The Strength of Categorical Data Analysis

Bernard S. Gorman
SUNY Distinguished Professor
Nassau Community College and
Adjunct Professor, Hofstra University

American Statistical Association Refresher Course
New York, March 28, 2011

Previous version delivered with
Louis H. Primavera
The Derner Institute Adelphi University
Eastern Psychological Association
2003
Baltimore, MD

Abstract: Many years ago, S.S. Stevens and others drew a distinction between categorical (nominal) scale data and ordinal, interval, and ratio scale data. While categorical data imposes fewer restrictions than other data forms, it is not an “inferior” form of data. In fact, many phenomena of interest to psychologists are naturally categorical. Unfortunately, the statistical training of most psychologists emphasizes data models and methods geared toward interval and ratio scale data. While some statistic courses give cursory coverage of contingency tables and a few nonparametric techniques, most psychologists are unaware of the rich possibilities that other research fields have enjoyed through the use of categorical data method. Advances in both statistical theory and software development can deliver the promises of strong categorical analyses.

The symposium will show participants how to gain the most information from two-way categorical contingency table analyses and will then provide glimpses and overviews of advanced categorical procedures such as loglinear modeling, categorical logistic regression, configural frequency analysis, correspondence analysis, categorical regression trees, and latent class analyses.

The Strength of Categorical Data
I. Levels of Data
   A. Stevens’ Typology
      1. Nominal
      2. Ordinal
      3. Interval
      4. Ratio
      5. Unfortunately, nominal became thought of as "inferior"
      6. Realistically, many phenomena of interest are naturally categorical and many things considered to be at other levels are actually nominal
   B. Coombs' Data Theory
   C. The nonparametric movement
      1. Fisher
      2. Siegel
      3. Goodman
      4. Kruskal
      5. Wilkinson
      6. Unfortunately, these models often have limited power as direct replacements for parametric techniques. However, they are probably the best ones when assumptions about levels of measurement, normality, and random assignment are violated.

II. A Simple Two-way Contingency Table
   A. Layout of table
      1. Rows
      2. Columns
      3. Cells
      4. Marginals
   B. Cell Contents
      1. Raw counts
         a. Observed frequencies
      2. Row Percents
      3. Column Percents
      4. Total Percents
   C. Marginals
   D. Some Simple probability theory
      1. Probability of independent events
         a. "Chance" events
         b. Our "model" of independence
            (1) Model formula: Freq = A + B + e
      2. Expected proportions
      3. Expected Frequencies
   E. What if Independence doesn't hold?
1. Observed frequencies will be different from expected frequencies
2. Row scores will be correlated with column scores

F. How do we know if independence is not tenable?
1. The chi square statistic
2. The chi-square distribution
3. Degrees of freedom
4. The chi-square table

G. Interpreting cells
1. Residuals
2. Standardized residuals
3. Adjusted standardized residuals

H. Strength of effects
1. Contingency coefficient
2. Phi
3. Cramer’s V
4. Odds Ratios
   a. Relative risk
   b. Logarithmic Odds
5. Asymmetrical measures
   a. Somer’s D
   b. Goodman and Kruskal’s Lambdas
6. Matching Coefficients
   a. Kappa
   b. Others
7. Ordinal Measures
   a. Gamma
   b. Tau

III. The Loglinear Model
A. The cross-tab as a prediction model
B. The General Linear model
C. The Saturated Hierarchical Loglinear model
1. Frequency = A + B + AB
2. Reduces to Frequency = AB
3. Logarithmic form = \log(Freq) = \log(A) + \log(b) + \log(AB)

D. Unsaturated models
E. Converting loglinear models to logit models
1. Logs odds (b) = b*\log(a) + E
2. Rules for Converting loglinear to logit
IV. Logit Models
   A. Additive effects
   B. Interactions
      1. Moderator variables

V. Prediction Logic

VI. Regression and Classification Trees

VII. Configural Frequency Analysis

VIII. Correspondence Analysis

IX. Item Response Theory and Binary Factor Analysis

X. Latent Class Analysis

XI. Path and Graphical Models

XII. Sources of Free Software

XIII. References
Examples and Exhibits
## Overall Taxonomy

### A Taxonomy of Categorical Methods

<table>
<thead>
<tr>
<th></th>
<th>Set I</th>
<th>Set II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One Manifest</td>
<td>One Manifest</td>
</tr>
<tr>
<td>Variable Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Manifest</td>
<td>Frequency Distribution Information statistics</td>
<td>Chi-Square Prediction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic Regression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Path Models</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Graphical Modeling</td>
</tr>
<tr>
<td>Several Manifest</td>
<td>Configural Frequency Analysis</td>
<td>Loglinear Modeling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic Regression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Regression Trees</td>
</tr>
<tr>
<td>One Latent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Several Latent</td>
<td></td>
<td>Binary Factor Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple Correspondence Analysis Latent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Class</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Path Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Graphical Models</td>
</tr>
</tbody>
</table>
Our Basic Tables
Simple Crosstab

Gender * TV Show Preference Crosstabulation

<table>
<thead>
<tr>
<th></th>
<th>TV Show</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sports</td>
<td>Romance</td>
</tr>
<tr>
<td>Gender</td>
<td>Count</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Expected Count</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>70.0%</td>
</tr>
<tr>
<td></td>
<td>% within TV Show Preference</td>
<td>70.0%</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>35.0%</td>
</tr>
<tr>
<td>Femal</td>
<td>Count</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Expected Count</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>30.0%</td>
</tr>
<tr>
<td></td>
<td>% within TV Show Preference</td>
<td>30.0%</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>15.0%</td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Expected Count</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>50.0%</td>
</tr>
<tr>
<td></td>
<td>% within TV Show Preference</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>50.0%</td>
</tr>
</tbody>
</table>

Chi-Square Tests

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>df</th>
<th>Asymp. Sig. (2-sided)</th>
<th>Exact Sig. (2-sided)</th>
<th>Exact Sig. (1-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>32.000(b)</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuity</td>
<td>30.420</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>32.913</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fisher's Exact Test</td>
<td></td>
<td></td>
<td></td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>Linear-by-Linear</td>
<td>31.840</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Association</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N of Valid Cases</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Computed only for a 2x2 table
b 0 cells (.0%) have expected count less than 5. The minimum expected count is 50.00.
### Directional Measures

<table>
<thead>
<tr>
<th>Nominal by</th>
<th>Value</th>
<th>Asymp. Std.</th>
<th>Approx. T(b)</th>
<th>Approx. Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambda</td>
<td>.400</td>
<td>.074</td>
<td>4.558</td>
<td>.000</td>
</tr>
<tr>
<td>Gender</td>
<td>.400</td>
<td>.077</td>
<td>4.170</td>
<td>.000</td>
</tr>
<tr>
<td>TV Show Preference</td>
<td>.400</td>
<td>.077</td>
<td>4.170</td>
<td>.000</td>
</tr>
<tr>
<td>Goodman and Kruskal tau</td>
<td>.160</td>
<td>.052</td>
<td>.000(c)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>.160</td>
<td>.052</td>
<td></td>
<td>.000(c)</td>
</tr>
<tr>
<td>TV Show Preference</td>
<td>.160</td>
<td>.052</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* a Not assuming the null hypothesis.
* b Using the asymptotic standard error assuming the null hypothesis.
* c Based on chi-square approximation.

### Symmetric Measures

<table>
<thead>
<tr>
<th>Nominal by</th>
<th>Value</th>
<th>Approx. Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phi</td>
<td>.400</td>
<td>.000</td>
</tr>
<tr>
<td>Cramer’s V</td>
<td>.400</td>
<td>.000</td>
</tr>
<tr>
<td>Contingency</td>
<td>.371</td>
<td>.000</td>
</tr>
</tbody>
</table>

N of Valid Cases

<table>
<thead>
<tr>
<th>Value</th>
<th>95% Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>Odds Ratio for Gender (Male / Female)</td>
<td>5.444</td>
</tr>
<tr>
<td>For cohort TV Show Preference = Sports</td>
<td>2.333</td>
</tr>
<tr>
<td>For cohort TV Show Preference = Romance</td>
<td>.429</td>
</tr>
<tr>
<td>N of Valid Cases</td>
<td>200</td>
</tr>
</tbody>
</table>
### Complex Table: Gender * TV Show Preference * Age Crosstabulation

<table>
<thead>
<tr>
<th>Age</th>
<th>Sports</th>
<th>Romance</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected Count</td>
<td>40</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>% within Gender</td>
<td>80.0%</td>
<td>20.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% within TV Show Preference</td>
<td>80.0%</td>
<td>20.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>% of Total</td>
<td>40.0%</td>
<td>10.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>Adjusted Residual</td>
<td>6.0</td>
<td>-6.0</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected Count</td>
<td>25.0</td>
<td>25.0</td>
<td>50.0</td>
</tr>
<tr>
<td>% within Gender</td>
<td>20.0%</td>
<td>80.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% within TV Show Preference</td>
<td>20.0%</td>
<td>80.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>% of Total</td>
<td>10.0%</td>
<td>40.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>Adjusted Residual</td>
<td>-6.0</td>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected Count</td>
<td>50.0</td>
<td>50.0</td>
<td>100.0</td>
</tr>
<tr>
<td>% within Gender</td>
<td>50.0%</td>
<td>50.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% within TV Show Preference</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% of Total</td>
<td>50.0%</td>
<td>50.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Adjusted Residual</td>
<td>2.0</td>
<td>-2.0</td>
<td></td>
</tr>
<tr>
<td>Older</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected Count</td>
<td>25.0</td>
<td>25.0</td>
<td>50.0</td>
</tr>
<tr>
<td>% within Gender</td>
<td>60.0%</td>
<td>40.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% within TV Show</td>
<td>60.0%</td>
<td>40.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>% of Total</td>
<td>30.0%</td>
<td>20.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>Adjusted Residual</td>
<td>2.0</td>
<td>-2.0</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected Count</td>
<td>25.0</td>
<td>25.0</td>
<td>50.0</td>
</tr>
<tr>
<td>% within Gender</td>
<td>40.0%</td>
<td>60.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% within TV Show</td>
<td>40.0%</td>
<td>60.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>% of Total</td>
<td>20.0%</td>
<td>30.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>Adjusted Residual</td>
<td>-2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected Count</td>
<td>50.0</td>
<td>50.0</td>
<td>100.0</td>
</tr>
<tr>
<td>% within Gender</td>
<td>50.0%</td>
<td>50.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% within TV Show</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>% of Total</td>
<td>50.0%</td>
<td>50.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
### Chi-Square Tests

<table>
<thead>
<tr>
<th>Age</th>
<th></th>
<th>Value</th>
<th>df</th>
<th>Asymp. Sig. (2-sided)</th>
<th>Exact Sig. (2-sided)</th>
<th>Exact Sig. (1-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger</td>
<td>Pearson Chi-</td>
<td>36.000(b)</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Continuity</td>
<td>33.640</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Likelihood</td>
<td>38.549</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fisher's Exact</td>
<td></td>
<td></td>
<td>.000 .000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Linear-by-Linear</td>
<td>35.640</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older</td>
<td>N of Valid</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pearson Chi-</td>
<td>4.000(b)</td>
<td>1</td>
<td>.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Continuity</td>
<td>3.240</td>
<td>1</td>
<td>.072</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Likelihood</td>
<td>4.027</td>
<td>1</td>
<td>.045</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fisher's Exact</td>
<td></td>
<td></td>
<td>.071 .036</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Linear-by-Linear</td>
<td>3.960</td>
<td>1</td>
<td>.047</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N of Valid</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a  Computed only for a 2x2 table  
b  0 cells (.0%) have expected count less than 5. The minimum expected count is 25.00.

### Symmetric Measures

<table>
<thead>
<tr>
<th>Age</th>
<th></th>
<th>Value</th>
<th>Appro x. Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger</td>
<td>Nominal by</td>
<td>Phi</td>
<td>.600 .000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cramer's V</td>
<td>.600 .000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contingency</td>
<td>.514 .000</td>
</tr>
<tr>
<td></td>
<td>N of Valid Cases</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Older</td>
<td>Nominal by</td>
<td>Phi</td>
<td>.200 .046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cramer's V</td>
<td>.200 .046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contingency</td>
<td>.196 .046</td>
</tr>
<tr>
<td></td>
<td>N of Valid Cases</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

a  Not assuming the null hypothesis.  
b  Using the asymptotic standard error assuming the null hypothesis.
### Risk Estimate

<table>
<thead>
<tr>
<th>Age</th>
<th>Value</th>
<th>95% Confidence</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>Younger</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Odds Ratio for Gender (Male / For cohort TV Show)</td>
<td>16.000</td>
<td>6.005</td>
<td>42.631</td>
<td></td>
</tr>
<tr>
<td>For cohort TV Show</td>
<td>4.000</td>
<td>2.259</td>
<td>7.083</td>
<td></td>
</tr>
<tr>
<td>For cohort TV Show</td>
<td>.250</td>
<td>.141</td>
<td>.443</td>
<td></td>
</tr>
<tr>
<td>Older</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Odds Ratio for Gender (Male / For cohort TV Show)</td>
<td>2.250</td>
<td>1.011</td>
<td>5.008</td>
<td></td>
</tr>
<tr>
<td>For cohort TV Show</td>
<td>1.500</td>
<td>.997</td>
<td>2.256</td>
<td></td>
</tr>
<tr>
<td>For cohort TV Show</td>
<td>.667</td>
<td>.443</td>
<td>1.003</td>
<td></td>
</tr>
<tr>
<td>N of Valid</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Odds Ratio for Gender (Male / For cohort TV Show)</td>
<td>2.250</td>
<td>1.011</td>
<td>5.008</td>
<td></td>
</tr>
<tr>
<td>For cohort TV Show</td>
<td>1.500</td>
<td>.997</td>
<td>2.256</td>
<td></td>
</tr>
<tr>
<td>For cohort TV Show</td>
<td>.667</td>
<td>.443</td>
<td>1.003</td>
<td></td>
</tr>
<tr>
<td>N of Valid</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Loglinear Model
Essential Formulae for Log-linear Analysis

I. Logarithmic Form

$$\log(n_{ij}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

$$n_{ij} = \text{actual cell frequency}$$

$$\lambda_i^A = \text{effect for category } i \text{ of Factor } A \text{ log units}$$

$$\sum_1^C \lambda_i^A = 0$$

II. Frequency Form

$$\eta = \sqrt{\prod n_{ij}}$$

$$\tau = e^\Lambda$$

$$\Lambda = \log(\tau)$$

$$n_{ij} = \eta_0 \tau_i^A \tau_j^B \tau_{ij}^{AB}$$

III. Test of Fit

$$G^2 = 2 \sum (O \ln \frac{O}{E})$$
Simple Loglinear Model
LEM: log-linear and event history analysis with missing data.
Developed by Jeroen Vermunt (c), Tilburg University, The Netherlands.
Version 1.0 (September 18, 1997).
*** INPUT ***
*Simp  e Loglinear Model
* GenTV.dat
man 2
dim 2 2
lab G T
mod {TG}
rec 200
dat genTV.dat
*** STATISTICS ***
Number of iterations = 2
Converge criterion = 0.0000000000
X-squared = 0.0000 (0.0000)
L-squared = 0.0000 (0.0000)
Cressie-Read = 0.0000 (0.0000)
Dissimilarity index = 0.0000
Degrees of freedom = 0
Log-likelihood = -260.80230
Number of parameters = 3 (+1)
Sample size = 200.0
BIC(L-squared) = 0.0000
AIC(L-squared) = 0.0000
BIC(log-likelihood) = 537.4995
AIC(log-likelihood) = 527.6046
Eigenvalues information matrix
280.0000 167.9798 120.0000
*** FREQUENCIES ***
G T observed estimated std. res.
1 1 70.000 70.000 0.000
1 2 30.000 30.000 0.000
2 1 30.000 30.000 0.000
2 2 70.000 70.000 0.000
*** LOG-LINEAR PARAMETERS ***
* TABLE GT [or P(GT)] *
effect  main  beta  std err  z-value  exp(beta)  Wald  df  prob
G     1   -0.0000    0.0772    0.000    1.0000
2   -0.0000                         1.0000
T     1   -0.0000    0.0772    0.000    1.0000
2   -0.0000                         1.0000
GT    1 1  0.4236    0.0772   5.491   1.5275
1 2  -0.4236                         0.6547
2 1  -0.4236                         0.6547
2 2   0.4236                         1.5275
*** (CONDITIONAL) PROBABILITIES ***
* P(GT) *
1 1 0.3500 (0.0337)
1 2 0.1500 (0.0252)
2 1 0.1500 (0.0252)
2 2 0.3500 (0.0337)
Complex Table
LEM: log-linear Model
*** INPUT ***
*loglinear
* Clasex dat
man 3
dim 2 2 2
lab A G T
mod {AGT}
rec 8
rco
dat crosex1.dat

*** STATISTICS ***
Number of iterations = 2
Converge criterion = 0.0000000000
X-squared = 0.0000 (0.0000)
L-squared = 0.0000 (0.0000)
Cressie-Read = 0.0000 (0.0000)
Dissimilarity index = 0.0000
Degrees of freedom = 0
Log-likelihood = -394.60028
Number of parameters = 7 (+1)
Sample size = 200.0
BIC(L-squared) = 0.0000
AIC(L-squared) = 0.0000
BIC(log-likelihood) = 826.2888
AIC(log-likelihood) = 803.2006

*** FREQUENCIES ***
A G T observed estimated std. res.
1 1 1 40.000 40.000 0.000
1 1 2 10.000 10.000 0.000
1 2 1 10.000 10.000 0.000
1 2 2 40.000 40.000 0.000
2 1 1 30.000 30.000 0.000
2 1 2 20.000 20.000 0.000
2 2 1 20.000 20.000 0.000
2 2 2 30.000 30.000 0.000
### LOG-LINEAR PARAMETERS

**TABLE AGT [or P(AGT)]**

<table>
<thead>
<tr>
<th>effect</th>
<th>beta</th>
<th>std err</th>
<th>z-value</th>
<th>exp(beta)</th>
<th>Wald</th>
<th>df</th>
<th>prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>main</td>
<td>3.0971</td>
<td></td>
<td></td>
<td>22.1336</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.1014</td>
<td>0.0807</td>
<td>-1.256</td>
<td>0.9036</td>
<td></td>
<td></td>
<td>0.209</td>
</tr>
<tr>
<td>2</td>
<td>0.1014</td>
<td></td>
<td></td>
<td>1.1067</td>
<td>1.58</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0000</td>
<td>0.0807</td>
<td>0.000</td>
<td>1.0000</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>0.0000</td>
<td></td>
<td></td>
<td>1.0000</td>
<td>0.00</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0000</td>
<td>0.0807</td>
<td>0.000</td>
<td>1.0000</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>0.0000</td>
<td></td>
<td></td>
<td>1.0000</td>
<td>0.00</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>AG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>-0.0000</td>
<td>0.0807</td>
<td>0.000</td>
<td>1.0000</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>1 2</td>
<td>0.0000</td>
<td></td>
<td></td>
<td>1.0000</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>2 1</td>
<td>0.0000</td>
<td></td>
<td></td>
<td>1.0000</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>2 2</td>
<td>0.0000</td>
<td></td>
<td></td>
<td>1.0000</td>
<td>0.00</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>AT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>-0.0000</td>
<td>0.0807</td>
<td>0.000</td>
<td>1.0000</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>1 2</td>
<td>0.0000</td>
<td></td>
<td></td>
<td>1.0000</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>2 1</td>
<td>0.0000</td>
<td></td>
<td></td>
<td>1.0000</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>2 2</td>
<td>0.0000</td>
<td></td>
<td></td>
<td>1.0000</td>
<td>0.00</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>GT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>0.4479</td>
<td>0.0807</td>
<td>5.551</td>
<td>1.5651</td>
<td>30.82</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>1 2</td>
<td>-0.4479</td>
<td></td>
<td></td>
<td>0.6389</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 1</td>
<td>-0.4479</td>
<td></td>
<td></td>
<td>0.6389</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2</td>
<td>0.4479</td>
<td></td>
<td></td>
<td>1.5651</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1 1</td>
<td>0.2452</td>
<td>0.0807</td>
<td>3.039</td>
<td>1.2779</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1 2</td>
<td>-0.2452</td>
<td></td>
<td></td>
<td>0.7825</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 1</td>
<td>-0.2452</td>
<td></td>
<td></td>
<td>0.7825</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 2</td>
<td>0.2452</td>
<td></td>
<td></td>
<td>1.2779</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 1 1</td>
<td>-0.2452</td>
<td></td>
<td></td>
<td>0.7825</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 1 2</td>
<td>0.2452</td>
<td></td>
<td></td>
<td>1.2779</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2 1</td>
<td>0.2452</td>
<td></td>
<td></td>
<td>1.2779</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2 2</td>
<td>-0.2452</td>
<td></td>
<td></td>
<td>0.7825</td>
<td>9.23</td>
<td>1</td>
<td>0.002</td>
</tr>
</tbody>
</table>

### (CONDITIONAL) PROBABILITIES

**P(AGT)**

| 1 1 1  | 0.2000 (0.0283) |
| 1 1 2  | 0.0500 (0.0154) |
| 1 2 1  | 0.0500 (0.0154) |
| 1 2 2  | 0.2000 (0.0283) |
| 2 1 1  | 0.1500 (0.0252) |
| 2 1 2  | 0.1000 (0.0212) |
| 2 2 1  | 0.1000 (0.0212) |
| 2 2 2  | 0.1500 (0.0252) |
Authors on Loglinear Modelling: See full references

Agresti, A.
Aldrich,
Andersen, E. B
Badsberg, J. H.
Bishop,...et. al.,
Christensen,
DeMaris,
Edwards
Elliot
Everitt
Feinberg
Fox,
Goodman,
Grizzle, J. E., Starmer, C. F. & Koch, G. G.
Haberman
Hagenaars
Ishii-Kuntz
Jaccard
Knoke,
Kriska, S. D. & Milligan, G. W.
Kreiner
Landwehr, J. M.
McCullagh, P. & Nelder, J. A.
Reynolds
Wickens
The Logit Model and Logistic Regression
The Logit Model and Logistic Regression

IV. The Logit Model

A. The Odds ratio

\[ \Omega = \frac{n_{il}}{n_{i2}} \]

\[ \phi = \psi = \log \frac{n_{il}}{n_{i2}} \]

\[ \psi = \log \left( \frac{p_{il}}{1-p_{il}} \right) \]

\[ \psi = \log \left( \frac{n_{il}}{N-n_{il}} \right) \]

\[ \psi = \log \Omega \]

B. The Logit Equation

\[ \Phi_i = \beta_0 + \beta_i A \]

\[ \beta = 2\lambda \]
Logistic Regression, Logit, and Probit Examples

log: C:\classdat\logistics.log
log type: text
opened on: 2 Mar 2003, 12:29:51

. logit ntv ngen

Iteration 0:  log likelihood = -138.62944
Iteration 1:  log likelihood = -122.22013
Iteration 2:  log likelihood = -122.17286
Iteration 3:  log likelihood = -122.17286

Logit estimates

Number of obs   =        200
LR chi2(1)      =      32.91
Prob > chi2     =     0.0000
Log likelihood = -122.17286
Pseudo R2       =     0.1187

------------------------------------------------------------------------------
ntv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
gen |   1.694596   .3086067     5.49   0.000     1.089738    2.299454
_cons |  -.8472979   .2182179    -3.88   0.000    -1.274997   -.4195987
------------------------------------------------------------------------------

. logistic ntv  ngen

Logit estimates

Number of obs   =        200
LR chi2(1)      =      32.91
Prob > chi2     =     0.0000
Log likelihood = -122.17286
Pseudo R2       =     0.1187

------------------------------------------------------------------------------
ntv | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
gen |   5.444444   1.680192     5.49   0.000     2.973494    9.968735
------------------------------------------------------------------------------
. logit ntv nage ngen

Iteration 0:  log likelihood = -138.62944
Iteration 1:  log likelihood = -122.22013
Iteration 2:  log likelihood = -122.17286
Iteration 3:  log likelihood = -122.17286

Logit estimates
Number of obs = 200
LR chi2(2) = 32.91
Prob > chi2 = 0.0000
Log likelihood = -122.17286  Pseudo R2 = 0.1187

------------------------------------------------------------------------------
      ntv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      nage |  -2.59e-17   .3086067    -0.00   1.000     -.604858     .604858
     ngen |   1.694596   .3086067     5.49   0.000     1.089738    2.299454
     _cons |  -.8472979   .2672612    -3.17   0.002     -1.37112   -.3234755
------------------------------------------------------------------------------

. logistic ntv nage ngen

Logit estimates
Number of obs = 200
LR chi2(2) = 32.91
Prob > chi2 = 0.0000
Log likelihood = -122.17286  Pseudo R2 = 0.1187

------------------------------------------------------------------------------
      ntv |   Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+---------------------------------------------------------------
      nage |          1   .3086067    -0.00   1.000      .546152    1.830992
     ngen |   5.444444   1.680192     5.49   0.000     2.973494    9.968735
     _cons |          1   2.672612    -3.17   0.002     -1.37112   -.3234755
------------------------------------------------------------------------------
. probit ntv ngen

Iteration 0:  log likelihood = -138.62944
Iteration 1:  log likelihood = -122.20357
Iteration 2:  log likelihood = -122.17286
Iteration 3:  log likelihood = -122.17286

Probit estimates

|         | Coef.   | Std. Err. |     z  |   P>|z|  |   [95% Conf. Interval] |
|---------|---------|-----------|-------|------|-----------------------|
| ntv     | 1.0488  | 0.1864    | 5.61  | 0.00 | 0.6835-1.4141         |
| ngen    |         |           |       |      |                       |
| _cons   | -0.52   | 0.13      | -3.93 | 0.00 | -0.84-0.20            |

Probit estimates

Number of obs   = 200  
LR chi2(1)      = 32.91  
Prob > chi2     = 0.000  
Pseudo R2       = 0.1187

------------------------------------------------------------------------------
|         | Coef.   | Std. Err. |     z  |   P>|z|  |   [95% Conf. Interval] |
|---------|---------|-----------|-------|------|-----------------------|
| ntv     |         |           |       |      |                       |
| nage    | -2.12e-17 | 0.1865    | -0.00 | 1.00 | -0.37-0.37            |
| ngen    | 1.0488  | 0.1864    | 5.61  | 0.00 | 0.6835-1.4141         |
| _cons   | -0.52   | 0.16      | -3.21 | 0.00 | -0.85-0.21            |

------------------------------------------------------------------------------
Using Interaction Terms

```
.xi: logistic ntv i.ingen*i.nage
i.ingen    _Ingen_0-1    (naturally coded; _Ingen_0 omitted)
i.nage     _Inage_0-1    (naturally coded; _Inage_0 omitted)
i.ingen*i.nage   _IngeXnag_#_#    (coded as above)
```

Logit estimates

|             | Odds Ratio | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|-------------|------------|-----------|------|-----|----------------------|
| _Ingen_1    | 16         | 7.999937  | 5.55 | 0.000 | 6.005132 - 42.6302   |
| _Inage_1    | 2.666667   | 1.217156  | 2.15 | 0.032 | 1.090064 - 6.523572  |
| _IngeXnag_~1| 0.140625   | 0.0907726 | -3.04| 0.002 | 0.0396841 - 0.4983198|

Log likelihood = -117.34141

```
.xi: logit ntv i.ingen*i.nage
i.ingen    _Ingen_0-1    (naturally coded; _Ingen_0 omitted)
i.nage     _Inage_0-1    (naturally coded; _Inage_0 omitted)
i.ingen*i.nage   _IngeXnag_#_#    (coded as above)
```

Iteration 0:   log likelihood = -138.62944
Iteration 1:   log likelihood = -117.62977
Iteration 2:   log likelihood = -117.34211
Iteration 3:   log likelihood = -117.34141

Logit estimates

|             | Coef.        | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|-------------|--------------|-----------|------|-----|----------------------|
| _Ingen_1    | 2.772589     | .4999961  | 5.55 | 0.000 | 1.792614 - 3.752563  |
| _Inage_1    | .9808293     | .4564333  | 2.15 | 0.032 | .0862364 - 1.875422  |
| _IngeXnag_~1| -1.961659    | .6454942  | -3.04| 0.002 | -3.226804 - .6965132 |
| _cons       | -1.386294    | .3535506  | -3.92| 0.000 | -2.079241 - .6933479 |

Log likelihood = -117.34141

```
.xi: probit ntv i.ingen*i.nage
i.ingen    _Ingen_0-1    (naturally coded; _Ingen_0 omitted)
i.nage     _Inage_0-1    (naturally coded; _Inage_0 omitted)
i.ingen*i.nage   _IngeXnag_#_#    (coded as above)
```

Iteration 0:   log likelihood = -138.62944
Iteration 1:   log likelihood = -117.5409
Iteration 2:   log likelihood = -117.34147
Iteration 3:   log likelihood = -117.34141

Probit estimates

|             | Coef.        | Std. Err. | z    | P>|z|  | [95% Conf. Interval] |
|-------------|--------------|-----------|------|-----|----------------------|
| _Ingen_1    | 1.683242     | .2857531  | 5.89 | 0.000 | 1.123177 - 2.243308  |
| _Inage_1    | .5882741     | .2701591  | 2.18 | 0.029 | .058772 - 1.117776  |
| _IngeXnag_~1| -1.176548    | .3820627  | -3.08| 0.002 | -1.925377 - .4277191|
| _cons       | -.8416212    | .202058   | -4.17| 0.000 | -1.237648 - .4455949|

Log likelihood = -117.34141

Authors on Logistic Regression and Logit Models. See full references.

Agresti
Prediction Logic
Prediction Logic

This approach, developed by Hildebrand and Liang, defines certain cells as "error" cells that should not fit the predicted pattern. Choose other cells as "correct". That is, as those that do fit the pattern. Assign a weight $W_{ij}$ to each "error" and zero to each "correct" cell.

General Equation

$$\nabla = 1 - \frac{\text{observed errors}}{\text{expected errors}}$$

Equation in Frequency Form

$$\nabla = 1 - \frac{\sum \sum w_{ij}n_{ij}}{\sum \sum \sum w_{ij}n_{i}n_{j}}$$

Equation in Proportion Form

$$\nabla = 1 - \frac{\sum \sum w_{ij}p_{ij}}{\sum \sum \sum w_{ij}p_{i}p_{j}}$$

(12) General Equation
Variance of Delta

Subterm 1

$$a_{ij} = \left[ \sum \sum w_{ij} P_i P_j \right]^{-1} \left[ w_{ij} 1 - \nabla \right] \left( \pi_i + \pi_j \right)$$

Subterm 2

$$\Pi_i = \sum_{j=1}^{t} w_{ij} P_j$$
$$\Pi_j = \sum_{i=1}^{t} w_{ij} P_i$$

Full Variance Equation

$$Var_\varphi = \frac{1}{N-1} \left[ \sum \sum a_\varphi^2 p_\varphi - (\sum \sum a_\varphi p_\varphi)^2 \right]$$
PREDICTION ANALYSIS OF CROSS-CLASSIFICATIONS
FORTAN90 program; author: Alexander von Eye

Simple Example: I hypothesized that Male would go with Sports and Female would go with Romance.

number of predictor levels, p = 2
number of criterion levels, k = 2
N = 200.

Observed Cell Frequencies
-------- ---- -----------
70.    30.   100.
30.    70.   100.
100.   100.   200.

Expected Frequencies
--------  -----------
50.00  50.00
50.00  50.00

Matrix of Hit and Error Cells
1.00   .00
.00   1.00

Table of Results
there is strong support of the hypothesis

Descriptive Measures
-----------  --------
PIH1 = .4000  
PIH2 = .4000  
Del = .4000  
Precision = .5000  
Scope = 1.0000

Overall Significance Test
----------  ------------ ----
Test statistic z = 5.6569
Tail probability for z = .0000

Chi-square tests for model [Predictors][Criteria]
----------  -----  ---  ----------------------
Pearson Chi-square = 32.0000
                     df = 1.   p = .000000
Likelihood ratio Chi-square = 32.9132
                     df = 1.   p = .000000

Evaluation of partial hypotheses
----------  ----  -----  -----  -------  --------

<table>
<thead>
<tr>
<th>partial hyp.</th>
<th>fo</th>
<th>fe</th>
<th>del</th>
<th>precis. del</th>
<th>del(cum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.00</td>
<td>50.00</td>
<td>.400</td>
<td>.250</td>
<td>.200</td>
</tr>
<tr>
<td>2</td>
<td>30.00</td>
<td>50.00</td>
<td>.400</td>
<td>.250</td>
<td>.400</td>
</tr>
</tbody>
</table>
Regression and Classification Trees
Zhao's RTREE

There are 2 covariates
Var 2  gender
cvar1  age

Original status of variables are
3 3
1 refers to as an ordinal covariate and 3 means a nominal one
For an ordinal covariate, the min. and max. will be given;
For a nominal one, the counts corresponding to each level will be listed.
1: 100(1) 100(2)
2: 100(1) 100(2)

The initial tree:

node  #cases   left   right  split var  cut-off
1     200      2       3      2 G     {1}
2     100      4       5      1 A     {1}
3     100      6       7      1 A     {1}
4      50  terminal node with distribution: 20 30
5      50  terminal node with distribution: 10 40
6      50  terminal node with distribution: 30 20
7      50  terminal node with distribution: 40 10

The pruned tree at p-value of 0.010000:

node  #cases   left   right  split var  cut-off
1     200      2       3      2       {1}
2     100  terminal node with distribution: 30 70
3     100  terminal node with distribution: 70 30

The chi-square value for nodes after pruned are shown below
32.000000 4.761905 4.761905

The pruned tree at p-value of 0.005000:

node  #cases   left   right  split var  cut-off
1     200      2       3      2       {1}
2     100  terminal node with distribution: 30 70
3     100  terminal node with distribution: 70 30

The chi-square value for nodes after pruned are shown below
32.000000 4.761905 4.761905

The pruned tree at p-value of 0.001000:

node  #cases   left   right  split var  cut-off
1     200      2       3      2       {1}
2     100  terminal node with distribution: 30 70
3     100  terminal node with distribution: 70 30

The chi-square value for nodes after pruned are shown below
32.000000 4.761905 4.761905

The pruned tree at p-value of 0.000500:

node  #cases   left   right  split var  cut-off
1     200      2       3      2       {1}
2     100  terminal node with distribution: 30 70
3     100  terminal node with distribution: 70 30

The chi-square value for nodes after pruned are shown below
32.000000 4.761905 4.761905

The pruned tree at p-value of 0.000100:

node  #cases   left   right  split var  cut-off
1     200      2       3      2       {1}
2     100  terminal node with distribution: 30 70
3     100  terminal node with distribution: 70 30

The chi-square value for nodes after pruned are shown below
32.000000 4.761905 4.761905

The pruned tree at p-value of 0.000050:

node  #cases   left   right  split var  cut-off
1     200      2       3      2       {1}
2     100  terminal node with distribution: 30 70
3     100  terminal node with distribution: 70 30

The chi-square value for nodes after pruned are shown below
32.000000 4.761905 4.761905
The pruned tree at p-value of 0.000100:

<table>
<thead>
<tr>
<th>node</th>
<th>#cases</th>
<th>left</th>
<th>right</th>
<th>split var</th>
<th>cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>{1}</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>terminal node with distribution: 30 70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>terminal node with distribution: 70 30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The chi-square value for nodes after pruned are shown below

32.000000 4.761905 4.761905

The pruned tree at p-value of 0.000010:

<table>
<thead>
<tr>
<th>node</th>
<th>#cases</th>
<th>left</th>
<th>right</th>
<th>split var</th>
<th>cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>{1}</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>terminal node with distribution: 30 70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>terminal node with distribution: 70 30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The chi-square value for nodes after pruned are shown below

32.000000 4.761905 4.761905

The pruned tree at p-value of 0.000001:

<table>
<thead>
<tr>
<th>node</th>
<th>#cases</th>
<th>left</th>
<th>right</th>
<th>split var</th>
<th>cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>{1}</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>terminal node with distribution: 30 70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>terminal node with distribution: 70 30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The chi-square value for nodes after pruned are shown below

32.000000 4.761905 4.761905
RTREE Analysis Of TV Preference by Gender and Age
Variable description file: tv1.dsc
Learning sample data file is: tv1.dat
Code for missing values: ?
Variables in data file (d=dependent, c=categorical, n=numerical, x=excluded):

<table>
<thead>
<tr>
<th>Column #</th>
<th>Variable name</th>
<th>Variable type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>id</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>age</td>
<td>c</td>
</tr>
<tr>
<td>3</td>
<td>gender</td>
<td>c</td>
</tr>
<tr>
<td>4</td>
<td>tv</td>
<td>d</td>
</tr>
</tbody>
</table>

Number of learning samples = 200
Number of classes = 2
Number of numerical variables = 0
Number of categorical variables = 2

Split type: Univariate splits
Variable Selection method: 1D
alpha-threshold used = 0.500E-01
Split method: Linear discriminant analysis
Prior selection: Estimated prior probabilities
Class prior (scaled to have sum 1)
1.0 0.500
2.0 0.500
Cost selection: Equal misclassification costs
Imputation option for building trees: No missing value was found

Tree size determination: Pruning by cross-validation
S.E. rule used = 0.0
Minimum size of each node (mindat): 3
Number of folds for cross-validation: 10

Subtree pruning sequence:

<table>
<thead>
<tr>
<th>Subtree</th>
<th>True alpha</th>
<th># Terminal Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(largest)</td>
<td>0.0000</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>0.0000</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0.2000</td>
<td>1</td>
</tr>
</tbody>
</table>

CV misclassification cost and SE of subtrees:

<table>
<thead>
<tr>
<th>Subtree</th>
<th>CV R(t)</th>
<th>CV SE</th>
<th># Terminal Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(largest)</td>
<td>0.310000</td>
<td>0.3240E-01</td>
<td>4</td>
</tr>
<tr>
<td>1*</td>
<td>0.300000</td>
<td>0.3240E-01</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0.500000</td>
<td>0.3536E-01</td>
<td>1</td>
</tr>
</tbody>
</table>

* denotes 0-SE Tree
** denotes given-SE Tree
* tree is same as ** tree
Following tree is based on **

Splits of the Tree:

Node  Split variable
1  gender
2  * terminal *
3  * terminal *

Tree Structure:

Node 1: gender = 1.0

Node 2: Terminal Node, predicted class = 1.0
Class label : 1.0 2.0
Class size : 70 30

Node 1: gender = 2.0

Node 3: Terminal Node, predicted class = 2.0
Class label : 1.0 2.0
Class size : 70 30

Detailed Description of the Tree:

Nodes No. Subnode Split Stat. Split Split
label cases label stat. variable value
1 200 2  F  gender  = 1.0
3
# obs mean/mode of gender
Class 1.0 : 100 1.0
Class 2.0 : 100 2.0

2 100 **** terminal, predicted class: 1.0
# obs
Class 1.0 : 70
Class 2.0 : 30

3 100 **** terminal, predicted class: 2.0
# obs
Class 1.0 : 30
Class 2.0 : 70

Number of nodes in maximum tree = 7
Number of nodes in final tree = 3
Number of terminal nodes in final tree = 2

Classification Matrix: Predicted class

<table>
<thead>
<tr>
<th></th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual class</td>
<td>1.0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>100</td>
</tr>
</tbody>
</table>
Total obs = 200, # correct = 140

Resubstitution misclassification cost = 0.3000
S.E. of Resubstitution misclassification cost = 0.3240E-01
Cross-validation error cost from pruning = 0.3000
S.E. of CV misclassification cost = 0.3240E-01
Elapsed system time in seconds: 1.51

Configural Frequency Analysis
Configural Frequency Analysis: Complex Set

author of program: Alexander von Eye, 1997

Marginal Frequencies

Variable Frequencies

<table>
<thead>
<tr>
<th></th>
<th>100.</th>
<th>100.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100.</td>
<td>100.</td>
</tr>
<tr>
<td>G</td>
<td>100.</td>
<td>100.</td>
</tr>
<tr>
<td>T</td>
<td>100.</td>
<td>100.</td>
</tr>
</tbody>
</table>

sample size N = 200

Lehmachers test with continuity correction was used

Bonferroni-adjusted alpha = .0062500

Table of results

<table>
<thead>
<tr>
<th>Configuration</th>
<th>fo</th>
<th>fe</th>
<th>statistic</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGT</td>
<td>111</td>
<td>40.</td>
<td>25.000</td>
<td>4.094</td>
</tr>
<tr>
<td></td>
<td>112</td>
<td>10.</td>
<td>25.000</td>
<td>-4.094</td>
</tr>
<tr>
<td></td>
<td>121</td>
<td>10.</td>
<td>25.000</td>
<td>-4.094</td>
</tr>
<tr>
<td></td>
<td>122</td>
<td>40.</td>
<td>25.000</td>
<td>4.094</td>
</tr>
<tr>
<td></td>
<td>211</td>
<td>30.</td>
<td>25.000</td>
<td>1.270</td>
</tr>
<tr>
<td></td>
<td>212</td>
<td>20.</td>
<td>25.000</td>
<td>-1.270</td>
</tr>
<tr>
<td></td>
<td>221</td>
<td>20.</td>
<td>25.000</td>
<td>-1.270</td>
</tr>
<tr>
<td></td>
<td>222</td>
<td>30.</td>
<td>25.000</td>
<td>1.270</td>
</tr>
</tbody>
</table>

Pearson Chi2 for CFA-model = 40.0000

df = 4  p = .0000004
Correspondence Analysis
From: A MANUAL FOR CORV: A MICROCOMPUTER PROGRAM FOR SEVEN SCALINGS OF CORRESPONDENCE ANALYSIS OF CONTINGENCY TABLES.

Bernard S. Gorman
Nassau Community College, Garden City, NY 11530
and Hofstra University, Hempstead, NY 11550

and

Louis H. Primavera
St. John's University, Jamaica, NY 11439

Correspondence analysis (CA) refers to a series of techniques for the factor analysis of data in contingency tables so that the rows and columns of a table can be geometrically represented as points in multidimensional space. In recent years, CA has gained in popularity. However, a variety of scalings of CA have appeared in the statistical literature for over 50 years under such names as: biplots (Gabriel, 1971), reciprocal averaging (Hill, 1974), dual scaling or optimal scaling (Nishisato, 1980), correspondence analysis (Greenacre, 1984), quantification of qualitative data (Guttman, 1941). The present program, CORV, is a microcomputer program, written in compiled BASIC, to implement seven major scalings of correspondence analysis of two-way tables.

Theory

All scalings of correspondence analysis start by factoring a contingency table into matrices of characteristic roots (eigenvalues) and characteristic vectors (eigenvectors). The scalings differ in the ways that they scale these matrices. To start, we take a contingency table, $A$, with $m$ rows, $n$ columns and elements $a_{ij}$ and form an $m$ by $m$ diagonal matrix $D_r$ in which the diagonal elements are the marginal sums for each row. We also form $D_c$, an $n$ by $n$ diagonal matrix in which the diagonal elements are the marginal sums for each column. A constant, $N$, the sum of all elements is found by

$$N = \sum_{1}^{m} \sum_{1}^{n} a_{ij}$$

(16)

We then form a matrix $G$ so that:

$$G = D_r^{-\frac{1}{2}} A D_c^{-\frac{1}{2}}$$

(17)

That is, each element of $G$, $g_{ij}$ is:
We find two sets of eigenvectors: R, for the rows of G, and C for the columns of G by means of a singular value decomposition. This decomposition will also provide a matrix of eigenvalues, S, so that:

\[ G = R S C' \]  \hspace{1cm} (19)

The first eigenvalue in S is trivial, its value will always be 1.0 and it makes the elements in the first columns of R and C proportional to matrix A's, marginal row and column frequencies, respectively. The subsequent eigenvalue elements \( s_{ii} \) in S can be tested for statistical significance by a Chi-square test (Maxwell, 1977) with \( (n + m - 2i) \) degrees of freedom. If significant, you can state that there is an association between some row elements of A and some column elements.

\[ \chi^2_{i(i+1)} = N s_i^2 \]  \hspace{1cm} (20)

The chi-square test for the entire table is the sum of squared elements in S (excluding the first trivial root) multiplied by N, the total sample size. The degrees of freedom are \((m-1)(n-1)\).

\[ \chi^2 = N \sum_{i=2}^{k} s_i^2 \]  \hspace{1cm} (21)

Some authors, such as Nishisato (1984), prefer to use Bartlett's test of the remaining roots.

**Scalings**

So far, we have obtained the characteristic roots and vectors of G and have shown that G can be reproduced by the matrices. In traditional factor analysis, eigenvector elements are scaled by multiplying them by the square roots of their corresponding eigenvalues to form orthogonal factor loadings. These loadings typically represent the Pearson product-moment correlation between a variable with the new, underlying, factor. In correspondence analysis, we rescale the eigenvectors so that they represent the coordinates of row and column categories on underlying dimensions. The five scalings of correspondence analysis reported by the CORV program will apply different scalings.

The Biplot Scaling: This scaling provides factor loadings of rows and columns that are decompositions of the standardized chi-square residuals; the standardized difference between observed and expected frequencies for each cell in the table. This scaling appeared in Gabriel (1971) and Maxwell (1977) and provides a strong statistical rationale for correspondence analysis.

The biplot scaling stays very close to traditional factor analysis by scaling the row eigenvectors and the column eigenvectors by the square roots of their respective eigenvalues to form row factors, \( F_r \), and column factors \( F_c \) so that:

\[ g_{ij} = \frac{a_{ij}}{\sqrt{a_{ii} a_{jj}}} \]  \hspace{1cm} (18)
Let $F_{ri}$ be the loading of row element $i$, on factor 1, the trivial factor. Let $F_{cj}$ be the loading of column element $j$, on factor 1, the trivial factor. Let $a_i$ be the row marginal frequencies and $a_j$ be the column marginal frequencies for contingency table $A$. It can be shown that $e_{ij}$, the expected frequency, or frequency for a contingency table cell that can be predicted from chance alone is:

\[
e_{ij} = \left( \sqrt{a_i a_j} \right) \left( f_{ri} f_{cj} \right)
\]  

(23)

It can also be shown that $d_{ij}$, the deviation of the expected frequency from the observed frequency for a cell can be formed by:

\[
d_{ij} = \left( \sqrt{a_i a_j} \right) \sum_{k=2}^{n} \left( f_{rik} f_{gkh} \right)
\]  

(24)

Since the observed frequency of any cell in the contingency table, $o_{ij}$, is equal to the sum of its expected frequency and the deviation of the cell’s frequency from the expected frequency, this decomposition provides a very strong proof that this scaling preserves the information in the original contingency table $A$.

Program users should retain factors that are statistically significant and, if desired, can rotate the factor loadings to simple structure by means of a varimax rotation. As in any factor analysis or multidimensional scaling, the loadings on the factors may be plotted to form a graphic representation of the relationship among rows and rows, columns and columns, and rows and columns. In this scaling, it is correct to say that rows that are close to other rows in the factor space are similar to each other. It is also legitimate to state that columns that are close to other columns in the factor space are similar to each other. However, it is not correct to say that rows that are close to columns in the factor space bear a strong relationship to each other. Rather, row-to-column relationships are angular relationships. If one were to draw a vector; a line from the origin of a factor plot to the point that represented a row or column factor loading, then rows that are strongly related to columns will have small acute angles between their vectors. Rows and columns that are unrelated will have right angles between their vectors and columns that are negatively related will have large obtuse angles between the vectors.

Greenacre’s and Benzecri’s (Greenacre, 1984) Scaling: This scaling is probably the most popular of the scalings. It is sometimes called the "French Scaling" and "Principal Scaling". As in the biplot, it permits comparisons of distances of rows to rows and columns to columns. The relationships between rows and columns, however, are not represented as point-to-point distance relationships but, rather, as vector relationships in which row vectors point in the direction of columns to which they are most strongly related and vice-versa. That is, if the angle between any two vectors is small, then vectors are similar. If the angle approaches 90 degrees, the vectors are unrelated (orthogonal). If the vectors approach 180 degrees, the are
inversely related. Vector cosines can be treated as correlations

For this scaling and for subsequent scalings, define two matrices, X and Y that rescale the eigenvectors. Their formulae are:

\[ X = \left( N^{-1} D_r \right)^{\frac{1}{2}} \]
\[ Y = \left( N^{-1} D_c \right)^{\frac{1}{2}} \]  
(25)

Using these two matrices to rescale the row and column eigenvectors of G, gives the following coordinate matrices:

\[ F_r = X R S \]
\[ F_c = Y C S \]  
(26)

Unlike the biplot scaling, the product of the row and column matrices will not directly reproduce expected or observed frequencies.

Carroll, Green, and Schaffer’s (1986) distance scaling: Although all scalings of CA permit users to compare interpoint distances of rows to other rows and columns to other columns, most literature cautions readers not to estimate distances between row points and column points. However, Carroll, Green, and Schaffer believe that their CA scaling permits column-row comparisons, although Greenacre (1989) debated the feasibility of this property. Their scaling differs from Greenacre’s scaling by the introduction of two identity matrices, \( I_r \) an m by m identity matrix for the rows and \( I_c \) an n by n identity matrix for the columns. Their formula is:

\[ F_r = X R \left( S + I_r \right)^{\frac{1}{2}} \]
\[ F_c = Y C \left( S + I_c \right)^{\frac{1}{2}} \]  
(27)

Kendall and Stuart’s (1979) and Nishisato’s (1980) ‘Optimal Scaling’ : This solution provides weights for each row category and each column category within a factor so the qualitative categories may be quantitatively weighted to form maximum Pearson product-moment correlations between the rows and columns within each factor. Unlike previous scalings that use the eigenvalue information in matrix S, this scaling only uses the row and column eigenvectors and the X and Y matrices. The formula is

\[ F_r = X R \]
\[ F_c = Y C \]  
(28)

The scalings produced by this method match those of Thurstone’s and Torgerson’s Successive Category Scaling Methods. It has the added bonus of being able to find dimensions beyond the first dimension, which hints at curvilinearity or other sources of variance.
Hill’s (1974) Reciprocal Averaging Scaling (Column Principal): As in the Kendall and Stuart and Nishisato scalings, this scaling provides weighting of row and column categories. The formulae for the coordinates are:

\[
F_r = X \cdot R \\
F_c = Y \cdot C \cdot S
\]

This scaling is useful for comparing columns to columns but does a poor job of comparing rows. It weighs columns by marginal frequencies and eigenvalues. However, it does not weigh rows by eigenvalues. In doing so, the column coordinates are shrunken relative to the row coordinates. This is so because squared singular values are always less than 1.0.

Row Principal Scalling: This scaling is useful for comparing rows to rows but it does a poor job of comparing columns. It weighs the rows by marginal frequencies and eigenvalues. However, it does not weigh columns by eigenvalues. The row coordinates are shrunken relative to column coordinates. This is so because squared singular values are always less than 1.0.

\[
F_r = X \cdot R \cdot S \\
F_c = Y \cdot C
\]
Canonical Scaling: This solution is the default solution for SPSS’ s Categories Program (SPSS, Inc., 1993). It is like the Biplot but it scales the row coordinates by their marginal proportions.

\[
F_r = X R S^\frac{1}{2}
\]

\[
F_c = Y C S^\frac{1}{2}
\]
General Correspondence Analysis

(C) 1984-1994 Bernard S. Gorman, Ph.D. and Louis H. Primavera, Ph.D.

Original Table

<table>
<thead>
<tr>
<th>Row</th>
<th>70</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>70</td>
</tr>
</tbody>
</table>

Guttman Contingency Coefficients

<table>
<thead>
<tr>
<th>Row</th>
<th>0.700</th>
<th>0.300</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.300</td>
<td>0.700</td>
</tr>
</tbody>
</table>

Total N = 200

Eigenvalues and Tests

Squared Eigenvalue (Inertia) 1 = 1.000
Cumulative Proportion of Variance = 0.000
   Eigenvalue (Singular Value) 1 = 1.000

Squared Eigenvalue (Inertia) 2 = 0.160
Proportion of Variance = 1.000
Cumulative Proportion of Variance = 1.000
   Eigenvalue (Singular Value) 2 = 0.400
   Chi-Square = 32.000  df = 1
   Probability = 0.000

Overall Chi-square = 32.000  df = 1
   Probability = 0.000

Column Eigenvectors

Col. 1  0.707  -0.707
Col. 2  0.707   0.707

Row Eigenvectors

Row 1  0.707  -0.707
Row 2  0.707   0.707

Unrotated Interbattery Loadings Including 'Trivial Factor'

Sports  0.707  -0.447
Romance 0.707   0.447
Male  0.707  -0.447
Female 0.707   0.447
In a contingency table analysis, the first root is trivial. Therefore, the first column will be eliminated.

**Unrotated Carroll/Green/Schaffer Coordinates**
- Sports: -1.183
- Romance: 1.183
- Male: -1.183
- Female: 1.183

**Unrotated Greenacre Coordinates (Principal)**
- Sports: -0.400
- Romance: 0.400
- Male: -0.400
- Female: 0.400

**Unrotated Kendall/Nishisato Optimal Scaling Coordinates**
- Sports: -1.000
- Romance: 1.000
- Male: -1.000
- Female: 1.000

**Reciprocal Averaging (Hill) Coordinates (Column Principal)**
- Sports: -0.400
- Romance: 0.400
- Male: -1.000
- Female: 1.000

**Canonical Coordinates**
- Sports: -0.632
- Romance: 0.632
- Male: -0.632
- Female: 0.632

**Row Principal Coordinates**
- Sports: -1.000
- Romance: 1.000
- Male: -0.400
- Female: 0.400

**Non-Trivial Biplot Loadings**
- Sports: -0.447
- Romance: 0.447
- Male: -0.447
- Female: 0.447
Multiple Correspondence Analysis
From: Manual for MCA: A microcomputer program for multiple correspondence analysis

Bernard S. Gorman
Nassau Community College, Garden City, NY 11530
and Hofstra University, Hempstead, NY 11550

and

Louis H. Primavera
St. John’s University, Jamaica, NY 11439

Theory

Multiple correspondence analysis is a multivariate extension of correspondence analysis that performs a principal components analysis of multiple categorical data. Although mathematical derivations of the technique have been available for over forty years (Burt, 1950; Guttman, 1941; Greenacre, 1985; Lebart, Morineau, & Warwick; 1984; Lingoes, 1968; Nishisato, 1980); it has seen relatively few applications. The present program MCA.EXE was written for MS/DOS-compatible microcomputers and employs the formulas provided by Hoffman and Franke (1986) and Carroll and Green (1988).

Multiple correspondence analysis starts out with a response matrix, A, shown in Table 1, in which the rows are subjects and the columns are categorical (nominal scale) variables. There may be several categories within each variable. The response matrix may be transformed into, B, an indicator matrix, a set of dummy variables in which there are separate dichotomous (1,0) indicator variables to represent each category of each original variable. Thus, if the response matrix had three variables: a, b, and c with 2, 2, and 2 categories, respectively, the indicator matrix would have 6 indicator variables. An indicator variable can have only two values: 1 if the subject fits within the category and zero otherwise. Table 2 presents an indicator matrix corresponding to the response matrix in Table 1.

Table 1

A Response Matrix, A, with Three Categorical Variables and 10 Ss

<table>
<thead>
<tr>
<th>Ss</th>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>s2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>s3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s4</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>s5</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>s6</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>s7</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s8</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>s9</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>s10</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 2

Indicator Matrix, B, With 6 Category Indicators and 10 Ss

<table>
<thead>
<tr>
<th>B</th>
<th>a1</th>
<th>a2</th>
<th>b1</th>
<th>b2</th>
<th>c1</th>
<th>c2</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>s2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>s3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>s4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>s5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>s6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>s7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>s8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>s9</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>s10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

A cross-product matrix, G, called the Burt matrix (after Cyril Burt, 1950) is formed from the matrix product

\[ G = B' \cdot B \]

G, which is shown in Table 3, is a square, symmetrical matrix in which the cell entries are cross-tabulations of the categories of each variable with the categories of all other variables. The matrix may be considered to be a supermatrix in which the categories of any single variable form a diagonal submatrix containing the marginal frequencies for the categories of the variable. The non-diagonal submatrices are the cross-tabulations of the categories from different variables with each other.
Table 3
The Burt Matrix

\[
G = \begin{pmatrix}
  a1 & a2 & b1 & b2 & c1 & c2 \\
  a1 & 5 & 0 & 4 & 1 & 1 & 4 \\
a2 & 0 & 5 & 1 & 4 & 5 & 0 \\
b1 & 4 & 1 & 5 & 0 & 2 & 3 \\
b2 & 1 & 4 & 0 & 5 & 4 & 1 \\
c1 & 1 & 5 & 2 & 4 & 6 & 0 \\
c2 & 4 & 0 & 3 & 1 & 0 & 4 \\
\end{pmatrix}
\]

G is then scaled so that:

\[
G^* = H^{-1/2}GH^{-1/2}
\]

H is a diagonal matrix in which there are zeros in every cell except the diagonal cells, which contain the diagonal elements of G. Table 4 presents the G* matrix.

Table 4
The G* Matrix

\[
G^* = \begin{pmatrix}
  a1 & a2 & b1 & b2 & c1 & c2 \\
  a1 & 1.0000 & 0.0000 & 0.8000 & 0.2000 & 0.1826 & 0.8944 \\
a2 & 0.0000 & 1.0000 & 0.2000 & 0.8000 & 0.9129 & 0.0000 \\
b1 & 0.8000 & 0.2000 & 1.0000 & 0.0000 & 0.3651 & 0.6708 \\
b2 & 0.2000 & 0.8000 & 0.0000 & 1.0000 & 0.7303 & 0.2236 \\
c1 & 0.1826 & 0.9129 & 0.3651 & 0.7303 & 1.0000 & 0.0000 \\
c2 & 0.8944 & 0.0000 & 0.6708 & 0.2236 & 0.0000 & 1.0000 \\
\end{pmatrix}
\]

It can be seen that each element of G* is

\[
g_{ij}^* = \frac{g_{ij}}{\sqrt{h_{ii} h_{jj}}}
\]
Readers may recognize the formula above as similar to the formula by which a correlation coefficient can be formed from a variance covariance matrix. The G matrix is analogous to a variance/covariance matrix and the G* matrix is analogous to a correlation matrix. As such, the G* matrix can be factored by principal components analysis to form a set of eigenvectors, U, and a set of eigenvalues, S.

\[
G^* = USU^T
\]
Tables 5 and 6 present the eigenvalues, $S$, and the eigenvectors, $U$ of $G^*$. 

Table 5

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>3.000</td>
<td>.000</td>
<td>.000</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>II</td>
<td>.000</td>
<td>2.23</td>
<td>.000</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>III</td>
<td>.000</td>
<td>.000</td>
<td>.618</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>IV</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.14</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>V</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>VI</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

Table 6

<table>
<thead>
<tr>
<th></th>
<th>a1</th>
<th>a2</th>
<th>b1</th>
<th>b2</th>
<th>c1</th>
<th>c2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>.4082</td>
<td>-.4484</td>
<td>.1174</td>
<td>-.5340</td>
<td>-.5714</td>
<td>.0826</td>
</tr>
<tr>
<td>II</td>
<td>.4082</td>
<td>.4484</td>
<td>-.1174</td>
<td>.5340</td>
<td>-.5714</td>
<td>.0826</td>
</tr>
<tr>
<td>III</td>
<td>.4082</td>
<td>-.3559</td>
<td>-.5870</td>
<td>.1697</td>
<td>.2142</td>
<td>-.5361</td>
</tr>
<tr>
<td>IV</td>
<td>.4082</td>
<td>.3559</td>
<td>.5870</td>
<td>-.1697</td>
<td>.2142</td>
<td>-.5361</td>
</tr>
<tr>
<td>V</td>
<td>.3651</td>
<td>-.4547</td>
<td>.4123</td>
<td>.4725</td>
<td>.3195</td>
<td>.4057</td>
</tr>
</tbody>
</table>

Factor loadings are formed as

$$X = H^{-1/2}US^{1/2}$$

where $H$ is, as before, the diagonal elements of $G$. 
Table 7 presents the factor loadings, $X$, from $G^*$

Table 7

Factor Loadings, Including the "Trivial" First Factor

$$X = \begin{array}{cccccc}
\text{Factors} & \text{I} & \text{II} & \text{III} & \text{IV} & \text{V} & \text{VI} \\
\text{Indicators} & a1 & a2 & b1 & b2 & c1 & c2 \\
a1 & .3162 & -.2996 & .0413 & -.0924 & .0000 & .0000 \\
a2 & .3162 & .2996 & -.0413 & .0924 & .0000 & .0000 \\
b1 & .3162 & -.2378 & -.2064 & .0294 & .0000 & .0000 \\
b2 & .3162 & .2378 & .2064 & -.0294 & .0000 & .0000 \\
c1 & .3162 & .2264 & -.1081 & -.0610 & .0000 & .0000 \\
c2 & .3162 & -.3397 & .1621 & .0914 & .0000 & .0000 \\
\end{array}$$

The first factor of $G$ is "trivial" in that it contains constants that are proportional to the marginal categories of each category. To be consistent with previous literature, the first eigenvalue of $S$ and the first eigenvector of $X$ will be discarded. It should also be noted that the last two diagonal elements of $S$ and the last two columns of $X$ are filled with zeros. This is not accidental, as $G^*$ is a singular (rank deficient) matrix due to the fact that dummy variable indicator columns for each category carry more information than is necessary. That is, in most multiple regression or MANOVA analyses, one would create one less dummy variables than there are levels of a factor in order to avoid linear dependencies and singularity. In general, there will be as many meaningful factors (after the first trivial factor has been removed) as there are original categorical variables. As we started with three categorical variables in the present case, we have three factors.

If desired, the non-trivial factors may be rotated by varimax rotation to form a simple structure solution. In the present case, two factors were retained and their loadings are presented in Table 8. A plot of the loadings of the first two non-trivial factors is presented in Figure 1.
Table 8
Varimax Rotation of Non-Trivial Factors

<table>
<thead>
<tr>
<th>Categories</th>
<th>I</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>0.270</td>
<td>-0.136</td>
</tr>
<tr>
<td>a2</td>
<td>-0.270</td>
<td>0.137</td>
</tr>
<tr>
<td>b1</td>
<td>0.078</td>
<td>-0.305</td>
</tr>
<tr>
<td>c2</td>
<td>-0.078</td>
<td>0.305</td>
</tr>
<tr>
<td>c1</td>
<td>-0.248</td>
<td>0.040</td>
</tr>
<tr>
<td>c2</td>
<td>0.371</td>
<td>-0.060</td>
</tr>
</tbody>
</table>
Carroll and Greene (1988) suggested that the G* matrices could be combined and compared in an INDSCAL (Carroll and Chang, 1970) analysis if the effect of the trivial factor were removed. In order to remove the effect, a new matrix, G**, is formed from the remaining non-trivial loadings of from the factorization above. Thus:
\[ G^{**} = X'X \]

Table 9 presents the resulting G** matrix.

Table 9
Reduced G** Matrix

<table>
<thead>
<tr>
<th>Category Indicators</th>
<th>a1</th>
<th>a2</th>
<th>b1</th>
<th>b2</th>
<th>c1</th>
<th>c2</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>0.457</td>
<td>-0.457</td>
<td>0.313</td>
<td>-0.314</td>
<td>-0.396</td>
<td>0.485</td>
</tr>
<tr>
<td>a2</td>
<td>-0.457</td>
<td>0.458</td>
<td>-0.314</td>
<td>0.314</td>
<td>0.396</td>
<td>-0.485</td>
</tr>
<tr>
<td>b1</td>
<td>0.313</td>
<td>-0.314</td>
<td>0.495</td>
<td>-0.496</td>
<td>-0.173</td>
<td>0.211</td>
</tr>
<tr>
<td>b2</td>
<td>-0.314</td>
<td>0.314</td>
<td>-0.496</td>
<td>0.496</td>
<td>0.173</td>
<td>-0.212</td>
</tr>
<tr>
<td>c1</td>
<td>-0.396</td>
<td>0.396</td>
<td>-0.173</td>
<td>0.173</td>
<td>0.378</td>
<td>-0.463</td>
</tr>
<tr>
<td>c2</td>
<td>0.485</td>
<td>-0.485</td>
<td>0.211</td>
<td>-0.212</td>
<td>-0.463</td>
<td>0.566</td>
</tr>
</tbody>
</table>

Some users may wish only to perform single-sample sample analyses. Therefore, they probably will not need the G** matrix. However, even in the single sample case, it might be useful to split the sample at random into sub-samples; obtain the G** matrix for each one; and then use INDSCAL to obtain a synthesized configuration. The presence of any substantial difference in the individual weights for the solution might be an indicator of instability.
Multiple Correspondence Analysis Example
OPTS.EXE - OPTIMAL SCALING with optional ANOVA by Alexei van Baaren
Sunday January 26 2003

OPTS-E Optimal Scaling (HOMALS, MCA) Analysis of Age, Gender, and TV Preference

# of items = 3
#s of options  2  2  2 TOTAL  6

*** RAW DATA ***
1 0 1 0 1 0 3
1 0 1 0 1 0 3
1 1 1 1 1 1 6
1 1 1 1 1 1 6
1 1 1 1 1 1 6
1 1 1 1 1 1 6
1 1 1 1 1 1 6

etc...

STATISTICS    Sol.1  Sol.2
delta(var)%    70.3   95.6
# iterations   12      6
correlation   0.340  0.204
C H I - S Q  111.14  38.54
D.F.            203    201
probability   1.000  1.000

*** Optimal Scale Values ***

<table>
<thead>
<tr>
<th>Item</th>
<th>Option</th>
<th>Dim1</th>
<th>Dim2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1</td>
<td>-0.285</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.745</td>
<td>0.774</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>-0.285</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.364</td>
<td>-0.414</td>
</tr>
<tr>
<td>TV</td>
<td>1</td>
<td>-0.285</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.261</td>
<td>-0.451</td>
</tr>
</tbody>
</table>

*** Standard Deviations ***

<table>
<thead>
<tr>
<th>Item</th>
<th>Sol.1</th>
<th>Sol.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.267</td>
<td>0.247</td>
</tr>
<tr>
<td>2</td>
<td>0.189</td>
<td>0.165</td>
</tr>
<tr>
<td>3</td>
<td>0.164</td>
<td>0.184</td>
</tr>
</tbody>
</table>

********************************************************************
| Joint Correspondence Analysis -- Iterative (Greenacre Algorithm) |

------------------------------------------
CorAna
Simple and Joint Correspondence Analysis
**Burt Matrix**

young     old    male  female  sports romance
young       100.
old           0.    100.
male         50.     50.    100.
female       50.     50.      0.    100.
sports       50.     50.     70.     30.    100.
romance      50.     50.     30.     70.      0.    100.

Masses:
0.500   0.500   0.500   0.500   0.500   0.500

Number of Cases:
200.

**Correlation Matrix of Transformed Variables**

1.000
0.000   1.000
0.000   0.400   1.000

Standard Eigenvalues:
1.400   1.000   0.600

Discrepancy Function (Before Iteration): 0.308148791101958D-032

Iteration Details:
1  0.00000000000000000000  0.00000000000000000000
1  0.00000000000000000000  0.00000000000000000000

**Standard Coordinates**

young      0.000   0.000
old        0.000   0.000
male       1.225   0.000
female    -1.225   0.000
sports  1.225  0.000
romance -1.225  0.000

Square roots of inertias:
0.267  0.000

| Principal Coordinates |

| young     | 0.000 | 0.000 |
| old       | 0.000 | 0.000 |
| male      | 0.327 | 0.000 |
| female    | -0.327| 0.000 |
| sports    | 0.327 | 0.000 |
| romance   | -0.327| 0.000 |

| Discrepancy Function Value |

0.90876762814606D-029

Quality of Solution:
100.0000000000000

-----CorAna: End of run-----

CPU Seconds: 0.60000000000000D-001
Authors on Correspondence Analysis: See full references

Benzecri, J.P.
Everitt
Friendly
Goodman
Gorman and Primavera
Greenacre, M.
Hill
Hirschfield
Nishisato
Takane
Van de Geer
van der Heijden
von Eye.
Wickens
Weller and Romney
Item Response Theory and Binary Factor Analysis
Some Notes on Item Response Theory

XIV. The Two-Parameter Model

\[ P_i(\theta) = \frac{1}{1 + e^{-Da(\theta - b_i)}} \]
A. Definitions of terms
1. \( e = 2.718 \)
2. \( b = \) difficulty level; the point on the ability scale at which the probability of a correct response is 0.5; Normally, \(-3 < b < +3\)
3. \( a = \) Item discrimination parameter. This is the slope (tangent) of the curve at \( b \).
4. \( D = \) a constant. For 2- and 3-parameter models = 1.7, 1 for 1 parameter
5. \( \theta = \) Ability level in standard score units. The are in the range of \(-\infty < \theta < +\infty\)

XV. The One-parameter or Rasch Model

\[
P_i(\theta) = \frac{1}{1 + e^{-D(\theta - b_i)}}
\]

A. Only \( b \), the difficulty parameter, is important in this model \( D \) is assumed to be 1.0.
Item Response Theory

1. Item response theory (IRT) is built around the notion of the item characteristic curve (ICC).

   a. The curve is a function that relates a person's position on the underlying "Latent" trait to the probability that a person will endorse the item.

   b. The ordinate of the curve is the probability that a person will endorse an item.

   c. The abscissa of the curve is the ability level of the person in standardized units. We use the term "theta".

   d. Typically, the curve is a monotonically increasing function of ability levels, so that people who have more of the underlying trait will have a higher probability of passing or endorsing the item.

   e. We can describe item characteristic curves in terms of three parameters:

      (1) The difficulty level

          (a) This is the point on the ability axis that corresponds to 50% endorsement

      (2) The slope

          (a) This is the slope of the curve at the 50% point

          (b) An item with a steep slope has good discriminating ability

          (c) Conversely, an item with a little slope does a poor job of discriminating among individuals of different ability levels.

      (3) A guessing parameter

          (a) This is the probability of endorsing the item by chance alone. That is by guessing.

          (b) The is represented by the intercept of the curve on the ordinate.

(4) In practice, programs can be used for 1-, 2-, and 3-
The one-parameter model is called the Rasch model.

Thus, each item can be calibrated in terms of its difficulty level, its slope or discriminability, and a guessing parameter.

In effect, a test could consist of a single item because a person who passes the item could be estimated to have a corresponding position on the underlying trait continuum.

A scale developed from IRT approaches contains an assembly of calibrated items.

If the items are arranged in order and the items have reasonably steep slopes, it is not necessary to administer all items to a person because passing an item of a given difficulty level implies that the person would generally pass all items easier than that item.

Thus, adaptive testing strategies can be developed so that a person can take a minimal number of items.

A person’s score can be defined as the difficulty level of the highest difficulty item endorsed by the person.

In classical test theory, however, a person’s score is simply the number of items endorsed.

Although it is beyond the scope of the presentation time allotted, item response theory holds out the prospect that the same items can be calibrated in groups that differ widely in abilities. Furthermore, the calibration of items is somewhat independent of the rest of the items in the test.

Therefore, items can be "banked" and reassembled in different tests. Furthermore, norms tables can be assembled beforehand with knowledge of the ICC parameters.
Item Response and Factor Model
Item Response Example: NOHARM (PC version)
Fitting a (multidimensional) Normal Ogive
by Harmonic Analysis - Robust Method

Input File: class2a.in
Title: Three-item binary FA
Number of items = 3
Number of dimensions = 1
Number of subjects = 200
Convergence criterion = 0.0000100
Maximum function calls = 1000
An exploratory solution has been requested.

SAMPLE RAW PRODUCT-MOMENT MATRIX

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
</table>
| 1 | 0.500
| 2 | 0.250 | 0.500 |
| 3 | 0.250 | 0.350 | 0.500 |

ITEM COVARIANCE MATRIX

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.000</td>
<td>0.250</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>0.100</td>
<td>0.250</td>
</tr>
</tbody>
</table>

FIXED GUESSES

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

INITIAL CONSTANTS

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
</tbody>
</table>
FINAL CONSTANTS
1  2  3
-0.000 -0.000 -0.000

FINAL COEFFICIENTS OF THETA
1
1  1.6E-04
2  1.208
3  1.208

RESIDUAL MATRIX (lower off-diagonals)
1  2
2  1.9E-05
3  1.9E-05  1.7E-05

Sum of squares of residuals (lower off-diagonals) = 1.01587961284763E-0009
Root mean square of residuals (lower off-diagonals) = 1.84018079985603E-0005
Tanaka index of goodness of fit = 9.99999991790872E-0001

THRESHOLD VALUES
1  2  3
-0.000 -0.000 -0.000

UNIQUE VARIANCES
1  2  3
1.000 0.406 0.406

FACTOR LOADINGS
1
1  1.6E-04
2  0.770
3  0.770

(c) LORD’s PARAMETERIZATION - for the unidimensional case

VECTOR A : Discrimination parameters
1  2  3
1.6E-04  1.208  1.208

VECTOR B : Difficulty parameters
1  2  3
0.000  0.000  0.000
Authors on Item Response Theory and Binary Factor Analysis: See Full Refs.

Item Response Theory:

Andrich
Birnbaum
Bock
Embretson,
Goodman,
Hambleton, R. K. & Swaminathan, H.
Lazarsfeld, P. F.
Lord,
Lord, F. M. & Novick, M. R
Rasch, G..
Thissen,
Wright

Binary Factor Analysis
Batholomew
McDonald
Muthéén
Latent Class Analysis
Latent Class Model

Latent Class with Two Levels

Age

Gender

TV

e1

e2

e3
Simple Latent Class Example
LEM: log-linear and event history analysis with missing data.
Developed by Jeroen Vermunt (c), Tilburg University, The Netherlands.
Version 1.0 (September 18, 1997).
*** INPUT ***
*Simple Latent Class Model
* GenTV.dat
lat 1
  man 2
dim 2 2 2
lab X G T
mod {XG,XT}
rec 200
dat gentv.dat

*** STATISTICS ***
Number of iterations = 20
Converge criterion = 0.0000001974
Seed random values = 1561
X-squared = 0.0000 (0.0000)
L-squared = 0.0000 (0.0000)
Cressie-Read = 0.0000 (0.0000)
Dissimilarity index = 0.0000
Degrees of freedom = -2
Log-likelihood = -260.80230
Number of parameters = 5 (+1)
Sample size = 200.0
BIC(L-squared) = 548.0962
AIC(L-squared) = 531.6046
BIC(log-likelihood) = 10.5966
AIC(log-likelihood) = 4.0000
Eigenvalues information matrix
  394.2514   120.3975    68.6209     0.0265     0.0010

*** FREQUENCIES ***
G T observed estimated std. res.
1 1 70.000 69.999 0.000
1 2 30.000 30.001 -0.000
2 1 30.000 30.001 -0.000
2 2 70.000 69.999 0.000
### LOG-LINEAR PARAMETERS

* TABLE XGT [or P(XGT)] *

<table>
<thead>
<tr>
<th>effect</th>
<th>beta</th>
<th>std err</th>
<th>z-value</th>
<th>exp(beta)</th>
<th>Wald</th>
<th>df</th>
<th>prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>main</td>
<td>2.7064</td>
<td></td>
<td></td>
<td>14.9758</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.0573</td>
<td>26.7397</td>
<td>-0.002</td>
<td>0.9443</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0573</td>
<td></td>
<td>1.0590</td>
<td>0.00</td>
<td>1 0.998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.0260</td>
<td>12.1165</td>
<td>-0.002</td>
<td>0.9744</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0260</td>
<td></td>
<td>1.0263</td>
<td>0.00</td>
<td>1 0.998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.0258</td>
<td>12.0558</td>
<td>-0.002</td>
<td>0.9745</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0258</td>
<td></td>
<td>1.0262</td>
<td>0.00</td>
<td>1 0.998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>-0.7473</td>
<td>4.4107</td>
<td>-0.169</td>
<td>0.4736</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2</td>
<td>0.7473</td>
<td></td>
<td>2.1113</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 1</td>
<td>0.7473</td>
<td></td>
<td>2.1113</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2</td>
<td>-0.7473</td>
<td></td>
<td>0.4736</td>
<td>0.03</td>
<td>1 0.865</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>-0.7451</td>
<td>4.3809</td>
<td>-0.170</td>
<td>0.4747</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2</td>
<td>0.7451</td>
<td></td>
<td>2.1067</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 1</td>
<td>0.7451</td>
<td></td>
<td>2.1067</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2</td>
<td>-0.7451</td>
<td></td>
<td>0.4747</td>
<td>0.03</td>
<td>1 0.865</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### (CONDITIONAL) PROBABILITIES

* P(XGT) *

| 1 1 1  | 0.151  | (0.8125) |
| 1 1 2  | 0.0705 | (2.1055) |
| 1 2 1  | 0.0709 | (2.0980) |
| 1 2 2  | 0.3312 | (0.9013) |
| 2 1 1  | 0.3349 | (0.8135) |
| 2 1 2  | 0.0795 | (2.1053) |
| 2 2 1  | 0.0791 | (2.0978) |
| 2 2 2  | 0.0188 | (0.9010) |

### LATENT CLASS OUTPUT

<table>
<thead>
<tr>
<th>X 1</th>
<th>X 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4877</td>
<td>0.5123</td>
</tr>
<tr>
<td>G 1</td>
<td>G 2</td>
</tr>
<tr>
<td>0.1756</td>
<td>0.8089</td>
</tr>
<tr>
<td>T 1</td>
<td>T 2</td>
</tr>
<tr>
<td>0.1763</td>
<td>0.8082</td>
</tr>
<tr>
<td>E = 0.1753, lambda = 0.6406</td>
<td></td>
</tr>
</tbody>
</table>
Latent Class Authors: See Full Refs.
Bartholomew
Clogg,
Clogg, C. C. & Shihadeh
Goodman,
Hagenaars
Heinen
Lazarsfeld,
McCutcheon,
Vermunt
Wickens,
Path and Graphical Models
Path Analysis

Age

Gender

TV Preference

Other
Simple Path Analysis Problem
LEM: log-linear and event history analysis with missing data.
Developed by Jeroen Vermunt (c), Tilburg University, The Netherlands.
Version 1.0 (September 18, 1997).

*** INPUT ***
* path
 * Clasex dat
 man 3
dim 2 2 2
lab A G T
mod GA
 T|GA
 rec 8
 rco
dat crosex1.dat

*** STATISTICS ***
Number of iterations = 2
Converge criterion = 0.0000000000
X-squared = 0.0000 (0.0000)
L-squared = 0.0000 (0.0000)
Cressie-Read = 0.0000 (0.0000)
Dissimilarity index = 0.0000
Degrees of freedom = 0
Log-likelihood = -394.60028
Number of parameters = 7 (+1)
Sample size = 200.0
BIC(L-squared) = 0.0000
AIC(L-squared) = 0.0000
BIC(log-likelihood) = 826.2888
AIC(log-likelihood) = 803.2006

Eigenvalues information matrix
200.0000 200.0000 200.0000 192.0000 192.0000 128.0048
127.9722

*** FREQUENCIES ***
A G T observed estimated std. res.
1 1 1 40.000 40.000 0.000
1 1 2 10.000 10.000 0.000
1 2 1 10.000 10.000 0.000
1 2 2 40.000 40.000 0.000
2 1 1 30.000 30.000 0.000
2 1 2 20.000 20.000 0.000
2 2 1 20.000 20.000 0.000
2 2 2 30.000 30.000 0.000
### *** PSEUDO R-SQUARED MEASURES ***

#### * P(T|AG) *

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Fitted</th>
<th>R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>0.6931</td>
<td>0.5867</td>
<td>0.1536</td>
</tr>
<tr>
<td>Qualitative variance</td>
<td>0.2500</td>
<td>0.2000</td>
<td>0.2000</td>
</tr>
<tr>
<td>Classification error</td>
<td>0.5000</td>
<td>0.3000</td>
<td>0.4000</td>
</tr>
<tr>
<td>(-2/N \times \log \text{-likelihood})</td>
<td>1.3863</td>
<td>1.1734</td>
<td>0.1536/0.1755</td>
</tr>
<tr>
<td>Likelihood^(-2/N)</td>
<td>4.0000</td>
<td>3.2330</td>
<td>0.1917/0.2557</td>
</tr>
</tbody>
</table>

### *** LOG-LINEAR PARAMETERS ***

#### * TABLE AG [or P(AG)] *

<table>
<thead>
<tr>
<th>Effect</th>
<th>Beta</th>
<th>Std Err</th>
<th>Z-value</th>
<th>Exp(Beta)</th>
<th>Wald</th>
<th>df</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.0000</td>
<td>0.0707</td>
<td>0.0000</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0.0000</td>
<td>0.0707</td>
<td>0.0000</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>0.0000</td>
<td>0.0707</td>
<td>0.0000</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### * TABLE AGT [or P(T|AG)] *

<table>
<thead>
<tr>
<th>Effect</th>
<th>Beta</th>
<th>Std Err</th>
<th>Z-value</th>
<th>Exp(Beta)</th>
<th>Wald</th>
<th>df</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>-0.0000</td>
<td>0.0807</td>
<td>0.0000</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT</td>
<td>-0.0000</td>
<td>0.0807</td>
<td>0.0000</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GT</td>
<td>0.4479</td>
<td>0.0807</td>
<td>5.551</td>
<td>1.5651</td>
<td>30.82</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>AGT</td>
<td>0.2452</td>
<td>0.0807</td>
<td>3.039</td>
<td>1.2779</td>
<td>9.23</td>
<td>1</td>
<td>0.002</td>
</tr>
</tbody>
</table>
### (Conditional) Probabilities

#### $P(AG)$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Probability</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.2500</td>
<td>0.0306</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.2500</td>
<td>0.0306</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.2500</td>
<td>0.0306</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.2500</td>
<td>0.0306</td>
</tr>
</tbody>
</table>

#### $P(T|AG)$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Probability</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.8000</td>
<td>0.0566</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.2000</td>
<td>0.0566</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.2000</td>
<td>0.0566</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.8000</td>
<td>0.0566</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.6000</td>
<td>0.0693</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.4000</td>
<td>0.0693</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.6000</td>
<td>0.0693</td>
</tr>
</tbody>
</table>
Authors on Graphical Modeling: See full references

Badsberg,
Christensen
Edwards
Kreiner
Landwehr
Lauritzen,
Wickens
Sources of Free Software

Analytitech

http://www.analytictech.com/

Steve Borgatti has an incredible site for social network analysis. Many of the programs are free and the “commercial” ones are a bargain.

Badberg’s CoCo site:

http://www.math.auc.dk/~jhb/CoCo/cocoinfo.html

CoCo is a powerful graphical modeling program

Michael Browne’s site:

http://quantrm2.psy.ohio-state.edu/browne/

A great program, CORANA, for correspondence analysis.

Scott Eliason’s site:

http://www.soc.umn.edu/~eliason/CDAS.htm

Eliason was a colleague and student of Clogg’s. His site has an excellent package, CDAS, which computes loglinear, correspondence, and latent class models.

Michael Friendly’s site:

http://www.math.yorku.ca/SCS/friendly.html

This has a huge amount of information on graphic display of categorical data and loads of links to other places.

Linda Collin’s site:

http://methodology.psu.edu/winlta.html

Collins is a leader in latent class analysis of change data. Her program, WINLTA, computes many latent class models.
Svend Kreiner’s site:

http://www.biostat.ku.dk/staff/skm-e.htm

Kreiner is a major figure in graphical modelling of categorical data. His DIGRAM program can compute many path models.

Pierre: Legendre’s site:

http://www.fas.umontreal.ca/BIOL/legendre/

A great place for some good correspondence analysis programs.

Roderick McDonald’s Software.

http://kiptron.psyc.virginia.edu/disclaimer.html

The best place for this is from Jack McCardle’s site at University of Virginia. Look for the NOHARM program. It does IRT and exploratory and confirmatory binary factor analysis.

STATA Corporation’s Links to Software Providers

www.stata.com

STATA is a major statistical package. The company also has numerous links to other free and commercial software providers.

John Uebersax’s Site:

http://ourworld.compuserve.com/homepages/jsuebersax

This is a goldmine for finding out about latent class analysis. In addition, it has links to dozens of other sites and to free latent class software.
Vermunt’s LEM site:

http://www.uvt.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html

LEM is a program than can compute nearly any categorical analysis problem. Its program commands are very logical. The program and it’s manuals are free. Vermunt also co-authored an extremely sophisticated nw latent class analysis program LatentGOLD (J. VERMUNT & J.Magidson) which you can see (and buy) at www.latentgold.com

Alexander von Eye’s site:

http://www.msu.edu/user/voneye/cfa.htm

Alexander von Eye wrote numerous books and articles on configural frequency analysis and prediction logic. His site has several useful programs.

Wright’s MESA site:

www.mesa.com

Benjamin Wright is a major writer in latent trait and item responses theory. His site has a wealth of literature on Item Response theory and and some free and commercial programs.

Zhang (and others) Regression Tree Site:

http://www.kdcentral.com/Software/Machine_Learning/Regression_Trees/more2.shtml

This site has links to several very good (and free) regression tree programs.
References


Bock, R. D. (1972). Estimating item parameters and latent ability when the responses are scored
in two or more nominal categories. *Psychometrika*, 37, 29-51.


What it says.


Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests . (Copenhagen: Danmarks Paedogogiske Institut.)


