TP3_Gaudre_Boisson

Gaudré Boisson

16 mai 2019

0.

Obtention des données du fichier fiabilites.csv

1.

Générer et stocker dans une matrice 1000 échantillons de taille n = 30 à partir de la population de données initiale.

2.

Écrire une fonction permettant d'effectuer un test de nullité de la moyenne. Cette fonction aura en paramètre un échantillon de données, l'écart-type σ et le risque de première espèce α , et retournera 0 si H0 est rejetée et 1 sinon.

Nous avons :

- X suit une loi normale
- "$H0 = 0$" et "$H1 \neq 0$"
- $\sigma^2 = 60^2$

Donc on accepte $H0$ si $\frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}} \in [-U_{1-\frac{\alpha]}{2}}, U_{1-\frac{\alpha]}{2}}]$

3.

Parmi les 1000 échantillons générés, combien de fois peut-on affirmer au risque de $\alpha = 5\%$ que l'indemnisation moyenne est différente de 650 euros.

Nous avons juste à appliquer la fonction créée à la question 2 :

```
## [1] 43
```

Dans les échantillons que nous avons générer, nous pouvons affirmer dans 43 cas sur 1000 que la moyenne est différente de 650 euros au risque $\alpha = 5\%$.

4.

Retrouver la formule explicite de la puissance de ce test. Écrire une fonction qui à partir de n, σ, α et de $\delta = \mu_1 - \mu_0$ (où μ_1 est la "vrai" valeur de μ sous $H1$) calcule la puissance du test.

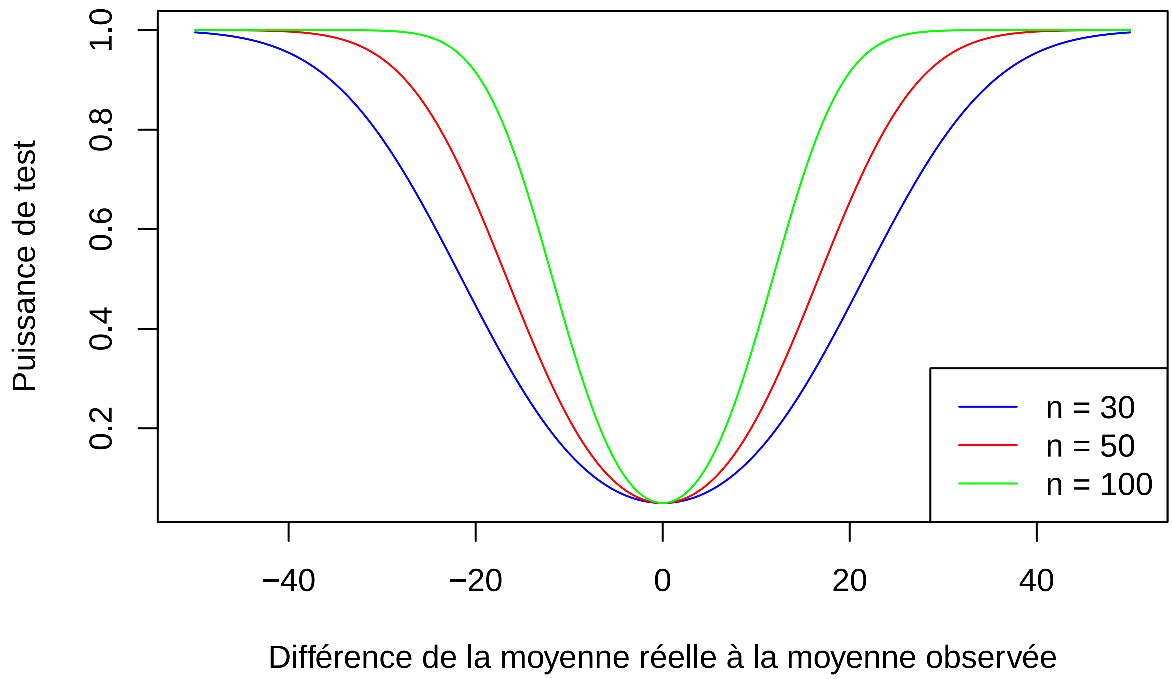
Formule de la puissance de test : $1 - P_m(S \notin W) = 1 - P_m(\mu_1 - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \leq \overline{X} \leq \mu_1 + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}})$
Nous obtenons donc : $1 - P_m(S \notin W) = 1 - [F_{N(0,1)}(u_{1-\frac{\alpha}{2}} + \frac{\mu_1-\mu_0}{\frac{\sigma}{\sqrt{n}}}) - F_{N(0,1)}(-u_{1-\frac{\alpha}{2}} + \frac{\mu_1-\mu_0}{\frac{\sigma}{\sqrt{n}}})]$

```
## [1] 0.2790905
```

Pour un essais de moyenne selon le premier échantillon : la puissance de test atteint 27,9%. Nous ne pouvons donc pas rejeter H_0 avec un risque de deuxième espèce $\beta = 5\%$.

5.

En fixant $\alpha = 5\%$: Tracer la puissance du test pour $n \in \{30; 50; 100\}$ en fonctio de $\delta in[-50; 50]$



On remarque que plus le nombre d'observation est élevé, plus la zone d'incertitude (ou de transition) est faible. Donc si la moyenne observée est écartée de la moyenne réelle moins on a de risque de se tromper si on rejette H_0. Et inversement, si l'écart est faible on a plus le risque de se tromper si on rejette H_0.

6.

On suppose maintenant la variance inconnue.Refaire les questions 2–5 :
- En utilisant les formules explicites comme précédemment,
- En utilisant les fonctions t.test et power.t.test.

Formules Explicites

2.

La variance corrigée est un estimateur sans biais de la variance. Donc, on utilise la variance corrigée à défaut d'utiliser la variance réelle comme la variance réelle est inconnue.

3.

```
## [1] 43
```

```
## [1] 43
```

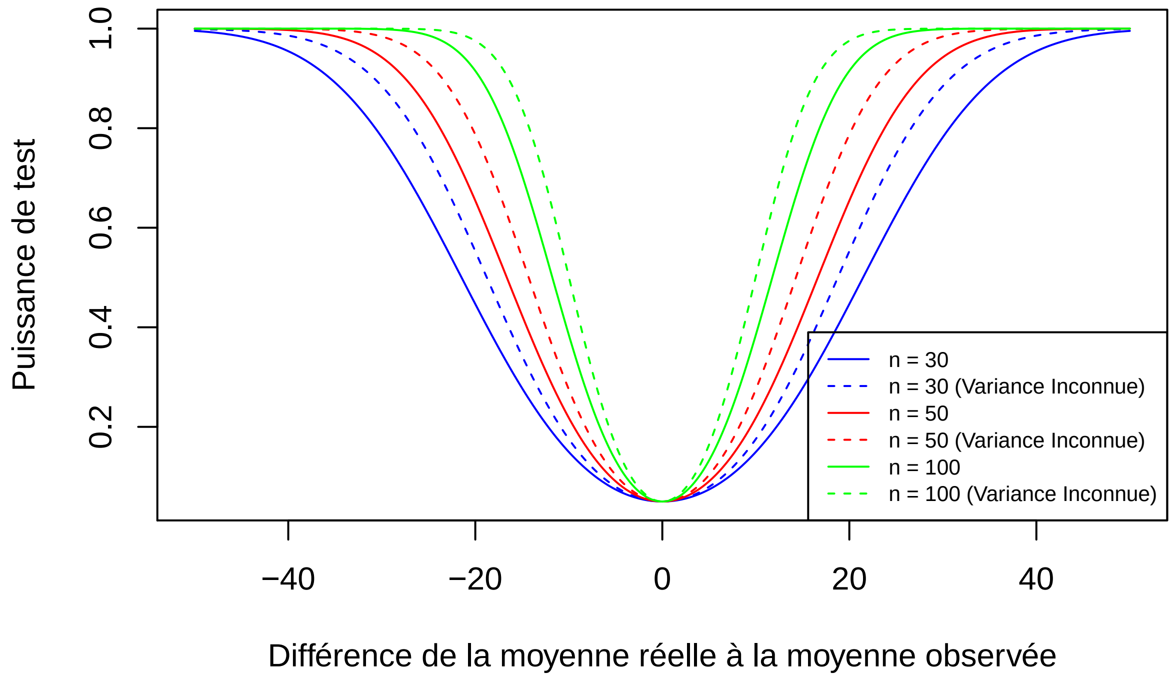
Nous obtenons plus de rejets de H_0 lorsque la variance est connue pour les échantillons que nous avons générés.

4.

```
## [1] 0.2790905
```

```
## [1] 0.3441273
```

Pour le même échantillon, la puissance de test est plus élevée si l'on ne connais pas la variance. Cela veut dire que l'on est plus sûr de rejetter à raison H_O. La question 3 nous donne donc plus de rejets de H_0 lorsque la variance est inconnue, parce qu'on atteint plus rapidement un risque faible de se tromper l'orsque l'on rejette H_0.



5.

Le graphique nous confirme la question 4. On peut voir qu'à une même taille d'échantillon donnée, on est plus sûr de rejetter H_0 à raison lorsque nous ne connaissons pas la variance.

Cependant, le test à la variance inconnue se base sur une estimation de la variance via un petit échantillon. Donc si la variance du petit échantillon est inférieure à la variance réelle : le test sera de meilleure puissance, et inversement si la variance estimée est plus importante que la variance réelle. On pense toutefois, comme l'estimateur de la variance n'est pas biaisé, que le test aura une puissance de test qui se situera autour de la puissance du test à la variance connue.

Fonction t.test et power.t.test

2.

La fonction t.test nous donne la statistique, ainsi que les bornes de l'intervalle de confiance.

3.

```
## [1] 28
```

```
## [1] 43
```

```
## [1] 43
```

Nous obtenons moins de rejets avec la fonction t.test. La différence pourrait s'expliquer par une méthode différente pour calculer la statistique ou les bornes de l'intervalle de confiance.

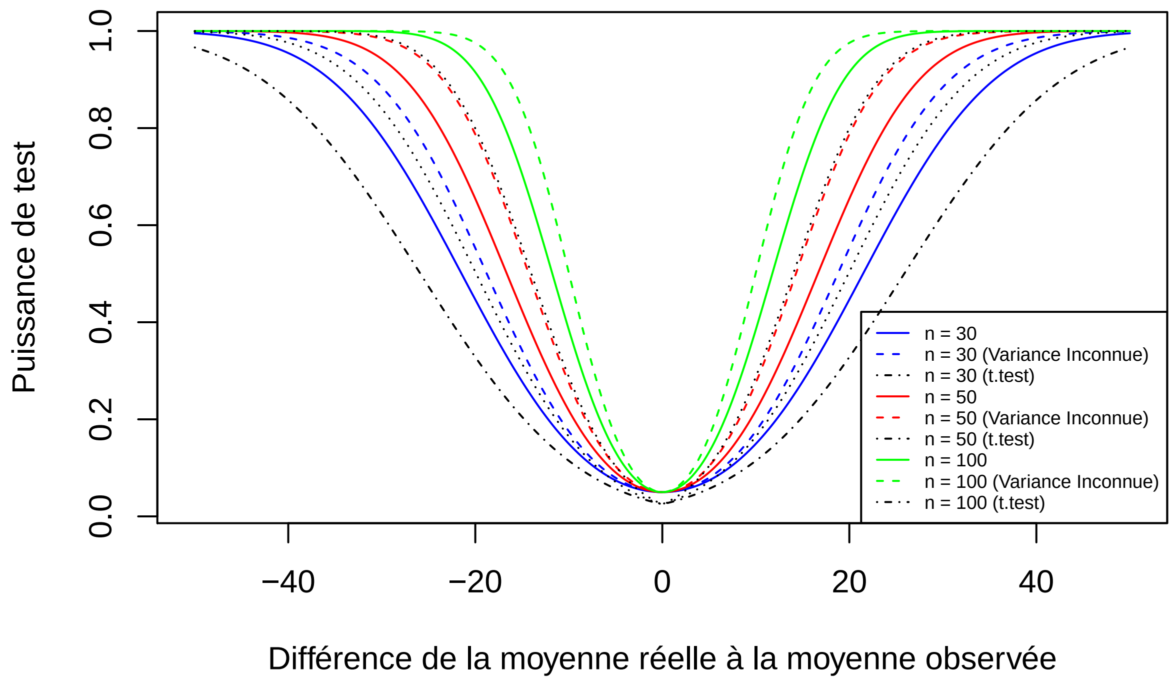
4.

```
## [1] 0.2790905
```

```
## [1] 0.3441273
```

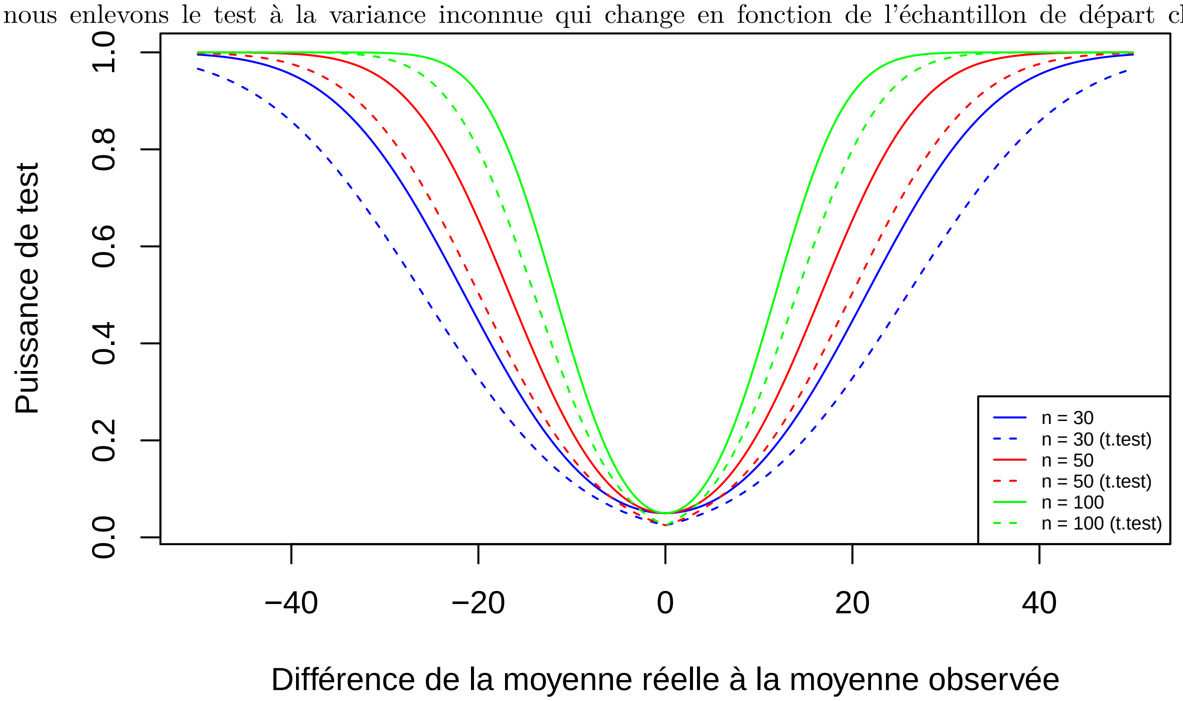
```
## [1] 0.2060017
```

On observe une puissance de test inférieure au 2 autres tests. Le test est donc moins sûr de rejetter à raison H_0.



5.

Ceci est une comparaison des 3 tests sur le risque de deuxième espèce. Mais le graphe trop dense, nous enlevons le test à la variance inconnue qui change en fonction de l'échantillon de départ choisi :



Si nous comparons les valeurs obtenues de la fonction t.test avec les valeurs qui sont issues du test avec la variance inconnue. Nous voyons que la probabilité de rejeter H_0 à raison est forcément moins bonne avec la fonction t.test qu'avec le test à la variance connue. Ce qui nous donnerais plus l'envie d'utiliser le test avec la variance inconnue qui possède une meilleur puissance de test que t.test (alors que tous les 2 utilisent la même estimation de la variance).

7.

Calculer le nombre d'observations nécessaires pour que le risque de seconde espèce ne dépasse pas $\beta = 5\%$ pour $\delta = 10$.

Nous savons que $n \geq \frac{sigma^2}{delta^2}(\mu_{1-\frac{\alpha}{2}} + \mu_{1-\beta})^2$ pour une loi normale et un test bilatéral centré sur la moyenne.

Avec la fonction power.t.test et la variance inconnue (utilisation d'une estimation de la variance avec un premier petit échantillon ou $n = 30$):

```
## [1] 657.9311
```

Avec la variance connue:

```
## [1] 467.8096
```

Avec la variance inconnue (utilisation d'une estimation de la variance avec un premier petit échantillon ou n = 30):

```
## [1] 328.4846
```

Nous observons que le nombre d'observation nécessaire diffère selon la fonction utilisée. Le nombre d'observation nécessaire suit bien les résultats que nous avons obtenu aux questions précédentes. C'est à dire que la fonction t.test est moins efficace que le test avec la variance connue qui est aussi moins efficace que le test à la variance inconnue (pour notre estimation de la variance). Ce qui rend le nombre d'observations nécessaire plus grand pour atteindre un même niveau et une même puissance de test pour le test t.test et le test à la variance connue.

Nous pouvons aussi vérifier nos résultats :

```
## [1] 0.9500195
```

```
## [1] 0.9500756
```

```
## [1] 0.9501975
```

Nous atteignons bien $beta = 5\%$ dans chacun des test. Les estimations de tailles d'échantillons sont donc correctes. (En prenant en compte que si l'on utilise un estimateur de la variance, ce n'est qu'une valeur approchée du n optimal qui est trouvé)