TABLES FOR USE IN RANK CORRELATION

by

L. Kaarsemaker and A. van Wijngaarden.

Report R 73

of the

Computation Department of the Mathematical Centre.

Amsterdam.

1 9 5 2 .

1. Introduction.

In the so-called rank-correlation methods of modern statistics, treated extensively by M.G. Kendall in "Rank Correlation Methods", London 1948, two series of the same n objects but ranked according to two different properties of those objects are correlated in order to establish evidence on possible correlation between those properties. Therefore, a score S is allotted in the following way. For any pair of objects 1 is scored if the differences of the ranks in the two rankings have the same sign, and -1 is scored if they have opposite sign. The sum of all these $\binom{n}{2}$ partial scores is the total score S. Apparently holds $-\binom{n}{2} \leq S \leq \binom{n}{2}$. Moreover, if $n \equiv 0,1 \pmod 4$, S is always even, whereas, if $n \equiv 2,3 \pmod 4$, S is always odd. It is easy to see that rearranging the objects in such a way that the first ranking becomes the natural order 1,2,...,n does not influence the score. Hence, without loss of generality the first ranking can be assumed to be the natural order. There arises a complication if some objects are so similar with respect to a property that they cannot be ranked with respect to that property. They are said to be tied. Here only the untied case is considered.

Now, if the two properties are independent, all rankings are equally probable, and the probability $P_n[\underline{S} \geq S]$ that the random variable $\underline{S}$ (the random character of a variable is denoted by underlining its symbol) attains or exceeds a fixed value S can be calculated on the basis of which a null-hypothesis can be rejected or nonrejected.

Kendall (loc.cit.p.141) gives a table of $P_n[\underline{S} \geq S]$ for small values of n (up to 10). As the distribution of $\underline{S}$ is asymptotically normal with known standard deviation (Kendall, loc.cit.p.39), it is possible for higher values of n to calculate the probability with sufficient accuracy from a normal distribution after application of a correction for continuity. Quite apart from the fact that not an exact value is obtained in this way it is very annoying to have to perform these calculations regularly, and they are, moreover, a constant source of errors. Therefore, tables of the exact values are presented here for values of n up to 40 and for all corresponding values of S, for which the probability differs from zero in three decimals. For n = 40 up to 100 a table gives the standard deviation of the corresponding (and the practically identical) normal distribution.

2. Description of the tables.

Table I gives $10^3.P_n[\underline{S} \geq S]$ for n = 4,5, 8,9, 12, 13,...,36,37,40 and corresponding even values of S. Table II gives the same for

$n = 6,7,10,11,\ldots 38,39$ and corresponding odd values of S. If 000 is given the probability vanishes in three decimals, and if 0 is given, it is exactly zero.

Table III gives for $n = 4(1)40$, and for $\alpha = 0.005$, 0.010, 0.025, 0.050 and 0.100 the smallest value of S for which $P_n[\underline{S} \geq S] \leq \alpha$.

Table IV gives for $n = 40(1)100$ the standarddeviation $\sigma = \{n(n-1)(2n+5)\}^{\frac{1}{2}}$ of the corresponding normal distribution.

3. Mathematical background.

The probabilities are computed by means of a recurrence-relation that is a modification of the one given by Kendall (log.cit.pp.55,66). In order to derive this relation it is appropriate to change the notations for a while. If instead of scoring 1 or -1, one scores 1 or 0 under the same conditions, a score s is obtained that is related to S by the formula

$$S = 2\,s - \binom{n}{2} \tag{3.1}$$

Be the natural order correlated with all n! possible rankings. Then $u(n,s)$ denotes the number of times that $\underline{s} = s$, and $v(n,s) = u(n,0) + u(n,1) + \ldots u(n,s-1)$ the number of times that $\underline{s} < s$. Again, $F(n,s-1) = v(n,s)/n!$ is the probability that $\underline{s} < s$. This $F(n,s)$ is related to the probability $P_n[\underline{S} \geq S]$ bij the formula

$$P_n[\underline{S} \geq S] = 1-F(n,s-1) = F(n,\binom{n}{2}-s), \tag{3.2}$$

together with (3.1)

Now, consider the case that the two rankings contain only n-1 elements. The original rankings are obtained by adding the element n at the end of the first ranking and inserting n somewhere in the second one. If the element is introduced into the second ranking before the first element, the score is unaltered. If it is inserted just after the k'th element, the score is increased by k.

Hence,

$$u(n,s) = u(n-1,s) + u(n-1,s-1) + \ldots u(n-1,s-n+1),$$

and by summation

$$v(n,s) = v(n-1,s) + v(n-1,s-1) + \ldots . v(n-1,s-n+1),$$

whence

$$v(n,s) - v(n,s-1) = v(n-1,s) - v(n-1,s-n)$$

or

$$F(n,s) - F(n,s-1) = \frac{F(n-1,s) - F(n-1,s-n)}{n} . \tag{3.3}$$

with

$$F(n,s) = 0 \quad \text{if} \quad s < 0, \tag{3.4}$$

$$F(1,s) = 1 \quad \text{if} \quad s \geq 0. \tag{3.5}$$

Equation (3.3) lends itself very well to numerical computation. It is selfchecking as $F(n,s)= 1$ for $s = \binom{n}{2}$, and in general the symmetry, expressed by (3.2) provides checks for much lower values of s already. Equation (3.3), moreover, expresses a simple and nice geometrical property of $F(n,s)$-"curves" for two consecutive values of n.

Tables I, II and III were computed by means of (3.3), and only some rearrangement had to be done in order to get them into the final form. The number of decimals used in the calculations was seven, gradually decreasing to five with increasing n. The results were finally rounded off to three decimals in Tables I and II, but Table III is, of course, computed on the basis of the non-rounded off values.

For larger values of n, the error in replacing the exact probability distribution by a normal one, applying a correction for continuity, is at most about one unit of the third decimal. The formulae to be used are

$$\sigma^2 = \sigma_n^2(\underline{S}) = \frac{1}{18} n(n-1)(2n+5), \qquad (3.6)$$

$$P(x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-\frac{1}{2}t^2} dt, \qquad (3.7)$$

$$P_n[\underline{S} \geqslant S] = \frac{1}{2}\left\{1 - P\left(\frac{S-1}{\sigma}\right)\right\} . \qquad (3.8)$$

$P(x)$ is to be found, for instance, in the Tables of Probability Functions, Vol.II, National Bureau of Standards. Table IV gives, for the sake of convenience, σ for $n = 40(1)100$.

TABLE I

Values of $10^3 P_n[\underline{S} \geqslant S]$ for

$n \equiv 0,1 \pmod 4$ and even S.

Values of $10^3 P_n[\underline{S} \geq S]$

n	S=0	2	4	6	8	10	12	14	16	18
4	625	375	167	42	0	0	0	0	0	0
5	592	408	242	117	42	8	0	0	0	0
8	548	452	360	274	199	138	89	54	31	16
9	540	460	381	306	238	179	130	90	60	38
12	527	473	420	369	319	273	230	190	155	125
13	524	476	429	383	338	295	255	218	184	153
17	516	484	452	420	388	358	328	299	271	245
20	513	487	462	436	411	387	362	339	315	293
21	512	488	464	441	417	394	371	349	327	306
24	510	490	471	451	432	413	394	375	356	338
25	509	491	472	454	436	418	400	382	364	347
28	508	492	477	461	446	430	415	400	385	370
29	507	493	478	463	448	434	419	405	390	376
32	506	494	481	468	455	442	430	417	405	392
33	506	494	482	469	457	445	433	421	409	397
36	505	495	484	473	462	452	441	430	420	409
37	505	495	484	474	464	453	443	433	423	413
40	505	495	486	477	468	459	449	440	431	422

n	20	22	24	26	28	30	32	34	36	38
8	7	2	1	000	000	0	0	0	0	0
9	22	12	6	3	1	000	000	000	000	0
12	98	76	58	43	31	22	16	10	7	4
13	126	102	82	64	50	38	29	21	15	11
16	199	175	153	133	114	97	83	70	58	48
17	220	196	174	154	135	118	102	88	76	64
20	271	250	230	211	193	176	159	144	130	117
21	285	265	246	228	210	193	177	162	147	134
24	320	303	286	270	254	238	223	209	195	181
25	330	314	297	282	266	251	237	222	209	196
28	355	341	326	312	298	285	272	259	246	234
29	362	348	334	321	308	295	282	270	257	246
32	380	368	356	344	332	320	309	298	287	276
33	385	373	362	350	339	328	317	306	295	285
36	399	388	378	368	358	347	338	328	318	308
37	403	393	383	373	363	353	344	334	325	315
40	413	404	395	386	377	369	360	351	343	334

n	40	42	44	46	48	50	52	54	56	58
12	3	2	1	000	000	000	000	000	000	000
13	7	5	3	2	1	1	000	000	000	000
16	39	32	26	21	16	13	10	8	6	4
17	54	46	38	32	26	21	17	14	11	9
20	104	93	82	73	64	56	49	43	37	32
21	121	109	98	88	79	70	62	55	49	43
24	169	156	145	134	123	113	104	95	87	79
25	183	171	159	148	138	128	118	109	101	93
28	222	211	200	189	178	168	158	149	140	131
29	234	223	212	201	191	181	171	162	153	144
32	265	255	244	234	224	215	206	197	188	179
33	274	264	254	244	235	225	216	207	199	190
36	299	290	280	271	262	254	245	237	228	220
37	306	297	288	279	271	262	254	245	237	229
40	326	318	309	301	293	285	277	270	262	255

n	60	62	64	66	68	70	72	74	76	78
16	3	2	2	1	1	1	000	000	000	000
17	7	5	4	3	2	2	1	1	1	000
20	27	23	20	17	14	12	10	8	7	6
21	37	32	28	24	21	18	15	13	11	9
24	72	66	59	54	48	44	39	35	31	28
25	85	78	71	65	59	54	49	44	40	36
28	123	115	108	101	94	87	81	75	70	65
29	136	128	120	112	105	99	92	86	80	75
32	171	163	155	147	140	133	126	119	113	107
33	182	174	166	158	151	144	137	130	124	117
36	212	204	197	189	182	175	168	161	155	148
37	222	214	206	199	192	185	178	171	165	158
40	247	240	233	226	219	212	205	199	192	186

n	80	82	84	86	88	90	92	94	96	98
20	5	4	3	2	2	2	1	1	1	1
21	8	7	5	5	4	3	2	2	2	1
24	25	22	19	17	15	13	11	10	9	7
25	32	29	26	23	21	18	16	14	13	11
28	60	55	51	47	43	39	36	33	30	27
29	70	65	60	56	52	48	44	41	37	34
32	101	95	90	85	80	75	70	66	62	58
33	111	106	100	95	90	85	80	75	71	67
36	142	136	130	124	119	114	108	103	99	94
37	152	146	140	134	129	123	118	113	108	103
40	180	174	168	162	156	151	146	140	135	130

n	100	102	104	106	108	110	112	114	116	118
21	1	1	1	1	000	000	000	000	000	000
24	6	6	5	4	3	3	3	2	2	1
25	10	9	8	7	6	5	4	4	3	3
28	25	23	21	19	17	15	14	12	11	10
29	31	29	26	24	22	20	18	17	15	14
32	54	51	48	44	41	39	36	33	31	29
33	63	59	55	52	49	46	43	40	37	35
36	89	85	81	77	73	69	66	62	59	56
37	98	94	90	85	81	77	74	70	67	63
40	125	121	116	111	107	103	99	95	91	87

n	120	122	124	126	128	130	132	134	136	138
24	1	1	1	1	1	000	000	000	000	000
25	2	2	2	1	1	1	1	1	1	000
28	9	8	7	6	6	5	4	4	3	3
29	12	11	10	9	8	7	7	6	5	5
32	27	25	23	21	19	18	16	15	14	13
33	32	30	28	26	24	23	21	19	18	17
36	53	50	47	44	42	39	37	35	33	31
37	60	57	54	51	48	46	43	41	39	37
40	83	80	76	73	70	67	64	61	58	55

n	140	142	144	146	148	150	152	154	156	158
28	3	2	2	2	2	1	1	1	1	1
29	4	4	3	3	3	2	2	2	2	1
32	12	11	10	9	8	7	7	6	6	5
33	15	14	13	12	11	10	9	8	8	7
36	29	27	25	24	22	21	20	18	17	16
37	34	32	31	29	27	25	24	22	21	20
40	53	50	48	46	43	41	39	37	35	34

n	160	162	164	166	168	170	172	174	176	178
28	1	1	000	000	000	000	000	000	000	000
29	1	1	1	1	1	1	1	0	000	000
32	5	4	4	3	3	3	2	2	2	2
33	6	6	5	5	4	4	4	3	3	3
36	15	14	13	12	11	10	10	9	8	8
37	18	17	16	15	14	13	12	11	11	10
40	32	30	29	27	26	24	23	22	20	19

n	180	182	184	186	188	190	192	194	196	198
32	2	1	1	1	1	1	1	1	1	1
33	2	2	2	2	2	1	1	1	1	1
36	7	6	6	6	5	5	4	4	4	3
37	9	9	8	7	7	6	6	5	5	5
40	18	17	16	15	14	14	13	12	11	11

n	200	202	204	206	208	210	212	214	216	218
33	1	1	1	1	1	000	000	000	000	000
36	3	3	3	2	2	2	2	2	1	1
37	4	4	4	3	3	3	3	2	2	2
40	10	9	9	8	8	7	7	6	6	5

n	220	222	224	226	228	230	232	234	236	238
36	1	1	1	1	1	1	1	1	1	000
37	2	2	2	1	1	1	1	1	1	1
40	5	5	4	4	4	4	3	3	3	3

n	240	242	244	246	248	250	252	254	256	258
37	1	1	1	1	000	000	000	000	000	000
40	2	2	2	2	2	2	2	1	1	1

n	260	262	264	266	268	270	272	274	276	278
40	1	1	1	1	1	1	1	1	1	1

TABLE II

Values of $10^3\, P_n[\underline{S} \geqslant S]$ for

$n \equiv 2, 3 \pmod 4$ and odd S.

Values of $10^3 P_n[\underline{S} \geq s]$

n	S=1	3	5	7	9	11	13	15	17	19
6	500	360	235	136	68	28	8	1	0	0
7	500	386	281	191	119	68	35	15	5	1
10	500	431	364	300	242	190	146	108	78	54
11	500	440	381	324	271	223	179	141	109	82
14	500	457	415	374	334	295	259	225	194	165
15	500	461	423	385	349	313	279	248	218	190
18	500	470	441	411	383	354	327	300	275	250
19	500	473	445	418	391	365	339	314	290	267
22	500	478	456	434	412	390	369	348	328	308
23	500	479	458	438	417	397	377	357	338	319
26	500	483	465	448	431	414	397	380	363	347
27	500	484	467	451	434	418	402	386	371	355
30	500	486	472	458	444	430	416	402	389	375
31	500	487	473	460	446	433	420	407	394	381
34	500	488	477	465	453	442	430	418	407	396
35	500	489	478	466	455	444	433	422	411	400
38	500	490	480	470	460	450	440	431	421	411
39	500	490	481	472	462	452	443	433	424	414

n	21	23	25	27	29	31	33	35	37	39
10	36	23	14	8	5	2	1	000	000	000
11	60	43	30	20	13	8	5	3	2	1
14	140	117	96	79	63	50	40	31	24	18
15	164	141	120	101	84	70	57	46	37	29
18	227	205	184	165	147	130	115	100	88	76
19	245	223	203	184	166	149	133	119	105	93
22	289	270	252	234	217	201	186	171	157	144
23	301	283	265	248	232	216	201	187	173	160
26	331	316	300	285	270	256	242	229	216	203
27	340	325	310	296	281	268	254	241	228	216
30	362	349	336	323	310	298	286	274	262	251
31	368	355	343	331	318	306	295	283	272	261
34	384	373	362	351	340	329	319	308	298	288
35	389	378	368	357	347	336	326	316	306	296
38	401	392	382	373	363	354	345	336	327	318
39	405	396	387	377	368	359	350	341	333	324

n	41	43	45	47	49	51	53	55	57	59
14	13	10	7	5	3	2	2	1	1	000
15	23	18	14	10	8	6	4	3	2	1
18	66	56	48	41	34	29	24	20	16	13
19	82	72	62	54	47	40	34	29	25	21
22	131	120	109	99	89	80	72	64	58	51
23	147	135	124	114	104	94	86	78	70	63
26	191	179	168	157	147	137	127	118	110	102
27	204	192	181	170	160	150	141	132	123	115
30	239	228	218	208	198	188	178	169	160	152
31	250	239	229	219	209	199	190	181	172	164
34	278	268	259	249	240	231	222	213	205	196
35	286	277	267	258	249	240	232	223	215	206
38	309	300	291	283	274	266	258	250	242	234
39	315	307	298	290	282	274	266	258	250	243

n	61	63	65	67	69	71	73	75	77	79
15	1	1	000	000	000	000	000	000	000	000
18	11	9	7	5	4	3	3	2	1	1
19	17	14	12	10	8	6	5	4	3	3
22	45	40	35	31	27	24	21	18	15	13
23	57	51	46	41	36	32	28	25	22	19
26	94	87	80	73	67	62	57	52	47	43
27	107	99	92	85	79	73	67	62	57	52
30	144	136	128	121	114	107	100	94	88	83
31	155	147	140	132	125	118	112	105	99	93
34	188	180	173	165	158	151	144	137	131	125
35	198	191	183	176	168	161	154	148	141	135
38	227	219	212	205	198	191	184	177	171	165
39	235	228	221	214	207	200	193	187	180	174

n	81	83	85	87	89	91	93	95	97	99
18	1	1	000	000	000	000	000	000	000	000
19	2	2	1	1	1	1	000	000	000	000
22	11	10	8	7	6	5	4	3	3	2
23	17	15	13	11	9	8	7	6	5	4
26	39	35	32	29	26	23	21	19	17	15
27	48	44	40	36	33	30	27	25	22	20
30	77	72	67	63	59	54	51	47	43	40
31	88	82	77	72	68	63	59	55	52	48
34	119	113	107	102	97	92	87	82	78	74
35	129	123	117	112	107	101	96	92	87	83
38	158	152	147	141	135	130	125	120	115	110
39	168	162	156	150	145	139	134	129	124	119

n	101	103	105	107	109	111	113	115	117	119
22	2	2	1	1	1	1	1	000	000	000
23	4	3	3	2	2	1	1	1	1	1
26	13	12	10	9	8	7	6	5	5	4
27	18	16	15	13	12	10	9	8	7	6
30	37	34	32	29	27	25	23	21	19	17
31	45	41	38	36	33	31	28	26	24	22
34	70	66	62	58	55	52	49	46	43	40
35	78	74	70	66	63	59	56	53	50	47
38	105	101	96	92	88	84	80	76	73	69
39	114	109	105	101	96	92	88	84	81	77

n	121	123	125	127	129	131	133	135	137	139
26	4	3	3	2	2	2	1	1	1	1
27	6	5	4	4	3	3	3	2	2	2
30	16	14	13	12	11	10	9	8	7	6
31	20	19	17	16	14	13	12	11	10	9
34	38	35	33	31	29	27	25	23	22	20
35	44	42	39	37	34	32	30	28	26	25
38	66	63	60	57	54	51	49	46	44	41
39	74	70	67	64	61	58	55	53	50	48

n	141	143	145	147	149	151	153	155	157	159
26	1	1	1	000	000	000	000	000	000	000
27	1	1	1	1	1	1	1	000	000	000
30	6	5	5	4	4	3	3	3	2	2
31	8	7	7	6	6	5	4	4	4	3
34	19	17	16	15	14	13	12	11	10	9
35	23	22	20	19	17	16	15	14	13	12
38	39	37	35	33	31	29	28	26	25	23
39	45	43	41	39	37	35	33	31	29	28

n	161	163	165	167	169	171	173	175	177	179
30	2	2	1	1	1	1	1	1	1	1
31	3	3	2	2	2	2	1	1	1	1
34	8	8	7	7	6	5	5	5	4	4
35	11	10	10	9	8	7	7	6	6	5
38	22	21	19	18	17	16	15	14	13	12
39	26	25	23	22	21	20	18	17	16	15

n	181	183	185	187	189	191	193	195	197	199
31	1	1	1	1	1	000	000	000	000	000
34	3	3	3	3	2	2	2	2	2	1
35	5	5	4	4	3	3	3	3	2	2
38	11	11	10	9	9	8	8	7	7	6
39	14	14	13	12	11	10	10	9	9	8

n	201	203	205	207	209	211	213	215	217	219
34	1	1	1	1	1	1	1	1	1	000
35	2	2	2	1	1	1	1	1	1	1
38	6	5	5	4	4	4	4	3	3	3
39	7	7	6	6	6	5	5	4	4	4

n	221	223	225	227	229	231	233	235	237	239
35	1	1	1	1	000	000	000	000	000	000
38	3	2	2	2	2	2	2	1	1	1
39	4	3	3	3	3	2	2	2	2	2

n	241	243	245	247	249	251	253	255	257	259
38	1	1	1	1	1	1	1	1	1	000
39	2	1	1	1	1	1	1	1	1	1

n	261	263	265	267	269	271	273	275	277	279
39	1	1	1	1	000	000	000	000	000	000

TABLE III

Smallest value of S for which $P[S \geq S_*] \leq \alpha$.

Smallest value of S for which P[$\underline{S} \geq S$] $\leq \alpha$.

n	$\alpha = 0.005$	$\alpha = 0.010$	$\alpha = 0.025$	$\alpha = 0.050$	$\alpha = 0.100$
4	8	8	8	6	6
5	12	10	10	8	8
6	15	13	13	11	9
7	19	17	15	13	11
8	22	20	18	16	12
9	26	24	20	18	14
10	29	27	23	21	17
11	33	31	27	23	19
12	38	36	30	26	20
13	44	40	34	28	24
14	47	43	37	33	25
15	53	49	41	35	29
16	58	52	46	38	30
17	64	58	50	42	34
18	69	63	53	45	37
19	75	67	57	49	39
20	80	72	62	52	42
21	86	78	66	56	44
22	91	83	71	61	47
23	99	89	75	65	51
24	104	94	80	68	54
25	110	100	86	72	58
26	117	107	91	77	61
27	125	113	95	81	63
28	130	118	100	86	68
29	138	126	106	90	70
30	145	131	111	95	75
31	151	137	117	99	77
32	160	144	122	104	82
33	166	152	128	108	86
34	175	157	133	113	89
35	181	165	139	117	93
36	190	172	146	122	96
37	198	178	152	128	100
38	205	185	157	133	105
39	213	193	163	139	109
40	222	200	170	144	112

TABLE IV

Values of $\sigma = \left\{ \frac{n(n-1)(2n+5)}{18} \right\}^{\frac{1}{2}}$ for n = 40(1)100.

n	σ_n	n	σ_n
40	85,82929	71	201,46546
41	89,03183	72	205,70853
42	92,27315	73	209,98095
43	95,55278	74	214,28252
44	98,87029	75	218,61305
45	102,22524	76	222,97235
46	105,61723	77	227,36021
47	109,04586	78	231,77647
48	112,51074	79	236,22094
49	116,01149	80	240,69344
50	119,54776	81	245,19380
51	123,11918	82	249,72185
52	126,72542	83	254,27741
53	130,36615	84	258,86032
54	134,04104	85	263,47043
55	137,74977	86	268,10757
56	141,49205	87	272,77158
57	145,26757	88	277,46231
58	149,07604	89	282,17961
59	152,91719	90	286,92333
60	156,79073	91	291,69333
61	160,69640	92	296,48946
62	164,63394	93	301,31158
63	168,60308	94	306,15954
64	172,60359	95	311,03322
65	176,63522	96	315,93248
66	180,69772	97	320,85719
67	184,79087	98	325,80720
68	188,91444	99	330,78241
69	193,06821	100	335,78267
70	197,25195		