Parametric and nonparametric tests

A parametric statistical test is a test whose model specifies certain conditions about the parameters of the population from which the research sample was drawn. Since these conditions are not ordinarily tested, they are assumed to hold. The meaningfulness of the results of a parametric test depends on the validity of these assumptions. Parametric tests also require that the scores under analysis result from measurement in the strength of at least an interval scale.

A nonparametric statistical test is a test whose model does not specify conditions about the parameters of the the population from which the sample was drawn. Certain assumptions are associated with most nonparametric statistical tests, i.e that the observations are independent and that the variable under study has underlying continuity, but these assumptions are fewer and much weaker than those associated with parametric tests. Moreover, nonparametric tests do not require measurement so strong as that required for the parametric tests; most nonparametric tests apply to data in an ordinal scale, and some apply also to data in a nominal scale.

dvantages of Nonparametric tests

Probability statements obtained from most nonparametric stical tests are exact probabilities (except in the case of large samples, where excellent approximations are available), regardless of the shape of the population distribution from which the random sample was drawn.

2. If sample sizes as small as $N = 6$ are used, there is NO alternative to using a nonparametric statistical test unless the nature of the population distribution is known exactly.

3. Nonparametric statistical tests are available to treat data which are inherently in ranks as well as data whose seemingly numerical scores have the strength of ranks. That is, the researcher may only be able to say of his subjects that one has more or less of the characteristic than another, without being able to say how much more or less.

4. Nonparametric methods are available to treat data which are simply classificatory, i.e. measured in nominal scale. No parametric test technique applies to such data.

5. Nonparametric statistical tests are typically much easier to learn & to apply than are parametric tests.

Disadvantages of nonparametric Stat, test

1. If all the assumptions of the parametric statistical model are in fact met in the data, and if the measurement is of the required strength, then nonparametric stat. tests are wasteful of data.

2 There are as yet NO nonparametric methods for testing interactions in ANOVA model, unless special assumptions are made about additivity.

I) Chi-square One-sample test

Frequently research is undertaken in which the researcher is interested in the number of subjects, objects, or responses which fall in various categories. e.g. persons may be categorized according to whether they are 'favor of', 'indifferent to' or 'opposed to' some statement of opinion, to enable the researcher to test the hypothesis that these responses will differ in frequency.

Method :

In order to be able to compare an observed with an expected group of frequencies, we must of course be able to state what frequencies would be expected. The null hypothesis states the proportion of objects falling in each of the categories in the presumed population. The χ^2 technique tests whether the observed frequencies are sufficiently close to the expected ones to be likely to have occurred under Ho.

7.4

(4)

The null hypothesis may be tested by

$$\chi^2 = \frac{\sum_{i=1}^{k} (O_i - E_i)^2}{E_i} \sim \chi^2_{k-1}$$

where O_i : observed # of cases categorized in i^{th} categ
E_i : expected # of cases categorized in i^{th} categ.
$\sum_{i=1}^{k}$ directs one to sum over all (k) categories

If the agreement between the observed and expected frequencies is close, the difference ($O_i - E_i$) will be small & consequently χ^2 will be small. Roughly speaking, the larger χ^2 is, the more likely it is that the observed frequencies did not come from the population on which null hypothesis is based.

e.g. $\alpha = 0.01$

Ho: there is NO difference in expected # of winners starting from each of the post positions

1 day has 8 races, 144 is the total # of winners in 18 days of races

Post Position

	1	2	3	4	5	6	7	8	Total
expected # wins	18	18	18	18	18	18	18	18	
obs. No. of wins	29	19	18	25	17	10	15	11	144

$$\chi^2_{obs} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(29-18)^2}{18} + \frac{(19-18)^2}{18} + \frac{(18-18)^2}{18}$$
$$+ \frac{(25-18)^2}{18} + \frac{(17-18)^2}{18} + \frac{(10-18)^2}{18} + \frac{(15-18)^2}{18}$$
$$+ \frac{(11-18)^2}{18} = 16.3$$

$$\chi^2_{0.01}(7) = 18.48$$

$\chi^2_{obs} = 16.3 < 18.48$ Do <u>NOT</u> reject H_0 at 0.01 level.

II) <u>Mc Neman test for significance of changes</u>

The Mc Neman test for the significance of <u>changes</u> is particularly applicable to those "before and after" designs in which each person is used as his own control and in which measurement is in the strength of either a nominal or ordinal scale.

		After +	After −
Before	+	A	B
	−	C	D

H_0: No change before & after

Notice that those cases which show changes between the first and second response appear in cells B and C. An individual is tallied in cell B if he/she changed from + to −. He/she is tallied in C if he/she changed from − to +. If no change is observed, he/she is tallied in either cell A (+ responses both before & after) or cell D (− responses both before/after)

Since $B+C$ represents the total number of persons who changed, the expectation under the null hypothesis would be that $\frac{1}{2}(B+C)$ cases changed in one direction and $\frac{1}{2}(B+C)$ cases changed in the other. In other words $\frac{1}{2}(B+C)$ is the expected frequency under H_0 in both cell B and cell C.

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{\left(B - \frac{B+C}{2}\right)^2}{\left(\frac{B+C}{2}\right)} + \frac{\left(C - \frac{B+C}{2}\right)^2}{\left(\frac{B+C}{2}\right)}$$

$$= \frac{(B-C)^2}{(B+C)} \quad \text{with } df = 1 \quad \text{———} \quad (1)$$

That is, the sampling distribution under H_0 (No change between before and after) of χ^2 is distributed as approximately as chi-square with $df = 1$.

<u>Correction for continuity</u>. The approximation by the chi-square distribution of the sampling distribution of eqt (1) becomes an excellent one if a correction for continuity is performed.

$$\chi^2 = \frac{\left(|B-C|-1\right)^2}{B+C} \quad \text{with } d.f. = 1$$

e.g. (Polit's book p207)

		BSE After intervention	
		Yes	NO
BSE Before intervention	Yes	15	0
	NO	5	30

$$\chi^2 = \frac{(|5-0|-1)^2}{5+0} = \frac{16}{5} = 3.2$$

For the Mc Neman test, there is always 1 degree of freedom. $\chi^2_{0.05}(1) = 3.84$, $3.2 < 3.84$

We DO NOT reject null hypothesis. The change in % of women who practised BSE is NOT stat. sig. at the 0.05 level.

Wilcoxon Signed-Rank test

The Wilcoxon signed-rank test uses a random sample of matched pairs of observations. We wish to test the null hypothesis that the two population distributions are identical or the null hypothesis that the distribution of differences is centred at 0.

Procedure:

(a) discard pairs for which the difference is 0 and rank the absolute differences in ascending order.

(b) Calculate the sum of the ranks for +ve differences and for −ve differences. The observed Wilcoxon signed-rank test statistic t_0 is the smaller of these two sums.

(c) When sample size ≥ 15, the observed statistic

$$z = \frac{t_0 - \mu_T}{\sigma_T} \sim N(0,1), \quad t_0 = \min(T_P, T_N)$$

where $\mu_T = \frac{n(n+1)}{4}$, $\sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$

Case 1 : If the alternative hypothesis is one sided, reject Ho if $z < -z_\alpha$, α: level of significance

Case 2 : If the alternative hypothesis is two-sided,

reject Ho if $z < -z_{\alpha/2}$.

Any tied values are given the same rank, which is the average rank of the tied values.

e.g. A soft drink bottler has produced a new drink using two different recipes, one of which is much sweeter. The bottler asks 20 individuals to test both drinks and rate the drinks on a scale of 1 to 10, where 10 means the individual likes the drink very much. Use the Wilcoxon signed-rank test to test the null hypothesis that neither drink is prefered over the other. Use a two-tailed test and a 5% level of significance.

Sol. : The sum of the ranks for the positive difference is $T_p = 154$ and for negative differences, $T_n = 17$. The Wilcoxon signed-rank test statistic is the smaller of these two values,

$$t_o = \min(T_p, T_n) = \min(154, 17) = 17$$

$$\mu_T = \frac{n(n+1)}{4} = \frac{18\ (19)}{4} = 85.5$$

$n = 18$ because 2 pairs of observation deleted due to 0 diff

$$\sigma_T^2 = \frac{n(n+1)(2n+1)}{24} = \frac{(18)(19)(37)}{24} = 527.25$$

$$z = \frac{t_0 - \mu_T}{\sigma_T} = \frac{17 - 85.5}{\sqrt{527.25}} = -2.98$$

$z_{\alpha/2} = 1.96$,

critical value of the test $= -z_{\alpha/2} = -1.96$

Since $-2.98 < -z_{\alpha/2} = -1.96$ reject H_0 at 0.05 level.

Drink A	Drink B	Difference	Sign A − B	Rank +ve	Ra −v
10	6	4	+	13	
8	5	3	+	8.5	
6	2	4	+	13	
8	2	6	+	16	
7	4	3	+	8.5	
5	6	−1	−		2
1	4	−3	−		8.5
3	5	−2	−		4.5
9	9	0	Omit		
7	8	−1	−		2
4	2	2	+	4.5	
5	2	3	+	8.5	
8	1	7	+	18	
6	3	3	+	8.5	
8	2	6	+	16	
7	6	1	+	2	
4	1	3	+	8.5	
8	2	6	+	16	
9	5	4	+	13	
3	3	0	Omit		
				$T_p = 154$	$T_n = 17$

7.9a

SPSS example for Chapter 7.9 data:

```
-> get file='d:\stat601.14\polit\17.9sav'.

-> npar tests wilcoxon=drinka with drinkb.

- - - - - Wilcoxon Matched-Pairs Signed-Ranks Test

     DRINKA
with DRINKB

     Mean Rank    Sum of Ranks  Cases      T+

       11.00        154.0        14   - Ranks (DRINKB LT DRINKA)
        4.25         17.00        4   + Ranks (DRINKB GT DRINKA)
                                      2   0 Ties  (DRINKB EQ DRINKA)
                                     --
                                     20     Total
                              T-
       Z =    -3.0003                2-Tailed P =  .0027


-> t-test pairs=drinka with drinkb.

t-tests for Paired Samples
```

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
DRINKA	20	.256	.276	6.3000	2.386	.534
DRINKB				3.9000	2.292	.512

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
2.4000	2.854	.638	3.76	19	.001
95% CI (1.064, 3.736)					

10

When $n \leq 15$, the normal distribution does not necessarily provide a good approximation to the distribution of random variable T. For small values of n, tables of probabilities for T are available. For given values of n and α

$P(T \leq T_\alpha) = \alpha$. Reject H_0 if $t_0 < T_\alpha$

e.g. Two makes of tires are tested on the rear wheel of 6 different cars. The number of miles traveled until a tire fails is recorded. Because one tire of each make is used on each car, the observations occurred in matched pairs. Use the Wilcoxon signed-rank test and a 5% level of significance to test the null hypothesis

H_0: The population means are equal

vs H_A: The populat. mean are NOT equal.

Car	Tire A (in miles)	Tire B (in miles)	A-B	Rank +ve	Rank -ve
1	20 000	19000	1000	1	
2	24600	23000	1600	2	
3	32500	37 000	-4500		3
4	36000	30100	6500	4	
5	37 000	25500	11700	5	
6	23 000	39500	-16500		6
				$T_p = 12$	$T_n = 9$

Soln: $T_p = 12$, $T_n = 9$, $t_0 = \min(12, 9) = 9$

$T_{.05} = 1$ $\quad t_0 = 9 > 1$ DO NOT reject H_0 at .05 level.

7.10a

TABLE A.10 **Critical values of the Wilcoxon test statistic**

The following table gives critical values of T in the Wilcoxon matched-pair signed-rank test.

$n = 5, 6, 7, \ldots, 50$

1-sided	2-sided	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$\alpha = .05$	$\alpha = .10$	1	2	4	6	8	11
$\alpha = .025$	$\alpha = .05$		1	2	4	6	8
$\alpha = .01$	$\alpha = .02$			0	2	3	5
$\alpha = .005$	$\alpha = .01$				0	2	3

1-sided	2-sided	$n = 11$	$n = 12$	$n = 13$	$n = 14$	$n = 15$	$n = 16$
$\alpha = .05$	$\alpha = .10$	14	17	21	26	30	36
$\alpha = .025$	$\alpha = .05$	11	14	17	21	25	30
$\alpha = .01$	$\alpha = .02$	7	10	13	16	20	24
$\alpha = .005$	$\alpha = .01$	5	7	10	13	16	19

1-sided	2-sided	$n = 17$	$n = 18$	$n = 19$	$n = 20$	$n = 21$	$n = 22$
$\alpha = .05$	$\alpha = .10$	41	47	54	60	68	75
$\alpha = .025$	$\alpha = .05$	35	40	46	52	59	66
$\alpha = .01$	$\alpha = .02$	28	33	38	43	49	56
$\alpha = .005$	$\alpha = .01$	23	28	32	37	43	49

1-sided	2-sided	$n = 23$	$n = 24$	$n = 25$	$n = 26$	$n = 27$	$n = 28$
$\alpha = .05$	$\alpha = .10$	83	92	101	110	120	130
$\alpha = .025$	$\alpha = .05$	73	81	90	98	107	117
$\alpha = .01$	$\alpha = .02$	62	69	77	85	93	102
$\alpha = .005$	$\alpha = .01$	55	61	68	76	84	92

1-sided	2-sided	$n = 29$	$n = 30$	$n = 31$	$n = 32$	$n = 33$	$n = 34$
$\alpha = .05$	$\alpha = .10$	141	152	163	175	199	201
$\alpha = .025$	$\alpha = .05$	127	137	148	159	171	183
$\alpha = .01$	$\alpha = .02$	111	120	130	141	151	162
$\alpha = .005$	$\alpha = .01$	100	109	118	128	138	149

1-sided	2-sided	$n = 35$	$n = 36$	$n = 37$	$n = 38$	$n = 39$	
$\alpha = .05$	$\alpha = .10$	214	228	242	256	271	
$\alpha = .025$	$\alpha = .05$	195	208	222	235	250	
$\alpha = .01$	$\alpha = .02$	174	186	198	211	224	
$\alpha = .005$	$\alpha = .01$	160	171	183	195	208	

Example: If $n = 30$, then $P(T \geqslant 120) = .01$ and $P(T \geqslant 109) = .005$.

continued

ABLE A.10 **Critical values of the Wilcoxon test statistic (*continued*)**

$n = 5, 6, 7, \ldots, 50$

1-sided	2-sided	$n = 40$	$n = 41$	$n = 42$	$n = 43$	$n = 44$	$n = 45$
$\alpha = .05$	$\alpha = .10$	287	303	319	336	353	371
$\alpha = .025$	$\alpha = .05$	264	279	295	311	327	344
$\alpha = .01$	$\alpha = .02$	238	252	267	281	297	313
$\alpha = .005$	$\alpha = .01$	221	234	248	262	277	292
1-sided	2-sided	$n = 46$	$n = 47$	$n = 48$	$n = 49$	$n = 50$	
$\alpha = .05$	$\alpha = .10$	389	408	427	446	466	
$\alpha = .025$	$\alpha = .05$	361	379	397	415	434	
$\alpha = .01$	$\alpha = .02$	329	345	362	380	398	
$\alpha = .005$	$\alpha = .01$	307	323	339	356	373	

m Wilcoxon, F. and R. A. Wilcox. "Some Rapid Approximate Statistical Procedures," 1964. Reprinted by permission of Lederle Labs, a sion of the American Cyanamid Co.

Cochran's Q-Test (Polit p209)

Cochran's Q-test can be used to test for population differences in proportions when the dependant variable is dichotomous and when there are three or more repeated observations or correlated groups. For example, suppose a sample of 10 elderly patients with constipation problems was put on a special fiber-rich diet, and bowel movement were recorded for 3 conservative days, beginning with the day the treatment was initiated.

Patient	Day 1	Day 2	Day 3	Row Sum	Row sum ^{2}
1	1	0	1	2	4
2	0	1	1	2	4
3	0	1	0	1	1
4	0	1	1	2	4
5	1	1	1	3	9
6	0	0	1	1	1
7	0	1	0	1	1
8	1	0	1	2	4
9	0	1	1	2	4
10	0	1	1	2	4
	$S_{C_1} = 3$	$S_{C_2} = 7$	$S_{C_3} = 8$	$S_R = 18$	$S_R^2 = 36$

Codes: 0 = No bowel movement
1 = bowel movement

According to this table, a total of <u>three</u> patients had bowel movements on Day 1, 7 patients & 8 patients had bowel movements on Day 2 & Day 3 respectively. The research question is whether the differences reflect true population changes or are the result of sampling fluctuations.

$$Q = \frac{k(k-1) \sum_{i=1}^{k} \left(S_{C_i} - M_R\right)^2}{k(S_R) - S_{R^2}}$$

where k : # of times of obs.
S_C : sum of each column
S_R : sum of all summed rows.
$M_R = \frac{S_R}{k}$

$$M_R = \frac{18}{3} = 6$$

$$Q = \frac{3(3-1)\left[(3-6)^2 + (7-6)^2 + (8-6)^2\right]}{3(18) - 36}$$

$$= 4.67$$

When sample size ≥ 10, $Q \sim \chi^2_{(k-1)}$

$\chi^2_{0.05}(2) = 5.99$

$Q < 5.99$ DO NOT reject H_0 at 0.05 level.

The Friedman Test (Polit p211)

Like Cochran's Q, the Friedman Test is used when there are three or more correlated groups or more sets of observations for the same subjects. However, the Friedman test is used when the dependent variable

```
get file='d:\stat601.14\polit\p210.sav'.
npar tests cochran=day1 day2 day3.
```

7.12a

NPar Tests:p210.spo

Cochran Test

Frequencies

	Value	
	0	1
day1	7	3
day2	3	7
day3	2	8

Test Statistics

N	10
Cochran's Q	4.667[a]
df	2
Asymp. Sig.	.097

a. 1 is treated as a success.