Практична робота № 6

Тема: Багатофакторний дисперсійний аналіз

Мета роботи: Зробити багатофакторний дисперсійний аналіз, використовуючи пакет SPSS або Excel

Стислі теоретичні відомості

Припустимо: ми обстежуємо k факторів, які можуть впливати на результат спостережень. Необхідно з'ясувати: який з факторів впливає на результат, який не впливає. Для відповіді на це питання проводять $k+1$ груп експериментів (по n експериментів у кожній групі). У кожної з груп визначають значення результатів: y_{ij}; $i = 1 \ldots n$ у групах при наявності j-фактора, та y_{i0}; $i = 1 \ldots n$ у групі при відсутності аналізованих факторів (контрольна група). Потім для кожного фактора визначають парні різниці між відповідними спостереженнями з цим фактором та спостереженнями контрольної групи $dif_{ij} = y_{ij} - y_{ij}$.

Результати зводять у таблицю (табл. 4).

Таблиця 4

Спостереження	Фактор x_1	Фактор x_2	……	Контрольна група (без фактора)	Різниця №1	Різниця №2
1	y_{11}	y_{12}	…	y_{10}	$dif_{11} = y_{11} - y_{10}$	$dif_{12} = y_{11} - y_{10}$
2	y_{21}	y_{22}	…	y_{20}	$dif_{21} = y_{21} - y_{20}$	$dif_{22} = y_{22} - y_{20}$
3	y_{31}	y_{32}	…	y_{30}	$dif_{31} = y_{31} - y_{30}$	$dif_{32} = y_{32} - y_{30}$

Середня j-та різниця для кожного фактора: $\widetilde{dif}_{\text{ср}j} = \dfrac{\sum_{i=1}^{n} dif_{ij}}{n}$.

Точкова оцінка математичного сподівання середньої j-ї різниці є середня j-та різниця: $\widetilde{dif}_{\text{ср}j} = \dfrac{\sum_{i=1}^{n} dif_{ij}}{n}$.

Точкова оцінка середньоквадратичного відхилення середньої j-ї різниці:

$$\sigma^*_{\text{ср}j} = \frac{1}{\sqrt{n}} * \sqrt{\frac{\sum_{i=1}^{n}(dif_{IJ} - dif_{\text{ср}j})^2}{n-1}}.$$

Або: $\sigma^*_{\text{ср}j} = \frac{1}{\sqrt{n}} * \sigma^*_J$, де σ^*_J – оцінка середньоквадратичного відхилення j-ї різниці (не середньої).

Для того, щоб визначити вплив j-го фактора, перевіряємо на значимість середню різницю між спостереженнями з цим фактором та спостереженнями контрольної групи. Якщо різниця значима, то фактор впливає, якщо різниця не значима, то фактор не впливає, а якесь середнє значення різниці отрималось внаслідок розбіжності та недостатку даних.

Для цього робимо такі дії.

Обираємо ступінь довіри нашим висновкам (1–α) = 0,95 – 0,99 та висуваємо нульову гіпотезу, що фактор не впливає на результат і математичне сподівання середньої j-ї різниці $M(dif_{\text{ср}j})=0$ (тобто вона не значима).

Знаходимо статистику Ст'юдента j-ї $t_j = \frac{dif_{\text{ср}j} - M(dif_{\text{ср}j})}{\sigma^*_{\text{ср}j}}$ різниці, яка при виконанні нульової гіпотези має вигляд $t_j = \frac{dif_{\text{ср}j}}{\sigma^*_{\text{ср}j}} = \frac{dif_{\text{ср}j}}{\sigma^*_j} * \sqrt{n}$.

Випадкова величина – t_j має розподіл ймовірності Ст'юдента з (n-1) ступенями вільності. Вона визначає значення середньої j-ї різниці відносно її відхилення.

Знаходимо квантиль $\frac{\alpha}{2}$ Ст'юдента $t_{\frac{\alpha}{2}}$ – тобто таке значення випадкової величини, яке відсікає під кривою щільності розподілу ймовірностей Ст'юдента площу $\frac{(1-\alpha)}{2}$ з обох боків та є розв'язком рівняння:

$$P(-t_{\frac{\alpha}{2}} < t_j < t_{\frac{\alpha}{2}}) = (1-\alpha)$$

де $P(-t_{\frac{\alpha}{2}} < t_j < t_{\frac{\alpha}{2}})$ – ймовірність влучення випадкової величини t_j в інтервал між квантилями $t_{\frac{\alpha}{2}}$.

Звідси ми отримуємо критерії для визначення невпливовості фактора з імовірністю 1-α:

– якщо $\left|t_j\right| < t_{\frac{\alpha}{2}}$, тобто влучається в інтервал між квантилями, то імовірність нульової гіпотези велика, різниця незначима і фактор не впливає;
– якщо $\left|t_j\right| < t_{\frac{\alpha}{2}}$, тобто влучається в інтервали за квантилі (у хвости), то імовірність нульової гіпотези мала (менше $\frac{\alpha}{2}$), різниця значима і фактор впливає (рис. 22).

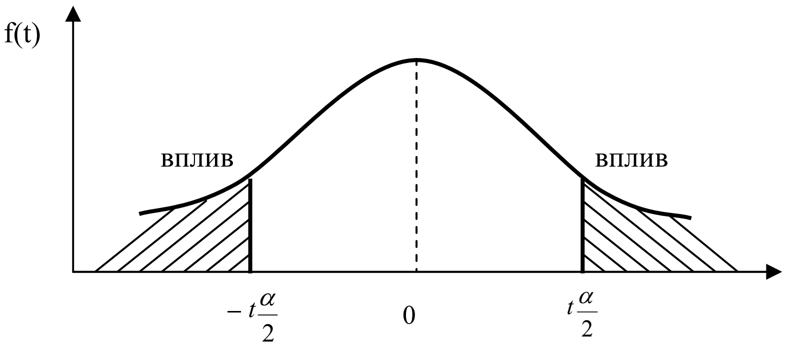

Рис. 22. Впливова зона при *t*-тесті

Отже, для перевірки впливовості фактора X_j виконуємо такі дії:

– знаходимо різниці між результатами з фактором та контрольної групи $dif_{ij} = y_{ij} - y_{i0}$;
– знаходимо середні *j*-ті різниці $dif_{\text{ср}j}$, та точкові оцінки середньоквадратичного відхилення *j*-ї різниці σ^*_j;
– обираємо ступінь довіри $(1 - \alpha)$;
– висуваємо нульову гіпотезу, що фактор не впливає на результат і $M(dif_{\text{ср}j}) = 0$;
– знаходимо статистику Ст'юдента $t_j = \frac{dif_{\text{ср}j}}{\sigma^*_j} * \sqrt{n}$;
– якщо $\left|\frac{dif_{\text{ср}j} * \sqrt{n}}{\sigma^*_j}\right| < t_{\frac{\alpha}{2}}$, то ймовірність нульової гіпотези більше, ніж $\frac{\alpha}{2}$, тобто велика, і фактор не впливає на результат;

– якщо $\left|\frac{dif_{\text{ср}j} * \sqrt{n}}{\sigma_j^*}\right| > t_{\frac{\alpha}{2}}$, то ймовірність нульової гіпотези менше $\frac{\alpha}{2}$ і фактор впливає на результат.

Використання пакета SPSS для багатофакторного дисперсійного аналізу

Головне меню пакета:

Statistics → Compare Means → Pairs Samples T Test.

Обрати та занести у вікно Paired difference назву колонок, які містять дані з фактором, що аналізується, та дані контрольної групи.

У протоколі для кожного аналізованого фактора містяться точкові оцінки математичного сподівання (mean), середньоквадратичного відхилення (SE) для даних з аналізованим фактором та даних контрольної групи, а також для їхніх парних різниць (paired differences), статистика Ст'юдента парної різниці t_{VALUE}, ймовірність нульової гіпотези (2–tail Sig).

Використання пакета Excel для багатофакторного дисперсійного аналізу

У пакеті EXCEL виконання факторного аналізу у головному меню обрати

Tools → Data Analisis → t-test: Pared Two Samples for Means.

Далі у віконці заповнюються комірки даних з фактором (група 1) контрольної групи (група 2). Обирається інтервал помилки (α = 0,05), очікуване математичне сподівання різниці при виконанні нульової гіпотези (Hypot. Means Difference) = 0, місце, де будуть знаходитися вихідні результати.

У протоколі для кожного фактора, який аналізується, містяться точкові оцінки математичного сподівання (mean), середньоквадратичного відхилення (SE) для даних з фактором та даних контрольної групи, а також для їхніх парних різниць (paired differences), статистика Ст'юдента парної різниці t_{VALUE}, квантиль Ст'юдента $\frac{\alpha}{2}$ t_{VALUE}, ймовірність нульової гіпотези (2–tail Sig).

Контрольні запитання

1. В чому полягає основна ідея багатофакторного дисперсійного аналізу?
2. На чому заснована перевірка значимості в дисперсійному аналізі?
3. Яка послідовність дій при дисперсійному аналізі?
4. Як формулюється нульова гіпотеза?
5. Як визначається статистика Ст'юдента?
6. Як обирається ступінь довіри?
7. Як визначається квантиль $t_{\alpha/2}$?
8. Як використати пакет SPSS для багатофакторного дисперсійного аналізу?