blocking ←cf.→ Matched Pairs Design (MPD)

block 1 (pair 1) hetro-geneous block b (pair b)

EU11 A, EU12 B . . . EUb1 A, EUb2 B

- Design # of blocks DOE homogeneous
  - A block factor: y_{++} levels, each level represents a block, each block of size 2, i.e., 2 experimental units (EUs) in one block

factor of no interest, but has effect on response

successful blocking: • between-block variation large • within-block variation: small

EUs: hetrogeneous between blocks, homogeneous within blocks

  - A treatment factor: 2 levels A and B, randomly assigned to the 2 EUs in each blocks

factor of main interest

  - A response variable: categorical same characteristic
  - 2 formats of representing data

block	X_1	X_2
pair 1	1	3
pair 2	2	1
⋮	⋮	⋮

response Y_x

block	treatment	X	1	2	3
pair 1	A	1	1	0	0
pair 1	B	3	0	0	1
pair 2	A	2	0	1	0
pair 2	B	1	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮

count data

∵ X_1, X_2 same characteristic

  - Comparison 1: MPD ↔ MCCD → (prospective vs. retrospective studies)
  - Comparison 2: MPD ↔ Paired sample t-test → (qualitative vs. quantitative response)

apply 1-sample t-test on $X_1 - X_2$

- Data for contingency table: observe *one* type of categorical measure on *two matched* objects (EUs) in one block

use ★

  - In contrast, in the typical 2-way contingency table, observe *two* (different) types of categorical measures (X_1 and X_2) on *one* object

check graph in LNp.5-1

object block

  - e.g., left (X_1) and right (X_2) eye performance of a person

X_1 \ X_2	1	$\cdots$	I	
1	π_{11}	$\cdots$	π_{1I}	π_{1+}
$\vdots$	$\cdots$		$\cdots$	$\cdots$
I	π_{I1}	$\cdots$	π_{II}	π_{I+}
	π_{+1}	$\cdots$	π_{+I}	1

same [Y_{ij}] a square matrix

- Contingency table for matched pair data is a square matrix and

[Y_{ij}] → marginals Y_{i+}, Y_{+j}, Y_{++} X_1 and X_2 have same # of categories

defined in terms of π_{ij}'s

  - no marginal totals are fixed in advance → row & column totals are random
  - grand total Y_{++} could be random or fixed → modeling (Δ) — scheme 1 (LNp.5-5), scheme 2 (LNp.5-8) → can get all information of π_{ij}'s

- **Q**: what questions are of interest for matched pair data?

A & B no difference on ...

  - row and column marginals are *homogeneous*, i.e., $\pi_{i+}=\pi_{+i}$? → X_1 & X_2 have same marginal dist ←cf.→ 2-sample t-test (check the table & ★ in LNp.5-18)
  - $[\pi_{ij}]_{I\times I}$ is a *symmetric* matrix, i.e., $\pi_{ij}=\pi_{ji}$?
    - symmetry implies marginal homogeneity (MH), but, the reverse statement not necessarily true (except for 2×2 table)

$\pi_{i+} = \sum_j \pi_{ij} = \sum_j \pi_{ji} = \pi_{+i}$ ∵ MH ∵ symmetry

$\pi_{11} + \pi_{12} = \pi_{1+} = \pi_{+1} = \pi_{11} + \pi_{21}$

$\log(\gamma_{ij}) = \log(\pi_{ij}) - \log(\pi_{i+}) - \log(\pi_{+j})$ interaction ⇕ association (check LNp.5-6)

    - **Q**: how to interpret symmetry? → $\pi_{ij}/\pi_{i+} = \pi_{ji}/\pi_{+i} \Rightarrow \pi_{X_2=j|X_1=i} = \pi_{X_1=j|X_2=i}$ ex.

  - When row and column marginal totals are quite different, might be interested in whether i.e., MH does not hold ⇒ symmetry does not hold

symmetry MH QS (*)

$\gamma_{ij} \equiv \frac{\pi_{ij}}{\pi_{i+}\pi_{+j}} \Leftrightarrow \pi_{ij}=\pi_{i+}\pi_{+j}\gamma_{ij}$, where $\gamma_{ij}=\gamma_{ji}$? → $[\gamma_{ij}]$ is a symmetric matrix

A & B have no difference on association

  - It is called *quasi-symmetry* (QS)

(⇐) $\gamma_{ij} = \frac{\pi_{ij}}{\pi_{i+}\pi_{+j}} = \frac{\pi_{ji}}{\pi_{+i}\pi_{j+}} = \gamma_{ji}$ (⇒) $\pi_{ij} = \pi_{i+}\pi_{+j}\gamma_{ij}$, $\pi_{ji} = \pi_{+i}\pi_{j+}\gamma_{ji}$

    - MH + QS ⇔ symmetry

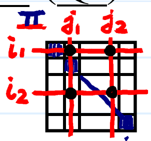
usually rejected because π_{ij}'s usually large due to blocking

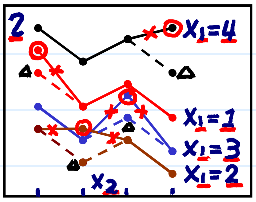
  - X_1 and X_2 are *independent*, i.e., $\pi_{ij}=\pi_{i+}\pi_{+j}$ for all i and j?

- If not independent, whether $\pi_{ij}=a_i \times b_j$ for $i \neq j$? It is called *quasi-independent* (QI).

often ∵ π_{ii}'s are large.

Note. $a_i \neq \pi_{i+}$, $b_j \neq \pi_{+j}$ | ignore diagonal components





- Q: how to interpret QI?

$i_1, i_2 \neq j_1, j_2$ | $\pi_{i_1j_1} \pi_{i_2j_2} = \pi_{i_2j_1} \pi_{i_1j_2} \Rightarrow$ odds ratio $\Delta = 1$ for any 4 such cells not including diagonals

- Tests for these hypotheses based on log-linear model, e.g., Poisson GLM (why? check (Δ) in LNp.5-19)

X_1 \ X_2	1	2	3	
1	y_{11}	y_{12}	y_{13}	y_{1+}
2	y_{21}	y_{22}	y_{23}	y_{2+}
3	y_{31}	y_{32}	y_{33}	y_{3+}
	y_{+1}	y_{+2}	y_{+3}	y_{++}

$$\underset{\text{response vector}}{Y} = (\underbrace{y_{11}, y_{21}, y_{31}}_{\text{1st column}}, \underbrace{y_{12}, y_{22}, y_{32}}_{\text{2nd column}}, \underbrace{y_{13}, y_{23}, y_{33}}_{\text{3rd column}})^T$$

- Test for symmetry (H_0) hypothesis: H_1: saturated model

$\frac{I^2-I}{2}+I$: # of levels

- Generate a vector with I^2 components for an $(I(I+1)/2)$-level nominal factor with the structure: $\eta_{ij} \propto \log(\pi_{ij})$

l_1	l_2	l_3
l_2	l_4	l_5
l_3	l_5	l_6

$\eta_{ij} = \log(y_{++}) + \log(\pi_{ij}) = \beta_0 + \Sigma_k \beta_k d_k$, function of i & j

$$\text{sym-factor} = (\underbrace{l_1, l_2, l_3}_{\text{1st}}, \underbrace{l_2, l_4, l_5}_{\text{2nd}}, \underbrace{l_3, l_5, l_6}_{\text{3rd}})^T$$

# of dummy variable d_k's $= \frac{I(I+1)}{2} - 1$

- $Y \sim$ sym-factor $\equiv S_{\text{sym}}$
- Deviance-based/Pearson X^2 goodness-of-fit test for S_{sym}

- Test for QS (H_0) hypothesis H_1: saturated

$$\eta_{ij} - \beta_0 = \log(\pi_{ij}) = \log(\pi_{i+}\, \pi_{+j}\, \gamma_{ij}) = \underset{\text{function of } X_1 \text{ only}}{\log(\pi_{i+})} + \underset{\text{function of } X_2 \text{ only}}{\log(\pi_{+j})} + \underset{\text{(symmetric) } \gamma_{ji}}{\log(\gamma_{ij})}$$

$S_{\text{sym}} \subseteq S_{\text{qsym}}$

- $Y \sim X_1 + X_2 + \text{sym-factor} \equiv S_{\text{qsym}}$ ← could be an over-parameterized model, e.g., $1+2+2+5 > 9$, but OK for
- Deviance-based/Pearson X^2 goodness-of-fit test for S_{qsym}

- Test for MH (H_0) hypothesis
  - No log-linear models that directly corresponds to MH

Consider the log-linear model: $Y \sim X_1 + X_2 + X_1{:}X_2$, where the parameters corresponding to X_1 & X_2 respectively are assumed to be identical ⇒ cannot guarantee MH ∵ the term $X_1{:}X_2$

row totals Y_{i+}'s ~ Poisson/multinomial
column totals Y_{+j}'s ~ Poisson/multinomial
problem: Y_{i+}'s & Y_{+j}'s not indep. | Test whether the 2 joint dist. are identical (cf. 2-sample t)

not a universal MH test

  - An indirect test using log-linear models **when S_{qsym} already holds**
  - $H_0 \cup H_1 = S_{\text{qsym}}$ ← Deviance-based test for H_0: S_{sym} vs. H_1: $S_{\text{qsym}} \backslash S_{\text{sym}}$ ← all the models in this H_0 are of MH ((*), LNp.5-10)
  - Other approaches, see Agresti (2013), 11.3

- Test for QI (H_0) hypothesis
  - Approach 1

maximizing scheme 2 (LNp 5-8) multinomial likelihood under MH constraints to get $\hat{\mu}_{ij} = y_{++}\hat{\pi}_{ij}$. Then, use X^2 or deviance to compare Y_{ij}'s and $\hat{\mu}_{ij}$'s

   - Omit the diagonal data, i.e., let

$$Y' = (y_{21}, y_{31}, y_{12}, y_{32}, y_{13}, y_{23})^T \quad \cancel{y_{11}}\ \cancel{y_{22}}\ \cancel{y_{33}}$$

main-effects model ⇔ no interactions ⇔ independent on Y' components

   - $Y' \sim X_1 + X_2 \equiv S_{\text{qindep1}}$
   - Deviance-based/Pearson X^2 goodness-of-fit test for S_{qindep1}

  - Approach 2 ← The 2 approaches would generate same test result

I dummy variable, each for 1 diagonal component, e.g.,
- l_0: reference
- l_1, l_2, l_3: each coded as 1 in 1 dummy variable

- Generate a vector with I^2 components for an $(I+1)$-level nominal factor with the structure:

l_1	l_0	l_0
l_0	l_2	l_0
l_0	l_0	l_3

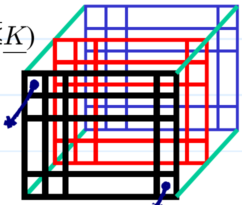
$$\text{QI-factor} = (\underbrace{l_1, l_0, l_0}_{\text{1st}}, \underbrace{l_0, l_2, l_0}_{\text{2nd}}, \underbrace{l_0, l_0, l_3}_{\text{3rd}})^T$$

not Y'

- $Y \sim X_1 + X_2 + \text{QI-factor} \equiv S_{\text{qindep2}}$
- Deviance-based/Pearson X^2 goodness-of-fit test for S_{qindep2}

Under S_{qindep2} model.
① $\hat{\pi}_{ii} = y_{ii}/y_{++} \Leftarrow \hat{\mu}_{ii} = y_{++}\hat{\pi}_{ii} = y_{ii}$
$\Rightarrow$ diagonals y_{ii}'s have no contribution to deviance or X^2
② add I more counts in Y, but also add I more parameters in $\eta = X\beta$
$\Rightarrow$ cancelling out & same d.f.

❖ **Reading**: Faraway (2006, 1st ed.), 4.3



X_3 $(1 \le k \le K)$

X_1 $(1 \le i \le I)$

X_2 $(1 \le j \le J)$

3 crossing covariates

Three-Way Contingency Table

- The π's and Y's are defined in the same manner as in the 2-way table

5/5

π_{ijk} → marginal probabilities

$\boxed{X_1}$ $\boxed{X_2}$ $\boxed{X_3}$

Y_{ijk} → marginal totals

- Poisson GLM approach to investigate how X_1, X_2, X_3 interact

(LNp.5-4) Recall Poisson GLM can be applied to SS1~4 but some information could be meaningless (LNp.5-13)

➢ Mutual independence (X_1, X_2, X_3 are independent)

$Y_{ijk} \overset{\text{indep}}{\sim} P(\mu_{ijk})$, $\mu_{ijk} = t\,\pi_{ijk}$ $\xrightarrow{\log}$ $X\beta = \eta_{ijk}$

joint = π marginals

- $\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k} \Leftrightarrow P(X_1=i, X_2=j, X_3=k) = P(X_1=i)P(X_2=j)P(X_3=k)$

function of X_1 only

- $\log(\pi_{ijk}) = \log(\pi_{i++}\pi_{+j+}\pi_{++k}) = \log(\pi_{i++}) + \log(\pi_{+j+}) + \log(\pi_{++k})$ $\boxed{X_3}$ $\boxed{X_2}$

$= \eta_{ijk} - \beta_0$