Contingency Table Analysis of Categorical Variables (Crosstabulation)

Experimental Condition	Observed Frequency: Incorrect 0	Observed Frequency: Correct 1	Marginal Frequencies	Mean Proportion Correct -*p* (SD)
Control - 0	25	30	n_0 = 55	0.54545 (0.5025)
Treatment 1	20	25	n_1 = 45	0.55556 (0.5025)
Treatment 2	10	40	n_2 = 50	0.80000 (0.4041)
Marginal Frequencies	55	95	N = 150	0.63333 (0.4835)

Note. Standard Deviation of a proportion can be computed by $SD_p = \sqrt{(p(1-p)N)/(N-1)}$.

For example for the Condition Treatment 2, $SD_2 = \sqrt{(0.8(.2)50)/(49)}$ = 0.4041

Computation of Expected Values is completed by Multiplying the Row Marginal Total by the Column Marginal Frequency then divided by the Total (*N*).

Computation of Chi-Square (χ^2) and phi (ϕ) statistics

Expected Values	Incorrect 0	Correct 1			Incorrect 0	Correct 1
Control - 0	E = (55*55) / 150 E = 20.1667	E = (95*55) / 150 E = 34.8333	55	(O - E) $(O - E)^2$ / E	(25-20.167) = 4.833 $4.833^2/20.167$ = **1.158**	(30-34.833) = -4.833 $-4.833^2/34.833$ = **0.671**
Treatment 1	E = (55*45) / 150 E = 16.50	E = (95*45) / 150 E = 28.50	45	(O - E) $(O - E)^2$ / E	(20-16.50) = 3.50 $3.50^2/16.50$ = **0.742**	(25-28.50) = -3.50 $-3.50^2/28.50$ = **0.430**
Treatment 2	E = (55*50) / 150 E = 18.3333	E = (95*50) / 150 E = 31.6667	50	(O - E) $(O - E)^2$ / E	(10-18.333) = -8.333 $-8.333^2/18.333$ = **3.788**	(40-31.667) = 8.333 $-8.333^2/31.667$ = **2.193**
Marginal Totals	55	95	150			

$\chi^2 = \Sigma(O - E)^2/E$ = (1.158 + 0.671 + 0.742 + 0.430 +3.788 + 2.193) = 8.982. $df = (R - 1)(C - 1) = 2x1 = 2$

From Beasley & Schumacker (1995), $\phi^2 = \chi^2/N$ = 8.982/150 = 0.05988.

Thus $\phi = \sqrt{0.05988}$ = 0.2447. The ***p***-value associated with χ^2 = 8.982 (df = 2) is p = 0.011

These values can also be located in an SPSS Crosstabs analysis printout.

With the response variable being dichotomous (i.e., 0 = Incorrect or 1 = Correct), it is simple to demonstrate that using an ANOVA model with the dichotomous outcome as the dependent variable yields an identical statistic concerning the proportion of variance explained.

$SS_{Between} = \sum n_j (\bar{p}_j - \bar{p}_*)^2$ = $55(0.545-0.633)^2 + 45(0.556-0.633)^2 + 50(0.800-0.633)^2$ = 2.086.

where, $\bar{p}_*$ = 95/150 = 0.633.

$SS_{Within} = \Sigma(S_j^2)(n_j\text{-}1)$ = (0.5025x54) + (0.5025x44) + (0.4041x49) = 32.747

Since, $SD_p = \sqrt{(p(1-p)N)/(N-1)}$, $SSp = p(1\text{-}p)N$.

Thus, $SS_{Total} = (\bar{p}_*(1\text{-}\bar{p}_*)N)$ = (0.633x(0.367)x150 = 34.833.

ANOVA Source Table **$H_0: \mu_1 = \mu_2 = \mu_3$** or **$H_0: \pi_1 = \pi_2 = \pi_3$**

Source	*SS*	*df*	*MS*	*F*
Between (Explained)	2.086	J-1 = 2	2.086/2 = 1.043	1.043/0.223 = 4.682
Within (Error)	32.747	N-J = 150-3 = 147	32.747/149 = 0.223	
Total	34.833	N - 1 = 149	34.833/149 = 0.234	

η^2 = 2.086/34.833 = 0.0599.

Thus, $F(2,147) = 4.682$, $p = 0.011$, $\eta^2 = 0.0599$.

Moreover, $\eta^2 = \phi^2$, thus demonstrating the equivalence of the Contingency Table Analysis and ANOVA Models.