

Data Analysis

Тест до лекції 1

Бойченко Вікторія

1. Чому необхідно мати декілька характеристик центральної тенденції? Яка з них стійка до викидів?

Бо може виникнути ситуація, коли викид сильно впливає на обрану тенденцію. Тобто, на прикладі обчислення середнього віку групи з вибіркою 18, 20, 19, 20, 23... при введенні одного неправильного значення 2002 (рік замість віку) середнє значення буде аномальним, коли інші характеристики - медіана чи мода - вкажуть на спотворення значення та наявність викиду.

Стійка до викидів - медіана

2. Яка характеристика розсіювання застосовується до вибіркового середнього?

Дисперсію та середньоквадратичне відхилення

3. Яка характеристика розсіювання дозволяє оцінити розкид 50% вибірки і не враховує вплив викидів?

Міжквартильний розмах

4. Якщо розподіл має довгий правий хвіст, то коефіцієнт асиметрії додатний чи від'ємний?

Додатний

5. Якщо коефіцієнт асиметрії від'ємний, то вибіркє середнє менше за медіану. Це твердження справедливе?

Так

6. Якщо розподіл плосковершинний, то яким має бути коефіцієнт ексцесу? А якщо гостровершинний?

Для плосковершинного коефіцієнт ексцесу < 3 , для гостровершинного > 3

7. Як і для чого здійснюється z-стандартизація даних?

За формулою зводимо до нормального розподілу $\mathcal{N} \sim (0, 1)$

$$z = \frac{x - \mu}{\sigma}$$

де μ - середнє значення,

σ - стандартне відхилення;

Зокрема стандартизація дозволяє:

- (а) порівнювати нормально розподілені набори даних
- (б) визначати нормальність та викиди