

REVUE DE STATISTIQUE APPLIQUÉE

G. HISLEUR

Détermination de la taille de l'échantillon dans un test Student

Revue de statistique appliquée, tome 17, n° 1 (1969), p. 69-77

http://www.numdam.org/item?id=RSA_1969__17_1_69_0

© Société française de statistique, 1969, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

DÉTERMINATION DE LA TAILLE DE L'ÉCHANTILLON DANS UN TEST STUDENT

G. HISLEUR

Laboratoire de Mathématiques Appliquées de GRENOBLE

1 - INTRODUCTION

Un problème que l'on rencontre fréquemment en statistique est celui de la taille requise d'un échantillon pour obtenir un résultat déterminé. Nous nous proposons ici de donner huit tables qui permettront la résolution du problème suivant :

Disposant d'une estimation préalable s_0^2 avec $n_0 - 1$ degrés de liberté de la variance d'une population normale, quel doit être l'effectif n de l'échantillon à prendre pour que, dans le test de l'hypothèse $H_0 : m = m_0$ contre l'hypothèse $H : m > m_0$ au niveau de signification α , on accepte H_0 avec la probabilité β alors qu'en fait $m = m_0 + a$?

2 - TABLES

Ce problème a été traité numériquement dans les seuls cas particuliers $\alpha = 0.05$; $\beta = 0.20$ et $\alpha = 0.05$; $\beta = 0.05$ par M. HARRIS, D.G. HORVITZ et A.M. MOOD [1]. Nous avons écrit une procédure complète calquée sur celle qu'ils utilisèrent et qui permet de trouver la solution pour tous les couples (α, β) . Nous en rappellerons l'essentiel au paragraphe 4.

Nous donnons ici les tables qui correspondent à $\alpha = 0.01$; $\beta = 0.50, 0.20, 0.10, 0.05$, et $\alpha = 0.05$; $\beta = 0.50, 0.20, 0.10, 0.05$.

3 - UTILISATION

Soit donc à déterminer le nombre $n - 1$, nombre de degrés de liberté lors de la seconde évaluation s^2 de la variance.

Calculons d'abord le rapport

$$k = a/s_0$$

puis cherchons dans la table cette valeur qui se trouve à l'intersection de la colonne qui correspond à $n_0 - 1$ et de la ligne qui correspond au $n - 1$ cherché. La taille de l'échantillon à prendre est alors n .

Lorsque k est inférieur à la dernière entrée de la colonne d'entête $n_0 - 1$ la taille requise est alors supérieure à 120 mais on peut la déterminer de la manière suivante :

$$n - 1 = 100 (K/k)^2$$

où K est la valeur tabulée pour $n - 1 = 120$ dans la colonne d'entête $n_0 - 1$.

Signalons que dans les tables l'interpolation est pratiquement linéaire par rapport à l'inverse de $n_0 - 1$ et également par rapport à l'inverse de la racine carrée de $n - 1$.

Les colonnes qui correspondent à $n - 1$ infini ont été obtenues à partir des tables de la fonction puissance du test de Student de NEYMAN et TOKARSKA [2].

4 - PROBLEME THEORIQUE

Soit une population normale de variance inconnue, on se propose de tester l'hypothèse $H_0 : m = m_0$ contre l'hypothèse $H : m > m_0$, au niveau de signification α .

La méthode usuelle consiste alors à accepter H_0 si la condition suivante est réalisée :

$$\frac{\bar{x} - m_0}{s} \sqrt{n} < t_{1-\alpha}(n-1),$$

\bar{x} et s^2 étant la moyenne et la variance empiriques estimées à l'aide d'un échantillon de taille n et $t_{1-\alpha}(n-1)$ la valeur pour $1 - \alpha$ de la variable de Student à $n - 1$ degrés de liberté.

Avoir une probabilité β d'accepter H_0 , si en fait $m = m_0 + a$, revient à écrire :

$$\begin{aligned} \beta &= \Pr \left[\frac{\bar{x} - m_0}{s} \sqrt{n} < t_{1-\alpha}(n-1) \right] \\ &= \Pr \left[\frac{\bar{x} - m}{s} \sqrt{n} + \frac{a}{s} \sqrt{n} < t_{1-\alpha}(n-1) \right] \end{aligned}$$

où la probabilité dépend à la fois de m , connu, et σ , inconnu.

Or on dispose d'une information sur σ , de la forme d'une estimation préalable s_0^2 avec $n_0 - 1$ degrés de liberté.

On peut alors écrire :

$$\beta = \Pr [t < t_{1-\alpha}(n-1) - D \sqrt{F} \sqrt{n}]$$

en posant $F = s_0^2/s^2$, variable qui suit une loi de Fisher-Snedecor à $n_0 - 1$ et $n - 1$ degrés de liberté, et $D = a/s_0$.

C'est la relation utilisée par M. HARRIS, D.G. HORVITZ et A.M. MOOD [1], qui ont calculé les valeurs de $D = \frac{m - m_0}{s_0}$ correspondant à des valeurs données de n_0 et n , pour $\alpha = 0.05$; $\beta = 0.05$ et 0.20 .

Quand D est fixé la loi de probabilité qui intervient dans l'équation ci-dessus est la loi des deux variables t et F conditionnée par s_0 et on peut démontrer en utilisant un nouveau changement de variables que, considérée comme fonction de D, $\Pr [t < t_{1-\alpha}(n) - D \sqrt{F} \sqrt{n}]$ ne dépend pas de σ et est monotone et strictement croissante.

Il suffit donc de rechercher deux valeurs encadrant la solution, qu'un banal processus de bisection permet alors de calculer.

5 - COMPARAISON AVEC LA METHODE DE STEIN

Nous pouvons utiliser une autre méthode bien connue qui est due à C. STEIN [4].

Le premier échantillon de taille n permet le calcul de s_0^2 . La taille du second échantillon à prélever est alors $n - n_0$, le nombre entier n étant défini au moyen de l'égalité :

$$n = \max \left\{ \left[\frac{s_0^2}{z} \right] + 1, n_0 + 1 \right\}$$

où z est une constante positive qui, ainsi que nous le verrons plus loin, dépend directement de β ; $[q]$ représente la partie entière du nombre q.

Il faut ensuite les nombres réels b_i ($i = 1, 2, \dots, n$) tels que :

$$\sum_{i=1}^n b_i = 1 ; b_1 = b_2 = \dots = b_{n_0} ; \sum_{i=1}^n b_i^2 = \frac{z}{s_0^2}$$

Introduisons maintenant la statistique t définie par :

$$t = \frac{\sum_{i=1}^n b_i x_i - m_0}{\sqrt{z}} = \frac{\sum_{i=1}^n b_i x_i - m}{\sqrt{z}} + \frac{m - m_0}{\sqrt{z}}$$

où

$$u = \frac{\sum_{i=1}^n b_i x_i - m}{\sqrt{z}}$$

a la distribution de Student à $n_0 - 1$ degrés de liberté.

Pour tester $H_0 : m = m_0$ contre $H = m > m_0$, la région critique est donnée par :

$$\frac{\sum_{i=1}^n b_i x_i - m_0}{\sqrt{z}} > t_{1-\alpha}(n_0 - 1)$$

Accepter H_0 avec une probabilité β , si en fait $m = m_0 + a$, revient à écrire :

$$\beta = \Pr \left[u + \frac{m - m_0}{\sqrt{z}} < t_{1-\alpha}(n_0 - 1) \right]$$

Cette relation est utilisée pour déterminer la valeur de la constante z , qui permet ensuite de trouver la taille requise.

A titre comparatif pour $n_0 = 7$, $s_0^2 = 4$, $a = 3.7$, $\alpha = 0.01$ et $\beta = 0.10$ nous avons obtenu par :

- a) la méthode de STEIN $n - n_0 = 1$
- b) la méthode de HARRIS, HORVITZ et MOOD $n = 9$.

Nous pouvons remarquer que la première méthode tient certainement plus compte de l'information apportée par le premier échantillon. Par contre elle introduit une complication supplémentaire puisqu'elle nécessite le calcul des nombres b_1 qui dépendent de n_0 , n et z/s_0^2 .

6 - TEST BILATERAL

Les tables sont encore utilisables si l'on désire effectuer un test bilatéral, c'est-à-dire tester l'hypothèse $H_0 : m = m_0$ contre l'hypothèse $H : m \neq m_0$. Dans ce cas le niveau de signification est 2α et les tables correspondent à $\alpha = 0.02 ; \beta = 0.50, 0.20, 0.10, 0.05$, et $\alpha = 0.10 ; \beta = 0.50, 0.20, 0.10, 0.05$.

Les tableaux suivants donnent les valeurs de $k = a/s_0$ pour ALPHA := 0.01, 0.05 ; BETA := 0.50, 0.20, 0.10, 0.05.

ALPHA = 0.01

BETA = 0.50

$n - 1 \backslash n_0 - 1$	2	4	6	12	24	Infini
2	4.022	3.670	3.565	3.464	3.413	3.366
4	1.827	1.677	1.632	1.588	1.569	1.552
6	1.329	1.219	1.187	1.157	1.145	1.130
8	1.094	1.004	0.979	0.955	0.943	0.933
12	0.852	0.782	0.762	0.743	0.736	0.727
16	0.722	0.662	0.646	0.630	0.622	0.616
20	0.640	0.587	0.572	0.559	0.552	0.546
25	0.567	0.520	0.507	0.495	0.489	0.482
30	0.512	0.471	0.458	0.448	0.442	0.438
40	0.441	0.406	0.394	0.386	0.381	0.364
50	0.393	0.361	0.351	0.343	0.339	0.326
60	0.357	0.328	0.319	0.312	0.308	0.298
80	0.308	0.283	0.276	0.269	0.266	0.259
100	0.275	0.253	0.246	0.240	0.237	0.232
120	0.251	0.230	0.224	0.219	0.216	0.212

ALPHA = 0.05

BETA = 0.50

$n - 1 \backslash n_0 - 1$	2	4	6	12	24	Infini
2	1.687	1.552	1.509	1.475	1.455	1.438
4	1.032	0.952	0.931	0.909	0.898	0.890
6	0.815	0.753	0.736	0.718	0.710	0.703
8	0.694	0.642	0.627	0.613	0.605	0.600
12	0.561	0.518	0.506	0.495	0.490	0.483
16	0.483	0.447	0.436	0.426	0.420	0.417
20	0.431	0.399	0.389	0.380	0.375	0.371
25	0.384	0.356	0.347	0.339	0.335	0.331
30	0.350	0.324	0.316	0.309	0.306	0.302
40	0.303	0.280	0.273	0.267	0.264	0.256
50	0.271	0.250	0.244	0.239	0.236	0.230
60	0.247	0.228	0.223	0.218	0.215	0.210
80	0.214	0.197	0.193	0.189	0.186	0.182
100	0.192	0.176	0.172	0.169	0.167	0.163
120	0.175	0.161	0.157	0.154	0.152	0.149

ALPHA = 0.01

BETA = 0.20

$n - 1 \backslash n_0 - 1$	2	4	6	12	24	Infini
2	8.104	6.393	5.940	5.353	5.349	5.173
4	3.528	2.746	2.537	2.350	2.263	2.182
6	2.544	1.976	1.823	1.686	1.623	1.561
8	2.088	1.619	1.494	1.381	1.329	1.280
12	1.625	1.258	1.161	1.072	1.033	0.993
16	1.375	1.065	0.982	0.908	0.873	0.842
20	1.219	0.941	0.870	0.805	0.774	0.742
25	1.079	0.833	0.771	0.712	0.685	0.657
30	0.976	0.755	0.697	0.644	0.619	0.596
40	0.841	0.651	0.600	0.554	0.533	0.495
50	0.748	0.580	0.534	0.493	0.474	0.444
60	0.680	0.528	0.485	0.448	0.431	0.406
80	0.585	0.455	0.419	0.387	0.372	0.352
100	0.521	0.406	0.374	0.345	0.332	0.315
120	0.473	0.370	0.341	0.315	0.303	0.288

ALPHA = 0.05

BETA = 0.20

$n - 1 \backslash n_0 - 1$	2	4	6	12	24	Infini
2	3.510	2.805	2.620	2.448	2.371	2.298
4	2.098	1.669	1.554	1.452	1.405	1.360
6	1.649	1.311	1.220	1.139	1.101	1.066
8	1.405	1.117	1.041	0.970	0.939	0.910
12	1.114	0.902	0.840	0.784	0.758	0.732
16	0.976	0.775	0.722	0.673	0.651	0.631
20	0.873	0.693	0.646	0.602	0.583	0.563
25	0.779	0.619	0.577	0.538	0.520	0.502
30	0.708	0.563	0.525	0.489	0.473	0.456
40	0.613	0.486	0.454	0.423	0.409	0.395
50	0.548	0.435	0.405	0.378	0.365	0.353
60	0.499	0.396	0.369	0.344	0.333	0.322
80	0.432	0.342	0.319	0.298	0.288	0.278
100	0.385	0.306	0.285	0.266	0.257	0.249
120	0.351	0.279	0.260	0.243	0.235	0.227

ALPHA = 0.01

BETA = 0.10

$n - 1 \backslash n_0 - 1$	2	4	6	12	24	Infini
2	12.173	8.470	7.588	6.838	6.506	6.201
4	5.245	3.568	3.166	2.820	2.666	2.522
6	3.776	2.557	2.265	2.011	1.899	1.792
8	3.096	2.092	1.850	1.642	1.550	1.463
12	2.406	1.625	1.437	1.274	1.201	1.134
16	2.038	1.376	1.215	1.078	1.016	0.958
20	1.807	1.219	1.077	0.955	0.900	0.847
25	1.600	1.079	0.953	0.845	0.796	0.749
30	1.447	0.975	0.862	0.764	0.720	0.679
40	1.241	0.840	0.742	0.658	0.620	0.564
50	1.100	0.747	0.660	0.585	0.551	0.506
60	0.996	0.679	0.600	0.532	0.501	0.462
80	0.849	0.587	0.518	0.460	0.433	0.401
100	0.748	0.524	0.462	0.410	0.386	0.359
120	0.674	0.477	0.421	0.374	0.352	0.328

ALPHA = 0.05

BETA = 0.10

$n - 1 \backslash n_0 - 1$	2	4	6	12	24	Infini
2	5.308	3.747	3.372	3.051	2.909	2.778
4	3.152	2.204	1.975	1.779	1.691	1.610
6	2.474	1.727	1.547	1.393	1.324	1.259
8	2.109	1.471	1.317	1.186	1.127	1.070
12	1.701	1.184	1.059	0.953	0.906	0.863
16	1.464	1.020	0.913	0.821	0.780	0.742
20	1.307	0.910	0.815	0.733	0.696	0.661
25	1.167	0.812	0.727	0.653	0.621	0.590
30	1.063	0.739	0.662	0.595	0.565	0.537
40	0.920	0.639	0.572	0.514	0.489	0.458
50	0.822	0.571	0.511	0.460	0.436	0.410
60	0.750	0.521	0.466	0.419	0.398	0.375
80	0.650	0.451	0.403	0.363	0.344	0.326
100	0.581	0.403	0.361	0.324	0.308	0.292
120	0.531	0.368	0.329	0.296	0.281	0.266

ALPHA = 0.01

BETA = 0.05

$n - 1 \backslash n_0 - 1$	2	4	6	12	24	Infini
2	17.698	10.743	9.256	8.063	7.541	7.078
4	7.589	4.474	3.805	3.259	3.023	2.808
6	5.460	3.199	2.711	2.312	2.141	1.980
8	4.475	2.616	2.213	1.885	1.744	1.613
12	3.481	2.030	1.178	1.462	1.349	1.248
16	2.949	1.717	1.452	1.235	1.141	1.055
20	2.601	1.522	1.286	1.094	1.011	0.932
25	2.290	1.347	1.138	0.969	0.895	0.824
30	2.061	1.218	1.029	0.876	0.809	0.747
40	1.749	1.049	0.886	0.754	0.696	0.620
50	1.536	0.933	0.789	0.671	0.620	0.556
60	1.379	0.848	0.717	0.610	0.563	0.508
80	1.196	0.734	0.619	0.527	0.486	0.441
100	1.071	0.656	0.552	0.470	0.434	0.395
120	0.979	0.598	0.503	0.428	0.395	0.361

ALPHA = 0.05

BETA = 0.05

$n - 1 \backslash n_0 - 1$	2	4	6	12	24	Infini
2	7.734	4.771	4.129	3.610	3.385	3.187
4	4.582	2.789	2.397	2.080	1.944	1.820
6	3.595	2.182	1.876	1.623	1.516	1.417
8	3.066	1.858	1.597	1.382	1.290	1.207
12	2.470	1.495	1.285	1.112	1.038	0.971
16	2.129	1.287	1.106	0.956	0.892	0.834
20	1.900	1.149	0.991	0.858	0.798	0.744
25	1.694	1.026	0.884	0.765	0.712	0.663
30	1.543	0.935	0.804	0.695	0.647	0.605
40	1.338	0.809	0.696	0.601	0.560	0.525
50	1.198	0.722	0.622	0.537	0.500	0.469
60	1.095	0.658	0.567	0.490	0.456	0.428
80	0.943	0.569	0.490	0.423	0.395	0.369
100	0.840	0.508	0.438	0.378	0.353	0.329
120	0.767	0.465	0.399	0.345	0.322	0.300

BIBLIOGRAPHIE

- [1] Marilyn HARRIS, D.G. HORVITZ, and A.M. MOOD - "On the détermination of sample sizes in designing experiments", Journal of the American Statistical Association, Septembre 1948, p. 391.
- [2] J. NEYMAN and B. TOKARSKA - "Errors of the Second Kind in Testing Student's Hypothesis", Journal of the American Statistical Association, Vol. 31, 1936, p. 318.
- [3] E. MORICE - "Puissance de quelques tests classiques", Revue de Statistique appliquée (XVI), 1968 n° 1, p. 90. (On y trouvera des graphiques correspondant aux tests $\alpha = 0,05$ et $\beta = 0,10, 0,20$).
- [4] C. STEIN - "A two sample test for a linear hypothesis whose power is independant of the variance", Annals of Mathematical Statistics, Vol. 16, 1945, p. 243.