A: 8.5.5

2. Measure reliability using proportion of agreement.
   a. Excess of observed proportion agreeing over $p_e = \sum_i \pi_{i+}\pi_{+i} =$ expected under independence.
      i. Expectation same as for χ^2 test
      ii. All divided by its maximal value $1 - p_e$
      iii. Result is called *kappa statistic*. R Code

:

G. Reliability

1. Reliability Question
   a. Do various measurements (items) contributing to scale all measure the same thing?
2. Notation:
   a. I items over all.
   b. X_{ji} is measurement of item i from subject j
   c. $X_{j.} = \sum_{i=1}^{I} X_{ji}$
   d. s_i is standard deviation of item i, with σ_i the population version
   e. $s.$ is standard deviation of total, with $\sigma.$ the population version

3. In this context, redundant variables are a good sign.
   a. If X_{ji} are independent, $\mathrm{Var}\left[X_{j}.\right] = \sum_{i=1}^{I} \mathrm{Var}\left[X_{ji}\right]$
   b. If X_{ji} are exact copies of one another, $\mathrm{Var}\left[X_{j}.\right] = I^2 \mathrm{Var}\left[X_{ji}\right] \forall i$.
   c. More generally, $\mathrm{Var}\left[X_{j}.\right] = \sum_{i=1}^{I} \mathrm{Var}\left[X_{ji}\right] + \sum_{i=1}^{I-1} \sum_{\ell=i+1}^{I}$ Cov $\left[X_{ji}, X_{j,\ell}\right]$
   d. The more the items contain redundant information, the smaller is $\sum_{i=1}^{I} \mathrm{Var}\left[X_{ji}\right] / \mathrm{Var}\left[X_{j}.\right]$
   e. If variables are negatively correlated, chance teh sign of these items.
      i. Otherwise summation cancels data out.
4. Measure reliability
   a. Use $1 - \sum_{i=1}^{I} \mathrm{Var}\left[X_{ji}\right] / \mathrm{Var}\left[X_{j}.\right]$
      i. By same reasoning as for dividing by $n-1$ rather than n for the SD, rescaled by $I/(I-1)$
      ii. Generally replace σ_i and $\sigma.$ by s_i and $s.$ resp.
   b. Define Cronbach's $\alpha = \frac{I}{I-1}(1 - \sum_{i=1}^{I} s_i^2/s_{\cdot}^2)$
      i. Also called Guttman's γ_3, K-R 20, etc.
   c. A collection of measurements might measure the same thing,

but might be on different scales.

i. In this case, apply alpha to the standardized values

ii. There's a more direct formula in terms of the correlations. R Code

A: 5.4.6

XI. Exact Methods

A. Contingency tables:

1. Model:

a. $X_{ij} \sim \mathsf{P}(\lambda_{ij})$

b. $\log(\lambda_{ij}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$

c. $\alpha_0 = 0$, $\beta_0 = 0$, $\gamma_{i0} = \gamma_{0j} = 0 \forall i, j$.

d. $H_0 : \gamma_{ij} = 0 \forall i, j$ vs. $H_A : \gamma_{ij} \neq 0$ for some i, j.

2. Test statistics:

a. Score statistic is Pearson χ^2: $T = \sum_{i,j}(X_{ij} - X_{i\cdot}X_{\cdot j}/X_{\cdot\cdot})^2/(X_{i\cdot}X_{\cdot j}/X_{\cdot\cdot})$.

b. LR statistic

c. Fisher's statistic $1/\mathsf{P}[\boldsymbol{X}]$

3. Remove effect of unknown parameters:

a. Remove α_i by conditioning on $X_{i\cdot}$

i. Reduces $I \times J$ independent Poisson variables to J independent multinomials, each with I bins.

ii. $X_{i.}$ are exactly ancillary

iii. Little loss due to discreteness

b. Remove β_j by conditioning on $X_{.j}$

i. Reduces J independent multinomials, each with I bins, to generalized geometric

ii. Probabilities are $(\prod_{i=1}^{I} x_{i.}!)(\prod_{j=1}^{J} x_{.j}!)/(x_{..}! \prod_{i=1}^{I} \prod_{j=1}^{J} x_{ij}!)$

iii. Violates conditionality principal: column totals are not ancillary

iv. Bigger problem: discreteness

4. Computation

a. Either enumerate all tables, and calculate probabilities straight–forwardly, or

b. (Pagano and Halvorsen, 1981) calculate recursively

i. $\mathsf{P}\left[X_{11} = x_{11} | X_{1.}, X_{.1}, X_{..}\right]$

$$= \frac{x_{1.}! x_{.1}! (x_{..} - x_{1.})! (x_{..} - x_{.1})!}{x_{..}! x_{11}! (x_{.1} - x_{11})! (x_{1.} - x_{11})! (x_{..} - x_{1.} - x_{.1} + x_{11})!}$$

ii. Probabilities do not depend on other aspects of conditioning event.

iii. Hence same expression holds for $\mathsf{P}\left[X_{11} = x_{11} | X_{i\cdot}, X_{\cdot j} \forall i, j\right]$

iv. $\mathsf{P}\left[X_{21} = x_{21} | X_{11} = x_{11}, X_{1\cdot}, X_{\cdot 2}, X_{\cdot 1}, X_{\cdot\cdot}\right]$
$= x_{2\cdot}!(x_{\cdot\cdot} - x_{1\cdot} - x_{2\cdot})!(x_{\cdot 1} - x_{11})!(x_{\cdot\cdot} - x_{2\cdot} - x_{\cdot 1} + x_{11})!/$

$[(x_{\cdot\cdot} - x_{1\cdot})!x_{21}!(x_{\cdot 1} - x_{11} - x_{21})!(x_{2\cdot} - x_{21})!$

$\times (x_{\cdot\cdot} - x_{1\cdot} - x_{\cdot 1} - x_{2\cdot} + x_{11} + x_{21})!]$

v. More generally, split table into 9 bits, some possibly empty: $\begin{pmatrix} * & * & \ddagger \\ * & X_{ij} & \dagger \\ * & \dagger & \dagger \end{pmatrix}$

- Condition on all marginals of collapsed table, and on totals marked $*$.
- $\ddagger$ cell fixed by conditioning event as well
- Result is hypergeometric distribution with rows and columns containing $\dagger$.
- As before, collapse over rows and columns containing $\dagger$ to obtain a hypergeometric distribution from 2×2 table.
- Hence $\mathsf{P}\left[\boldsymbol{X} = \boldsymbol{x} | X_{\cdot j} = x_{\cdot j} \forall j, X_{i\cdot} = x_{i\cdot} \forall i\right] = \prod_{ij} p_{ij}$ for hypergeometric probabilities p_{ij}. SAS Code R Code R Code

A: 5.4.4–5.4.5

B. Logistic Regression

1. The Logistic Regression Model and Notation

a. $X_j \sim \mathsf{Bin}(\exp(\boldsymbol{z}_j\boldsymbol{\theta})/(1+\exp(\boldsymbol{z}_j\boldsymbol{\theta}), n_j)$

b. $\boldsymbol{T} = \boldsymbol{Z}^\top \boldsymbol{X}$

c. Probabilities $\exp(\boldsymbol{t}^\top\boldsymbol{\theta} - \sum_j n_j \log(1+\exp(\boldsymbol{z}_j\boldsymbol{\theta}))c(\boldsymbol{t})$, for

i. $c(\boldsymbol{t})$ the number of $\boldsymbol{x}$ vectors with $\boldsymbol{Z}^\top\boldsymbol{x} = \boldsymbol{t}$.

2. Conditional probabilties

a. Calculate these to remove effect of parameter not of immediate interest.

b. $\boldsymbol{T} = (\boldsymbol{U}, \boldsymbol{V})$, $\boldsymbol{\theta} = (\boldsymbol{\omega}, \boldsymbol{\tau})$

c. $\mathsf{P}[\boldsymbol{V} = \boldsymbol{v}|\boldsymbol{U} = \boldsymbol{u}] = c(\boldsymbol{u}, \boldsymbol{v})\exp(\boldsymbol{v}\boldsymbol{\tau})/\sum_{\boldsymbol{v}} c(\boldsymbol{u}, \boldsymbol{v})\exp(\boldsymbol{v}\boldsymbol{\tau})$

d. Allows construction of null distribution for tests of $\boldsymbol{\tau}$ without knowing $\boldsymbol{\omega}$.