` **R** réalise en fait un test de Student. Pourquoi peut-on aussi réaliser un test avec une statistique de test normale ? Concluez sur l'honnêteté du producteur.
9. Que représente le risque de première espèce α ? Quelle sera la conséquence d'une diminution du risque de première espèce ?
10. Vous voulez maintenant savoir quel est votre risque de passer à côté d'une fraude, c'est-à-dire de ne pas rejeter $\mathcal{H}_0$ alors que celle-ci est fausse.
    (a) Rappeler la règle de décision.
    (b) Soit x la valeur du volume moyen des bouteilles réglé par le producteur. On suppose maintenant que le producteur est fraudeur, et donc μ est inférieur à 75 cl. Calculer l'erreur de seconde espèce en fonction de μ.
    (c) Pour des valeurs de μ allant de 74 cl a 74.99 cl tracer l'erreur de seconde espèce en fonction de μ. Qu'en concluez-vous quant à vos chances de ne pas repérer une fraude ?

2 Exercice 2

On considère une classe de 20 élèves dans laquelle il y a 8 filles. On se demande si les filles sont sous-représentées dans cette classe. Soit X le nombre d'élèves dans la classe.

1. Donner la loi de X.
2. Enoncer les hypothèses nulle et alternative qui permettent d'attester que les filles ne sont pas sous-représentées.
3. Donner la loi de X sous $\mathcal{H}_0$.
4. Que fait la fonction `pbinom()` ? A l'aide de cette fonction, complétez le tableau suivant

x	0	1	2	$\cdots$	20
$\mathbf{P}_{\mathcal{H}_0}(X \leq x)$					

5. Utiliser le tableau ci-dessus pour définir la région de rejet du test au niveau 5% et donner la probabilité critique. Conclure.
6. Reprendre le test ci-dessus en approximant la distribution de $\bar{X}$ par une distribution normale. Quelle zone de rejet obtenez-vous ? Comparer les deux résultats.
7. La fonction `binom.test()` permet d'exécuter un test dont la loi sous $\mathcal{H}_0$ est une loi binomiale. Exécuter ce test et interpréter le résultat.