Nonparametric or Distribution-Free Methods

Methods of data analysis we have studied thus far generally involve the assumption of normality of observations or random errors. There are a class of procedures that are applicable without assuming a specific parametric family of distributions underlying the model, so they are called *nonparametric* or *distribution-free* methods of statistical inference. Most introductory statistics books provide an introduction to nonparametric methods.

The Sign Test: for a single sample or paired samples

Perhaps the simplest nonparametric test is the *sign test*. Given a random sample $Y_1, \ldots, Y_n$ from a continuous distribution with median $\tilde{\mu}$, consider testing the null hypothesis $H_o : \tilde{\mu} = \tilde{\mu}_o$. Let X be the number of observations greater than $\tilde{\mu}_o$. Then $X \sim \text{Bin}(n, p)$, where $p = 0.5$ if H_o is true. One can set up exact one- and two-tailed tests based on the binomial distribution, though there are only select natural choices for α.

This method can also be applied to paired sample data by analyzing the paired differences.

Example: Consider the data for training methods 1 and 2 of problem 21.5, p912, and consider conducting a 2-tailed test of the null hypothesis that the median of the median difference is zero. The differences are as follows.

Block	Method		
i	1	2	Δ
1	73	81	-8
2	76	78	-2
3	75	76	-1
4	74	77	-3
5	76	71	5
6	73	75	-2
7	68	72	-4
8	64	74	-10
9	65	73	-8
10	62	69	-7

Exact test: For a 2-tailed test, we would reject the null hypothesis if x is either large or small compared to $n/2$. Here, $x = 1$, a small value, since $E[X] = 5$ under the null hypothesis. Now, $P(X \leq 1 \mid p = 0.5) \approx 0.011$, and doubling this gives the observed significance level of 0.022. Based on this, one might reasonably reject the null hypothesis that the median difference is zero.

The only possible p-values in this case are 0.002, 0.022, 0.110, 0.344, and 0.754, unless one ventures into the realm of *randomization tests*.

Asymptotic test: For large n, one can base the test on $z = (x - 0.5n)/\sqrt{0.25n} = (\hat{p} - 0.5)/\sqrt{0.25/n}$, for $\hat{p} = x/n$, and use standard normal critical values.

Here, let's use the continuity correction since n is small. So, $p = 2P(X \leq 1 \mid p = 0.5)$
$= 2P(X \leq 1.5 \mid p = 0.5) \approx 2P[Z \leq (1.5 - 5)/\sqrt{2.5}] \approx 2P(Z \leq -2.21) \approx 2(0.0136) = 0.0272$.

The Wilcoxon Signed-Rank Test: for a single sample or paired samples

Given a random sample $Y_1, \ldots, Y_n$ from a symmetric, continuous distribution with median $\tilde{\mu}$, consider testing the null hypothesis $H_o : \tilde{\mu} = \tilde{\mu}_o$. Note: with the added assumption that the distribution is symmetric, the mean and median are the same. The method is as follows.

Step 1. Rank the absolute differences, $|Y_i - \tilde{\mu}_o|$.

Step 2. Use as the test statistic, S_+ say, equal to the sum of the ranks associated with the positive differences $Y_i - \tilde{\mu}_o$.

Example: Using the above data,

Block	Training Method					
i	1	2	Δ	$\lvert\Delta\rvert$	Rank	
1	73	81	-8	8	8.5	
2	76	78	-2	2	2.5	
3	75	76	-1	1	1	
4	74	77	-3	3	4	
5	76	71	5	5	6	6
6	73	75	-2	2	2.5	
7	68	72	-4	4	5	
8	64	74	-10	10	10	
9	65	73	-8	8	8.5	
10	62	69	-7	7	7	
						6

Hence, $S_+ = 6$. Is this significant?

Exact test: If the null hypothesis is true, then each of the ranks $1, 2, \dots, n$ is equally likely to be associated with a positive or negative sign, and these signs are independent. Thus, there are 2^n equally likely possible outcomes, each yielding a value of S_+ from 0 to $n(n+1)/2$, from which one can determine the exact null distribution of S_+ on which to base the test.

So, again, is $S_+ = 6$ significant? The only way to get S_+ this small or smaller, ignoring tied ranks, is if the following collections of ranks are associated with positive outcomes.

None;
1; 2; 3; 4; 5; 6;
1,2; 1,3; 1,4; 1,5; 2,3; 2,4;
1,2,3

These are 14 of 2^{10} equally likely possible outcomes if one ignores ties, so the observed significance level for a 2-tailed test is $p = 2 \times 14 \times (1/2)^{10} \approx 0.0274$, ignoring ties. One should however adjust for ties, in which case the possibility "2,4" would apparently be lost, as well as "8,9" at the other extreme, giving $p = 2 \times 13 \times (1/2)^{10} \approx 0.0254$.

Asymptotic test: Under the null hypothesis, $S_+ = \sum_{i=1}^{n} W_i$, where $P(W_i = i) = P(W_i = 0) = 0.5$, and the W_i are independent. Hence,

$$E[S_+] = \sum_i E[W_i] = \sum_i i/2 = n(n+1)/4\,,$$

and

$$\sigma^2(S_+) = \sum_i \sigma^2(W_i) = \sum_i E[(W_i - i/2)^2] = \sum_i (i/2)^2 = \cdots = n(n+1)(2n+1)/24\,.$$

So, for large n, one can base the test on $z = [S_+ - n(n+1)/4]/\sqrt{n(n+1)(2n+1)/24}$.

Here, $z = [6 - 27.5]/\sqrt{96.25} \approx -2.19$, so $p \approx 2P(Z \le -2.19) = 2(0.0143) = 0.0286$.

Using SAS: Proc UNIVARIATE provides three tests for location: Student's t test, the sign test, and the Wilcoxon signed rank test. In the SAS documentation, go to Procedures → Proc UNIVARIATE → Details → Tests for Location, to see details concerning the SAS implementation of these procedures. Note that SAS uses variations on the above test statistics.

Example: ch.nonpar.txt

The Wilcoxon Rank-Sum Test: for comparing means or medians of two populations based on independent random samples

(aka the Mann-Whitney-Wilcoxon test)

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from continuous distributions with means μ_1 and μ_2, respectively, where the two distributions are the same except for location.

To test $H_o : \mu_1 - \mu_2 = \Delta_o$ (or equivalently, $H_o : \tilde{\mu}_1 - \tilde{\mu}_2 = \Delta_o$), consider the adjusted values $X_i - \Delta_o$ ($i = 1, \ldots, m$) and Y_i ($i = 1, \ldots, n$), rank these $m + n$ adjusted values from smallest (rank 1) to largest (rank $m+n$), and let W denote the sum of the ranks associated with the X_i in the combined sample.

Exact test: If H_o is true, then these $m + n$ adjusted values constitute a random sample of size $m + n$ from the same continuous distribution, (i.e. with mean μ_2), and there are $\binom{m+n}{m}$ equally likely choices for the ranks associated with the X_i. This provides the null distribution for the test.

Asymptotic test: One can also obtain a large sample test. Since $W = \sum_{i=1}^{m} R_i$ for R_i the rank associated with X_i, it follows directly that $E[W] = m(m+n+1)/2$, and less directly that $\sigma^2(W) = mn(m+n+1)/12$, since $\sigma^2(R_i) = (m+n+1)(m+n-1)/12$ and $\mathrm{Cov}(R_i, R_j) = -(m+n+1)/12$. Hence, for large samples, one can base the test on $z = (w - E[W])/\sigma(W)$. If there are tied ranks, there is an adjusted formula for $\sigma^2(W)$.

Example Using SAS: See Proc NPAR1WAY → Examples → Example 52.2: The Exact Wilcoxon Two-Sample Test. The following description is from the SAS documentation.

"Researchers conducted an experiment to compare the effects of two stimulants. Thirteen randomly selected subjects received the first stimulant and six randomly selected subjects received the second stimulant. The reaction times (in minutes) were measured while the subjects were under the influence of the stimulants."

wilcoxon.txt

RECALL (with updates/revisions) from chapter 18:

18.7 Nonparametric Rank F Test and Confidence Interval Procedures

Nonparametric statistical methods are methods that do not depend on parametric model assumptions such as normality. As such, they are often suggested inappropriately as a remedial measure when variances are unequal. The Nonparametric Rank F Test considered here does not assume or require normality, but it does assume that the r treatment distributions are continuous and differ only with respect to location. Note: this implies equal variances!

Test Procedure: Rank the observations in ascending order from 1 to n_T, and let R_{ij} denote the rank of Y_{ij}. Conduct the usual ANOVA F-test, but using the statistic F_R^*, say, computed from the ranks rather than the original data. When the treatment distributions are the same, the statistic F_R^* follows approximately the $F(r-1, n_T - r)$ distribution if the sample sizes n_i are not very small. One can view this as a test for equality of treatment means or effects, or equivalently as a test of equality of treatment medians.

Example Using SAS: See PROC NPAR1WAY → Getting Started, for an example. The ANOVA option in the Proc NPAR1WAY statement does the above analysis, though there are other options/procedures one could choose to use instead. For example, The WILCOXON option requests an analysis using Wilcoxon scores, which are simply the ranks. When there are two classification levels, or two samples, this option produces the Wilcoxon rank-sum test. For any number of classification levels, this option produces the Kruskal-Wallis test based on the Chi-squared distribution. You can use the EXACT statement to request exact or Monte Carlo based p-values for any of the location or scale tests available in PROC NPAR1WAY—see PROC NPAR1WAY → Syntax → Exact, for details.

RECALL from chapter 21:

Nonparametric methods can be obtained by ranking the observations separately within each block, (see pp1138–1139). The F-statistics F_R^* computed from the ranks has asymptotically an $F(r-1,(r-1)(n_b-1))$ distribution. Asymptotically, the minimum significant difference for the Bonferroni method for pairwise differences based on treatment mean ranks is $B[r(r+1)/6n_b]^{1/2}$, where $B = z(1-\alpha/2g)$ for $g = \binom{r}{2} = r(r-1)/2$.

The test statistic has other forms, but the analysis is based on the sums of ranks for each of the treatments—aka the Friedman Rank Sums.

References

Hollander, M. and Wolfe, D. A. (1973), *Nonparametric Statistics Methods*, Wiley: New York.

Devore, J. L. (2000), *Probability and Statistics for Engineers and the Sciences*, Fifth Edition, Duxbury (Brooks/Cole): California.

Homework: Find the exact null distribution for the Wilcoxon Rank Sum test for two independent samples each of size three. What is the smallest observed significance level one could obtain for a two-tailed test? What is the smallest observed significance level one could obtain for a two-tailed test if both samples were of size 4? 5? 6? What is the smallest observed significance level one could obtain for a one-tailed test if both samples are of size: 3, 4, 5 and 6?