TD 6 : tests d'hypothèses

1. La durée de vie annoncée d'un composant est de 75 h. On souhaite vérifier ces valeurs.
   Les mesures étant destructrices et donc coûteuses, trois seulement sont réalisées, dont les résultats sont 70 h, 72 h et 74 h. Peut-on, au risque de 5% de se tromper, dénoncer la publicité du vendeur ?
2. 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, le taux moyen des volontaires 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% ?
3. 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 ?
4. 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 discrète « nombre d'atterrissages par jour entre 14h et 15h ». Donner l'estimation ponctuelle de la variance et de l'écart-type de X.
   (b) Estimer $E(X)$ et $\text{Var}(X)$ par des intervalles de confiance à 95%. Ces résultats sont-ils compatibles avec une loi de poisson ? Quel serait son paramètre ?
   (c) Tester la validité de ce modèle (test du χ^2 au risque 5%).
   (d) 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.
5. 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.
   (d) A l'aide de la droite de Henry, dire s'il semble possible de représenter le phénomène par la loi normale. Donner l'estimation graphique de ses paramètres.

TD 6 : corrigé

1. Notons X la variable aléatoire (**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 : $n = 3$, $\bar{x} = 72$ et $\sigma'^2 = 8/3$ 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 $\alpha = 0.05$, on a $t(\alpha) = 4.3027$.

   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 72 mesurée sur l'échantillon étant bien dans cet intervalle, on accepte H_0.

2. (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 $s_1 = \sqrt{100/99} \times 0.2 = 0.201$ estimation ponctuelle de l'écart-type de X_1.

   Ainsi, dans 95% des cas le taux moyen observé sur un échantillon sera compris dans $[2 - 1.96 \times 0.201/10, 2 + 1.96 \times 0.201/10] = [1.96, 2.04]$.

   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 $n_2 = 100$.

   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 $s_2 = \sqrt{100/99} \times 0.25 = 0.251$), et par conséquent on détermine l'intervalle de confiance au risque 5% de $\overline{X_1} - \overline{X_2}$: $[-1.96(\sqrt{s_1^2/n_1} + \sqrt{s_2^2/n_2}), 1.96(\sqrt{s_1^2/n_1} + \sqrt{s_2^2/n_2})] = [-0.063, 0.063]$.

   Comme la différence entre les taux moyens mesurés $2.02 - 2.09 = 0.07$ n'est pas dans cet intervalle, on rejette H_0 au risque 5%, donc on considère, au risque 5%, que le médicament a un effet secondaire.

   En revanche, l'intervalle de confiance au risque 1% est $[-2.33(\sqrt{s_1^2/n_1} + \sqrt{s_2^2/n_2}), 2.33(\sqrt{s_1^2/n_1} + \sqrt{s_2^2/n_2})] = [-0.075, 0.075]$. Ainsi, la différence n'est pas significative au risque 1%.

3. (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 $[\bar{x} - t(0.05)\dfrac{s}{\sqrt{16}}; \bar{x} + t(0.05)\dfrac{s}{\sqrt{16}}]$.

   On lit $t(0.05)$ étant lu 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.

4. (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'espérance et la moyenne 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) L'effectif est suffisant pour assimiler la loi de $\sqrt{n}\frac{\bar{X}-\mu}{s}$ 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 $[\frac{199}{c_1^2}s^2, \frac{199}{c_2^2}s^2]$, 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 $[0.825s^2, 1.223s^2] = [2.47, 3.66]$.

(c) 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 au moins égal à 5. Or ici ce n'est pas le cas : les valeurs 7, 8, 9, 10 ont des effectifs respectifs 0, 2, 1, 1. On doit regrouper en une seule classe les valeurs 6, 7, 8, 9, 10 : $[6, 10]$ est une classe d'effectif 11.

Les probabilités des évènements $X = 0, X = 1, ..., X = 5, X \in [6, 10]$ sont déterminées par lecture de la table de la loi de Poisson : $p(X = 0) = 0.0498$, $p(X = 1) = 0.1494$, $p(X = 2) = 0.224$, $p(X = 3) = 0.224$, $p(X = 4) = 0.168$, $p(X = 5) = 0.1008$, $p(X \in [6, 10]) = p(X \leq 10) - p(X \leq 5) = 0.9997 - 0.9161 = 0.0836$. 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	[6,10]
Effectif mesuré	11	28	43	47	32	28	11
Effectif théorique	9.96	29.88	44.8	44.8	33.6	20.16	16.72

et en en déduit la valeur de $\chi_o^2 = \frac{(11-9.96)^2}{9.96} + \frac{(28-29.88)^2}{29.88} + \ldots + \frac{(11-16.72)^2}{16.72} = 5.52.$

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 5 degrés de liberté (on a 7 classes, et le test porte sur une loi de Poisson : le nombre de d.d.l.à considérer est donc 7-2). On lit dans la table, dans la colonne $\alpha = 0.05$ et la ligne 5 : $c^2 = 11.0705$.

Ainsi, on accepte H_0.

(d) 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'atterissages entre 14h et 15h sur trois jours distincts, si on note X_1, X_2 et X_3 les variables donnant le nombre d'atterissage 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, $p(X_1 + X_2 + X_3 = 2) = 0.5\%$.

5. (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...