

EXAMEN DE STATISTIQUES

Tous documents interdits - Calculatrice non programmable autorisée

Questions de cours

Formulez des réponses courtes aux 3 questions suivantes.

1) Voici plusieurs variables. Indiquez, pour chacune d'entre elles, la notation utilisée en cours et s'il s'agit d'une variable aléatoire ou pas.

- Ecart-type de la moyenne empirique
- Moyenne sur la population
- Espérance de la variable aléatoire parente

2) Décrivez ce que signifie le quantile de niveau 0,95 d'une loi normale centrée réduite (phrase et équation) ?

3) A quoi sert la méthode du maximum de vraisemblance ?

Répondez de façon détaillée à la question 4 suivante.

4) Expliquez la logique d'un test statistique de conformité à une valeur de type

$$\begin{cases} H_0: \theta = 5 \\ H_1: \theta = 10 \end{cases}$$

Plus précisément à quoi correspondent les notions d'erreurs de type I, II, la puissance du test et expliquez comment la région critique est calculée.

Exercice 1

Une étude a montré que la durée de réponse à un test d'anglais contenant 20 questions parmi l'ensemble des lycéens de France est en moyenne de 45 minutes avec un écart-type de 14 minutes. On souhaite calibrer la durée de la prochaine épreuve. En supposant que les durées de réponse sont identiquement et indépendamment distribués selon une loi normale, indiquez (*sans démontrer les lois*) :

1) Quelle est la probabilité qu'un lycéen tiré au hasard réponde en plus de 60 minutes.

2) Quelle est la probabilité pour que dans un échantillon aléatoire simple de taille $n=30$, la durée moyenne de réponse soit inférieure à 40 minutes

3) Comment calibrer la durée **maximale** de l'épreuve pour que 99% des étudiants aient suffisamment de temps pour répondre au test ?

4) Définir un seuil de durée tel que la durée moyenne de réalisation de l'épreuve soit inférieure à ce seuil avec une probabilité de 99%.

Indication :

Pour cette dernière question et uniquement pour celle-ci vous devez définir un intervalle de pari unilatéral.

Exercice 2 : Panneaux Solaires

Une entreprise de panneaux solaires dispose de points de vente répartis sur la France entière. Chaque vente de panneaux solaires rapporte à l'entreprise 3000 € environ en moyenne en région parisienne entre 2013 et 2016 et 2500 € environ en moyenne dans les autres régions (province). Le suivi des ventes obtenu à partir d'un échantillon de magasins (Tableau 1a) indique une progression hétérogène au sein des régions depuis 2010. Sur la base des tendances passées, la PDG anticipe une augmentation du chiffre d'affaire issu de ses points de vente en province de l'ordre de 20% entre 2016 et 2020. Toutefois, le directeur des ventes objecte que la hausse du chiffre d'affaire peut être imputable en partie à une hausse des coûts moyens de production répercutés dans les prix (tableau 1d). En outre, les évolutions des ventes sont calculées sur un échantillon restreint de magasins en province ce qui est problématique : la valeur produite pour l'année 2016 ne lui semble pas « réaliste ». Son intuition est que les quantités vendues par l'ensemble des magasins ne sont pas différentes de 2013.

Qui a raison : le PDG ou la directrice des ventes ? Répondre en calculant un intervalle de confiance de niveau 95% pour le nombre de ventes en Province en 2016.

Pour rédiger votre réponse respectez les étapes de rédaction et de notations suivantes :

Étape 1 : Définissez les populations mères et le caractère statistique étudié

Étape 2 : Modélisez : Définissez la variable aléatoire X , l'échantillon et spécifiez le processus générateur des données de loi inconnue mais d'espérances μ et d'écart-type σ . Par souci de simplification, on supposera que l'écart-type est connu : $\sigma = 1130$.

Étape 3 : Donnez un estimateur adéquat pour la moyenne et indiquez sa loi de probabilité (preuve *complète*)

Étape 4 : Calculez l'intervalle de confiance, fournissez une phrase d'interprétation et conclure

Tab. 1a) Nombre de magasins

	2013	2014	2015	2016	2017
Province	25	24	25	26	23
Paris	32	31	32	33	30
France	97	95	97	99	98

Tab 1c) Chiffre d'Affaire en Millions d'€

	2013	2014	2015	2016	Moyenne	Evolution
Province	442	440	440	547	467	23,89%
Paris	105	111	122	124	115	18,37%
France	976	994	1027	1192	1047	22,13%

Tab 1b) Nombre de Ventes (en moyenne) par magasin

	2013	2014	2015	2016	Evolution
Province	2830	2835	2751	3150	11,31%
Paris	1127	1226	1224	1230	9,14%
France	3957	4061	3975	4380	10,69%

Tab 1d) Prix de vente par magasin (Chiffre d'Affaire moyen)

	2013	2014	2015	2016	Moyenne	Evolution
Province	2402	2424	2460	2633	2480	9,62%
Paris	2900	2930	3120	3050	3000	5,17%
France	2544	2577	2663	2750	2633	8,11%

Variable de FISHER : $F \approx F(n_1, n_2)$

$$F = \frac{Y_1/n_1}{Y_2/n_2} \quad \text{ou} \quad Y_1 \approx \chi^2(n_1) \text{ et } Y_2 \approx \chi^2(n_2) \text{ indépendants}$$

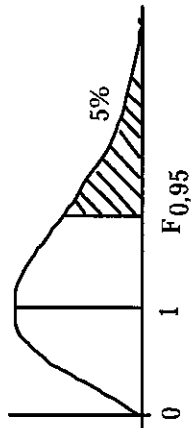


Table du quantile $F_{0,95}$ en fonction de n_1 et n_2 : $P(F > F_{0,95}) = 5\%$

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10	12	16	20	24	30	40	50	100	200	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,43	19,4	19,5	19,5	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,69	8,66	8,64	8,62	8,59	8,58	8,56	8,54	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,84	5,80	5,77	5,75	5,72	5,70	5,66	5,65	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,60	4,56	4,53	4,50	4,46	4,44	4,40	4,38	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,92	3,87	3,84	3,81	3,77	3,75	3,71	3,69	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,49	3,44	3,41	3,38	3,34	3,32	3,28	3,25	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,20	3,15	3,12	3,08	3,04	3,03	2,98	2,96	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	2,98	2,94	2,90	2,86	2,83	2,80	2,76	2,73	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,82	2,77	2,74	2,70	2,66	2,64	2,59	2,56	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,70	2,65	2,61	2,57	2,53	2,50	2,45	2,42	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,60	2,54	2,51	2,47	2,43	2,40	2,35	2,32	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,51	2,46	2,42	2,38	2,34	2,32	2,26	2,24	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,44	2,39	2,35	2,31	2,27	2,24	2,19	2,16	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,39	2,33	2,29	2,25	2,20	2,18	2,12	2,10	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,33	2,28	2,24	2,19	2,15	2,13	2,07	2,04	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,29	2,23	2,19	2,15	2,10	2,08	2,02	1,99	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,25	2,19	2,15	2,11	2,06	2,04	1,98	1,95	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,21	2,16	2,11	2,07	2,03	2,00	1,94	1,91	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,18	2,12	2,08	2,04	1,99	1,96	1,90	1,87	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,15	2,10	2,05	2,01	1,96	1,93	1,87	1,84	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,13	2,07	2,03	1,98	1,94	1,91	1,84	1,81	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,10	2,05	2,01	1,96	1,91	1,88	1,82	1,79	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,09	2,03	1,98	1,94	1,89	1,86	1,80	1,76	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,06	2,01	1,96	1,92	1,87	1,84	1,77	1,74	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	1,99	1,93	1,89	1,84	1,79	1,76	1,69	1,66	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,90	1,84	1,79	1,74	1,69	1,66	1,59	1,55	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,95	1,85	1,78	1,74	1,69	1,63	1,60	1,52	1,48	1,44
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,85	1,75	1,68	1,63	1,57	1,51	1,48	1,39	1,34	1,28
200	3,89	3,04	2,65	2,41	2,26	2,14	2,05	1,98	1,92	1,87	1,80	1,69	1,62	1,57	1,52	1,45	1,42	1,32	1,26	1,19
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,64	1,57	1,52	1,46	1,40	1,35	1,24	1,17	1,00

Correction Session 1 (Eco G)

Questions de cours

- ① - Ecart type de la moyenne empirique: $\frac{\sigma}{\sqrt{n}}$ non aléatoire
 - Moyenne sur la population: m non aléatoire
 - Espérance de la variable aléatoire parente: μ non aléatoire
- ② $z_{0,95}$ ou $z_{95\%}$ le quantile de niveau 0,95 d'une loi normale centrée réduite: $z_{0,95} = \Phi^{-1}(0,95)$ est la valeur, z , telle qu'il y a 95% de chances pour que la v.a. X réalise en-dessous de $z_{0,95}$.
- ③ la méthode du maximum de vraisemblance permet de construire des estimateurs, de trouver leur forme fonctionnelle prédéfinie.
- ④ VOIR COURS.

Exercice 1

la population est connue: ens. de lycéens de France.

X^A caractéristique stat: durée de réponse au test

unité stat: un lycéen.

m : moyenne de $X^A = 45$ min

σ : écart type de $X^A = 14$ min.

Ex 1 (suite)

soit la v. a. X_i : durée de réponse d'un lycéen à titre au hasard.

$$X_i \sim N(\mu, \sigma) \quad \text{où } \mu = m = 45 \\ \text{et } \sigma = e = 14.$$

$$\textcircled{1} P(X_i > 60) = P\left(z > \frac{60 - \mu}{\sigma}\right) \quad \text{où } z = \frac{X_i - \mu}{\sigma} \sim N(0,1).$$

$$\Rightarrow P\left(z > \frac{60 - 45}{14}\right) = P\left(z > \frac{15}{14}\right) = P(z > 1,07) = 1 - P(z < 1,07) \\ = 1 - \Phi(1,07).$$

Dans la table $\Phi(1,07) = 0,8577$ Donc $P(X_i > 60) = 0,1423$

\textcircled{2} on sait que $\bar{X}_m \sim N\left(\mu, \frac{\sigma}{\sqrt{m}}\right)$ où $\mu = 45$, $\sigma = 14$ et $m = 30$

$$P(\bar{X}_m < 40) = P\left(\frac{\bar{X}_m - \mu}{\frac{\sigma}{\sqrt{m}}} < \frac{40 - \mu}{\frac{\sigma}{\sqrt{m}}}\right) = P\left(z < \frac{40 - 45}{\frac{14}{\sqrt{30}}}\right)$$

$$= P(z < -1,96) = P(z > 1,96) = 1 - \Phi(1,96).$$

Dans la table $\Phi(1,96) = 0,9750 \Rightarrow P(\bar{X}_m < 40) = 0,0250 = 2,5\%$

\textcircled{3} soit max la durée telle que

$$P(X_i < \text{max}) = 99\% = 0,99.$$

$$\Leftrightarrow P\left(\frac{X_i - \mu}{\sigma} < \frac{\text{max} - \mu}{\sigma}\right) = 0,99 = \Phi\left(\frac{\text{max} - \mu}{\sigma}\right)$$

$$\frac{\text{max} - \mu}{\sigma} = \Phi^{-1}(0,99) \Rightarrow \text{max} = \Phi^{-1}(0,99) \times \sigma + \mu \\ = 2,33 \times 14 + 45 = \underline{\underline{77,62 \text{ euros}}}$$

Ex 1 (suite)

④ soit "seuil" la valeur telle que $P(\bar{X}_m < \text{seuil}) = 0,99$

$$\Leftrightarrow P\left(\frac{\bar{X}_m - \mu}{\frac{\sigma}{\sqrt{m}}} < \frac{\text{seuil} - \mu}{\frac{\sigma}{\sqrt{m}}}\right) = P\left(Z < \frac{\text{seuil} - \mu}{\frac{\sigma}{\sqrt{m}}}\right) = \Phi\left(\frac{\text{seuil} - \mu}{\frac{\sigma}{\sqrt{m}}}\right)$$

$$\frac{\text{seuil} - \mu}{\frac{\sigma}{\sqrt{m}}} = \Phi^{-1}(0,99) \Leftrightarrow \text{seuil} = \Phi^{-1}(0,99) \times \frac{\sigma}{\sqrt{m}} + \mu$$
$$= 2,33 \times \frac{14}{\sqrt{30}} + 45 = \underline{\underline{50,9}} \text{ euron}$$

Exercice 2

Etape 1: Pop. Réel: ens. de magasins en 2016 en Provence.

Caractéristique X^i : Nbre de ventes

unité statistique: 1 magasin

moyenne μ inconnue

écart type σ connue ($1130 = \sigma$).

Etape 2: la v. a. X_i : Nbre de ventes du magasin i en 2016 en Provence

X_i a loi inconnue ($\mu, \sigma = 1130$) $\mu = m$ inconnue.

on dispose d'un échantillon iid (X_1, \dots, X_m) où $m = 26$.

la moyenne d'échantillon $\bar{X}_m = 3150$.

Etape 3: la moyenne empirique est une estimation de l'espérance μ

\bar{X}_m permet donc d'approcher $\mu = m$.

Preuve: $E(\bar{X}_m) = \mu$ en effet: $E(\bar{X}_m) = E\left(\frac{1}{m} \sum_{i=1}^m X_i\right) = \frac{1}{m} \sum_{i=1}^m E(X_i)$

or $E(X_i) = \mu \Rightarrow E(\bar{X}_m) = \frac{1}{m} \sum_{i=1}^m \mu = \frac{m\mu}{m} = \mu$

estimateur sans biais.

Ex 2 (suite)

Soit une suite iid de var. aléa (X_1, \dots, X_m) d'espérance μ et d'écart-type σ alors, par le théorème central limite :

$$\bar{X}_m \underset{\infty}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{m}}\right)$$

Preuve pour la variance :

$$V(\bar{X}_m) = V\left(\frac{1}{m} \sum_{i=1}^m X_i\right) = \frac{1}{m^2} V\left(\sum_{i=1}^m X_i\right) = \frac{1}{m^2} \sum_{i=1}^m V(X_i)$$

Car indépendance

$$\Rightarrow V(\bar{X}_m) = \frac{1}{m^2} \sum_{i=1}^m \sigma^2 = \frac{m \times \sigma^2}{m^2} = \frac{\sigma^2}{m}$$

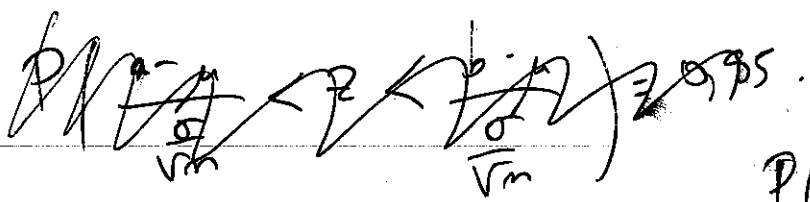
$$\Rightarrow \sigma(\bar{X}_m) = \sqrt{V(\bar{X}_m)} = \frac{\sigma}{\sqrt{m}}$$

Étape 4 :

on cherche un intervalle de confiance de niveau 95% pour μ

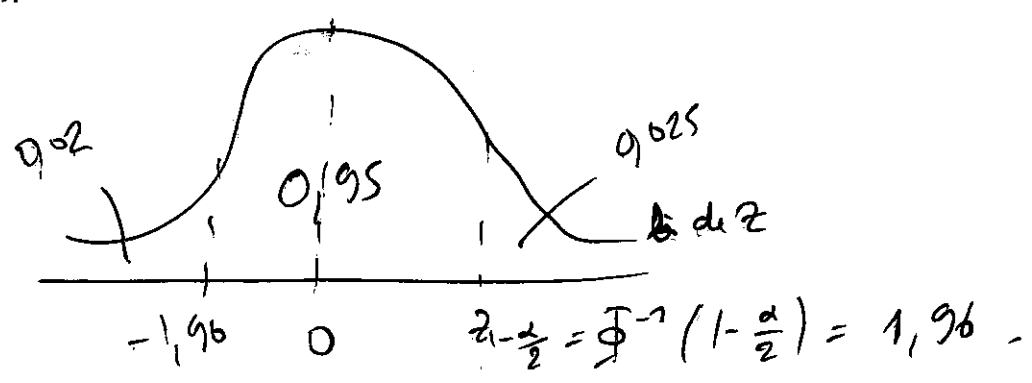
(a, b) tq $P(a < \mu < b) = 0,95$.

• on centre et écarte l'estimateur : $Z = \frac{\bar{X}_m - \mu}{\frac{\sigma}{\sqrt{m}}} \sim N(0,1)$



$$P\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 0,95$$

• donc :



Ex 2 (suite)

$$P(-1,96 < \frac{\bar{X}_m - \mu}{\frac{\sigma}{\sqrt{m}}} < 1,96) = 0,95$$

$$\Leftrightarrow P\left(\underbrace{-1,96 \times \frac{\sigma}{\sqrt{m}} + \bar{X}_m}_a < \mu < \underbrace{1,96 \times \frac{\sigma}{\sqrt{m}} + \bar{X}_m}_b\right) = 0,95.$$

$$a = 3150 - 1,96 \times \frac{1130}{\sqrt{26}} = 2715,6$$

$$b = 3150 + 1,96 \times \frac{1130}{\sqrt{26}} = 3584,3$$

Il y a 95% pour que le nombre de ventes en moyenne dans la population de l'ens. des magasins soit comprise en 2016, entre 2715 et 3584.

on constate que le nombre de vents moyen en 2013 : 2830, appartient à l'intervalle de confiance, il n'y a donc pas eu de progression des ventes avec 5% de chances de se tromper. la direction des ventes a raison.