# Question 1:

Soit X1,X2,…,X100 un échantillon aléatoire d'une population X de loi normale de moyenne μ=1 et de variance σ2=1, c'est-à-dire X~N(1,1)

a) Calculer P(1≤X̄≤1,05)

b) On définit deux variables T et U par

$$T = \sum_{i=1}^{6} X_i \quad \text{et} \quad U = \sum_{i=7}^{9} X_i.$$

Calculez P(T>U)

c) On définit $Y = \sum_{i=1}^{100} (X_i - 1)^2.$

1.) Donner le nom de la loi exacte de Y ainsi que les valeurs de ses paramètres.

2.) Donner le nom de la loi approximative de Y selon le théorème central limite ainsi que les valeurs de ses paramètres.

d) On définit

$$G = \sum_{i=1}^{6} (X_i - 1)^2 \quad \text{et} \quad H = \sum_{i=7}^{9} (X_i - 1)^2.$$

Déterminer la valeur de m telle que P(G>mH)=0,01

Question #1 (6 points)

a) $X \sim N\left(\mu = 1 ; \sigma^2 = \frac{1}{100}\right)$ $Z \sim N(0,1)$ où $Z = \frac{\bar{X} - \mu}{\sigma}$

$$P(1 \leq \bar{X} \leq 1{,}05) = P(0 \leq Z \leq 0{,}5)$$
$$= \Phi(0{,}5) - \Phi(0)$$
$$= 0{,}69146 - 0{,}5$$
$$= 0{,}19146$$

b) $P(T > U) = P(T - U > 0) = P(V > 0)$ où $V = T - U \Rightarrow V \sim N(3, 9)$
$$= P(Z > -1)$$
$$= 1 - \Phi(-1)$$
$$= \Phi(1)$$
$$= 0{,}84134$$

c) $y = \sum_{i=1}^{100} (x_i - 1)^2$

1.c) $Y \sim \chi^2_{100}$

2.c) $Y \sim N(k, 2k)$
$Y \sim N(100, 200)$

d) $G = \sum_{i=1}^{6} (X_i - 1)^2$ et $H = \sum_{i=7}^{9} (X_i - 1)^2$

5582

Par la loi de Fisher:

$$F = \frac{G/u}{H/v} = \frac{G/6}{H/3} = \underline{\frac{G}{2H}}$$

On a:

$$P(G > mH) = 0{,}01$$
$$= P\left(\underline{\frac{G}{2H}} > \frac{m}{2}\right) = 0{,}01$$

$$\Rightarrow \frac{m}{2} = F_{0{,}01;\,6{,}3} = 27{,}91 \Rightarrow m = 55{,}82$$

## Question n° 2 : (6 points)

Le tableau ci-dessous présente la distribution de fréquence du nombre de défauts $X$ sur le fini d'une pièce observés dans un échantillon aléatoire de 100 pièces d'une production.

| Nombre de défauts ($x_i$) | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Nombre de pièces ($O_i$) | 15 | 45 | 25 | 15 |

En utilisant le test d'ajustement du Khi-deux, on veut tester l'hypothèse selon laquelle la variable $X$ est distribuée selon une loi de Poisson.

a) **(1 point)** Formuler les hypothèses $H_0$ et $H_1$ du test à effectuer.

$H_0 : X \sim Poi(c)$

$H_1 : X \not\sim Poi(c)$ ✓

b) **(5 points)** Effectuer le test et conclure au seuil $\alpha = 5\%$.

**Rappel :** si $X \sim \text{Poisson}(c)$, alors $P(X = x) = \dfrac{c^x e^{-c}}{x!}$, pour $x = 0, 1, \ldots$

Question #2 (6 points)

a) $H_0 : X \sim Poi(c)$
$H_1 : X \nsim Poi(c)$

b) On rejette $H_0$ si
$\mu_0 > \chi^2_{0{,}05;\, k-p-1}$

$\hat{c} = \frac{(15 \cdot 0 + 45 \cdot 1 + 25 \cdot 2 + 15 \cdot 3)}{15 + 45 + 25 + 15} = 1{,}4$

1)

| X | $O_i$ | $P(X=k)$ | E |
|---|---|---|---|
| 0 | 15 | $\frac{1.4^0 \cdot e^{-1.4}}{0!} = 0{,}2466$ | 24,66 |
| 1 | 45 | 0,3452 | 34.52 |
| 2 | 25 | 0,2417 | 24,17 |
| 3 | 15 | 0,1128 | 11.28 |

2) $\mu_0 = \frac{(15-24.66)^2}{24.66} + \frac{(45-34.52)^2}{34.52} + \frac{(25-24.17)^2}{24.17} + \frac{(15-11.28)^2}{11.28}$

$= 3{,}7841 + 3{,}1816 + 0{,}0285 + 1{,}2268$

$= 8{,}22$

3) $\chi^2_{0{,}05;2} = 5{,}99$

On rejette $H_0$, car $\mu_0 > \chi^2_{0.05;2}$

# Question 3:

Le service du contrôle de la qualité veut étudier (estimer et tester) la proportion actuelle d'unités non conformes d'une production. Pour cette étude, le service compte utiliser un échantillon de n unités prélevées au hasard de la production.

a) Quelle doit être la plus petite valeur de n pour que, avec un niveau de confiance de 95%, l'erreur d'estimation soit inférieure à 0,045 ?

Dans le passé, la proportion d'unités non conformes de la production était de 10%. Le service du contrôle de la qualité veut effectuer un test statistique, de seuil a = 5%, afin de vérifier si cette proportion est à présent plus élevée.

b) Définir le paramètre d'intérêt et formuler les hypothèses Ho et H1 du test à effectuer.
c) Avec un échantillon de 400 unités prélevées de la production, on obtient une valeur-p (ou «p-value») de 0,0024 pour le test des hypothèses définies en b).

- Quelle doit être la conclusion du test?

- Quel est le nombre d'unités non conformes observées dans cet échantillon?

d) Supposons que la proportion réelle d'unités non conformes soit actuellement de 11% et qu'on utilise un échantillon aléatoire de 500 unités pour le test. Quelle serait alors la probabilité de conclure à tort que la proportion d'unités non conformes actuelle n'est pas plus élevée que dans le passé?

Question #3 (8 points)

a) $n \geq \left(\frac{Z_{\alpha/2}}{e}\right)^2 \cdot 0,5(1-0,5)$

$n \geq \left(\frac{Z_{\alpha/2}}{e}\right)^2 \cdot \frac{1}{4}$

$n \geq \left(\frac{1,96}{0,045}\right)^2 \cdot \frac{1}{4}$

$n \geq 474,27$

$n = 475$

b) p: proportion actuelle d'unités non-conformes d'une production

$H_0: p = 0,10 \quad H_1: p > 0,10$

c) • Rejetter $H_0$, car $0,0024 < 0,05$

• 1) $1 - 0,0024 = 0,9976 \Rightarrow \Phi(Z_0) = 0,9976$
$Z_0 = 2,82$

2) $Z_0 = \frac{\hat{p} - 0,10}{\sqrt{\frac{0,10(1-0,10)}{400}}} = 2,82 \Rightarrow \hat{p} = 0,1423$

3) $400 \cdot 0,1423 = 56,92 \approx 57$ unités

d) $\beta = P(\text{Erreur de deuxième espace})$

$\beta = P\left(p_0 - p_1 + Z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} \mid H_1\right)$

$= P\left(Z < \frac{0,10 - 0,11 + 1,645\sqrt{\frac{0,10 \cdot 0,90}{500}}}{\sqrt{\frac{0,11 \cdot 0,89}{500}}}\right)$

$= \Phi(0,86)$

$= 0,80511$

$= 80,511\%$

# Question 4:

Dans un réseau informatique, un ingénieur mesure les délais de communication (en millisecondes) entre deux ordinateurs avec un protocole standard (population X1) et avec un protocole expé- rimental (population X2). L'ingénieur prélève un échantillon aléatoire de taille n1 = 16 de la population X1 et un échantillon aléatoire de taille n2 9 de la population X2. On suppose que les deux échantillons sont indépendants. Les modèles utilisés et les résultats obtenus sont décrits dans le tableau suivant.

| Population $X_i$ | Taille $n_i$ | Moyenne $\bar{x}_i$ | Variance $s_i^2$ |
|---|---|---|---|
| $X_1 \sim N(\mu_1,\ \sigma_1^2)$ | 16 | 52,4 | 0,37 |
| $X_2 \sim N(\mu_2,\ \sigma_2^2)$ | 9 | 50,3 | 2,18 |

Important: formuler les hypothèses Ho et H1 pour chaque test.

a) Donner un intervalle de confiance de niveau 90% pour le paramètre μ2.
b) Donner un intervalle de prévision de niveau 90% pour une dixième mesure de délai du protocole expérimental.
c) Peut-on rejeter l’affirmation selon laquelle σ1^2 = 0,5? Donner la conclusion d’un test d’hypothèse au seuil α = 0,05.
d) Peut-on dire que les variances σ1^2 et σ2^2 sont différentes? Donner la conclusion d’un test d’hypothèse au seuil α = 0,05.
e) Peut-on affirmer qu’en moyenne les délais du protocole expérimental sont plus courts que ceux du protocole standard? Donner la conclusion d’un test d’hypothèses au seuil α = 0,05.

Question #4 (12 points)

a) On sait que $X_2 \sim N(\mu_2, \sigma_2^2)$

$$IDC(\mu_2) = \bar{X} \pm t_{\frac{\alpha}{2};n-1} \cdot \frac{S}{\sqrt{n}}$$

$$= 50{,}3 \pm t_{0{,}05;8} \cdot \frac{1.48}{3}$$

$$= 50{,}3 \pm 1{,}86 \cdot \frac{1.48}{3}$$

$$= [49{,}3824 ; 51{,}2176]$$

b) $X_{n+1} \in \bar{X} \pm t_{n-1}\left(\frac{\alpha}{2}\right) S \sqrt{1+\frac{1}{n}}$

$X_{n+1} \in 50{,}3 \pm 1{,}86 \cdot 1{,}476 \cdot \sqrt{1+\frac{1}{9}}$

$X_{n+1} \in 50{,}3 \pm 2{,}89$

$X_{n+1} \in [47{,}41 ; 53{,}13]$

c) $H_0 : \sigma_1^2 = 0{,}5$
$H_1 : \sigma_1^2 \neq 0{,}5$

Rejetter $H_0$ si $\chi_0^2 < \chi^2_{\frac{\alpha}{2};n-1}$

1) $\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{15 \cdot 0{,}37}{0{,}5} = 11{,}1$

2) $\chi^2_{n-1}\left(\frac{\alpha}{2}\right) = 27{,}49 \quad \chi_0^2 < \chi^2_{15;\frac{\alpha}{2}}$

3) $\chi^2_{(1-\frac{\alpha}{2});n-1} = 6{,}27 \quad \chi_0^2 > \chi^2_{(1-\ldots}$

On ne peut pas rejetter.

d) $H_0 : \sigma_1^2 = \sigma_2^2$
$\sigma_1^2 \neq \sigma_2^2$

Rejetter si $F_0 < F_{1-\frac{\alpha}{2};n_1-1;n_2-1}$ ou $F_0 > F_{\frac{\alpha}{2};n_1-1;n_2-1}$

1) $F_0 = \dfrac{S_1^2}{S_2^2} = \dfrac{0,37}{2,18} = 0,17$

2) $F_{1-\frac{\alpha}{2};n_1-1;n_2-1} = \dfrac{1}{3,20} = 0,3125$ $F_0 < F_{1-\frac{\alpha}{2}}\ldots$ (rejetter $H_0$)

3) $F_{\frac{\alpha}{2};n_1-1;n_2-1} = 4,10$

e) $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

Rejetter $H_0$ si $T_0 > t_{\alpha;v}$

1) $v = \dfrac{(S_1^2/n_1 + S_2^2/n_2)^2}{\dfrac{(S_1^2/n_1)^2}{n_1+1} + \dfrac{(S_2^2/n_2)^2}{n_2+1}} - 2$
$= 9,94$
$\approx 10$

2) $T_0 = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$
$= 4,07673$

3) $t_{0,05;10} = 1.81$

On rejette $H_0$, oui on peut affirmer qu'en moyenne les délais du protocole expérimental sont plus courts que ceux du protocole standard.

# Question 5:

Pour des raisons de santé publique, on cherche à déterminer le lien entre la concentration d'ozone O3 dans l'air d'une journée et la température à midi de la même journée. Pour ce faire, on envisage un modèle de régression linéaire simple d'équation $Y = \beta_0 + \beta_1 X + \epsilon,$

où X représente la température à midi (en °C) et Y, la concentration d'ozone 03 dans l'air (en microgrammes par millilitre), Bo et Bi sont des paramètres et ε, une erreur aléatoire que l'on suppose de loi N(0, o2).

Les données recueillies pour un échantillon de 12 jours sont présentées dans le tableau suivant.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 10,0 | 23,8 | 16,3 | 27,2 | 7,1 | 25,1 | 27,5 | 19,4 | 15,0 | 19,8 | 32,2 | 20,7 |
| $y_i$ | 71,3 | 115,4 | 76,8 | 113,8 | 81,6 | 115,4 | 125,0 | 83,6 | 84,5 | 75,2 | 136,8 | 102,8 |

a) Calculer β^0 et β^1. Donner l'équation de la droite des moindres carrés.

b) Compléter le tableau d'analyse de la variance ci-dessous. Quelle conclusion peut-on tirer au seuil α = 0,05?

c) Donner une estimation ponctuelle de la variance σ^2

d) À un niveau de 95%, donner un intervalle de prévision pour la concentration d'ozone O3 d'une journée dont la température à midi est de 20C.

e) Calculer le coefficient de détermination R^2 et interpréter ce résultat.

Question # 5 (14 points)

a) $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

1) $S_{xy} = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}$

1.1) $\bar{X} = \frac{1}{n} \sum x_i = \frac{244,10}{12} = 20,34$

1.2) $\bar{Y} = \frac{1}{n} \sum Y_i = \frac{1182,20}{12} = 98,52$

Reprenons $S_{xy} = 25\,631.34 - 12(20,34)(98,52)$
$= 1\,584,58$

2) $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{X}^2 = 5567.37 - 12(20.34)^2 = 602.78$

3) $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{1\,584,58}{602,78} = 2.63$

4) $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 98.52 - 2.63 \cdot 20,34 = 45,02$

$$Y = 45,02 + 2,63X + \epsilon$$

b)

| Source de variation | Somme des carrés | Degrés de liberté | Moyenne des carrés | $F_0$ |
|---|---|---|---|---|
| Régression | SSR = 4167,45 | 1 | MSR = SSR/1 = 4167,45 | $\frac{MSR}{MSE} = 32,69$ |
| Erreur | SSE = 1274,86 | n-2 = 10 | MSE = SSE/(n-2) ≐ 127,49 | — |
| Totale | 5450,197 | n-1 = 9 | — | — |

1) $SS_R = \hat{\beta}_1 S_{xy} = 2,63 \cdot 1\,584,58 = 4167,45$

2) $SS_E = SST - SS_R$
$= S_{yy} - SS_R$
$= \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 - 4167,45$
$= 121\,916,60 - 12(98,52)^2 - 4167,45$
$= 1274,86$

$\longrightarrow$

$H_0 : \beta_1 = 0$ (aucun lien)
$H_1 : \beta_1 \neq 0$ (lien)

On rejette $H_0$ si $f_0 > F_{\alpha;1;n-2}$

3) $F_{\alpha;1;n-2} = 4{,}67$

On rejette $H_0$, alors c'est possible qu'il y ait une rég. entre les 2

c) $\hat{\sigma}^2 = MS_E = 128{,}0415$

d) 1) $\hat{Y}_0 \pm t_{\frac{\alpha}{2};n-2} \sqrt{MSE\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$ 2) $\hat{Y}_0 = 45{,}02 + 2{,}63(20) = 97{,}62$

$\hat{Y}_0 \pm 2{,}23 \sqrt{127{,}49\left(1 + \frac{1}{12} + \frac{(20 - 20{,}91)^2}{602{,}78}\right)}$

$\hat{Y}_0 \pm 26{,}21$

$97{,}62 \pm 26{,}21 \Rightarrow [71{,}41 , 123{,}83]$

e) $R^2 = \hat{\beta}_1^2 \frac{S_{xx}}{S_{yy}} = 2{,}63^2 \cdot \frac{602{,}78}{121916{,}60 - 12(98{,}52)^2} = 0{,}7661$

$R^2$ est proche de 1, alors il y a une certaine régression linéaire.