**Pearson *R* and Phi (*φ*)**

 Pearson *r* computed on two dichotomous variables is a phi coefficient. Have a look at the data at [Phi\_r.xlsx](http://core.ecu.edu/psyc/wuenschk/docs30/Phi_r.xlsx). Notice that both variables are dichotomous. When variable A has value 1, 5/15 = 33% of the cases have B = 1. When variable A has value 2, 8/15 = 53% of the cases have B = 2. When the value of A increases so does the value of B, so A and B are positively correlated.

To test the significance of such a phi coefficient one generally uses a chi-square statistic, which can be computed as . For the contingency table presented below,  (*r2* is the ratio of the regression *SS* to the total *SS*). This chi-square is evaluated on one degree of freedom. Do notice that the *p* value provided with the usual test of significance for a Pearson correlation coefficient is off a bit (.285 as compared to the .269 obtained from the chi-square). You should use the *p* value from chi-square.

**Regression**

****

****

**Correlations**

****

**φ = .202**

**Crosstabs**

****

|  |
| --- |
| **Chi-Square Tests** |
|  | Value | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 1.222a | 1 | .269 |
| Likelihood Ratio | 1.231 | 1 | .267 |

|  |
| --- |
| **Symmetric Measures** |
|  | Value | Approximate Significance |
| Nominal by Nominal | Phi | .202 | .269 |

 I find **phi** an appealing estimate of the magnitude of effect of the relationship between two dichotomous variables and **Cramér’s phi** appealing for use with tables where at least one of the variables has more than two levels.

|  |
| --- |
| **Risk Estimate** |
|  | Value | 95% Confidence Interval |
| Lower | Upper |
| Odds Ratio for A (1 / 2) | 2.286 | .522 | 10.011 |

 When A is 2 the odds that B is 2 are 8 to 7 = 1.143. When A is 1 the odds that B is 2 is 5 to 10 = .5. As A changes from 1 tp 2 the odds that B is 2 (rather than 1) increase. The ratio of these two conditional odds is the odds ratio – here (8/7)/(5/10) = 2.286. The odds of B being 2 are 2.286 times higher when A is 2 than when A is 1.

 An alternative analysis of the relationship between two dichotomous variables is a logistic regression or a log-linear analysis, both of which are based on a generalized linear model. Here is output from a logistic regression on these data:

|  |
| --- |
| **Omnibus Tests of Model Coefficients** |
|  | Chi-square | df | Sig. |
|  | Model | 1.231 | 1 | .267 |

 Notice that the chi-square from the logistic regression is the likelihood ratio chi-square reported earlier.

|  |
| --- |
| **Variables in the Equation** |
|  | B | S.E. | Wald | df | Sig. | Exp(B) |
| Step 1a | A | .827 | .754 | 1.203 | 1 | .273 | 2.286 |
| Constant | -1.520 | 1.212 | 1.574 | 1 | .210 | .219 |

 The slope for predicting B from A is .827. When we exponentiate that we get the odds ratio, 2.286.

 Here is output from the log-linear model:

|  |
| --- |
| **K-Way and Higher-Order Effects** |
|  | K | df | Likelihood Ratio | Pearson | Number of Iterations |
|  | Chi-Square | Sig. | Chi-Square | Sig. |
| K-way Effectsb | 1 | 2 | .535 | .765 | .512 | .774 | 0 |
| 2 | 1 | 1.231 | .267 | 1.222 | .269 | 0 |
| a. Tests that k-way and higher order effects are zero. |
| b. Tests that k-way effects are zero. |

 **Odds ratios** can also be very useful. Consider the results of some of my research on attitudes about animals (Wuensch, K. L., & Poteat, G. M. Evaluating the morality of animal research: Effects of ethical ideology, gender, and purpose. *Journal of Social Behavior and Personality*, 1998, *13,* 139-150. Participants were pretending to be members of a university research ethics committee charged with deciding whether or not to stop a particular piece of animal research which was alleged, by an animal rights group, to be evil. After hearing the evidence and arguments of both sides, 140 female participants decided to stop the research and 60 decided to let it continue. That is, the odds that a female participant would stop the research were 140/60 = 2.33. Among male participants, 47 decided to stop the research and 68 decided to let it continue, for odds of 47/68 = 0.69. The ratio of these two odds is 2.33 / .69 = 3.38. In other words, the women were more than 3 times as likely as the men to decide to stop the research.

|  |  |  |
| --- | --- | --- |
| Size of effect | *w* = φ | odds ratio |
| small | .1 | 1.49 |
| medium | .3 | 3.45 |
| large | .5 | 9 |

\*For a 2 x 2 table with both marginals distributed uniformly.

 Why form ratios of odds rather than ratios of probabilities? See my document [Odds Ratios and Probability Ratios](http://core.ecu.edu/psyc/wuenschk/StatHelp/OddsRatios.htm).

[Karl L. Wuensch](http://core.ecu.edu/psyc/wuenschk/klw.htm), July, 2021.