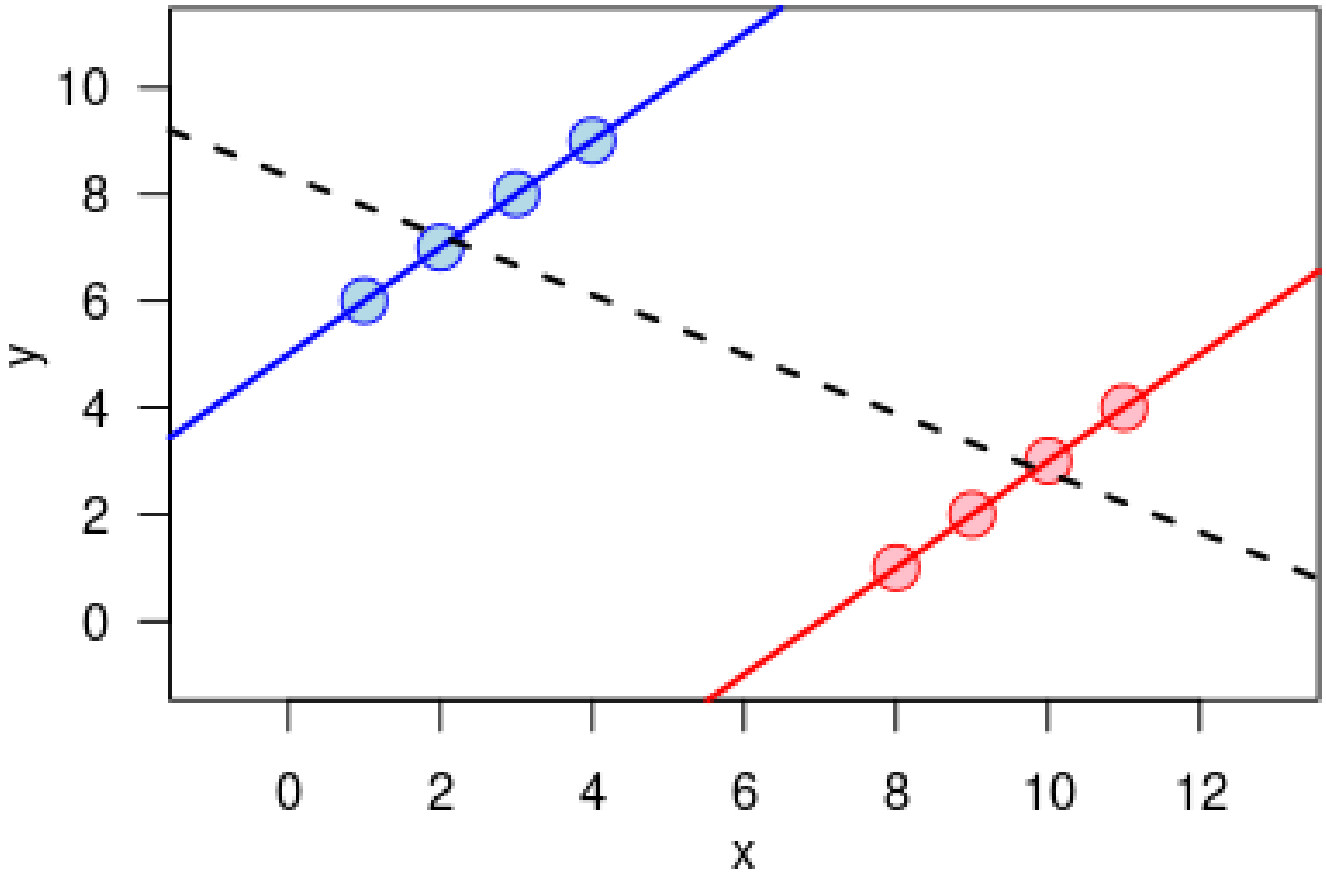
Парадокс Симпсона



Доклад студента группы 131-ПИо Пестерева В.О.

Парадокс Симпсона

- эффект, явление в статистике, когда при наличии двух групп данных, в каждой из которых наблюдается одинаково направленная зависимость, при объединении этих групп направление зависимости меняется на противоположное

Пример М. Гарднера с камнями

Пусть мы имеем четыре набора камней.

Вероятность вытащить чёрный камень набора № 1 выше, чем из набора № 2. В свою очередь, вероятность вытащить чёрный камень из набора № 3 больше, чем из набора № 4.

Объединим набор № 1 с набором № 3 (получим набор I), а набор № 2 — с набором № 4 (набор II). Интуитивно можно ожидать, что вероятность вытащить чёрный камень из набора I будет выше, чем из набора II. Однако, в общем случае такое утверждение неверно.

Пример М. Гарднера с камнями (доказательство)

Пусть n_i — число чёрных камней в i-ом наборе (выборке),

m_i — общее число камней в i-ом наборе при i = 1,2,3,4. По условию:

$$\frac{n_1}{m_1} > \frac{n_2}{m_3}, \frac{n_3}{m_3} > \frac{n_4}{m_4}.$$

Вероятность вытащить чёрный камень из наборов I и II, соответственно:

$$\frac{n_1 + n_3}{m_1 + m_3}, \frac{n_2 + n_4}{m_2 + m_4}.$$

Пример М. Гарднера с камнями (доказательство)

Выражение для набора I не всегда больше выражения для набора II. Например:

$$n_1 = 6,\ m_1 = 13,\ n_2 = 4,\ m_2 = 9,\ n_3 = 6,\ m_3 = 9,\ n_4 = 9,\ m_4 = 14$$

Легко проверить, что $6/13 > 4/9,\ 6/9 > 9/14$

В то время как $12/22 < 13/23$

Причина

- Некорректное усреднение 2х групп данных с различной долей контрольных наблюдений

Начальные данные

Мужчины	Принимавшие лекарство	Не принимавшие лекарство
Выздоровевшие	700	80
Невыздоровевшие	800	130
Соотношение	0.875	0.615

Женщины	Принимавшие лекарство	Не принимавшие лекарство
Выздоровевшие	150	400
Невыздоровевшие	70	280
Соотношение	2.142	1.429

Закономерность не сохраняется

Сумма	Принимавшие лекарство	Не принимавшие лекарство
Выздоровевшие	850	480
Невыздоровевшие	870	410
Соотношение	0.977	1.171

Итоговые данные

Мужчины	Принимавшие лекарство	Не принимавшие лекарство	
		исходные	с весом x22.07
Выздоровевшие	700	80	1765
Невыздоровевшие	800	130	2869
Соотношение	0.875	0.615	

Сумма	Принимавшие лекарство	Не принимавшие лекарство	
		исходные	с весом x22.07
Выздоровевшие	850	480	2165
Невыздоровевшие	870	410	3149
Соотношение	0.977	1.171	0.685