RELAN - Relation Analysis

RELAN Statistics

Since both the elementary and the complex statistical evaluations in the RELAN program are based on dichotomous data, all kinds of statistics can be clearly visually displayed using a four-field contingency table (see also at the end of this text). All four resulting frequencies concerning the combinations of manifestations of two variables (X, Y) are symbolized by F1, F2, F3, F4 and the corresponding expected values by E1, E2, E3, E4, while F13, F14, F13, F24 are the boundary sums of the contingency cells and N is the sample size. For the calculation of the statistics of a single (complex) relation, only its frequency F in positive minterms (minterms: all combinations of variable values), its expected value E, its simulated optimal value O, as well as their sample size N is needed for the statistical evaluation.

T: Truth values of the hypothetical relation

F: Observed frequencies of cases assigned to the minterms

O: Relation-optimal frequencies (simulated values)

E: Expected values (values expected by chance)

RE: Relative frequencies of expected values (E/N)

RF: Relative frequencies of observed values (F/N)

SD: Standard deviation of the binomial function ($\sqrt{}$(RE*(1-RE)*N))

Z: Z-value of the approximation of the binomial function (F – E)/SD

P(Z): P-value of the Z-value of the approximation of the binomial function

Z-T: Z-value of the tested or simulated relation

***P-T**: P-value of the* Z-value of the tested or simulated relation

HI = ld(1/RF): Information content of the single manifestations in the minterms

CH2 = $(F - E)^2/E$: Chi-square of each minterm. The sum of CH2 is asymptotically chi-square distributed, df = (2**M)- M-1.

LR(C) = F· *ln*(F/E): Likelihood-Ratio-Statistic (deviation of the observed frequencies from the expected frequencies). The double of the sum of LR(C) is asymptotically chi-square distributed, df = (2**M)- M-1

LR(R) = F· ln(F/O): Likelihood-Ratio-Statistic (deviation of the observed frequencies from the simulated frequencies). The double of the sum of LR(R) is asymptotically chi-square distributed, df = (2**M)- M-1

PTZ= (F – E)/E: Proportional increase of hits. PTZ_{max} = (N - E)/E; PTZ_{min} = (0 – E)/E

=-1.

PTZM = (F – E)/(N - E): Relative proportional increase of hits in relation to the possible maximum of hits (Max. = 1)

PTZR = $(PTZ-PTZ_{min})/(PTZ_{max}-PTZ_{min})$: Relative proportional increase of hits in relation to the possible minimum and to the possible maximum of hits (Min. = 0, Max. = 1)

PREC (precision): This statistic quantifies the explanatory value of a relation (Max. = 100, Min. = 0)

POWE: Power score (by using the standard deviation of the null hypothesis)

POWF: Power score (by using the standard deviation of the alternative hypothesis)

Z: (Unnormalised) Z-value of the approximated binomial test

P(Z): P-value of the (unnormalised) Z-value of the approximated binomial test

Z*: Normalised Z-value of the approximated binomial test. This statistic allows a fair comparison of relations which have different numbers of truth values

P(Z*): P-value of the normalised Z-value of the approximated binomial test

LR(cha): Likelihood-Ratio-Statistic of the deviation of the hypothetical relation from a relation by chance (deviation of observed frequency from expected frequency). The double of LR(cha) is asymptotically chi-square distributed, df = 1

LR(rel): Likelihood-Ratio-Statistic of the deviation of the hypothetical relation from the simulated optimal relation (deviation of observed frequency from simulated frequency). The double of LR(rel) is asymptotically chi-square distributed, df = 1

CHI2**: Chi-square test for the hypothetical relation

F+: Number of observed cases in minterms with positive truth values of the relation

E+: Number of expected cases in minterms with positive truth values of the relation

F+ - E+: Difference of cases between observed and expected cases (in minterms with positive truth values oft the relation)

POWER (freq): Power score for the hypothetical relation (by using the standard deviation of the null hypothesis)

POWER(expe): Power score for the hypothetical relation (by using the standard deviation of the alternative hypothesis)

R+(%): Percent score of cases which are found in minterms with positive truth values (R+(%) = PTZR · 100) (Max. = 100, Min. = 0)

T-F: Relative frequency of minterms in which positive truth values are in accordance with cases and non-positive minterms with no cases (Max. = 1, Min = 0).

T-Z: Difference between correspondence (+T+Z) (-T-Z) and non-correspondence (+T-Z) (-T+Z) of truth scores and z-values in the mintermes (+T+Z = relative

frequency of cases in positive minterms with positive Z-values; -**T-Z** = relative frequency of cases in non-positive minterms with negative z-values (Max = 1.0, Min = -1.0)

Z(T) = Mean among all absolute Z-values of minterms

SENS (Sensitivity): Relative frequency of cases with positive Z-values which also have positive truth values (Max = 1.0, Min = 0.0).

SPEZ (Specificity): Relative frequency of cases with negative Z-values which have also non-positive truth values (Max = 1.0, Min = 0.0).

PPV (positive predictive value): Relative frequency of cases with positive truth values which also have positive Z-values (Max = 1.0, Min = 0.0).

NPV (negative predictive value): Relative frequency of cases with non-positive truth values which also have negative z-values (Max = 1.0, Min = 0.0).

D(Cohen): Cohens effect size $(Z_{pos} - Z_{neg})/SD_Z$)

R(Cohen): Cohens proposal of transforming the effect size to a correlation statistic

CA(relation): Test of causality for each subrelation and the final relation (yes: relation can be accepted in accordance to the preconditions of causality)

LR(F-E): Likelihood-Ratio-Statistic of the deviation of the observed frequencies and expected (by chance) frequencies in the minterms of the hypothetical relation. The double of LR(F-E) is asymptotically chi-square distributed, df = (2**M)-M-1

LR(F-O): Likelihood-Ratio-Statistic of the deviation of the observed frequencies and the (optimal) simulated frequencies in the minterms of the hypothetical relation. The double of LR(cha) is asymptotically chi-square distributed; df = (2**M)-M-1

P(R/D): The probability of a relation under the condition of the data. This statistic increases the more the data are „included" in the relation (Max. = 1.0, Min. = 0.0).

P(D/R): The probability of the data under the condition of the hypothesised relation. This statistic increases the more the relation is „included" in the data (Max = 1.0, Min = 0.0).

Z(Imp): Z-value of the right implication

Z(Bij): Z-value of the bijunction

PTZ-B: PTZ-value of the bijunction

PTZM-B: PTZM-value of the bijunction

P(Y/X): Probability of Y under the condition of X

p(Y=X): (F1 + F4)/N. Probability (relative frequency) of equivalence between Y and X

ER1 = p(Y/X): F4/(F34) Experimental event rate (effect condition) which is equal to the positive predictive value (PPV)

ER0 = F2/(F12): Control event rate (control condition)

ERD = ER1 - ER0: Event rate difference

RED = (ER1 - ER0)/ER0: Relative event rate difference

ERQ = ER1/ER0: Event rate quotient (event rate ratio)

OD1 = F4/F3: Effect event odd. Ratio between effect (Y = 1) and no-effect (Y = 0) under the condition of X (X = 1)

OD0 = F2/F1: Control event odd. Ratio between effect (Y = 1) and no-effect (Y = 0) under the condition of X is not given (X = 0).

OR = OD1/OD0: Odds ratio = Cross-Product Ratio: (F1 · F4)/(F2 · F3)

Q = (OR – 1)/(OR + 1): Yules Q

MI = Σ[p(x,y) · ld(p(x,y)/p(x) · p(y))]: Mutual Information (transinformation) between X and Y

PHI = ((F1 · F4) - (F2 · F3))/SQRT(F12 · F13 · F23 · F24): Correlation (Pearson)

CHIQ: Chi-square statistic

C: Contingency coefficient

V: Cramers V

LR+ = Sensitivity / (1 - Specificity): „Likelihood ratio positive"

LR- = (1 – Sensitivity) / Specificity: „Likelihood ratio negative"

LAMB (X > Y): (Lambda = (p error apriori - p error aposteriori) / (p error apriori)

SEN = F4/F24: Sensitivity (probability for Y = 1 under the condition of X = 1)

SPE = F1/F13: Specificity (probability for Y = 0 under the condition of X = 0)

PPV: Positive predictive value

NPV: Negative predictive Value

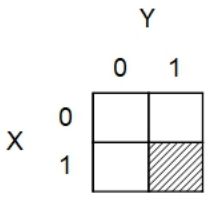
PRAE: Precision

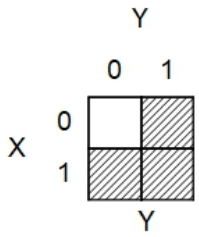
BESD = (0.5 + r/2) – (0.5 - r/2): Binomial Effect Size Display

Appendix: Graphical Illustration of the Basic Logical Functions

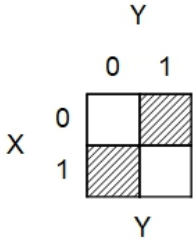
The different frequencies in the cells of the four-field contingency table of two variables characterises the kind of statistically defined logical function (relation) between two variables. If there are in one cell, in two or three (shaded) cells significantly more cases than in the other cells then the corresponding logical function should be called significantly true.

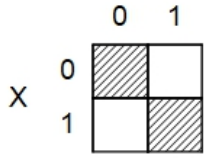
The same principle is used for the definition and verification of complex relations by analysing multidimensional contingency tables regarding their frequency distribution. If there are more cases in the cells (hatched cells) of the adjunctive normal form than expected by chance, then the respectively relation should be called statistically verified.

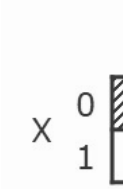
Conjunction 

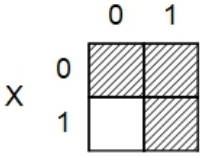
Adjunction 

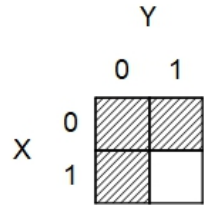
NAND 

Disjunction 

Bijunction 

NOR 

Implication 

Nonjunction 

Hatched cells symbolize overall a significantly higher cell frequency than would be expected by chance.

REFERENCE

Maderthaner, R. (2022). RELATIONSANALYSE (RELAN) - Aussagenlogische, statistische und kausale Analyse von Daten. Springer Spektrum, Heidelberg.