CHAPTER 12: NONPARAMETRIC TESTS

IN PREVIOUS CHAPTERS, WE WERE NOT ABLE TO DEAL WITH SITUATIONS WHICH INVOLVED SMALL SAMPLES FROM POPULATIONS WHICH WERE NOT NORMAL.

THIS CHAPTER WILL TRY TO CORRECT THAT BY LEARNING A FEW TESTS WHICH DON'T REQUIRE THE NORMAL ASSUMPTION ('NONPARAMETRIC' SIMPLY MEANS: THE POPULATION CAN BE OF ANY SHAPE).

FURTHERMORE, THESE NEW TESTS ARE USUALLY LESS ELABORATE THAN THE OLD 'PARAMETRIC' ONES - WE CAN THUS USE THEM EVEN WITH LARGE SAMPLES, JUST TO KEEP THINGS SIMPLE.

< SIGN TEST

THIS IS A MODIFICATION OF THE 'PAIRED DIFFERENCE' TEST ('BEFORE AND AFTER').

IF WE CANNOT ASSUME THAT THE DIFFERENCES HAVE A NORMAL DISTRIBUTION, WE REPLACE THEM BY A SIMPLE INDICATION OF WHETHER THE 'AFTER' VALUE HAS INCREASED OR NOT (COMPARED TO 'BEFORE') - THIS IS USUALLY DONE BY THE **SIGN** OF THE DIFFERENCE (+ WHEN THE VALUE HAS GONE UP, ! FOR GOING DOWN). THE CASES WHERE THE VALUE STAYED THE SAME ARE MARKED ACCORDINGLY, AND EXCLUDED FROM A FINAL TALLY.

THE PROPORTION OF + SIGNS (OUT OF THE TOTAL OF + AND ! SIGNS, LET US CALL THIS NUMBER n) IS DENOTED x (THE TEXTBOOK HAS A KNACK FOR INCONSISTENT NOTATION).

THE POPULATION PROBABILITY OF A PLUS (+) SIGN IS CALLED p.

THE NULL HYPOTHESIS CLAIMS THAT THIS $p = \frac{1}{2}$ (PILL AS LIKELY TO INCREASE BLOOD PRESSURE AS TO REDUCE IT - THUS NOT EFFECTIVE AS MEDICINE).

THE DISTRIBUTION OF $\boldsymbol{x}$ IS, UNDER $\boldsymbol{H_0}$, BINOMIAL, BUT SINCE THE $n \times \frac{1}{2} > 5$ CONDITION IS MET WHENEVER $n > 10$, WE CAN TAKE IT TO BE APPROXIMATELY NORMAL, WITH THE MEAN OF **½** AND THE STANDARD DEVIATION OF $\sqrt{\frac{0.25}{n}}$.

THE **TEST STATISTICS** IS THUS EQUAL TO

$$\frac{x - 0.5}{\sqrt{\frac{0.25}{n}}}$$

AND HAS, UNDER THE $\boldsymbol{H_0}$ HYPOTHESIS, THE USUAL STANDARD NORMAL DISTRIBUTION OF OUR TABLES.

LET US GO BACK TO ONE OF OUR PREVIOUS EXAMPLES (CHAPTER 9):

A CERTAIN BLOOD-PRESSURE MEDICATION IS BEING TESTED ON 12 RANDOMLY SELECTED INDIVIDUALS. THEIR BLOOD PRESSURE IS RECORDED BEFORE AND AFTER THEY TAKE THIS MEDICATION; THESE ARE THE RESULTS:

B:	143	128	160	148	139	172	144	150	138	153	180	163
A:	128	132	144	139	137	140	125	138	139	139	161	129
SIGN	!	%	!	!	!	!	!	!	%	!	!	!

ALL WE NEED TO KNOW NOW IS THAT THE BLOOD PRESSURE HAS INCREASED IN ONLY 2 OUT OF 12 CASES (OTHERWISE, IT HAS ALWAYS DROPPED - NO N.D. VALUES)

TO TEST $\boldsymbol{H_0}$: **p = ½** AGAINST $\boldsymbol{H_1}$: **p < ½**

WE COMPUTE THE VALUE OF THE TEST STATISTICS, THUS:

$$\frac{\frac{2}{12}-0.5}{\sqrt{\frac{0.25}{12}}}=-2.309$$

USING THE SAME 1% LEVEL OF SIGNIFICANCE AS WE USED ORIGINALLY, THE CORRESPONDING CRITICAL VALUE (THE LAST ROW OF TABLE 6) EQUALS !2.326 (LEFT-TAIL TEST)!

CONCLUSION: BASED ON THE SIGN TEST, WE DON'T HAVE SIGNIFICANT ENOUGH PROOF THAT THE MEDICATION IS EFFECTIVE (WHEN FLIPPING A FAIR COIN 12 TIMES, GETTING ONLY TWO HEADS IS STILL FEASIBLE - HAS MORE THAN 1% PROBABILITY).

HOW COME THAT THE SAME DATA RESULTED IN A HIGHLY SIGNIFICANT REJECTION OF THE NULL HYPOTHESIS LAST TIME? IT'S BECAUSE NONPARAMETRIC TEST ARE, IN GENERAL, LESS POWERFUL THAN THEIR PARAMETRIC COUNTERPARTS (THAT'S THE PRICE TO PAY FOR SIMPLICITY, AND A LOSS OF THE 'NORMAL' ASSUMPTION).

ANOTHER EXAMPLE:

THE FOLLOWING IS THE NUMBER OF MIGRAINE HEADACHES 15 PATIENTS HAVE SUFFERED IN A MONTH BEFORE AND AFTER THEY STARTED TO TAKE A CERTAIN MEDICATION:

BEFORE:	4	2	9	8	3	6	3	7	3	4	2	8	3	5	4
AFTER:	2	0	4	8	0	3	4	4	1	1	3	4	3	1	0
SIGN:	!	!	!	0	!	!	%	!	!	!	%	!	0	!	!

THE TWO HYPOTHESES ARE THE SAME AS IN THE PREVIOUS EXAMPLE (A LEFT-TAIL ALTERNATIVE), THE TEST STATISTIC IS COMPUTED AS FOLLOWS:

$$\frac{\frac{2}{13}-0.5}{\sqrt{\frac{0.25}{13}}}=-2.496$$

NOTE THAT n IS NOW EQUAL TO 13 (NOT 15)!

USING THE SAME SIGNIFICANCE LEVEL OF 1% (AND THEREFORE THE SAME CRITICAL VALUE OF 2.326), WE CAN NOW CLAIM TO HAVE A STATISTICALLY SIGNIFICANT EVIDENCE OF THE MEDICATION BEING EFFECTIVE.

< **RANK-SUM (MANN-WHITNEY) TEST**

THIS TEST IS USED IN THE SITUATION OF TWO <u>INDEPENDENT</u> SAMPLES, TAKEN FROM POPULATIONS OF <u>IDENTICAL SHAPE</u>, BUT NOT NECESSARILY OF THE SAME MEAN.

IT WILL ENABLE US TO TEST THE USUAL NULL HYPOTHESIS $\mu_1 = \mu_2$ AGAINST POSSIBLE ALTERNATIVES, <u>WITHOUT</u> HAVING TO ASSUME THAT THE POPULATIONS ARE NORMAL (OR THAT EACH SAMPLE SIZE IS BIGGER THAN 30).

IT WORKS AS FOLLOWS:

FIRST, WE POOL THE TWO SAMPLES INTO ONE AND ASSIGN A **RANK** TO EACH OBSERVATION.

THIS MEANS ARRANGING THE OBSERVATIONS FROM THE SMALLEST TO THE LARGEST, AND RANKING THEM: 1, 2, 3, ... N, WHERE N IS THE TOTAL SAMPLE SIZE (I.E. $N = n_1 + n_2$).

IN CASE OF **TIES** (IDENTICAL OBSERVATIONS), EACH GET THE AVERAGE RANK THEY WOULD BE GETTING INDIVIDUALLY.

EXAMPLE:

OBSERVATION:	23	47	15	29	18	33	29	18	31	42	29
RANK:	4	11	1	6	2.5	9	6	2.5	8	10	6

WE THEN COMPUTE THE SUM OF RANKS (DENOTED R) OF ALL n_1 OBSERVATIONS COMING FROM THE FIRST SAMPLE.

THE FINAL TEST STATISTIC IS

$$\frac{R - n_1(n_1 + n_2 + 1)/2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}$$

WHEN EACH SAMPLE SIZE (n_1 AND n_2) IS BIGGER THAN 7, THE TEST STATISTIC HAS (TO A GOOD APPROXIMATION) THE STANDARD NORMAL DISTRIBUTION.

EXAMPLE: LIGHT BULBS OF TWO TRADEMARKS ARE TESTED BY APPLYING HIGH VOLTAGE AND RECORDING THE TIME TILL THEY BURN. THESE ARE THE RESULTS:

TYPE A: 1.2, 3.0, 1.4, 0.3, 4.7, 2.2, 0.7, 2.7, 3.9 HOURS

TYPE B: 1.9, 5.4, 6.1, 3.8, 2.5, 8.4, 2.8 HOURS

CAN WE CONCLUDE THAT EITHER TYPE HAS A HIGHER DURABILITY THAN THE OTHER (USE 10% LEVEL OF SIGNIFICANCE).

TESTING $\boldsymbol{H_0}$: $:_1 = :_2$ AGAINST $\boldsymbol{H_1}$: $:_1 \neq :_2$

FIRST WE REPLACE THE ORIGINAL DATA WITH RANKS:

TYPE A: 3, 10, 4, 1, 13, 6, 2, 8, 12

TYPE B: 5, 14, 15, 11, 7, 16, 9

R IS THEREFORE EQUAL TO 3 + 10 + ... + 12 = 59, $n_1 = 9$, $n_2 = 7$.

THE VALUE OF THE TEST STATISTIC IS THUS EQUAL TO

$$\frac{59 - 9 \times 17/2}{\sqrt{9 \times 7 \times 17/12}} = -1.852$$

CRITICAL VALUES ARE ± 1.645.

CONCLUSION: YES, THERE IS STATISTICALLY SIGNIFICANT EVIDENCE THAT TYPE B BULBS LAST LONGER.

Second example:

12 families were randomly selected from US and Japan, and the number of children recorded, as follows:

US	0	3	1	1	2	2	0	4	2	1	1	0
JAPAN	2	5	0	3	5	3	2	0	1	3	4	2

Does this data constitute a statistically significant evidence (using α = 10%) that families in Japan are larger than those in US?

Now, we test $\boldsymbol{H_0}$: $\mu_1 = \mu_2$ against $\boldsymbol{H_1}$: $\mu_1 < \mu_2$

Again, converting to ranks:

US	3	18.5	8	8	13.5	13.5	3	21.5	13.5	8	8	3
JAPAN	13.5	23.5	3	18.5	23.5	18.5	13.5	3	8	18.5	21.5	13.5

R = 3 + 18.5 + 8 + ... + 3 = 121.5

Value of test statistic: $\dfrac{121.5 - 12 \times 12.5}{\sqrt{12 \times 12 \times 25/12}}$ = - 1.645

The critical value is - 1.282.

Conclusion: Yes, we have statistically significant evidence that families in Japan are larger.

< SPEARMAN CORRELATION

WHEN THE EXPECTED VALUE OF THE DEPENDENT VARIABLE Y INCREASES (DECREASES) WITH X, BUT THE RELATIONSHIP IS NOT NECESSARILY LINEAR (STRAIGHT LINE), THE USUAL CORRELATION COEFFICIENT r OF CHAPTER 10 IS NO LONGER APPROPRIATE AS A MEASURE OF STRENGTH OF THIS RELATIONSHIP.

WHAT WE USE INSTEAD IS THE SO CALLED SPEARMAN (RANK) CORRELATION COEFFICIENT r_s, COMPUTED BY FIRST **RANKING,** INDIVIDUALLY, THE x AND y OBSERVATIONS (TO SIMPLIFY THE ISSUE, YOUR TEXTBOOK USUALLY PRESENTS ITS DATA IN THIS RANKED FORM ALREADY), AND THEN SUBSTITUTING THESE RANKS INTO THE OLD FORMULA FOR COMPUTING r. ONE CAN PROVE THAT, AS A RESULT, WE GET:

$$r_s \equiv 1 - \frac{6\sum d^2}{n(n^2-1)}$$

WHERE $\boldsymbol{d = x - y}$ (WE WILL HAVE n OF THESE, n BEING THE NUMBER OF x-y PAIRS).

EXAMPLE: THE FOLLOWING TABLE GIVES THE AGE OF A RANDOM EMPLOYEE, TOGETHER WITH HIS/HER SALARY (ROUNDED OFF TO THE NEAREST THOUSAND):

AGE	27	45	31	49	62	55	39	24	33
SALARY	31	37	29	38	43	47	30	28	35

TO COMPUTE r_s, WE MUST FIRST REPLACE THIS DATA BY THE CORRESPONDING RANKS, THUS:

AGE	2	6	3	7	9	8	5	1	4
SALARY	4	6	2	7	8	9	3	1	5
d	-2	0	1	0	1	-1	2	0	-1

WHICH YIELDS: $r_s = 1 - \frac{6\times(4+1+1+1+4+1)}{9\times(81-1)} = 0.900$

BASED ON THIS <u>SAMPLE</u> CORRELATION COEFFICIENT r_s, WE CAN ALSO **TEST** WHETHER THE CORRESPONDING

POPULATION CORRELATION COEFFICIENT D_s IS NON-ZERO OR NOT (THE CRITICAL VALUES OF r_s ARE LISTED IN TABLE 9).

CONTINUATION OF THE PREVIOUS EXAMPLE: DOES THIS DATA PROVIDE A STATISTICALLY SIGNIFICANT EVIDENCE (USING " = ½ %) THAT SALARIES INCREASE WITH AGE?

TESTING **H_0: $D_s = 0$** AGAINST **H_1: $D_s > 0$**

BASED ON TABLE 9, THE CRITICAL VALUE OF r_s IS 0.834 .

CONCLUSION: YES, WE HAVE A HIGHLY SIGNIFICANT EVIDENCE THAT, IN THE PARTICULAR COMPANY FROM WHICH THIS SAMPLE WAS TAKEN, SALARIES DO INCREASE WITH AGE.