TABLE VI—*Sex distribution of responders giving various specialties as first choice of career*

First choice of career	Men	Women	Sex not stated	% women (of responders of known sex)
Medicine	347	98	9	22·0
Surgery	290	21	10	6·8
Obstetrics and gynaecology	56	22		28·2
General practice	433	226	6	34·3
General practice with other specialty	20	9		31·0
Psychiatry	49	24	1	32·9
Community medicine	3	6	1	66·7
Pathology	27	24	2	47·1
Anaesthetics	62	24		27·9
Radiology	20	11	1	35·5
Other (medical)	47	17		26·6
Paediatrics	68	59	2	46·5
Other (non-medical)	4	4		50·0
Not indicated	16	3		15·8
Total	1442	548	32	27·5

TABLE VIII—*Percentages of responders estimating various chances of success in different specialties*

	Certain	>75%	<75%	<25%	Total number giving specialty as first choice
Medicine	7	32	46	9	454
Paediatrics	5	38	40	13	129
Surgery	7	40	41	10	321
Obstetrics and gynaecology	4	38	37	17	78
General practice	55	40	2	1	665
Psychiatry	20	62	11	3	74
Community medicine	30	20	40	10	10
Pathology	22	43	23	9	53
Anaesthetics	20	71	9	0	86
Radiology and radiotherapy	9	72	19	0	32

of the responders rated their success as certain, and 40·5% thought that they had a better than 75% chance, 24·3% less than a 75% chance, and 6·7% less than a 25% chance. Over half of the responders who gave general practice as their first choice estimated their chances of success as certain (table VIII) compared with less than 10% of those preferring medicine, surgery, obstetrics, and radiology. Despite the large number of responders giving paediatrics as their first choice only 5% felt sure of success.

There was some evidence of variation between medical schools: among Manchester graduates only eight out of 21 preferring medicine considered their chances of success to be certain or above 75%, and likewise only four out of 18 preferring surgery. The comparable estimates of career prospects from King's College Hospital graduates were 11 out of 17 and 5 out of 7 for medicine and surgery respectively. In the case of general practice, 48 out of 52 Manchester graduates felt that they had a certain or better than 75% assured career, and 19 out of 20 King's College graduates felt the same.

Conclusion

The response rate to this inquiry was encouragingly high, but too many conclusions should not be drawn from a single sample of data. We are currently conducting a similar survey of doctors who graduated in 1975 and hope that the high rate of response will be maintained. Further studies will also be made, and we intend to follow up the careers of our original responders to see how their progress relates to their original preferences and intentions.

We are grateful to the deans and all those in the medical schools who supplied lists of names and addresses. We appreciate their interest in the project.

References

[1] McLaughlin, C, and Parkhouse, J, *Lancet*, 1972, **2**, 1018.
[2] McLaughlin, C, and Parkhouse, J, *Lancet*, 1974, **1**, 870.
[3] Parkhouse, J, and McLaughlin, C, *Lancet*, 1975, **1**, 1342.
[4] Royal Commission on Medical Education, 1965-8, *Report*, Cmnd 3569. London, HMSO, 1968.

Statistics at Square One

XVII—Some non-parametric tests

T D V SWINSCOW

British Medical Journal, 1976, **2**, 632-634

Values such as the mean and standard deviation that characterise a normal distribution are known as its parameters. Some data that we require to analyse do not conform to a normal distribution. Tests have been devised specially for situations such as this where we do not wish to make any assumption about the nature of the distribution. They are known as non-parametric tests. However, they can be applied also to normally distributed data.

British Medical Journal
T D V SWINSCOW, MSC, MB, deputy editor

Wilcoxon's rank sum tests

Wilcoxon's rank sum tests[1] are procedures of this kind. Examples of these tests follow, first on paired data (the signed rank test), then on unpaired data (the two-sample test). The Mann-Whitney U test has a similar approach to the problem and gives entirely equivalent results to the Wilcoxon two-sample test.

Dr Rosemary Mauve, a senior registrar in the rheumatology clinic of a district hospital, is studying the possible benefits of a new drug for rheumatoid arthritis. Among other properties its effect on the course of the disease is measured by the concentration of a globulin fraction in the plasma. Twenty patients have been selected for trial of the new drug (treatment B) against a standard form of treatment (treatment A).

Owing to the nature of the disease and its response to treatment

a cross-over trial on the same patients is thought to be unsuitable, so Dr Mauve divides the patients into two groups of 10 pairs, each member of a pair matching the other so far as possible in severity of the disease and prognosis. Again, as in the paired *t* test (Part XIII), it is essential to construct the pairs before treatment is given, because the choice of pairs must be uninfluenced by the effects of treatment. The intention of treatment is to lower the plasma concentration of this globulin. The reductions of concentration that treatment is known to cause are believed not to conform to a normal distribution. The actual results are set out in table 17.1.

TABLE 17.1—*Wilcoxon test of significance of difference between results (on a plasma globulin fraction) of treating two groups of patients in matched pairs*

Pairs of patients (1)	Globulin fraction, g/l		Difference (4)	Rank (5)	Signed rank (6)
	Treatment A (2)	Treatment B (3)			
1	38	45	−7	6½	−6½
2	26	28	−2	2½	−2½
3	29	27	2	2½	2½
4	41	38	3	4	4
5	36	40	−4	5	−5
6	31	42	−11	9	−9
7	32	39	−7	6½	−6½
8	30	39	−9	8	−8
9	35	34	1	1	1
10	33	45	−12	10	−10

Totals: plus ranks, 7½; minus ranks, 47½.

In columns (2) and (3) of table 17.1 are listed the plasma concentrations of globulin fraction after treatment A and treatment B respectively. The difference between each pair is shown in col (4). The differences are then ranked in order in col (5), the smallest being given rank 1. When two or more differences are identical they are each allotted the half-way point between the ranks they would fill if distinct. This is done irrespective of the plus or minus sign. For instance, the differences of −2 (patient 2) and +2 (patient 3) fill ranks 2 and 3. Since (2+3)/2=2½ they are allotted rank 2½. Had there been three differences of size 2, irrespective of sign, they would have filled in this particular series of results ranks 2, 3, and 4. They would all have been allotted the halfway rank, (2+4)/2=3. And so on however many values are identical. The ranks are listed in col (5), and a useful check is that they must add up to the same total as n(n+1)/2 patients; in this case 10(10+1)/2=55. In column (6) the ranks are repeated from column (5), but to each is now added the sign of the difference from column (4).

The numbers representing the plus ranks and the minus ranks in column (6) are now added up separately, and only the smaller of the two totals is used. Irrespective of its sign it is referred to table 17.2 against the number of pairs used in the investigation. *Larger* rank totals than those in the table are *non*-significant at the levels of probability shown.

In this case Dr Mauve has 10 pairs of patients and the smaller rank total is 7½. This is *smaller* than the figure 8 against 10 pairs in table 17.2. So the result is just *significant* at the 5% level. It is larger than the figure 3 shown for this number of pairs at the 1% level, so does not reach that level of significance.

TABLE 17.2—*Wilcoxon test on paired samples: 5% and 1% levels of P*

Number of pairs	5% Level	1% Level
7	2	0
8	2	0
9	6	2
10	8	3
11	11	5
12	14	7
13	17	10
14	21	13
15	25	16
16	30	19

Reprinted (slightly abbreviated) by permission of the publisher from *Statistical Methods*, by George W Snedecor and William G Cochran, 6th edition (copyright), 1967, Iowa State University Press, Ames, Iowa, USA.

Unpaired samples

A similar test developed by Wilcoxon is applicable to unpaired samples, and they need not be of equal size. As an example we can take Dr Grey's data and treat them as though they had not been paired.

The observations in the two samples are combined into a single series and ranked in order; but, in the ranking, the figures from one sample must be distinguished from those of the other

TABLE 17.3—*Wilcoxon test of significance of difference between results (on a plasma globulin fraction) of treating two unpaired groups of patients, figures for sample B in underlined type*

Globulin fraction g/l	Rank	Globulin fraction g/l	Rank
26	1	36	11
27	2	38	12½
28	3	**38**	12½
29	4	**39**	14½
30	5	**39**	14½
31	6	**40**	16
32	7	41	17
33	8	**42**	18
34	9	**45**	19½
35	10	**45**	19½

Totals of ranks: sample A, 81·5; sample B, 128·5.

TABLE 17.4—*Wilcoxon test on unpaired samples: 5% and 1% levels of P*

5% Critical points of rank sums

n_2 ↓ n_1 →	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4			10											
5		6	11	17										
6		7	12	18	26									
7		7	13	20	27	36								
8	3	8	14	21	29	38	49							
9	3	8	15	22	31	40	51	63						
10	3	9	15	23	32	42	53	65	78					
11	4	9	16	24	34	44	55	68	81	96				
12	4	10	17	26	35	46	58	71	85	99	115			
13	4	10	18	27	37	48	60	73	88	103	119	137		
14	4	11	19	28	38	50	63	76	91	106	123	141	160	
15	4	11	20	29	40	52	65	79	94	110	127	145	164	185
16	4	12	21	31	42	54	67	82	97	114	131	150	169	
17	5	12	21	32	43	56	70	84	100	117	135	154		
18	5	13	22	33	45	58	72	87	103	121	139			
19	5	13	23	34	46	60	74	90	107	124				
20	5	14	24	35	48	62	77	93	110					
21	6	14	25	37	50	64	79	95						
22	6	15	26	38	51	66	82							
23	6	15	27	39	53	68								
24	6	16	28	40	55									
25	6	16	28	42										
26	7	17	29											
27	7	17												
28	7													

1% Critical points of rank sums

n_2 ↓ n_1 →	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5				15										
6			10	16	23									
7			10	17	24	32								
8			11	17	25	34	43							
9		6	11	18	26	35	45	56						
10		6	12	19	27	37	47	58	71					
11		6	12	20	28	38	49	61	74	87				
12		7	13	21	30	40	51	63	76	90	106			
13		7	14	22	31	41	53	65	79	93	109	125		
14		7	14	22	32	43	54	67	81	96	112	129	147	
15		8	15	23	33	44	56	70	84	99	115	133	151	171
16		8	15	24	34	46	58	72	86	102	119	137	155	
17		8	16	25	36	47	60	74	89	105	122	140		
18		8	16	26	37	49	62	76	92	108	125			
19	3	9	17	27	38	50	64	78	94	111				
20	3	9	18	28	39	52	66	81	97					
21	3	9	18	29	40	53	68	83						
22	3	10	19	29	42	55	70							
23	3	10	19	30	43	57								
24	3	10	20	31	44									
25	3	11	20	32										
26	3	11	21											
27	4	11												
28	4													

n_1 and n_2 are the numbers of cases in the two groups. If the groups are unequal in size, n_1 refers to the smaller.

Reproduced by permission of the author and publisher from White, C, *Biometrics*, 1952, **8**, 33.

sample—for instance, by underlining. The data now appear as in table 17.3. To save space they have been set out in two columns, but a single ranking is done. The figures for sample B are in underlined type. Again the sum of the ranks is $n(n+1)/2$.

The ranks for the two samples are now added separately, and the *smaller* total is used. It is referred to table 17.4, with n_1 equal to the number of observations in one sample and n_2 to the number of observations in the other sample. In this case they both equal 10. At $n_1=10$ and $n_2=10$ the upper part of the table shows the figure 78. The smaller total of Dr Mauve's ranks is 81·5. Since this is slightly *larger* than 78 it does *not* reach the 5% level of probability. The result is therefore not significant at that level. In the lower part of table 17·4, which gives the figures for the 1% level of probability, the figure for $n_1=10$ and $n_2=10$ is 71. As expected, Dr Mauve's result is further from that than from the 5% figure of 78.

It is worth noting that the same data when paired produce a just significant result at the 5% level and when unpaired a just non-significant result. This disparity illustrates the value of pairing, when it can appropriately be done, in disclosing the existence of a difference that is obscured by lack of matching in the sources of the data. Pairing leads to greater sensitivity of the test because it eliminates some confusing variables.

The advantages of these tests based on ranking are that they can be safely used on data that are not at all normally distributed. Moreover, they are quick to carry out, and no calculator is needed. Their disadvantages are, firstly, that they are generally not so sensitive as a *t* test, since they do not use so much information. Non-normally distributed data can sometimes be transformed by the use of logarithms or some other method to make them normally distributed, and a *t* test performed on them then. Consequently the best procedure to adopt may require careful thought. Secondly, the extent and nature of the difference between two samples is often brought out more clearly by standard deviations and *t* tests than by non-parametric tests. The latter are simply tests at preselected levels of probability.

Exercise 17. A new treatment in the form of tablets for the prophylaxis of migraine has been introduced, to be taken before an impending attack. Twelve patients agree to try this remedy in addition to the usual general measures they take, subject to advice from their doctor on the taking of analgesics also. A crossover trial with identical placebo tablets is carried out over a period of eight months. The numbers of attacks experienced by each patient on, firstly, the new treatment and, secondly, the placebo were as follows: Patient (1), 4 and 2. Patient (2). 12 and 6. Patient (3), 6 and 6. Patient (4), 3 and 5. Patient (5), 15 and 9, Patient (6), 10 and 11. Patient (7), 2 and 4. Patient (8), 5 and 6. Patient (9), 11 and 3. Patient (10), 4 and 7. Patient (11), 6 and 0. Patient (12), 2 and 5. In a Wilcoxon rank sum test what is the smaller total of ranks? Is it significant at the 5% level? *Answer*: −30; no.

Another doctor carried out a similar pilot study with this preparation on 12 patients, giving the identical placebo to 10 other patients. The numbers of migraine attacks experienced by the patients over a period of six months were as follows. Group receiving new preparation: Patient (1), 8; (2), 6; (3), 0; (4), 3; (5), 14; (6), 5; (7), 11; (8), 2; (9), 1; (10), 9; (11), 12; (12), 4. Group receiving placebo: Patient (1), 7; (2), 10; (3), 4; (4), 12; (5), 2; (6), 8; (7), 8; (8), 6; (9), 0; (10), 5. In a Wilcoxon two-sample test what is the smaller total of ranks? Which sample of patients provides it? Is it significant at the 5% level? *Answer*: 116; the group on the placebo; no.

Reference

[1] Wilcoxon, F, *Biometrics Bulletin*, 1945, **1**, 80.

If poultry are fed antibiotics do the latter occur in the eggs?

The antibiotics that are fed to poultry as growth promoters are not used in man for medical purposes. They include such substances as virginiamycin and bacitracin and are added to the poultry rations at the rate of 10 g per 1000 kg of feed. At such levels there is little risk of the antibiotic occurring in the eggs. When antibiotics are fed at high levels for treatment (which is unusual) there could be a risk of drug residues in the eggs; but in this instance it is highly unlikely that the hens concerned would be laying any eggs, because of the nature of their illness.

Does progesterone have a place in treating premenstrual and postpartum depression?

Although several views are held many psychiatrists regard a progestogen or progesterone as essential in treating purely premenstrual depression. As progestogens are simpler to administer they are usually tried first. Progesterone may be given as suppositories or pessaries, or by intramuscular injection for 10 days before menstruation if other approaches fail.[1] As the symptoms and repercussions of premenstrual depression may be severe, many psychiatrists give an appropriate antidepressant drug continuously for several cycles. In established postpartum depressive illness, progesterone should be given either within 10 days of delivery (given that the illness started very early in the puerperium) or, as is often the case, if considerable premenstrual exacerbation of symptoms occurs once menstruation is re-established. In the former case a course of daily intramuscular injections for a week is often given. For the latter, progesterone is given during the 10 premenstrual days for three cycles. For women likely to be vulnerable to depression in the puerperium daily intramuscular injections of progesterone for a week post partum reduce their vulnerability and the risk of developing depressive illness.[2]

[1] Dalton, K, *The Premenstrual Syndrome*. London, Heinemann, 1964.
[2] Stafford-Clark, D, *Psychiatry for Students*, vol 1. London, George Allen and Unwin, 1974.

Should nuts, currants, raisins, tomatoes, pears, and bananas be allowed in a diet recommended for diverticulosis?

There is absolutely no contraindication to using these fruits in diverticulosis except in so far as they exclude more useful forms of "dietary fibre" such as bran. It should be remembered, however, that with the possible exception of tomatoes all are high in calories.

Could the use and subsequent ingestion of appreciable amounts of a proprietary denture-fixative powder over three to four years cause or contribute to diverticulosis?

Almost all the proprietary denture-fixative powders appear to be based on tragacanth. It would have no aetiological role in diverticulosis. In so far as this is a storage polysaccharide it would probably have some slight therapeutic effect in diverticulosis but unless the use of the powder is idiosyncratic the effect of eating it is likely to be trivial.

Are Toxoplasma gondii visible in a blood film?

No. *Toxoplasma* may produce significant blood changes which resemble those of glandular fever in some detail but the Paul-Bunnell test is negative. Less commonly a severe haemolytic anaemia may occur. In neither case can the parasite be demonstrated in blood. The acquired form of toxoplasmosis is usually seen in young adults, who acquire it from cats or other animals excreting the cysts in the faeces. It may be mild and often asymptomatic. In others, lymph nodes, particularly in the neck, are enlarged. A rising titre in the toxoplasma dye test is suggestive, and serum samples are required to carry out this test. The congenital form may occur if the mother has been infected during pregnancy and can produce hepatosplenomegaly, anaemia, jaundice, and rashes in the newborn. *Toxoplasma* is an obligatory intracellular parasite and rarely can be demonstrated in lymph nodes.