

Aplikace matematiky

Zbyněk Šidák

Tables for the two-sample location E -test based on exceeding observations

Aplikace matematiky, Vol. 22 (1977), No. 3, 166–175

Persistent URL: <http://dml.cz/dmlcz/103689>

Terms of use:

© Institute of Mathematics AS CR, 1977

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://dml.cz>

TABLES FOR THE TWO-SAMPLE LOCATION *E*-TEST BASED
ON EXCEEDING OBSERVATIONS

ZBYNĚK ŠIDÁK

(Received December 29, 1975)

Summary. The rank statistic *E*, based on the minimum number of exceeding observations in two samples, gives rise to a quick and easy *E*-test, which is suitable for the two-sample location problem. This paper contains tables of the one-sided significance levels $P\{E \geq k\}$ for $2 \leq k \leq 6$ (which includes usual significance levels) for sizes *m*, *n* of the two samples satisfying $3 \leq m \leq n \leq 25$.

Description of the test. Let us have two random samples X_1, \dots, X_m and Y_1, \dots, Y_n with densities f_1 and f_2 , respectively. Suppose that the notation is chosen so that $m \leq n$. We wish to test the hypothesis H_0 that f_1 and f_2 are identical but otherwise arbitrary against the alternatives of shift in location which may be expressed by $f_1(x) = f(x - \Delta)$, $f_2(x) = f(x)$, where $\Delta > 0$, or $\Delta < 0$ (one-sided alternatives), or $\Delta \neq 0$ (two-sided alternative).

The *E*-test is performed very quickly and simply. First, we find the quantities *A*, and *B'*, being equal to the number of observations among X_1, \dots, X_m larger than $\max_{1 \leq j \leq n} Y_j$, or smaller than $\min_{1 \leq j \leq n} Y_j$, respectively, and the quantities *A'*, and *B*, being equal to the number of observations among Y_1, \dots, Y_n larger than $\max_{1 \leq i \leq m} X_i$, or smaller than $\min_{1 \leq i \leq m} X_i$, respectively. (Of course, only one of the numbers *A*, *A'* (or *B*, *B'*) is positive, while the other must be zero.) Our test statistic is then

$$E = \min(A, B) - \min(A', B').$$

The test based on *E* can be used against both two-sided and one-sided alternatives. However, let us remark that against the one-sided alternatives $\Delta > 0$ we might use the statistic $\min(A, B)$; this latter statistic generates the locally most powerful rank test for testing H_0 against a shift Δ of a uniform distribution for Δ close to 0, i.e. in some region $0 < \Delta < \varepsilon$. (Cf. Hájek-Šidák [2], Section III.1.2 and Problem II.13.)

Description of the table. In Table 1 we tabulate (in per cents) the upper tails $100 \cdot P\{E \geq k\}$ of the distribution of E under H_0 , i.e. the one-sided significance levels of the E -test in per cents, for $2 \leq k \leq 6$ and for the sample sizes m, n satisfying $3 \leq m \leq n \leq 25$. This range of levels includes all commonly used significance levels. Let us also remark that it can be easily proved that, for each n fixed and for $1 \leq k \leq n - 1$, the probabilities $P\{E \geq k\}$ for $m = n - 1$ and for $m = n$ coincide.

For the computation of Table 1 we used the formula

$$P\{E \geq k\} = \binom{m+n-2k}{m-k} \binom{m+n}{m}^{-1}$$

valid for $k = 1, 2, \dots, \min(m, n)$. (Cf. Hájek-Šidák [2], Theorem IV.2.2.b.)

The distribution of E under H_0 is symmetric about 0 so that $P\{E \leq -k\} = P\{E \geq k\}$.

If we are testing H_0 against the one-sided alternative $\Delta > 0$ (i.e. f_1 is shifted to the right with respect to f_2), we use a critical region $\{E \geq k\}$, and its significance level may be found in Table 1. If we are testing against $\Delta < 0$ (i.e. f_1 is shifted to the left), we use a critical region $\{E \leq -k\}$, but its significance level is equal to that of $\{E \geq k\}$ so that it again may be found directly in Table 1. If we are testing against the two-sided alternative $\Delta \neq 0$, we use a critical region $\{|E| \geq k\}$ whose significance level is equal to $2P\{E \geq k\}$.

Example. Let one sample contain the values 42; 28; 49; 36; 48; the other sample the values 34; 22; 14; 24. For the application of Table 1 we must have $m \leq n$, therefore the latter sample with 4 values must be denoted as X_1, \dots, X_4 , and the former sample with 5 values as Y_1, \dots, Y_5 . We see that $E = -\min(4, 3) = -3$. If we wish to test against $\Delta < 0$, we may use e.g. the critical region $\{E \leq -3\}$, and in Table 1 for $n = 5, m = 4, k = 3$ we find its significance level 2.381%. If we wish to test against $\Delta \neq 0$, we may use the critical region $\{|E| \geq 3\}$, and its significance level is twice the tabulated number, that is 4.762%.

Remark on the asymptotic distribution. Let $m, n \rightarrow \infty$, and $m/n \rightarrow r, 0 < r < \infty$. Then it is easy to prove that

$$\lim_{m, n \rightarrow \infty} P\{E \geq k\} = [(1 + r^{-1})(1 + r)]^{-k}$$

for $k = 1, 2, \dots$. In particular, if $m = n \rightarrow \infty$, we have

$$\lim_{m=n \rightarrow \infty} P\{E \geq k\} = 1/4^k$$

for $k = 1, 2, \dots$; these numbers for $2 \leq k \leq 6$ are given at the end of Table 1.

Table 1
 One-sided significance levels $100P\{E \geq k\}$ (i.e. in percents)

n	m	k				
		2	3	4	5	6
3	3	10.000	5.000	—	—	—
4	3	8.571	2.857	—	—	—
	4	8.571	2.857	1.429	—	—
5	3	7.143	1.786	—	—	—
	4	7.937	2.381	0.794	—	—
	5	7.937	2.381	0.794	0.397	—
6	3	5.952	1.190	—	—	—
	4	7.143	1.905	0.476	—	—
	5	7.576	2.164	0.649	0.216	—
	6	7.576	2.164	0.649	0.216	0.108
7	3	5.000	0.833	—	—	—
	4	6.364	1.515	0.303	—	—
	5	7.071	1.894	0.505	0.126	—
	6 or 7	7.343	2.040	0.583	0.175	0.058
8	3	4.242	0.606	—	—	—
	4	5.657	1.212	0.202	—	—
	5	6.527	1.632	0.388	0.078	—
	6	6.993	1.865	0.500	0.133	0.033
	7 or 8	7.179	1.958	0.544	0.155	0.047
9	3	3.636	0.455	—	—	—
	4	5.035	0.979	0.140	—	—
	5	5.994	1.399	0.300	0.050	—
	6	6.593	1.678	0.420	0.100	0.020
	7	6.923	1.836	0.490	0.131	0.035
	8 or 9	7.059	1.900	0.518	0.144	0.041
10	3	3.147	0.350	—	—	—
	4	4.496	0.799	0.100	—	—
	5	5.494	1.199	0.233	0.033	—
	6	6.181	1.498	0.350	0.075	0.012
	7	6.618	1.697	0.432	0.108	0.026
	8	6.863	1.810	0.480	0.128	0.034
	9 or 10	6.966	1.858	0.500	0.136	0.038
11	3	2.747	0.275	—	—	—
	4	4.029	0.659	0.073	—	—
	5	5.037	1.030	0.183	0.023	—

Table 1 — continued

<i>n</i>	<i>m</i>	<i>k</i>					
		2	3	4	5	6	
11	6	5.777	1.333	0.291	0.057	0.008	
	7	6.291	1.555	0.377	0.088	0.019	
	8	6.622	1.703	0.437	0.111	0.028	
	9	6.811	1.788	0.472	0.125	0.033	
	10 or 11	6.892	1.824	0.487	0.131	0.036	
12	3	2.418	0.220	—	—	—	
	4	3.626	0.549	0.055	—	—	
	5	4.622	0.889	0.145	0.016	—	
	6	5.392	1.185	0.242	0.043	0.005	
	7	5.960	1.419	0.327	0.071	0.014	
	8	6.357	1.589	0.393	0.095	0.022	
	9	6.617	1.703	0.438	0.112	0.029	
	10	6.767	1.769	0.464	0.122	0.032	
	11 or 12	6.832	1.798	0.476	0.127	0.034	
	13	3	2.143	0.179	—	—	—
4		3.277	0.462	0.042	—	—	
5		4.248	0.770	0.117	0.012	—	
6		5.031	1.054	0.203	0.033	0.004	
7		5.635	1.291	0.284	0.058	0.010	
8		6.082	1.476	0.351	0.081	0.018	
9		6.398	1.610	0.402	0.100	0.024	
10		6.606	1.700	0.437	0.112	0.029	
11		6.729	1.753	0.458	0.120	0.032	
12 or 13		6.783	1.776	0.467	0.124	0.033	
14		3	1.912	0.147	—	—	—
	4	2.974	0.392	0.033	—	—	
	5	3.913	0.671	0.095	0.009	—	
	6	4.696	0.939	0.170	0.026	0.003	
	7	5.322	1.174	0.246	0.047	0.008	
	8	5.805	1.366	0.313	0.069	0.014	
	9	6.166	1.514	0.367	0.087	0.020	
	10	6.423	1.623	0.408	0.102	0.025	
	11	6.594	1.696	0.436	0.112	0.029	
	12	6.696	1.739	0.453	0.118	0.031	
	13 or 14	6.741	1.758	0.461	0.121	0.032	
	15	3	1.716	0.123	—	—	—
		4	2.709	0.335	0.026	—	—
5		3.612	0.587	0.077	0.006	—	
6		4.386	0.838	0.144	0.020	0.002	

Table 1 — continued

<i>n</i>	<i>m</i>	<i>k</i>					
		2	3	4	5	6	
15	7	5.024	1.067	0.213	0.039	0.006	
	8	5.534	1.262	0.278	0.058	0.011	
	9	5.929	1.420	0.334	0.077	0.017	
	10	6.225	1.542	0.379	0.092	0.022	
	11	6.438	1.630	0.412	0.104	0.026	
	12	6.581	1.691	0.435	0.112	0.029	
	13	6.667	1.727	0.449	0.117	0.031	
	14 or 15	6.705	1.743	0.455	0.119	0.031	
16	3	1.548	0.103	—	—	—	
	4	2.477	0.289	0.021	—	—	
	5	3.342	0.516	0.064	0.005	—	
	6	4.101	0.751	0.122	0.016	0.001	
	7	4.743	0.971	0.186	0.032	0.004	
	8	5.270	1.165	0.247	0.049	0.009	
	9	5.692	1.328	0.303	0.067	0.014	
	10	6.020	1.459	0.349	0.082	0.019	
	11	6.268	1.561	0.386	0.095	0.023	
	12	6.447	1.635	0.414	0.105	0.026	
	13	6.568	1.686	0.433	0.111	0.029	
	14	6.641	1.716	0.445	0.116	0.030	
	15 or 16	6.674	1.730	0.450	0.117	0.031	
	17	3	1.404	0.088	—	—	—
		4	2.272	0.251	0.017	—	—
5		3.099	0.456	0.053	0.004	—	
6		3.840	0.674	0.104	0.013	0.001	
7		4.480	0.884	0.162	0.026	0.003	
8		5.017	1.075	0.220	0.042	0.007	
9		5.458	1.240	0.274	0.058	0.012	
10		5.812	1.378	0.322	0.073	0.016	
11		6.089	1.489	0.361	0.086	0.020	
12		6.299	1.575	0.392	0.097	0.024	
13		6.451	1.638	0.415	0.105	0.027	
14		6.555	1.681	0.431	0.111	0.028	
15		6.618	1.707	0.441	0.114	0.030	
16 or 17		6.647	1.719	0.446	0.116	0.030	
18		3	1.278	0.075	—	—	—
		4	2.092	0.219	0.014	—	—
	5	2.880	0.404	0.045	0.003	—	
	6	3.600	0.606	0.089	0.010	0.001	
	7	4.233	0.806	0.141	0.022	0.003	

Table 1 — continued

<i>n</i>	<i>m</i>	<i>k</i>				
		2	3	4	5	6
18	8	4.776	0.992	0.196	0.036	0.006
	9	5.231	1.158	0.248	0.051	0.010
	10	5.604	1.300	0.295	0.065	0.014
	11	5.905	1.417	0.336	0.078	0.018
	12	6.141	1.512	0.370	0.090	0.021
	13	6.321	1.585	0.396	0.099	0.024
	14	6.453	1.639	0.416	0.106	0.027
	15	6.543	1.676	0.430	0.110	0.028
	16	6.598	1.699	0.438	0.113	0.029
	17 or 18	6.623	1.709	0.442	0.115	0.030
19	3	1.169	0.065	—	—	—
	4	1.931	0.192	0.011	—	—
	5	2.682	0.360	0.038	0.002	—
	6	3.379	0.547	0.077	0.008	0.001
	7	4.003	0.737	0.124	0.018	0.002
	8	4.547	0.917	0.175	0.031	0.005
	9	5.011	1.080	0.224	0.044	0.008
	10	5.400	1.224	0.271	0.058	0.012
	11	5.720	1.346	0.312	0.071	0.016
	12	5.978	1.448	0.347	0.082	0.019
	13	6.182	1.529	0.376	0.092	0.022
	14	6.338	1.592	0.399	0.100	0.025
	15	6.453	1.639	0.416	0.106	0.027
	16	6.532	1.672	0.428	0.110	0.028
	17	6.580	1.691	0.436	0.112	0.029
18 or 19	6.602	1.701	0.439	0.114	0.029	
20	3	1.073	0.056	—	—	—
	4	1.788	0.169	0.009	—	—
	5	2.503	0.322	0.032	0.002	—
	6	3.177	0.495	0.066	0.007	—
	7	3.789	0.674	0.109	0.015	0.002
	8	4.330	0.847	0.156	0.026	0.004
	9	4.800	1.008	0.203	0.039	0.007
	10	5.200	1.152	0.248	0.052	0.010
	11	5.535	1.277	0.290	0.064	0.014
	12	5.812	1.384	0.326	0.076	0.017
	13	6.036	1.472	0.356	0.086	0.020
	14	6.214	1.543	0.382	0.094	0.023
	15	6.350	1.598	0.401	0.101	0.025
	16	6.451	1.639	0.416	0.106	0.027
17	6.521	1.667	0.427	0.109	0.028	

Table 1 — continued

<i>n</i>	<i>m</i>	<i>k</i>				
		2	3	4	5	6
20	18	6.564	1.685	0.433	0.112	0.029
	19 or 20	6.584	1.693	0.436	0.113	0.029
21	3	0.988	0.049	—	—	—
	4	1.660	0.150	0.008	—	—
	5	2.341	0.289	0.027	0.002	—
	6	2.991	0.449	0.058	0.006	—
	7	3.590	0.618	0.096	0.013	0.001
	8	4.126	0.784	0.139	0.023	0.003
	9	4.598	0.941	0.184	0.034	0.006
	10	5.006	1.084	0.228	0.046	0.009
	11	5.353	1.211	0.268	0.058	0.012
	12	5.645	1.321	0.305	0.069	0.015
	13	5.887	1.414	0.337	0.079	0.018
	14	6.083	1.491	0.364	0.088	0.021
	15	6.239	1.553	0.386	0.095	0.023
	16	6.359	1.602	0.403	0.101	0.025
	17	6.449	1.638	0.416	0.106	0.027
	18	6.511	1.663	0.425	0.109	0.028
	19	6.549	1.679	0.431	0.111	0.029
	20 or 21	6.567	1.686	0.434	0.112	0.029
	22	3	0.913	0.043	—	—
4		1.545	0.134	0.007	—	—
5		2.194	0.260	0.024	0.001	—
6		2.821	0.409	0.050	0.005	—
7		3.404	0.567	0.085	0.011	0.001
8		3.934	0.726	0.125	0.019	0.003
9		4.405	0.878	0.167	0.030	0.005
10		4.818	1.020	0.209	0.041	0.008
11		5.175	1.147	0.248	0.052	0.011
12		5.479	1.260	0.285	0.063	0.014
13		5.735	1.357	0.317	0.073	0.017
14		5.948	1.439	0.346	0.082	0.019
15		6.121	1.507	0.369	0.090	0.022
16		6.259	1.562	0.389	0.097	0.024
17		6.366	1.605	0.404	0.102	0.026
18		6.445	1.637	0.416	0.106	0.027
19		6.501	1.659	0.424	0.108	0.028
20		6.535	1.673	0.429	0.110	0.028
21 or 22		6.551	1.680	0.431	0.111	0.029
23	3	0.846	0.038	—	—	—

Table I — continued

<i>n</i>	<i>m</i>	<i>k</i>					
		2	3	4	5	6	
23	4	1.442	0.120	0.006	—	—	
	5	2.059	0.235	0.020	0.001	—	
	6	2.663	0.373	0.044	0.004	—	
	7	3.231	0.522	0.076	0.009	0.001	
	8	3.752	0.673	0.112	0.017	0.002	
	9	4.221	0.821	0.152	0.026	0.004	
	10	4.637	0.959	0.191	0.036	0.006	
	11	5.001	1.086	0.230	0.047	0.009	
	12	5.315	1.200	0.266	0.058	0.012	
	13	5.584	1.300	0.299	0.068	0.015	
	14	5.810	1.386	0.328	0.077	0.018	
	15	5.998	1.459	0.353	0.085	0.020	
	16	6.152	1.520	0.374	0.092	0.022	
	17	6.275	1.569	0.391	0.097	0.024	
	18	6.371	1.607	0.405	0.102	0.026	
	19	6.442	1.636	0.415	0.106	0.027	
	20	6.492	1.656	0.423	0.108	0.028	
	21	6.523	1.668	0.427	0.110	0.028	
	22 or 23	6.537	1.674	0.429	0.110	0.028	
	24	3	0.786	0.034	—	—	—
		4	1.348	0.107	0.005	—	—
5		1.937	0.213	0.018	0.001	—	
6		2.518	0.341	0.039	0.003	—	
7		3.070	0.481	0.067	0.008	0.001	
8		3.582	0.625	0.101	0.015	0.002	
9		4.047	0.768	0.138	0.023	0.003	
10		4.464	0.903	0.176	0.032	0.006	
11		4.832	1.029	0.213	0.042	0.008	
12		5.154	1.143	0.248	0.053	0.011	
13		5.433	1.245	0.281	0.062	0.013	
14		5.671	1.334	0.311	0.071	0.016	
15		5.872	1.411	0.337	0.080	0.019	
16		6.040	1.476	0.359	0.087	0.021	
17		6.178	1.530	0.378	0.093	0.023	
18		6.288	1.574	0.394	0.098	0.024	
19		6.374	1.609	0.406	0.102	0.026	
20		6.438	1.634	0.415	0.105	0.027	
21		6.483	1.652	0.421	0.108	0.027	
22		6.512	1.664	0.426	0.109	0.028	
23 or 24		6.525	1.669	0.427	0.110	0.028	
25		3	0.733	0.031	—	—	—

Table 1 -- continued

n	m	k					
		2	3	4	5	6	
25	4	1.263	0.097	0.004	—	—	
	5	1.824	0.194	0.015	0.001	—	
	6	2.384	0.312	0.034	0.003	—	
	7	2.920	0.444	0.060	0.007	0.001	
	8	3.421	0.581	0.091	0.013	0.002	
	9	3.881	0.718	0.125	0.020	0.003	
	10	4.297	0.850	0.161	0.029	0.005	
	11	4.669	0.974	0.197	0.038	0.007	
	12	4.997	1.088	0.232	0.048	0.010	
	13	5.283	1.191	0.264	0.057	0.012	
	14	5.532	1.283	0.294	0.066	0.015	
	15	5.745	1.363	0.321	0.075	0.017	
	16	5.925	1.432	0.344	0.082	0.019	
	17	6.075	1.491	0.364	0.089	0.021	
	18	6.199	1.539	0.381	0.094	0.023	
	19	6.298	1.579	0.395	0.099	0.025	
	20	6.376	1.610	0.406	0.102	0.026	
	21	6.434	1.633	0.414	0.105	0.027	
	22	6.475	1.649	0.420	0.107	0.027	
	23	6.501	1.660	0.424	0.108	0.028	
		24 or 25	6.513	1.664	0.426	0.109	0.028
	$n = m$ $n, m \rightarrow \infty$		6.250	1.562	0.391	0.098	0.024

Remark on ties. The preceding discussion and Table 1 concern the cases where the underlying distributions have some densities, so that all observations are distinct with probability 1. For other cases in practice in which some equal observations occur we can add the following two remarks. First, we may randomize the ordering of equal observations in these latter cases, and Table 1 then gives again precise significance levels. Second, if we follow strictly the above recipe for the E -test also in these latter cases, the procedure becomes a conservative one (because it uses, in fact, the "least favourable" configuration of equal observations, cf. Hájek-Šidák [2], Section III.8.2); in other words, the nominal levels in Table 1 become now upper bounds for the actual significance levels.

Bibliographical remarks. The E -test described above was introduced for the first time by Hájek-Šidák [2], Section III.1.2. Other closely related location tests based on the numbers of exceeding observations are as follows: the simplest test based

only on A was suggested by Rosenbaum [3], the test based on $A + B$ by Šidák-Vondráček [6] and by Tukey [7], the best based on $A + B - A' - B'$ by Haga [1]. For some more information cf. Hájek-Šidák [2], Section III.1.2, and a survey paper by Rosenbaum [4].

Finally, it might be mentioned that the power functions of the present E -test, the Haga test, and the Wilcoxon test have been compared for the uniform distribution in [5].

References

- [1] *T. Haga*: A two-sample rank test on location. *Ann. Inst. Statist. Math.* 11 (1959/60), 211–219.
- [2] *J. Hájek, Z. Šidák*: Theory of rank tests. Academia, Prague & Academic Press, New York—London, 1967.
- [3] *S. Rosenbaum*: Tables for a nonparametric test of location. *Ann. Math. Statist.* 25 (1954), 146–150.
- [4] *S. Rosenbaum*: On some two-sample non-parametric tests. *J. Amer. Statist. Assoc.* 60 (1965), 1118–1126.
- [5] *Z. Šidák, S. Hojek*: Monte Carlo comparisons of some rank tests optimal for uniform distributions. In *Contributions to Statistics — J. Hájek Memorial Volume* (in print).
- [6] *Z. Šidák, J. Vondráček*: A simple non-parametric test of the difference in location of two populations. (In Czech.) *Aplikace matematiky* 2 (1957), 215–221.
- [7] *J. W. Tukey*: A quick, compact, two-sample test to Duckworth's specifications. *Technometrics* 1 (1959), 31–48.

Souhrn

TABULKY PRO E -TEST POLOHY DVOU VÝBĚRŮ, ZALOŽENÝ NA PŘEKRAČUJÍCÍCH POZOROVÁNÍCH

ZBYNĚK ŠIDÁK

Pořadová statistika E , založená na minimu počtů překračujících pozorování ve dvou výběrech, dává vznik rychlému a snadnému E -testu, který je vhodný pro problém polohy dvou výběrů. Článek obsahuje tabulky jednostranných hladin významnosti $P\{E \geq k\}$ pro $2 \leq k \leq 6$ (což zahrnuje obvykle používané hladiny) pro rozsahy výběrů m, n splňující $3 \leq m \leq n \leq 25$.

Author's address: RNDr. Zbyněk Šidák, DrSc., Matematický ústav ČSAV, Žitná 25, 115 67 Praha 1.