Echantillonange et Estimation

Pr: Youssef TIDLI

Loi Normale

Loi Normale centrée réduite

Définition

On dit que la v. a X suit loi normale centrée réduite, notée $\mathcal{N}(0,1)$, si X prend ses valeurs dans tout $\mathbb{R}$ et que sa fonction de densité est

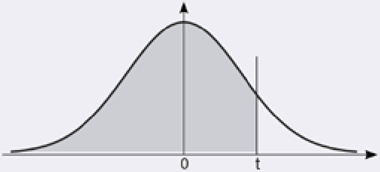
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- $\mathbb{E}(X) = 0$ et $\mathrm{Var}(x) = 1$.
- La fonction de répartition de X , est alors

$$\phi(x) = \mathbb{P}\left(X \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Cette fonction ne peut s'exprimer, ni se calculer de façon simple mais ses valeurs sont tabl´

Table statistique de la loi Normale centrée réduite



	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015

Table statistique de la loi Normale centrée réduite

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952

Remarque

La loi Normale est symétrique, c.à.d pour tout $x \in \mathbb{R}$

$$\mathbb{P}(X \leq x) = \mathbb{P}(X \geq -x).$$

En particulier $\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = \frac{1}{2}$.
On peut aussi écrire que

$$\phi(-x) = 1 - \phi(x), \quad \forall x \in \mathbb{R}.$$

Remarque

Dans une table statistique de la loi centrée réduite, on ne peut lire que $\phi(x) = \mathbb{P}(X \leq x)$ *pour* $x \geq 0$.
Comment calculer par exemple $\mathbb{P}(X \geq -1,87)$ *et* $\mathbb{P}(X \leq -0,93)$ *?*
Il suffit d'appliquer la remarque précédente et utiliser la table statistique :

- *D'une part,* $\mathbb{P}(X \geq -1,87) = \mathbb{P}(X \leq 1,87) = 0,97$.
- *D'autre part,*

$$\begin{aligned} \mathbb{P}(X \leq -0,93) = \mathbb{P}(X \geq 0,93) &= 1 - \mathbb{P}(X < 0,93) \\ &= 1 - 0,82 = 0,18 \end{aligned}$$

Exercice

Soit X une variable normale centrée réduite.

1. *À partir de la table statistique de la loi $\mathcal{N}(0,1)$, trouver $\mathbb{P}(X \leq 1,74)$ et $\mathbb{P}(X \leq 0,96)$.*
2. *En déduire*
   1. $\mathbb{P}(X \geq -1,74)$ *et* $\mathbb{P}(X < -0,96)$.
   2. $\mathbb{P}(0,96 < X \leq 1,74)$ *et* $\mathbb{P}(X < -0,96$ *ou* $X > 1,74)$.

Loi Normale

Définition

On dit que la v.a. X suit loi normale d'espérance $\mu \in \mathbb{R}$ et d'écart-type $\sigma > 0$, notée $\mathcal{N}(\mu, \sigma)$, si la v.a. $Y = \dfrac{X - \mu}{\sigma}$ suit loi normale centrée réduite.
Sa fonction de densité est donnée par

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

La fonction de répartition F de X, de loi $\mathcal{N}(\mu, \sigma)$ s'obtient à partir de ϕ, fonction de répartition de $\mathcal{N}(0, 1)$ de la façon suivant

$$\mathrm{F}(x) = \phi\left(\frac{x-\mu}{\sigma}\right)$$

Loi Normale

Proposition

- *Si N suit une loi $\mathcal{N}(0,1)$ alors la v.a. $X = \sigma N + \mu$ suit une loi $\mathcal{N}(\mu, \sigma)$.*
- *Si X suit $\mathcal{N}(\mu, \sigma)$ et si $Z = aX + b$ pour $a \neq 0$ et $b \in \mathbb{R}$ alors Z suit $\mathcal{N}(a\mu + b, a^2\sigma^2)$.*
- *Soit X_1 et X_2 deux variables aléatoires <u>indépendantes</u> telles que $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$. Alors $X_1 + X_2$ est aussi une variable normale de moyenne $\mu_1 + \mu_2$ et de variance $\sigma_1^2 + \sigma_2^2$ càd $X_1 + X_2 \sim \mathcal{N}\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$.*

Calcul de probabilité pour les v.a. Normales

On peut exprimer la fonction de répartition d'une variable X qui suit une loi $\mathcal{N}(\mu, \sigma)$ par

$$\begin{aligned} F_x(t) &= \mathbb{P}(X \leq t) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{t-\mu}{\sigma}\right) \\ &= \mathbb{P}\left(N \leq \frac{t-\mu}{\sigma}\right) =: \phi\left(\frac{t-\mu}{\sigma}\right) \end{aligned}$$

où $N \sim \mathcal{N}(0,1)$.
Conclusion : Pour calculer les probabilités de la loi $X \sim \mathcal{N}(\mu, \sigma)$, il suffit de la transformer en une variable $N \sim \mathcal{N}(0,1)$ puis utiliser la table statistique de la loi normale centrée r´

Exercice

La quantité annuelle de précipitation (en mm) suit une loi Normale de moyenne $\mu = 150$ *et de variance* $\sigma^2 = 64$*. Quelle est la probabilité d'avoir une quantité de pluie supérieure à* 140 *mm ?.*

Exercice

Imaginons que l'on étudie chez une petite entreprise le fond propre dont le montant suit une loi normale de moyenne $\mu = 15000$ *et d'écart-type* $\sigma = 4870$*.*

1. *Quelle est la probabilité que l'entreprise ait un fond propre supérieure à* $16387,95$ *?*
2. *Quelle est la probabilit´*

Lois issues de la loi Normale

Loi Khi-deux à n degré de liberté $\chi^2(n)$

Cette loi joue un rôle important dans la théorie des tests statistiques. La loi Khi-deux est obtenue en additions des carrées de variables aléatoires Normales, alors elle ne prend que des valeurs positives.

Définition

Soient $X_1, ..., X_n$ des variables aléatoires indépendantes de loi Normale $\mathcal{N}(0,1)$. Alors

$$X_1^2 + X_2^2 + X_3^2 + ... + X_n^2$$

suit une loi Khi-deux de n degrés de liberté notée χ^2

La loi $\chi^2(n)$ a la fonction de densité suivante

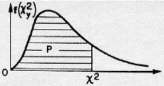
$$f_{\chi^2(n)}(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{n/2-1} e^{-x/2}.$$

où la fonction Γ est définie par $\Gamma(n) = \int_0^{+\infty} t^{n-1} e^{-t} dt$.

Proposition

$\mathbb{E}(X) = n$ *et* $Var(X) = 2n$.

Table statistique de la loi $\chi^2(n)$



P	0,00050	0,0010	0,0050	0,010	0,0250	0,050	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	0,950	0,9750	0,990	0,9950	0,9990	0,99950
ν																					
1	0,000000393	0,00000157	0,0000393	0,000157	0,000982	0,00393	0,0158	0,0642	0,148	0,275	0,455	0,708	1,074	1,642	2,706	3,841	5,024	6,635	7,879	10,828	12,116
2	0,00100	0,00200	0,0100	0,0201	0,0506	0,103	0,211	0,446	0,713	1,022	1,386	1,833	2,408	3,219	4,605	5,991	7,378	9,210	10,597	13,816	15,202
3	0,0153	0,0243	0,0717	0,115	0,216	0,352	0,584	1,005	1,424	1,869	2,366	2,946	3,665	4,642	6,251	7,815	9,348	11,345	12,838	16,266	17,730
4	0,0639	0,0908	0,207	0,297	0,484	0,711	1,064	1,649	2,195	2,753	3,357	4,045	4,878	5,989	7,779	9,488	11,143	13,277	14,860	18,467	19,998
5	0,158	0,210	0,412	0,554	0,831	1,145	1,610	2,343	3,000	3,655	4,351	5,132	6,064	7,289	9,236	11,070	12,832	15,086	16,750	20,515	22,105
6	0,299	0,381	0,676	0,872	1,237	1,635	2,204	3,070	3,828	4,570	5,348	6,211	7,231	8,558	10,645	12,592	14,449	16,812	18,548	22,458	24,103
7	0,485	0,598	0,989	1,239	1,690	2,167	2,833	3,822	4,671	5,493	6,346	7,283	8,383	9,803	12,017	14,067	16,013	18,475	20,278	24,322	26,018
8	0,710	0,857	1,344	1,646	2,180	2,733	3,490	4,594	5,527	6,423	7,344	8,351	9,524	11,030	13,362	15,507	17,535	20,090	21,955	26,125	27,868
9	0,972	1,153	1,735	2,088	2,700	3,325	4,168	5,380	6,393	7,357	8,343	9,414	10,656	12,242	14,684	16,919	19,023	21,666	23,589	27,877	29,666
10	1,265	1,479	2,156	2,558	3,247	3,940	4,865	6,179	7,267	8,295	9,342	10,473	11,781	13,442	15,987	18,307	20,483	23,209	25,188	29,588	31,419
11	1,587	1,834	2,603	3,053	3,816	4,575	5,578	6,989	8,148	9,237	10,341	11,530	12,899	14,631	17,275	19,675	21,920	24,725	26,757	31,264	33,136
12	1,934	2,214	3,074	3,571	4,404	5,226	6,304	7,807	9,034	10,182	11,340	12,584	14,011	15,812	18,549	21,026	23,336	26,217	28,300	32,909	34,821
13	2,305	2,617	3,565	4,107	5,009	5,892	7,042	8,634	9,926	11,129	12,340	13,636	15,119	16,985	19,812	22,362	24,736	27,688	29,819	34,528	36,478
14	2,697	3,041	4,075	4,660	5,629	6,571	7,790	9,467	10,821	12,079	13,339	14,685	16,222	18,151	21,064	23,685	26,119	29,141	31,319	36,123	38,109
15	3,108	3,483	4,601	5,229	6,262	7,261	8,547	10,307	11,721	13,030	14,339	15,733	17,322	19,311	22,307	24,996	27,488	30,578	32,801	37,697	39,719
16	3,536	3,942	5,142	5,812	6,908	7,962	9,312	11,152	12,624	13,983	15,338	16,780	18,418	20,465	23,542	26,296	28,845	32,000	34,267	39,252	41,308
17	3,980	4,416	5,697	6,408	7,564	8,672	10,085	12,002	13,531	14,937	16,338	17,824	19,511	21,615	24,769	27,587	30,191	33,409	35,718	40,790	42,879
18	4,439	4,905	6,265	7,015	8,231	9,390	10,865	12,857	14,440	15,893	17,338	18,868	20,601	22,760	25,989	28,869	31,526	34,805	37,156	42,312	44,434
19	4,912	5,407	6,844	7,633	8,907	10,117	11,651	13,716	15,352	16,850	18,338	19,910	21,689	23,900	27,204	30,144	32,852	36,191	38,582	43,820	45,973
20	5,398	5,921	7,434	8,260	9,591	10,851	12,443	14,578	16,266	17,809	19,337	20,951	22,775	25,038	28,412	31,410	34,170	37,566	39,997	45,315	47,498
21	5,896	6,447	8,034	8,897	10,283	11,591	13,240	15,445	17,182	18,768	20,337	21,991	23,858	26,171	29,615	32,671	35,479	38,932	41,401	46,797	49,010
22	6,405	6,983	8,643	9,542	10,982	12,338	14,041	16,314	18,101	19,729	21,337	23,031	24,939	27,301	30,813	33,924	36,781	40,289	42,796	48,268	50,511
23	6,924	7,529	9,260	10,196	11,688	13,091	14,848	17,187	19,021	20,690	22,337	24,069	26,018	28,429	32,007	35,172	38,076	41,638	44,181	49,728	52,000
24	7,453	8,085	9,886	10,856	12,401	13,848	15,659	18,062	19,943	21,652	23,337	25,106	27,096	29,553	33,196	36,415	39,364	42,980	45,558	51,179	53,479
25	7,991	8,649	10,520	11,524	13,120	14,611	16,473	18,940	20,867	22,616	24,337	26,143	28,172	30,675	34,382	37,652	40,646	44,314	46,928	52,620	54,947
26	8,538	9,222	11,160	12,198	13,844	15,379	17,292	19,820	21,792	23,579	25,336	27,179	29,246	31,795	35,563	38,885	41,923	45,642	48,290	54,052	56,407
27	9,093	9,803	11,808	12,879	14,573	16,151	18,114	20,703	22,719	24,544	26,336	28,214	30,319	32,912	36,741	40,113	43,194	46,963	49,645	55,476	57,858
28	9,656	10,391	12,461	13,565	15,308	16,928	18,939	21,588	23,647	25,509	27,336	29,249	31,391	34,027	37,916	41,337	44,461	48,278	50,993	56,892	59,300
29	10,227	10,986	13,121	14,256	16,047	17,708	19,768	22,475	24,577	26,475	28,336	30,283	32,461	35,139	39,087	42,557	45,722	49,588	52,336	58,302	60,734
30	10,804	11,588	13,787	14,953	16,791	18,493	20,599	23,364	25,508	27,442	29,336	31,316	33,530	36,250	40,256	43,773	46,979	50,892	53,672	59,703	62,161
31	11,389	12,196	14,458	15,655	17,539	19,281	21,434	24,255	26,440	28,409	30,336	32,349	34,598	37,359	41,422	44,985	48,232	52,191	55,003	61,098	63,582
32	11,979	12,811	15,134	16,362	18,291	20,072	22,271	25,148	27,373	29,376	31,336	33,381	35,665	38,466	42,585	46,194	49,480	53,486	56,328	62,487	64,995
33	12,576	13,431	15,815	17,073	19,047	20,867	23,110	26,042	28,307	30,344	32,336	34,413	36,731	39,572	43,745	47,400	50,725	54,776	57,648	63,870	66,402
34	13,179	14,057	16,501	17,789	19,806	21,664	23,952	26,938	29,242	31,313	33,336	35,444	37,795	40,676	44,903	48,602	51,966	56,061	58,964	65,247	67,803
35	13,788	14,688	17,192	18,509	20,569	22,465	24,797	27,836	30,178	32,282	34,336	36,475	38,859	41,778	46,059	49,802	53,203	57,342	60,275	66,619	69,198
36	14,401	15,324	17,887	19,233	21,336	23,269	25,643	28,735	31,115	33,252	35,336	37,505	39,922	42,879	47,212	50,998	54,437	58,619	61,581	67,985	70,588
37	15,020	15,965	18,586	19,960	22,106	24,075	26,492	29,635	32,053	34,222	36,336	38,535	40,984	43,978	48,363	52,192	55,668	59,892	62,883	69,346	71,972
38	15,644	16,611	19,289	20,691	22,878	24,884	27,343	30,537	32,992	35,192	37,335	39,564	42,045	45,076	49,513	53,384	56,895	61,162	64,181	70,703	73,351
39	16,273	17,261	19,996	21,426	23,654	25,695	28,196	31,441	33,932	36,163	38,335	40,593	43,105	46,173	50,660	54,572	58,120	62,428	65,476	72,055	74,725
40	16,906	17,916	20,707	22,164	24,433	26,509	29,051	32,345	34,872	37,134	39,335	41,622	44,165	47,269	51,805	55,758	59,342	63,691	66,766	73,402	76,095
41	17,544	18,575	21,421	22,906	25,215	27,326	29,907	33,251	35,813	38,105	40,335	42,651	45,224	48,363	52,949	56,942	60,561	64,950	68,053	74,745	77,459
42	18,186	19,238	22,138	23,650	25,999	28,144	30,765	34,157	36,755	39,077	41,335	43,679	46,282	49,456	54,090	58,124	61,777	66,206	69,336	76,084	78,820
43	18,832	19,905	22,859	24,398	26,785	28,965	31,625	35,065	37,698	40,050	42,335	44,706	47,339	50,548	55,230	59,304	62,990	67,459	70,616	77,418	80,176
44	19,482	20,576	23,584	25,148	27,575	29,787	32,487	35,974	38,641	41,022	43,335	45,734	48,396	51,639	56,369	60,481	64,201	68,709	71,893	78,749	81,528
45	20,136	21,251	24,311	25,901	28,366	30,612	33,350	36,884	39,585	41,995	44,335	46,761	49,452	52,729	57,505	61,656	65,410	69,957	73,166	80,077	82,876
46	20,794	21,929	25,041	26,657	29,160	31,439	34,215	37,795	40,529	42,968	45,335	47,787	50,507	53,818	58,641	62,830	66,617	71,201	74,437	81,400	84,220
47	21,456	22,610	25,774	27,416	29,956	32,268	35,081	38,708	41,474	43,942	46,335	48,814	51,562	54,906	59,774	64,001	67,821	72,443	75,704	82,720	85,560
48	22,121	23,295	26,511	28,177	30,755	33,098	35,949	39,621	42,420	44,915	47,335	49,840	52,616	55,993	60,907	65,171	69,023	73,683	[illegible]	[illegible]	[illegible]

Loi de Student à n degré de liberté $\mathcal{T}(n)$

Cette loi joue un rôle important dans l'estimation par intervalle de confiance. Elle est symétrique, de moyenne nulle et dépend d'un seul paramètre n appelé nombre de degrés de liberté.

Définition

Soit $X \sim \mathcal{N}(0,1)$ *et* $Y \sim \chi^2(n)$*, alors la variable*

$$T = \frac{X}{\sqrt{Y/n}}$$

suit une loi dite de Student, notée $\mathcal{T}(n)$*.*

La fonction densité d'une loi de Student à n degré de liberté $\mathcal{T}(n)$ est

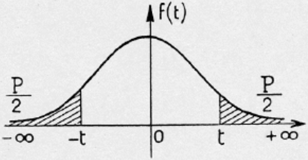
$$f_{\mathcal{T}_n}(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}.$$

Proposition

Si X suit une loi de Student $\mathcal{T}_n$, alors

1. $\mathbb{E}(X) = 0$ *si* $n > 1$.
2. $Var(X) = \frac{n}{n-2}$, *si* $n > 2$

Table statistique de la loi $\mathcal{T}_n$



ν \ P	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850

Théorème central limite

Théorème (Théorème central limite)

Si $X_1, X_2, ..., X_n$ sont des variables aléatoires (discrètes ou continues) indépendantes, de même loi, d'espérance μ et d'écart-type σ , on démontre que, quand $n \longrightarrow +\infty$, la loi de :

$$\frac{X_1 + X_2 + ... + X_n - n\mu}{\sigma\sqrt{n}}$$

tend vers la loi $\mathcal{N}(0, 1)$.

Conséquence du théorème central limite

Proposition (Convergence de la loi Binômiale vers la loi Normale)

Si X suit une loi $\mathcal{B}(n, p)$ avec n est grand (en pratique dès que $np > 18$), alors la loi de X tend vers une loi Normale $\mathcal{N}(np, np(1-p))$.

Remarque

En pratique, on peut approcher une loi Binômiale (n, p) par une loi Normale $(np, \sqrt{np(1-p)})$ lorsque

- $n \gg 30$ *et* $0,1 \leq p \leq 0,9$.
- $np(1-p) > 9$.
- $np > 5$ *ou* $n(1-p) > 5$.

Conséquence du théorème central limite

Proposition (Convergence de la loi khi-deux vers la loi Normale)

Soit X une variable aléatoire de loi $\chi^2(n)$, alors, quand n devient grand ($n \to +\infty$),

$$\frac{X-n}{\sqrt{2n}} \longrightarrow \mathcal{N}(0,1),$$

ou bien

$$X \approx \mathcal{N}(n, \sqrt{2n}).$$

(en pratique l'approximation est satisfaisante quand $n \ggg 30$)

Proposition (Convergence de la loi Student vers la loi Normale)

Soit X une variable aléatoire de loi $\mathcal{T}_n$, alors, quand n devient grand ($n \to +\infty$),

$$X \longrightarrow \mathcal{N}(0,1),$$

(en pratique l'approximation est satisfaisante quand $n \ggg 30$)

Échantillonnage

Introduction

Le calcul des probabilités apporte les outils nécessaires aux techniques de la statistique mathématique, c'est à dire les modèles qui vont être utilisés pour décrire des phénomènes réels où le hasard intervient.

Exemples :

1.
   - Étude du nombre de vacanciers pendant une période déterminée dans la station "Lexus" $\Longrightarrow$ Statistique Descriptive.
   - Prévoir le nombre de lits nécessaires pour l'hébergement $\Longrightarrow$ Statistique Mathématique.
2.
   - Étude de données économiques sur les dépenses des ménages $\Longrightarrow$ Statistique Descriptive.
   - Prévoir l'évolution de la vente d'un produit $\Longrightarrow$ Statistique Mathématique.

Introduction

En résumé :

- La mise en ordre des données relève des techniques de la statistique descriptive (caractéristiques numériques ou graphique).
- La prévision de l'évolution d'un phénomène réel, à partir des données numériques et des lois de probabilité théoriques, relève de la statistique mathématique.

Introduction

Une étude statistique portant sur tous les éléments d'une population peut être impossible à réaliser pour divers raisons :

- la population considérée peut contenir une infinité d'unités,
- le coût d'une mesure est considérable,
- la mesure peut détruire dans certains cas l'objet mesuré ;

Question : Comment obtenir des résultats fiables sur les caractéristiques d'une population en se limitant à l'étude des éléments d'un échantillon ?.

Introduction

L'échantillonnage est bien souvent incontournable, mais il faut prendre conscience d'une de ses caractéristiques essentielles, c'est le fait qu'elle se trompe

$$\Downarrow$$

elle apporte une information partielle sur la population

$$\Updownarrow$$

Il y a toujours un écart entre le résultat obtenu sur l'échantillon et celui qu'on aurait eu si on avait mesuré toute la population.

Définition

- *Un échantillon est un sous ensemble d'unités de la population étudiée.*
- *Une base de sondage est la liste des unités de la population.*
- *La taille de la population, N, est le nombre d'unités de la population.*
- *La taille de l'échantillon, n, est le nombre d'unités de l'échantillon.*
- *Le taux de sondage est le rapport $f = n/N$.*

Échantillonnage non aléatoire

Ces méthodes sont beaucoup moins coûteuses, plus rapides et plus simples. Il est par contre, peu recommandé de généraliser les résultats provenant de ces méthodes à l'ensemble de la population, puisque toutes les unités statistiques n'ont pas la même chance d'être choisi ce qui influence la représentativité de l'échantillon.
On peut citer deux exemples de ce type d'échantillonnage :

- échantillonnage à l'aveuglette
- échantillonnage au volontariat

L'échantillonnage à l'aveuglette est une technique simple et peu coûteuse. Cet échantillonnage n'est pas normalement représentatif de la population cible, parce qu'on ne sélectionne des unités d'échantillonnage dans son cas que si on peut y avoir facilement et commodément accès. Les reporters des stations de télévision sont, en outre, souvent à la recherche de soi-disant ≪ interviews de gens de la rue ≫ pour déterminer comment la population perçoit un enjeu ou une question.

C'est une des méthodes les plus utilisées actuellement sur le marché des médicaments. Les compagnies pharmaceutiques sont les pionniers en la matière. Les unités statistiques décident de faire partie de l'étude de leur propre gré.

Échantillonnage aléatoire

Pour qu'un échantillon soit représentatif de la population, il faut que chaque individus de la population ait la même chance d'être choisit dans cet échantillon. On dit dans ce cas que l'échantillonnage est aléatoire.

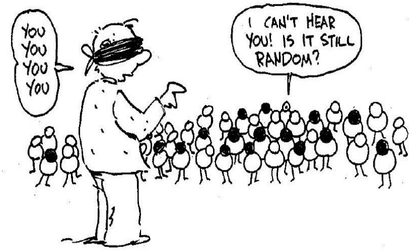


FIGURE: Echantillonnage aléatoire

Il consiste simplement à choisir des individus au hasard parmi ceux de la base de sondage (liste des individus à partir de laquelle on prélève un échantillon par exemple l'annuaire téléphonique). Les étapes sont les suivantes :

1. Numéroter les unités statistiques de 1 à N.
2. Tirer au hasard des unités statistiques de la population qui feront partie de l'échantillon.

C'est une technique où les unités statistiques sont choisis à intervalle régulier dans la base de sondage. Les étapes de cette cette technique sont les suivantes :

1. Numéroter les unités statistiques de 1 à N.
2. Calculer l'intervalle de sélection que l'on appelle aussi le pas de sondage. On le calcule en divisant la taille totale de la population observée par la taille de l'échantillon recherchée $k = \frac{N}{n}$.
3. Tirer au hasard une unité statistique entre la première et la k^{ime} unité. Par exemple la $i^{\text{ème}}$ unité avec $1 \leq i \leq k$.
4. Pour compléter l'échantillon, on choisit la $(i+k)^{ime}$ unité, et la $(i+2k)^{ime}$.....jusqu'a $(i+(n-1)k)^{ime}$. On constitue ainsi un échantillon de taille (n-1+1=n) unités.

Il consiste à choisir des groupes (toute une grappe de raisin) plutôt que de choisir des unités statistiques isolées (un seul raisin).

Définition

Une grappe est un sous-ensemble non homogènes de la population défini selon la proximité.

Il est plus facile de faire une liste des groupes et de choisir au hasard parmi ces dizaines de groupes et d'interroger toutes les unités statistiques du groupe (par exemple un groupe d'élèves faisant partie de la même classe, des habitants du même immeuble, des habitants du même quartier ou même des équipes sportives

Cette méthode permet de sauver beaucoup de temps en déplacement. Les étapes de cette techniques sont :

1. Diviser la population en grappes.
2. Choisir de façon aléatoire simple un certain nombre de grappes.
3. L'échantillon sera alors composé de toutes les unités statistiques appartenant aux grappes choisies.

la méthode peut entraîner des résultats imprécis (moins précis que les méthodes précédentes) puisque les unités voisines ont tendance se rassembler. De plus elle ne permet pas de contrôler la taille finale de l'´

L'échantillonnage stratifié est une technique qui consiste à subdiviser une population hétérogène en sous groupes plus homogènes selon un critère (Caractère qualitatif ou quantitatif : des étudiants par leur sexe, âge ou diplôme préparé, des entreprises par chiffre d'affaires ou secteur d'activités . . .) lié à la nature et aux objectifs de l'étude. Ces différents groupes sont appelés des strates.

Définition

Les strates sont des sous-ensembles de la population ayant des caractéristiques communes.

Les étapes de cette méthode sont les suivantes :

1. Diviser la population en strates mutuellement exclusives.
2. Dresser la liste la plus complète possible (base de sondage) constituant chacune des strates.
3. Proportionnellement à son importance dans la population, on calcule combien il faut d'individus au sein de l'échantillon pour représenter chaque strate.
4. Dans chaque strate, choisir de façon aléatoire simple un nombre d'unités statistiques pour constituer l'échantillon de telle sorte que le pourcentage d'unités dans chacune des strates de l'échantillon soit le plus près possible du pourcentage d'unités dans chacune des strates de la population.

Distributions d'échantillonnage

Distributions d'échantillonnage

Dans toute la suite, on traite le cas de l'échantillonnage aléatoire simple. Les concepts fondamentaux et les formules importantes découlent de cette méthode.

L'échantillonnage simple consiste à extraire un échantillon de taille n dans une population de taille N par des tirages aléatoires équiprobables et indépendants (tirages avec remise).

- $\Omega = \{\omega_1, ..., \omega_N\}$: la population constituée d'éléments appelés unités d'observation.
- X : le caractère que l'on voudrait étudier sur l'ensemble de cette population.
- X_k : le résultat aléatoire du k ème tirage est une v.a qui suit la même loi que X.
- $(X_1, ..., X_n)$: les résultats aléatoires des n tirages modélisant l'échantillon étudié.
- x_k : une réalisation du k ème tirage X_k.
- $(x_1, ..., x_n)$: une réalisation de l'échantillon $(X_1, ..., X_n)$.

Définition

$X_1, ..., X_n$ sont n v.a. indépendantes et de même loi (celle de X) ; il est appelé n-échantillon ou échantillon de taille n de X. Une réalisation $(x_1, ..., x_n)$ de l'échantillon $(X_1, ..., X_n)$ est l'ensemble des valeurs observées.

Définition

Une statistique Y sur un échantillon $(X_1, ..., X_n)$ est une v.a., fonction mesurable des X_k :

$$Y = \varphi(X_1, ..., X_n).$$

Après réalisation, la statistique Y prend la valeur $\varphi(x_1, ..., x_n)$.

Soit X le caractère quantitatif que l'on voudrait étudier sur une population infinie et soit $(X_1, X_2, ..., X_n)$ un n-échantillon de X.

Définition

La statistique $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ *est appelée moyenne empirique de* X*.*

Remarque

La moyenne empirique est une variable aléatoire qui prend des valeurs différentes sur chaque échantillon appelées moyennes observ´

Proposition

Soit X une variable aléatoire de moyenne μ et de variance σ^2 et soit $(X_1, X_2, ..., X_n)$ un n-échantillon de X. Alors

- $\mathbb{E}(\bar{X}) = \mu$,
- $Var(\bar{X}) = \dfrac{\sigma^2}{n}$.

Proposition

La distribution d'échantillonnage de la moyenne est donnée par :

- *Si X suit une loi normale $\mathcal{N}(\mu, \sigma)$, alors $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$.*
- *Si la loi de X est quelconque avec $n \ggg 30$, par le théorème central limite, $\bar{X}$ suit approximativement une loi $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}}$*

Exercice

1. *X La taille des étudiants de SEG(3) suit une loi $\mathcal{N}(1,62;0,2)$. On considère $(X_1, X_2, \ldots, X_{25})$ un échantillon de X. Quelle est la loi de $\bar{X}$ la taille moyenne de 25 étudiants choisis hasard ?*
2. *Dans une grande entreprise, les salaires sont distribués suivant une loi Y de moyenne 4000 et d'écart-type 1600. Soit $(Y_1, Y_2, \ldots, Y_{150})$ un échantillon de Y. Quelle est la loi de $\bar{Y}$ le salaire moyen de 150 salariés pris au hasard ?.*

Soit $(X_1, X_2, ..., X_n)$ un n-échantillon d'une variable aléatoire X de moyenne μ et d'écart-type σ. On définit la statistique

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \bar{X}^2.$$

Alors

$$\mathbb{E}(S^2) = (1 - \frac{1}{n})\sigma^2.$$

Remarque

Comme $1 - \frac{1}{n} < 1$; alors $\mathbb{E}(S^2) < \sigma^2$. En moyenne, la variance dans l'échantillon est plus faible que dans la population-m`

Proposition

Si le caractère X à étudier suit une loi normale $\mathcal{N}(\mu, \sigma)$ alors $n\dfrac{S^2}{\sigma^2}$ suit une loi de khi-deux à $(n-1)$ degrés de liberté càd

$$n\frac{S^2}{\sigma^2} \sim \chi^2(n-1).$$

Soit une population comportant deux modalités A et B.
Soit p la proportion d'individus de la population possédant la modalité A. $1 - p$ est donc la proportion des individus de la population possédant la modalité B.
On extrait de la population un échantillon de taille n.
Soit K_n la v.a qui représente le nombre d'individus dans l'échantillon ayant la modalité A.

Définition

La variable aléatoire $\hat{p} = \dfrac{K_n}{n}$ *s'appelle la fréquence empirique. Sa réalisation* f *est la proportion d'individus dans l'échantillon ayant la modalit´*

Proposition

- *Si* $n \gg 30$, $np \geq 5$ *ou* $n(1-p) \geq 5$, *par le théorème central limite,*
$$\hat{P} \sim \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$
- *Sinon (le cas ou* $n < 30$*), la variable* K_n *suit une loi Binômiale* $\mathcal{B}(n, p)$*.*

Exercice

On suppose que la distribution des salaires dans une entreprise est telle que 20% touchent moins que 2000 DH. On tire un échantillon de 100 salariés au hasard. Quelle est la loi de $\hat{p}$ *la proportion des salari´*

Estimation

Objectif :
On s'interesse à l'estimation des principales caractéristiques (ou paramètres) d'un caractère dans une population, à savoir la moyenne, la variance et la fréquence, à partir des valeurs calculées sur les échantillons.
Question :
Quelle est l'estimation la plus bonne ? Et bonne dans quel sense ?
Méthodes d'estimation :

1. Estimation ponctuelle
2. Estimation par intervalle de confiance

Les paramètres à estimer seront notés les par des lettres grecques minuscules

- μ pour la moyenne de la population.
- σ pour l'écart type de la population.
- σ^2 pour la variance de la population.
- p pour la proportion dans la population.

Estimation ponctuelle

Définition

Un estimateur $T = f(X_1, ..., X_n)$ d'un paramètre θ est une statistique et sa réalisation $f(x_1, ..., x_n)$ sera appelée estimation ponctuelle de θ.

Définition

On appelle erreur d'estimation la différence entre l'estimateur et le paramètre : Erreur $= T - \theta$.

Cette Erreur peut être décomposer de la façon suivante :

$$T - \theta = \overbrace{T - \mathbb{E}(T)}^{\text{fluctuation autour de la moyenne}} + \underbrace{\mathbb{E}(T) - \theta}_{\text{Biais de l'estimateur}}$$

Définition

1. *Un estimateur T de θ est dit* **sans biais** *si $\mathbb{E}(T) = \theta$.*
2. *Sinon, on dit que c'est un estimateur* **biaisé** *:*
    1. *Si le biais $\mathbb{E}(T) - \theta$ est* **positif**, *($\mathbb{E}(T) > \theta$), alors l'estimateur* **surestime** *la valeur du paramètre.*
    2. *Si le biais $\mathbb{E}(T) - \theta$ est* **négatif**, *($\mathbb{E}(T) < \theta$), alors l'estimateur* **sousestime** *la valeur du paramètre.*

Définition

Un estimateur T de θ est dit asymptotiquement sans biais si $\mathbb{E}(T) \to \theta$ quand $n \to \infty$.

Définition

Un estimateur T de θ est dit convergent si $Var(T) \longrightarrow 0$ quand $n \to \infty$.

Définition

- *Soit T et T' deux estimateurs sans biais de θ. On dit que T est plus efficace que T' si $Var(T) \leq Var(T')$.*
- *L'estimateur sans biais et de variance minimale est appelé* **estimateur efficace***.*

Exercice

Soit X une variable aléatoire de moyenne μ et de variance σ^2. On considère $(X_1, X_2, ..., X_n)$ est n-échantillon de X.

On pose $Y = \dfrac{X_1 + X_n}{2}$ et $Z = \dfrac{1}{3}(X_1 - 2X_2 + 3X_3 - 4X_4 + 5X_5)$.

1. *Vérifier que Y et Z sont deux estimateurs sans biais de μ.*
2. *Lequel de ces deux estimateurs est le plus efficace ?*

Soit X un caractère (une variable aléatoire) dont on veut estimer la moyenne μ à partir d'un échantillon $(X_1, ..., X_n)$ de même loi que X. La loi de X est inconnue.

Théorème

La moyenne empirique $\overline{X} = \dfrac{X_1 + X_2 + ... + X_n}{n}$ *est un estimateur efficace de la moyenne* μ.

En effet, l'estimateur $\overline{X}$ est

- sans biais car $\mathbb{E}(\overline{X}) = \mu$.
- convergeant car $Var(\overline{X}) = \dfrac{Var(X)}{n} \to 0$, quand n tend vers l'infini.
- On peut montrer qu'il est de variance minimale.

Soit X une variable aléatoire de moyenne μ et d'écart type σ). On veut estimer la variance de X. Deux cas de figure se présentent :

1. La moyenne μ de la population est connue
2. La moyenne μ est inconnue

Si la moyenne de la population est **connue** alors

Proposition

La statistique $T^2 = \dfrac{1}{n}\sum_{j=1}^{n}(X_i - \mu)^2$ *est un estimateur sans biais de la variance* σ^2.

En effet,

$$\begin{aligned}
\mathbb{E}(T^2) &= \frac{1}{n}\mathbb{E}\sum_{j=1}^{n}(X_i-\mu)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(X_i-\mu)^2, \qquad \text{avec } \mathbb{E}(X_i)=\mu, \\
&= \frac{\sigma^2+\ldots+\sigma^2}{n} = \sigma^2
\end{aligned}$$

Si la moyenne de la population est **inconnue** alors

Proposition

La statistique $S^2 = \frac{1}{n}\sum_{j=1}^{n}(X_i - \overline{X})^2$ *est un estimateur biaisé de la variance* σ^2.

En effet,

$$\mathbb{E}(S^2) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(X_i - \overline{X})^2 = \frac{n-1}{n}\sigma^2.$$

Pour corriger le biais on prend l'estimateur

$$\overline{S}^2 = \frac{n}{n-1} S^2.$$

On remarque que

$$\mathbb{E}(\overline{S}^2) = \frac{n}{n-1}\mathbb{E}(S^2) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2.$$

Proposition

La statistique (la variance corrigée)

$$\overline{S}^2 = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_i - \overline{X})^2$$

est un estimateur de la variance σ^2 qui est sans biais et convergeant.

Soit une population ayant des individus possédant une certaine caractéristique A. On veut estimer à partir d'un échantillon de taille n la proportion d'individus possédant cette caractéristique A.
Soit K la v.a qui représente le nombre d'individus dans l'échantillon possédant la caractéristique A.
On rappelle que p est la proportion d'individus de la population possédant la modalité A.

Proposition

La fréquence empirique $\widehat{p} = \frac{K}{n}$ est l'estimateur efficace de p.

En effet, $\widehat{p} = \frac{K}{n}$ est un estimateur sans biais car, comme $X_1, X_2, ..., X_n$ sont des variables de Bernoulli, alors

$$\begin{aligned}\mathbb{E}(\widehat{p}) &= \frac{\mathbb{E}(X_1) + \mathbb{E}(X_2) + ... + \mathbb{E}(X_n)}{n} \\ &= \frac{p + p + ... + p}{n} \\ &= \frac{n \times p}{n}\end{aligned}$$

En plus $\widehat{p}$ est un estimateur convergeant, car

$$\begin{aligned} Var(\widehat{p}) &= \frac{Var(X_1) + Var(X_2) + ... + Var(X_n)}{n^2} \\ &= \frac{p(1-p) + p(1-p) + ... + p(1-p)}{n^2} \\ &= \frac{p(1-p)}{n}. \end{aligned}$$

alors $Var(\widehat{p}) \longrightarrow 0$, quand $n \to +\infty$.

Exercice

Afin d'étudier le salaire horaire des ouvriers d'un secteur d'activité, on prélève un échantillon aléatoire $(X_1, X_2, ..., X_{15})$ *de réalisation (en dhs)*

9,8	8,6	9,7	10,2	8,9	10,1	9,9	9,7
8,7	9,8	10,2	9,3	10,4	9,5	10,9	

On suppose que la loi suivie par le salaire horaire est normale de moyenne μ *et variance* σ^2.

1. *Donner un estimateur sans biais de* μ *et calculer sa réalisation.*
2. *Donner un estimateur sans biais de* σ^2 *et calculer sa réalisation*
3. *Supposons que* $\mu = 9,7$. *Donner un estimateur de* σ^2 *et une*

Exercice

L'objet d'une enquête menée sur la région Larache-El Ksar El Kebir est d'étudier la variable X représentant la dépense moyenne de foyer. On choisit au hasard 100 foyers, les réalisation $(x_i)_{i=1}^{100}$ de l'échantillon $(X_i)_{i=1}^{100}$ permettent d'affirmer que

$$\sum_{i=1}^{100} x_i = 3007 \quad et \quad \frac{1}{100}\sum_{i=1}^{100} x_i^2 = 940,4.$$

1. *Donner un estimateur sans biais de la moyenne m et calculer sa réalisation $\hat{m}$.*
2. *Donner un estimateur sans biais de la variance σ^2 et calculer sa réalisation $\hat{\sigma}^2$.*
3. *En admettant que $X \sim \mathcal{N}(\hat{m}; \hat{\sigma}^2)$, calculer la probabilité que la dépense d'un foyer soit supérieur ou égale 32,65 DH.*

Exercice

Dans une étude réalisée l'an dernier, on a observé que la consommation en électricité des habitants d'une résidence composée d'appartements de trois pièces est en moyenne de 215 *dhs et d'écart-type* 27 *dhs. Cette année, sur un échantillon de* 35 *appartements, les consommation en électricité* $(x_i)_{i=1}^{35}$ *vérifient*

$$\sum_{i=1}^{35} x_i = 7700 \quad et \quad \frac{1}{35}\sum_{i=1}^{35} x_i^2 = 49007.$$

Supposons que l'échantillon $(X_i)_{i=1}^{35}$ *suit une loi* $\mathcal{N}(\mu, \sigma)$.

1. *Donner un estimateur sans biais de* μ *et calculer sa réalisation.*
2. *Donner deux estimateurs de* σ^2 *et calculer leurs r´*

Exercice

On s'interesse à la proportion p des étudiants ayant un Baccalauréat Sciences-économiques inscrit en première année à la FP Larache. On a prélevé indépendamment deux échantillons de tailles $n_1 = 120$ et $n_2 = 150$. On constate que 48 étudiants du premier échantillon et 66 du second ont une un bac Sciences économiques. Calculer 3 estimations ponctuelles de p.

Exercice

La distribution du nombre de pannes observées dans le fonctionnement d'une machine au cours de 100 journées de travail est résumée dans le tableau suivant :

Nombre de pannes	*0*	*1*	*2*	*3*	*4*	*Total*
Nombre de jours	*53*	*32*	*11*	*3*	*1*	*100*

Donner une estimation du nombre moyen de pannes par jour.

Exercice

La distribution du nombre de pannes observées dans le fonctionnement d'une machine au cours de 100 journées de travail est résumée dans le tableau suivant :

Nombre de pannes	0	1	2	3	4	Total
Nombre de jours	53	32	11	3	1	100

Donner une estimation du nombre moyen de pannes par jour.

Exercice

On a mesuré le poids de raisin produit par pied sur 10 pieds pris au hasard dans une vigne. On a obtenu les résultats suivants exprimés en kilogrammes :

2.4	3.4	3.6	4.1	4.3	4.7	5.4	5.9	6.5	6.9

1. *Donner une estimation de m moyenne de l'échantillon.*
2. *On modélise le poids de raisin produit par une souche de cette vigne par une variable aléatoire de loi $\mathcal{N}(m, \sigma)$. Donner une estimation de l'´*

Exercice

On a mesuré le poids de raisin produit par pied sur 10 pieds pris au hasard dans une vigne. On a obtenu les résultats suivants exprimés en kilogrammes :

$$2.4 \quad 3.4 \quad 3.6 \quad 4.1 \quad 4.3 \quad 4.7 \quad 5.4 \quad 5.9 \quad 6.5 \quad 6.9$$

On modélise le poids de raisin produit par une souche de cette vigne par une variable aléatoire de loi $\mathcal{N}(m, \sigma)$.

1. *Donner une estimation de m moyenne de l'échantillon.*
2. *Donner une estimation de l' ´*

Exercice

Pour un sondage électoral, on constitue deux échantillons d'électeurs de tailles 300 et 200 respectivement dans deux circonscriptions indépendantes A et B. Cela met en évidence des intentions de vote de 168 *et* 96 *électeurs pour un candidat X donné. Donner une estimation du pourcentage des électeurs votant pour le candidat X*

1. *dans la circonscription A*
2. *dans la circonscription B*
3. *dans les circonscription A et B*

Exercice

Soit X une variable aléatoire de loi de poisson de paramètre λ. Soit $(X_1, X_2, \dots, X_{25})$ un échantillon de X de réalisations $(x_1, x_2, \dots, x_{25})$ vérifiant

$$\sum_{i=1}^{25} x_i = 186 \quad et \quad \sum_{i=1}^{25} x_i^2 = 1385,84.$$

1. *Donner deux estimateurs sans biais de λ.*
2. *Lequel de ces estimateurs est le plus efficace ? justifier*
3. *Donner une estimation de λ.*

Estimation par la méthode du maximum de vraisemblance

Soit X une variable aléatoire à valeurs dans E. On suppose que la loi de X dépend d'un paramètre θ à estimer. On considère $x_1, x_2, ..., x_n$ des éléments de E.

Définition (fonction de Vraisemblance)

On appelle fonction de vraisemblance la fonction L_θ définie par

$$L_\theta\left(x_1, x_2, ..., x_n\right) = \begin{cases} \displaystyle\prod_{i=1}^{n} \mathbb{P}(X = x_i) & \text{si } X \text{ est un variable discrète} \\ \displaystyle\prod_{i=1}^{n} f_X(x_i) & \text{si } X \text{ est un variable continue} \end{cases}$$

Exemple 1 : loi de Bernoulli de paramètre p.
Remarquons que pour $x \in \{0; 1\}$

$$\mathbb{P}(X = x) = p^x (1 - p)^{1-x}.$$

Donc, pour $x_1, x_2, ..., x_n$ éléments de $\{0; 1\}$

$$\begin{aligned} L_p(x_1, ..., x_n) &= \prod_{i=1}^{n} p^{x_i} (1 - p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n} x_i} (1 - p)^{n-\sum_{i=1}^{n} x_i}. \end{aligned}$$

Exemple 2 : loi de Poisson de paramètre λ.

Rappelons que, pour tout $x \in \mathbb{N}$,

$$\mathbb{P}(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}.$$

Donc, pour $x_1, x_2, ..., x_n$ éléments de $\mathbb{N}$

$$\begin{aligned} L_\lambda(x_1, ..., x_n) &= \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} \\ &= e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1!x_2!...x_n!}. \end{aligned}$$

Exemple 3 : loi Normale de paramètre (μ, σ).

Rappelons que, pour tout $x \in \mathbb{N}$,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Donc, pour $x_1, x_2, ..., x_n$ éléments de $\mathbb{R}$

$$\begin{aligned} L_{(\mu,\sigma)}(x_1, ..., x_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}. \end{aligned}$$

Définition

On dit que la variable aléatoire $\hat{\theta}$ est un estimateur de maximum de vraisemblance d'un paramètre θ si sa réalisation maximise la fonction de vraisemblance L_θ :

$$L_{\hat{\theta}}(x_1, x_2, ..., x_n) = \max_{\theta} L_\theta(x_1, x_2, ..., x_n)$$

Proposition

Soit $x_1, x_2, ..., x_n \in E$. Supposons que la fonction $\theta \longmapsto L_\theta(x_1, x_2, ..., x_n)$ est différentielle et strictement positive. Soit $\hat{\theta}$ l'estimateur de vraisemblance de θ. Alors

$$\frac{\partial \ln(L_\theta)}{\partial \theta}(x_1, x_2, ..., x_n$$

Exemple 1 : loi de Bernoulli de paramètre p. Pour $x_1, x_2, ..., x_n$ éléments de $\{0; 1\}$

$$L_p\left(x_1, ..., x_n\right) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}.$$

Donc

$$\ln(L_p)\left(x_1, ..., x_n\right) = \ln\left(\frac{p}{1-p}\right)\sum_{i=1}^n x_i + n\ln(1-p).$$

D'où

$$\frac{\partial \ln(L_p)}{\partial p}\left(x_1, ..., x_n\right) = \frac{1}{p(1-p)}\sum_{i=1}^n x_i - \frac{n}{1-p}.$$

Par conséquent, l'estimateur de maximum de vraisemblance de p est

$$\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}.$$

Exemple 2 : loi de Poisson de paramètre λ. Pour $x_1, x_2, ..., x_n$ éléments de $\mathbb{N}$

$$L_\lambda(x_1, ..., x_n) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{x_1!x_2!...x_n!}.$$

Donc

$$\ln(L_\lambda)(x_1, ..., x_n) = -n\lambda + \ln(\lambda) \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!).$$

D'où

$$\frac{\partial \ln(L_\lambda)}{\partial \lambda}(x_1, ..., x_n) = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i.$$

Par conséquent, l'estimateur de maximum de vraisemblance de λ est

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

Exemple 3 : loi Normale de paramètre (μ, σ). Pour $x_1, x_2, ..., x_n$ éléments de $\mathbb{R}$

$$L_{(\mu,\sigma)}(x_1, ..., x_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}.$$

Donc

$$\ln\left(L_{(\mu,\sigma)}\right)(x_1, ..., x_n) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

D'une part, si on dérive par rapport à μ

$$\frac{\partial \ln(L_{(\mu,\sigma)}}{\partial\mu}(x_1, ..., x_n) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu).$$

D'où, l'estimateur de maximum de vraisemblance de μ est

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{}$$

D'autre part :

- Si μ est connue, alors

$$\frac{\partial \ln(L_{(\mu,\sigma)})}{\partial \sigma^2} = 0 \Leftrightarrow -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0.$$

  D'où l'estimateur de maximum de vraisemblance de σ^2 est

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 = T^2$$

Si μ est inconnue, alors

$$\left\{ \begin{array}{ccc} \dfrac{\partial \ln \left(L_{(\mu,\sigma)}\right)}{\partial \mu} & = & 0 \\ \dfrac{\partial \ln \left(L_{(\mu,\sigma)}\right)}{\partial \sigma^2} & = & 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{c} \displaystyle\sum_{i=1}^{n} x_i - n\mu = 0 \\ -n + \dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{n} (x_i - \mu)^2 = 0 \end{array} \right.$$

D'où l'estimateur de maximum de vraisemblance de σ^2 est

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 = S^2$$

Proposition

L'estimateur obtenu par la méthode du maximum de vraisemblance est :

- *convergent*
- *asymptotiquement efficace*
- *asymptotiquement distribué selon une loi Normale*

Exercice

La durée de vie d'une certaine marque d'ampoules est modélisée par la variable aléatoire X de densité

$$f_X(x) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) \ \textit{si } x \geq 0 \qquad \textit{et} \qquad f_X(x) = 0 \ \textit{sinon}$$

où λ est un paramètre positif.

1. *calculer $\mathbb{E}(X)$ et $Var(X)$*
2. *Trouver l'estimateur de maximum de vraisemblance du paramètre λ.*
3. *Que peut-on dire sur cet estimateur ?*

Estimation par intervalle de confiance

Introduction

- L'estimation d'un paramètre inconnu par une seule valeur est quelque fois insuffisante, on préfère souvent donner un intervalle de valeurs.
- On cherche des intervalles dit "intervalle de confiance" qui contiennent
  - la moyenne μ inconnue
  - l'écart-type σ inconnu
  - le pourcentage p d'une certaine propriété que possède la population.

Définition

Soit X une v.a. dont la loi dépend d'un paramètre inconnu θ.
On appelle intervalle de confiance pour un de niveau $1-\alpha$ (ou de seuil α), un intervalle qui a la probabilité $1-\alpha$ de contenir la vraie valeur de θ.
Autrement dit, $[t_1, t_2]$ est intervalle de confiance pour un niveau $1-\alpha$ du paramètre θ si

$$\mathbb{P}\left\{t_1 < \theta < t_2\right\} = 1-\alpha.$$

Introduction

Remarque

- *α est appelé le seuil ou le risque, et $1-\alpha$ est le niveau de confiance.*
- *Plus le niveau de confiance est élevé, plus la certitude est grande que la méthode d'estimation produira une estimation contenant la vraie valeur de θ.*
- *Les niveaux de confiance les plus utilisés sont 90%, 95% et 99%.*

Intervalle de confiance pour une moyenne

Soit X une variable aléatoire et $(X_1, ..., X_n)$ un échantillon de X.
Nous avons vu que la moyenne $\overline{X}$ d'un échantillon aléatoire permet d'estimer la vraie moyenne de la population. Nous voudrions estimer également la précision de cette moyenne, c'est-à-dire donner une marge d'erreur ou un intervalle de confiance.
On peut distinguer deux cas :

1. la taille de l'échantillon est petite $(n < 30)$ et $X \sim \mathcal{N}(\mu, \sigma)$:
    1. l'écart type σ est connu
    2. l'écart type σ est inconnu
2. la taille de l'échantillon est grande $(n \ggg 30)$ et X de loi quelconque

Intervalle de confiance pour une moyenne

$n < 30$ et $X \sim \mathcal{N}(\mu, \sigma)$

<u>σ **est connu :**</u> On se fixe le risque α et on cherche dans la table de la loi normale la valeur de $u_{1-\frac{\alpha}{2}}$ le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite. On peut écrire donc que

$$\mathbb{P}\left(-u_{1-\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

Ceci est équivalent à

$$\mathbb{P}\left(\overline{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Conclusion : Si $\overline{x}$ est une réalisation de $\overline{X}$, l'intervalle de confiance de la moyenne μ de seuil $1 - \alpha$ est

$$IC = \left[\overline{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

Intervalle de confiance pour une moyenne

$n < 30$ et $X \sim \mathcal{N}(\mu, \sigma)$

Exemple

Des mesures d'un échantillon $(X_1, ..., X_{20})$ *de poids* X *ont donné les résultats* $(x_1, ..., x_{20})$ *vérifiant* $\sum_{i=1}^{20} x_i = 80$ *et* $\sum_{i=1}^{20} x_i^2 = 323$*. On suppose que ces poids sont distribués selon une loi Normale de moyenne* μ *et d'écart-type* $\sigma = 0.17$*. On cherche un intervalle de confiance de la moyenne à un niveau de confiance* 90% *:*

$$IC = \left[\overline{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \ ; \ \overline{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

AN : $\overline{x} = 4$*,* $\alpha = 0.1$*,* $u_{0.95} = 1.64$*,* $n = 20$ *et* $\sigma = 0.17$*. Donc*

$$IC = [3.853; 4.147]$$

<u>**σ est inconnu :**</u> L'écart type σ étant inconnu, on l'estime par S ou $\bar{S}$. On a recours à la loi de Student.

Théorème

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n-1}}} = \frac{\overline{X}-\mu}{\frac{\bar{S}}{\sqrt{n}}}$$

suit une loi de Student de degré de liberté $n-1$.

Intervalle de confiance pour une moyenne

$n < 30$ et $X \sim \mathcal{N}(\mu, \sigma)$

En appelant t_{n-1} le quantile d'ordre $1 - \frac{\alpha}{2}$, on peut écrire que

$$\mathbb{P}\left(-t_{n-1} \leq \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n-1}}} \leq t_{n-1}\right) = 1 - \alpha.$$

Par suite

$$\mathbb{P}\left(\overline{X} - t_{n-1}\frac{S}{\sqrt{n-1}} \leq \mu \leq \overline{X} + t_{n-1}\frac{S}{\sqrt{n-1}}\right) = 1 - \alpha.$$

En remplaçant $\overline{X}$ et S par leurs valeurs calculées sur l'échantillon, on obtient l'intervalle de confiance sur la moyenne μ :

$$IC = \left[\overline{x} - t_{n-1}\frac{s}{\sqrt{n-1}}, \overline{x} + t_{n-1}\frac{s}{\sqrt{n-1}}\right]$$

Intervalle de confiance pour une moyenne
$n < 30$ et $X \sim \mathcal{N}(\mu, \sigma)$

Exemple

Des mesures d'un échantillon $(X_1, ..., X_{20})$ *de poids* X *ont donné les résultats* $(x_1, ..., x_{20})$ *vérifiant* $\sum_{i=1}^{20} x_i = 80$ *et* $\sum_{i=1}^{20} x_i^2 = 323$. *On suppose que ces poids sont distribués selon une loi Normale de moyenne* μ *et d'écart-type* σ. *On cherche un intervalle de confiance de la moyenne à un niveau de confiance 90% :*

$$IC = \left[\overline{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}} \ ; \ \overline{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}\right]$$

AN : $\overline{x} = 4$, $\alpha = 0.1$, $t_{0.95} = 1.72$, $n = 20$ *et* $s = 0.39$. *Donc*

$$IC = [3.85; 4.15]$$

Lorsque la taille n de l'échantillon est grande (pratiquement dès que $n > 30$), on appliquera les formules de l'intervalle de confiance sur μ, même si l'échantillon n'est pas issu d'une population normale.
En effet, le théorème central limite nous permet de dire que $\overline{X}$ est approximativement de loi $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$ lorsque n est grand.

Intervalle de confiance pour une moyenne

$n \ggg 30$ et X de loi quelconque

l'intervalle de confiance de la moyenne μ de seuil $1-\alpha$ est :

- Si σ est connue : $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ et on a

$$IC = \left[\overline{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right].$$

- Si σ est inconnue : $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim \mathcal{N}(0,1)$ et on a

$$IC = \left[\overline{x} - u_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \overline{x} + u_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right].$$

$u_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite.

Exemple

Des mesures d'un échantillon de 60 *poids de* X *ont donné une estimation de la moyenne* $\bar{x} = 4$ *et une estimation de l'écart-type* $s = 0.17$*. On cherche un intervalle de confiance de la moyenne à un niveau de confiance* 90% *:*

$$IC = \left[\overline{x} - u_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \; ; \; \overline{x} + u_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$$

AN : $\overline{x} = 4$, $\alpha = 0.1$, $u_{0.95} = 1.64$, $n = 60$ *et* $s = 0.17$*. Donc*

$$IC = [3.957; 4.043]$$

Intervalle de confiance pour la variance d'une variable Gaussienne

Soit X une variable aléatoire de loi normale de moyenne μ et de variance σ^2 et soit $(X_1, ..., X_n)$ un échantillon de X.
Pour un seuil α, on désire estimer la variance σ^2.
On distingue deux cas :

- μ connue
- μ inconnue

Intervalle de confiance pour la variance d'une variable Gaussienne

<u>μ **connue :**</u> Rappelons que

$$\frac{nT^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_2(n).$$

Soit $k_{n(\frac{\alpha}{2})}$ et $k_{n(1-\frac{\alpha}{2})}$ les quantiles de la loi $\chi_2(n)$ associés respectivement aux valeurs $\frac{\alpha}{2}$ et $1 - \frac{\alpha}{2}$. càd

$$\mathbb{P}\left(n\frac{T^2}{\sigma^2} \leq k_{n(\frac{\alpha}{2})}\right) = \frac{\alpha}{2}$$

et

$$\mathbb{P}\left(n\frac{T^2}{\sigma^2} \leq k_{n(1-\frac{\alpha}{2})}\right) = 1 - \frac{\alpha}{2}$$

Intervalle de confiance pour la variance d'une variable Gaussienne

Donc

$$\mathbb{P}\left(k_{n(\frac{\alpha}{2})} \leq n\frac{T^2}{\sigma^2} \leq k_{n(1-\frac{\alpha}{2})}\right) = 1-\alpha.$$

Par conséquent

$$\mathbb{P}\left(\frac{nT^2}{k_{n(1-\frac{\alpha}{2})}} \leq \sigma^2 \leq \frac{nT^2}{k_{n(\frac{\alpha}{2})}}\right) = 1-\alpha.$$

On calcule t^2 la réalisation de T^2.
L'intervalle de confiance de σ^2 est donné par

$$IC = \left[\frac{nt^2}{k_{n(1-\frac{\alpha}{2})}} ; \frac{nt^2}{k_{n(\frac{\alpha}{2})}}\right]$$

Intervalle de confiance pour la variance d'une variable Gaussienne

Exemple

Des mesures d'un échantillon $(X_1, ..., X_{20})$ *de poids* X *ont donné les résultats* $(x_1, ..., x_{20})$ *vérifiant* $\sum_{i=1}^{20} x_i = 80$ *et* $\sum_{i=1}^{20} x_i^2 = 323$. *On suppose que ces poids sont distribués selon une loi Normale de moyenne* $\mu = 3.98$ *et d'écart-type* σ. *On cherche un intervalle de confiance de la variance à un niveau de confiance* 90% *:*

$$IC = \left[\frac{nt^2}{k_{n(1-\frac{\alpha}{2})}} ; \frac{nt^2}{k_{n(\frac{\alpha}{2})}} \right].$$

AN : $n = 20$, $t^2 = 0.1504$, $\alpha = 0.1$, $k_{20(0.5)} = 19.337$ *et* $k_{20(0.95)} = 31.410$. *Donc*

$$IC = [0.0958; 0.1556]$$

Intervalle de confiance pour la variance d'une variable Gaussienne

μ **inconnue :** Rappelons que

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)\overline{S}^2}{\sigma^2} \sim \chi_2(n-1).$$

Soient $k_{1-\frac{\alpha}{2}}$ et $k_{\frac{\alpha}{2}}$ les quantiles d'ordre $1-\alpha/2$ et $\alpha/2$ de la loi χ^2_{n-1}. C'est à dire

$$\mathbb{P}\left(\frac{nS^2}{\sigma^2} \leq k_{\frac{\alpha}{2}}\right) = \alpha/2$$

et

$$\mathbb{P}\left(\frac{nS^2}{\sigma^2} \leq k_{1-\frac{\alpha}{2}}\right) = 1-\alpha/2$$

Alors

$$\mathbb{P}\left(k_{\frac{\alpha}{2}} \leq \frac{nS^2}{\sigma^2} \leq k_{1-\frac{\alpha}{2}}\right)$$

Intervalle de confiance pour la variance d'une variable Gaussienne

Donc

$$\mathbb{P}\left(\frac{nS^2}{k_{1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{nS^2}{k_{\frac{\alpha}{2}}}\right) = 1-\alpha.$$

On calcule s^2 la réalisation de S^2. L'intervalle de confiance de la variance est

$$\left[\frac{ns^2}{k_{n-1(1-\frac{\alpha}{2})}}; \frac{ns^2}{k_{n-1(\frac{\alpha}{2})}}\right].$$

Remarque

On peut procéder de la même manière si on estime σ^2 par $\overline{S}^2$. L'intervalle de confiance de la variance peut s'écrire comme :

$$\left[\frac{(n-1)\overline{s}^2}{k_{n-1(1-\frac{\alpha}{2})}}; \frac{(n-1)\overline{s}^2}{k_{n-1(\frac{\alpha}{2})}}\right].$$

Intervalle de confiance pour la variance d'une variable Gaussienne

Exemple

Des mesures d'un échantillon $(X_1, \ldots, X_{20})$ *de poids* X *ont donné les résultats* $(x_1, \ldots., x_{20})$ *vérifiant* $\sum_{i=1}^{20} x_i = 80$ *et* $\sum_{i=1}^{20} x_i^2 = 323$. *On suppose que ces poids sont distribués selon une loi Normale de moyenne* μ *et d'écart-type* σ. *On cherche un intervalle de confiance de la variance à un niveau de confiance* 90% *:*

$$IC = \left[\frac{ns^2}{k_{n-1(1-\frac{\alpha}{2})}}; \frac{ns^2}{k_{n-1(\frac{\alpha}{2})}}\right].$$

AN : $n-1 = 19$, $s^2 = 0.15$, $k_{19(0.5)} = 10.117$ *et* $k_{19(0.95)} = 30.144$. *Donc*

$$IC = [0.0995; 0.297].$$

Intervalle de confiance pour une proportion

Soit une population ayant des individus possédant une certaine caractéristique A. On veut estimer à partir d'un échantillon de taille n la proportion p d'individus possédant cette caractéristique A. Soit $\widehat{p}$ l'estimateur sans biais de p et f sa réalisation.
Si n est grand ($n \ggg 30$) ou si la fréquence observée $0 < f < 1$ telle que $nf \geq 5$ ou $n(1-f) \geq 5$, alors

$$\widehat{p} \sim \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

On cherche dans la table de la loi normale $u_{1-\frac{\alpha}{2}}$ telle que

$$\mathbb{P}\left(-u_{1-\frac{\alpha}{2}} \leq \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} < u_{1-\frac{\alpha}{2}}\right)$$

Intervalle de confiance pour une proportion

L'intervalle de confiance au risque α pour la proportion p est donné par

$$IC = \left[f - u_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}}\ ,\ f + u_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}} \right].$$

Intervalle de confiance pour une proportion

Exemple

Un échantillon de 100 *votants choisis au hasard dans un référendum a montré que* 55% *d'entre eux étaient favorable. On cherche un intervalle de confiance à* 90% *de la proportion* p *des votant favorablement :*

$$IC = \left[f - u_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}} \;,\; f + u_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}} \right].$$

AN : $f = 0.55$, $\alpha = 0.1$, $u_{0.95} = 1.64$ *et* $n = 100$. *Donc*

Exercice

Afin d'étudier le salaire horaire des ouvriers d'un secteur d'activité, on prélève un échantillon aléatoire $(X_1, X_2, ..., X_{15})$ *de réalisation (en dhs)*

$$\begin{array}{cccccccc} 9,8 & 8,6 & 9,7 & 10,2 & 8,9 & 10,1 & 9,9 & 9,7 \\ 8,7 & 9,8 & 10,2 & 9,3 & 10,4 & 9,5 & 10,9 & \end{array}$$

On suppose que la loi suivie par le salaire horaire est normale de moyenne μ *et variance* σ^2.

1. *Supposons que* $\sigma = 0,4$. *Donner l'intervalle de confiance de* μ *à un niveau de confiance* 95%.
2. *Supposons que* $\mu = 9,7$. *Donner l'intervalle de confiance de* σ *à un seuil de* 5%.

Exercice

Pour estimer la précision d'un thermomètre, on réalise 100 *mesures indépendantes de* X *la température d'un liquide maintenu à une température constante. Les observations* x_i *ont conduit aux valeurs suivantes*

$$\sum_{i=1}^{100} x_i = 1995 \quad et \quad \sum_{i=1}^{100} x_i^2 = 40011.$$

À 99%*, donner l'intervalle de confiance de la moyenne* μ*.*

Exercice

L'objet d'une enquête menée sur la région Larache-Ksar El Kebir est d'étudier la variable X représentant la dépense moyenne de foyer (en Dirhams par jour). On choisit au hasard 100 *foyers, les réalisation* $(x_i)_{i=1}^{100}$ *de l'échantillon* $(X_i)_{i=1}^{100}$ *permettent d'affirmer que*

$$\sum_{i=1}^{100} x_i = 3007 \quad et \quad \frac{1}{100}\sum_{i=1}^{100} x_i^2 = 940,4.$$

On suppose que les dépenses des foyers suivent une loi Normale de moyenne m et d'écart-type σ.

1. *Déterminer un intervalle de confiance de niveau* 90% *pour* σ^2

Le responsable sur l'enquête souhaite tester les résultats obtenus sur la ville de Larache. Il suppose que les dépenses moyennes des foyers de la ville de Larache sont modélisées par une Loi Normale de moyenne μ et d'écart-type $\sigma = 6,05$. Les dépenses moyennes de 10 foyers de la ville de Larache choisis au hasard sont

$25,8$	37	$31,95$	$26,55$	$32,9$	$34,95$	$31,8$	$36,2$	$22,3$	$31,05$

2. Déterminer un intervalle de confiance de niveau 90% pour la moyenne μ et donner une conclusion.

Exercice

Pour estimer le marché potentiel d'un nouveau produit laitier, un sondage est effectué auprès de 90 consommateurs de ce produit. Il en résulte que l'estimation de la consommation moyenne est de $7,4$ *unités par semaine avec un écart-type estimé de* $3,1$. *Donner un intervalle de confiance de la consommation moyenne par semaine avec un risque de* 1%.

Exercice

Les notes d'une épreuve de statistique inférentielle d'un échantillon de 25 étudiants sont :

12	10	13	15	18	9	12	6	14	7	13	13	12
10	15	11	12	14	13	15	13	15	14	10	12	

Supposons que les notes sont la réalisation d'une variable aléatoire normale d'écart-type σ.
Donner l'intervalle de confiance de σ *à un niveau de 99%.*

Exercice

Pour un référendums sur l'indépendance d'une petite île de l'ukraine, on constitue un échantillon aléatoire de 1000 électeurs qui donnent leurs avis par internet. On constate que 655 ont voté 'oui' pour l'indépendance. Soit p la proportion des habitants de l'île désirant l'indépendance.

1. *Donner un estimateur et une estimation de p.*
2. *Donner un intervalle de confiance de p à un risque de 5%.*

Taille de l'échantillon

Taille d'échantillon

Durant la préparation de l'enquête, le chercheur doit à un moment décider de la taille de l'échantillon. Cette décision est importante car elle a une incidence sur

- les coûts de l'étude.
- la précision des résultats.

Une première approche consiste à utiliser le Budget disponible :

$$\begin{aligned} \text{Budget} &= \text{Coûts fixes} \\ &\quad + (\text{taille de l'échantillon x Coût d'un Questionnaire}) \end{aligned}$$

Taille d'échantillon

On trouve ainsi une taille de l'échantillon imposée par la contrainte budgétaire. Mais est ce que cet échantillon est suffisant pour représenter la population entière ? ! ! !

.

Il faut préciser que "plus on souhaite des résultats précis, plus l'échantillon nécessaire est important"

.

Cependant une deuxième approche (Plus rationnelle) consiste à utiliser la marge d'erreur tolérée (la précision de l'étude) pour calculer la taille minimale de l'échantillon afin qu'il représente la population.

Écart type connu : Pour trouver la taille d'échantillon il faut résoudre l'équation

$$u\frac{\sigma}{\sqrt{n}} \leq e,$$

où e est la marge d'erreur fixé à l'avance, u le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi Normale et σ l'écart type de la population. On peut écrire donc que

$$\sqrt{n} \geq \frac{\sigma u}{e},$$

d'où

$$n \geq \left(\frac{u\sigma}{e}\right)^2.$$

Écart type inconnu : On utilise une étude pilote c.à.d On distribue un questionnaire d'essai et on calcul l'écart type corrigé sur l'échantillon. Ensuite, on fixe une marge d'erreur e qu'on peut tolérée et le reste resemble au premier cas :

$$u\frac{s}{\sqrt{n}} \leq e,$$

u est le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi Normale et s l'estimation de l'écart type. Alors

$$\sqrt{n} \geq \frac{su}{e},$$

d'où

$$n \geq \left(\frac{us}{e}\right)^2.$$

En pratique, on utilise le fait que les valeur de la loi Normale ne s'étendent pas plus loin que 4σ. On prend alors

$$\sigma = \frac{lv}{4} = \frac{\text{Valeur maximale - valeur minimale}}{4}.$$

On peut considérer que

$$n \geq \left(\frac{u\ lv}{4e}\right)^2.$$

De la même manière que dans le cas de la moyenne, si l'on se fixe la marge d'erreur e à ne pas dépasser (avec une probabilité $1-\alpha$), on cherche n tel que

$$u\sqrt{\frac{f(1-f)}{n}} \leq e.$$

ce qui implique que

$$n \geq \frac{u^2 f(1-f)}{e^2}.$$

Mais comme on n'a pas encore tiré l'échantillon, la fréquence dans l'échantillon est inconnue. Alors comment peut-on procéder ?

On peut procéder de deux manières :

- On trouve une estimation de f à l'aide d'une enquête pilote.
- On prend comme valeur de f celle qui nous donne la plus grande taille d'échantillon. Ceci est réalisé si $f(1-f)$ prend sa valeur maximale qui est 0,25. Et on a

$$n \geq \frac{u^2 \times 0,25}{e^2} = \frac{u^2}{4e^2}.$$

Exercice

Afin d'estimer le revenu mensuel moyen dans un secteur de production. Quelle doit être la taille de l'échantillon de salariés à interroger pour que la moyenne empirique ne s'éloigne pas de la moyenne de la population de $100dh$ *avec une probabilité au moins égale à* $0,95$ *avec un écart type est de* $500dh$ *par salarié ?*

Exercice

Un magasin réalise un chiffre d'affaire d'au moins $5000dh$ *et d'au plus* $12000dh$ *par jour. On cherche à estimer le chiffre d'affaire moyen par jour de ce magasin. Quel est la taille de l'échantillon minimale pour une marge erreur maximale de* $500dh$ *et un niveau de confiance de* 90% *?.*

Exercice

Dans le but d'étudier l'intention d'achat d'un produit, on décide de réaliser un sondage. Combien de personnes doit-on interroger pour que la fréquence empirique ne s'éloigne pas de la vraie proportion de 1% *et ce avec une probabilité au moins égale à* 95% *?*

Exercice

Un parc de loisirs souhaite estimer la proportion des visiteurs qui font des achats à une marge d'erreur de 2.5%. *Une enquête pilote a estimé cette proportion à* 65%. *Quelle est la taille de l'échantillon minimale à considérer pour un niveau de confiance de* 95% *?.*

Tests Statistiques

Introduction

- Un test statistique est une méthodes permettant de prendre une décision d'accepter ou de rejeter une hypothèse à partir des informations fournies par un ou plusieurs échantillons.
- On formule une hypothèse de départ, appelée hypothèse nulle et souvent notée (H_0). Il s'agit de décider si on rejette ou non cette hypothèse par opposition à une contre-hypothèse appelée hypothèse alternative souvent notée (H_1).

Introduction

Exemple : Un contrôleur de réception a reçu un lot de pièces sensées être de 5 mm de diamètre ; mais il se demande si, par suite d'un étiquetage douteux, on ne lui a pas livré par erreur des pièces de 6 mm de diamètre.

On sait que la machine fournie une légères variation et que le diamètre des pièces est en fait distribué selon une loi normale N(m ; 0, 6). Le problème est de savoir si on a bien $m = 5$, et pas plutôt $m = 6$.

1. Si une pièce prise au hasard dans le lot mesure exactement 5 mm, est-on sûr que le lot est bon ?
2. Si elle fait exactement 5.8 mm, est-on sûr que le lot est mauvais ?
3. Est-ce la même chose si, sur 10 pièces prises au hasard, on a un diamètre moyen de 5.8 mm ?
4. A partir de quelle valeur du diamètre moyen peut on dire que le lot est mauvais ?

Mécanique des tests d'hypothèses

Le principe général du test statistique peut s'énoncer comme suit :

1. Formuler les hypothèses (H_0) et (H_1) : Soit un caractère d'une population dont la valeur du paramètre (moyenne, proportion, variance, etc.) est inconnue. Une hypothèse "nulle" est alors formulée sur ce paramètre inconnu, elle résulte de considérations théoriques, pratiques ou encore plus simplement basée sur un pressentiment.
2. Choisir le risque ou le niveau de confiance : Pour effectuer le test statistique, il faudra choisir un certain risque d'erreur qui est la probabilité de se tromper en prenant la décision retenue.
3. Il faut définir la statistique qui convient pour effectuer le test appelée "variable de décision".

Mécanique des tests d'hypothèses

4. Déterminer la région critique ou région de rejet I_{rejet} qui est l'ensemble des valeurs prises par la variable de décision qui conduiront à rejeter (H_0) ou déterminer I_{accept} le complémentaire de I_{rejet} est appelé région d'acceptation de (H_0).
5. On calcule la réalisation de la variable de décision à partir des valeurs observées.
6. Conclusion finale.

- Hypothèse nulle H_0 : hypothèse selon laquelle on fixe a priori un paramètre de la population à une valeur particulière.
- Hypothèse alternative H_1 (ou contre-hypothèse) : n'importe quelle autre hypothèse qui diffère de l'hypothèse nulle H_0.
- La formulation des hypothèses peut être écrite comme

$$\begin{cases} \theta = \theta_0 & (H_0) \\ \theta \neq \theta_0 & (H_1) \end{cases} \qquad \begin{cases} \theta = \theta_0 & (H_0) \\ \theta > \theta_0 & (H_1) \end{cases} \qquad \begin{cases} \theta = \theta_0 & (H_0) \\ \theta < \theta_0 & (H_1) \end{cases}$$

$$\begin{cases} \theta = \theta_0 & (H_0) \\ \theta = \theta_1 & (H_1) \end{cases} \qquad \begin{cases} \theta \leq \theta_0 & (H_0) \\ \theta > \theta_0 & (H_1) \end{cases} \qquad \begin{cases} \theta \geq \theta_0 & (H_0) \\ \theta < \theta_0 & (H_1) \end{cases}$$

- On ne pourra jamais conclure avec certitude dans un test statistique. Il y aura toujours des erreurs de décision.
- Pour effectuer le test statistique, il faudra choisir un certain risque d'erreur qui est la probabilité de se tromper en prenant la décision retenue.

Il existe deux types d'erreurs :

- On appelle erreur de première espèce ou erreur de type I, notée α, la probabilité de rejeter H_0 alors qu'elle est vraie. α est aussi appelée niveau ou seuil de signification.
- On appelle erreur de deuxième espèce ou erreur de type II, notée β, la probabilité d'accepter H_0 alors qu'elle est fausse.

- α est aussi appelé niveau ou seuil de signification.

$$\alpha = \mathbb{P}\left(\textit{Rejeter } H_0 \big/ H_0 \textit{ vraie}\right) = \mathbb{P}\left(\textit{Choisir } H_1 \big/ H_0 \textit{ vraie}\right).$$

- $(1-\beta)$ est appelé puissance du test pour H_1, c'est la probabilité de retenir H_1 alors qu'elle est vraie.

$$1-\beta = \mathbb{P}\left(\textit{Rejetet } H_0 \big/ H_1 \textit{ vraie}\right).$$

On peut résumer les types d'erreurs susceptibles de survenir dans le tableau suivant :

Décision suite au test	Réalité	
	H_0 vraie	H_0 fausse
Rejet de H_0	Erreur de premier espèce α	Bonne décision
Acceptation de H_0	Bonne décision	Erreur de deuxième espèce β

Test bilatéral : Un test est dit bilatéral si la condition de rejet est indépendante du signe de l'écart observé entre les caractéristiques comparées θ et θ_0. Les hypothèses formulées du test bilatéral sont :

$$\begin{cases} H_0 & \theta = \theta_0 \\ H_1 & \theta \neq \theta_0. \end{cases}$$

La règle de décision peut être représentée ainsi :

$\theta \neq \theta_0$	$\theta = \theta_0$	$\theta \neq \theta_0$
Rejet de H_0	Acceptation de H_0	Rejet de H_0

z_1 z_2

z_1 et z_2 sont les valeurs critiques qui déterminent la région d'acceptation.

Test Unilatéral à droite : Les hypothèses formulées du test unilatéral à droite sont

$$\begin{cases} H_0 & \theta = \theta_0 \\ H_1 & \theta > \theta_0. \end{cases}$$

La règle de décision peut être représentée ainsi :

$\theta \leq \theta_0$	$\theta > \theta_0$
Acceptation de H_0	Rejet de H_0

z

z est la valeur critique qui détermine la région d'acceptation.

Test Unilatéral à gauche : Les hypothèses formulées du test unilatéral à gauche sont

$$\begin{cases} H_0 & \theta = \theta_0 \\ H_1 & \theta < \theta_0. \end{cases}$$

La règle de décision peut être représentée ainsi :

$\theta < \theta_0$	$\theta \geq \theta_0$
Rejet de H_0	Acceptation de H_0

z

z est la valeur critique qui détermine la région d'acceptation.

Test de conformité sur la moyenne

On suppose que X suit une loi Normale de moyenne μ et d'écart type σ.

σ est connu : La variable X étudiée au niveau de la population suit une loi Normale $\mathcal{N}(\mu, \sigma)$ avec σ connu. Ainsi la distribution de $\overline{X}$ au niveau de l'échantillon sera :

$$\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$$

et

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1).$$

Test de conformité sur une moyenne

Test bilateral :

Les hypothèses du test se présentent sous la forme :

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu \neq \mu_0 \end{cases}$$

On considère comme variable de décision $\overline{X}$. La région d'acceptation du test comme un intervalle symétrique autour de μ_0 de la forme $I_{accept} = [c_1, c_2]$, où

$$c_1 = \mu_0 - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ et } c_2 = \mu_0 + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

On détermine la valeur de $u_{1-\frac{\alpha}{2}}$ à partir de la table de la loi Normale centrée et réduite

Conclusion du test : Si $\overline{x}$, la valeur de la moyenne sur l'échantillon, appartient à la zone d'acceptation ($\overline{x} \in [c_1, c_2]$,) alors on accepte (H_0), sinon, on rejette H_0

Test de conformité sur une moyenne

Si on prend comme variable de décision

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

alors la région d'acceptation est :

$$\widetilde{I}_{accept} = [-u_{1-\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}}].$$

c'est à dire on accept H_0 si la valeur observée

$$z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \in \widetilde{I}_{accept}.$$

Exemple

Des mesures d'un échantillon $(X_1, ..., X_{20})$ *de poids* X *ont donné les résultats* $(x_1, ..., x_{20})$ *vérifiant*

$$\sum_{i=1}^{20} x_i = 80 \quad et \quad \sum_{i=1}^{20} x_i^2 = 323.$$

On suppose que ces poids sont distribués selon une loi Normale de moyenne μ *et d'écart-type* $\sigma = 0.17$.
Peut-on dire, à un niveau de 90%, *que le poids moyen est égal à* 3.8 *?*

1. Hypothèse de test : $\begin{cases} \mu = 3.8 & (H_0) \\ \mu \neq 3.8 & (H_1) \end{cases}$
2. Seuil de test : $\alpha = 0.1$
3. Statistique de décision : $\overline{X} = \frac{1}{20} \sum_{i=1}^{20} X_i$
4. Intervalle d'acceptation :

$$I_{accep} = \left[\mu_0 - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \mu_0 + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

avec $\mu_0 = 3.8$, $u_{0.95} = 1.64$, $n = 20$ et $\sigma = 0.17$. Donc $I_{accep} = [3.74; 3.86]$.

5. $\overline{x} = 4 \notin I_{accep}$ donc n'accepte pas (H_0)
6. On peut dire, à 90%, que le poids moyen est différent de 3.8.

Test de conformité sur une moyenne

Test Unilateral à droite :

Les hypothèses du test se présentent sous la forme :

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu > \mu_0 \end{cases}$$

On considère comme variable de décision $\overline{X}$. La région critique (de rejet) du test est de la forme $I_{rejet} =]c, +\infty[$ ou la frontière de la région critique aura pour expression :

$$c = \mu_0 + u_{1-\alpha}\frac{\sigma}{\sqrt{n}}.$$

$u_{1-\alpha}$ est le quantile de la loi Normale centrée réduite associé à la valeur $1-\alpha$ (càd $\mathbb{P}(N \leq u_{1-\alpha}) = 1-\alpha$.

Conclusion du test : Si $\overline{x}$, la valeur de la moyenne sur l'échantillon, appartient à la zone de rejet, alors on rejette (H_0), sinon, on accepte H_0

Test de conformité sur une moyenne

Si on prend comme variable de décision

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

alors la région de rejet sera de la forme :

$$\widetilde{I}_{rejet} =]u_{1-\alpha}, +\infty[.$$

c'est à dire on rejette H_0 si la valeur observée

$$z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \in \widetilde{I}_{rejet}.$$

Exemple

Des mesures d'un échantillon $(X_1, ..., X_{20})$ *de poids* X *ont donné les résultats* $(x_1, ..., x_{20})$ *vérifiant*

$$\sum_{i=1}^{20} x_i = 80 \quad \textit{et} \quad \sum_{i=1}^{20} x_i^2 = 323.$$

On suppose que ces poids sont distribués selon une loi Normale de moyenne μ *et d'écart-type* $\sigma = 0.17$.
Peut-on dire, à un niveau de 90%, *que le poids moyen est supérieur à* 3.8 *?*

1. Hypothèse de test : $\begin{cases} \mu = 3.8 & (H_0) \\ \mu > 3.8 & (H_1) \end{cases}$ ou $\begin{cases} \mu \leq 3.8 & (H_0) \\ \mu > 3.8 & (H_1) \end{cases}$
2. Seuil de test : $\alpha = 0.1$
3. Statistique de décision : $\overline{X} = \frac{1}{20} \sum_{i=1}^{20} X_i$
4. Intervalle de rejet :

$$I_{rejet} = \left] \mu_0 + u_{1-\alpha} \frac{\sigma}{\sqrt{n}} ; +\infty \right[$$

   avec $\mu_0 = 3.8$, $u_{0.90} = 1.28$, $n = 20$ et $\sigma = 0.17$. Donc $I_{rejet} =]3.85; +\infty[$.
5. $\overline{x} = 4 \in I_{rejet}$ donc on rejette (H_0) et on accepte (H_1)
6. On peut dire, à 90%, que le poids moyen est supérieur à 3.8.

Test de conformité sur une moyenne

Test Unilateral à gauche :

Les hypothèses du test se présentent sous la forme :

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu < \mu_0 \end{cases}$$

On considère comme variable de décision $\overline{X}$. La région critique (de rejet) du test est de la forme $I_{rejet} =]-\infty, c[$, ou la frontière de la région critique aura pour expression

$$c = \mu_0 - u_{1-\alpha}\frac{\sigma}{\sqrt{n}}.$$

Conclusion du test : Si $\overline{x}$, la valeur de la moyenne sur l'échantillon, appartient à la zone de rejet, alors on rejette (H_0), sinon, on ne la rejette pas (on accepte H_0

Test de conformité sur une moyenne

Si on prend comme variable de décision

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}},$$

alors la région de rejet sera de la forme :

$$\widetilde{I}_{rejet} =]-\infty, -u_{1-\alpha}[.$$

c'est à dire on rejette H_0 si la valeur observée

$$z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \in \widetilde{I}_{rejet}.$$

Exemple

Des mesures d'un échantillon $(X_1, ..., X_{20})$ *de poids* X *ont donné les résultats* $(x_1, ..., x_{20})$ *vérifiant*

$$\sum_{i=1}^{20} x_i = 80 \quad et \quad \sum_{i=1}^{20} x_i^2 = 323.$$

On suppose que ces poids sont distribués selon une loi Normale de moyenne μ *et d'écart-type* $\sigma = 0.17$.
Peut-on dire, à un niveau de 80%, *que le poids moyen est inférieur à* 3.8 *?*

1. Hypothèse de test : $\begin{cases} \mu = 3.8 & (H_0) \\ \mu < 3.8 & (H_1) \end{cases}$ ou $\begin{cases} \mu \geq 3.8 & (H_0) \\ \mu < 3.8 & (H_1) \end{cases}$
2. Seuil de test : $\alpha = 0.2$
3. Statistique de décision : $\overline{X} = \frac{1}{20} \sum_{i=1}^{20} X_i$
4. Intervalle de rejet :

$$I_{rejet} = \left] -\infty ; \mu_0 - u_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right[$$

avec $\mu_0 = 3.8$, $u_{0.80} = 1.04$, $n = 20$ et $\sigma = 0.17$. Donc $I_{rejet} =] - \infty ; 3,76[$.

5. $\overline{x} = 4 \notin I_{rejet}$ donc on rejette (H_1) et on accepte (H_0)
6. On peut dire, à 80%, que le poids moyenne est supérieur ou égale à 3.8.

Test de conformité sur une moyenne

σ est inconnu : La démarche est la même que pour le test précédent mais la variance de la population n'étant pas connue, elle est estimée par la variance corrigée S^2. Rappelons que la variable

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n-1}}}$$

suit une loi de Student à $(n-1)$ degrés de liberté.

Test de conformité sur une moyenne

Test bilateral : Les hypothèses du test se présentent sous la forme :

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu \neq \mu_0 \end{cases}$$

On considère comme variable de décision $\overline{X}$. La région d'acceptation du test comme un intervalle symétrique autour de μ_0 de la forme : $I_{accept} = [c_1, c_2]$, ou :

$$c_1 = \mu_0 - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}} \text{ et } c_2 = \mu_0 + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}$$

avec $t_{1-\frac{\alpha}{2}}$ est le quantile de la loi Student à $n-1$ degré de liberté associé à la probabilité $1-\alpha/2$

Conclusion du test : Si $\overline{x}$, la valeur de la moyenne sur l'échantillon, appartient à l'intervalle d'acceptation, on accepte (H_0). Sinon, on rejette H_0

Test de conformité sur une moyenne

Si on prend comme variable de décision

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n-1}}}$$

alors la région d'acceptation est :

$$\widetilde{I}_{accept} = \left[-t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}}\right].$$

c'est à dire on accept H_0 si la valeur observée

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n-1}}} \in \widetilde{I}_{accept}$$

Exemple

Dans une étude réalisée l'an dernier, on a observé que la consommation en électricité des habitants d'une résidence composée d'appartements de trois pièces est en moyenne de 215 *dh et d'écart-type* 27 *dh. Cette année, sur un échantillon de* 35 *appartements, les consommation en électricité* $(x_i)_{i=1}^{35}$ *vérifient*

$$\sum_{i=1}^{35} x_i = 7700 \quad \textit{et} \quad \frac{1}{35}\sum_{i=1}^{35} x_i^2 = 49007.$$

Supposons que l'échantillon $(X_i)_{i=1}^{35}$ *suit une loi* $\mathcal{N}(\mu, \sigma)$.
À 95%*, peut on dire que la consommation moyenne est égale* 215 *dh ?.*

1. Hypothèse de test : $\begin{cases} \mu = 215 & (H_0) \\ \mu \neq 215 & (H_1) \end{cases}$
2. Seuil de test : $\alpha = 0.05$
3. Statistique de décision : $\overline{X} = \frac{1}{35} \sum_{i=1}^{35} X_i$
4. Intervalle d'acceptation :

$$I_{accep} = \left[\mu_0 - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}} ; \mu_0 + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}\right]$$

   avec $\mu_0 = 215$, $t_{0.975}(34) = 2.03$, $n = 35$, $\overline{x} = 220$ et $s = 24.64$. Donc $I_{accep} = [206.42; 223.58]$.
5. $\overline{x} \in I_{accep}$ donc on accepte (H_0).
6. On peut dire, à 95%, que la consommation moyenne est égale 215 dh.

Test de conformité sur une moyenne

Test Unilateral à droite : Les hypothèses du test se présentent sous la forme

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu > \mu_0 \end{cases}$$

On considère comme variable de décision $\overline{X}$. La région critique (de rejet) du test est de la forme :$I_{rejet} =]c, +\infty[$, ou la frontière de la région critique aura pour expression : $c = \mu_0 + t_{1-\alpha}\frac{s}{\sqrt{n-1}}$.
$t_{1-\alpha}$ est quantile de la loi Student à $n-1$ dll associé à la valeur $1-\alpha$.
<u>**Conclusion du test :**</u> Si $\overline{x}$, la valeur de la moyenne sur l'échantillon, appartient à la zone de rejet, on rejette (H_0). Sinon, on ne la rejette pas (on accepte H_0

Test de conformité sur une moyenne

Si on prend comme variable de décision

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n-1}}}$$

alors la région de rejet sera de la forme : $\widetilde{I}_{rejet} =]t_{1-\alpha}, +\infty[$. On rejette H_0 si la valeur observée

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n-1}}} \in \widetilde{I}_{rejet}.$$

Exemple

Dans une étude réalisée l'an dernier, on a observé que la consommation en électricité des habitants d'une résidence composée d'appartements de trois pièces est en moyenne de 215 *dh et d'écart-type* 27 *dh. Cette année, sur un échantillon de* 35 *appartements, les consommation en électricité* $(x_i)_{i=1}^{35}$ *vérifient*

$$\sum_{i=1}^{35} x_i = 7700 \quad et \quad \frac{1}{35}\sum_{i=1}^{35} x_i^2 = 49007.$$

Supposons que l'échantillon $(X_i)_{i=1}^{35}$ *suit une loi* $\mathcal{N}(\mu, \sigma)$.
À 99%, *peut on dire que la consommation moyenne est inférieure ou égale à* 215 *dh ?.*

1. Hypothèse de test : $\begin{cases} \mu \leq 215 & (H_0) \\ \mu > 215 & (H_1) \end{cases}$
2. Seuil de test : $\alpha = 0.01$
3. Statistique de décision : $\overline{X} = \frac{1}{35} \sum_{i=1}^{35} X_i$
4. Intervalle de rejet :

$$I_{rejet} = \left] \mu_0 + t_{1-\alpha} \frac{s}{\sqrt{n-1}} ; +\infty \right[$$

   avec $\mu_0 = 215$, $t_{0.99}(34) = 2.44$, $n = 35$, $\overline{x} = 220$ et $s = 24.64$. Donc $I_{rejet} =]225.31; +\infty[$.
5. $\overline{x} \notin I_{rejet}$ donc on accepte (H_0).
6. On peut dire, à 99%, que la consommation moyenne est inférieure ou égale à 215 *dh*.

Test de conformité sur une moyenne

Test Unilateral à gauche : Les hypothèses du test se présentent sous la forme

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu < \mu_0 \end{cases}$$

On considère comme variable de décision $\overline{X}$. La région critique du test est de la forme : $I_{rejet} =]-\infty, c[$, ou la frontière de la région critique aura pour expression :

$$c = \mu_0 - t_{1-\alpha} \frac{\sigma}{\sqrt{n-1}}.$$

Conclusion du test : Si $\overline{x}$, la valeur de la moyenne sur l'échantillon, appartient à la zone de rejet, alors on rejette (H_0). Sinon, on ne la rejette pas (on accepte H_0

Test de conformité sur une moyenne

Si on prend comme variable de décision

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n-1}}}$$

alors la région de rejet sera de la forme :

$$\widetilde{I}_{rejet} =]-\infty, -t_{1-\alpha}[.$$

On rejette H_0 si la valeur observée

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n-1}}} \in \widetilde{I}_{rejet}.$$

Exemple

Dans une étude réalisée l'an dernier, on a observé que la consommation en électricité des habitants d'une résidence composée d'appartements de trois pièces est en moyenne de 215 *dh et d'écart-type* 27 *dh. Cette année, sur un échantillon de* 35 *appartements, les consommation en électricité* $(x_i)_{i=1}^{35}$ *vérifient*

$$\sum_{i=1}^{35} x_i = 7700 \quad et \quad \frac{1}{35}\sum_{i=1}^{35} x_i^2 = 49007.$$

Supposons que l'échantillon $(X_i)_{i=1}^{35}$ *suit une loi* $\mathcal{N}(\mu, \sigma)$.
À 99%, *peut on dire que la consommation moyenne est inférieure à* 215 *dh ?.*

1. Hypothèse de test : $\begin{cases} \mu = 215 & (H_0) \\ \mu < 215 & (H_1) \end{cases}$ ou $\begin{cases} \mu \geq 215 & (H_0) \\ \mu < 215 & (H_1) \end{cases}$
2. Seuil de test : $\alpha = 0.01$
3. Statistique de décision : $\overline{X} = \frac{1}{35} \sum_{i=1}^{35} X_i$
4. Intervalle d'acceptation :

$$I_{accep} = \left] \mu_0 + t_{1-\alpha} \frac{s}{\sqrt{n-1}}; +\infty \right[$$

avec $\mu_0 = 215$, $t_{0.99}(34) = 2.44$, $n = 35$, $\overline{x} = 220$ et $s = 24.64$. Donc $I_{rejet} =]-\infty; 204.69[$.

5. $\overline{x} \notin I_{rejet}$ donc on accepte (H_0).
6. On peut dire, à 99%, que la consommation moyenne n'est pas inférieure à 215 *dh*.

Si la taille de l'échantillon est grande (en pratique $n \ggg 30$), on distingue deux cas :

- σ est connu : les résultats du paragraphe précédent restent valables.
- σ est inconnu : on l'estime par s, dans les résultats du paragraphe on remplace les quantiles de la loi Student par ceux de la loi Normale centrée réduite.

Test de conformité sur une moyenne

Cas d'un échantillon de grande taille

σ est connu :

- Test bilatéral :

$$I_{accept} = \left[\mu_0 - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \mu_0 + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

- Test unilateral à droite

$$I_{rejet} = \left]\mu_0 + u_{1-\alpha} \frac{\sigma}{\sqrt{n}} ; +\infty\right[.$$

- Test unilateral à gauche :

$$I_{rejet} = \left]-\infty ; \mu_0 - u_{1-\alpha} \frac{\sigma}{\sqrt{n}}\right[$$

σ est inconnu :

- Test bilatéral :

$$I_{accept} = \left[\mu_0 - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}; \mu_0 + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]$$

- Test unilateral à droite

$$I_{rejet} = \left]\mu_0 + u_{1-\alpha}\frac{s}{\sqrt{n}}; +\infty\right[.$$

- Test unilateral à gauche :

$$I_{rejet} = \left]-\infty; \mu_0 - u_{1-\alpha}\frac{s}{\sqrt{n}}\right[$$

Exercice

Une société s'approvisionne en pièces brutes, qui conformément aux conditions fixées par le fournisseurs, doivent avoir une masse moyenne de 780g.
On prélève au hasard un échantillon de 36 *pièces dont on mesure la masse. La masse moyenne des pièces de l'échantillon est de* 775g.
En supposant que la masse d'une pièces est Normale de moyenne μ *et d'écart-type* $\sigma = 12,5g$, *peut on considérer que la masse moyenne est comme le prévoient les conditions fixées par le fournisseur*

1. *à un seuil de signification de* 5% *?*
2. *à un seuil de signification de* 1% *?*

Exercice

La presse affirme que parmi les femmes regardant la télévision, la durée moyenne devant un poste est supérieure à 3h/*jour*.

Sur un échantillon de 100 *femmes, on a observé les données suivantes :*

Durée	[0 ;1[	[1 ;2[	[2 ;3[	[3 ;4[	[4 ;5[	[5 ;6[	≥ 6
Nombre de femmes	15	12	19	25	16	8	5

Que pensez vous de cette affirmation de la presse au seuil de 10% *?.*

Exercice

On veut étudier X le salaire horaire des ouvriers d'un secteur privé. On choisit au hasard d'un échantillon aléatoire de taille n.

1. *On suppose que $n = 20$ et que X suit une loi normale de moyenne μ et de variance $\sigma^2 = 4$. Sachant que*

$$\sum_{i=1}^{20} x_i = 217\ DH, \quad \sum_{i=1}^{20} x_i^2 = 2308$$

   1. *le salaire horaire est-il en moyenne égal à 10 DH au seuil de signification $\alpha = 0,1$?*
   2. *le salaire horaire est-il en moyenne supérieur à 12 DH au seuil de signification $\alpha = 0,01$?.*

suite

2. On suppose que $n = 50$ et que X est une v.a. de variance $\sigma^2 = 4$ telle que

$$\sum_{i=1}^{50} x_i = 529 \ DH, \quad \sum_{i=1}^{50} x_i^2 = 5606.$$

le salaire horaire est-il en moyenne inférieure ou égal à 10 DH au seuil de signification 99% ?

Exercice

Pour estimer le marché potentiel d'un nouveau produit laitier, un sondage est effectué auprès de 90 consommateurs de ce produit de la ville de Larache. Il en résulte que l'estimation de la consommation moyenne est de $7,4$ *unités par semaine avec un écart-type estimé de* $3,1$*. Peut on dire que, à 80% la consommation moyenne de ce produit par les habitants de Larache est égale à* $7,4$ *unités par semaine ?*

Tests de conformité sur une variance d'une v.a Gaussienne

Si X suit une loi normale $N(\mu, \sigma)$, alors on peut les tests suivants

$$\begin{cases} H_0 : & \sigma^2 = \sigma_0^2 \\ H_1 : & \sigma^2 \neq \sigma_0^2 \end{cases} \quad \begin{cases} H_0 : & \sigma^2 = \sigma_0^2 \\ H_1 : & \sigma^2 > \sigma_0^2 \end{cases} \quad \begin{cases} H_0 : & \sigma^2 = \sigma_0^2 \\ H_1 : & \sigma^2 < \sigma_0^2 \end{cases}$$

On se fixe un seuil α ou un niveau de confiance $1 - \alpha$. On distingue deux cas :

- μ connue
- μ inconnue

Tests de conformité sur une variance d'une v.a Gaussienne

μ est connue : On prend comme variable de décision :

$$T^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Si $\sigma^2 = \sigma_0^2$, alors $\frac{nT^2}{\sigma_0^2}$ suit une loi $\chi_2(n)$.

Tests de conformité sur une variance d'une v.a Gaussienne

Test bilatéral : On cherche la région d'acceptation sous la forme $[c_1, c_2]$. Soient $k_{n(\frac{\alpha}{2})}$ et $k_{n(\frac{1-\alpha}{2})}$ les réels déterminés dans la table de la loi $\chi_2(n)$, tels que

$$\left\{ \begin{array}{c} \mathbb{P}\left(\dfrac{nT^2}{\sigma_0^2} \leq k_{n(1-\alpha/2)}\right) = 1-\alpha/2 \\ \mathbb{P}\left(\frac{nT^2}{\sigma_0^2} \leq k_{n(\alpha/2)}\right) = \alpha/2 \end{array} \right.$$

L'intervalle d'acceptation pour T^2 au risque α est

$$I_{accept} = \left[\frac{\sigma_0^2}{n} k_{n(\alpha/2)}, \frac{\sigma_0^2}{n} k_{n(1-\alpha/2)}\right].$$

Conclusion : Si t^2, la réalisation de T^2 appartient à I_{accept}, on accept (H_0). Sinon, on la rejette.

Exemple

On veut étudier X le salaire horaire des ouvriers d'un secteur privé. On choisit au hasard 20 salariés. On suppose que la moyenne des salaires est $12 dh/h$ et que les salaires sont supposés distribués suivant une loi Normale de variance σ^2. La réalisation de l'échantillon donne l'information suivante :

$$\sum_{i=1}^{20} x_i = 236, \qquad \sum_{i=1}^{20} x_i^2 = 2960.$$

Au seuil $\alpha = 10\%$, la variance est elle égale à 4 ?

1. Hypothèses de test : $\begin{cases} \sigma^2 = 4 & (H_0) \\ \sigma^2 \neq 4 & (H_1) \end{cases}$
2. Seuil $\alpha = 0,1$
3. Variable de décision : $T^2 = \frac{1}{20}\sum_{i=1}^{20}(X_i - \mu)^2$
4. Intervalle d'acceptation

$$I_{accep} = \left[\frac{\sigma_0^2}{n} k_{n(\alpha/2)}; \frac{\sigma_0^2}{n} k_{n(1-\alpha/2)}\right]$$

avec $\sigma_0^2 = 4,\ n = 20,\ k_{20(0{,}05)} = 10,85$ et $k_{20(0{,}95)} = 31,41$.
Donc $I_{accep} = [2,17; 6,28]$.

5. Réalisation de T^2 :

$$t^2 = \frac{1}{20}\sum_{i=1}^{20}(x_i - \mu)^2 = \frac{1}{20}\sum_{i=1}^{20} x_i^2 - 2\mu\bar{x} + \mu^2 = 8,8$$

Puisque $t^2 \notin I_{accep}$ on rejette (H_0).

6. À 90%, on ne peut pas dire que la variance des salaires est égale à 4.

Tests de conformité sur une variance d'une v.a Gaussienne

Test Unilatéral à droite : On cherche la région critique sous la forme $]t_1, +\infty[$. Soit $k_{n(1-\alpha)}$ le réel déterminé dans la table de la loi $\chi_2(n)$ par

$$\mathbb{P}\left(\frac{nT^2}{\sigma_0^2} < k_{n(1-\alpha)}\right) = 1 - \alpha.$$

La région critique ou intervalle de rejet pour T^2 au risque α est

$$I_{rejet} = \left]\frac{\sigma_0^2}{n} k_{n(1-\alpha)}, +\infty\right[.$$

Conclusion : Si t^2, la réalisation de T^2, appartient à I_{rejet} on accepte (H_1). Sinon on la rejette.

Exemple

On veut étudier X le salaire horaire des ouvriers d'un secteur privé. On choisit au hasard 20 salariés. On suppose que la moyenne des salaires est $12dh/h$ et que les salaires sont supposés distribués suivant une loi Normale de variance σ^2. La réalisation de l'échantillon donne l'information suivante :

$$\sum_{i=1}^{20} x_i = 236, \qquad \sum_{i=1}^{20} x_i^2 = 2960.$$

À un niveau de confiance de 90%, la variance est elle supérieure à 4 ?

1. Hypothèses de test : $\begin{cases} \sigma^2 = 4 & (H_0) \\ \sigma^2 > 4 & (H_1) \end{cases}$
2. Seuil $\alpha = 0,1$
3. Variable de décision : $T^2 = \frac{1}{20}\sum_{i=1}^{20}(X_i - \mu)^2$
4. Intervalle d'acceptation

$$I_{rejet} = \left] \frac{\sigma_0^2}{n} k_{n(1-\alpha)} ; +\infty \right[$$

avec $\sigma_0^2 = 4$, $n = 20$, $k_{20(0,90)} = 28,41$. Donc $I_{rejet} =]5,68; +\infty[$.

5. Réalisation de T^2 :

$$t^2 = \frac{1}{20}\sum_{i=1}^{20}(x_i - \mu)^2 = \frac{1}{20}\sum_{i=1}^{20} x_i^2 - 2\mu\bar{x} + \mu^2 = 8,8$$

Puisque $t^2 \in I_{rejet}$, on accepte (H_1).

6. À 90%, on peut dire que la variance des salaires est supérieure à 4.

Tests de conformité sur une variance d'une v.a Gaussienne

Test Unilatéral à gauche : On cherche la région critique sous la forme $]-\infty; t_1[$. Soit $k_{n(\alpha)}$ le réel déterminé dans la table de la loi $\chi_2(n)$ par

$$\mathbb{P}\left(\frac{nT^2}{\sigma_0^2} < k_{n(\alpha)}\right) = \alpha.$$

La région critique ou intervalle de rejet pour T^2 au risque α est

$$I_{rejet} = \left]-\infty; \frac{\sigma_0^2}{n} k_{n(\alpha)}\right[.$$

Conclusion : Si t^2, la réalisation de T^2 appartient à I_{rejet} on accepte (H_1). Sinon on la rejette.

Exemple

On veut étudier X le salaire horaire des ouvriers d'un secteur privé. On choisit au hasard 20 salariés. On suppose que la moyenne des salaires est $12dh/h$ et que les salaires sont supposés distribués suivant une loi Normale de variance σ^2. La réalisation de l'échantillon donne l'information suivante :

$$\sum_{i=1}^{20} x_i = 236, \qquad \sum_{i=1}^{20} x_i^2 = 2960.$$

À un niveau de confiance de 90%, la variance est elle est inférieure à $9,5$?.

1. Hypothèses de test : $\begin{cases} \sigma^2 = 4 & (H_0) \\ \sigma^2 < 9,5 & (H_1) \end{cases}$
2. Seuil $\alpha = 0,1$
3. Variable de décision : $T^2 = \frac{1}{20}\sum_{i=1}^{20}(X_i - \mu)^2$
4. Intervalle d'acceptation

$$I_{rejet} = \left]-\infty; \frac{\sigma_0^2}{n} k_{n(\alpha)}\right[$$

avec $\sigma_0^2 = 9,5$ $n = 20$, $k_{20(0,10)} = 12,44$. Donc $I_{rejet} =]-\infty; 5,91[$.

5. Réalisation de T^2 :

$$t^2 = \frac{1}{20}\sum_{i=1}^{20}(x_i - \mu)^2 = \frac{1}{20}\sum_{i=1}^{20} x_i^2 - 2\mu\bar{x} + \mu^2 = 8,8$$

Puisque $t^2 \notin I_{rejet}$, on rejette (H_1).

6. À 90%, on peut dire que la variance des salaires n'est pas inférieure à 9, 5.

Tests de conformité sur une variance d'une v.a Gaussienne

Remarque

- *Si on choisit comme variable de décision* $\frac{nT^2}{\sigma_0^2}$*, l'intervalle d'acceptation pour* $\frac{nT^2}{\sigma_0^2}$ *au risque* α *pour un test bilatéral est*

$$I_{accep} = \left[k_{n(\frac{\alpha}{2})}, k_{n(1-\frac{\alpha}{2})}\right].$$

- *L'intervalle de rejet pour* $\frac{nT^2}{\sigma_0^2}$ *au risque* α*, pour une test unilateral à droite et à gauche est respectivement*

$$\left]k_{n(1-\alpha)}, +\infty\right[\qquad et \qquad \left]-\infty, k_{n(\alpha)}\right[.$$

Tests de conformité sur une variance d'une v.a Gaussienne

μ est inconnue : On a

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)\bar{S}^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi_2(n-1).$$

On reprend les résultats de précédents en remplaçant T^2 par S^2 et $\chi_2(n)$ par $\chi_2(n-1)$.

Tests de conformité sur une variance d'une v.a Gaussienne

- Intervalle d'acceptation pour S^2 dans un test bilatéral :

$$I_{accept} = \left[\frac{\sigma_0^2}{n} k_{n-1(\frac{\alpha}{2})} ; \frac{\sigma_0^2}{n} k_{n-1(1-\frac{\alpha}{2})} \right].$$

- Intervalle de rejet pour S^2 dans un test unilatéral à droite :

$$I_{rejet} = \left] \frac{\sigma_0^2}{n} k_{n-1(1-\alpha)} ; +\infty \right[.$$

- Intervalle de rejet pour S^2 dans un test unilatéral à gauche :

$$I_{rejet} = \left] -\infty ; \frac{\sigma_0^2}{n} k_{n-1(\alpha)} \right[.$$

Tests de conformité sur une variance d'une v.a Gaussienne

Exemple

On veut étudier X le salaire horaire des ouvriers d'un secteur privé. On choisit au hasard 20 salariés. On suppose que les salaires sont supposés distribués suivant une loi Normale $\mathcal{N}(\mu, \sigma^2)$. La réalisation de l'échantillon donne l'information suivante :

$$\sum_{i=1}^{20} x_i = 236, \qquad \sum_{i=1}^{20} x_i^2 = 2960.$$

Au seuil $\alpha = 10\%$, la variance est elle égale à 9.5 ?

1. Hypothèses de test : $\begin{cases} \sigma^2 = 9.5 & (H_0) \\ \sigma^2 \neq 9.5 & (H_1) \end{cases}$
2. Seuil $\alpha = 0,1$
3. Variable de décision : $S^2 = \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2$
4. Intervalle d'acceptation

$$I_{accep} = \left[\frac{\sigma_0^2}{n} k_{n-1(\alpha/2)}; \frac{\sigma_0^2}{n} k_{n-1(1-\alpha/2)}\right]$$

avec $\sigma_0^2 = 9.5$, $n = 20$, $k_{19(0,05)} = 10,12$ et $k_{19(0,95)} = 30,14$.
Donc $I_{accep} = [4.81; 14.32]$.

5. Réalisation de S^2 :

$$s^2 = \frac{1}{20}\sum_{i=1}^{20}(x_i - \bar{x})^2 = \frac{1}{20}\sum_{i=1}^{20} x_i^2 - \bar{x}^2 = 8,76$$

Puisque $s^2 \in I_{accep}$ on accepte (H_0).

6. À 90%, on peut dire que la variance des salaires est égale `

Exercice

En 2010, le bénéfice moyen par action pour la population des sociétés de services financiers était de moyenne $22,5$ *DH et d'écart-type* $3,5$ *DH. Pour le premier trimestre de 2011, un échantillon de sociétés de services financiers* $(X_i)_{i=1}^{10}$ *a fourni des bénéfices par actions* $(x_i)_{i=1}^{10}$ *tels que* $\sum_{i=1}^{10} x_i = 240$ *et* $\sum_{i=1}^{10} x_i^2 = 5920$*. On suppose de plus que les bénéfices suivent une loi Normale de moyenne* μ *et d'écart-type* σ*.*
Au seuil de signification $\alpha = 1\%$*, peut-on dire que l'écart-type est plus grand que* $3,5$ *DH ?.*

Exercice

Une grande surface hésite de mener une compagne publicitaire pour la vente des TV-LED. On a choisit 7 semaines au hasard. Les ventes $(X_i)_{i=1}^{7}$ des TV-LED et de réalisation $(x_i)_{i=1}^{7}$ telles que

$$\sum_{i=1}^{7} x_i = 106 \quad et \quad \sum_{i=1}^{7} x_i^2 = 1636.$$

Supposons que les ventes des TV-LED sont de loi Normale de moyenne μ et de variance σ^2. Au seuil de 5%, peut on dire que l'écart-type des ventes est égale à 3 TV-LED/semaine ?

Tests de conformité sur une proportion

Soit p la proportion de la population possédant le caractère considéré. On veut effectuer un test

$$\left\{ \begin{array}{ll} H_0 : & \qquad\qquad p = p_0 \\ H_1 : & p > p_0, \ p \neq p_0, \quad p < p_0. \end{array} \right.$$

On prend comme variable de décision $\widehat{p}$. Si $p = p_0$, alors la loi de $\widehat{p}$ est normale $N(p_0, \sqrt{\frac{p_0(1-p_0)}{n}})$.

Tests de conformité sur une proportion

Test bilateral : L'intervalle d'acceptation pour $\widehat{p}$ au risque α est

$$I_{accept} = \left[p_0 - u_{1-\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}} \;\; ; \;\; p_0 + u_{1-\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}} \right].$$

Conclusion : Si la fréquence f sur l'échantillon appartient à I_{accept}, on accept (H_0). Sinon, on accepte (H_1).

Tests de conformité sur une proportion

Test Unilateral à droite : L'intervalle de rejet de $\widehat{p}$ au risque α est

$$I_{rejet} = \left] p_0 + u_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}} \;\; ; \;\; 1 \right].$$

Conclusion : Si la fréquence f sur l'échantillon appartient à I_{rejet}, on accepte (H_1). Sinon, on la rejette.

Tests de conformité sur une proportion

Test Unilateral à gauche : L'intervalle de rejet de $\widehat{p}$ au risque α est

$$I_{rejet} = \left[0, p_0 - u_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}\right[.$$

Conclusion **:** Si la fréquence f sur l'échantillon appartient à I_{rejet}, on accepte (H_1). Sinon, on la rejette.

Exercice

Pour évaluer chez les enfants de dix ans les capacités de mémorisation à partir d'un test écrit, on a demandé à 30 *enfants de lire une liste de* 15 *mots, puis quelques minutes après d'écrire ceux dont ils se souviennent.*
On a compté le nombre des mots mémorisés par les enfants de l'échantillon que voici les résultats obtenues :

4	6	2	11	5	13	14	7	8	2
6	11	8	13	13	3	11	9	8	9
12	7	11	15	15	8	14	4	12	5

Peut on affirmer que, au seuil de 5%,

- 25% *d'enfants trouvent au plus* 5 *mots ?*
- *au moins* 45% *d'enfants trouvent entre* 6 *et* 10 *mots ?*
- *plus de* 30% *d'enfants trouvent de* 11 *à* 15 *mots ?.*

Exercice

Deux semaines avant le contrôle final, le service des affaires pédagogique de notre faculté a réalisé une enquête auprès des étudiants afin de savoir est ce que les étudiants sont prêt ou non à passer le contrôle. 70% *des étudiants interrogés ont affirmé qu'ils sont prêt.*
Nous avons mené la même enquête une semaine avant le contrôle de rattrapage auprès d'un échantillon aléatoire de 24 étudiants. 17 *d'entre eux sont prêt à passer le contrôle.*
Au seuil $\alpha = 1\%$, *peut-on dire que le pourcentage des étudiants qui étaient prêt à passer le contrôle finale est le même que celui de ceux qui vont passer le contrôle de rattrapage ?.*

Exercices de révision

Exercice 1 : Après la correction de certaines copies d'un épreuve de contrôle continu, on a constaté que la moyenne des notes est de 12 avec un écart-type égal à 3.68. On considère que la note d'un étudiant suit une loi Normale de moyenne μ et de variance σ^2. On désire, à 99%, avoir une idée générale sur la note moyenne de tous les étudiants de la section avec une marge d'erreur de 30%. Soit $(X_1, X_2, ..., X_n)$ un échatillon de X de réalisation $(x_1, x_2, ..., x_n)$.

1. Donner la taille minimale de l'échantillon considéré.

Supposons que $n = 1000$. Les réalisation $(x_i)_i$ vérifient

$$\sum_{i=1}^{1000} x_i = 12521 \quad \text{et} \quad \frac{1}{1000}\sum_{i=1}^{1000} x_i^2 = 217.4.$$

1. Donner un estimateur sans bias de σ^2 et donner une estimation de σ
2. Donner un intervalle de confiance de la note moyenne μ.

Exercice 2 : Le service pédagogique de la faculté désire étudier le taux d'absentéisme des étudiants de première année dans les cours de la langue française. Soit X la variable aléatoire nombre de séance ratées et $(X_1, X_2, ..., X_{90})$ un échantillon de X réalisant $\sum_{i=1}^{90} x_i = 382.5$ et $\sum_{i=1}^{90} x_i^2 = 1822.5$. Supposons que X suit une loi $\mathcal{N}(\mu, \sigma)$.

1. Supposons que $\mu = 4$.
    1. Donner un estimateur sans biais de σ^2 et donner ρ une estimation de σ
    2. Si on considère que $\sigma = \rho$, calculer la probabilité que le nombre de séances ratées par un étudiant soit supérieur à 5.
2. Trouver un intervalle de confiance de σ^2 à un niveau de 95%.

Exercice 3 : Une chaîne de télévision publie chaque mois la côte de popularité du premier ministre. Un sondage auprès de 1024 à montré que seulement 350 déclaraient lui faire confiance. Soit p le pourcentage de la population faisant confiance au premier ministre.

1. Donner un estimateur sans bias et une estimation de p.
2. Trouver l'intervalle de confiance de p à un seuil de 10%.

Exercice 4 : Après la correction de certaines copies d'un épreuve de contrôle continu, on a constaté que la moyenne des notes est de 11 avec un écart-type égal à 3.84. On considère que X la note d'un étudiant suit une loi Normale de moyenne μ et de variance σ^2. On désire, à 99%, avoir une idée générale sur la note moyenne de tous les étudiants. On considère $(X_i)_{i=1}^{81}$ un échatillon de X de réalisation $(x_i)_{i=1}^{81}$ vérifiant

$$\sum_{i=1}^{81} x_i = 818 \quad \text{et} \quad \frac{1}{81}\sum_{i=1}^{81} x_i^2 = 116.74$$

1. Peut-on dire que l'écart-type σ est égal à 3.84 ?.
2. Supposons que $\sigma = 3.84$. Peut-on dire que la note moyenne des étudiants est égale `

Exercice 5 : Un échantillon de 676 abonnés à un site forum a révélé que chaque membre de forum passait en moyenne 3.4 heures par jours sur le forum avec un écart-type égale à 1.3 heures. Peut-on dire, à un risque de 1%, que le temps moyen passé par les membres de forum est supérieur à 3 heures par jour ?.

Exercice 6 : Un grand quotidien publie chaque mois la côte de popularité de l'entraîneur de l'équipe nationale de football. Le mois précédent, 51% des personnes déclarent lui faire confiance. Ce mois-ci, à la suite d'un sondage auprès de 1024 personnes, seulement 492 ont encore confiance à l'entraîneur national. Peut-on affirmer, à un niveau de confiance de 95%, que la côte de popularité de l'entraîneur de l'équipe nationale de football a baissé ?.

Fin