# Corrigé Exam Stats



## QCM

Q1 – $\boxed{C}$

Q2 – $\boxed{A}$

Q3 – $\boxed{D}$

## Exo 1

1° I.D.C. pour une moyenne avec $\sigma = 10$ connu et $n = 100$ $\Rightarrow$ loi normale.

à 5%, $I.D.C = \bar{x} \pm 1{,}96 \frac{\sigma}{\sqrt{n}} = 45 \pm 1{,}96 \frac{10}{\sqrt{100}}$

$I.D.C \simeq [43{,}04 \ ; \ 46{,}96]$.

2° $\sigma$ inconnu, $s = 10$

loi de STUDENT à $99 = 100 - 1$ d.d.l

Confondue avec $\mathcal{N}(0,1)$.

Donc l'I.D.C. ne change pas.

3° $\alpha = 1\%$ $\Rightarrow$ $I.D.C = \bar{x} \pm 2{,}58 \frac{\sigma}{\sqrt{n}}$

l'intervalle est plus large

4° On diminue $n$ de moitié $\Rightarrow$ on multiplie l'amplitude par $\sqrt{2}$.

**Exo 2**

1er échantillon : $X_1, X_2, \dots, X_{n_1}$, $X_i \sim \mathcal{B}(p_1)$

2e éch. $Y_1, Y_2, \dots, Y_{n_2}$, $Y_i \sim \mathcal{B}(p_2)$.

$p_1 = p_2 = p$.

1° $F_1 = \dfrac{\sum X_i}{n_1}$ $F_2 = \dfrac{\sum Y_i}{n_2}$ $F_3 = \dfrac{F_1 + F_2}{2}$

Les 2 premiers estimateurs sont sans biais, car moyennes empiriques.

Le troisième également car : $E(F_3) = \dfrac{E(F_1) + E(F_2)}{2} = p$.

2° Pour choisir le meilleur, il faut comparer les variances.

$Var(F_1) = \dfrac{p(1-p)}{n_1}$ $Var(F_2) = \dfrac{p(1-p)}{n_2}$

$$Var(F_3) = \frac{Var(F_1) + Var(F_2)}{2^2} = \frac{p(1-p)}{4}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

Supposons, pour simplifier, $n_1 < n_2$.

Alors : $Var(F_2) < Var(F_1)$.

$$Var(F_3) - Var(F_2) = \frac{p(1-p)}{4}\left[\frac{n_1 + n_2}{n_1 n_2} - \frac{1}{n_2}\right] = \frac{p(1-p)}{4}\left[\frac{n_2 - 3n_1}{n_1 n_2}\right]$$

1er Cas : $n_1 < n_2 < 3n_1 \Longrightarrow Var(F_3) < Var(F_2)$

Il faut choisir $F_3$.

2e Cas : $n_1 < 3n_1 < n_2 \Longrightarrow Var(F_3) > Var(F_2)$

Il faut choisir $F_2$.

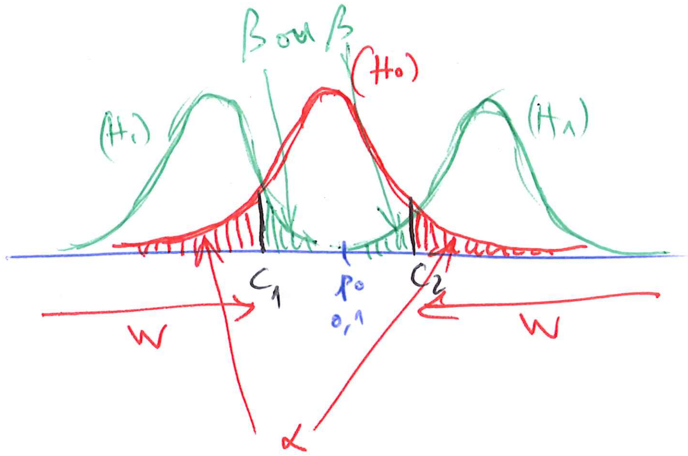
## Exo 3

1° $(H_0)$ $p = 0,1 = p_0$

$(H_1)$ $p = p_1 \neq 0,1$.

2° $n = 50 > 30$, fréquence empirique : $F_n \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$

3°




4°

$$C_1 = p_0 - 1,96\sqrt{\frac{p_0(1-p_0)}{n}} = 0,1 - 1,96\sqrt{\frac{0,1 \times 0,9}{50}} = 0,02$$

$$C_2 = p_0 + 1,96\sqrt{\frac{p_0(1-p_0)}{n}} = 0,18.$$

5° $f_n = \frac{10}{50} = 0,2 > C_2$. $f_n \in W$.

On rejette $(H_0)$.

Il y a bien un effet des rejets chimiques sur l'apparition de ces algues.

## $\boxed{\text{Exo 4}}$

**1°** **a)** Modèle : $Perf_i = \beta_0 + \beta_1 \times Taille_i + \varepsilon_i$

avec $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

et les $\varepsilon_i$ indépdts.

**b)** Il s'agit d'1 régression linéaire simple. Le test de significativité du modèle global se confond avec le test de non nullité du coef. de la variable explicative.

**c)** La p-valeur de la constante indique qu'elle est significativement non nulle.

**d)** Le modèle est significatif, en raison des p-valeurs. Mais il n'explique qu'une partie de la variabilité de la performance $\simeq 40\%$.

**2°** **a)** Dans la colonne Sum Sq, la première valeur est la somme des carrés expliqués par la variable provenance, la deuxième est la somme des carrés résiduels.

~~[illegible]~~

**b)** $(H_0)$ Performance indépdte de la provenance
(les échantillons sont tous issus de la m̂ population)
même moyenne

$(H_1)$ Performance liée à la provenance.

**c)** Echantillons gaussiens.

**d)** p-valeur $= 0{,}47 \gg 0{,}05$. On ne peut pas rejeter $(H_0)$
La performance semble indépendante de la provenance.