1 Дисперсионный анализ на блок-схемах

1.1 Блок-схемы

В комбинаторике размещение v элементов по b блокам размера k, что каждый элемент встречается r раз, а каждая пара λ раз, называется *неполной сбалансированной блок-схемой* $D(v, b, r, k, \lambda)$ или коротко *дизайном.* Если $v = b$, $r = k$, то дизайн называется *симметричным* и обозначается $D(v, k, \lambda)$. Например, симметричный дизайн $D(7, 3, 1)$ образуют семь ненулевых элементов поля F_8, которые для удобства закодируем числами от $i = 1, 2, \ldots, 7$, $i = a_1X_1 + a_2X_2 + a_3X_3$, $a_j, X_j \in \{0, 1\}$, организованные в семь блоков по три элемента так, чтобы сумма этих трех элементов над полем F_8 была равна нулю:

$$123,\ 145,\ 167,\ 246,\ 257,\ 347,\ 356$$

Для $D(v, b, r, k, \lambda)$ справедливы два соотношения баланса:

$$vr = bk, \qquad r(k-1) = \lambda(v-1)\,.$$

Первое указывает на то, что общее количество элементов в размещении может быть вычислено двумя способами: либо число всевозможных элементов v умножается на r (сколько раз каждый элемент встречается), либо количество блоков b умножается на число элементов в каждом блоке k. Для доказательства второго выделим какой-то элемент, например, a, который встречается r раз. Число пар с участием a равно $r(k-1)$. С другой стороны, число пар вида $(a*)$, которая встречается λ раз, равно $\lambda(v-1)$, так как для $*$ имеет место $v-1$ вариант.

Если пронумеровать b блоков дизайна и собрать из этих номеров блок-схему таким образом, чтобы в один блок входили номера блоков, содержащие один из v элементов, то мы получим "двойственный"дизайн $D^*(v, b, r, k, \lambda)$, состоящий из v блоков, содержащих r элементов. Что касается параметра λ, то в двойственном дизайне он означает не количество пар, а то, сколько общих элементов у каждой пары блоков.

В качестве примера рассмотрим дизайн $D(4, 6, 3, 2, 1)$, состоящий из 6 блоков b_i, образованнными 4-мя элементами, и двойственный дизайн $D^*(4, 6, 3, 2, 1)$, состоящий из 4-х блоков, образованными 6-ю элементами.

$D(4, 6, 3, 2, 1)$

i	b_i
1	12
2	13
3	14
4	23
5	24
6	34

$D^*(4, 6, 3, 2, 1)$

j	γ_j
1	123
2	145
3	246
4	356

1.2 Модель дисперсионного анализа на блок-схемах

Рассмотрим пример исследования двух факторов: метод обработки и состояние объекта на характеристику эффективности, например, оценка параметра масштаба в модели гамма распределения для высоты секреторного эпителия молочной железы у кормящих мышей. В качестве фактора обработки можно взять а) инъекция тиреолиберина в дозировке 10^{-6}, б) инъекция тиреолиберина в дозировке 10^{-4}, в) инъекцияфизраствора, г) контроль; в качестве фактора состояния среднюю массу дитенышей в граммах: 1 - ≥ 2.72, 2 - $(2.22, 2.72]$, 3 - $(2.1, 2.22]$, 4 - $(1.9, 2.1]$, 5 - $(1.8, 1.9]$, 6 - $(1.2, 1.8]$. Для того чтобы проверить значимость обоих факторов при помощи двухфакторного дисперсионного анализа, необходимо применить 4 вида обработки к 6 видам мышей, но это не всегда возможно. Выход заключается в том, чтобы применить методы обработки только к определенным видам мышей. Например, для мышей первого вида обработки 1 и 2, для второго вида 1 и 3, для третьего 1 и 4, и так далее в соответствии с блоками дизайна $D(4, 6, 3, 2, 1)$.

$v_i \mid b_j$	1	2	3	4	5	6
1	1.22	1.26	0.99	–	–	–
2	1.22	–	–	1.92	1.92	–
3	–	1.45	–	1.31	–	2.02
4	–	–	0.9	–	1.31	0.96

Модель дисперсионного анализа имеет вид

$$x_{ij} = \mu + v_i + b_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma)\,.$$

Основные статистики – это суммы по методам обработки

$$V_i = \sum_{j \in \gamma_i} x_{ij}$$

и по блокам состояния

$$B_j = \sum_{i \in \beta_j} x_{ij}$$

1.2.1 Оценка параметров модели

Оценим параметры модели по методу минимума суммы квадратов отклонений.

$$\mathcal{L} = \sum_{i,j} (x_{ij} - \mu - v_i - b_j)^2 \ \rightarrow \ min$$

при ограничениях $\sum\limits_{i=1}^{v} v_i = 0, \quad \sum\limits_{j=1}^{b} b_j = 0$. Суммирование (ij) означает или $\sum\limits_{i=1}^{v} \sum\limits_{j \in \gamma_i}$, или $\sum\limits_{j=1}^{b} \sum\limits_{i \in \beta_j}$.

$$\frac{\partial \mathcal{L}}{\partial \mu} = -2\sum_{i,j}(x_{ij} - \mu - v_i - b_j) = 0 \;\Leftrightarrow\; \hat{\mu} = \frac{1}{bk}\sum_{i,j} x_{ij}$$

Дифференцируем по параметрам v_i.

$$\frac{\partial \mathcal{L}}{\partial v_i} = -2\sum_{j\in\gamma_i}(x_{ij} - \mu - v_i - b_j) = 0 \;\Leftrightarrow\; \sum_{j\in\gamma_i} b_j = V_i - r\mu - rv_i$$

Дифференцируем по параметрам b_j.

$$\frac{\partial \mathcal{L}}{\partial b_j} = -2\sum_{i\in\beta_j}(x_{ij} - \mu - v_i - b_j) = 0 \;\Leftrightarrow\; B_j = \sum_{i\in\beta_j} v_i + k\mu + kb_j$$

Отсюда

$$T_i = \sum_{j\in\gamma_i} B_j = \sum_{j\in\gamma_i}\left(\sum_{i\in\beta_j} v_i + k\mu + kb_j\right) =$$
$$= \sum_{j\in\gamma_i}\sum_{i\in\beta_j} v_i + rk\mu + k\left(V_i - r\mu - rv_i\right) = \sum_{j\in\gamma_i}\sum_{i\in\beta_j} v_i + kV_i - rkv_i$$

Лемма 1. $\sum\limits_{j\in\gamma_i}\sum\limits_{i\in\beta_j} v_i = (r-\lambda)v_i$

Доказательство. Рассматриваем все блоки, содержащие, например, первый элемент (123), (145),(167), – в каждом из них есть 1, который встречается r раз, и пары вида $(1*)$, количество которых равно $\lambda(k-1)$:

$$\sum_{j\in\gamma_i}\sum_{i\in\beta_j} v_i = rv_i + \lambda\sum_{m\neq i} v_m = rv_i + \lambda(-v_i) = (r-\lambda)v_i\,.$$

Таким образом используя $r(k-1) = \lambda(v-1)$, получаем

$$v_i(r - \lambda - rk) = v_i(-\lambda v) = T_i - kV_i, \Longrightarrow \hat{v}_i = \frac{kV_i - T_i}{\lambda v}$$

Аналогично

$$\hat{b_j} = \frac{1}{k}(B_j - \sum_{i\in b_j}\hat{v}_i) - k\hat{\mu}$$

1.2.2 Разложение суммы квадратов

Обозначим через

$$M_{ij} = \hat{\mu} + \hat{b_j} + \hat{v}_i = \frac{B_j}{k} + \left(\hat{v}_i - \frac{1}{k} \sum_{l \in \beta_j} \hat{v}_l \right) .$$

$$\sum_{ij} \frac{B_j}{k} \left(\hat{v}_i - \frac{1}{k} \sum_{l \in \beta_j} \hat{v}_l \right) = 0, \text{так как}$$

$$\sum_{ij} B_j \sum_{l \in \beta_j} \hat{v}_l = \sum_{j=1}^{b} \sum_{i \in \beta_j} B_j \sum_{l \in \beta_j} \hat{v}_l = \sum_{j=1}^{b} k B_j \sum_{l \in \beta_j} \hat{v}_l = \sum_{l=1}^{v} \sum_{j \in \gamma_l} k B_j \hat{v}_l = k \sum_{l=1}^{v} \hat{v}_l \sum_{j \in \gamma_l} B_j = 0 .$$

Рассмотрим

$$\sum_{ij} x_{ij}^2 = \sum_{ij} (x_{ij} - M_{ij} + M_{ij})^2 = \sum_{ij} (x_{ij} - M_{ij})^2 + \sum_{ij} \left(\hat{v}_i - \frac{1}{k} \sum_{l \in \beta_j} \hat{v}_l \right)^2 + \sum_{ij} \frac{B_j^2}{k^2}$$

Сначала покажем, что

$$\sum_{ij} \frac{B_j^2}{k^2} = \sum_{j=1}^{b} k \left(\frac{B_j}{k} - \hat{\mu} \right)^2 + bk\hat{\mu}^2,$$

Действительно,

$$\sum_{ij} \left(\frac{B_j}{k} - \hat{\mu} + \hat{\mu} \right)^2 = \sum_{ij} \left(\frac{B_j}{k} - \hat{\mu} \right)^2 + bk\hat{\mu}^2 = \sum_{j=1}^{b} k \left(\frac{B_j}{k} - \hat{\mu} \right)^2 + bk\hat{\mu}^2,$$

суммы перекрестных произведений равны 0 из-за

$$\sum_{j=1}^{b} k \left(\frac{B_j}{k} - \hat{\mu} \right) \hat{\mu} = \hat{\mu} \sum_{j=1}^{b} (B_j - k\hat{\mu}) = 0 .$$

Лемма 2. $\sum\limits_{ij} \left(\hat{v}_i - \frac{1}{k} \sum\limits_{i \in \beta_j} \hat{v}_i \right)^2 = \frac{\lambda v}{k} \sum\limits_{i=1}^{v} \hat{v}_i^2$

Доказательство

$$\sum_{ij} \left(\hat{v}_i - \frac{1}{k} \sum_{l \in \beta_j} \hat{v}_l \right)^2 = \sum_{ij} \hat{v}_i^2 - \frac{2}{k} \sum_{ij} \hat{v}_i \sum_{l \in \beta_j} \hat{v}_l + \frac{1}{k^2} \sum_{ij} \left(\sum_{l \in \beta_j} \hat{v}_l \right)^2 =$$

$$= \sum_{i=1}^{v} \underbrace{\sum_{j \in \gamma_i}} \hat{v}_i^2 - \frac{2}{k} \sum_{j=1}^{b} \sum_{i \in \beta_j} \hat{v}_i \sum_{l \in \beta_j} \hat{v}_l + \frac{1}{k^2} \sum_{j=1}^{b} \underbrace{\sum_{i \in \beta_j}} \left(\sum_{l \in \beta_j} \hat{v}_l \right)^2 =$$

$$\sum_{i=1}^{v} r\hat{v}_i^2 - \frac{1}{k}\sum_{j=1}^{b}\left(\sum_{l\in\beta_j}\hat{v}_l\right)^2 \tag{1}$$

Отдельно рассмотрим

$$\sum_{j=1}^{b}\left(\sum_{l\in\beta_j}\hat{v}_l\right)^2 = \sum_{j=1}^{b}\left(\sum_{l\in\beta_j}\hat{v}_l^2 + 2\sum_{l,f\in\beta_j,\ l\neq f}\hat{v}_l\hat{v}_f\right) =$$

$$= \sum_{i=1}^{v}\underbrace{\sum_{j\in\gamma_i}}\hat{v}_l^2 + 2\sum_{l,f\in\beta_j,\ l\neq f}\lambda\hat{v}_l\hat{v}_f$$

В последнем равенстве пересчитываются всевозможные пары. Поскольку $\sum\limits_{i=1}^{v}\hat{v}_i = 0$, то

$$0 = \left(\sum_{i=1}^{v}\hat{v}_i\right)^2 = \sum_{i=1}^{v}\hat{v}_i^2 + 2\sum_{l,f\in\beta_j,\ l\neq f}\hat{v}_l\hat{v}_f,$$

$$\sum_{i=1}^{v}\underbrace{\sum_{j\in\gamma_i}}\hat{v}_l^2 + 2\sum_{l,f\in\beta_j,\ l\neq f}\lambda\hat{v}_l\hat{v}_f = \sum_{i=1}^{v} r\hat{v}_l^2 - \lambda\sum_{i=1}^{v}\hat{v}_i^2$$

Продолжаем выражение (1).

$$\sum_{i=1}^{v} r\hat{v}_i^2 - \frac{1}{k}\sum_{j=1}^{b}\left(\sum_{l\in\beta_j}\hat{v}_l\right)^2 = \sum_{i=1}^{v} r\hat{v}_i^2 - \frac{1}{k}\left(\sum_{i=1}^{v} r\hat{v}_l^2 - \lambda\sum_{i=1}^{v}\hat{v}_i^2\right) =$$

$$= \left(\frac{r(k-1)+\lambda}{k}\right)\sum_{i=1}^{v} r\hat{v}_l^2 = \left(\frac{\lambda(v-1)+\lambda}{k}\right)\sum_{i=1}^{v} r\hat{v}_l^2 = \frac{\lambda v}{k}\sum_{i=1}^{v} r\hat{v}_l^2$$

1.2.3 Равенство нулю сумм перекрестных произведений

Первая сумма перекрестных произведений имеет вид

$$1)\sum_{ij}(x_{ij} - M_{ij})\left(\hat{v}_i - \frac{1}{k}\sum_{l\in\beta_j}\hat{v}_l\right) = \sum_{ij}\left(x_{ij} - \frac{B_j}{k} - \left(\hat{v}_i - \frac{1}{k}\sum_{l\in\beta_j}\hat{v}_l\right)\right)\left(\hat{v}_i - \frac{1}{k}\sum_{l\in\beta_j}\hat{v}_l\right);$$

$$a)\sum_{ij}\hat{v}_i\left(x_{ij} - \frac{B_j}{k} - \left(\hat{v}_i - \frac{1}{k}\sum_{l\in\beta_j}\hat{v}_l\right)\right) = \sum_{i=1}^{v}\hat{v}_i\sum_{j\in\gamma_i}\left(x_{ij} - \frac{B_j}{k} - \left(\hat{v}_i - \frac{1}{k}\sum_{l\in\beta_j}\hat{v}_l\right)\right) =$$

$$= \sum_{i=1}^{v}\hat{v}_i\left(V_i - \frac{T_i}{k} - r\hat{v}_i + \frac{1}{k}\sum_{j\in\gamma_i}\sum_{l\in\beta_j}\hat{v}_l\right) =$$

Воспользуемся тем, что

$$\frac{T_i}{k} = \frac{1}{k}\sum_{j\in\gamma_i}\sum_{i\in\beta_j}\hat{v}_i + V_i - r\hat{v}_i, \Longrightarrow$$

$$= \sum_{i=1}^{v}\hat{v}_i\left(V_i - \frac{1}{k}\sum_{j\in\gamma_i}\sum_{i\in\beta_j}\hat{v}_i n - V_i + r\hat{v}_i - r\hat{v}_i + \frac{1}{k}\sum_{j\in\gamma_i}\sum_{l\in\beta_j}\hat{v}_l\right) = 0\,.$$

$$b)\frac{1}{k}\sum_{j=1}^{b}\sum_{l\in\beta_j}\hat{v}_l\sum_{i\in\beta_j}\left(x_{ij} - \frac{B_j}{k} - \hat{v}_i + \frac{1}{k}\sum_{l\in\beta_j}\hat{v}_l\right) =$$

$$= \frac{1}{k}\sum_{j=1}^{b}\sum_{l\in\beta_j}\hat{v}_l\left(B_j - B_j - \sum_{i\in\beta_j}\hat{v}_i + \frac{k}{k}\sum_{l\in\beta_j}\hat{v}_l\right) = 0\,.$$

$$2)\sum_{ij}(x_{ij} - M_{ij})\frac{B_j^2}{k^2} = \sum_{ij}\left(x_{ij} - \frac{B_j}{k} - \hat{v}_i + \frac{1}{k}\sum_{l\in\beta_j}\hat{v}_l\right)\frac{B_j^2}{k^2} =$$

$$= \sum_{j=1}^{b}\frac{B_j^2}{k^2}\sum_{i\in\beta_j}\left(x_{ij} - \frac{B_j}{k} - \hat{v}_i + \frac{1}{k}\sum_{l\in\beta_j}\hat{v}_l\right) =$$

$$= \sum_{j=1}^{b}\frac{B_j^2}{k^2}\left(B_j - \frac{kB_j}{k} - \sum_{l\in\beta_j}\hat{v}_i + \frac{k}{k}\sum_{l\in\beta_j}\hat{v}_l\right) = 0\,.$$

$$3)\sum_{ij}\left(\hat{v}_i - \frac{1}{k}\sum_{l\in\beta_j}\hat{v}_l\right)\frac{B_j^2}{k^2} = \sum_{j=1}^{b}\frac{B_j^2}{k^2}\sum_{i\in\beta_j}\left(\hat{v}_i - \frac{1}{k}\sum_{l\in\beta_j}\hat{v}_l\right) = 0\,.$$

1.2.4 Статистические критерии

Таким образом,

$$\sum_{i,j}(x_{ij}-\hat{\mu})^2 = S_e^2 + S_v^2 + S_b^2,$$

$$S_v^2 = \sum_{ij}\left(\hat{v}_i - \frac{1}{k}\sum_{i\in\beta_j}\hat{v}_i\right)^2 = \frac{\lambda v}{k}\sum_{i=1}^{v}\hat{v}_i^2, \quad df_v = v-1,$$

$$S_b^2 = \sum_{j=1}^{b} k\left(\frac{B_j}{k} - \hat{\mu}^2\right)^2, \quad df_b = b-1;$$

$$S_e^2 = \sum_{i,j}\left(x_{ij} - \hat{v}_i + \frac{1}{k}\sum_{l\in b_j}\hat{v}_i - \frac{B_j}{k}\right)^2,$$

$$df_e = bk-1-(v-1)-(b-1) = bk-v-b+1\,.$$

Содержание