KEMORECEPCIJA U GUŠTERA

Uvod: Iako su se prva istraživanja komunikacije među gušterima uglavnom usmjerila na vidnu komunikaciju, posljednjih 20-tak godina istraživanje kemijske komunikacije među gušterima postalo je vrlo popularno. Osim u guštera, istraživanja kemijske komunikacije vrlo se intenzivno provode i na zmijama, ali i na mnogim sisavcima koji uglavnom "žive u mirisnom svijetu". Tome u prilog govori i činjenica da su "*Journal of Chemical Ecology*" i "*Chemical Senses*" u posljednjih 15 godina postali vrlo prestižni znanstveni časopisi.

Mnoge vrste guštera izbacuju jezik da bi njime ispitali okoliš. To izbacivanje jezika zovemo palucanje. U nekim slučajevima, jezik se kratkotrajno samo izbaci izvan usta i uvuče, dok u drugim jezik aktivno dodirne predmete. U oba slučaja, čestice koje se vežu za jezik unose se u vomeronazalni (ili Jacobsonov) organ. To je parni organ smješten u krovu tvrdog nepca. Posebice su u ispitivanju okoline jezikom aktivni gušteri porodice varana (*Varanidae*) i rovaša (*Scincidae*).

Istraživanja ponašanja ukazuju da izbacivanje jezika u gmazova može imati različite funkcije (Simon 1983; Halpern 1992). Zanimljivo je da postoji vrlo malo kvantitativnih istraživanja kemorecepcije u guštera. Ona je dobro proučena samo u nekoliko vrsta kao što su *Sceloporus jarrovi* (Defazio i sur. 1977; Simon i sur. 1981), *Dipsosaurus dorsalis* (Krekorian 1989; Cooper i Alberts 1990; Dussault i Krekorian 1991; Bealor i Krekorian 2002) i *Eumeces laticeps* (Cooper i Vitt 1984; Cooper 1990). Ova istraživanja pokazuju da izbacivanje jezika služi da bi se dobile različite informacije. Neke od funkcija sustava "jezik - vomeronazalni organ" su: Prepoznavanje primjeraka iste vrste, prepoznavanje drugih vrsta, prepoznavanje suprotnog spola, traženje hrane, traženje spolnog partnera, prepoznavanje predatora te opće ispitivanje okoliša.

Materijali: U ovoj vježbi koristit ćete 8 - 16 guštera vrste primorska gušterica (*Podarcis sicula*), terarije i grijalice. Gušteri se mogu označiti bojanjem na leđima. Kao podlogu u terarije treba staviti pijesak, šljunak ili papirnate ručnike. U terarije treba staviti jedan do dva manja kamena ili komada stijene. Za izvođenje vježbe potrebna je i štoperica.

Metode: Radi se u skupinama od 4 - 5 studenata. Vježbu treba izvoditi tako da se izbjegava glasni govor i nagli pokreti, kako se gušteri ne bi preplašili. Jedna osoba treba mjeriti latenciju prvog izbačaja jezika tj. vrijeme koje je prošlo dok jezik ne bude izbačen prvi put te broj izbačaja jezika tijekom pokusnog perioda. Na kraju se skupe svi podaci, kako bi se dobilo dovoljno veliki uzorak za statističku analizu.

Promatrajte guštere tijekom 10 minuta. Pokušajte uočiti koje sve predmete dodiruje jezikom, kada samo izbacuje jezik u zrak. Da li gušteri uopće dodiruju jezikom predmete iz okoliša? Da li izbacuju jezik dok miruju ili dok se kreću? Nakon ovog početnog razdoblja od 10 minuta, započnite pokus. Pokus se sastoji od tri dijela. U prvom dijelu mjerite temeljnu razinu palucanja jezikom, razinu palucanja tijekom tretmana te razinu palucanja tijekom kontrolnog perioda. Za svaki period odredite vrijeme latencije (vrijeme proteklo do prvog palucaja jezikom), broj palucaja jezikom kada je jezik samo izbačen u zrak te broj palucaja kada jezik dodiruje neki predmet. Mjerenje palucanja traje 15 minuta, a počinje prvim palucajem jezika. Neke će životinje vrlo sporo reagirati u ovom testu tj. imati će vrlo dugo vrijeme latencije. Za te životinje primjenjuje se tzv. pravilo maksimalne latencije od 10 minuta, a nakon toga počinje se mjerenje.

Temeljna razina palucanja jezikom - Izvadite guštera iz kaveza, držite ga u ruci minutu te ga vratite u kavez na *p r i b l i ž n o* isto mjesto sa kojeg ste ga uzeli (tko ovo uspije izvesti sa primorskom guštericom, svaka mu čast!). Odmah počnite mjeriti slijedeće varijable: latenciju palucanja, broj palucaja u zrak, broj palucaja sa dodirom predmeta te objekte koji su dodirnuti jezikom. Zapamtite: *15 - minutni preriod vremena počinje sa prvim palucajem!!!*

Tretman - Uzmite guštera iz njegovog kaveza te ga stavite u novi terarij. Odmah počnite mjerenja varijabli koje su već navedene.

Kontrola - nakon tretmana, vratite guštera u njegov kavez i mjerite gore navedene varijable. Svaka skupina radi pokus sa dva različita guštera tj. radi dvije replikacije.

Nježno s gušterima! Ne hvatajte guštera za rep, jer lako puca, a takav gušter neće dati dobre razultate u pokusu!

Nakon završenog pokusa, popunite tablicu na kraju ovog predloška.

Hipoteze i predviđanja: Opće hipoteza koju izvodimo iz ovog pokusa je da broj palucaja ovisi o tome koliko je okoliš poznat životinji. Ako je to točno, onda će broj palucaja u tretmanu biti veći ili manji nego broj palucaja u kontroli i temeljnoj razini palucaja. Ako je kemorecepcija temeljni način ispitivanja okoliša u guštera, kako će tretman utjecati na njihovo ponašanje?

Bilježenje i analiza podataka: Treba skupiti sve podatke u grupi da bi izračunali srednju vrijednost latencije te srednju vrijednost palucaja (za zrak i za dodirivanje predmeta) za sva tri slučaja. Nul hipoteza je da se vrijeme latencije ne razlikuje između tretmana, kontrole i temeljne razine palucanja jezikom. Alternativne hipoteze su da se vrijeme latencije razlikuje između tri slučaja. Rezultati se mogu analizirati Friedman-ovim testom za analizu varijance zavisnih uzoraka (engl. Friedman two-way ANOVA by ranks).

Literatura:

Bealor M. T., Krekorian C. O. (2002) Chemosensory identification of lizard-eating snakes in the desert iguana, *Dipsosaurus dorsalis* (Squamata: Iguanidae). *J. Herpetol.* **36:** 9 - 15

Cooper W. E. (1990) Chemical detection of predators by a lizard, the broad-headed skink, *Eumeces laticeps*. *J. Herpetol.* **25:** 126 - 127

Cooper W. E., Alberts A. C. (1990) Responses to chemical food stimuli by an herbivorous actively foraging lizard, *Dipsosaurus dorsalis*. *Herpetologica* **46:** 259 - 266

Cooper W. E., Alberts A. C. (1991) Tongue-flicking and biting in response to chemical food stimuli by an iguanid lizard (*Dipsosaurus dorsalis*) having sealed vomeronasal ducts: Vomerolfaction may mediate these chemical responses. *J. Chem. Ecol.* **17:** 135 - 145

Cooper W. E., Vitt L. J. (1984) Conspecific odor detection by the male broad-headed skink, *Eumeces laticeps*: Effects of sex and site of odor source and of male reproductive condition. *J. Exp. Zool.* **230:** 199 - 209

DeFazio A., Simon A. C., Middendorf G. A., Romano D. (1977) Iguanid substrate licking: A respose to novel situation by *Sceloporus jarrovi*. *Copeia* **1977:** 706 - 709

Dussault M. H., Krekorian C. O. (1991) Conspecific discrimination by chemoreception in the desert iguana, *Dipsosaurus dorsalis*. *Herpetologica* **47:** 82 - 88

Halpern M.: Nasal chemical senses in reptiles: Structure and function. In: “Biology of the Reptilia”, Vol. 18, (C. Gans, D. Crews, eds.), University of Chicago Press, Chicago, 1992, pp 423 - 523

Krekorian C. O. (1989) Field and laboratory observations on chemoreception in the desert iguana, *Dipsosaurus dorsalis*. *J. Herpetol.* **23:** 267 - 273

Pedersen J. M. (1988) Laboratory observations on the function of tongue extrusions in the desert iguana, *Dipsosaurus dorsalis*. *J. Comp. Psychol.* **102:** 1 - 4

Pedersen J. M. (1992) Field observations on the role of tongue extrusion in the social behavior of the desert iguana (*Dipsosaurus dorsalis*). *J. Comp. Psychol.* **106:** 287 - 294

Siegel S., Castellan N. J., Jr.: "Nonparametric Statistics for the Behavioral Science", 2nd ed., McGraw-Hill, New York, 1988

Simon C. A. A review of lizard chemoreception. In: "Lizard ecology: Studies of a model organism" (R. B. Huey, E. R. Pianka, T. W. Schoener, eds.), Harvard University Press, Cambridge, 1983, pp 119 - 133

Simon C. A., Gravelle K., Bissinger B. E., Eiss I., Ruibal R. (1981) The role of chemoreceptionin the iguanid lizard *Sceloporus jarrovi*. *Anim. Behav.* **29:** 46 - 54

Pitanja za raspravu:

1. Što je, prema vašim podacima, uloga palucanja jezikom?
2. Da li je spol ili veličina guštera čimbenik koji utječe na palucanje jezikom?
3. Koja je ovisna varijabla najbolja za analizu?
4. Da li su češći palucaji kojima se dodiruju predmeti ili oni kojima se samo "kuša" zrak?
5. Koje su sve promjene u ponašanju guštera kada ga se prebaci u terarij u kojem se testira? Da li se one mogu usporediti sa promjenama u ponašanju tijekom mjerenja temeljne razine palucanja i kontrolnog pokusa?
6. Da li bi se uočene promjene ponašanja mogle kvantificirati i koristiti kao ovisne varijable?

	Temeljna razina palucanja (predtest)			Tretman (pokus)			Kontrola (posttest)		
gušter	latencija (s)	izbačaji jezika		latencija (s)	izbačaji jezika		latencija (s)	izbačaji jezika	
		zrak	predmet		zrak	predmet		zrak	predmet
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									

7.2 THE FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE BY RANKS

7.2.1 Function

When the data from k matched samples are in at least an ordinal scale, the *Friedman two-way analysis of variance by ranks* is used to test the null hypothesis that the k samples have been drawn from the same population.

Since the k samples are matched, the number of cases N is the same in each of the samples. That matching may be achieved by studying the same group of subjects under each of the k conditions. Or the researcher may obtain N sets, each consisting of k matched subjects, and then randomly assign one subject in each set to the first condition, one subject in each set to the second condition, etc. For example, if one wished to study the differences in learning achieved under four teaching methods, one might obtain N sets of $k = 4$ pupils, each set consisting of children who are matched on the relevant variables (age, previous learning, intelligence, socioeconomic status, motivation, etc.), and then at random assign one child from each of the N sets to teaching method A, another from each set to method B, another from each set to C, and the fourth to D.

The Friedman two-way analysis of variance by ranks tests the null hypothesis that the k repeated measures or matched groups come from the same population or populations with the same median. To specify the null hypothesis more explicitly, let θ_j be the population median in the jth condition or group. Then we may write the null hypothesis that the medians are the same as H_0: $\theta_1 = \theta_2 = \cdots = \theta_k$. The alternative hypothesis is then H_1: $\theta_i \quad \theta_j$ for at least two conditions or groups i and j. That is, if the alternative hypothesis is true, at least one pair of conditions has different medians. Under the null hypothesis, the test assumes that the variables have the same underlying continuous distribution; thus it requires at least ordinal measurement of that variable.

7.2.2 Rationale and Method

For the Friedman test, the data are cast in a two-way table having N rows and k columns. The rows represent the subjects or matched sets of subjects, and the columns represent the various conditions. If the scores of subjects serving under all conditions are under study, then each row gives the scores of one subject under each of the k conditions.

The data of the test are ranks. The scores in each *row* are ranked separately. That is, with k conditions being studied, the ranks in any row range from 1 to k. The Friedman test determines the probability that the different *columns* of ranks (samples) come from the same population, that is, that the k variables have the same median.

For example, suppose we wish to study the scores of three groups under four conditions. Here $N = 3$ and $k = 4$. Each group contains four matched subjects, one being assigned to each of the four conditions. Suppose our scores for this study were those given in Table 7.2. To perform the Friedman test on those data, we first rank the scores *in each row*. We may give the lowest score in each row a rank of 1, the next lowest score in each row the rank of 2, etc. By doing this we obtain the data shown in Table 7.3. Observe that the ranks in each row of Table 7.3 range from 1 to $k = 4$.

Now if the null hypothesis (that all of the samples—columns—came from the same population) is in fact true, then the distribution of ranks *in each column* would be a matter of chance and, thus, we would expect the ranks of 1, 2, 3, and 4 to appear in each column with about equal frequency. That is, if the data were

TABLE 7.2
Scores of three matched groups under four conditions

	Conditions			
Groups	I	II	III	IV
A	9	4	1	7
B	6	5	2	8
C	9	1	2	6

TABLE 7.3
Ranks of three matched groups under four conditions

	Conditions			
Groups	I	II	III	IV
A	4	2	1	3
B	3	2	1	4
C	4	1	2	3
R_j	11	5	4	10

random, we would expect the sum of ranks in each column to be $N(k + 1)/2$. For the data in Table 7.3, the expected column sums would be $3(4 + 1)/2 = 7.5$. This indicates that for any group it is a matter of chance under which condition the highest score occurs and under which condition the lowest occurs—which would be the case if the conditions really did not differ.

If the subjects' scores were independent of the condition, the set of ranks in each column would represent a random sample from the discrete rectangular distribution of rank numbers 1, 2, 3, and 4, and the rank totals for the various columns would be about equal. If the subjects' scores were dependent on the conditions (i.e., if H_0 were false), then the rank totals would vary from one column to another. Inasmuch as the columns all contain an equal number of cases, an equivalent statement would be that under H_0 the average of the ranks in the various columns would be about equal.

The Friedman test determines whether the rank totals (denoted R_j) for each condition or variable differ significantly from the values which would be expected by chance. To do this test, we compute the value of the statistic which we shall denote as F_r.

$$F_r = \left[\frac{12}{Nk(k+1)} \sum_{j=1}^{k} R_j^2\right] - 3N(k+1) \qquad (7.3)$$

where N = number of rows (subjects)
k = number of columns (variables or conditions)
R_j = sum of ranks in the jth column (i.e., the sum of ranks for the jth variable)

and $\sum_{j=1}^{k}$ directs one to sum the squares of the sums of ranks over all conditions.

Probabilities associated with various values of F_r when H_0 is true have been tabulated for various sample sizes and various numbers of variables. Appendix Table M gives the probabilities associated with values of F_r as large or larger than the tabled values for various values of N and k. If the observed value of F_r is larger than the tabled value of F_r at the chosen significance level, then H_0 may be rejected in favor of H_1.

If the number of variables k exceeds five or the sample size N exceeds that for the tabled entries in Appendix Table M, then a large-sample approximation may be used. When the number of rows and/or columns is large, it can be shown that the statistic F_r as given in Eq. (7.3) is distributed approximately as χ^2 with $df = k - 1$. Thus, Appendix Table C may be used to determine the significance probability.

If the value of F_r as computed from Eq. (7.3) is equal to or larger than that given in Appendix Table M or Appendix Table C for a particular level of significance, then the implication is that the sums of ranks (or, equivalently, the average rank R_j/N) for the various columns differ significantly (which is to say that the size of the scores depends on the conditions under which the scores were obtained), and thus H_0 may be rejected at that level of significance.

To illustrate the computation of F_r and the use of Appendix Table M, we may test for significance of differences in the data shown in Table 7.3. By referring to that table, the reader may see that the number of conditions is $k = 4$ and the number of rows is $N = 3$. The successive sums of ranks R_j are 11, 5, 4, and 10 respectively. We may compute the value of F_r for the data in Table 7.3 by substituting these values into Eq. (7.3):

$$F_r = \left[\frac{12}{Nk(k+1)} \sum_{j=1}^{k} R_j^2\right] - 3N(k+1) \qquad (7.3)$$

$$= \frac{12}{(3)(4)(4+1)}(11^2 + 5^2 + 4^2 + 10^2) - (3)(3)(4+1)$$

$$= 7.4$$

We may determine the probability of occurrence when H_0 is true of $F_r \geq 7.4$ by turning to Appendix Table M which gives the selected critical values of observed F_r for $k = 4$. Reference to that table shows that the probability associated with $F_r \geq 6.5$ when $N = 3$ and $k = 4$ is $p \leq .05$. Thus, for these data, we may reject the null hypothesis that the four samples were drawn from a population with the same medians at the .05 level of significance since the observed value of F_r exceeds the critical value.

Example 7.2a For *N* and *k* large. In a study of the effect of three different patterns of reinforcement upon the extent of discrimination learning in rats[2] three matched samples ($k = 3$) of 18 rats ($N = 18$) were trained under three patterns of reinforcement. Matching was achieved by the use of the 18 sets of litter mates, 3 in each set. Although all of the 54 rats received the same quantity of reinforcement (reward), the patterning of the administration of reinforcement was different for each of the groups. One group was trained with 100 percent reinforcement (*RR*), a second matched group was trained with partial reinforcement in which each sequence of trials ended with an unreinforced trial (*RU*), and the third

[2] Grosslight, J. H. and Radlow, R. (1956). Patterning effect of the nonreinforcement-reinforcement sequence in a discrimination situation. *Journal of Comparative and Physiological Psychology*, **49**, 542–546.

matched group was trained with partial reinforcement in which each sequence of trials ended with a reinforced trial (*UR*).

After this training, the extent of learning was measured by the speed at which the various rats learned an "opposing" behavior—whereas they had been trained to run to white, the rats were now rewarded for running to black. The better the initial learning, the slower should be this transfer of learning. The prediction was that the different patterns of reinforcement used would result in differential learning as exhibited by ability to transfer.

i. *Null hypothesis.* H_0: the different patterns of reinforcement have no differential effect on the observed behavior. H_1: the different patterns of reinforcement have a differential effect.

ii. *Statistical test.* Because the number of errors in transfer of learning is probably not an interval measure of the strength of original learning, the Friedman two-way analysis of variance by ranks was chosen. Moreover, the use of the parametric analysis of variance is precluded because examination of the experimental situation suggested that one of the basic assumptions of the *F* test was probably untenable.

iii. *Significance level.* Let $\alpha = .05$ and *N* is the number of rats in each of the $k = 3$ matched groups = 18.

iv. *Sampling distribution.* As computed by Eq. (7.3) and because the sample size is large, F_r is distributed approximately as χ^2 with $df = k - 1$. Thus the probability associated with the occurrence under H_0 of a value as large as the observed value of F_r may be determined by reference to Appendix Table C.

v. *Rejection region.* The region of rejection consists of all values of F_r which are so large that the probability associated with their occurrence when H_0 is true is equal to or less than $\alpha = .05$.

vi. *Decision.* The number of errors committed by each rat in the transfer-of-learning situation was determined, and those scores were ranked for each of the 18 sets of 3 matched rats. These ranks are given in Table 7.4.

Observe that the sum of ranks for the *RR* groups is 39.5, the sum of ranks for the *RU* group is 42.5, and the sum or ranks for the *UR* groups is 26.0. A low rank signifies a high number of errors in transfer, i.e., strong original learning. We may compute the value of F_r by substituting our observed values into Eq. (7.3):

$$F_r = \left[\frac{12}{Nk(k+1)} \sum_{j=1}^{k} R_j^2\right] - 3N(k+1) \qquad (7.3)$$

$$= \frac{12}{(18)(3)(3+1)}(39.5^2 + 42.5^2 + 26^2) - (3)(18)(3+1)$$

$$= 8.58$$

Reference to Appendix Table C indicates that $F_r = 8.58$ when $df = k - 1 = 3 - 1 = 2$ is significant at between the .02 and .01 levels. Therefore, since $p < .02$ is less than our previously established significance level of $\alpha = .05$, the decision is to reject H_0. The conclusion is that the rats' scores on transfer of learning depend on the pattern of reinforcement in the original learning trials.

TIES When there are ties among the ranks for any given group (or row) the statistic F_r must be corrected to account for changes in the sampling distribution. Equation

TABLE 7.4
Ranks of 18 matched groups in transfer after training under three conditions of reinforcement

	Type of reinforcement		
Group	RR	RU	UR
1	1	3	2
2	2	3	1
3	1	3	2
4	1	2	3
5	3	1	2
6	2	3	1
7	3	2	1
8	1	3	2
9	3	1	2
10	3	1	2
11	2	3	1
12	2	3	1
13	3	2	1
14	2	3	1
15	2.5	2.5	1
16	3	2	1
17	3	2	1
18	2	3	1
R_j	39.5	42.5	26.0

(7.4) gives the value of F_r which is appropriate when ties occur. Although Eq. (7.4) can be used in general, that is, when there are no ties as well as when there are ties, the computation is somewhat more tedious.

$$F_r = \frac{12 \sum_{j=1}^{k} R_j^2 - 3N^2k(k+1)^2}{Nk(k+1) + \dfrac{\left(Nk - \sum_{i=1}^{N} \sum_{j=1}^{g_i} t_{i,j}^3\right)}{(k-1)}} \qquad (7.4)$$

where g_i is the number of sets of tied ranks in the *i*th group and $t_{i,j}$ is the size of the *j*th set of tied ranks in the *i*th group. We include the sets of size 1 in the count. As is the case with other corrections for tied data, the effect of tied ranks is to increase the size of the Friedman statistic F_r. If the correction for ties is made in the example given above, we note that there are two tied ranks in the 15th group. We note that there are 52 ties of size 1 and one tie of size 2. Therefore,

$$\sum_{i=1}^{N} \sum_{j=1}^{g_i} t_{i,j}^3 = 1 + 1 + 1 + \cdots + 1 + 8 + 1 + \cdots + 1$$

$$= 60$$

By using Eq. (7.4), we get $F_r = 8.70$, which is larger than the value (8.58) obtained without the correction. Obviously, since H_0 was rejected without the correction, it would be rejected with the correction as well. It should be remarked that in this example the effect of ties was very small; however, as the number of ties increases, the greater is the effect on F_r.

7.2.3 Multiple Comparisons between Groups or Conditions

When the obtained value of F_r is significant, it indicates that *at least one* of the conditions differs from *at least one* other condition. It does not tell the researcher which one is different, nor does it tell the researcher how many of the groups are different from each other. That is, when the obtained value of F_r is significant, we would like to test the hypothesis H_0: $\theta_u = \theta_v$ against the hypothesis H_1: $\theta_u \neq \theta_v$ for some conditions u and v. There is a simple procedure for determining which condition (or conditions) differ. We begin by determining the differences $|R_u - R_v|$ for all pairs of conditions or groups. When the sample size is large, these differences are approximately normally distributed. However, since there are a large number of differences and because the differences are not independent, the comparison procedure must be adjusted appropriately. Suppose the hypothesis of no difference between the k conditions or matched groups was tested and rejected at the α significance level. Then we can test the significance of individual pairs of differences by using the following inequality. That is, if

$$|R_u - R_v| \geq z_{\alpha/k(k-1)} \sqrt{\frac{Nk(k+1)}{6}} \tag{7.5a}$$

or if the data are expressed in terms of average ranks within each condition, and if

$$|\bar{R}_u - \bar{R}_v| \geq z_{\alpha/k(k-1)} \sqrt{\frac{k(k+1)}{6N}} \tag{7.5b}$$

then we may reject the hypothesis H_0: $\theta_u = \theta_v$ and conclude that $\theta_u \neq \theta_v$. Thus, if the difference between the rank sums (or average ranks) exceeds the corresponding critical value given in Eq. (7.5*a*) or (7.5*b*), then we may conclude that the two conditions are different. The value of $z_{\alpha/k(k-1)}$ is the abscissa value from the unit normal distribution above which lies $\alpha/k(k-1)$ percent of the distribution. The values of z can be obtained from Appendix Table A.

Because it is often necessary to obtain values based upon extremely small probabilities, especially when k is large, Appendix Table A_{II} may be used in place of Appendix Table A. This is a table of the standard normal distribution which has been arranged so that values used in multiple comparisons may be obtained easily. The table is arranged on the basis of the number of comparisons ($\#c$) that can be made. The tabled values are the *upper-tail* probabilities associated with various values of α. When there are k groups, there are $k(k-1)/2$ comparisons possible.[3]

Example 7.2*b* In the example above which analyzed the differences between patterns of reinforcement, the null hypothesis that there was no difference among the three training methods was rejected, and we concluded that there was a difference among the training methods. However, although we could conclude that there was a difference, we do not know whether there was a difference between one condition and the others or whether all three groups were different from each other. To find where the differences are, we shall determine the multiple comparisons for the three groups.

Since the $\alpha = .05$ level of significance was used in the initial analysis, we shall use the same level here as well. First, we determine the differences among the conditions. For convenience, we shall use the subscripts RR, RU, and UR to refer to the three groups. Then, since $R_{RR} = 39.5$, $R_{RU} = 42.5$, and $R_{UR} = 26.0$, we have the following differences:

$$|R_{RR} - R_{RU}| = |39.5 - 42.5| = 3.0$$
$$|R_{RR} - R_{UR}| = |39.5 - 26.0| = 13.5$$
$$|R_{RU} - R_{UR}| = |42.5 - 26.0| = 16.5$$

We then find the critical difference by using Eq. (7.5*a*). Since $\alpha = .05$ and $k = 3$, the number of comparisons $\#c$ is equal to $k(k-1)/2 = (3)(2)/2 = 3$. Referring to Appendix Table A_{II}, we see that the value of z is 2.394. [Alternatively, we could obtain the value of z from Appendix Table A. To use that table, we first compute $\alpha/k(k-1) = .05/(3)(2) = .00833$. Referring to Appendix Table A, we again find (after interpolation) that $z_{.00833} = 2.394$.] The critical difference is then

$$z_{\alpha/k(k-1)} \sqrt{\frac{Nk(k+1)}{6}} = 2.394 \sqrt{\frac{(18)(3)(3+1)}{6}}$$
$$= 2.394\sqrt{36}$$
$$= 14.36$$

Since only the third difference (16.5) exceeds the critical difference, we conclude that only the difference between conditions RU and UR is significant. Note that the second difference, although large, is not of a magnitude great enough to permit us to conclude that RR and UR are different when using the significance level we have chosen.

7.2.4 Comparison of Groups or Conditions with a Control

Sometimes a researcher may have a more specific comparison in mind than the set of multiple comparisons described above. For example, suppose one condition or group represented a *baseline* or control condition against which all of the other conditions should be compared. After applying the Friedman two-way analysis of

[3] Some readers will note a seeming discrepancy between the subscript for z, which is $\alpha/k(k-1)$, and the number of comparisons $\#c$, which is $k(k-1)/2$. Note that we are testing absolute differences and, therefore, use only the *upper tail* of tabled distribution. Hence the upper-tail probability, $\alpha/2$, is divided by the number of comparisons $k(k-1)/2$ which yields $\alpha/k(k-1)$.

variance by ranks test and noting that it is significant, the researcher may wish to compare all conditions against one. For convenience, we shall denote the control condition as condition 1. Then the hypothesis that the researcher would like to test is

$$H_0: \theta_1 = \theta_u \qquad \text{for } u = 2, 3, \ldots, k$$

against

$$H_1: \theta_1 \neq \theta_u \qquad \text{for } some\ u = 2, 3, \ldots, k$$

The following procedure permits the researcher to test a set of conditions against a control condition.

As with the multiple-comparison procedure described in the previous section, we compute the differences $|R_1 - R_u|$ between the treatment condition and each of the other conditions. When the sample sizes are moderate to large, these differences are approximately normally distributed. However, the comparisons are not independent and the critical values must be obtained by use of a special table (Appendix Table A_{III}). Then we can test the significance of differences between a treatment condition and other conditions by using the following inequality. That is, if

$$|R_1 - R_u| \geq q(\alpha, \#c)\sqrt{\frac{Nk(k+1)}{6}} \qquad (7.6a)$$

or if the data are expressed in terms of average ranks within each condition, and if

$$|\bar{R}_1 - \bar{R}_u| \geq q(\alpha, \#c)\sqrt{\frac{k(k+1)}{6N}} \qquad (7.6b)$$

then we can reject the hypothesis $H_0: \theta_1 = \theta_u$ in favor of $H_1: \theta_1 \neq \theta_u$. Values of $q(\alpha, \#c)$ are given in Appendix Table A_{III} for selected values of α and $\#c$, where $\#c = k - 1$, which is the number of comparisons.

Example 7.2.c As an example, suppose we had a set of $N = 12$ subjects measured at some baseline and in four other conditions; then $k = 5$. Suppose the values of $R_1 = 33$, $R_2 = 30$, $R_3 = 43$, $R_4 = 14$, and $R_5 = 60$. By using Eq. (7.3), the value of $F_r = 38.47$, which is significant beyond the $\alpha = .05$ level.[4] Suppose we then wish to test the difference between each condition and the baseline. The successive values of $|R_1 - R_u|$ are 3, 10, 19, and 27 respectively. By using Eq. (7.6a) we can find the limits for the differences. First, from Appendix A_{III}, we find that $q(\alpha, \#c) = 2.44$ for $\alpha = .05$ and $\#c = k - 1 = 4$. Then,

$$|R_1 - R_u| \geq q(\alpha, \#c)\sqrt{\frac{Nk(k+1)}{6}} \qquad (7.6a)$$

$$\geq 2.44\sqrt{\frac{(12)(5)(5+1)}{6}}$$

$$\geq 18.9$$

[4] The reader is encouraged to calculate the value of F_r in this example to ensure an understanding of its calculation from the data given.

Any difference which exceeds 18.9 will indicate a significant difference between that condition and the control condition. Only two of the differences exceed that limit. Therefore, we may conclude that conditions 4 and 5 are significantly different from control condition 1.

7.2.5 Summary of Procedure

These are the steps in the use of the Friedman two-way analysis of variance by ranks:

1. Cast the scores in a two-way table having N rows (subjects) and k columns (conditions or variables).
2. Rank the data in each row from 1 to k.
3. Determine the sum of the ranks in each column (R_j).
4. Compute the value of F_r with Eq. (7.3) if there are no ties or Eq. (7.4) if there are tied observations in any row.
5. The method for determining the probability of occurrence when H_0 is true of an observed value of F_r depends upon the sizes of N and k:
   (a) Appendix Table M gives selected critical values of F_r for small N and k.
   (b) For N and/or k larger than those used in Appendix Table M, the associated probability may be determined by reference to the χ^2 distribution (Appendix Table C) with $df = k - 1$.
6. If the probability yielded by the appropriate method in step 5 is equal to or less than α, reject H_0.
7. If H_0 is rejected, use multiple comparisons [Eq. (7.5)] to determine which differences among conditions are significant. If the differences among the various conditions and a control condition are to be tested, use Eq. (7.6).

7.2.6 Relative Efficiency

The power-efficiency of the Friedman two-way analysis of variance test for normally distributed data when compared with its normal counterpart (the F test) is $2/\pi = .64$ when $k = 2$ and increases with k to .80 for $k = 5$, .87 for $k = 10$, and .91 for $k = 20$. When compared with samples from uniform and exponential distributions the efficiency is greater.

7.2.7 References

Early discussion of the Friedman two-way analysis of variance by ranks may be found in Friedman (1937, 1940). More recent discussions include Lehmann (1975) and Randles and Wolfe (1979). The Friedman two-way analysis of variance by ranks is functionally related to another nonparametric test, the Kendall coefficient of concordance (W), which is discussed in Chap. 9.

NONPARAMETRIC STATISTICS FOR THE BEHAVIORAL SCIENCES

Second Edition

Sidney Siegel

Late Professor of the Pennsylvania State University

N. John Castellan, Jr.

Indiana University

McGRAW-HILL BOOK COMPANY

New York St. Louis San Francisco Auckland Bogotá Caracas Colorado Springs
Hamburg Lisbon London Madrid Mexico Milan Montreal New Delhi
Oklahoma City Panama Paris San Juan São Paulo Singapore Sydney Tokyo Toronto

NONPARAMETRIC SYSTEMS FOR THE BEHAVIORAL SCIENCES
INTERNATIONAL EDITION 1988

4 5 6 7 8 9 20 ANL 2002

This book was set in Times Roman by Syntax International.
The editor was James D. Anker.
The production supervisor was Denise L. Puryear.

Library of Congress Cataloging-in-Publication Data

Siegel, Sidney
Nonparametric statistics for the behavioral sciences.

Bibliography: p.
Includes index.
1. Psychometrics. 2. Nonparametric statistics.
3. Social sciences—Statistical methods. I. Castellan, N. John II. Title.
BF39.S464 1988 519.5 87-16923
ISBN 0-07-057357-3

When ordering this title use ISBN 0-07-100326-6

Printed in Singapore

ABOUT THE AUTHORS

Sidney Siegel was Research Professor of Psychology at the Pennsylvania State University at the time of his death in 1961 at age forty-five. A native of New York City, he was educated in California and received his Ph.D. in psychology from Stanford University in 1953. From then until his death he served on the Penn State faculty, except for a year as a Fellow at the Center for Advanced Study in the Behavioral Sciences, at Stanford.

He was the author or coauthor of four books published by McGraw-Hill. *Bargaining and Group Decision Making* (1960), with Lawrence E. Fouraker, was the recipient of the Monograph Prize of the American Academy of Arts and Sciences in 1959. Its sequel was *Bargaining Behavior* (1963), also with Fouraker. *Choice, Strategy, and Utility* was published by McGraw-Hill in 1964 after it was completed posthumously by Alberta E. Siegel and Julia McMichael Andrews. As well, McGraw-Hill published his collected writings in 1964, under the title *Decision and Choice*, edited by Samuel Messick and Arthur H. Brayfield. Included is a memoir by Ms. Siegel. The earliest of his books was *Nonparametric Statistics for the Behavioral Sciences* (1956), which has appeared in Japanese, Italian, German, and Spanish, as well as English.

N. John Castellan, Jr. is professor of psychology at Indiana University at Bloomington. He received his A.B. from Stanford University and his Ph.D. from the University of Colorado. He has served as visiting research associate at the Oregon Research Institute and visiting professor of computer science at the University of Colorado.

Professor Castellan has served as a consultant on statistics and computing to business and industry, and served as associate dean for research and graduate development at Indiana University. He has served on the editorial boards of several professional journals, is past president of the Society for Computers in Psychology,

TABLE M
Critical values for the Friedman two-way analysis of variance by ranks statistic, F_r*

k	N	$\alpha \leq .10$	$\alpha \leq .05$	$\alpha \leq .01$
3	3	6.00	6.00	–
	4	6.00	6.50	8.00
	5	5.20	6.40	8.40
	6	5.33	7.00	9.00
	7	5.43	7.14	8.86
	8	5.25	6.25	9.00
	9	5.56	6.22	8.67
	10	5.00	6.20	9.60
	11	4.91	6.54	8.91
	12	5.17	6.17	8.67
	13	4.77	6.00	9.39
	∞	4.61	5.99	9.21
4	2	6.00	6.00	–
	3	6.60	7.40	8.60
	4	6.30	7.80	9.60
	5	6.36	7.80	9.96
	6	6.40	7.60	10.00
	7	6.26	7.80	10.37
	8	6.30	7.50	10.35
	∞	6.25	7.82	11.34
5	3	7.47	8.53	10.13
	4	7.60	8.80	11.00
	5	7.68	8.96	11.52
	∞	7.78	9.49	13.28

* Some entries adapted and reproduced by permission of the publishers Charles Griffin & Co. Ltd., 16 Pembridge Road, London W11 3HL, from appendix table 5 of Kendall, M. G. (1970). *Rank correlation methods* (fourth edition). Other entries adapted from table A.15 of Hollander, M., and Wolfe, D. A. (1973). *Nonparametric statistics*. New York: J. Wiley. Reproduced by permission of the authors and publisher.