Université de Nice
Département de Mathématiques
NOM :
PRENOM :

L1MASS, année 2014-2015
Statistique
Date : 18 mars 2015 .
.

Feuille de la semaine 8 : Estimation par intervalle de confiance d'une proportion

1. Simuler un échantillon x de taille $n = 250$ de loi de Bernoulli de paramètres $p = 0.75$ (on pourra utiliser la commande `rbinom`). Tracer son histogramme et calculer sa moyenne et son écart type. On appelle *pchapeau* la proportion de 1 dans cet échantillon. Quelle valeur de *pchapeau* trouvez-vous ? Si vous faites une nouvelle simulation, trouverez-vous la même valeur ? Pourquoi n'est-elle pas égale à p ?

2. Répétez l'expérience une dizaine de fois et notez les *pchapeau* obtenus. En déduire un intervalle (approximatif) dans lequel *pchapeau* semble se situer le plus souvent. Expliquez-vous.

3. Recommencer les deux questions précédentes avec $n = 25000$ cette fois. Qu'observez-vous ?

4. Expliquer ce que produisent les instructions suivantes :

```
pchapeau=numeric(200)
for(i in 1:200) pchapeau[i]=sum(rbinom(n,1,p))/n
hist(pchapeau)
```

5. En déduire un intervalle dans lequel se trouve 90% des tirages (si l'on ne tient pas compte des 5% des valeurs les plus petites et des 5% des valeurs les plus grandes) lorsque $n = 250$ puis lorsque $n = 25000$. Comparer vos résultats avec les intervalles que vous avez proposez en 2 et 3.

6. Vérifier sur les deux exemples précédents (pour $n = 250$ et $n = 25000$) la formule théorique de l'intervalle de confiance symétrique au seuil 0.10 pour la probabilité p d'un caractère suivant une loi de Bernoulli.

7. Extraire du fichier Survey de la librairie MASS, la variable `W.Hnd` et calculer la proportion de gauchers dans l'échantillon. Expliquez le code que vous utilisez (à noter que `sum(y=="Left")` permet de dénombrer les gauchers du vecteur y).

8. Sachant que la proportion p de gauchers dans la population est égale à 0.12, simuler, comme ci-dessus, un échantillon de même taille que celui de Survey de loi de Bernoulli avec cette proportion et indiquer la valeur estimée de p par votre échantillon. Recommencer 100 fois et faites l'histogramme des valeurs obtenues. Au vu de votre histogramme, pensez-vous que l'échantillon qui a été utilisé pour Survey puisse provenir d'une population pour laquelle $p = 0.12$?