Statistique L3PRO

Josef WILCZEK

Planning

- **Séance 1 (07/11)**
- Rappels (distribution normale, échantillonnage)
- Intervalle de confiance
  - Moyenne
  - Variance et écart type
  - Pourcentage

- **Séance 2 - Test usuels (10/11)**
  - Hypothèses
  - Théorie de la statistique de décision
  - Comparaison de deux moyennes expérimentales (grands et petits échantillons)
  - Comparaison de moyennes de deux échantillons appariés
  - Comparaison de deux fréquences expérimentales
  - Comparaison de deux variances expérimentales

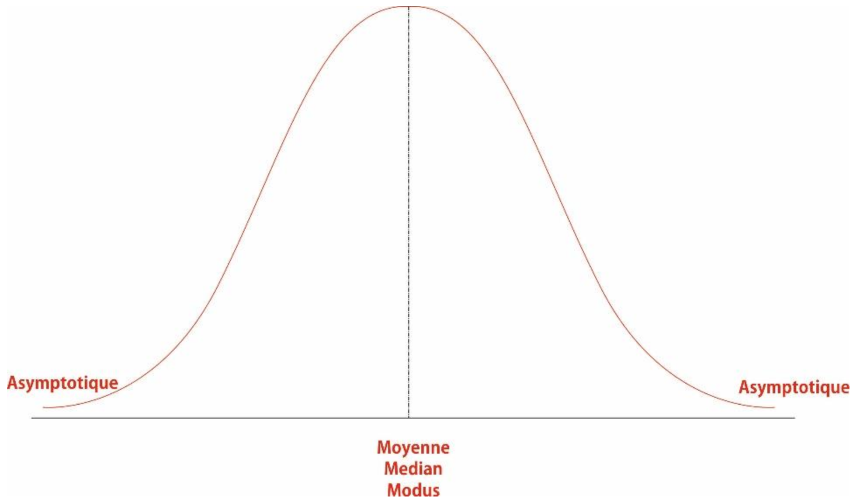
- **Séance 3 (14/11)**
  - Répétition, exemples…

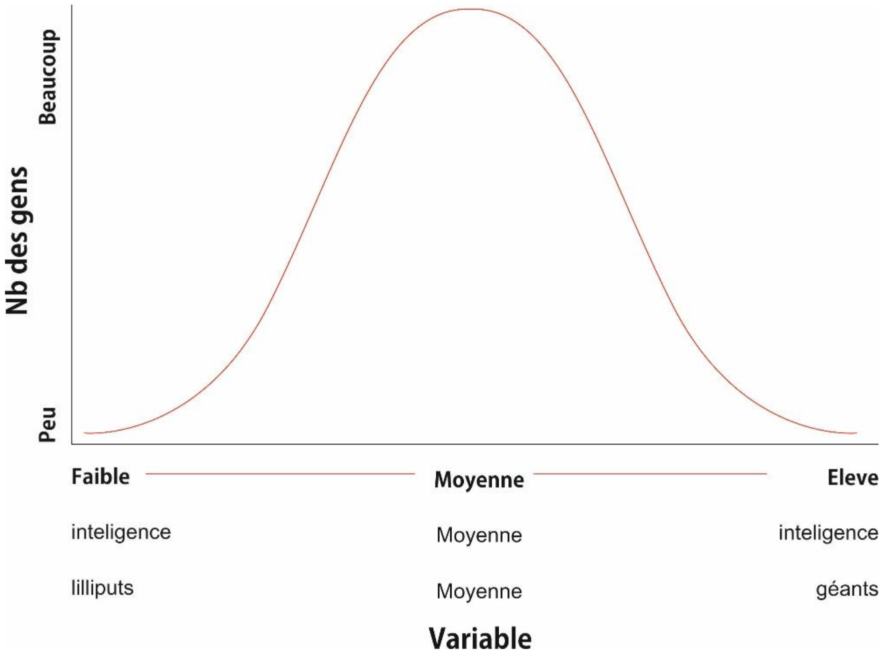
- **Matériels**
- Regarder les cours L1, L2
- Les présentations et la bibliographie sont en ligne:
- www.jwilczek.com
- www.fabricemonna.com

La distribution normale

La distribution normale
/ La courbe de la cloche / La courbe de Gausse

- La représentation visuelle des valeurs, ou les trois tendances centrales (moyenne = modus = médian) sont égales
- La courbe est parfaitement symétrique et asymptotique
- Exemple:
  - La distribution des tailles en population des gens



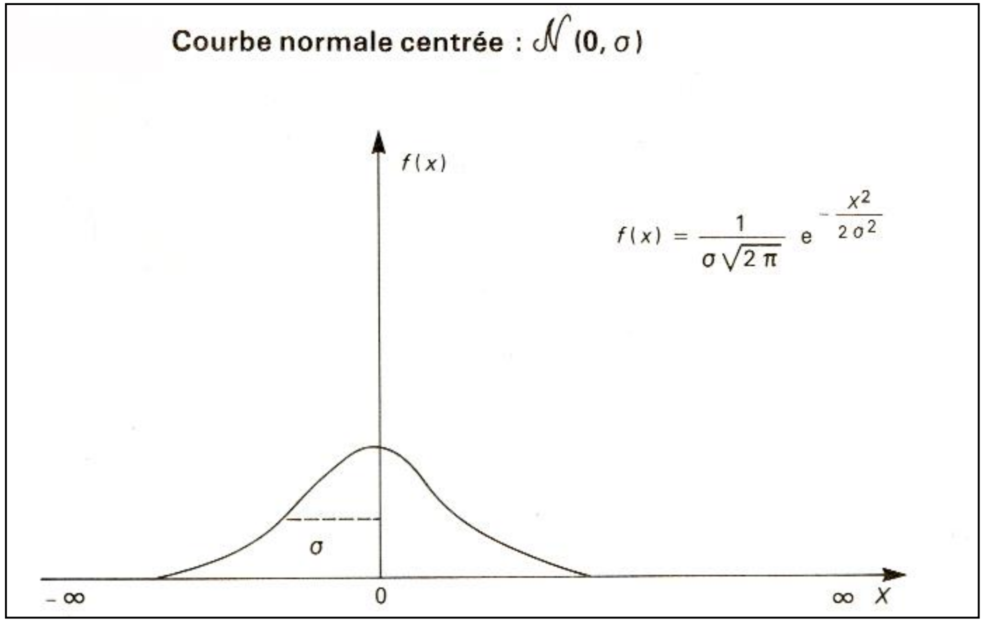


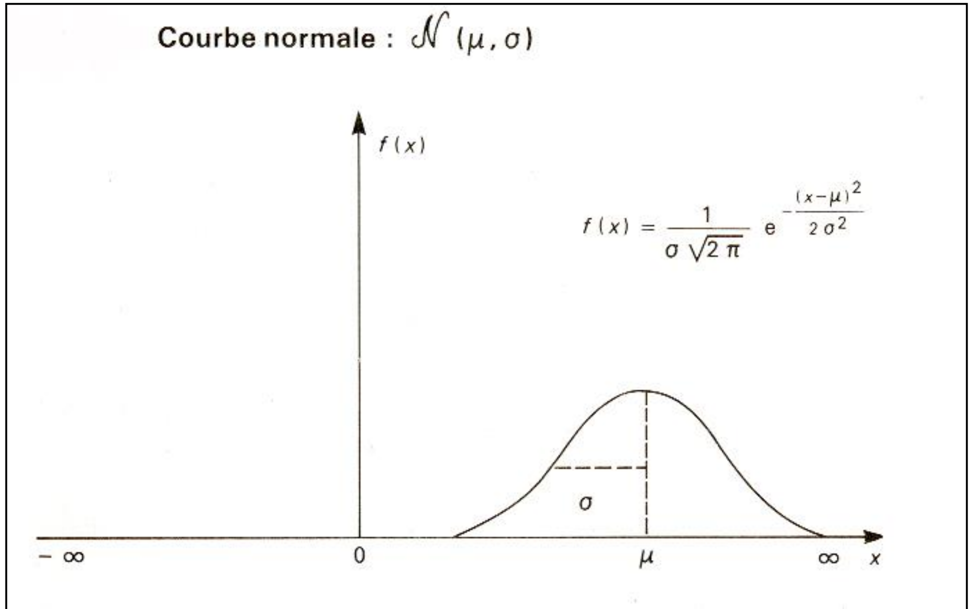
La distribution normale
/ La courbe de la cloche / La courbe de Gausse

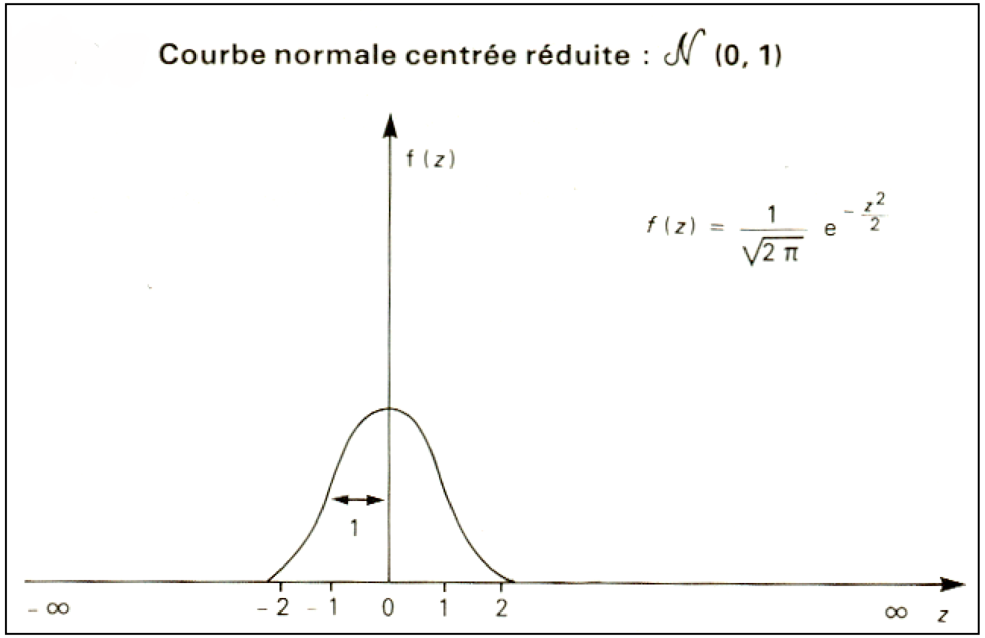
- Loi normale centrée réduite: moyenne = 0, écart type = 1
  - 1ere transformation: $X = x - \bar{x}$
  - 2eme transformation: $Z = X/s_x$

- $z = \frac{(x-\bar{x})}{s_x}$

- **z-score** de chaque valeur exprime, combiens des écartes types cette valeur est éloigné de la moyenne

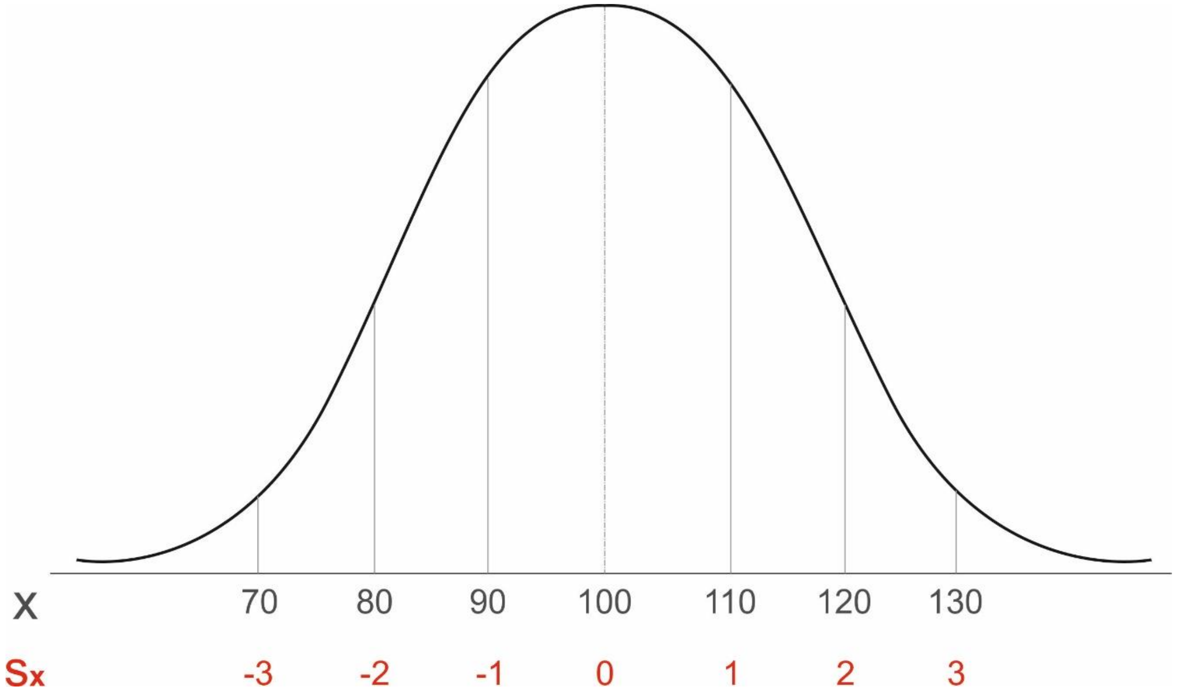






La distribution normale

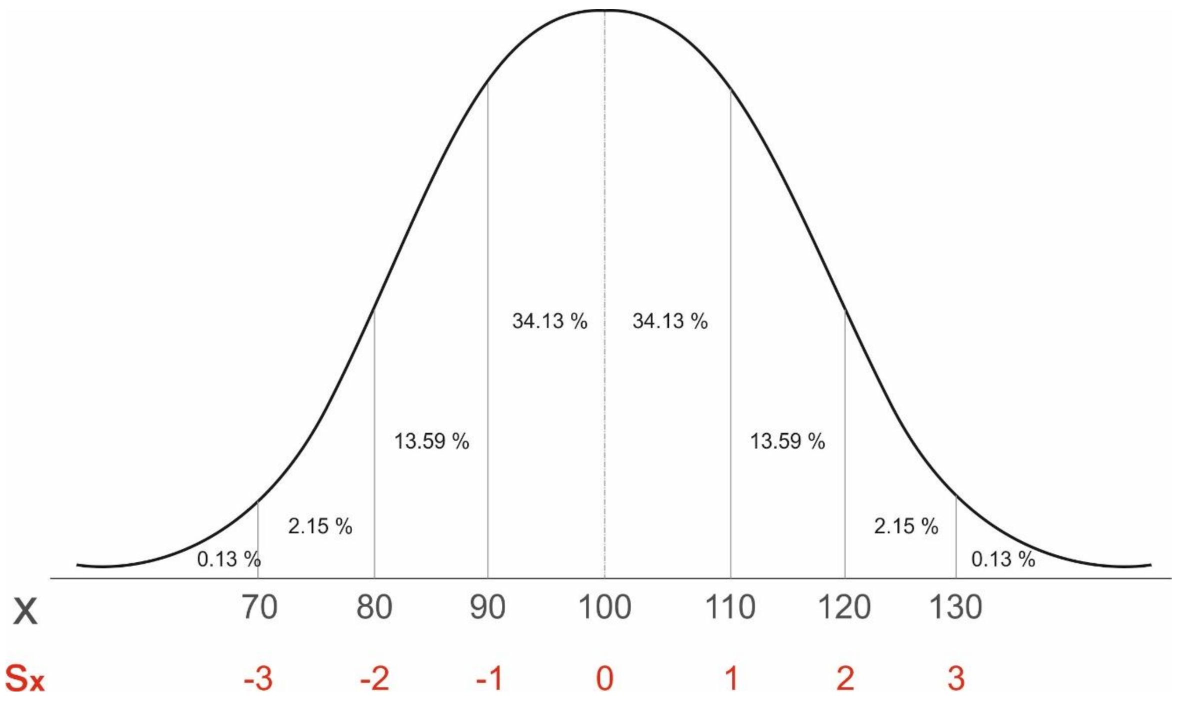
- $\bar{x} = 100$
- $s_x = 10$
- Sur l'axe X on a les valeurs distribuées toujours par les valeurs de l'écart type



La distribution normale

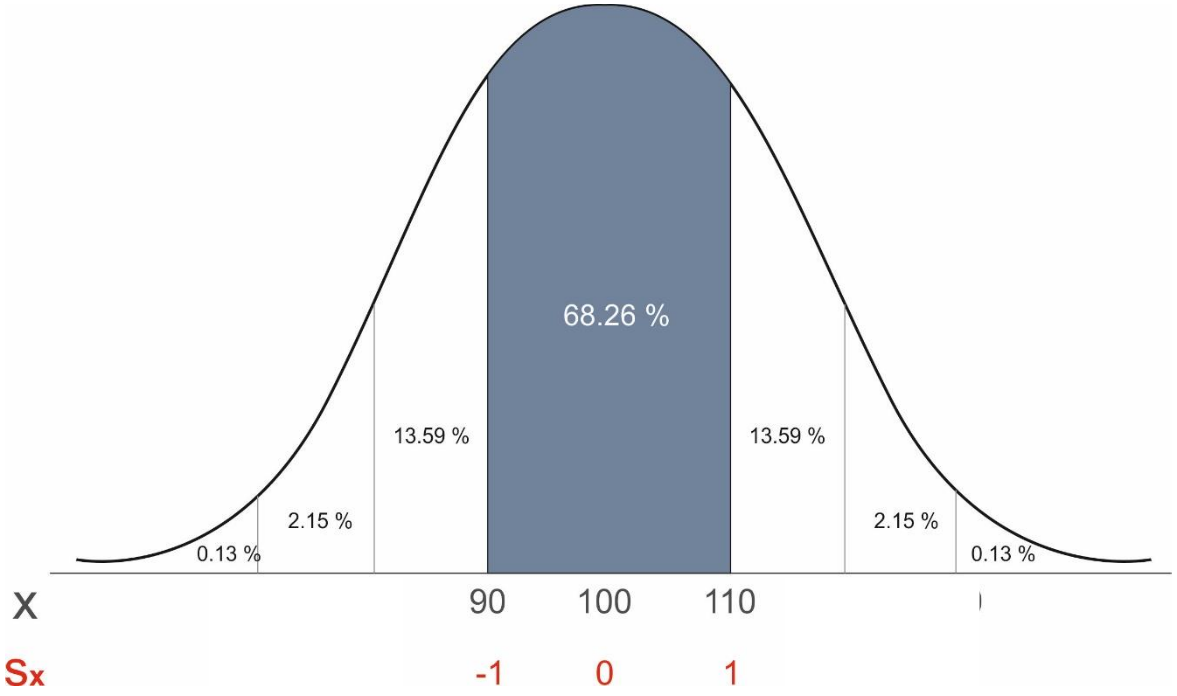
- $\bar{x} = 100$
- $s_x = 10$
- Sur l'axe X on a les valeurs distribuées toujours par les valeurs de l'écart type

Les surfaces au dessous de la courbe représentent la probabilité d'occurence d'un certain valeur.



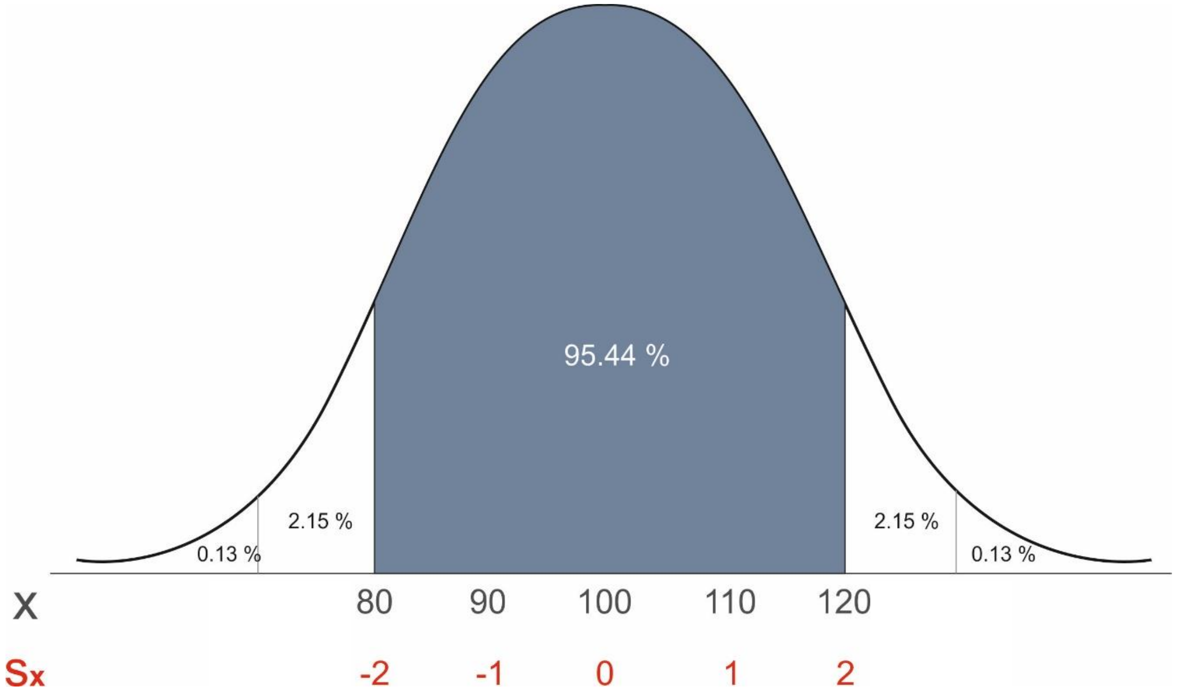
La distribution normale

- $\bar{x} = 100$
- $s_x = 10$
- Sur l'axe X on a les valeurs distribuées toujours par les valeurs de l'écart type
- Si les valeurs sont distribuées normalement
  - Presque 70 % des valeurs se trouvent entre -1 s_x et +1 s_x de la moyenne



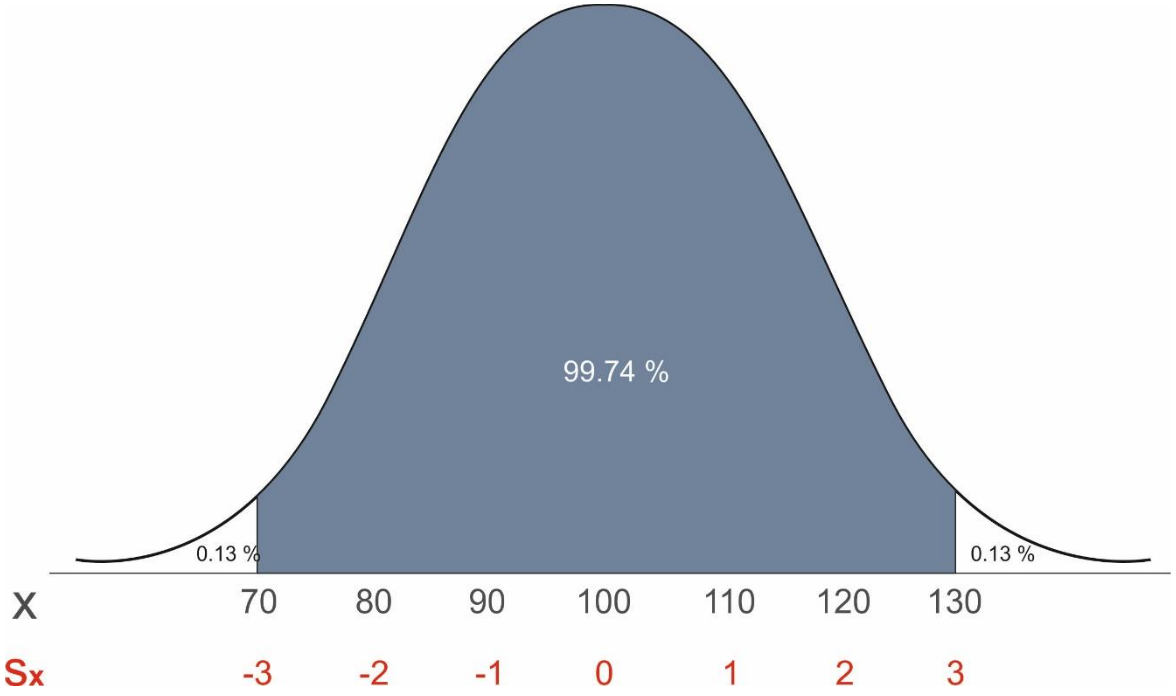
La distribution normale

- $\bar{x} = 100$
- $s_x = 10$
- Sur l’axe X on a les valeurs distribuées toujours par les valeurs de l’écart type
- Si les valeurs sont distribuées normalement
  - Presque 70 % des valeurs se trouvent entre -1 s_x et +1 s_x de la moyenne
  - 95% des valeurs se trouvent entre -2 et +2 s_x



La distribution normale

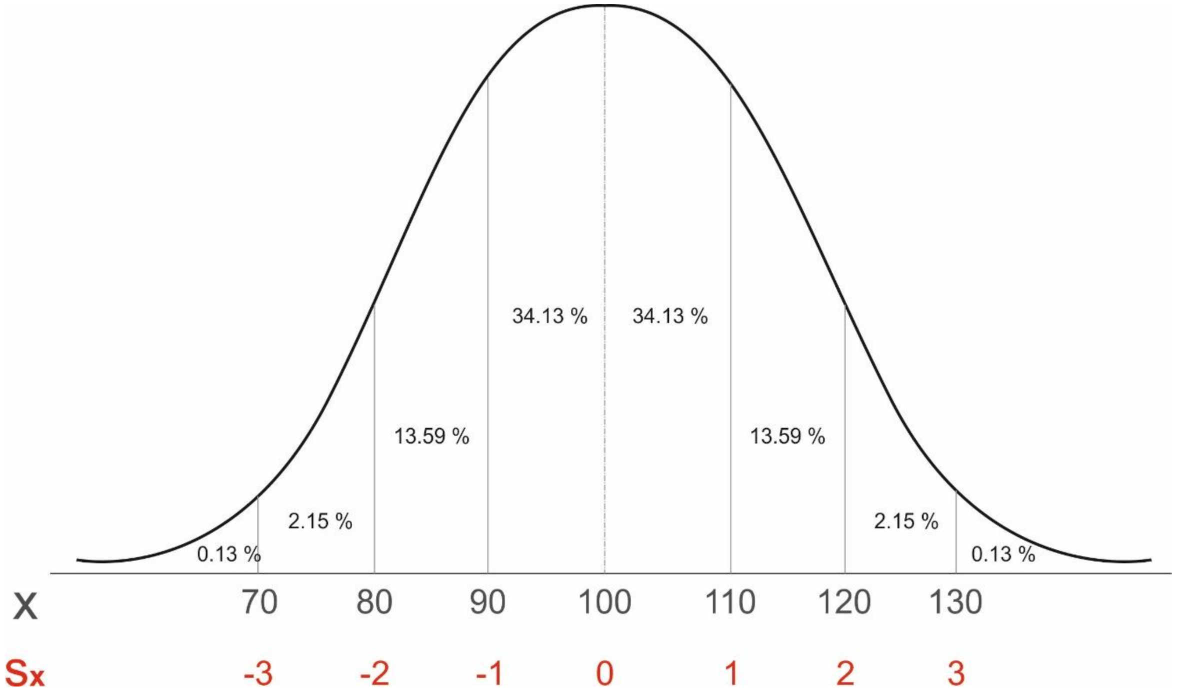
- $\bar{x} = 100$
- $s_x = 10$
- Sur l'axe X on a les valeurs distribuées toujours par les valeurs de l'écart type
- Si les valeurs sont distribuées normalement
    - Presque 70 % des valeurs se trouvent entre -1 s_x et +1 s_x de la moyenne
    - 95% des valeurs se trouvent entre -2 et +2 s_x
    - Presque 100 % des valeurs se trouvent entre -3 do +3 s_x



La distribution normale

- L'aire (pourcentage) au dessous de la courbe représentent **la probabilité de la presence d'une certaine valeur**

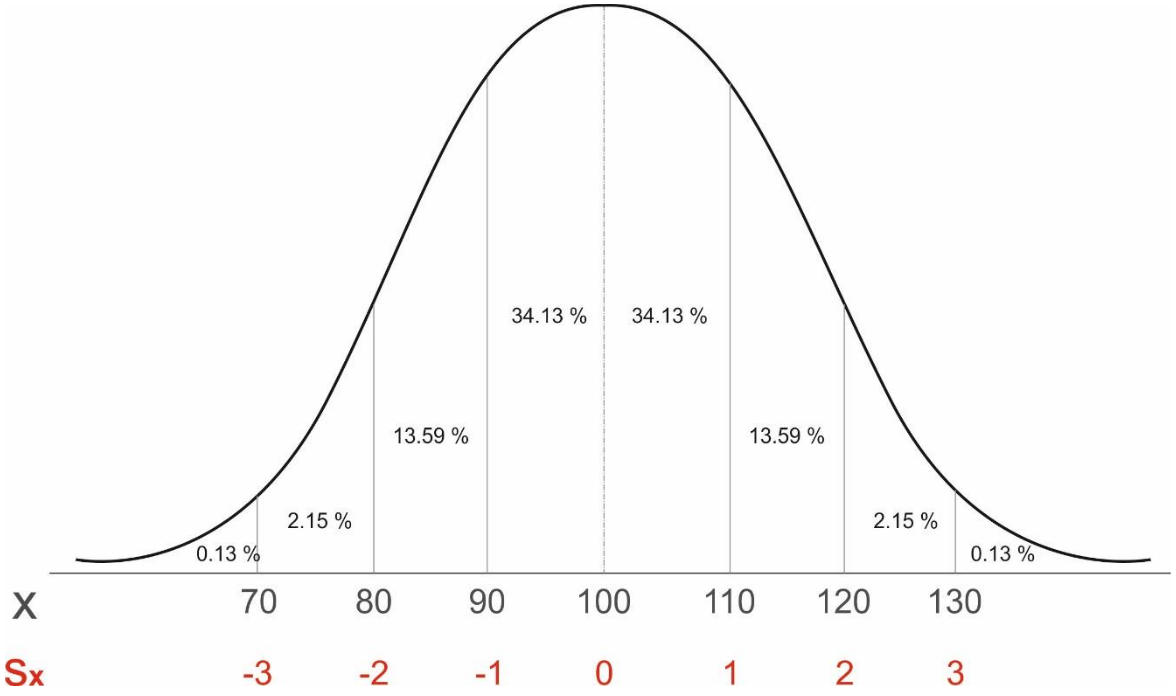
- Exemple: Quelle est la probabilité que dans cette distribution, une certaine valeur sera plus grande que 110 ?



La distribution normale

- L'aire (pourcentage) au dessous de la courbe représentent **la probabilité de la presence d'une certaine valeur**

- Exemple: Ou est-ce qu'on trouve les 95 % de gens de la taille « normale » ?

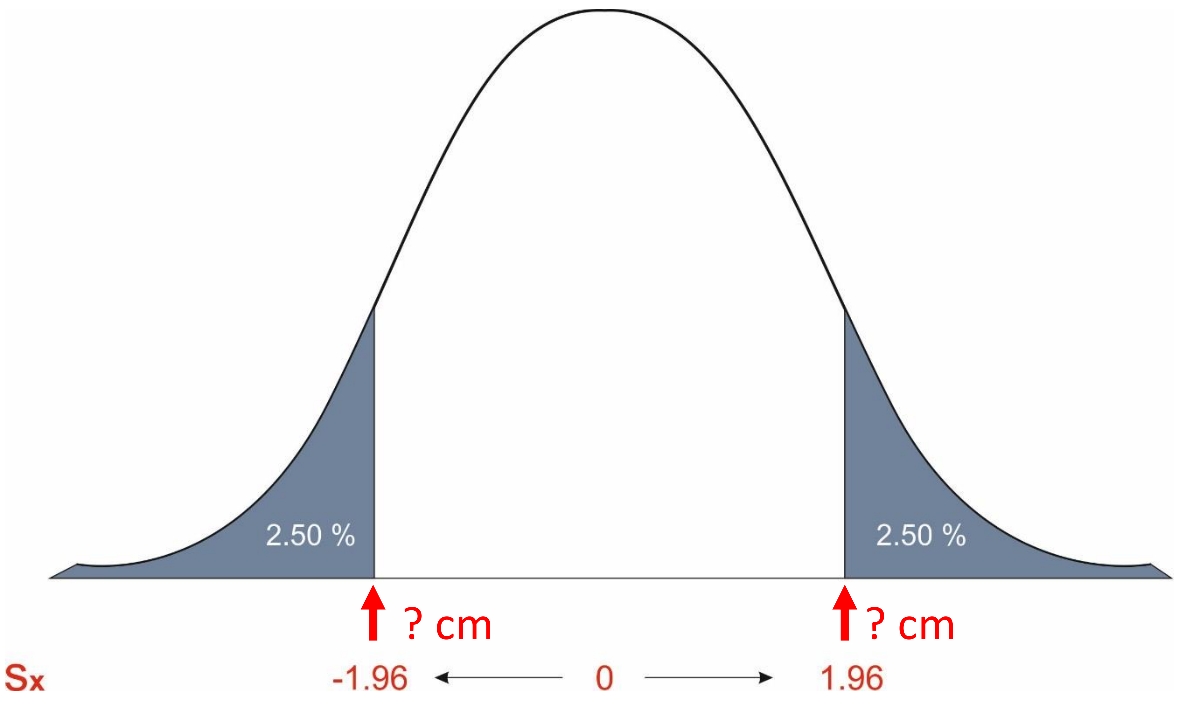


La distribution normale

- L'aire (pourcentage) au dessous de la courbe peut servir aussi pour **détermination d'une certaine valeur sur l'axe**

- Exemple: Ou est-ce qu'on trouve les 95 % de gens de la taille « normale » ?
- Et le reste (les gens trop grandes et trop petites) ?

- Oui mais se sont les quelles valeurs (en cm) sur l'axe x ?

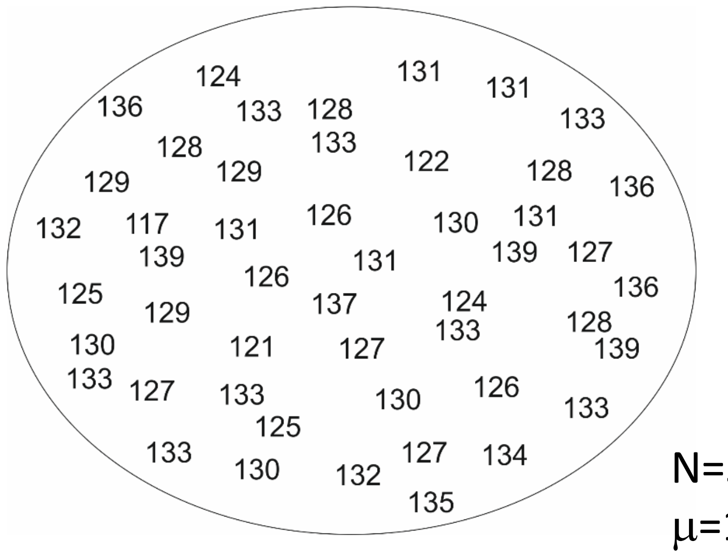


Intervalles de confiance

Echantillonnage – Estimation d'un paramètre

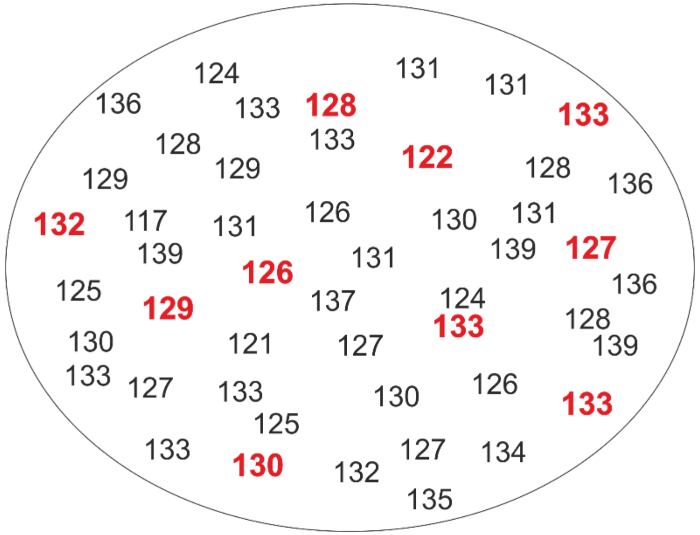
- D'une population *P* on peut extraire *N* échantillons
- Si l'on extrait plusieurs échantillons représentatifs de taille *n* fixée, les différences observées entre les résultats obtenus sont dues à des *fluctuations d'échantillonnage*. A partir d'un échantillon, on n'a donc pas de certitudes mais des *estimations de paramètres*.

Population



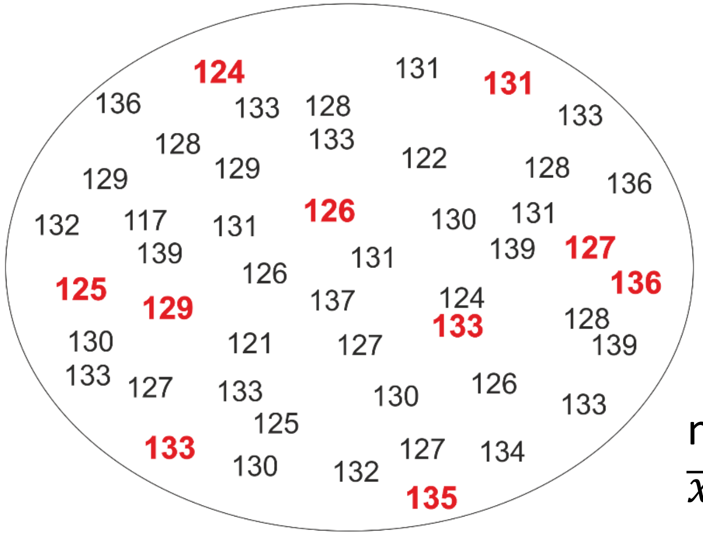
N=50
μ=131,1

1er échantillonnage



n=10
$\overline{x_1} = 128$

2ème échantillonnage



n=10
$\overline{x_2} = 124$

Echantillonnage – Estimation d’un paramètre

- D’une population *P* on peut extraire *N* échantillons

- Si l’on extrait plusieurs échantillons représentatifs de taille *n* fixée, les différences observées entre les résultats obtenus sont dues à des *fluctuations d’échantillonnage*. <u>A partir d’un échantillon, on n’a donc pas de certitudes mais des *estimations de paramètres*.</u>

- L'estimation d'un paramètre peut être faite
  - par un seul nombre: **estimation ponctuelle**
  - par 2 nombres entre lesquels le paramètre peut se trouver: **estimation par intervalle**

Echantillonnage – Estimation ponctuelle

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

<= moyenne

$$s_x{}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

<= variance

$$s_x = \sqrt{s_x{}^2}$$

<= écart type

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

<= écart type de la moyenne (« standard error »)

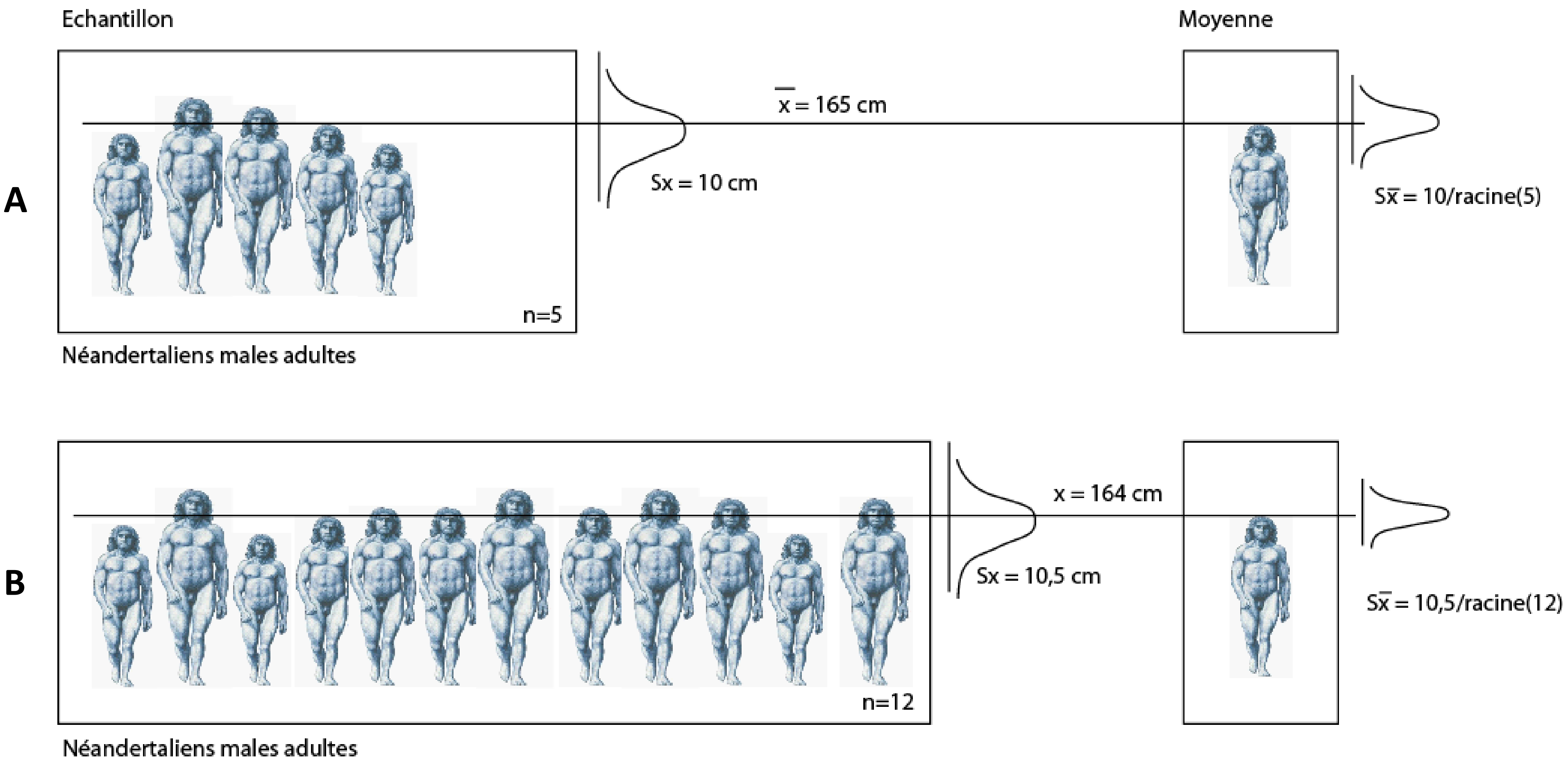
(i) Montre la qualité d'estimation ponctuelle (+petit = +précise)

(ii) Utilisé pour l'estimation d'intervalle de confiance

Echantillonnage – Estimation ponctuelle

écart type de la moyenne

$$\boxed{s_{\bar{x}} = \frac{s_x}{\sqrt{n}}}$$



$$s_{\bar{x}} = \frac{10}{\sqrt{5}} = 4{,}47$$

$$s_{\bar{x}} = 10{,}5/\sqrt{12} = 3{,}03$$

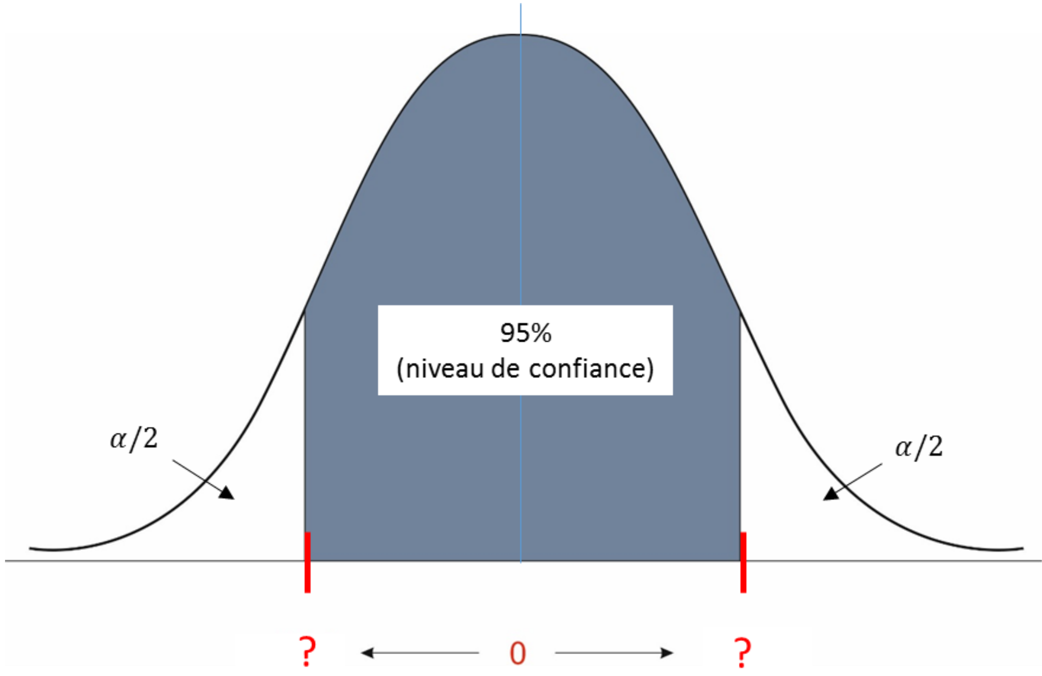
Pour améliorer la connaissance de la moyenne, il faut augmenter la taille de l'échantillon.

Echantillonnage – Estimation d'intervalle de confiance

- Un **intervalle de confiance** est un intervalle des valeurs utilisé pour estimer la vrai valeur d'un paramètre d'une population
  - De la moyenne
  - De la variance
  - D'un pourcentage

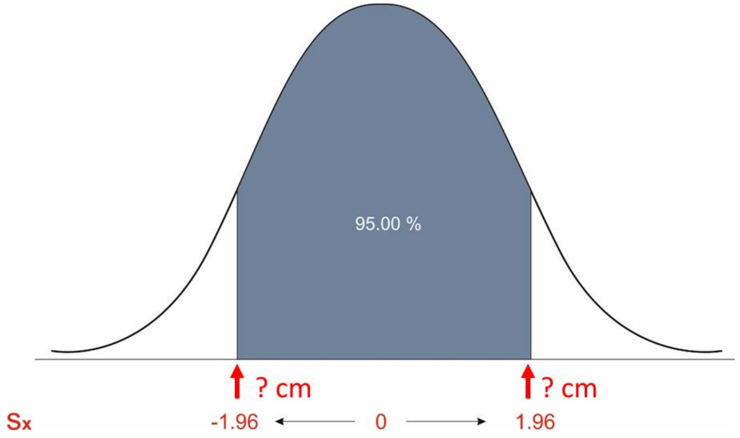
- Un **intervalle de confiance** est toujours associé a deux valeurs : **niveau de confiance** et **le risque**
  - **Niveau de confiance (*ndc*)**: c'est la certitude avec laquelle on veut estimer le paramètre
  - **Le risque ($\boldsymbol{\alpha}$):** est le risque que l'on décide de prendre pour estimer le paramètre ($\boldsymbol{\alpha = 1 - ndc}$)

  - c'est a nous de choisir ndc et α deux valeurs (souvent $ndc = 0.95$ et $\alpha = 0.05$)



Echantillonnage – Estimation d'intervalle de confiance de la moyenne (n>30)

- On peut estimer dans quel intervalle ce trouve la moyenne de la population
- **A) Cas des grands échantillons (variance connue)**
- Soit une population obéissant à une loi normale de moyenne μ et d'écart type σ

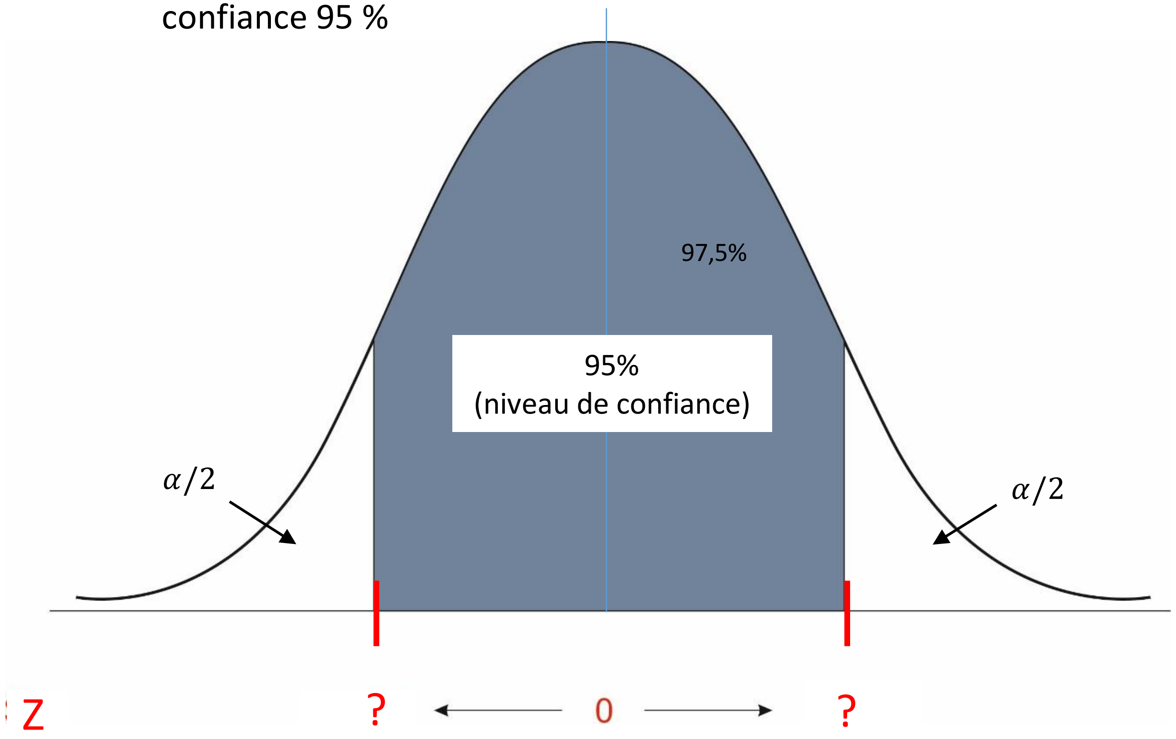


$$\boxed{\Pr(\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha}$$

- Comment?
  - Il nous faut connaitre/calculer: $\bar{x}$, σ et n
  - On défini niveau de confiance (notre certitude): $1 - \alpha$ (ex. 0.95)
  - On défini risque (α): $\alpha = 1 - ndc$ ($\alpha = 1 - 0.95 = 0.05$)
  - On trouve la valeur critique Z pour $\left(1 - \frac{\alpha}{2}\right)$

Echantillonnage – Estimation d'intervalle de

- Comment est-ce qu'on trouve la valeur critique Z pour niveau de confiance 95 %



!!! on cherche pas 0.95 mais $1 - \frac{\alpha}{2}$!!! => $\boxed{1 - \frac{0.05}{2} = 0.975}$

TABLE III — AIRES LIMITÉES PAR LA COURBE NORMALE CENTRÉE RÉDUITE

La table fournit les valeurs de $\phi(z)$ pour z positif. Lorsque z est négatif il faut calculer le complément à l'unité de la valeur lue dans la table. La première colonne indique la première décimale de z et la première rangée fournit la deuxième décimale.

Exemples : pour $z = 1{,}21$, $\phi(z) = 0{,}8869$ et pour $z = -1{,}21$, $\phi(z) = 0{,}1131$

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
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

z	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	0,9987	0,9990	0,9993	0,9995	0,9997	0,9998	0,9998	0,9999	0,9999	1,0000
4	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Echantillonnage – Estimation d’intervalle de confiance de la moyenne (n>30)

- Exemple:
- 45 hommes

$$\bar{x} = 164\,\mathrm{cm}$$

$$\sigma = 10\,\mathrm{cm}$$

$$\boxed{\Pr(\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha}$$

Echantillonnage – Estimation d'intervalle de confiance de la moyenne (n<30)

- D'un échantillon on peut estimer dans quel intervalle ce trouve la moyenne de la population
- **B) Cas des petits échantillons**
- Quand n<30 ou quand la variance est inconnue, on prend la loi de Student

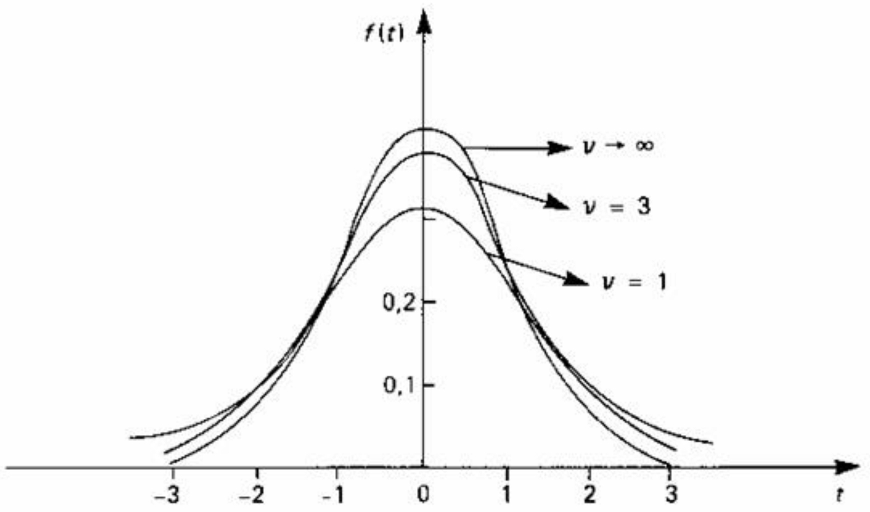
$$\boxed{\Pr(\bar{x} - t_{\alpha/2} \cdot \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \cdot \frac{s_x}{\sqrt{n}}) = 1 - \alpha}$$

- Comment?
  - Il nous faut calculer: $\bar{x}$, s_x et n
  - On défini niveau de confiance: $1 - \alpha$ (ex. 0.95)
  - On défini α: α ($\alpha = 1 - 0.95 = 0.05$)
  - Définir degrés de liberté v ($\mathrm{v} = n - 1$)
  - On trouve valeur critique t pour $(\alpha/2)$ ($t_{0.975}$)

Echantillonnage – Estimation d’intervalle de confiance de la moyenne (n<30)

- Pour chaque nombre des degrés de liberté existe une t-distribution diffèrent
- Plus il y a des individués/degrés de liberté, plus la courbe a tendance de devenir la courbe normal

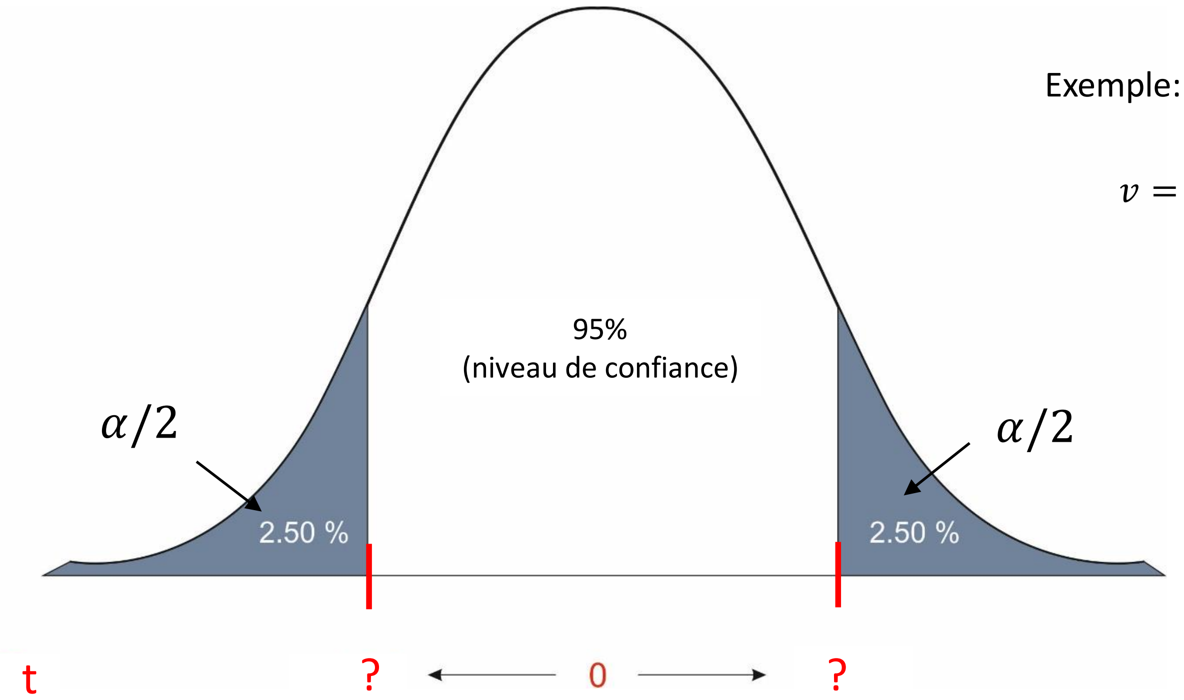
Famille de courbes de densité de probabilité obéissant à la loi de Student pour différents nombres de degrés de liberté



Pour l’estimation d’idcdlm on peut toujours utiliser la loi de Student puisque t tend vers la loi normale quand *n* est grand…

Echantillonnage – Estimation d'intervalle de confian

- Comment est-ce qu'on trouve la valeur critique t pour niveau de confiance 95 %
  - Pour chaque nombre des degrés de liberté existe une t-distribution diffèrent
  - Plus il y a des individués/degrés de liberté, plus la courbe a tendance de devenir la courbe normal



Exemple: $n = 22;\ \alpha = 0.05$

$$\nu = 21; \frac{\alpha}{2} = 0.025$$

Table de Student t

ν	α					
	0.100	0.050	0.025	0.010	0.005	0.001
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.611
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
100	1.290	1.660	1.984	2.365	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.090

La table de Student donne les valeurs $t_{(\alpha,\nu)}$ telles que

$$P\{T > t_{(\alpha,\nu)}\} = \alpha$$

Echantillonnage – Estimation d'intervalle de confiance de la moyenne (n<30)

- Exemple:
- 6 hommes

$$\bar{x} = 165\,\mathrm{cm}$$

$$s_x = 11\,\mathrm{cm}$$

$$\boxed{\Pr(\bar{x} - t_{\alpha/2} \cdot \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \cdot \frac{s_x}{\sqrt{n}}) = 1 - \alpha}$$

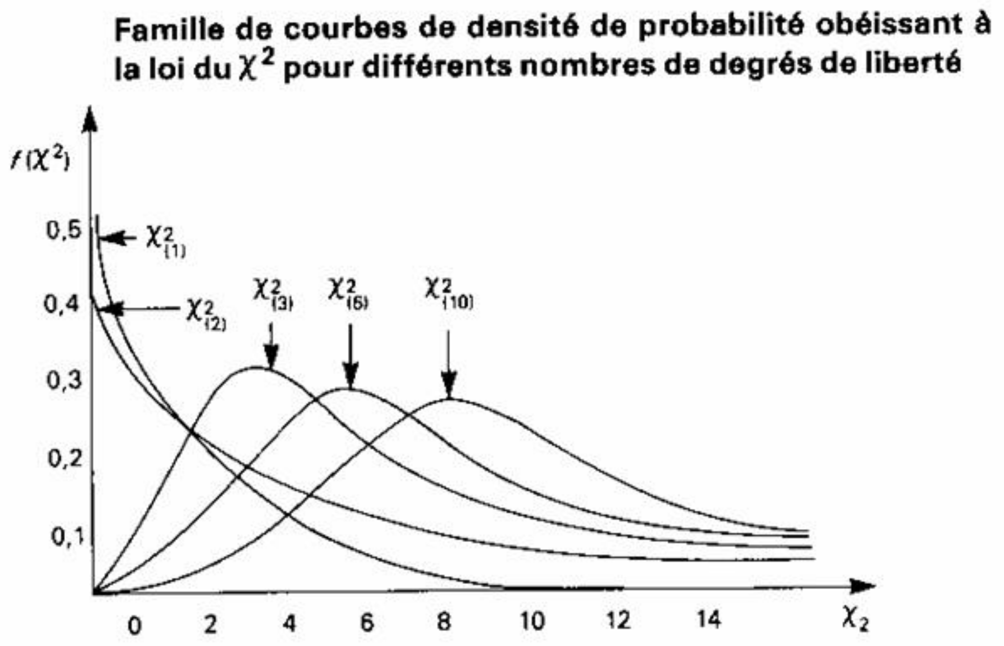
Echantillonnage – Estimation d’intervalle de confiance de la variance

- on peut aussi estimer dans quel intervalle ce trouve la variance
- Soit une population obéissant à une loi normale de moyenne μ (inconnue) et d’écart type σ (inconnue)

$$\boxed{\Pr\left(\frac{(n-1)\cdot s_x^2}{\chi^2_{(1-\alpha/2)}} < \sigma^2 < \frac{(n-1)\cdot s_x^2}{\chi^2_{\alpha/2}}\right) = 1-\alpha}$$

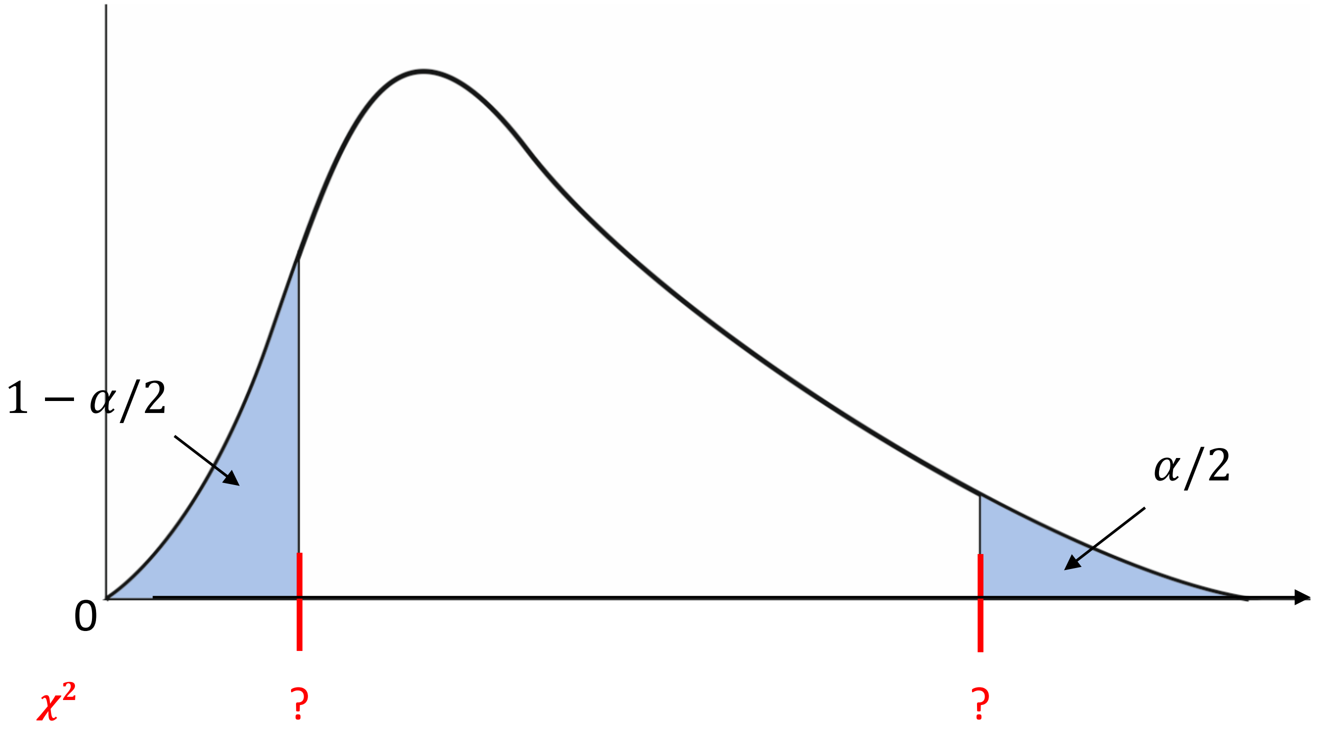
- Comment?
  - Il nous faut connaitre: $n, {s_x}^2$
  - On défini niveau de confiance: $1-\alpha$ (ex. 0.95)
  - On défini α: α ($\alpha = 1 - 0.95 = 0.05$)
  - Définir degrés de liberté v ($\mathrm{v} = n - 1$)
  - On trouve la valeur critique χ^2 pour $(1-\alpha/2)$ ($\chi^2{}_{0.975}$)
  - On trouve la valeur critique χ^2 pour $(\alpha/2)$ ($\chi^2{}_{0.025}$)

Echantillonnage – Estimation d'intervalle de confiance de la variance



Echantillonnage – Estimation d'intervalle de co

- Comment est-ce qu'on trouve les deux valeur critiques χ^2 pour niveau de confiance 95 % ?
- De la même façon qu'avec t-distribution…



v	$\alpha=0{,}990$	$\alpha=0{,}975$	$\alpha=0{,}950$	$\alpha=0{,}900$	$\alpha=0{,}100$	$\alpha=0{,}050$	$\alpha=0{,}025$	$\alpha=0{,}010$	$\alpha=0{,}001$
1	0,0002	0,0010	0,0039	0,0158	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,12	0,22	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	20,52
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	22,46
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,47	24,32
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	26,13
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	3,82	4,57	5,58	17,27	19,67	21,92	24,72	31,26
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	36,12
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	37,70
16	5,81	6,91	7,96	9,31	23,54	26,30	28,84	32,00	39,25
17	6,41	7,56	8,67	10,08	24,77	27,59	30,19	33,41	40,79
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,80	42,31
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	43,82
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	45,32
21	8,90	10,28	11,59	13,24	29,61	32,67	35,48	38,93	46,80
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	48,27
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	49,73
24	10,86	12,40	13,85	15,66	33,20	36,41	39,37	42,98	51,18
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	52,62
26	12,20	13,84	15,38	17,29	35,56	38,88	41,92	45,64	54,05
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	55,48
28	13,57	15,31	16,93	18,94	37,92	41,34	44,46	48,28	56,89
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	58,30
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	59,70

Echantillonnage – Estimation d'intervalle de confiance de l'écart type

- Presque pareille que pour la variance – pourquoi => $s_x = \sqrt{{s_x}^2}$
- Soit une population obéissant à une loi normale de moyenne μ et d'écart type σ

Écart type

$$\Pr\left(\sqrt{\frac{(n-1)\cdot s_x^2}{\chi^2_{(1-\alpha/2)}}} < \sigma < \sqrt{\frac{(n-1)\cdot s_x^2}{\chi^2_{\alpha/2}}}\right) = 1-\alpha$$

Variance

$$\Pr\left(\frac{(n-1)\cdot s_x^2}{\chi^2_{(1-\alpha/2)}} < \sigma^2 < \frac{(n-1)\cdot s_x^2}{\chi^2_{\alpha/2}}\right) = 1-\alpha$$

- Comment?
  - Il nous faut calculer: $n,\ {s_x}^2$
  - On défini intervalle de confiance: $1-\alpha$ (ex. 0.95)
  - On défini α: α ($\alpha = 1 - 0.95 = 0.05$)
  - On trouve la valeur critique χ^2 pour $(1-\alpha/2)$ ($\chi^2_{0.975}$)
  - On trouve la valeur critique χ^2 pour $(\alpha/2)$ ($\chi^2_{0.025}$)
  - Définir degrés de liberté v ($v = n - 1$)

Echantillonnage – Estimation ponctuelle d'un pourcentage

- La population est formée d'individus ayant ou non un caractère A. Soit p la probabilité pour qu'un individu pris au hasard dans la population présente le caractère A.

- $p = \frac{a}{n}$
- ${s_p}^2 = \frac{p(1-p)}{n-1}$

exemple:
100 ballons dans le sac: 80 noirs, 20 blanc
p est la probabilité qu'on prends le noir

$$p_{noir} = \frac{80}{100} = 0.8$$

$${s_p}^2 = \frac{0.8(1-0.8)}{100-1}$$

Probabilité qu'on prends noir

Probabilité qu'on prends blanc

Echantillonnage – Estimation d’intervalle de confiance d’un pourcentage

- Grands échantillons (n>30)
- La variable fréquence obéit à une loi normale centrée réduite

$$\Pr\left(p - Z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} < P < p + Z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$$

- Comment?
  - Il nous faut connaitre: p et n
  - On défini niveau de confiance: $1-\alpha$ (ex. 0.95)
  - On défini risque α: α ($\alpha = 1 - 0.95 = 0.05$)
  - On trouve la valeur critique Z pour $(1-\alpha/2)$ ($Z_{0.975}$)

TABLE III — AIRES LIMITÉES PAR LA COURBE NORMALE CENTRÉE RÉDUITE

La table fournit les valeurs de $\phi(z)$ pour z positif. Lorsque z est négatif il faut calculer le complément à l'unité de la valeur lue dans la table. La première colonne indique la première décimale de z et la première rangée fournit la deuxième décimale.

Exemples : pour z = 1,21, $\phi(z)$ = 0,8869 et pour z = − 1,21, $\phi(z)$ = 0,1131

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
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

z	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	0,9987	0,9990	0,9993	0,9995	0,9997	0,9998	0,9998	0,9999	0,9999	1,0000
4	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000