ISFA- M1 — **Semestre Automne 2017/2018**

STATISTIQUE INFÉRENTIELLE — lerouvillois@math.univ-lyon1.fr

Corrigé

Exercice 4, feuille 4 bis. (Intervalles de confiance pour la loi normale)

On s'interroge sur la comparaison des tailles moyennes des garçons et des filles de 6 ans dans une population, pour cela on a pris comme échantillon, jugé représentatif de cette tranche d'âge, une classe d'école primaire (niveau CP en France), et on a observé :

16 garçons : moyenne 126.5 cm, écart-type estimé 12.9 cm

15 filles : moyenne 136.9 cm, écart-type estimé 11.9 cm.

On admet que la distribution des tailles dans chacune des sous-populations (garçons, filles) suit une loi gaussienne.

1. Donner des intervalles de confiance à 95% pour les tailles moyennes des garçons et des filles.
2. Donner un intervalle de confiance à 95% pour l'écart-type de la taille des garçons. Même question pour les filles.

3) Donner un intervalle de confiance à 95% pour le rapport des deux écart-types. Peut-on conclure que les variances des deux populations sont différentes ? (On donne, pour la loi de Fisher-Snedecor à (15,14) degrés de libertés, les quantiles $f_{0.975} = 2.95$ et $f_{0.025} = \frac{1}{f_{0.975}} = 0.339$).

Corrigé : Soit $X = (X_1, ..., X_m)$ un m-échantillon de loi normale $\mathcal{N}(\mu, \sigma^2)$ représentant la taille des garçons mesurés et $X' = (X'_1, ..., X'_n)$ m-échantillon de loi normale $\mathcal{N}(\mu', \sigma'^2)$ représentant la tailles des filles mesurées avec X et X' indépendants. (Ici, $m = 16$ et $n = 15$).
Soit $\bar{X}_m = \frac{\sum_{i=1}^m X_i}{m}$ et $\bar{X}'_n = \frac{\sum_{i=1}^n X'_i}{n}$, les moyennes empiriques associées.
Soit $S_m^2 = \frac{\sum_{i=1}^m (X_i - \bar{X}_m)^2}{m-1}$ et ${S'_n}^2 = \frac{\sum_{i=1}^n (X'_i - \bar{X}'_n)^2}{n-1}$, les variances estimées associés.

1. **Comme σ est inconnu, pour déterminer un intervalle de confiance pour μ, on va utiliser une statistique de Student.** (Si σ était connu, on utiliserait directement la loi normale).

   Propriété du cours sur les lois normales : *Les statistiques* $\frac{\sqrt{m}(\bar{X}_m - \mu)}{\sigma}$ *et* $\frac{(m-1)S_m^2}{\sigma^2}$ *sont indépendantes et de lois respectives* $\mathcal{N}(0,1)$ *et* $\chi^2(m-1)$.

   On en déduit que l'estimateur :

$$T_m = \frac{\frac{\sqrt{m}(\bar{X}_m - \mu)}{\sigma}}{\sqrt{\frac{(m-1)S_m^2}{\sigma^2}/(m-1)}} = \frac{\sqrt{m}(\bar{X}_m - \mu)}{S_m} \sim \mathcal{T}(m-1)$$

suit une loi de student de paramètre $(m-1)$.

<u>Remarque</u> : *Lorsque σ est connue, on utilise $\frac{\sqrt{m}(\bar{X}_m-\mu)}{\sigma} \sim \mathcal{N}(0,1)$. Si σ est inconnu, "'on remplace"' σ par son écart-type estimé S_m pour obtenir la statistique $T_m \sim \mathcal{T}(m-1)$.*

On a donc $\mathbb{P}(|T_m| \leq t^{m-1}_{1-\alpha/2}) = 1-\alpha$. D'où l'intervalle de confiance de niveau $1-\alpha$ pour μ :

$$\bar{X}_m - \frac{t^{m-1}_{1-\alpha/2} S_m}{\sqrt{m}} \leq \mu \leq \bar{X}_m + \frac{t^{m-1}_{1-\alpha/2} S_m}{\sqrt{m}}.$$

De même, on obtient l'intervalle de confiance pour μ' :

$$\bar{X}'_n - \frac{t^{n-1}_{1-\alpha/2} S'_n}{\sqrt{n}} \leq \mu' \leq \bar{X}'_n + \frac{t^{n-1}_{1-\alpha/2} S'_n}{\sqrt{n}}.$$

<u>Application Numérique</u> : On lit sur la table de statistique pour $\alpha = 0.05$: $t^{15}_{0.975} = 2.1314$ et $t^{14}_{0.975} = 2.1448$ d'où les intervalles de confiance de niveau 95% :

$$\mu \in [119.63, 133.37],\ \mu' \in [130.31, 143.49].$$

2. **Pour déterminer un intervalle de confiance pour σ, on utilise l'estimateur S_m^2 et une loi du χ^2.** On sait que $\frac{(m-1)S_m^2}{\sigma^2} \sim \chi^2(m-1)$ d'où :

$$\mathbb{P}\Big(k^{m-1}_{\alpha/2} \leq \frac{(m-1)S_m^2}{\sigma^2} \leq k^{m-1}_{1-\alpha/2}\Big) = 1-\alpha.$$

En isolant σ, on obtient l'intervalle de confiance de niveau $1-\alpha$:

$$S_m\sqrt{\frac{m-1}{k^{m-1}_{1-\alpha/2}}} \leq \sigma \leq S_m\sqrt{\frac{m-1}{k^{m-1}_{\alpha/2}}},$$

et de même, on obtint l'intervalle de confiance pour σ' :

$$S'_n\sqrt{\frac{n-1}{k^{n-1}_{1-\alpha/2}}} \leq \sigma' \leq S'_n\sqrt{\frac{n-1}{k^{n-1}_{\alpha/2}}}.$$

<u>Application Numérique</u> : On lit sur la table de statistique $k^{15}_{0.025} = 6.2621$, $k^{15}_{0.975} = 27.4884$, $k^{14}_{0.025} = 5.6287$ et $k^{14}_{0.975} = 26.1186$ d'où les intervalles de confiance de niveau 95% :

$$\sigma \in [9.53, 19.97],\ \sigma' \in [8.712, 18.77].$$

3. **Pour déterminer un intervalle de confiance de σ'/σ, on utilise l'estimateur S'_n/S_m et une loi de Fisher-Snedecor.** On sait que $\frac{(m-1)S_m^2}{\sigma^2}$ et $\frac{(n-1){S'_n}^2}{\sigma'^2}$ sont indépendantes et de lois respectives $\chi^2(m-1)$ et $\chi^2(n-1)$. On en déduit que la statistique :

$$F_{m,n} = \frac{\frac{(m-1)S_m^2}{\sigma^2}/(m-1)}{\frac{(n-1){S'_n}^2}{\sigma'^2}/(n-1)} = \frac{{\sigma'}^2 S_m^2}{\sigma^2 {S'_n}^2} \sim \mathcal{F}(m-1, n-1),$$

suit une loi de Fisher-Snedecor à $(m-1, n-1)$ degrés de libertés.
Donc $\mathbb{P}(F_{m,n} \in [f_{\alpha/2}^{m-1,n-1}, f_{1-\alpha/2}^{m-1,n-1}]) = 1-\alpha$.
D'où l'intervalle de confiance pour σ'/σ de niveau $1-\alpha$:

$$\frac{S_n'}{S_m}\sqrt{f_{\alpha/2}^{m-1,n-1}} \leq \sigma'/\sigma \leq \frac{S_n'}{S_m}\sqrt{f_{1-\alpha/2}^{m-1,n-1}}.$$

Application Numérique : On nous donne $f_{0.025}^{15,14} = 0.339$ et $f_{0.975}^{15,14} = 2.95$ d'où l'intervalle de confiance de niveau 95% :

$$\sigma'/\sigma \in [0.537, 1.58].$$

Conclusion : la valeur $\sigma'/\sigma = 1$ appartient à l'intervalle de confiance de niveau 95% ci-dessus donc on ne peut pas rejeter l'hypothèse $\sigma' = \sigma$.