

УДК 618.142

А. Г. Шумілін, студент гр. ПВ-91МП, д.т.н., проф. Яремчук Н. А.
КПІ ім. Ігоря Сікорського

СПОСОБИ ВИЗНАЧЕННЯ ЦЕНТРАЛЬНОЇ ТЕНДЕНЦІЇ ВИБІРОК ВЕРБАЛЬНИХ ДАНИХ

Анотація. При опрацюванні вербальних даних, отриманих при візуальному контролі якості продукції і при експертному оцінюванні, виникає необхідність в визначенні центральної тенденції вербальних вибірок. В роботі проведено порівняльний аналіз оцінювання центральної тенденції вербальних вибірок за медіаною вибірки і за оператором OWA, і отримано рекомендації щодо вибору об'єму вибірок і кількості вербальних градацій шкали оцінювання.

Ключові слова: вербальна вибірка, медіана, оператор OWA

ВСТУП

При контролі якості продукції, стану об'єктів або характеристик технологічних процесів використовується така процедура експериментальної інформатики як спостереження. За [1] спостереження – це відображення властивості об'єкта або його стану словами або графічним описом. Тобто результатами спостереження є вербальні дані. Крім того, при вирішенні багатьох практичних проблем використовуються лінгвістичні шкали, шкали якості продукції або сервісів. Оскільки градації таких шкал, як правило, впорядковані, використовуються шкали квазіпорядку, за яких встановлено порядок на класах еквівалентності, тобто на терм-множині лінгвістичної змінної. Прикладами впорядкованих назв класів еквівалентності може бути для стану об'єкту: «незадовільний», «задовільний», «добрий». В [2] наведено результати візуального контролю процесу закорковування пляшок з алкогольними напоями за наступними класами еквівалентності: «зовсім відкинуто», «погана якість», «середня якість», «добра якість», «дуже добра якість». В [3] для побудови контрольної карти якості обробки поверхні використовувались 10 вибірок, кожна з яких складалась з чотирьох результатів спостережень, поданими за класами якості: «низька – L», «середня – M», «висока – H». Тобто приклад однієї вибірки мав вигляд: «HML». І тому методи опрацювання вербальних даних відрізняються від методів опрацювання числових даних. Проблема online моніторингу ординальних якісних характеристик потребує розвитку спеціальних методів опрацювання ординальних даних. Одним із важливих параметрів вибірки ординальних даних є центральна тенденція, для визначення якої можуть бути використана медіана вибірки та оператор OWA, що є емулятором середнього арифметичного вербальної вибірки. Дана робота присвячена порівнянню цих операторів з метою надання рекомендацій для їх практичного застосування

ОГЛЯД ПОПЕРЕДНІХ РОБІТ

При опрацюванні вербальних даних використовуються два підходи. Перший заснований на арифметизації вербальної шкали, що є впорядкованою послідовністю класів еквівалентності. Арифметизація шкали заснована на присвоєнні окремим класам еквівалентності числових рангів, значення яких розташовуються згідно з порядком класів еквівалентності. Вербальні вибірки

оцінюють за сумою рангів окремих елементів вибірки. Це дозволяє ранжувати вибірки, наприклад, за якістю технологічного процесу, тому що вищий ранг вибірки свідчить про вищу якість. Але багато вчених, що працюють в галузі опрацювання вербальних даних, вважають, що примусове введення рангів призводить до штучного введення відстані між класами еквівалентності. І паралельно до рангових способів опрацювання вербальних даних вводять процедури і оператори, що дозволяють ранжувати вибірки вербальних даних без арифметизації. Перший спосіб ранжування вербальних даних заснований на критерії домінування, що дозволяє будувати контрольні карти технологічних процесів. Але при використанні тільки критерію домінування частина вербальних вибірок залишається нерозрізненою. Тому доводиться використовувати кілька критеріїв одночасно [4]. Класифікувати вербальні вибірки за центральною тенденцією дозволяють такі оператори як емулятор середнього значення вербальної вибірки OWA [5] і медіана вибірки. Оператор OWA визначається як:

$$OWA = \underset{k=1}{\overset{n}{\text{Max}}}[\text{Min}\{Q(k), b_k\}] , \quad (1)$$

де $Q(k) = S_{f(k)}$, $K = 1, 2, \dots, n$;

$f(k) = \text{Int}\{1 + [k(t-1)/n]\}$;

Int – найближче ціле число;

t – кількість рівнів шкали;

n – об'єм вибірки;

B_k – k-тий елемент вибірки, ранжованої за зменшенням порядку.

Медіану ранжованої вибірки доцільно використовувати при непарному значенні n. При парному n результат отримання центральної тенденції може бути неоднозначним.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ.

В роботі [2] проведено аналіз способів побудови контрольних карт із застосуванням таких критеріїв і операторів опрацювання вербальних даних як критеріїв домінування і оператор OWA. Це дозволило ранжувати вербальні вибірки з метою їх розміщення на контрольній карті. Так як при побудові контрольних карт застосовуються послідовність вибірок малого об'єму, автори [2] вибрали для аналізу вибірки об'ємом 2,3,4 з кількістю рівнів шкали 3,5,7,9. Оскільки варіантів комбінування в такій ситуації було багато, то кількість класів еквівалентності становила, наприклад, (при n=4, t=9) 495.

Зважаючи на обмежену розрізняльну здатність при візуальному контролі або експертному оцінюванні (максимальне значення розрізняваних градацій або рівнів за [6] дорівнює 7) автори даної роботи обмежились кількістю розрізняваних рівнів 3,5,7. Крім того, об'єм вибірки було збільшено для отримання стійкого значення центральної тенденції. Задачею дослідження була класифікація вербальних вибірок за центральною тенденцією. Для визначення центральної тенденції використовувались два способи: за медіаною вибірки і за

оператором OWA. Щоб отримати однозначну оцінку за медіаною вибірки, кількість елементів вибірки була непарною.

Дослідження проводилось за наступними етапами: визначення вхідних даних, а саме об'єму вибірки n , кількості t і позначень рівнів шкали або класів еквівалентності; визначення медіани за центральним членом ранжованої вибірки; визначення вагової функції оператора OWA; визначення центральної тенденції вибірки за оператором OWA (формула 1); порівняння отриманих результатів.

Якщо обрати об'єм вибірки $n=3$, $t=3$ у відповідності з найменуванням класів еквівалентності $S_1=L$; $S_2=M$; $S_3=H$; тоді отримаємо вагову функцію оператора OWA за $f(k)$, $k=1,2,3$, тобто $f(1)=2$, $f(2)=2$, $f(3)=3$; $Q(k=1)=S_2$; $Q(k=2)=S_2$; $Q(k=3)=S_3$. Для першої і другої вибірки табл. 1 отримуємо:

$$OWA = \underset{k=1}{\overset{3}{\text{Max}}}[\text{Min}\{S_2, L\}; \text{Min}\{S_2, L\}; \text{Min}\{S_3, L\};] = L \quad (2)$$

$$OWA = \underset{k=1}{\overset{3}{\text{Max}}}[\text{Min}\{S_2, M\}; \text{Min}\{S_2, L\}; \text{Min}\{S_3, L\};] = M$$

Таблиця 1.

Результати визначення центральної тенденції для вербальних вибірок.

| Об'єм вибірки n | Кількість класів еквівалентності t | Досліджувані вибірки | Центральна тенденція | |
|-------------------|--------------------------------------|----------------------|----------------------|-------------------|
| | | | за медіаною вибірки | за оператором OWA |
| 3 | 3 | MLL | L | M |
| | | ... | ... | ... |
| | | HHM | H | M |
| 3 | 5 | HLL | L | LM |
| | | HHM | H | MH |
| 5 | 3 | HLLLL | L | L |
| | | HHHHM | H | M |
| 5 | 5 | HLLLL | L | LM |
| | | ... | ... | ... |
| | | HHHHM | H | MH |
| 7 | 3 | HLLLLLL | L | L |
| | | HHHHHMM | H | M |
| | | HHHHHHM | H | H |
| 7 | 5 | HLLLLLL | L | L |
| | | HHLLLLL | H | MH |
| | | HHHHHHM | H | H |

Підвищуємо розрізняльну здатність операторів, збільшуючи кількість класів еквівалентності до $t=5$ за рахунок двох проміжних класів. Тоді $S_1=L$; $S_2=LM$; $S_3=M$; $S_4=MH$; $S_5=H$; і при $n=3$ та $t=5$ збіжність результатів за медіаною і оператором OWA збільшується. Потім збільшуємо об'єми вибірок. Результати обчислень наведено в табл. 1.

ВИСНОВКИ.

На основі отриманих в таблиці 1 результатів дослідження можна зробити наступні висновки:

При об'ємах вибірок $n=3$ і кількості градацій вербальної шкали оцінювання $t=3$ на результати визначення центральної тенденції за оператором OWA суттєво впливає ступінчастість вагової функції, і це приводить до розбіжностей між результатами, отриманими за медіаною вибірки і оператором OWA, причому перевагу треба віддати медіані вибірки.

Розрізняльна здатність оператора OWA збільшується при збільшенні об'єму вибірки і кількості вербальних градацій шкали оцінювання, причому більша збіжність результатів, отриманих двома способами, спостерігається при збільшенні вербальних градацій шкали оцінювання при одному і тому об'ємі вибірки.

Найбільший ефект відповідності оцінок центральної тенденції за медіаною і оператора OWA спостерігається при пропорційному збільшенні і об'єму вибірки і кількості градацій вербальної шкали оцінювання, причому перевагу у цьому випадку має оцінювання центральної тенденції за оператором OWA.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Орнатський П.П. Вступ до методології науки про вимірювання. Навчальний посібник. К.:ІСЛО, 1994.-160с.
- [2] Qualitative Ordinal Scales: The concept of Ordinal Range. F.Franceschini, M.Galetto, M.Vazetto // Quality engineering. Vol.16 №4 pp. 515-524, 2004.
- [3] Ordered Samples Control Charts for Ordinal Variables. F.Franceschini, M.Galetto, M.Vazetto // Quality and Reliability Engineering International, 2005, 21:177-195.
- [4] Опрацювання вербальних даних за побудови контрольних карт. Н. Яремчук, Р. Семенюк // Метрологія та прилади №5-1(67), 2017 с.58-61.
- [5] Yager R., Filev DP. Essential of Fuzzy Modeling and Control. Wiley: New York, 1994.
- [6] Гнатієнко Г.М., Снитюк В.Е. Експертні технології прийняття рішень: Монографія. – К.ТОВ «Маклаут», - 2008.- 444с.

Наук. керівник – к.т.н., проф. Яремчук Н. А.