1. On veut savoir si la résistance moyenne de composants produits dans une usine est 400Ω. On considère que la distribution des résistances est normale, et on mesure pour 16 composants les valeurs 392, 396, 386, 389, 388, 387, 403, 397, 401, 391, 400, 402, 394, 406, 406, 400.

   (a) Donner les estimations ponctuelles des moyenne et variance.

   (b) Peut-on considérer, au seuil de signification $\alpha = 5\%$, que le lot respecte la norme de 400Ω ? Même question avec un seuil de $\alpha = 1\%$.

   corrigé succint :

   (a) On trouve $\boxed{\bar{x} = 396.125,\ s = 6.742 \text{ et } s^2 = 45.45.}$

   (b) Si l'on fait l'hypothèse H_0 : "le lot respecte la norme de 400Ω", alors dans 95% des cas la moyenne sur un échantillon d'effectif 16 se trouve dans l'intervalle $[400 - t * 6.742/4, 400 + t * 6.742/4]$, t étant lu dans la table de la loi de Student à 15 degrés de liberté : $t = 2.1314$. Ainsi l'intervalle de confiance 95% pour la résistance est $\boxed{[396.40, 403.59]\text{, et on peut donc, au risque } 5\%\text{, rejeter l'hypothèse.}}$

   Au seuil $\alpha = 1\%$, on a dans l'hypothèse H_0 un intervalle de confiance pour la moyenne $[400 - t * 6.742/4, 400 + t * 6.742/4]$, avec $t = 2.9467$. Ainsi, $\boxed{\text{l'intervalle est } [395.03, 404.97]\text{. Au risque } 1\%\text{, on ne rejette pas } H_0.}$

2. Un fabricant se vante de proposer des tubes à essai d'une durée de vie supérieure à 2000h de chauffage. A l'aide d'un échantillon de 100 tubes testés, on estime la durée de vie moyenne à 1975h, avec un écart-type de 130h. Peut-on affirmer, au risque 5%, que le fabriquant ment ?

   corrigé succint : Il s'agit ici d'un test unilatéral...H_0 est l'hypothèse : "la durée de vie moyenne vérifie $\mu \geq 2000$". On peut supposer, l'effectif de l'échantillon étudié étant grand, que $\sqrt{n}\dfrac{\bar{X} - \mu}{s}$ suit une loi normale centrée réduite.
   Si H_0 est vérifiée, on cherche t tel que $p(\mu - ts/\sqrt{n} \leq \bar{X}) = 0.95$, soit $p(-t \leq \sqrt{n}\dfrac{\bar{X} - \mu}{s}) = 0.95$, et donc $1 - F(-t) = F(t) = 0.95$: $t = 1.64$.
   Ainsi, dans l'hypothèse H_0, la durée de vie moyenne d'un échantillon d'effectif 100 se trouve, dans 95% des cas, dans l'intervalle $[2000 - 1.64 * 130/10, +\infty[= [1978.68, +\infty[$.
   La mesure de 1975h sur l'échantillon n'étant pas dans cet intervalle,
   $\boxed{H_0 \text{ doit être rejetée : il est probable que le fabriquant mente.}}$

3. Un fabricant annonce que la masse d'un composant de l'un de ses produits est de 75 mg. Les mesures pour le vérifier étant coûteuses, trois seulement sont réalisées, dont les résultats sont 70, 72 et 74 mg. Peut-on, au risque de 5% de se tromper, dénoncer la publicité du fabriquant ?

   corrigé succint : Notons X la variable aléatoire correspondante. (on doit supposer que la loi de X est une loi normale pour pouvoir appliquer les méthodes du cours). On note $\mu = E(X)$: il s'agit donc ici d'effectuer un test bilatéral de l'hypothèse $H_0 : \mu = 75$.
   On obtient sur un échantillon de 3 mesures : $n = 3$, $\bar{x} = 72$, $\sigma'^2 = 8/3$ et $s^2 = 8/2 = 4$, donc l'estimation ponctuelle de l'écart-type est $s = 2$.
   On sait que $\sqrt{n}\dfrac{\overline{X} - \mu}{s}$ suit une loi de Student à 2 degrés de liberté, donc si $\alpha = 0.05$, $t(\alpha) = 4.3027$.
   Ainsi, la moyenne des durées de vie mesurées sur un échantillon d'effectif 3 sera, dans 95% des cas, dans l'intervalle $[75 - 4.3027 \times s/\sqrt{3}, 75 + 4.3027 \times s/\sqrt{3}] = [70.03, 79.97]$.
   La valeur moyenne 72 mesurée sur l'échantillon étant bien dans cet intervalle,
   $\boxed{\text{on n'a pas de raisons, au vu de ces mesures, de rejeter } H_0.}$

4. Un laboratoire pharmaceutique désire étudier les effets secondaires potentiels d'un médicament sur le taux de cholestérol des patients. Cent volontaires sains sont donc choisis pour tester le médicament.

   (a) Avant l'expérience, le taux de cholestérol moyen de ces volontaires est de 2.02 ± 0.2g/l.
   Le taux de cholestérol moyen dans la population étant de 2 g/l, vérifier que cet échantillon est représentatif au risque 5%.

   (b) Après un mois de traitement, seuls 97 volontaires reviennent faire un test. Leur taux moyen de cholestérol est passé à 2.09 g/l avec un écart-type d'échantillon de 0.25g/l.
   La différence est-elle significative au risque 5% ? Au risque 1% ?

   corrigé succint :

   (a) Soit X_1 la variable aléatoire qui mesure le taux de cholestérol d'un individu ; $E(X_1) = \mu_1 = 2$.
   $\overline{X_1}$ est le taux moyen mesuré sur un échantillon de taille $n_1 = 100$.
   Alors n_1 étant plus grand que 30, on peut considérer que $\sqrt{n_1}\dfrac{\overline{X_1} - 2}{s_1}$ suit une loi normale, avec $s_1 = 0.2$ estimation ponctuelle de l'écart-type de X_1.
   Ainsi, dans 95% des cas le taux moyen observé sur un échantillon sera compris dans $[2 - 1.96 \times 0.2/10, 2 + 1.96 \times 0.2/10] = [1.961, 2.039]$.

Le taux de cholestérol moyen des volontaires étant bien dans cet intervalle, on peut considérer que cet échantillon est représentatif.

(b) Soit X_2 la variable aléatoire mesurant le taux de cholestérol d'un individu après un mois de traitement ; son espérance μ_2 est inconnue. $\overline{X_2}$ est le taux moyen d'un échantillon de taille $n_2 = 97$.

On fait l'hypothèse H_0 : « les taux de cholestérol moyens sont les mêmes avant et après traitement ». Alors $\mu_1 = \mu_2$, et on peut considérer que $\dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \sim N(0,1)$ (avec $s_1 = 0.2$, $s_2 = 0.25$), et par conséquent on détermine l'intervalle de confiance au risque 5% de $\overline{X_1} - \overline{X_2}$: $[-1.96\sqrt{s_1^2/n_1 + s_2^2/n_2}, 1.96\sqrt{s_1^2/n_1 + s_2^2/n_2}] = [-0.063, 0.063]$.

Comme la différence entre les taux moyens mesurés $2.02 - 2.09 = 0.07$ n'est pas dans cet intervalle, elle est significative, et on rejette H_0 donc

on considère, au risque 5% de se tromper, que le médicament a un effet.

En revanche, l'intervalle de confiance au risque 1% est $[-2.57\sqrt{s_1^2/n_1 + s_2^2/n_2}, 2.57\sqrt{s_1^2/n_1 + s_2^2/n_2}] = [-0.083, 0.083]$, intervalle qui contient la valeur $2.02-2.09 = 0.07$, donc la différence n'est pas significative au risque 1%.

5. Pour étudier un nouvel alliage métallique, on a soumis un échantillon aléatoire de 16 tiges aux essais pour obtenir les résistances suivantes en kg/cm^2 : 1895, 1920, 1886, 1890, 1864, 1880, 1875, 1915, 1850, 1927, 1910, 1912, 1886, 1903, 1854, 1880. On suppose la résistance distribuée normalement.

   (a) Estimer par intervalle avec un niveau de confiance de 95%, la résistance moyenne à la rupture

   (b) Avant l'introduction de ce nouvel alliage la résistance moyenne à la rupture des tiges était de 1840kg/cm^2. Que peut-on conclure des essais effectués avec le nouvel alliage ?

corrigé succint :

(a) L'effectif de l'échantillon est $n = 16$.

L'estimation ponctuelle de la moyenne, égale à la moyenne de l'échantillon, est $\bar{x} = 1890.44$,

Puis on calcule l'écart-type de l'échantillon, $\sigma' = 22.36$, et on en déduit l'estimation ponctuelle de l'écart-type de la population : $s = \sqrt{16/15}\ \sigma' = 23.09$.

Alors l'intervalle de confiance 95% de la moyenne est l'intervalle $[\overline{x} - t(0.05)\dfrac{s}{\sqrt{16}}; \overline{x} + t(0.05)\dfrac{s}{\sqrt{16}}]$.

On lit $t(0.05)$ dans la table de la loi de Student à 15 degrés de liberté : $t(0.05) = 2.1314$; par conséquent l'intervalle est $[1878.14, 1902.74]$.

(b) On peut affirmer (avec un risque de se tromper inférieur à 5%) que le nouvel alliage est plus résistant que l'ancien.

6. On relève chaque jour pendant 200 jours le nombre d'atterrissages entre 14h et 15h dans un aéroport :

Nb d'atterrissages	0	1	2	3	4	5	6	7	8	9	10
Nb de jours	11	28	43	47	32	28	7	0	2	1	1

   (a) Soit X la variable « nombre d'atterrissages par jour entre 14h et 15h ». Donner les estimations ponctuelles de $E(X)$ et $\text{Var}(X)$ et estimer $E(X)$ par un intervalle de confiance 95%. Ces résultats sont-ils compatibles avec une loi de poisson ? Quel serait son paramètre ?

   (b) Tester la validité de ce modèle (test du χ^2 au risque 5%).

   (c) Calculez la probabilité d'avoir dans cet aéroport, toujours entre 14h et 15h : 0 atterrissage un jour donné, 1 ou 2 atterrissages un jour donné, 2 atterrissages en tout sur 3 jours quelconques.

corrigé succint :

(a) L'effectif de l'échantillon est $n = 200$.

On détermine l'estimation ponctuelle de la moyenne $\bar{x} = 600/200 = 3$, et l'estimation ponctuelle de la variance $s^2 = 596/199 = 2.995$, soit $s = 1.73$.

L'effectif est suffisant pour assimiler la loi de $\sqrt{n}\dfrac{\bar{X} - \mu}{s}$ a une loi normale centrée réduite : l'intervalle de confiance 95% pour la moyenne est donc $[\bar{x} - 1.96\frac{s}{\sqrt{200}}, \bar{x} + 1.96\frac{s}{\sqrt{200}}]$, soit $[2.76, 3.24]$.

On sait que l'intervalle de confiance pour la variance est $[\dfrac{199}{c_1^2}s^2, \dfrac{199}{c_2^2}s^2]$, avec c_1^2 et c_2^2 lus dans la table de la loi du χ^2 à 199 degrés de liberté. En pratique on utilise la table de la loi du χ^2 à 200 degrés de liberté et on lit dans les colonnes $0.025 = 0.05/2$ et $0.975 = 1 - 0.05/2$: $c_1^2 = 241.1$ et $c_2^2 = 162.7$. Ainsi l'intervalle est $[0.825s^2, 1.223s^2] = [2.47, 3.66]$.

L'espérance et la variance ont (quasiment) la même estimation ponctuelle, égale à trois : les résultats sont compatibles avec le fait que X suive une loi de Poisson de paramètre 3.

(b) On définit l'hypothèse H_0 : « X suit une loi de Poisson de paramètre 3 ».

Pour utiliser un test du χ^2 pour accepter ou refuser H_0, on doit avoir des valeurs (ou classes) d'effectif "théorique" au moins égal à 5. Or ici ce n'est pas le cas : les valeurs 7, 8, 9, 10 ont des effectifs $np(X = 7)$, $np(X = 8)$, $np(X = 9)$, $np(X = 10)$ inférieurs : on doit regrouper en une seule classe les valeurs 7, 8, 9, 10.

Les probabilités des évènements $X = 0$, $X = 1$, ..., $X = 6$, $X \in [7, 10]$ sont déterminées par lecture de la table de la loi de Poisson : $p(X = 0) = 0.0498$, $p(X = 1) = 0.1494$, $p(X = 2) = 0.224$, $p(X = 3) = 0.224$, $p(X = 4) = 0.168$, $p(X = 5) = 0.1008$, $p(X = 6) = 0.0504$, $p(X \geq 7) = 1 - p(X \leq 6) = 1 - 0.9665 = 0.0335$. On obtient en multipliant ces valeurs par l'effectif total 200 les effectifs « théoriques » de chaque classe :

Nb d'atterrissages	0	1	2	3	4	5	6	7+
Effectif mesuré	11	28	43	47	32	28	7	4
Effectif théorique	9.96	29.88	44.8	44.8	33.6	20.16	10.08	6.7

et en en déduit la valeur de $\chi_o^2 = \dfrac{(11-9.96)^2}{9.96} + \dfrac{(28-29.88)^2}{29.88} + \ldots + \dfrac{(4-6.7)^2}{6.7} = 5.56$.

Le critère de décision pour accepter H_0 sera $\chi_o^2 \leq c^2$, avec c^2 lu dans la table du χ^2 à 6 degrés de liberté (on a 8 classes, et le test porte sur une loi de Poisson : le nombre de d.d.l.à considérer est donc 8-2). On lit dans la table, dans la colonne $\alpha = 0.05$ et la ligne 5 : $c^2 = 12.59$.

Ainsi, on accepte H_0.

(c) Si on admet, suite au test du χ^2, que X suit une loi de Poisson de paramètre 3, on lit directement dans la table les valeurs $p(X = 0) = 4.98\%$, $p(1 \leq X \leq 2) = p(X = 1) + p(X = 2) = 37.34\%$.

Si on s'intéresse aux nombre d'atterrissages entre 14h et 15h sur trois jours distincts, si on note X_1, X_2 et X_3 les variables donnant le nombre d'atterrissage chacun des trois jours, on sait que, X_1, X_2 et X_3 étant indépendantes, $X_1 + X_2 + X_3$ suit une loi de Poisson de paramètre 9. Par conséquent, $p(X_1 + X_2 + X_3 = 2) = 0.5\%$.

7. Un contrôle de qualité sur des échantillons issus d'une production conduit aux résultats suivants :

Nb de défauts	0	1	2	3	4	5	6
Nb de pièces	15	30	48	46	34	22	5

(a) Donner une estimation non biaisée de l'espérance et la variance du nombre de défauts X.

(b) Donner une estimation par un intervalle à 95% de confiance de l'espérance de X si l'on suppose que la variance de X est égale à l'estimation ponctuelle de l'espérance.

(c) Choisir une loi discrète pour représenter la variable X et faire un test du χ^2 à 95% de confiance.

corrigé succint :

(a) Ici l'effectif est $n = 200$. La moyenne d'échantillon est l'estimation ponctuelle de l'espérance (synonyme de « estimation non biaisée », voir le cours) : $\bar{x} = 540/200 = 2.7$. De même l'estimation non biaisée de la variance est $s^2 = 2.271$, et $s = 1.507$.

(b) Si l'on suppose que $\sigma^2 = 2.7$, on obtient (la méthode est habituelle maintenant...) l'intervalle $[\bar{x} - 1.96 \times \sqrt{2.7/200}, \bar{x} + 1.96 \times \sqrt{2.7/200}] = [2.47, 2.93]$

(c) La moyenne et de la variance estimées sont proches, et vue la nature du problème (évènements rares) on teste l'hypothèse que X suit une loi de Poisson : H_0 : « X suit une loi de Poisson de paramètre 2.7 ». La table ne donnant pas les valeurs d'une loi de Poisson de paramètre 2.7 on calcule directement les effectifs théoriques $200 \times p(X = k) = 200 \times e^{-2.7} 2.7^k / k!$:

Nombre de défauts	0	1	2	3	4	5	6
Effectif mesuré	15	30	48	46	34	22	5
Effectif théorique	13.4	36.3	49	44.1	29.8	16.1	7.2

Le χ^2 observé est $\chi_o^2 = 4.81$, alors que, sur la table du χ^2 à $7 - 2 = 5$ degrés de liberté on lit $c^2 = 11.07$. On accepte donc l'hypothèse H_0.

Remarque : un test du χ^2 validerait aussi l'hypothèse d'une loi de Poisson de paramètre 3...