DUT GEA semestre 3

Intervalles de confiance et tests statistiques

Intervalles de confiance & tests statistiques

- **Echantillonnage, rappels**
- **Intervalle de confiance**
  - Moyenne
  - Variance et écart type
  - Médiane
  - Pourcentage
- **Tests usuels**
  - Principe (rappels)
  - Théorie de la statistique de décision (rappels)
  - 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

Extraction de *n* échantillons d'une population P

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*.

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**

Estimation ponctuelle d'une moyenne

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

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

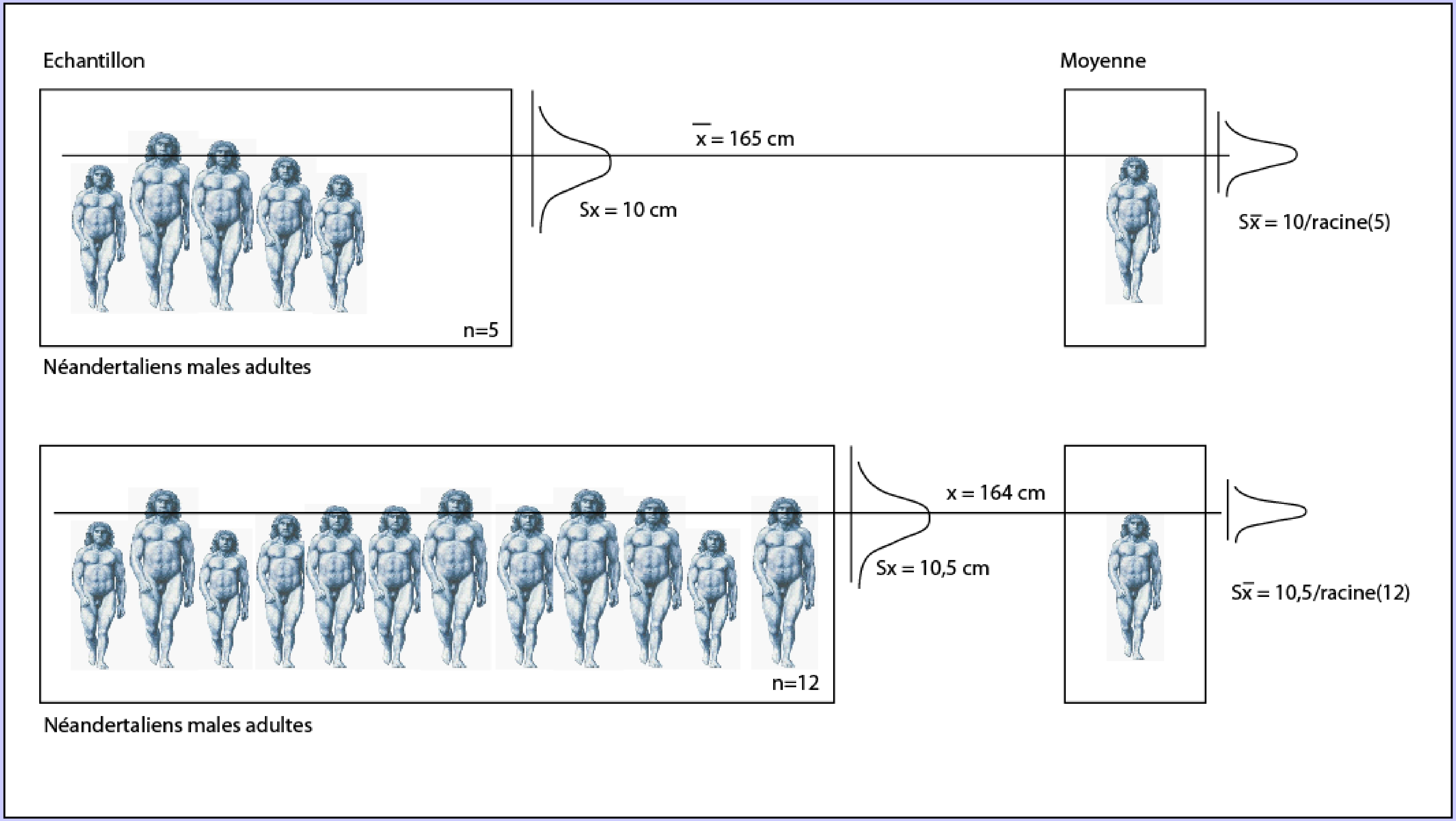
x barre

Estimateur sans biais

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

Ecart type de la moyenne

Echantillonnage – Estimation d'un paramètre



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

Intervalle de confiance de la moyenne

Cas des grands échantillons (variance connue):

Soit une population obéissant à une **loi normale** de moyenne μ et d'écart type σ.

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

Exemple:

45 hommes de Neandertal males adultes

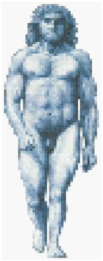
$$\bar{x} = 164 \text{ cm}$$

$$\sigma = 10 \text{ cm}$$

$$\mu \in \left[164 - 1.96 \cdot \frac{10}{\sqrt{45}} ; 164 + 1.96 \cdot \frac{10}{\sqrt{45}} \right]$$

$$\mu \in [161 ; 166.9]$$ à 95% de confiance

$$\mu = 164 \pm 2.9$$



Echantillonnage – Estimation d'un paramètre

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

Intervalle de confiance de la moyenne

Cas des petits échantillons:

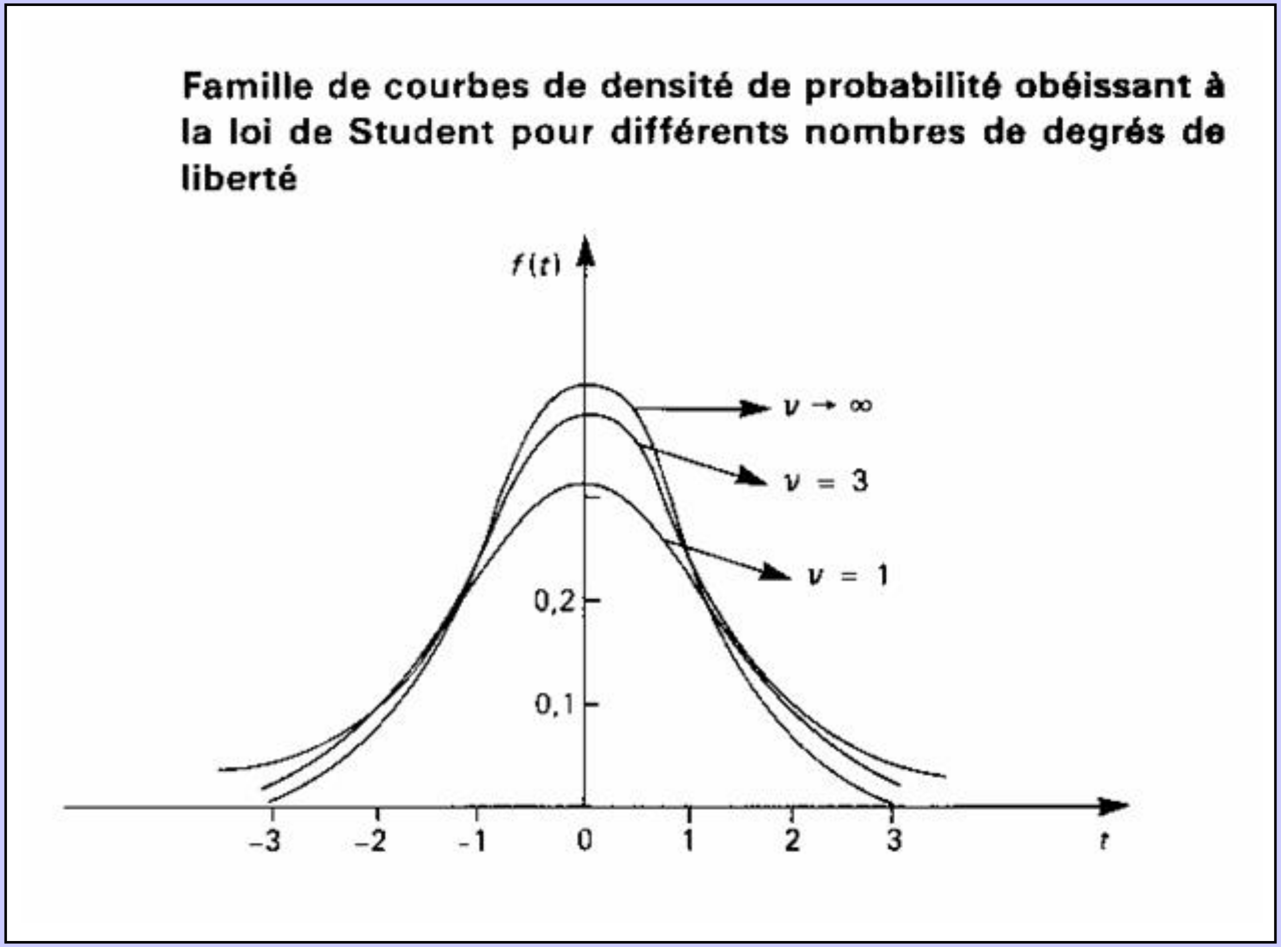
Quand $n<30$ ou quand la variance est inconnue, on prend la loi de Student.

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

Pour $\nu = n-1$ degrés de liberté

Finalement on peut toujours utiliser la loi de Student puisque t tend vers la loi normale quand n est grand…

La loi de Student: t(ν)



ν degrés de liberté

Converge vers **la loi Normale** quand ν augment.

La probabilité d'obtenir une valeur de t à l'extérieur de l'intervalle ($-t_{\alpha/2}$ et $t_{\alpha/2}$) -> TABLES.

$$P(|t| > t_{\alpha/2}) = \alpha$$

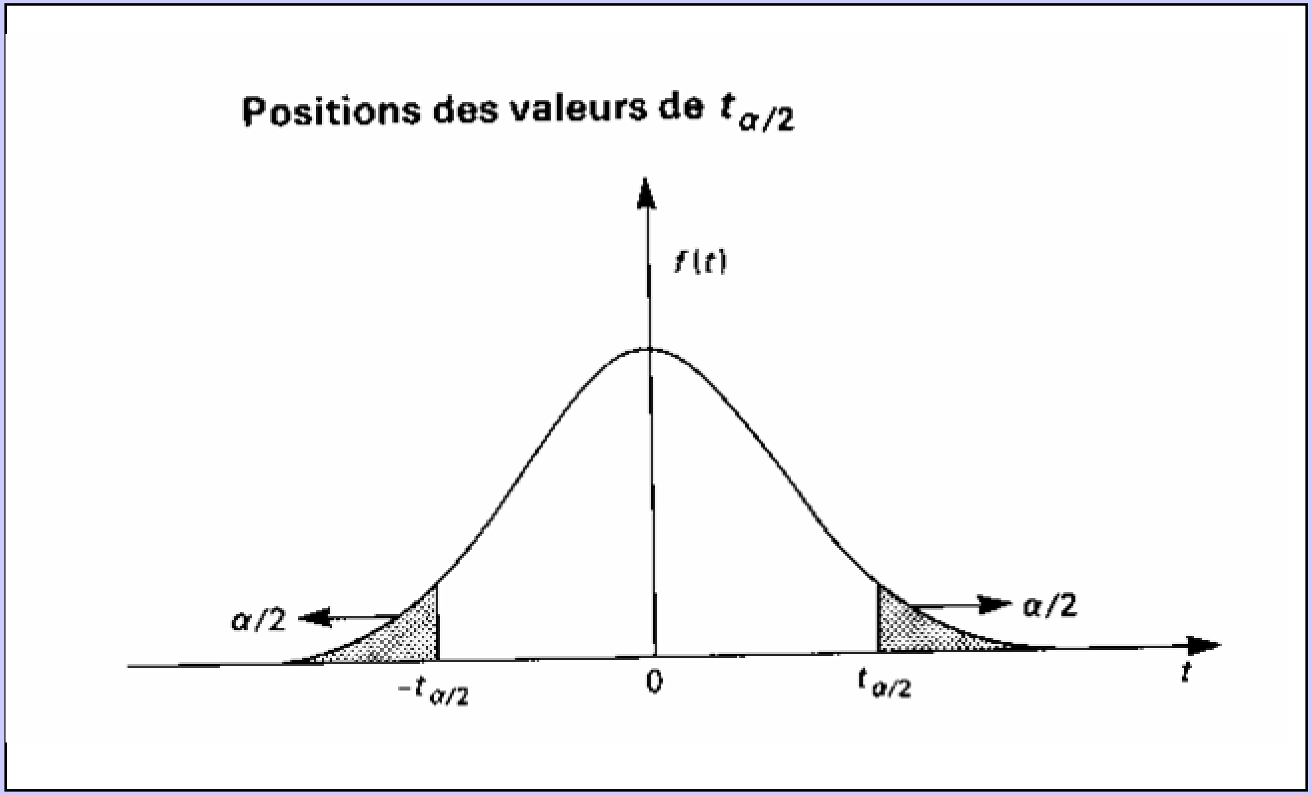


Table de Student t

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

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

Exemple:

6 hommes de Neandertal males adultes

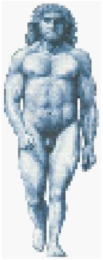
$$\bar{x} = 165\,\text{cm}$$

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

$$\mu \in \left[165 - 2.57 \cdot \frac{11}{\sqrt{6}} ; 165 + 2.57 \cdot \frac{11}{\sqrt{6}}\right]$$

$$\mu \in [153;177]$$ à 95% de confiance

$$\mu = 165 \pm 12$$



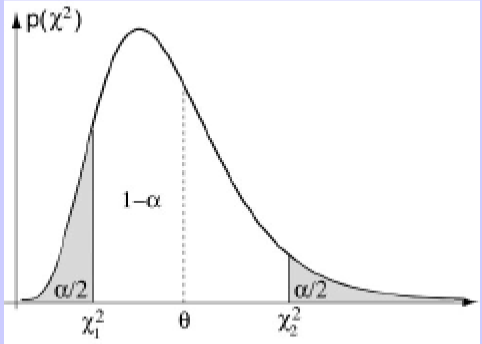
Finalement on peut toujours utiliser la loi de Student puisque t tend vers la loi normale quand n est grand…

Intervalle de confiance de la variance

Soit une population obéissant à une **loi normale** de moyenne μ (inconnue) et d'écart type σ (inconnu).

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

Pour ν = n-1 degrés de liberté



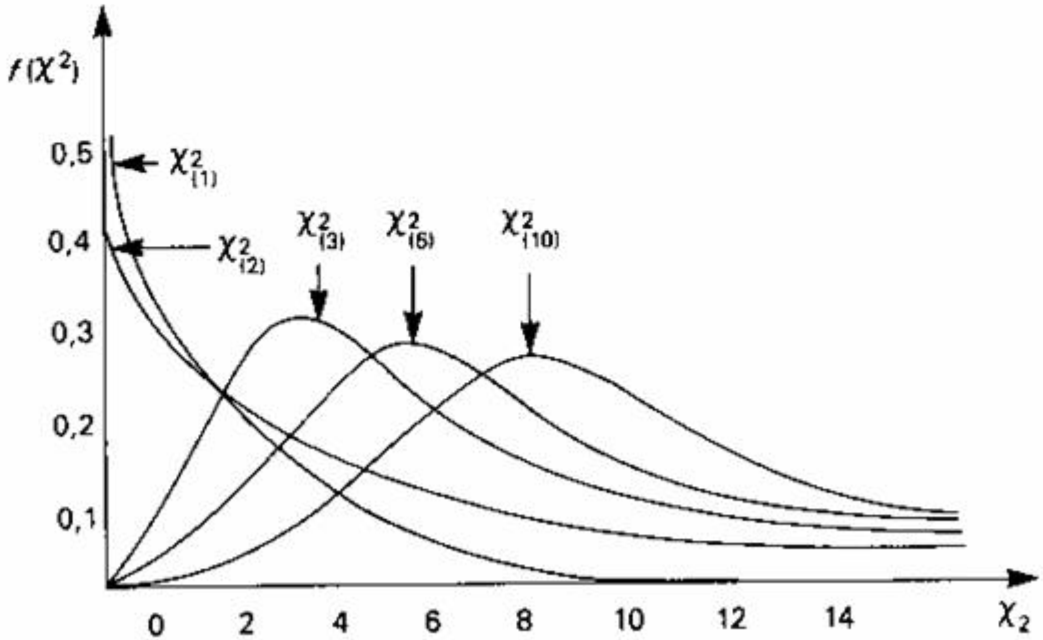
La loi du Khi carré: χ^2

Si Z_1, Z_2, Z_n sont des variables aléatoires normales centrées réduites et indépendantes entres elles, la somme des carrées de ces varaibles aléatoires obéit à la **loi du χ^2** à ν degrés de libertés

$$\boxed{\chi^2 = Z_1^2 + Z_2^2 + ... + Z_\nu^2}$$

La loi du Khi carré: χ^2

Famille de courbes de densité de probabilité obéissant à la loi du χ^2 pour différents nombres de degrés de liberté

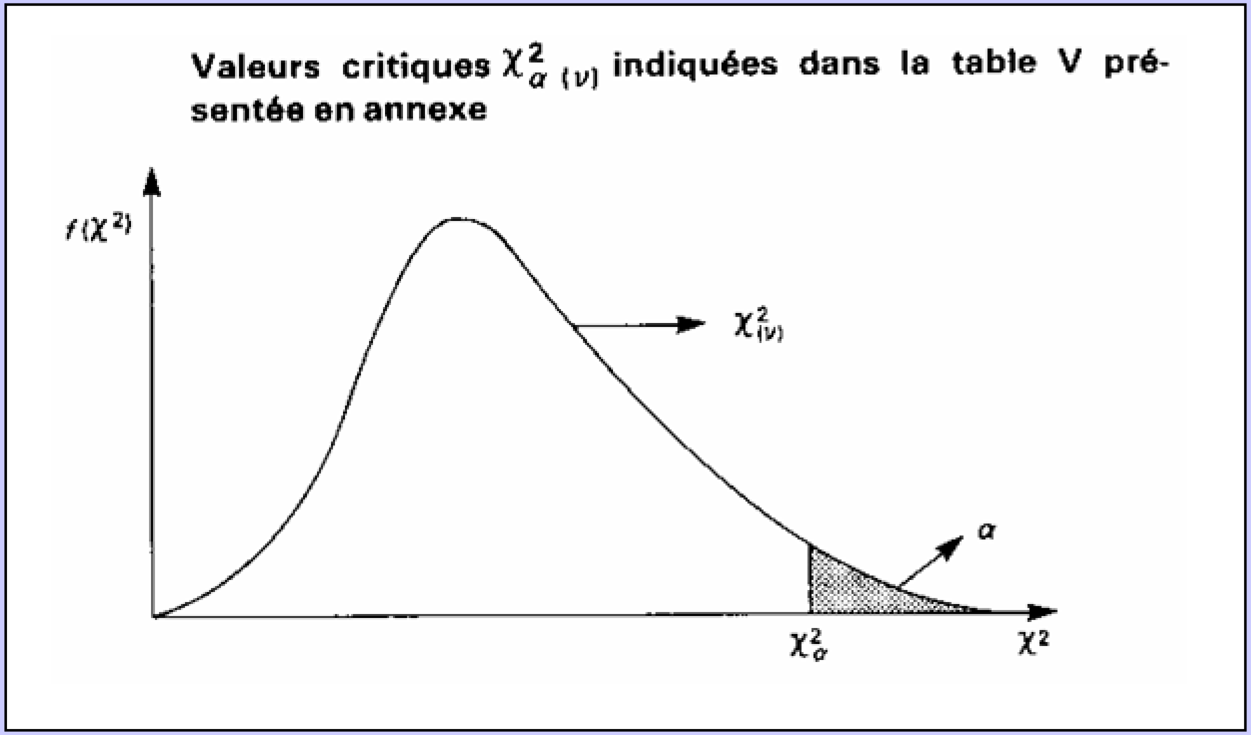


On remarquera que pour $\nu \geqslant 3$, la courbe passe par 0, atteint un maximum de densité de probabilité puis décroît jusqu'à l'infini.

La loi du Khi carré: χ^2

En fait, les calculs sont fastidueux -> TABLES

$$\alpha = P(\chi^2 > \chi^2_\alpha)$$



La loi du Khi carré: χ^2

Table du chi-carré χ^2

ν	α						
	0.900	0.700	0.500	0.300	0.100	0.050	0.010
1	0.016	0.15	0.46	1.07	2.71	3.84	6.63
2	0.21	0.71	1.39	2.41	4.60	5.99	9.21
3	0.58	1.42	2.37	3.67	6.25	7.81	11.34
4	1.06	2.19	3.36	4.88	7.78	9.49	13.28
5	1.61	3.00	4.35	6.06	9.24	11.07	15.09
6	2.20	3.83	5.35	7.23	10.65	12.59	16.81
7	2.83	4.67	6.35	8.38	12.02	14.07	18.48
8	3.49	5.53	7.34	9.52	13.36	15.51	20.09
9	4.17	6.39	8.34	10.66	14.68	16.92	21.67
10	4.87	7.27	9.34	11.78	15.99	18.31	23.21
11	5.58	8.15	10.34	12.90	17.28	19.68	24.73
12	6.30	9.03	11.34	14.01	18.55	21.03	26.22
13	7.04	9.93	12.34	15.12	19.81	22.36	27.69
14	7.79	10.82	13.34	16.22	21.06	23.69	29.14
15	8.55	11.72	14.34	17.32	22.31	25.00	30.58
16	9.31	12.62	15.34	18.42	23.54	26.30	32.00
17	10.09	13.53	16.34	19.51	24.77	27.59	33.41
18	10.87	14.44	17.34	20.60	25.99	28.87	34.81
19	11.65	15.35	18.34	21.69	27.20	30.14	36.19
20	12.44	16.27	19.34	22.78	28.41	31.41	37.57
25	16.47	20.87	24.34	28.17	34.38	37.65	44.31
30	20.60	25.51	29.34	33.53	40.26	43.77	50.89
35	24.80	30.18	34.34	38.86	46.06	49.80	57.34
45	33.35	39.58	44.34	49.45	57.50	61.66	69.96
55	42.06	49.06	54.33	59.98	68.80	73.31	82.29
65	50.88	58.57	64.33	70.46	79.98	84.82	94.42
75	59.79	68.13	74.33	80.91	91.06	96.22	106.39
85	68.77	77.71	84.33	91.32	102.08	107.52	118.24
95	77.82	87.32	94.33	101.72	113.04	118.75	129.97
120	100.62	111.42	119.33	127.61	140.23	146.57	158.95

La table du chi-carré donne les valeurs $\chi^2_{(\alpha,\nu)}$ telles que

$$P\{\chi^2 > \chi^2_{(\alpha,\nu)}\} = \alpha$$

Intervalle de confiance de l'écart type (idem)

Soit une population obéissant à une **loi normale** de moyenne μ et d'écart type σ.

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

Pour ν = n-1 degrés de liberté

Intervalle de confiance de la médiane

Si un échantillon est extrait d'une **population approximativement normale**, et si son effectif est relativement grand (n>60), la distribution d'échantillonnage de la médiane s'approche de la loi normale.

$$\boxed{s_{Me} = s_x \sqrt{\frac{\pi}{2n}}}$$

$$\boxed{\Pr(Me - Z_{\alpha/2} \cdot s_x \cdot \sqrt{\frac{\pi}{2n}} < Mediane < Me + Z_{\alpha/2} \cdot s_x \cdot \sqrt{\frac{\pi}{2n}}) = 1 - \alpha}$$

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.

$$p = a / n$$

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

Quand on dispose d'un seul échantillon de taille n, la meilleure estimation ponctuelle de P est donc la fréquence p observée sur l'échantillon.

Intervalle de confiance d'un pourcentage

Grands échantillons (n>30), p ni voisin de 0, ni voisin de 1, (np>5, n(1-p)>5)

La variable fréquence obéit à une loi normale centrée réduite

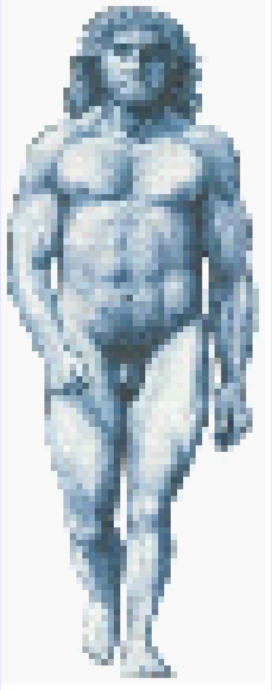
$$\Pr(p - Z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} < P < p + Z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}) = 1 - \alpha$$

Un problème très fréquent!
Un quotidien publie tous les mois la cote du chef du gouvernement à partir d'un sondage réalisé sur un échantillon représentatif de 1000 personnes. En janvier, la cote publiée était de 38% d'opinions favorables, en février de 36%. Un journaliste commente alors ces valeurs par "Le chef du gouvernement perd 2 points !!"

En fait: On construit un intervalle de confiance autour des proportions. Avec un seuil de 95%, on obtient respectivement [35;41] et [33;39] pour les valeurs 38% et 36%. Les deux intervalles ayant une intersection non vide, on ne peut pas conclure qu'il y ait eu baisse ou augmentation de la cote du chef de gouvernement.

Tests statistiques

Quel est le problème…?



On sait qu'un homme de Neandertal mesure en moyenne 165 cm.

Sur un site on trouve 16 hommes avec une moyenne de 167 et un écart type de 8 cm (e.t. échantillon).

Comparaison de la moyenne avec la valeur théorique de 165 cm

Possibilités:
Moyenne très élevée: Nous pourrons être amenés à croire que ces hommes ont des tailles différentes de 165 cm

Moyenne faiblement plus élevée: on ne pourra pas conclure si c'est significativement supérieur à la norme ou si c'est l'effet du hasard.

Question: à partir de quelle limite pouvons nous raisonnablement conclure à une différence?

H_0: μ=165 (il n'y pas de différence)
H_1: $\mu\neq$165

Calcul de $$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{8}{\sqrt{16}} = 2$$

Sur la table la probabilité pour que la moyenne d'échantillonnage soit différente celle de la population de plus 2,131 de écart-type est de 5%.

Les deux risques d’erreur dans un test.

Erreur de 2[nde] espèce (compliquée)

$1-\alpha$

Décision	H_0 est vraie	H_1 est vraie
H_0 acceptée	Bonne décision	Erreur β
H_0 rejetée	Erreur α	Bonne décision

$1-\beta$

Erreur de 1[ere] espèce

A priori on ne sait pas à quel type d’erreur on sera confronté:
Le résultat de l’échantillon a révélé 167 cm probablement par pur hasard. On conclue que la moyenne pourrait être 165 cm alors qu’en fait elle est mesurée à 167 cm.

H_0 : hypothèse nulle ou principale
Ex: Les haches de type A présentent les mêmes teneurs en Sn que les haches de type B.

H_1 : hypothèse alternative ou contraire …

Soumission à une épreuve de vérité!

Conclusion : différence attribuable aux fluctuations d’échantillonnage???

Niveau de signification : un peu arbitraire…
significatif : 0.05
hautement significatif : 0.01
très hautement significatif : 0.001.

Test bilatéral / unilatéral :
bilatéral : différence sans se préoccuper du sens.
Unilatéral : > ou <. Zone de rejet d'un seul coté de la distribution de probabilité de référence.

Echantillons indépendants ou appariés:
Indépendants : aucune influence du 1er ech sur le 2nd.
Appariés : prélèvements par paires. Ex : fumeurs H + F.

Comparaison des moyennes de 2 grands échantillons indépendants (n_1 et n_2 >30):

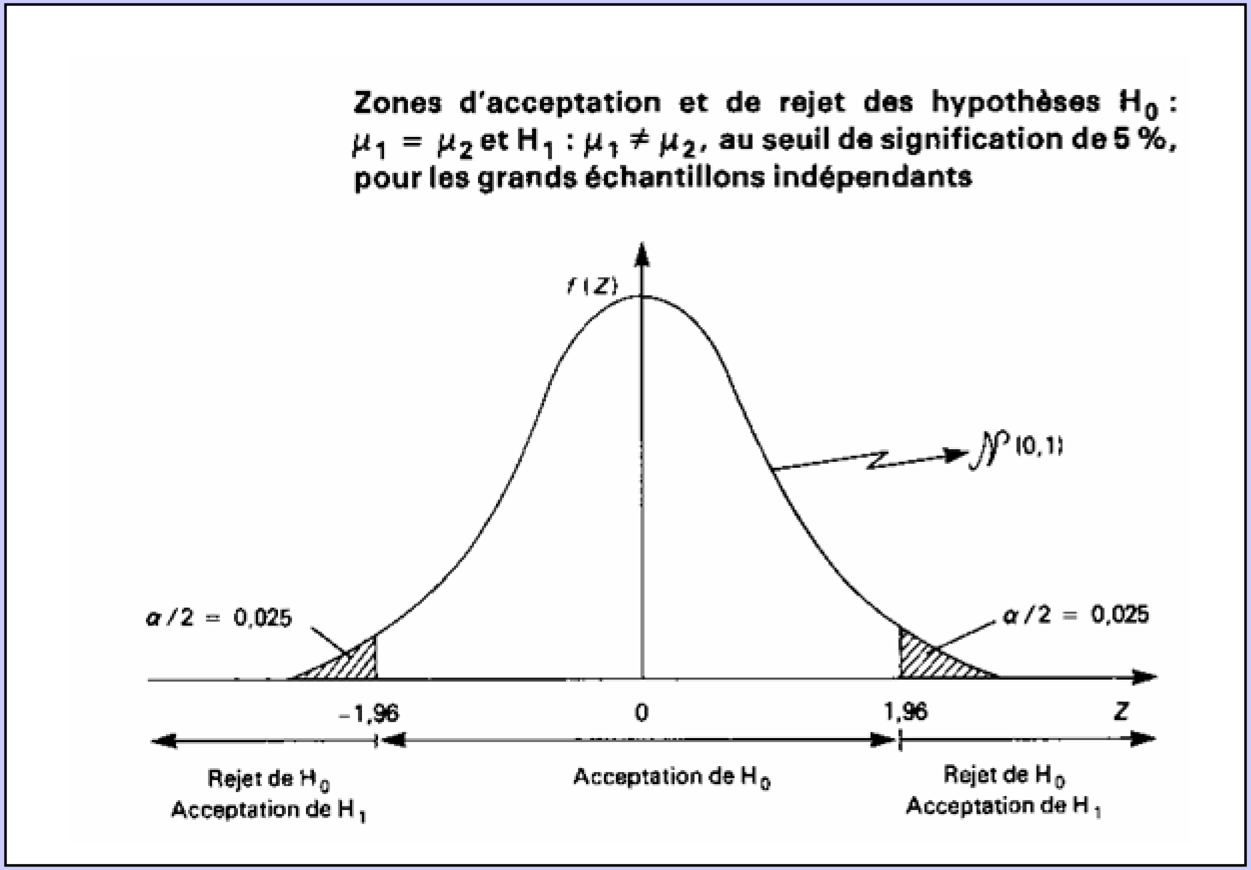
Deux échantillons qui suivent des **lois normales**: μ_1, σ^2_1; μ_2, σ^2_2

$H_0 : \mu_1 = \mu_2$

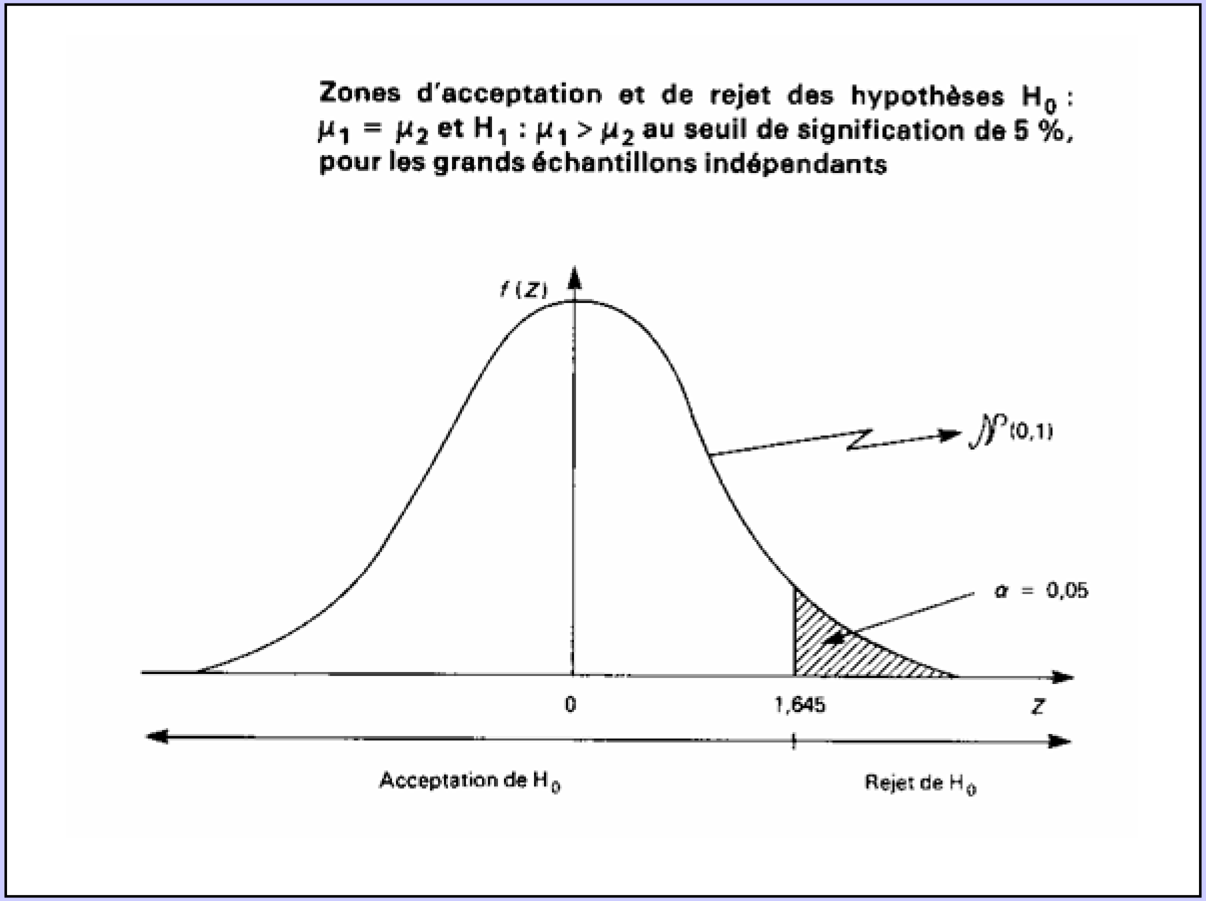
$$Z_c = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s^2_{x_1}}{n_1} + \frac{s^2_{x_2}}{n_2}}}$$

Si H_0 est vraie, Z_c suit une loi normale N(0,1)

$H_1 : \mu_1 \neq \mu_2$ bilatéral



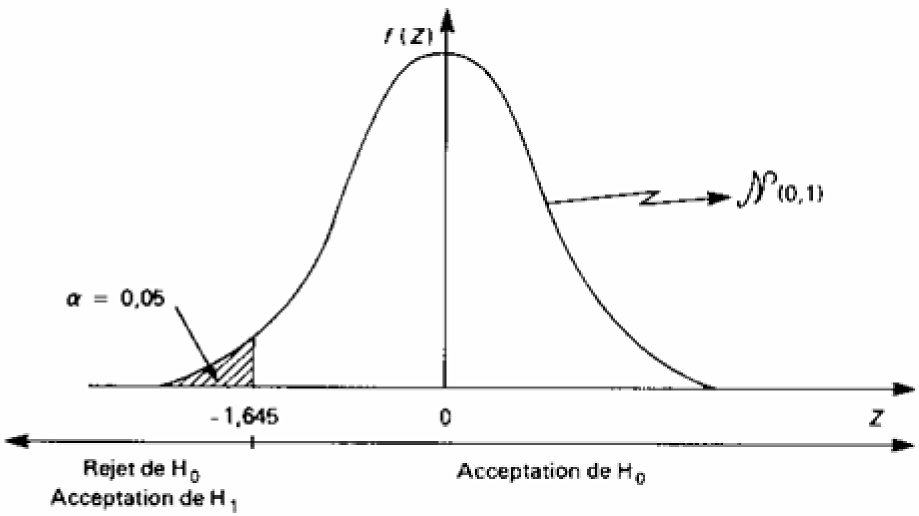
$H_1 : \mu_1 > \mu_2$ unilatéral



Comparaison de deux moyennes expérimentales–grands échantillons -

$H_1 : \mu_1 < \mu_2$ unilatéral

Zones d'acceptation et de rejet des hypothèses $H_0 : \mu_1 = \mu_2$ et $H_1 : \mu_1 < \mu_2$, au seuil de signification de 5 %, pour les grands échantillons indépendants



Comparaison de deux moyennes expérimentales–grands échantillons -

Pour résumer:

H_0	H_1	Rejet de H_0 si	$\alpha = 0.05$	$\alpha = 0.01$
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\lvert Z_c \rvert \geq \lvert z_{\alpha/2} \rvert$	$\lvert z_{\alpha/2} \rvert = 1.96$	$\lvert z_{\alpha/2} \rvert = 2.57$
	$\mu_1 > \mu_2$	$Z_c \geq z_\alpha$	$z_\alpha = 1.64$	$z_\alpha = 2.33$
	$\mu_1 < \mu_2$	$Z_c \leq z_\alpha$	$z_\alpha = 1.64$	$z_\alpha = 2.33$

Maintenant un exemple...

Taille des silex sur deux sites

$n_1 = 50$	$n_2 = 67$
$\bar{x}_1 = 158{,}86mm$	$\bar{x}_2 = 134{,}46mm$
$s_{x_1}^2 = 37{,}18mm^2$	$s_{x_2}^2 = 25{,}92mm^2$
$s_{x_1} = 6{,}09mm$	$s_{x_2} = 5{,}09mm$

Les moyennes de ces deux échantillons prélevés indépendamment l'un de l'autre diffèrent-elles d'une façon hautement significative?

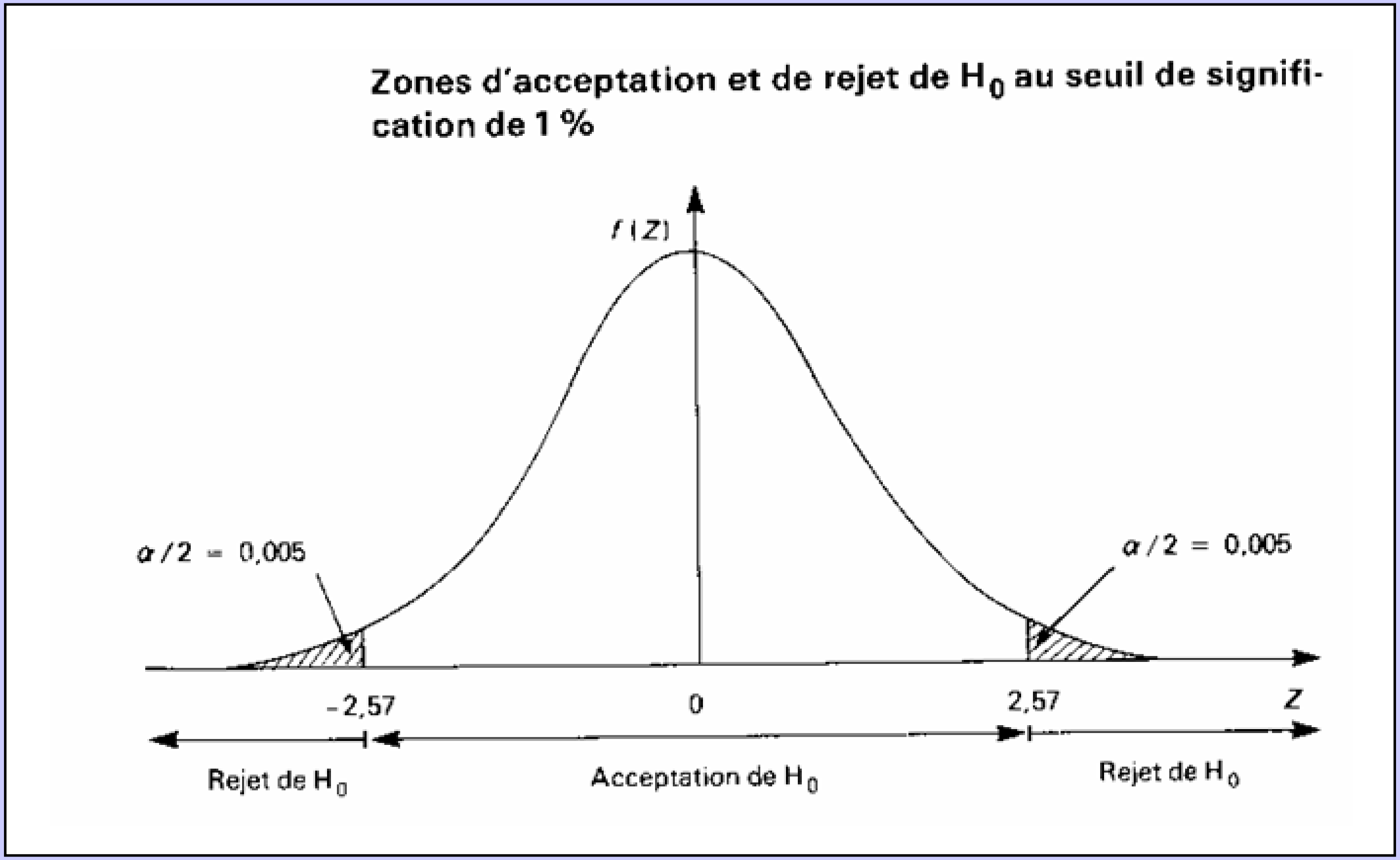
Comparaison de deux moyennes expérimentales–grands échantillons -

n_1 et n_2 grands -> test sur la **loi normale**

$H_0 : \mu_a = \mu_b$
$H_1 : \mu_a \neq \mu_b$ (bilatéral)

$$Z_c = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_{x1}^2}{n_1} + \frac{s_{x2}^2}{n_2}}} \qquad Z_c = \frac{158.86 - 134.66}{\sqrt{\frac{37.18}{50} + \frac{25.92}{67}}} = 22.9$$

$\alpha = 0.01$, $Z_{\alpha/2} = 2.57$



H_0 rejetée au seuil de signification de 1%

Comparaison d'une moyenne empirique à une moyenne théorique

Même principe que précédemment (quand n est grand):

H_0: $\mu=\mu_0$

$$Z_c = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}$$

que l'on teste sur la **loi normale N(0,1)**

Cas des petits échantillons: Test t

Deux **populations normales** μ_1 et μ_2 de **même variance (au moins approximativement)** σ^2. Si **n_1 et n_2 sont petits**, s^2_{x1} et s^2_{x2} sont des estimateurs peu précis de σ^2.

Dans ce cas, la variable différence centrée réduite n'obéit plus à une loi normale mais à une **loi de Student** à $\nu=n_1+n_2-2$ degrés de liberté.

La variance de la distribution des différences de moyennes est estimées par s^2_D

$$s_D^2 = s_{pd}^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

avec

$$s_{pd}^2 = \frac{(n_1 - 1)s_{x_1}^2 + (n_2 - 1)s_{x_2}^2}{n_1 + n_2 - 2}$$

Ce qui donne…

$H_0 : \mu_a = \mu_b$

$$t_c = \frac{\bar{x}_1 - \bar{x}_2}{s_D}$$

Avec $\nu = n_1 + n_2 - 2$

Si les **variances s'avèrent inégales** alors test t modifié.

$$t_{cm} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_{x_1}^2}{n_1} + \frac{s_{x_2}^2}{n_2}\right)}}$$

avec

$$\nu = \frac{\left[\frac{s_{x_1}^2}{n_1} + \frac{s_{x_2}^2}{n_2}\right]^2}{\frac{\left(\frac{s_{x_1}^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_{x_2}^2}{n_2}\right)^2}{n_2 - 1}}$$

Comparaison d'une moyenne empirique à une moyenne théorique

Même principe que précédemment. Suivant si n est petit ou grand, on calcule les variables auxiliaires suivantes:

H_0: $\mu=\mu_0$

$$t_c = \frac{\bar{x}-\mu_0}{\frac{s_x}{\sqrt{n}}} \qquad Z_c = \frac{\bar{x}-\mu_0}{\frac{s_x}{\sqrt{n}}}$$

que l'on teste sur la **loi de Student** ou **loi normale N(0,1)**

Comparaison de moyennes de deux échantillons appariés

Fondée sur les différences de chaque paire d'éléments

$$d_i = x_{i_1} - x_{i_2}$$

On imagine que la différence obéit à une **loi normale**, mais en général on utilise une **loi de Student** à n-1 degrés de liberté:

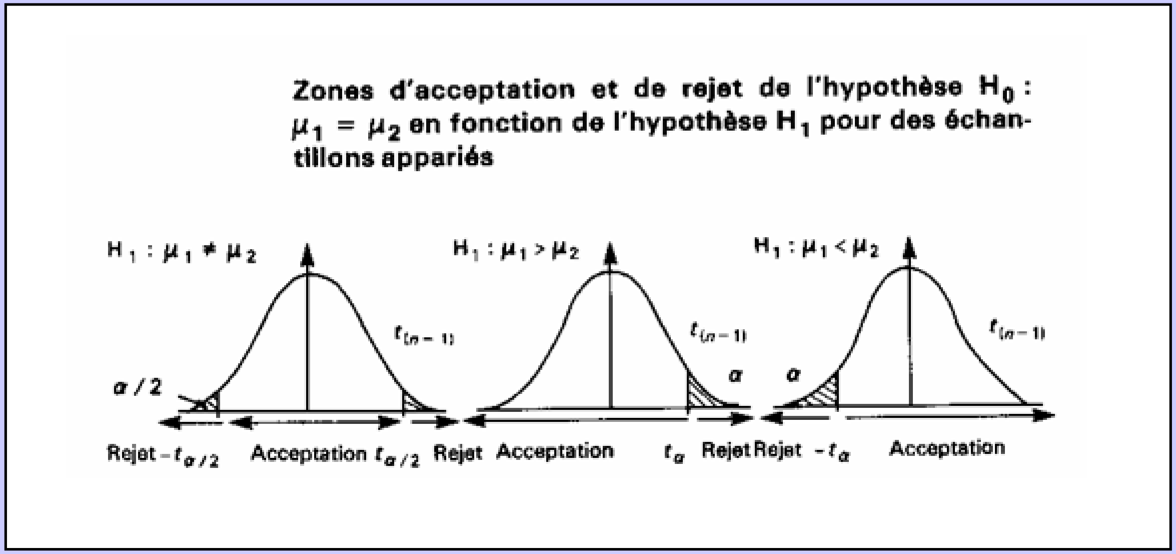
$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} \text{ et } s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$$

Comparaison de moyennes de deux échantillons appariés

$H_0 : \mu_1 = \mu_2$ ou $\mu_d = 0$

$$\boxed{t_c = \frac{\bar{d}}{s_{\bar{d}}}}$$

$H_1: \mu_1 \neq \mu_2$, bilatéral
$H_1: \mu_1 > \mu_2$, unilatéral
$H_1: \mu_1 < \mu_2$, unilatéral



t calculé pour ν = n-1 degrés de liberté

Comparaison des fréquences de 2 grands échantillons indépendants.

Deux échantillons : f_1, n_1; f_2, n_2
On approxime la loi binomiale par la loi normale mais:
$n_1>30$, $n_2>30$, $n_1f_1>5$, $n_2f_2>5$, $n_1(1-f_1)>5$, $n_2(1-f_2)>5$

$H_0 : p_1 = p_2 = p$

Sous H_0 on peut réunir les deux échantillons, et on est conduit à l'estimation de p

$$\hat{p} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$$

Z_c devient

$$Z_c = \frac{f_1 - f_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

H_1: $p_1 \neq p_2$
H_1: $p_1 > p_2$
H_1: $p_1 < p_2$

$\longrightarrow$ Test sur la **loi normale** N(0,1)

Comparaison d'une fréquence empirique et d'une fréquence théorique

La différence entre f (mesuré) et p (théorique) est-elle seulement explicable par les aléas dus à l'échantillonnage?

On approxime la loi binomiale par la loi normale mais:
$n>30$, $np>5$ et $nq>5$

H_0: f = p

$$\boxed{Z_c = \frac{f-p}{\sqrt{\frac{p(1-p)}{n}}}}$$

H_1: $f \neq p$
H_1: $f > p$
H_1: $f < p$

$\longrightarrow$ Test sur la **loi normale** N(0,1)

Comparaison de deux variances expérimentales

Deux échantillons qui suivent des **lois normales**: μ_1, σ^2_1; μ_2, σ^2_2

H_0: $\sigma^2_1=\sigma^2_2$

calcul de :

Plus grande variance

$$\boxed{F_c = \frac{s^2_{x_A}}{s^2_{x_B}}}$$

>1

Plus petite variance

Si H_0 est vraie, F_c suit une loi de **Fisher-Snedecor** avec $\nu_1=n_1-1$ et $\nu_2=n_2-1$

La loi de Fisher - Snedecor : $F(\nu_1,\nu_2)$

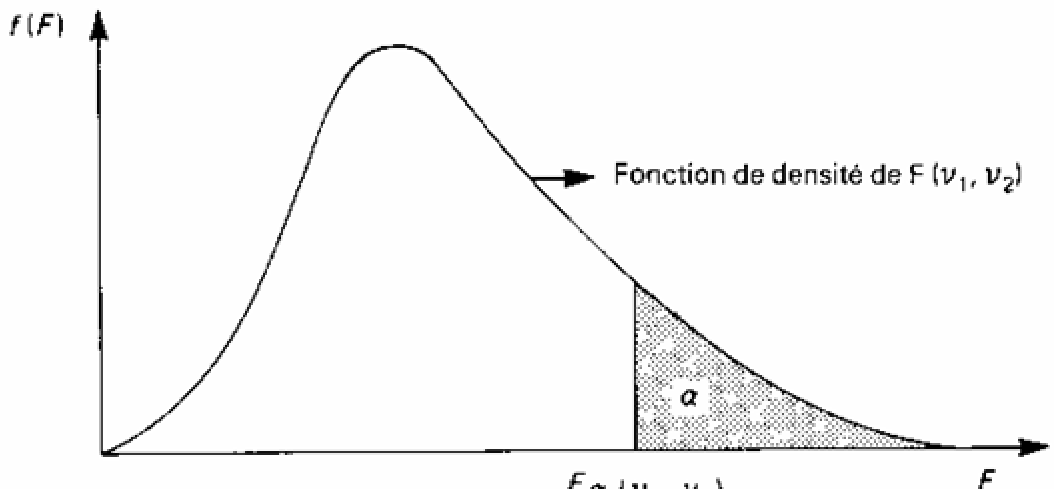
Soit χ^2_1 et χ^2_2, un couple de variables aléatoires indépendantes suivant respectivement des lois du χ^2 à ν_1 et ν_2 degrés de libertés.

$$\boxed{F = \frac{\chi_1^2 / \nu_1}{\chi_2^2 / \nu_2}}$$

Utile pour les tests de variance et de covariance

$$\alpha = P(F_{(\nu_1,\nu_2)} > F_{\alpha(\nu_1,\nu_2)})$$

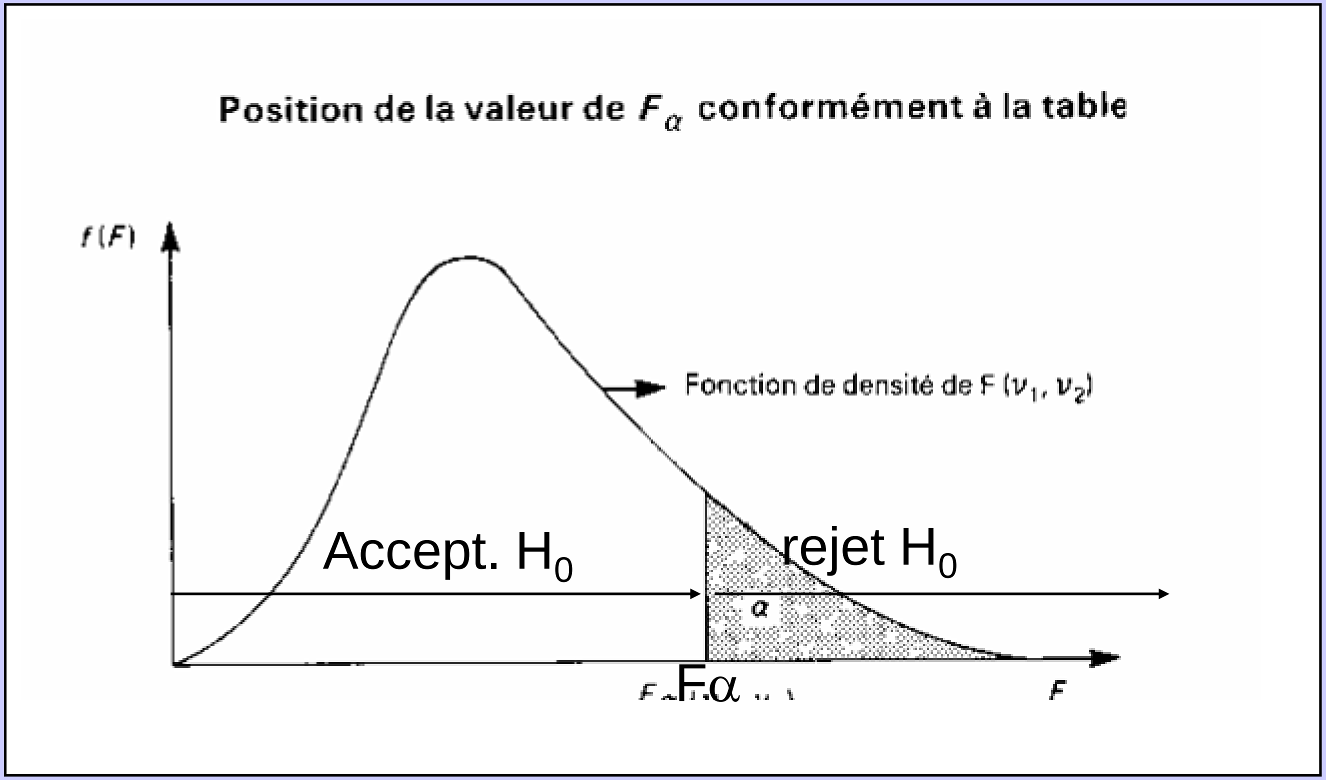
Position de la valeur de F_α conformément à la table



Comparaison de deux variances expérimentales

H_1: $\sigma^2_1 > \sigma^2_2$

Sous H_0: $Pr(F_c < F_\alpha) = 1 - \alpha$



Comparaison de deux variances expérimentales

H_1: $\sigma^2_1 \neq \sigma^2_2$

Sous H_0 : $Pr(F_c < F_{\alpha/2}) = 1-\alpha$

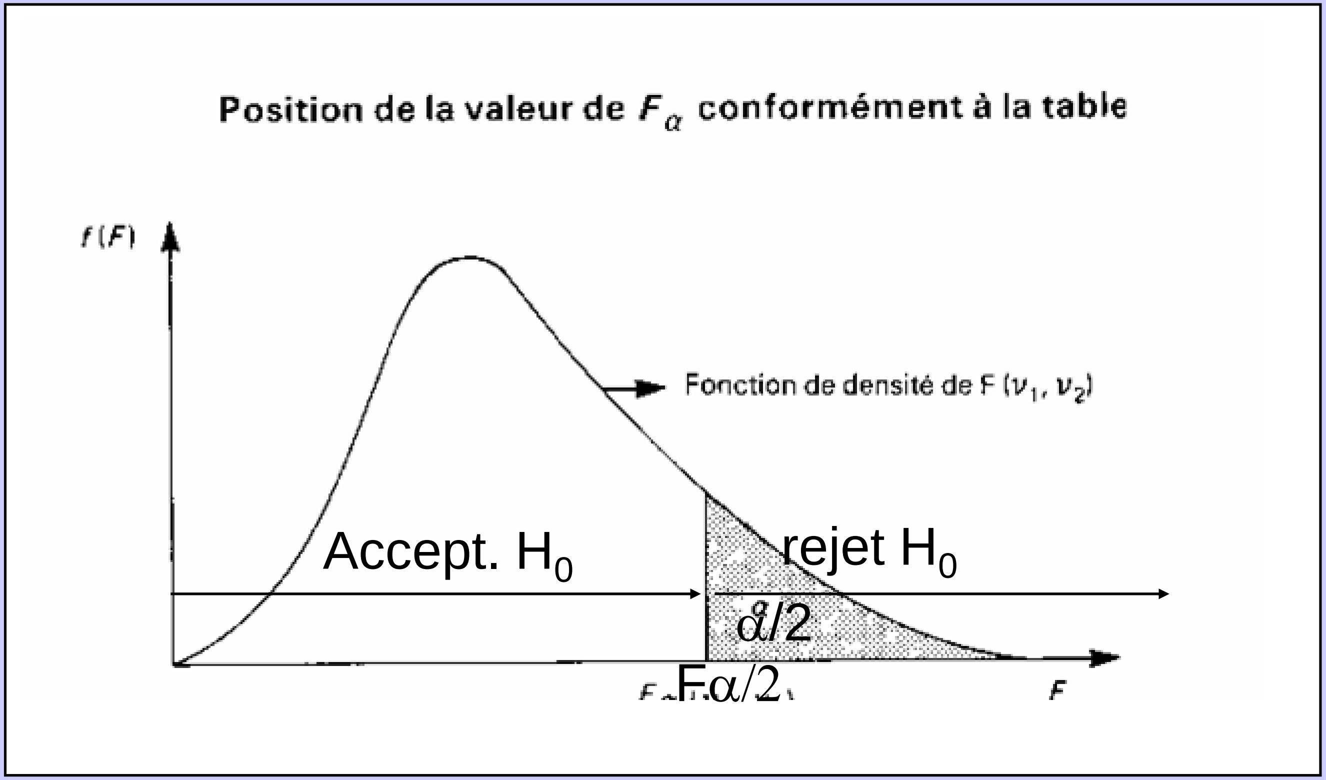


Table de Fisher-Snedecor

— ν_1 : nombre de degrés de liberté pour la plus faible des deux variances;

— ν_2 : nombre de degrés de liberté pour la plus forte des deux variances.

Pour $\alpha = 0{,}05$:

$\nu_1 \rightarrow$ / $\nu_2 \downarrow$	1	2	3	4	5	6	8	12	24	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,01	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00