TENTH LECTURE SUMMARY

TEST OF INDEPENDENCE

OF TWO VARIABLES (ATTRIBUTES) BOTH TYPICALLY (BUT NOT NECESSARILY) OF THE NOMINAL-SCALE TYPE.

SAMPLE IS USUALLY SUMMARIZED IN A CONTINGENCY TABLE, E.G.

GRADE6 9YEAR	A	B	C	D	
I	23 21.66	15 16.74	17 15.95	9 9.65	64
II	56 57.20	48 44.20	39 44.20	26 25.48	169
III	31 31.14	22 24.06	25 22.93	14 13.87	92
	110	85	81	49	325

WHERE WE CAN DISPLAY BOTH THE OBSERVED AND EXPECTED (CALCULATED) FREQUENCIES OF EACH CELL.

NULL HYPOTHESIS STATES THAT THE TWO VARIABLES ARE INDEPENDENT (THE DISTRIBUTION OF GRADES DOES NOT CHANGE WITH YEAR OF STUDY.

TEST STATISTIC COMPUTED BY $\Sigma\frac{(O-E)^2}{E}$

I.E. $\frac{(23-21.66)^2}{21.66}+\frac{(15-16.75)^2}{16.75}+\ldots+\frac{(14-13.87)^2}{13.87}$ = 1.335

AND COMPARED AGAINST THE CRITICAL VALUE OF THE CHI-SQUARE DISTRIBUTION WITH (#R - 1)×(#C - 1) DEGREES OF FREEDOM (ALWAYS A RIGHT TAIL TEST).

GOODNESS-OF-FIT TEST

CONCERNS A SINGLE NOMINAL-SCALE VARIABLE WITH A HANDFUL OF POTENTIAL VALUES. THE INDIVIDUAL PROBABILITIES OF THESE ARE SPECIFIED BY THE NULL HYPOTHESIS, E.G.

GRADE	A	B	C	D
OBSERVED	110	85	81	49
H_0: Prob =	35%	25%	25%	15%
EXPECTED	113.75	81.25	81.25	47.25

EXPECTED FREQUENCIES ARE COMPUTED BY MULTIPLYING EACH PROBABILITY OF H_0 BY n (TOTAL OF OBSERVED FREQUENCIES).

TEST STATISTIC IS COMPUTED USING THE SAME $\sum \frac{(O-E)^2}{E}$, I.E.

$$\frac{(110-113.75)^2}{113.75} + \ldots + \frac{(49-47.25)^2}{47.25} = 0.363$$

AND COMPARED TO CRITICAL VALUE OF CHI-SQUARE DISTRIBUTION WITH k - 1 DEGREES OF FREEDOM (CRITICAL REGION CONSISTS AGAIN OF THE RIGHT TAIL ONLY), WHERE k IS THE NUMBER OF 'CELLS', I.E. POTENTIAL VALUES.