ZUO_YOU_TPSTAT4

You ZUO

2019/4/19

Test de Student

<-1->

<-1.a->

α ici est la probabilité de l'erreur de première espèce, qui est la probabilité de rejeter à tort H_0 (ou encore de choisir H_1 alors que H_0 est vraie).

<-1.b->

D'après le cours, pour σ inconnu on a statistique de test:

$$T(x_1, ..., x_n) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim t(n-1)$$

et donc pour la zone de rejet:

$$W = \{(x_1, ..., x_n) \ ; \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} > t_{1-\alpha}\}$$

<-1.c->

```
decision <- function(S_n,alpha,mu0,mu1){
  mu_em <- mean(S_n)
  var_em <- var(S_n)
  borne <- (var_em*qt(1-alpha,length(S_n)-1))/sqrt(length(S_n))+mu0
  if(mu_em>borne) mu1 else mu0
}
decision(S_n,0.05,1,1.5)
```

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

<-2->

<-2.a->

```
## [1] 0.96
```

Le resultat veut dire qu'on a accepté H_0 finalement 96 fois pour le test de ces 100 groupes d'échantillons. Et on peut remarquer que la frequence est assez proche que $1-\alpha$. En fait, $\delta(S_n, \alpha, \mu_0, \mu_1)$ suit une loi $B(1-\alpha)$.

<-2.b->

```
##                  [,1]     [,2]     [,3]     [,4]
## alpha     0.2000000 0.100000 0.050000 0.010000
## t_1_alpha 0.8609506 1.327728 1.729133 2.539483
```

On peut voir que le K_a(ici $t_{1-\alpha}$) augmente en fonction de α selon sa décroissance, la zone de rejet devient donc plus petite quand α est plus petit. C'est à dire la probabilité de rejetter H_0 à tort est plus petit.

<-2.c->

```
## [1] 0.88 0.92 0.96 0.99
```

<-3->

<-3.a->

```
## [1] 0.74
```

Le resultat veut dire que la frequence d'accepter H_0 est 0.7, c'est égale à $P(choisir\ H_0|H_1\ est\ vraie)$.

<-3.b->

La puissance d'un test, notée β, est la probabilité de rejeter H_0 lorsque H_1 est vraie ($P(choisir\ H_0|H_1\ est\ vraie)$).
On a statistique de test: $\Lambda_n = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n}$ - région critique(de rejet): $W = \{\Lambda_n > K_\alpha\}$.
Il faut d'abord déterminer K_α:

$$\mathbb{P}_{H_0}(W) = \mathbb{P}_{H_0}(\Lambda_n > K_\alpha) = 1 - F_{T_{n-1}}(K_\alpha) = \alpha$$

on a donc $K_\alpha = F^{-1}_{T_{n-1}}(1-\alpha)$, pour la puissance:

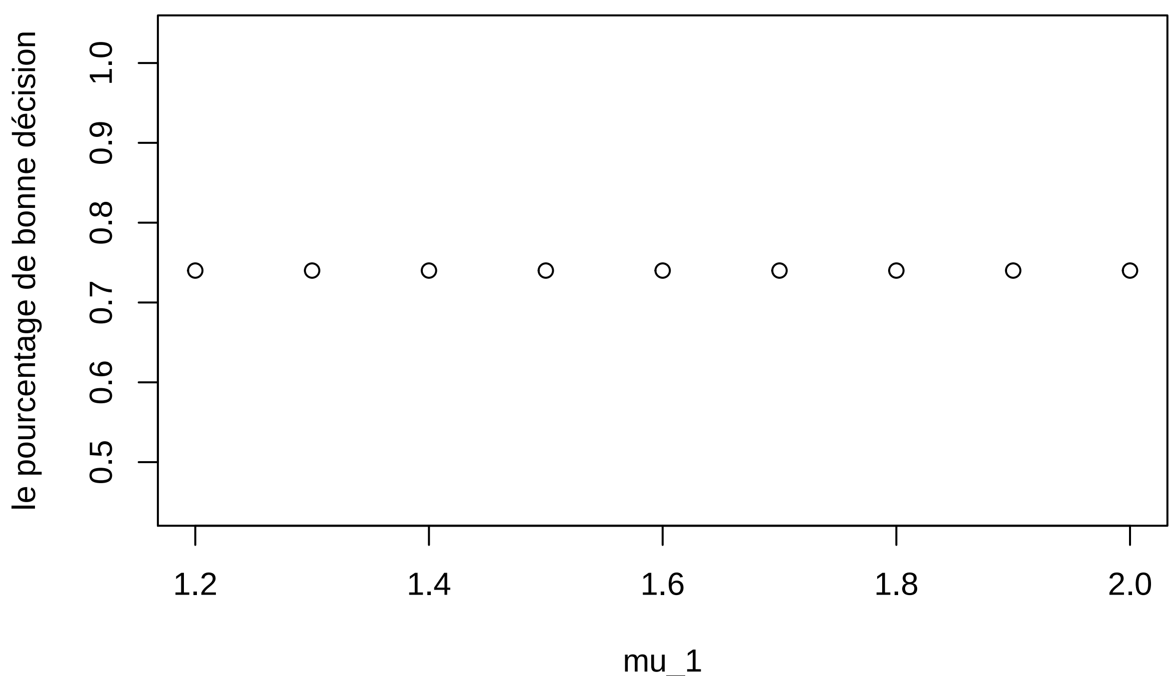
$\beta = \mathbb{P}_{H_1}(W) = \mathbb{P}_{H_1}(\Lambda_n > K_\alpha)$

$= \mathbb{P}_{H_1}(\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n} > F^{-1}_{T_{n-1}}(1-\alpha))$

$= \mathbb{P}_{H_1}(\frac{\sqrt{n}(\bar{X}_n - \mu_1)}{S_n} > \frac{\sqrt{n}(\mu_0 - \mu_1)}{S_n} + F^{-1}_{T_{n-1}}(1-\alpha))$

$= 1 - F_{T_{n-1}}(\frac{\sqrt{n}(\mu_0 - \mu_1)}{S_n} + F^{-1}_{T_{n-1}}(1-\alpha))$

<-3.c->



<-4->

<-4.a->

```
##
##  One Sample t-test
##
## data:  S_n
## t = 0.86619, df = 19, p-value = 0.3972
## alternative hypothesis: true mean is not equal to 1
## 95 percent confidence interval:
##  0.643341 1.860285
## sample estimates:
## mean of x
##  1.251813
```

Ici, la valuer t correspond à la valeur de statistique de test de Student:

$$t = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n}$$

df correspond à le degré de liberté $n-1$.

<-4.b->

Dans question 3.a on a obtenué que $K_\alpha = F^{-1}_{T_{n-1}}(1-\alpha)$ donc on a:

$$W = \{t > K_\alpha\} = \{t > F^{-1}_{T_{n-1}}(1-\alpha)\} = \{F_{T_{n-1}}(t) > 1-\alpha\} = \{1 - F_{T_{n-1}}(t) < \alpha\}$$

finalement notre zone de rejet est commme $W = \{p < \alpha\}$, c'est à dire on quand $p < \alpha$ on rejette l'hypothèse H_0.

<-4.c->

```
## [1]  76  89  95 100
```

On a trouvé que selon la décroissance de la valeur α, on a plus de probabilité d'accepter l'hypothèse H_0, c'est à dire moins de probabilité que $p \geq \alpha$.

<-4.d->

L'intervalle de confiance est une fonction $\mathcal{C}: \ (x_1, ..., x_n) \mapsto \mathcal{C}(x_1, ..., x_n)$ où $\mathcal{C}(x_1, ..., x_n)$ est un sous-ensemble de $\Theta \subset R^p$ et satisfaisant la propriété: pour tout $\theta \in \Theta$,

$$inf_\theta P_\theta(\theta \in \mathcal{C}(X_1, ..., X_n)) \geq 1 - \alpha$$

ici $\mathcal{C}(X_1, ..., X_n)$ est un ensemble qui est aléatoire (car sa définition depend des observations), et telle que quelque soit θ, il y ait une probabilité au moins supérieur à $1-\alpha$ que le vrai paramètre θ soit dedans.

```
## [1] 95
```

Pour les échantillons $S_n^1, ..., S_n^N$, il y a totalement presque 100 de cas que 1 est dans l'intervalle de confiance, ce qui est normal.