Département
Licence

MOSE2014 Probabilités et statistiques
Mathématiques TP machine 3
Ph. Thieullen

TP machine III
(Statistique inférentielle : intervalles de confiance)

Le TP en entier est à rendre. N'oubliez pas de commencer votre script par `# TP3.R` *et les noms de chaque étudiant :* `# noms, prenoms` *formant le binôme ou le trinôme.*

Dans toute la suite, on considère le poids (fictif) de 30 nourrissons dans une maternité A et de 50 autres nourrissons dans une maternité B :

```
poidsA <- c(rnorm(15,3050,450),rnorm(15,3400,250))
poidsB <- c(runif(20,2600,4500),runif(30,3150,3650))
```

1 Représentation graphique

– Afficher une figure contenant les deux histogrammes en indiquant bien en sur-titre les deux maternités. Utiliser `hist` et son aide en ligne. On devra trouver un figure semblable à la figure 1.

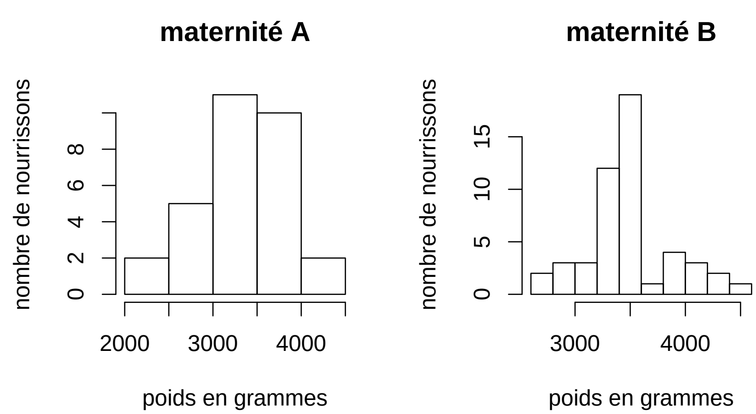

Figure 1 – Histogramme des naissances

(On laissera `R` choisir le nombre de colonnes).

– Afficher les quantiles de chaque distribution : on pourra essayer les deux méthodes

```
cat("quantiles de A : \n")
print(quantile(poidsA))
print(summary(poidsA))
```

– Remarque : pour arrêter le défilement des figures à l'affichage, on insérera en début de script la commande `devAskNewPage(ask = TRUE)`. Pour stopper l'exécution du script à tout moment, il peut être util d'utiliser `scan()`. La reprise de l'exécution du programme se fait par un retour à la ligne (sur PC) ou `OK` (sur Unix).

– Afficher les boxplots des 2 distributions sur un même graphique. On utilisera la commande `boxplot` et ses options `range` et `names`. Il est impératif d'aller voir l'aide en ligne de `boxplot`. Indiquer bien le nom de chaque boxplot A ou B ; indiquer aussi l'option choisie sur `range`. On devra trouver une figure semblale à la figure 2.

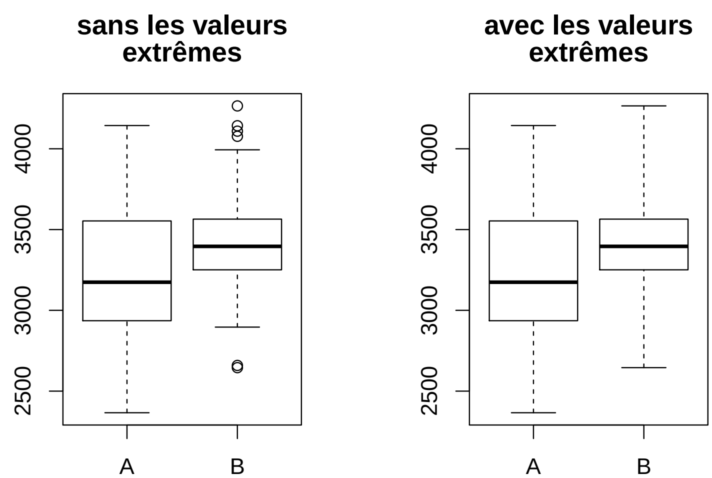

FIGURE 2 – Boxplot des naissances dans deux maternités

2 Intervalle de confiance

– En utilisant les formules donnant l'intervalle de confiance de la moyenne μ_A des poids de la maternité A, écrire un code de programme donnant les extrémités $[\mu_{min}, \mu_{max}]$ au seuil 2%. On rappelle la formule

$$\mu_{min} = \bar{x} - t_\alpha \frac{s_{n-1}}{\sqrt{n}} \leq \mu_A \leq \bar{x} + t_\alpha \frac{s_{n-1}}{\sqrt{n}} = \mu_{max}.$$

où $t_\alpha = q_{1-\alpha/2}$ et q_β est le quantile d'ordre β de la loi de Student à $n-1$ degrés de liberté. On utilisera pour cela les commandes `mean` (moyenne), `sd` (écart-type), `qt` (quantile de la loi de Student). On complétera les `...` du code suivant

```
muA <- mean(poidsA)
sigmaA <- sd(...)
alpha <- 0.02
nA <- length(...)
t_alpha <- qt(...,...)
mu_min <- ... # formule du cours
mu_max <- ... # formule du cours
cat("IC de muA : formule du cours\n")
cat("mu_min = ", mu_min, "\n")
cat("mu_max = ", mu_max, "\n")
```

`R` fait bien sûr tous ces calculs. Comparer les résultats précédents avec ceux du code suivant. Lire d'abord l'aide de `t.test`. Le résultat est une structure de type `list` ; les variables d'une liste sont accessibles avec l'oprérateur `$`.

```
icA <- t.test(poidsA,
      conf.level=1-alpha, alternative="two.sided")
mu_min_bis <- icA$conf.int[1]
mu_max_bis <- icA$conf.int[2]
cat("IC de muA : t.test \n")
cat("mu_min_bis = ", mu_min_bis, "\n")
cat("mu_max_bis = ", mu_max_bis, "\n")
```

– Déterminer un intervalle de confiance de $\mu_B - \mu_A$: on utilisera à nouveau `t.test` avec l'option `var.equal=TRUE` (donnant les formules du cours) puis avec `var.equal=FALSE` (non développé dans le cours). Formuler par exemple les résultats sous la forme

```
icAB <- t.test(...)
mu_min_AB <- ... # utilise icAB
mu_max_AB <- ... # utilise icAB
cat("IC de muB-muA : var.equal=TRUE \n")
cat("mu_min_AB = ", mu_min_AB, "\n")
cat("mu_max_AB = ", mu_max_AB, "\n")
```

S'agit-il d'un intervalle de confiance dans le cas apparié ou dans le cas indépendant ? Il est indispensable d'aller regarder l'aide en ligne de `t.test` et de son option `paired`. Répondez par `cat("interval dans le cas...\n")`.

– Pour les intervalles de confiance d'une proportion on utilise `binom.test`. Dans un échantillon de 197 pommes, on constate que 19 d'entre elles sont abîmées. Déterminer un intervalle de confiance de la proportion de pommes abîmées. On utilisera d'abord les formules du cours

$$p_{min} = \hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = p_{max}$$

```
n_tot <- 197
n_obs <- 19
alpha <- 0.05
z_alpha <- qnorm(...)
p_min <- ... # remplacer ... par les formule du cours
p_max <- ... # remplacer ... par les formule du cours
cat("IC de p : formule du cours \n")
cat("p_min = ", p_min, "\n")
cat("p_max = ", p_max, "\n")
```

Recommencer avec `binom.test` et comparer les résultats

```
icp <- binom.test(...)
p_min_binom <- ... # utilise icp
p_max_binom <- ... # utilise icp
cat("IC de p : binom.test \n")
cat("p_min_binom = ", p_min_binom, "\n")
cat("p_max_binom = ", p_max_binom, "\n")
```

Comparer ces derniers résultats avec ceux obtenus avec les formules du cours.