Examen de première session, Janvier 2016

Aucun document autorisé, calculatrice autorisée, durée : 02h00

Exercice n°1.

On s'intéresse, sur un site donné, à la concentration d'un polluant métallique que l'on note X $(\mu g/g)$ mesurée dans des sédiments. On suppose que la variable $X \rightsquigarrow \mathcal{N}\left(\mu_0, \sigma_0^2\right)$ où le paramètre populationnel $\sigma_0^2 = 100$ est connu. On dispose d'un échantillon $\{X_1, \cdots, X_n\}$ de taille $n = 100$.

1. On s'intéresse à $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, l'estimateur de la moyenne de la population. Avec les hypothèses précédentes, calculer son espérance et sa variance en fonction de μ_0, σ_0 et n. Qu'en déduisez-vous sur les qualités de cet estimateur?
2. Quelle loi suit la variable $\overline{X}$?
3. On suppose maintenant que $\mu_0 = 50\,\mu g/g$. Calculer les probabilités suivantes :
   (a) $P\left(\overline{X} \leq 52\right)$
   (b) $P\left(\overline{X} \leq 48\right)$
   (c) $P\left(48 \leq \overline{X} \leq 52\right)$
   (d) $P\left(48 \leq \overline{X} \leq 52/\overline{X} \geq 50\right)$
   (e) $P\left(\overline{X} = 48.25\right)$
4. En fait on a des doutes sur la valeur de μ_0. Sur un échantillon réalisé de taille $n = 100$, on a mesuré une concentration moyenne de $\overline{x} = 51.2\,\mu g/g$. Vérifier, au risque de $5\,\%$, si la valeur supposée de μ_0 est la bonne.
5. Construire un intervalle de confiance de la moyenne. Qu'en déduisez-vous?

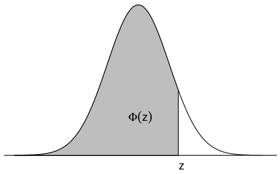
Exercice n°2.

Vous êtes sur la paillasse et vous utilisez un instrument de mesure qui vous permet de doser les nitrates. Soit X la concentration en nitrates $(\mu g/l)$, vous faites $n = 6$ mesures dans des conditions similaires et vous obtenez

102.94 103.75 103.89 96.17 102.94 112.08.

1. Vous vous penchez à l'arrière de votre appareil et vous lisez que la variance des mesures est de $\sigma_0^2 = 25$. Montrez qu'il n'y a pas de raison de remettre en doute la calibration de la machine.
2. Donner un intervalle de confiance à 95% de la variance de l'appareil.
   On donne $\chi^2_{5;0.025} = 0.83$, $\chi^2_{5;0.05} = 1.145$, $\chi^2_{5;0.95} = 11.07$, $\chi^2_{5;0.975} = 12.83$.

Distribution normale standard $\Phi(z)$



Pour $z < 0$ on utilise symmetrie: $P(Z \leq z) = \Phi(z) = 1 - \Phi(-z)$, $z \in \mathbb{R}$.

z	0	1	2	3	4	5	6	7	8	9
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56750	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84850	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92786	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

Correction exercice n°1.

1. On sait que $X \rightsquigarrow \mathcal{N}(\mu_0, \sigma_0^2)$. Calculons espérance et variance de X :
$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{n}{n}\mu_0 = \mu_0.$$
On en déduit que la moyenne empiriques est un estimateur sans biais de la moyenne μ_0 de la population. D'autre part :
$$V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i\right) = \frac{n}{n^2}\sigma_0^2 = \frac{\sigma_0^2}{n}.$$
La moyenne emprique est un estimateur sans biais et convergent car $\lim_{n\to+\infty} V\left(\overline{X}\right) = 0$. C'est un bon estimateur.

2. En vertu du thérorème central limite, la variable $\overline{X} \rightsquigarrow \mathcal{N}\left(\mu_0, \frac{\sigma_0^2}{n}\right)$.

3. Soit la variable centre réduite
$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma_0}{\sqrt{n}}} \rightsquigarrow \mathcal{N}\left(0, 1\right),$$
en utilisant la table de la loi normale centrée réduite, on obtient :
   (a) $P\left(\overline{X} \le 52\right) = P\left(Z \le \frac{52-50}{1}\right) = P\left(Z \le 2\right) = \Phi\left(2\right) = 0.97725$
   (b) $P\left(\overline{X} \le 48\right) = P\left(Z \le \frac{48-50}{1}\right) = P\left(Z \le -2\right) = \Phi\left(-2\right) = 1 - \Phi\left(2\right) = 1 - 0.97725 = 0.02275$
   (c) $P\left(48 \le \overline{X} \le 52\right) = P\left(-2 \le Z \le 2\right) = P\left(Z \le 2\right) - P\left(Z \le -2\right) = 0.97725 - 0.02275 = 0.9545$
   (d) $P\left(48 \le \overline{X} \le 52 / X \ge 50\right) = \frac{P(-2\le Z\le 2 \cap Z\ge 0)}{P(Z\ge 0)} = \frac{P(0\le Z\le 2)}{1-P(Z\le 0)} = \frac{P(Z\le 2)-P(Z\le 0)}{1-P(Z\le 0)} = \frac{\Phi(2)-\Phi(0)}{1-\Phi(0)} = \frac{0.97725-0.5}{1-0.5} = 0.9545$
   (e) $P\left(\overline{X} = 48.25\right) = 0$ car $\overline{X}$ est une variable aléatoire continue.

4. Nous sommes dans le cas où l'écart-type de la population est connu et l'effectif $n = 100$ est grand. On se propose de tester l'hypothèse $H_0 : \mu = \mu_0$ contre l'alternative $H_0 : \mu \neq \mu_0$. On est dans le cas d'un test gaussien bilatéral avec zone de rejet à droite et à gauche. Sous H_0, la variable
$$\overline{X} \rightsquigarrow \mathcal{N}\left(\mu = \mu_0, \frac{\sigma_0^2}{n}\right).$$
La zone de non-rejet de H_0 est donnée par
$$\overline{RH_0} = \left[\mu_0 - z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}; \mu_0 - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right] = \left[50 - 1.96 \times \frac{10}{\sqrt{100}}; 50 + 1.96 \times \frac{10}{\sqrt{100}}\right] = [48.04; 51.96]$$
où $z_{1-\alpha/2} = -z_{\alpha/2} = 1.96$ est le quantile d'ordre 1-$\alpha/2$ de la loi normale centrée réduite avec un risque $\alpha = 0.05$. On a observé la valeur
$$\overline{x} = 51.2 \in \overline{RH_0}.$$
L'hypothèse H_0 n'est donc pas rejettée : $\mu = \mu_0 = 50\,\mu g/g$ est acceptable.

5. L'IC_{95} d'une variable gaussienne est donné par
$$\left[\overline{x} - z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}; \overline{x} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right] = \left[51.2 - 1.96 \times \frac{10}{\sqrt{100}}; 51.2 + 1.96 \times \frac{10}{\sqrt{100}}\right] = [49.24; 53.16]$$
dans notre cas. Il y a 95% de chance de trouver la moyenne $\mu_0$de la population dans cet intervalle, étant donné la valeur de la moyenne observée $\overline{x} = 51.2$. Ce qui est en adéquation avec le test précédent.

Correction exercice n°1.

1. On suppose que $X \rightsquigarrow \mathcal{N}\left(\mu, \sigma^2\right)$, les paramètres de la gaussienne étant inconnus. On suppose l'échantillon $\{X_1, \cdots, X_6\}$ *i.i.d* et de même loi mère que X. On utilise les estimateurs classiques : pour estimer la moyenne populationnelle μ, on utilise la moyenne empirique $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ et pour estimer la variance σ^2, l'estimateur sans biais $S_{n-1}^2 = \frac{1}{n-1}\sum_i \left(X_i - \overline{X}\right)^2$, avec $n = 6$. On sait que la variable

$$Z = \frac{(n-1)\, S_{n-1}^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2.$$

   On veut tester l'hypothèse $H_0 : \sigma^2 = \sigma_0^2$ contre l'alternative $H_1 : \sigma^2 \neq \sigma_0^2$ avec $\sigma_0^2 = 25\,\mu g/l$. Sous H_0, la variable

$$Z = \frac{5 \times S_5^2}{\sigma_0^2} \rightsquigarrow \chi_5^2.$$

   Sous H_1, la variable S_5^2 et donc Z prendront des valeurs plus petites ou plus grandes puisque la machine serait mal calibrée. On est amené à construire un test bilatéral avec zone de rejet à gauche et à droite. Fixons le niveau du test à $\alpha = 0.05$. Les bornes de rejet du test sont données par le quantile d'ordre $\alpha/2$ du χ_5^2, c'est à dire $\chi_{5;0.025}^2 = 0.83$ et par le quantile d'ordre $1-\alpha/2$ soit $\chi_{5;0.975}^2 = 12.83$. La zone de non-rejet de H_0 est donc donnée par

$$\overline{RH_0} = [0.83;\, 12.83].$$

   On a observé les valeurs $\overline{x}_{obs} = 103.63\,\mu g/l$, $s_{obs}^2 = \frac{128}{5}\,mg^2/l^2$ et on en déduit $\chi_{obs}^2 = \frac{128}{25} = 5.12$. On constate que $\chi_{obs}^2 \in \overline{RH_0}$: l'hypothèse H_0 n'est pas rejetée. Avec une probabilité de 0.95, il n'y a pas de raison de considérer que la machine est mal calibrée.

2. On sait que $Z \rightsquigarrow \chi_5^2$. On peut calculer la probabilité que Z soit comprise entre deux quantiles d'ordre fixé :

$$P\left(\chi_{5;\alpha/2}^2 \leq Z \leq \chi_{5;1-\alpha/2}^2\right) = 1 - \alpha.$$

   Si on fixe le risque $\alpha = 0.05$, on obtient $\chi_{5;0.025}^2 = 0.83$ et $\chi_{5;0.975}^2 = 12.83$. L'$IC_{0.95}$ est donc le suivant

$$0.83 \leq \frac{5S_5^2}{\sigma^2} \leq 12.83$$
$$\frac{5S_5^2}{12.83} \leq \sigma^2 \leq \frac{5S_5^2}{0.83}.$$

   Les bornes de cet intervalle sont aléatoires et dépendent de la valeur prise par S_5^2. On a observé $s_{obs}^2 = \frac{128}{5}$ et donc

$$9.98 = \frac{128}{12.83} \leq \sigma^2 \leq \frac{128}{0.83} = 154.22.$$

   Il y a donc 95 % de chance d'avoir un écart-type de la population compris entre ces deux valeurs.