

Analysis of paired categorical data

E-mail distributed on 03-10-2022

Dear all,

Paired categorical data occur in a number of common situations, such as rater agreement, classification accuracy (i.e., confusion matrices), or interventional designs with a categorical pre-post measurement. In all these situations, we would have a square $K \times K$ frequency table, e.g., for a fictitious intervention study with emotion regulation choice as outcome (T1,T2):

T1	T2					
	Reappraisal	Relaxation	Acceptance	Rumination	Suppression	
Reappraisal	36	0	9	2	1	
Relaxation	5	22	15	1	0	
Acceptance	3	10	15	0	1	
Rumination	12	1	12	11	9	
Suppression	15	6	7	8	16	

Many metrics exist for quantifying the strength of the association in such matrices, e.g., **Cohen's Kappa** for rater agreement. As well, for the specific case of 2×2 paired categorical data, there is **McNemar's test** of marginal homogeneity, which tests whether the marginal distribution of choices differs between the row- and column-variable. In the above example, marginal heterogeneity would mean that the distribution of regulation choices changes between T1 and T2 (e.g., more reappraisal after intervention).

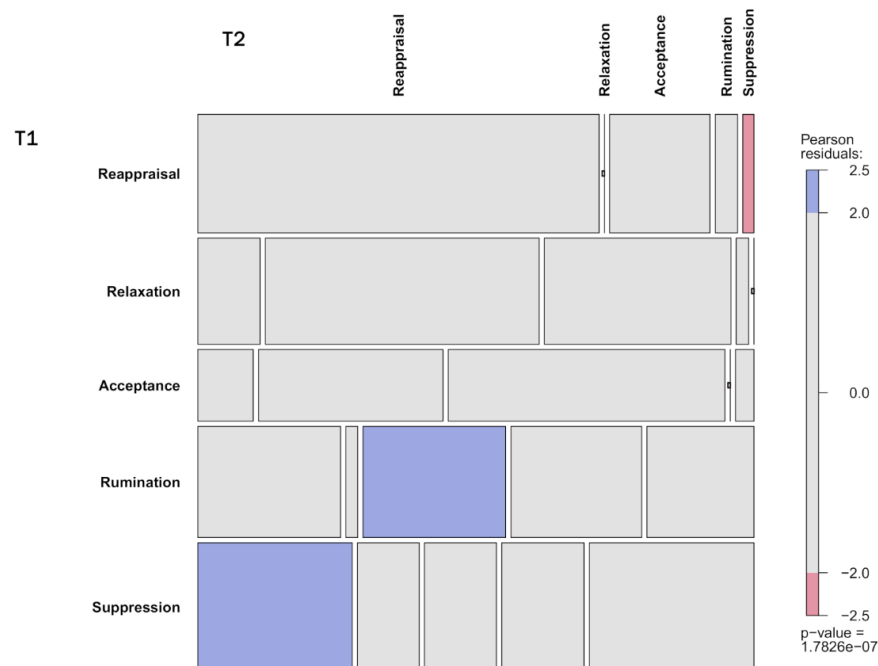
In general, there are a number of interesting tests that can be run on square frequency tables, which are not widely known to researchers but are variations of standard chi-square analysis:

- **Independence (chi-square):** The classical chi-square test for row-by-column independence. If the highest frequencies in the square table occur on the diagonal, then this test should be highly significant, and this is typically expected for paired data, so not a very informative analysis. If non-significant, it means there is no association between the row and column variables (e.g., random guessing, no agreement, no intervention effect).
- **Quasi-independence:** Variation of the chi-square test which analyzes only the off-diagonal frequencies. A non-significant test would mean that, excluding the diagonal frequencies, the pattern of off-diagonal confusions is independent (random). For data where specific confusion patterns are expected (e.g., anger and disgust confusion in emotion recognition), quasi-independence will normally be rejected.
- **Symmetry:** Analyzes if the off-diagonal frequencies are identical between the lower and upper triangle of the square table. A non-significant test here would mean that confusion is identical in both directions (e.g., anger is confused for disgust as often as disgust for anger). If

significant, it means that some confusions are not identical in both directions. However, the symmetry model also forces marginal homogeneity, which may be violated by design in some data (or may not be of interest).

- **Quasi-symmetry:** Analyzes symmetry while allowing marginal heterogeneity. Otherwise same interpretation as the symmetry model.
- **Marginal homogeneity:** Analyzes whether the marginal distribution of choices differs between the row and column variable. For rater agreement and confusion matrices, we normally wish to accept marginal homogeneity (indicating no significant disagreement). For interventional designs, we normally wish to reject marginal homogeneity, such that the shift in marginal distribution between pre- and post reflects the effect of the intervention.

All of these models are subsumed under **log-linear analysis**, and are in fact special cases of Poisson regression.* The attached R function can be run on any square frequency matrix and will return these 5 tests. In addition, the function returns fitted values and standardized Pearson residuals for the first 4 models, which can be inspected as a follow-up to a significant test, to clarify the pattern of confusions.



Large positive Pearson residuals indicate frequencies that are higher than expected under the fitted model, whereas large negative residuals indicate frequencies that are smaller than expected under the fitted model. Mosaic plots can further help to identify which cells of the tables helped reject the null hypothesis (see attached example for the symmetry model).

Best,
Ben

* In this framework, even more complex analyses can be accommodated, such as comparing square tables between several conditions.

Ben Meuleman, Ph.D.

Statistician

Swiss Center for Affective Sciences

University of Geneva | Campus Biotech

Chemin des Mines 9 | CH-1202 Genève

ben.meuleman@unige.ch | +41 (0)22 379 09 79

```
#####
## TESTS ON SQUARE CONTINGENCY TABLES
#####

square.tests <- function(V,correct=FALSE) {

  raw <- as.matrix(V)
  dims <- dim(raw)[1]
  dt <- as.data.frame.table(raw)
  dt$Diag <- factor(cumsum(ifelse(as.character(dt$Var1)==
    as.character(dt$Var2),1,0))*ifelse(as.character(dt$Var1)==
    as.character(dt$Var2),1,0))
  dt$Symm <- factor(paste(pmin(as.numeric(dt$Var1),as.numeric(dt$Var2)),
    pmax(as.numeric(dt$Var1),as.numeric(dt$Var2)), sep = "-"))

  if(correct==TRUE) { dt$Freq <- ifelse(dt$Freq==0,0.0001,dt$Freq) }

  indep <- glm(Freq~Var1+Var2, data=dt, family=poisson)
  qindep <- glm(Freq~Var1+Var2+Diag, data=dt, family=poisson)
  symm <- glm(Freq~Symm, data=dt, family=poisson)
  qsymm <- glm(Freq~Var1+Symm, data=dt, family=poisson)
  mhom <- anova(symm, qsymm, test="Chisq")

  fitted <- list(Independence=matrix(indep$fitted,nrow=dims),
    Qindependence=matrix(qindep$fitted,nrow=dims),
    Symmetry=matrix(symm$fitted,nrow=dims),
    Qsymmetry=matrix(qsymm$fitted,nrow=dims))

  residuals <- list(
    Independence=matrix(resid(indep,type="pearson"),nrow=dims)/
    sqrt(1-lm.influence(indep,do.coef=FALSE)$hat),
    Qindependence=matrix(resid(qindep,type="pearson"),nrow=dims)/
    sqrt(1-lm.influence(qindep,do.coef=FALSE)$hat),
    Symmetry=matrix(resid(symm,type="pearson"),nrow=dims)/
    sqrt(1-lm.influence(symm,do.coef=FALSE)$hat),
    Qsymmetry=matrix(resid(qsymm,type="pearson"),nrow=dims)/
    sqrt(1-lm.influence(qsymm,do.coef=FALSE)$hat))

  diag(residuals$Qindependence) <- 0
  diag(residuals$Symmetry) <- 0
  diag(residuals$Qsymmetry) <- 0

  tests <- data.frame(Test=c("Independence",
```

```

      "Quasi-independence", "Symmetry", "Quasi-symmetry", "Marginal homogeneity"),
      LRT=c(indep$deviance, qindep$deviance, symm$deviance, qsymm$deviance,
      mhom$Deviance[2]), DF=c(indep$df.residual, qindep$df.residual,
      symm$df.residual, qsymm$df.residual, mhom$Df[2]))

tests$P <- 1-pchisq(tests$LRT, tests$DF)
print(tests)
invisible(list(Data=dt, Tests=tests, Fitted=fitted, Residuals=residuals))
}

```