**p-value** – величина що використовується при тестуванні статистичних гіпотез.

Фактично це ймовірність помилки при відхиленні нульової гіпотези.

Це ймовірність отримати такі або більш виражені відмінності за умови, що в генеральній

сукупності ніяких відмінностей насправді немає.

**Регре́сія** — форма зв'язку між випадковими величинами. Закон зміни математичного очікування однієї випадкової величини залежно від значень іншої.

**Генеральна** **сукупність** – це множина об’єктів відносно яких ми хотіли би робити висновки в рамках дослідження певної наукової проблеми.

**Вибірка** – це [множина](https://uk.wikipedia.org/wiki/%D0%9C%D0%BD%D0%BE%D0%B6%D0%B8%D0%BD%D0%B0) об'єктів, подій, зразків або [сукупність вимірів](https://uk.wikipedia.org/wiki/%D0%A1%D1%83%D0%BA%D1%83%D0%BF%D0%BD%D1%96%D1%81%D1%82%D1%8C_%D0%B2%D0%B8%D0%BC%D1%96%D1%80%D1%96%D0%B2), за допомогою визначеної процедури вибраних з [статистичної популяції](https://uk.wikipedia.org/wiki/%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%BD%D0%B0_%D0%BF%D0%BE%D0%BF%D1%83%D0%BB%D1%8F%D1%86%D1%96%D1%8F) або [генеральної сукупності](https://uk.wikipedia.org/wiki/%D0%93%D0%B5%D0%BD%D0%B5%D1%80%D0%B0%D0%BB%D1%8C%D0%BD%D0%B0_%D1%81%D1%83%D0%BA%D1%83%D0%BF%D0%BD%D1%96%D1%81%D1%82%D1%8C_%D0%B2%D0%B8%D0%BC%D1%96%D1%80%D1%96%D0%B2) для участі в дослідженні

**Репрезентативність** **вибірки** — здатність вибіркової сукупності відтворювати основні характеристики генеральної сукупності.

Способи формування репрезентативної вибірки:

1. Simple random sample
2. Stratified sample
3. Claster sample

**Міри центральної тенденції**: Мода, Медіана, Середнє.

**Міри мінливості**: Розмах, Дисперсія, Середньоквадратичне відхилення.

**Розмах (Range)** – це різниця максимального і мінімального значення. R = Xmax – Xmin

**Дисперсія** (Variance) – середній квадрат відхилень індивідуальних значень ознаки від їх середньої величини.

Тобто на скільки наші значення відхиляються від середнього значення по вибірці.

Формула для Геральної сукупності: 

Формула для вибірки: 

**Середньоквадратичне відхилення** (δ – сігма) – це корінь з дисперсії.

Показує реальне середнє значення наших відхилень від середнього значення по вибірці.

**Позначки середньоквадратичного відхилення:**

- для Генеральної сукупності це сігма. δ

- для Вибірки це sd (standart deviation)

**z-стандартизація** – це таке перетворення даних, котре дозволяє нам будь-яку нашу шкалу привести до типу: середнє значення нуль, а стандартне відхилення (дисперсія) = 1. (Mz = 0 і Dz = 1)

формула: 

**Інтерквартильний розмах** – це різниця між 3 і 1 квартилем. IQR є х-кою розкиду розподілу величини.

**Квартилі** – це такі значення ознаки які ділять впорядковані дані на певне число рівних частин.

**Центральна гранична теорема** – теорема яка стверджує що сума достатньо великої кількості слабо залежних випадкових величин, котрі мають приблизно однакові масштаби має розподіл близький до нормального.

**Кореляція** - це статистична залежність між випадковими величинами, що носить імовірнісний характер. (ступінь лінійного зв’язку).

**Коваріа́ція** — це міра спільної мінливості двох [випадкових змінних](https://uk.wikipedia.org/wiki/%D0%92%D0%B8%D0%BF%D0%B0%D0%B4%D0%BA%D0%BE%D0%B2%D0%B0_%D0%B7%D0%BC%D1%96%D0%BD%D0%BD%D0%B0).

**Коефіцієнт детермінації** – це квадрат коефіцієнта кореляції. R2. Показує в які степені дисперсія

однієї змінної зумовлена впливом іншої змінної. [0, 1]

**МНК** – це метод знаходження оптимальних параметрів лінійної регресії, при яких сума квадратів помилок (залишків) буде мінімальна.

**Гомоскедастичність** – це постійна мінливість залишків на всіх рівнях незалежної змінної.