Exercice 1
On considère la série statistique à deux variables suivante :

x_i	1	2	3	4	5
y_i	0	0	1	1	2

1. Calculer les moyennes de x et de y.
2. Calculer les variances de x et de y, la covariance de x et y.
3. Donner une équation de la droite de régression de y sur x.
4. Tracer le nuage de points et la droite de régression.

Exercice 2
On se propose d'étudier l'influence de la température sur la durée d'incubation des oeufs de grenouilles. On choisit 6 échantillons de 200 oeufs chacun. Le nombre x d'éclosion au 22-ème jour est le suivant

température t_i d'incubation en degrés Celsius	6	6,4	6,8	7,2	7,6	8
nombre x_id'éclosions à la température t_i	131	144	157	170	190	189

a) Dessiner le nuage des données et tracer “à l'oeil” une droite D qui a l'air de bien approcher ce nuage.
b) Calculer le coefficient de corrélation observé et écrire l'équation de la droite de régression de x en t. Etudier la qualité de l'ajustement.
c) Calculer le nombre d'éclosions prédit pour un échantillon de 200 oeufs au 22-ème jour pour une température de 7,5 degrés.

Exercice 3
Soient $\{(x_i, y_i)\}_{i=1}^n$ une série statistique à deux variables. À quelle condition les deux droites de régression, de x sur y et de y sur x, coïncident-elles ?

Exercice 4
On a mesuré les variables x et y sur 10 individus et obtenu les résultats suivants :

individu $n^0 i$	1	2	3	4	5	6	7	8	9	10
x_i	13	16	23	29	35	43	49	55	58	63
y_i	16,5	17,9	20,3	22	23,5	25,3	26,5	27,6	28,2	29,1

1) Calculer la droite de régression linéaire de y en x.
2) On pose $z_i = \log x_i$ (logarithme en base 10). Chercher les valeurs de α, β, γ qui minimisent la somme $\sum_{i=1}^{10}(y_i - \alpha - \beta x_i - \gamma z_i)^2$.
3) Représenter sur le même graphique : le nuage de points (x_i, y_i), la droite de régression de la question 1), la courbe de la fonction $y = \alpha + \beta x + \gamma \log(x)$ avec les coefficients trouvés dans la question 2).

Probabilités

Exercice 5
Soit Y une variable aléatoire uniformément répartie sur l'intervalle $[1, 3]$, c'est-à-dire une variable aléatoire de loi de densité

$$f_Y(y) = \frac{1}{2} \text{ si } y \in [1, 3], \quad 0 \text{ sinon.}$$

Calculer l'espérance et la variance de Y.

Exercice 6
Soit Y une variable aléatoire de Bernouilli de paramètre (n, p) avec $p \in]0, 1[$ et $n \in \mathbb{N}^*$, c'est-à-dire une variable aléatoire telle que

$$\mathbb{P}(Y = k) = \tbinom{n}{k} p^k (1-p)^{n-k} \text{ si } 0 \leq k \leq n, \quad 0 \text{ sinon.}$$

Calculer l'espérance et la variance de Y.

Exercice 7

On lance douze fois une pièce équilibrée. Quelle est la probabilité d'obtenir six "pile" et six "face" ? Quelle est la probabilité d'obtenir au moins deux foix un "pile" ?

Exercice 8

Accident nucléaire : une certitude statistique. La probabilité d'un accident nucléaire majeur en Europe dans les trente prochaines années serait de plus de 100% (*Libération* le vendredi 3 juin 2011). L'article commence par estimer la probabilité d'un accident majeur par réacteur nucléaire et par année de fonctionnement. Selon l'article, le parc mondial actuel de réacteurs cumule 14000 réacteurs-ans (environ 450 réacteurs pendant 31 ans). Pendant cette période, il y a eu quatre accidents majeurs, ce qui mène à une probabilité d'accident majeur d'environ 0,0003 par an pour chaque réacteur. Les auteurs en « déduisent »donc que la probabilité d'un accident majeur en France (avec ses 58 réacteurs) pendant les trente prochaines années serait de 58 fois 30 fois 0,0003, donc d'environ 50%. Quant à la probabilité d'un accident en Europe (143 réacteurs) dans les trente prochaines années, elle « est »de 143 fois 30 fois 0,003, « donc » d'environ 129%. Commenter.

Exercice 9

On effectue un sondage auprès de 1000 personnes prises au hasard parmi 60 millions. La question posée admet deux réponses : A ou B. Pour simplifier, on suppose que toutes les personnes susceptibles d'être interrogées ont un avis stable sur la question posée et que tous les sondés répondent effectivement. Le nombre de personnes N préférant A est donc supposé bien défini. Le sondage a pour but d'estimer ce nombre. Nous ne nous intéressons ici qu'à la modélisation aléatoire du sondage.

1. Soit k un nombre compris entre 0 et 1000. Exprimer en fonction de k et N la probabilité que les sondeurs obtiennent k réponses A dans leur échantillon. Commencer par décrire l'ensemble des éventualités de l'expérience aléatoire que constitue le sondage.
2. À quelles conditions est-il raisonnable de considérer que les probabilités de la question précédente sont bien approchées par des probabilités associées à une loi binomiale ? Quelle loi binomiale ?

Lois normales

Exercice 10

1. Soit X une v.a. à densité de loi $N(0; 1)$. Donner des valeurs approchées de $\mathbb{P}(X > -1)$; $\mathbb{P}(X < -2)$; $\mathbb{P}(1 < X < 2)$ et $\mathbb{P}(|X| < 2)$.
2. Soit Z une v.a. à densité de loi $N(1.75\ ;\ 0.01)$. Donner une valeur approchée de $\mathbb{P}(Z > 1.9)$.

Exercice 11

On suppose que la taille mesurée en mètre des garçons de 20 ans suit une loi normale de moyenne m et d'écart-type σ. On sait que 84% des garçons de 20 ans mesurent moins de 1 m 86 et que 97% mesurent plus de 1 m 58. Déterminer m et σ.

Exercice 12

Pour un échantillon de 300 individus sains, on a étudié la glycémie ; on a constaté que 20% des glycémies sont inférieures à 0.82 g/l et que 30% des glycémies sont supérieures à 0.98 g/l. En supposant que la glycémie suit une loi normale, déterminer la moyenne et l'écart-type de cette loi.

Exercice 13

500 personnes ont postulé pour une place, mais 379 ont été refusées parce qu'elles n'étaient pas assez grandes. La taile d'un individu suivant une loi normale de moyenne 171.5 cm et d'écart-type 5 cm, estimer la taille minimale exigée.