CHAPTER 11: THREE MORE TESTS (INDEPENDENCE, GOODNESS OF FIT, SAME POPULATION MEANS)

< **CHI-SQUARE TEST OF INDEPENDENCE**

SUPPOSE WE HAVE TWO ATTRIBUTES (VARIABLES, USUALLY OF THE <u>NOMINAL</u> TYPE, E.G. UNIVERSITY STUDENT'S GENDER VERSUS THEIR FIELD OF STUDY).

SAMPLING FROM A POPULATION, THE RESULTS ARE CONVENIENTLY SUMMARIZED IN A **CONTINGENCY TABLE**, E.G.

SEX9 FIELD6	SCIENCE	ART	HUMANITIES
MALE	47	19	36
FEMALE	29	23	44

THE QUESTION IS:

IN THE ACTUAL POPULATION, IS THE WAY (THE 3 PROBABILITIES) HOW MALE AND FEMALE STUDENTS CHOOSE THEIR FIELD THE SAME FOR BOTH GROUPS?

IE. DO OUR SAMPLE DIFFERENCES REFLECT ONLY THE RANDOM FLUCTUATIONS, INHERENT IN ANY SAMPLING?

PUTTING IT YET ANOTHER WAY, ARE THE TWO VARIABLES INDEPENDENT OF EACH OTHER (THIS WILL BE OUR NULL HYPOTHESIS)?

OR (THE ALTERNATE HYPOTHESIS): IS THERE A REAL (POPULATION) DIFFERENCE IN HOW MALE AND FEMALE STUDENTS CHOOSE THEIR FILED (THE TWO VARIABLES BEING NOT INDEPENDENT)?

THE ACTUAL STATISTICAL TEST REQUIRES US TO FIRST COMPUTE THE SO CALLED **EXPECTED** FREQUENCIES (THOSE IN THE ORIGINAL TABLE ARE CALLED **OBSERVED** FREQUENCIES).

THIS IS DONE IN TWO STEPS:

i) COMPUTE THE ROW, COLUMN, AND FULL TABLE FREQUENCY TOTALS (THE LAST IS EFFECTIVELY THE SAMPLE SIZE n), THUS:

	SCIENCE	ART	HUMAN.	
MALE				102
FEMALE				96
	76	42	80	n=198

ii) THEN, COMPUTE THE EXPECTED FREQUENCIES (THEY DON'T NEED TO HAVE INTEGER VALUES) OF EACH **CELL** BY MULTIPLYING THE CORRESPONDING ROW AND COLUMN TOTALS, AND DIVIDING BY n, THUS:

102×76÷198 = 39.15	102×42÷198 = 21.64	102×80÷198 = 41.21	102
96×76÷198 = 36.85	96×42÷198 = 20.36	96×80÷198 = 38.79	96
76	42	80	198

(NOTE THAT THE EXPECTED FREQUENCIES HAVE THE SAME ROW AND COLUMN TOTALS AS THE OBSERVED ONES).

TO MAKE THIS TEST POSSIBLE, ALL THESE <u>EXPECTED</u> FREQUENCIES MUST BE <u>BIGGER THAN 5.</u>

THE TEST STATISTIC IS $\Sigma\frac{(O-E)^2}{E}$

WHERE O AND E STAND FOR THE OBSERVED AND EXPECTED FREQUENCIES, AND THE SUMMATION IS OVER THE WHOLE TABLE, E.G.

$$\frac{(47-39.15)^2}{39.15}+\frac{(19-21.64)^2}{21.64}+\frac{(36-41.21)^2}{41.21}+$$
$$\frac{(29-36.85)^2}{36.85}+\frac{(23-20.36)^2}{20.36}+\frac{(44-38.79)^2}{38.79}=5.27$$

UNDER H_0, THIS TEST STATISTIC HAS, TO A GOOD APPROXIMATION, THE P^2 (CHI-SQUARE) DISTRIBUTION WITH $(R - 1)\times(C - 1)$ $=$ (2-1)(3-1) $=$ 2 DEGREES OF FREEDOM.

WE REJECT H_0 WHEN THE RESULTING VALUE IS <u>BIGGER</u> THAN χ_c^2 (TABLE 7).

IN OUR CASE, USING A 5% LEVEL OF SIGNIFICANCE, THE CRITICAL VALUE IS 5.99. CONCLUSION: WE <u>DON'T</u> HAVE A SUFFICIENT EVIDENCE TO REJECT INDEPENDENCE (THAT MALE AND FEMALE STUDENTS' CHOICES FOLLOW THE SAME PATTERN).

Chi-Square Test: C1, C2, C3

```
Expected counts are printed below observed
counts

            C1         C2         C3    Total
    1       47         19         36      102
         39.15      21.64      41.21

    2       29         23         44       96
         36.85      20.36      38.79

Total       76         42         80      198

Chi-Sq =  1.573 +  0.321 +  0.659 +
          1.672 +  0.341 +  0.700 = 5.267
DF = 2, P-Value = 0.072
```

< GOODNESS-OF-FIT TEST

THIS IS USED IN SITUATIONS WHERE WE LOOK AT A <u>SINGLE</u> ATTRIBUTE (A NOMINAL-SCALE VARIABLE) HAVING A HANDFUL OF DISTINCT VALUES (CHOICES) OF GIVEN (HYPOTHESIZED) PROBABILITIES (E.G. WE MAY BELIEVE THAT UNIVERSITY STUDENTS

CHOOSE SCIENCE, ART OR HUMANITIES WITH THE PROBABILITY OF 2/5, 1/5 AND 2/5 RESPECTIVELY), CONSTITUTING OUR **NULL HYPOTHESIS**.

WE WANT TO CHECK WHETHER OUR EXPERIMENT (OBSERVED FREQUENCIES - ONLY ONE ROW OF THESE) 'SQUARE' WITH THESE PROBABILITIES.

THE **ALTERNATE** HYPOTHESIS WOULD NOW SIMPLY STATE 'NOT SO'.

THE **TEST STATISTIC** IS PRACTICALLY THE SAME AS IN THE PREVIOUS SECTION, EXCEPT NOW THE <u>EXPECTED</u> FREQUENCIES ARE COMPUTED BASED ON THE DISTRIBUTION SPECIFIED BY H_0 - WE SIMPLY MULTIPLY EACH PROBABILITY BY n.

THE NUMBER OF **DEGREES OF FREEDOM** IS EQUAL TO THE NUMBER OF 'CELLS' <u>MINUS ONE</u>.

EXAMPLE:

H_0: STUDENTS CHOOSE THEIR FIELD OF STUDY ACCORDING THE FOLLOWING DISTRIBUTION:

FIELD:	SCIENCE	ART	HUMANITIES
PROB:	0.4	0.2	0.4

H_1: ... SOME OTHER DISTRIBUTION.

THE OBSERVED AND EXPECTED FREQUENCIES ARE:

FIELD:	SCIENCE	ART	HUMANITIES
OBSERVED:	76	42	80
EXPECTED:	0.4×198 = 79.2	0.2×198 = 39.6	0.4×198 = 79.2

THE VALUE OF THE TEST STATISTIC IS:

$$\frac{(76-79.2)^2}{79.2}+\frac{(42-39.6)^2}{39.6}+\frac{(80-79.2)^2}{79.2}=0.283$$

USING THE USUAL 5% LEVEL OF SIGNIFICANCE, THE CRITICAL VALUE OF THE P^2 DISTRIBUTION WITH 2 DEGREES OF FREEDOM IS: 5.99

CONCLUSION: THE FIT IS EXTREMELY GOOD, THERE IS NO REASON TO REJECT H_0.

SECOND EXAMPLE:

ROLLING A DIE 100 TIMES AND OBTAINING THE FOLLOWING NUMBER OF

ONES	TWOS	THREES	FOURS	FIVES	SIXES
14	11	15	16	14	30

WE WOULD LIKE TO TEST WHETHER THE DIE IS 'FAIR' (I.E. YIELDS THE PROBABILITY OF 1/6 FOR EACH TYPE OF OUTCOME) - THIS IS OUR NULL HYPOTHESIS.

THE EXPECTED FREQUENCIES MUST BE THEN ALL EQUAL TO 100/6 = 16.67 .

THE VALUE OF THE TEST STATISTIC IS

$$\frac{(14-16.67)^2}{16.67}+\frac{(11-16.67)^2}{16.67}+\frac{(15-16.67)^2}{16.67}+\frac{(16-16.67)^2}{16.67}+$$
$$+\frac{(14-16.67)^2}{16.57}+\frac{(30-16.67)^2}{16.67}=13.64$$

THE CRITICAL VALUE (USING 5 DEGREES OF FREEDOM AND 5% LEVEL OF SIGNIFICANCE) IS 11.07.

CONCLUSION: AT THIS LEVEL OF SIGNIFICANCE, WE DO HAVE ENOUGH STATISTICAL EVIDENCE TO REJECT THE NULL HYPOTHESIS AND DECLARE THE DIE 'CROOKED' (NOTE THAT, AT THE 1% LEVEL OF SIGNIFICANCE, WE WOULD BE ACCEPTING H_0 - OUR REJECTION OF H_0 IS STILL NOT THAT CONVINCING - WE MAY BE COMMITTING TYPE I ERROR).

< ANALYSIS OF VARIANCE (ANOVA)

THIS IS A PECULIAR NAME FOR AN EXTENSION OF A TEST WE DID BEFORE:

DO TWO POPULATIONS HAVE THE SAME MEAN, OR ARE THE TWO MEANS DIFFERENT?

NOW WE HAVE THREE OR MORE POPULATIONS (SAY *k*), ASKING THE SAME QUESTION: DO THEY ALL HAVE THE SAME MEAN, OR IS AT LEAST ONE OF THEM DIFFERENT FROM THE REST (AT LEAST TWO OF THEM DIFFERENT FROM ONE ANOTHER?

THE ASSUMPTIONS ARE:

POPULATIONS (AT LEAST APPROXIMATELY) <u>NORMAL</u> (THIS IS LESS ESSENTIAL WHEN HAVING LARGE SAMPLES).

THEIR F 'S ARE IDENTICAL (THERE IS A WAY OF TESTING THIS, WHICH WE ARE SKIPPING).

THE k SAMPLES ARE <u>INDEPENDENT</u> (NO 'BEFORE AND AFTER' WITH THE SAME SUBJECT), AND NOT NECESSARILY OF THE SAME SIZE.

TO COMPUTE THE **TEST STATISTIC** IS, IN THIS CASE, RATHER ELABORATE:

FIRST, WE COMPUTE WHAT WE USED TO CALL 'SUM OF SQUARES ...' SS (BY THE USUAL $\Sigma x^2 - \frac{(\Sigma x)^2}{n}$ FORMULA), INDIVIDUALLY, FOR <u>EACH</u> OF THE k SAMPLES. WE THEN <u>ADD</u> THEM TOGETHER, CALLING THE RESULT **WITHING** (SAMPLES, OR GROUPS) VARIABILITY $SS_{W.}$

SECONDLY, WE **POOL** ALL k SAMPLES INTO ONE, AND COMPUTE THE CORRESPONDING SS (AS IF DEALING WITH <u>ONE</u> SAMPLE) CALLING IT SS_{TOT} (**TOTAL** VARIABILITY).

THIRDLY, WE COMPUTE SS_{BET} / SS_{TOT} - SS_{W}, CALLING IT **BETWEEN** (SAMPLES, OR GROUPS) VARIABILITY.

FINALLY, THE TESTS STATISTIC IS COMPUTED BY

$$\frac{SS_{BET}/(k-1)}{SS_{W}/(N-k)}$$

WHERE N IS THE TOTAL (POOLED) SAMPLE SIZE.

THE NUMERATOR IS CALLED **MEAN SQUARE BETWEEN** (MS_{BET}), THE DENOMINATOR **MEAN SQUARE WITHIN** (MS_{W}).

THE DISTRIBUTION OF THIS SAMPLE STATISTIC IS CALLED F (FISHER), AND ITS CRITICAL VALUES (ALWAYS THE RIGHT TAIL) ARE FOUND IN TABLE 8.

NOTE THAT NOW WE NEED TWO 'DEGREES OF FREEDOM', THE NUMERATOR *d.f.* EQUAL TO (k - 1), AND THE DENOMINATOR *d.f.* EQUAL TO (N - k).

EXAMPLE: SUPPOSE WE WANT TO TEST (USING 5% LEVEL OF SIGNIFICANCE) WHETHER SCIENCE, ART AND HUMANITIES STUDENTS AT A GIVEN UNIVERSITY ARE EQUALLY PROFICIENT IN ENGLISH. WE SELECT A RANDOM SAMPLE FROM EACH GROUP (EACH 'POPULATION') AND LET THEM TAKE AN ENGLISH-PROFICIENCY TEST. THE RESULTS ARE AS FOLLOWS:

SCIENCE: 68, 73, 80, 77, 86, 61

ART: 59, 88, 69, 72, 90

HUMANITIES: 73, 91, 84, 66, 72, 80, 87, 79

H_0: $:_1 = :_2 = :_3$

H_1: AT LEAST ONE OF THE MEANS IS DIFFERENT FROM THE REST.

FIRST WE COMPUTE THE INDIVIDUAL SAMPLE SUMS, SUM OF SQUARES AND THE CORRESPONDING SS:

	n	Σx	Σx^2		SS
	6	445	33399		394.83
	5	378	29270		693.2
	8	632	50416		488
				SS_{TOT}	SS_W
TOTAL:	19	1455	113085	1662.63	1576.03

THERE ARE TWO WAYS OF COMPUTING SS_{BET}:

SS_{TOT} - SS_W = 86.6 OR

$$SS_{BET} = \frac{445^2}{6} + \frac{378^2}{5} + \frac{632^2}{8} - \frac{1455^2}{19} = 86.6$$

THE REMAINING COMPUTATION OF THE TEST STATISTIC IS USUALLY ORGANIZED IN THE FOLLOWING TABLE:

SOURCE OF VARIATION	**SUM OF SQUARES**	**D.F.**	**MEAN SQUARE**	**F RATIO**
BETWEEN	*86.60*	*3 - 1*	*43.30*	*0.440*
WITHIN	*1576.03*	*19 - 3*	*98.50*	
TOTAL	*1662.63*	*19 - 1*		

SINCE THE CRITICAL VALUE OF $F_{2,16}$ IS *3.63* (NOTE THAT IT IS ALWAYS BIGGER THAN 1), WE DON'T HAVE ANY EVIDENCE TO REJECT THE CLAIM THAT THE THREE POPULATION MEANS ARE IDENTICAL.

One-way ANOVA: C1 versus C2

```
Analysis of Variance for C1
Source     DF        SS        MS        F        P
Factor      2      86.6      43.3     0.44    0.652
Error      16    1576.0      98.5
Total      18    1662.6
```