Статистический анализ внутригруппового плана

Лекция №4

1. Статистические основы внутригруппового эксперимента. Однофакторный дисперсионный анализ с повторным измерением.
2. Структурные модели однофакторного дисперсионного анализа с повторным измерением.

Вопросы для обсуждения

ВОПРОС №1

Внутригрупповой эксперимент…

Внутригрупповой план

- В отличие от *межгруппового* плана *внутригрупповой* экспериментальных план предполагает использование всего *одной группы* испытуемых
- *Внутригрупповым* называют экспериментальный план, в котором *каждому испытуемому* предъявляют *все* уровни независимой переменной
- Эксперимент, реализующий такую схему, принято называть *экспериментом с повторным измерением,* т. к. в ходе эксперимента измерение зависимой переменной у одно и того же испытуемого осуществляется более одного раза

Испытуемый	Условие 1	...	Условие j	...	Условие k	Сумма
1		…		…		P_1
...						...
i		…		…		P_i
...						...
n		…		…		P_k
Сумма	T_1	…	T_j	…	T_k	G
Среднее		...		...		

Повторные измерения

Общая дисперсия

Дисперсия внутри испытуемых (within subjects)

Дисперсия между испытуемыми (between subjects)

Дисперсия экспериментального воздействия (treatments)

Остаточная дисперсия (residual)

Анализ дисперсии

Суммарный квадрат *(SS)*	Степени свободы (*df*)
$SS_{between\_subjects} = k\sum\left(P_i - \overline{G}\right)^2$	

Между испытуемыми

Суммарный квадрат *(SS)*	Степени свободы (*df*)
$SS_{within\_subject} = \sum_i \sum_j \left(x_{ij} - \overline{P}\right)^2$	

$$SS_{within\_subject} = SS_{treatment} + SS_{residual}$$

Внутри испытуемых

Суммарный квадрат *(SS)*	Степени свободы (*df*)
$SS_{treatment} = n\sum\left(\overline{T_j} - \overline{G}\right)^2$	

Экспериментальное воздействие

Суммарный квадрат *(SS)*	Степени свободы (*df*)
$SS_{residual} = \sum\sum\left[x_{ij} - \left(\overline{P_i} - \overline{G}\right)\right]^2$	

Остаток

Суммарный квадрат *(SS)*	Степени свободы (*df*)
$SS_{total} = \sum\left(\sum X_{ij} - \overline{G}\right)^2$	

$$SS_{total} = SS_{between\ subjects} + SS_{within\ subjects} = SS_{between\ subjects} + SS_{treatment} + SS_{residual}$$

Всего

Источник дисперсии	Суммарный квадрат (*SS*)	Степени свободы (*df*)
Между испытуемыми	$k\sum\left(\overline{P}-\overline{G}\right)^2$	*(n-1)*
Внутри испытуемых	$\sum\sum\left(X-\overline{P}\right)^2$	*n(k-1)*
Экспериментальное воздействие	$n\sum\left(\overline{T}-\overline{G}\right)^2$	*(k-1)*
Ошибка (остаток)	$\sum\sum\left(X-\left(\overline{P}-\overline{G}\right)\right)^2$	*(n-1)(k-1)*
Общий	$\sum\sum\left(X-\overline{G}\right)^2$	*kn-1*

Оценка дисперсии

$$F[k-1,(k-1)(n-1)] = \frac{MS_{treatment}}{MS_{residual}}$$

F-отношение

ВОПРОС №2

Структурная модель однофакторного дисперсионного анализа с повторным измерением

- Будем предполагать, что результат измерения зависимой переменной может быть представлен через популяционную постоянную μ, эффект независимой переменной τ_j, специфичный и постоянный для каждого ее уровня,
- индивидуальную константу π_i, выражающую эффект отдельного испытуемого (предполагается, что ее популяционное значение равно 0) и неконтролируемую экспериментальную ошибку ε_{ij}
- Т.е.: $x_{ij} = \mu + \pi_i + \tau_j + \varepsilon_{ij} + \pi\tau_{ij}$

Структурная модель

- Экспериментальная ошибка представляет собой случайную величину, распределенную в соответствии с *нормальным* законом с *математическим ожиданием равным нулю*.
- Индивидуальная константа представляет собой также случайную величину, распределенную в популяции рассматриваемых данных в соответствии с *нормальным* законом с *математическим ожиданием равным нулю*.
- Эффект экспериментального воздействия представляет собой случайную величину, распределенную в соответствии с *нормальным* законом с заранее неизвестными параметрами

Допущения

- Будем предполагать, что эффект независимой переменной не взаимодействует с эффектом испытуемого
- Тогда: $x_{ij} = \mu + \pi_i + \tau_j + \varepsilon_{ij}$

Модель I

Испытуемый	Уровень 1 (T_1)	Уровень 2 (T_2)	Среднее
1			
…	…	…	…
i			
…	…	…	…
n			
Среднее			

Двухуровневый план

- Поскольку величины μ, τ_1 и τ_2 постоянны, дисперсия внутри экспериментального условия определяется дисперсией экспериментальной ошибки σ^2_ε и дисперсией индивидуального эффекта σ^2_π. Таким образом, справедливы следующие соотношения:

$$\sigma_1^2 = \sigma_\pi^2 + \sigma_\varepsilon^2$$
$$\sigma_2^2 = \sigma_\pi^2 + \sigma_\varepsilon^2$$

Дисперсия ЗП для каждого уровня НП

- Величины σ^2_ε являются статистически независимыми друг от друга в двух экспериментальных условиях, чего нельзя сказать о величинах σ^2_π. По сути дела величина σ^2_π определяет статистическую связь двух экспериментальных условий — T_1 и T_2. Иными словами,
- $\sigma^2_{12} = \sigma^2_\pi$, где σ^2_{12} — ковариация T_1 и T_2, $\mathrm{cov}(T_1, T_2)$

Тогда…

$$E\left(MS_{residual}\right)=\frac{\sigma_1^2+\sigma_2^2}{2}-\sigma_{12}=\sigma_\varepsilon^2$$

$$E\left(MS_{treatment}\right)=n\sigma_{\bar{T}_1-\bar{T}_2}^2=n\sigma_{\bar{T}}^2=n\sigma_\tau^2+\sigma_\varepsilon^2$$

$$E\left(F\right)=\frac{\sigma_\varepsilon^2+n\sigma_\tau^2}{\sigma_\varepsilon^2}$$

Следовательно…

Нулевая - H_0

- $\tau_1 = \tau_2$
- $\sigma_\tau = 0$

$$E(F) = \frac{\sigma_\varepsilon^2 + n\sigma_\tau^2}{\sigma_\varepsilon^2} = 1$$

Альтернативная – H_1

- $\tau_1 \neq \tau_2$
- $\sigma_\tau > 0$

$$E(F) = \frac{\sigma_\varepsilon^2 + n\sigma_\tau^2}{\sigma_\varepsilon^2} > 1$$

Модель I: гипотезы

- Будем предполагать, что эффект независимой переменной взаимодействует с эффектом
- испытуемого
- Т.е. вернемся к начальному предположению, что $x_{ij} = \mu + \pi_i + \tau_j + \varepsilon_{ij} + \pi\tau_{ij}$

Модель II

$$E(MS_{treatment}) = \sigma_{\varepsilon}^2 + n\sigma_{\pi\tau}^2 + n\sigma_{\tau}^2$$

$$E(MS_{residual}) = \sigma_{\varepsilon}^2 + n\sigma_{\pi\tau}^2$$

Тогда...

Нулевая - H_0

- $\tau_1 = \tau_2$
- $\sigma_\tau = 0$

$$E(F) = \frac{\sigma_\varepsilon^2 + n\sigma_{\tau\pi}^2 + n\sigma_\tau^2}{\sigma_\varepsilon^2 + n\sigma_{\pi\tau}^2} = \frac{\sigma_\varepsilon^2 + n\sigma_{\tau\pi}^2}{\sigma_\varepsilon^2 + n\sigma_{\tau\pi}^2} = 1$$

Альтернативная – H_1

- $\tau_1 \neq \tau_2$
- $\sigma_\tau > 0$

$$E(F) = \frac{\sigma_\varepsilon^2 + n\sigma_{\tau\pi}^2 + n\sigma_\tau^2}{\sigma_\varepsilon^2 + n\sigma_{\pi\tau}^2} > 1$$

Модель II: гипотезы

$$\sigma_{x_j}^2 = \frac{\sum (x_{ij} - \mu_j)^2}{n-1}$$

где μ_j – математическое ожидание значения зависимой переменной на уровне *j*

Тогда ковариация значений зависимой переменной на уровнях *j* и *j'* независимой переменной может быть найдена по формуле:

$$\sigma_{x_j x_{j'}} = \frac{\sum (x_{ij} - \mu_j)(x_{ij'} - \mu_{j'})}{n-1} = \rho_{x_j x_{j'}} \sigma_{x_j} \sigma_{x_{j'}}$$

Где ρ – корреляция значений зависимой переменной на уровнях *j* и *j'*

Многоуровневый план

Однородность матрицы ковариаций

Поскольку, согласно предположению модели, эффект испытуемого не взаимодействует с эффектами независимой переменной, матрица ковариаций должна быть однородной, т.е.

$$\sigma_{\pi}^{2} = \sigma_{x_j x_{j'}}$$

$$\begin{vmatrix}
\sigma_{\pi_1}^{2} & \cdots & \sigma_{x_1 x_j} & \cdots & \sigma_{x_1 x_k} \\
 & \ddots & \vdots & & \vdots \\
 & & \sigma_{\pi_j}^{2} & \cdots & \sigma_{x_j x_k} \\
 & & & \ddots & \vdots \\
 & & & & \sigma_{\pi_k}^{2}
\end{vmatrix}$$

$$\sigma_{\varepsilon}^{2} = \sigma_{x}^{2} - \sigma_{\pi}^{2} = \sigma_{x}^{2} - \sigma_{jj'}$$

$$s_x^2 = \frac{\sum\left(x - \overline{T}\right)^2}{n-1}$$

Оценка дисперсии для одного экспериментального условия

$$E\left(\overline{s}_x^{2}\right) = E\left(\frac{s_1^2 + \ldots + s_k^2}{k}\right) = \sigma_x^2$$

$$E(\overline{s}_x^{2} - \overline{\mathrm{cov}}) = \sigma_{\varepsilon}^{2}$$

Тогда...

$$E\left(MS_{residual}\right) = \sigma_{\varepsilon}^{2}$$

$$E\left(MS_{treatment}\right) = E\left(\frac{n\sum\left(\overline{T}_{j} - \overline{T}_{j'}\right)^{2}}{k-1}\right) = \sigma_{\varepsilon}^{2} + n\sigma_{\tau}^{2}$$

Наконец…

Нулевая - H_0

- $$SS_{total} = SS_{between\ subjects} + SS_{within\ subjects} =$$
  $$SS_{between\ subjects} + SS_{treatment} + SS_{residual}$$

Альтернативная – H_1

- $\tau_1 \neq \tau_2 \neq \ldots \neq \tau_j \neq .. \neq \tau_k$
- $\sigma_\tau > 0$
- $E(F) > 1$

Гипотезы

- Для оценки однородности ковариационной матрицы используют тест сферичности Моучли *(Mauchly)*.
- Если этот тест свидетельствует о значительной гетерогенности ковариационной матрицы, рекомендуется при статистической надежности анализе F-отношения, вычисленного по результатам эксперимента, уменьшить число степеней свободы
- Это обеспечивает большую степень консервативности при принятии решения о статистически надежных эффектах независимой переменной.

Оценка однородности ковариаций

- Уменьшить число степеней свободы можно, исходя из следующего правила:
- $df = (k-1)\theta$ для числителя
- $df = (k-1)(n-1)\theta$ для знаменателя
- Где θ принимается равной 1 в случае *полной гомогенности* ковариационной матрицы и $^{1}/_{(k-1)}$ - в случае *наименьшей гомогенности*.

Уменьшение *df*

www.ebbinghaus.ru