

Phi, Fei, Fo, Fum: Correlation Effect Sizes for Chi-squared Tests

Introduction

We review here the most common effect sizes for analysis of categorical variables, and introduce a new one – (Fei) – alongside R code from the `{effectsize}` package (Ben-Shachar, Lüdtke, and Makowski 2020) which implements these.

Tests Of Independence

The χ^2 test of independence between two categorical variables examines if the frequency distribution of one of the variables is dependent on the other.

Formally, the test examines how likely the observed conditional frequencies (cell frequencies) are under the null hypotheses of independence, by examining the degree of deviation of the observed cell frequencies from the frequencies expected if the variables were indeed independent.

The test statistic for these tests is the χ^2 , which is computed as:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where O_i are the *observed* frequencies and E_i are the frequencies *expected* under independence.

Instead of the deviations between the observed and expected frequencies, we can write χ^2 in terms of observed and expected cell probabilities and the total sample size N :

$$\chi^2 = N \sum \frac{(p_{O_i} - p_{E_i})^2}{p_{E_i}}$$

Where p_{O_i} are the *observed* cell probabilities and p_{E_i} are the probabilities *expected* under independence.

For example...

Phi

For a 2-by-2 contingency table analysis, the ϕ (*phi*) coefficient is a correlation-like measure of effect size indicating the strength of association between the two binary variables. One way to compute this effect size is to re-code the binary variables as dummy (0,1) variables, and computing the (absolute) Pearson correlation between them:

$$\phi = |r_{AB}|$$

Another way to compute ϕ is by using the χ^2 statistic:

$$\phi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\sum \frac{(p_{O_i} - p_{E_i})^2}{p_{E_i}}}$$

This value ranges between 0 (no association) and 1 (complete dependence), and its values can be interpreted the same as Person's correlation coefficient.

```
library(effectsize)
library(correlation)
```

```
data("RCT_table")
```

```
RCT_table
```

```
##           Group
## Diagnosis Treatment Control
## Sick           71         30
## Recovered      50        100
```

```
phi(RCT_table, adjust = FALSE)
```

```
## Phi |          95% CI
## -----
## 0.36 | [0.26, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
tidyr::uncount(as.data.frame(RCT_table), weights = Freq) |>
  transform(Diagnosis = Diagnosis == "Sick",
            Group = Group == "Treatment") |>
  correlation()
```

```
## # Correlation Matrix (pearson-method)
##
## Parameter1 | Parameter2 |    r |          95% CI | t(249) |          p
## -----
## Diagnosis |          Group | 0.36 | [0.25, 0.47] |    6.14 | < .001***
##
## p-value adjustment method: Holm (1979)
## Observations: 251
```

Cramér's V

These properties do not hold when the contingency table is larger than 2-by-2: $\sqrt{\chi^2/N}$ can be larger than 1.

Cramér showed (Cramér 1999) that while for 2-by-2 the maximal possible value of χ^2 is N , for larger tables the maximal possible value for χ^2 is $N \times (\min(k, l) - 1)$. Therefore, he suggested the V effect size (also sometimes known as Cramér's phi and denoted as ϕ_c):

$$\text{Cramer's } V = \sqrt{\frac{\chi^2}{N(\min(k, l) - 1)}}$$

V is 1 when the columns are completely dependent on the rows, or the row are completely dependent on the columns.

```
data("Music_preferences")
```

```
Music_preferences
```

```
##           Pop Rock Jazz Classic
## Psych 150  100  165    130
## Econ   50   65   35     10
## Law    2    55   40     25
```

```
cramers_v(Music_preferences, adjust = FALSE)
```

```
## Cramer's V |          95% CI
## -----
## 0.24        | [0.19, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

Tschuprow (REF) devised an alternative value, at , which is 1 only when the columns are completely dependent on the rows *and* the row are completely dependent on the columns, which is only possible when the contingency table is a square.

```
tschuprows_t(Music_preferences)
```

```
## Tschuprow's T |          95% CI
## -----
## 0.22          | [0.17, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

We can generalize both ϕ , V , and T to:

$$\sqrt{\frac{\chi^2}{\chi_{\max}^2}}$$

Also...

These can be used for confusion matrices...

Goodness of Fit

These tests compare an observed distribution of a multinomial variable to an expected distribution, using the same χ^2 statistic. Here too we can compute an effect size as $\sqrt{\frac{\chi^2}{\chi_{\max}^2}}$, all we need to find is χ_{\max}^2 .

Cohen's w

Cohen (REF) defined an effect size – w – for the goodness of fit test:

$$\text{Cohen's } w = \sqrt{\sum \frac{(p_{O_i} - p_{E_i})^2}{p_{E_i}}} = \sqrt{\frac{\chi^2}{N}}$$

Thus $\chi^2_{\max} = N$.

```
0 <- c(90, 10)
p_E <- c(0.5, 0.5)
cohens_w(0, p = p_E)
```

```
## Cohen's w |          95% CI
## -----
## 0.80      | [0.64, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

Unfortunately, w has an upper bound of 1 only when the variable is binomial (two categories) and the expected distribution is uniform ($p_{\text{class } 1} = p_{\text{class } 2} = 0.5$). If the distribution is none uniform (Rosenberg, 2010) or if there are more than 2 classes (Johnston et al., 2006), then $\chi^2_{\max} > N$, and so w can be larger than 1.

```
0 <- c(90, 10)
p_E <- c(0.35, 0.65)
cohens_w(0, p = p_E)
```

```
## Cohen's w |          95% CI
## -----
## 1.15      | [0.99, 1.36]
##
## - One-sided CIs: upper bound fixed at [1.36~].
```

```
0 <- c(10, 20, 80, 5)
p_E <- c(.25, .25, .25, .25)
cohens_w(0, p = p_E)
```

```
## Cohen's w |          95% CI
## -----
## 1.05      | [0.88, 1.73]
##
## - One-sided CIs: upper bound fixed at [1.73~].
```

Fei

We present here a new effect size, (Fei) which normalizes goodness-of-fit χ^2 by the proper χ^2_{\max} for non-uniform and/or multinomial variables.

The largest deviation from the expected probability would be if all observations are in the cell with the smallest expected probability.

$$p_O = \begin{cases} 1 & \text{if } p_i = \min(p) \\ 0 & \text{else} \end{cases}$$

Since $\chi^2 = N \sum \frac{(p_{E_i} - p_{O_i})^2}{p_{E_i}}$

We can find $\frac{(E_i - O_i)^2}{E_i}$ for each of these values:

$$\frac{(p_E - p_O)^2}{p_E} = \begin{cases} \frac{(p_i - 1)^2}{p_i} & \text{if } p_E = \min(p_E) \\ \frac{(p_i - 0)^2}{p_i} = p_i & \text{else} \end{cases}$$

Since $\sum p_i = 1$, therefor the χ^2 which is the sum of the expression above as:

$$\begin{aligned} \chi_{\max}^2 &= N \times (1 - \min(p_E) + \frac{(\min(p_E) - 1)^2}{\min(p_E)}) \\ &= N \times \frac{1 - \min(p_E)}{\min(p_E)} \\ &= N \times (\frac{1}{\min(p_E)} - 1) \end{aligned}$$

And so an effect size can be derived as:

$$\sqrt{\frac{\chi^2}{N \times (\frac{1}{\min(p_E)} - 1)}}$$

We call this effect size , which represents the voiceless bilabial fricative in the Hebrew language, keeping in line with ϕ (which in modern Greek marks the same sound) and V (which in English mark a voiced bilabial fricative) (W being derived from the letter V in modern Latin alphabet).

This will be 0 when the observed distribution matches the expected one perfectly, and will be 1 when the observed values are all of the same class - the one with the smallest expected probability.

```
0 <- c(90, 10)
p_E <- c(0.35, 0.65)
fei(0, p = p_E)
```

```
## Fei |          95% CI
## -----
## 0.85 | [0.73, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
0 <- c(10, 20, 80, 5)
p_E <- c(.25, .25, .25, .25)
fei(0, p = p_E)
```

```
## Fei |          95% CI
## -----
## 0.60 | [0.51, 1.00]
##
## - Adjusted for uniform expected probabilities.
## - One-sided CIs: upper bound fixed at [1.00].
```

When there are only 2 cells with uniform expected probabilities (50%), this expression reduces to N and $= w$.

```
0 <- c(90, 10)
p_E <- c(0.5, 0.5)

fei(0, p = p_E)
```

```
## Fei |          95% CI
## -----
## 0.80 | [0.64, 1.00]
##
## - Adjusted for uniform expected probabilities.
## - One-sided CIs: upper bound fixed at [1.00].
```

Summary

References

- Ben-Shachar, Mattan S., Daniel Lüdtke, and Dominique Makowski. 2020. “effectsize: Estimation of Effect Size Indices and Standardized Parameters.” *Journal of Open Source Software* 5 (56): 2815. <https://doi.org/10.21105/joss.02815>.
- Cramér, Harald. 1999. *Mathematical Methods of Statistics*. Vol. 43. Princeton University Press.