ACADEMIC WASTE
A platform for structurally failure and unnecessary thought.

RESEARCH ARTICLE

基于 MBTI-16 类型总体分布的联合正态分布 Probit 建模

*Spade 14**

MBTI 是什么

MBTI 把性格分成四个“偏好维度”，每个维度两种取向：外向 E-内向 I，感觉 S -直觉 N，思考 T -情感 F，判断 J-知觉 P。四个维度组合起来形成 16 种类型（如 INFJ、ESTP 等）。我们这里用到的是“总体分布”（每种类型占比多少），不是个人作答的原始数据。分布会因样本人群（国家、行业、年龄层）而不同。它是一套用来描述“我们比较倾向怎样做事/看世界”的性格偏好工具，不是考试，也没有好坏对错。本报告的目标是：在只有 16 类型“占比”的前提下，尽量还原这四个维度之间的“整体联动关系”。

1. 输入（比例）与要点

类型比例（原始，合计≈100.3%）：ISFJ 13.8%，ESFJ 12.3%，ISTJ 11.6%，ISFP 8.8%，ESTJ 8.7%，ESFP 8.5%，ENFP 8.1%，ISTP 5.4%，INFP 4.4%，ESTP 4.3%，INTP 3.3%，ENTP 3.2%，ENFJ 2.5%，INTJ 2.1%，ENTJ 1.8%，INFJ 1.5% 建模前归一化为概率（合计=1）。

四维边际（归一化后）：I(1) ≈ 0.507，N(1) ≈ 0.268，F(1) ≈ 0.597，J(1) ≈ 0.541。要点：N 比例偏少（≈27%），I/F/J 过半。

2. 方法（联合正态分布 Probit + 最大似然）

变量与阈值：I/N/F/J 记为 1（E/S/T/P 记为 0），INFJ = 1111。由观测边际 p_i 反推阈值 $\tau_i = \Phi^{-1}(1-p_i)$ 并固定，确保模型边际与观测边际一致。

似然与 MLE：对每个类型 $y \in \{0,1\}^4$，四维正态在对应长方体的概率为 $q_y(R)$。以归一化类型比例 p_y 为权重，最大化加权对数似然：$L(R) = \Sigma\, p_y \cdot \log q_y(R)$（采用最优化 MLE）。

未知数 VS 输入：阈值固定后，未知参数仅为相关矩阵 R 的 6 个非对角元素；输入为 16 个单元格比例，但“合计=1”⇒所以独立自由度 15，6 远小于 15 →识别性良好，采用 MLE 是合理的。

Probit 标准化：各维方差=1，因此估计的相关矩阵可直接视为协方差矩阵。

3.估计的相关矩阵

R（=协方差矩阵；顺序 I、N、F、J）

	I	N	F	J
I	1.000	-0.177	-0.122	0.090
N	-0.177	1.000	0.029	-0.487
F	-0.122	0.029	1.000	-0.150
J	0.090	-0.487	-0.150	1.000

要点：N–J 负相关最强（-0.487）；I–N 与 F–J 为轻度负相关；其余相关较弱

4. 16 类型“观测 vs 拟合”

(resid = p_obs − p_fit)

二进制	类型	p_obs	p_fit	resid
0000	ESTP	4.29%	4.08%	0.21%
0001	ESTJ	8.67%	8.79%	-0.11%
0010	ESFP	8.47%	8.61%	-0.14%
0011	ESFJ	12.26%	12.32%	-0.06%
0100	ENTP	3.19%	3.55%	-0.36%
0101	ENTJ	1.79%	1.85%	-0.06%
0110	ENFP	8.08%	7.48%	0.60%
0111	ENFJ	2.49%	2.57%	-0.08%
1000	ISTP	5.38%	5.31%	0.07%
1001	ISTJ	11.57%	12.06%	-0.50%
1010	ISFP	8.77%	8.76%	0.02%
1011	ISFJ	13.76%	13.25%	0.51%
1100	INTP	3.29%	3.06%	0.23%
1101	INTJ	2.09%	1.58%	0.52%

二进制	类型	p_obs	p_fit	resid
1110	INFP	4.39%	5.01%	-0.63%
1111	INFJ	1.50%	1.72%	-0.22%

大多数类型的拟合概率与观测比例仅存在微笑差距，少数（尤其含 N 类型）出现可关注偏的差，提示主要改进空间在依存结构而非边际上。

5. 边际一致（I/N/F/J 的 1 侧）

边际(1 侧)	obs	fit
I(1)	50.7%	50.7%
N(1)	26.8%	26.8%
F(1)	59.7%	59.7%
J(1)	54.1%	54.1%

各维度的一例比例在模型与观测之间几乎完全一致，阈值设定正确，整体框架稳定。

6. 主成分分析

固有值与方差贡献

主成分	固有值	解释率	累积解释率
PC1	1.596	39.91%	39.91%
PC2	1.020	25.50%	65.40%
PC3	0.899	22.48%	87.89%
PC4	0.485	12.11%	100.00%

变量	PC1	PC2	PC3	PC4
I	0.341	-0.507	0.774	0.165
N	-0.637	-0.337	-0.087	0.688
F	-0.263	0.755	0.568	0.197
J	0.639	0.245	-0.266	0.679

要点解读：PC1（39.9%）：N(−) vs J(+) 的强对立（I 小幅+，F 小幅-），与 N-J 的强负相关一致，可视为“N↔J 反向轴”。PC2（25.5%）：F(+) 显著，I (-) 与 N（-）偏弱，J 小幅+，可理解为“F 优势-I/N 抑制”的走向。PC3（22.5%）：I(+)与 F(+) 合轴，J 为-，N 极弱-，用于补充 I/F 的同向变化模式。PC4（12.1%）:N(+) 与 J(+) 同向(I/F 小幅+), 在与 PC1 直交的前提下提供 N/J 同向的辅助维度。

7. 结论

边际稳定：I/N/F/J 的 1 侧，观测与拟合一致。负相关是关键信号：N–J 最显著（约-0.487），I-N/F-J 为轻度负相关；总体匹配度较好。差异集中：N 相关类型（ENFP/INFP/INTJ 等）出现 0.5-0.6%差异→单靠线性、对称的联合正态不足以完全解释。

8. 后续研究：Copula

研究目标是在保持边际分布不变的前提下，放宽多变量的依存结构，以便更准确地刻画尾部相关与非对称性；为此，首先考虑以 t-copula 建模，以反映重尾与上下尾同时极端时的依赖增强，其次采用 vine copula 通过 pair copula 分解在"变量对"层面进行灵活指定，使依存结构在不同维度对上具有差异化与可解释性。推定上，鉴于离散格子数据的完全尤度计算成本较高，且在高次元下不稳定，我们将采用合成尤度 (composite likelihood) 或分布变换 (distributional transform)以实现稳健估计。评价方案则以联合正态分布 Probit 为基线, 对比 AIC/BIC 与残差总量/最大残差的缩小幅度，并在设定样本量的情形下使用 bootstrap 给出关键参数的区间估计；特别关注的检验点包括：包含 N 的类型 (如 ENFP/INFP/INTJ 等)的差异是否显著缩小，以及 N-J 负相关在更灵活依存结构下的稳健性与可迁移性。

基金支持：作者没有提到

利益冲突：看起来似乎没有

数据公开：N/A