Тема 6

Измерение связи

Вопросы:

- Генеральная совокупность и частотное распределение
- Измерение связи между количественными переменными
- Измерение связи между качественными переменными

Вопрос 1

Генеральная совокупность и частотное распределение

Генеральная совокупность и выборка

Рис. 2.2. Генеральная совокупность объектов (N) и выборка из нее (n) (n << N)

Основные понятия

- генеральная совокупность – множество элементов, обладающих каким-то одним или несколькими признаками (вариантами)
- признак = варианта – переменная величина, которой характеризуется каждый элемент генеральной совокупности
- количественная варианта может быть:
  - дискретной – которая может принимать только целочисленные значения
  - непрерывной – которая может принимать любые значения

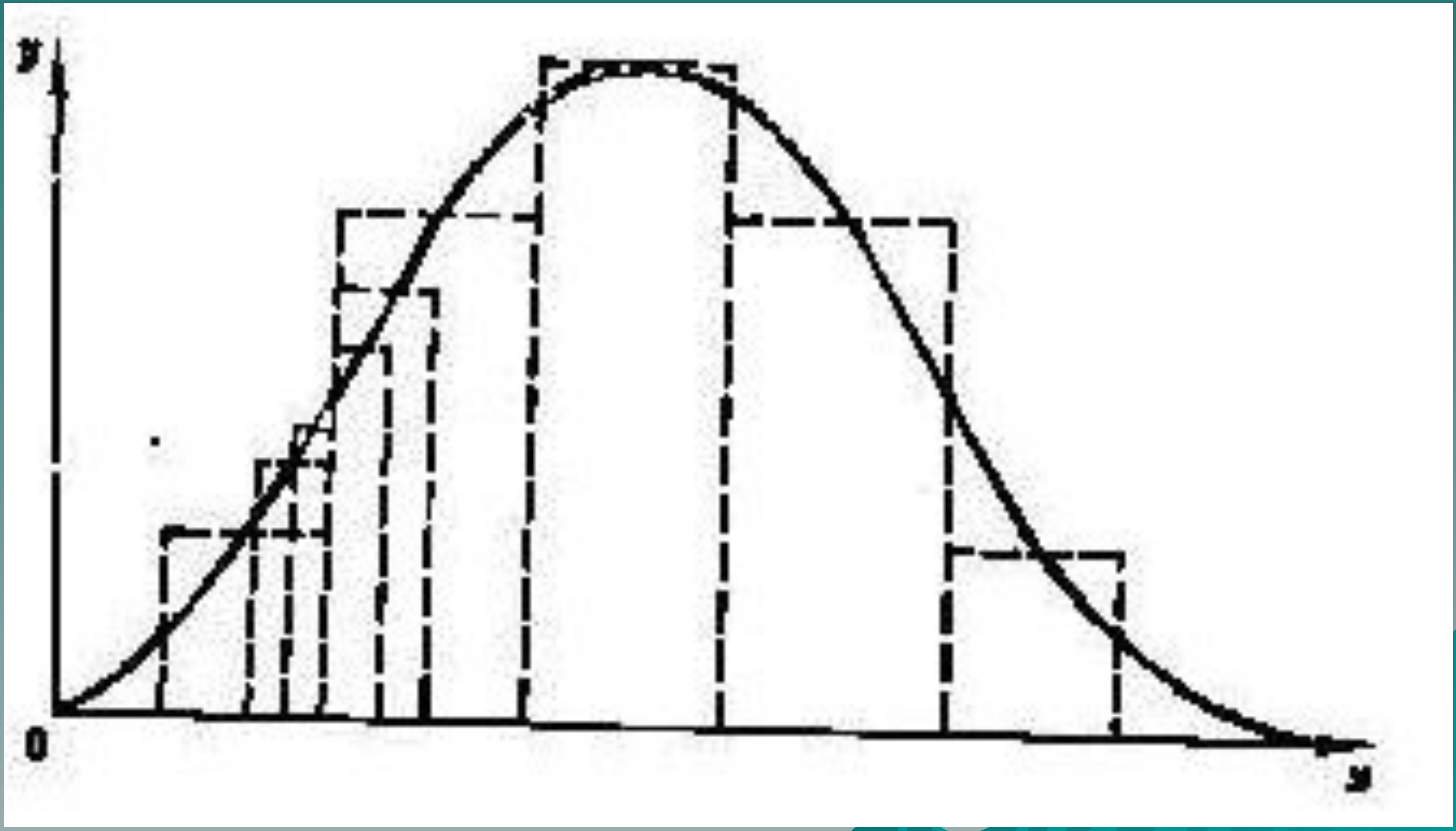
Распределение генеральной совокупности по дискретной варианте:

- сгруппировать все элементы ГС по признакам
- подсчитать количество элементов в каждой группе
- оформить результаты как два ряда чисел, которые дают частотное распределение:

Значение варианты	X_1	X_2	…	X_k
Количество случаев (частота)	n_1	n_2	…	n_k

- графическое представление дает ломаную линию = полигон распределения

Полигон распределения



Распределение генеральной совокупности по непрерывной варианте:

- весь диапазон значений варианты разбить на n класс-интервалов (их количество м.б. разным, но они должны быть равными)
- подсчитать количество элементов в каждом класс-интервале оценить частоту каждого класс-интервала

Класс-интервалы	X_1	X_2	…	X_k
Частота класс-интервала	n_1	n_2	…	n_k

- графическое представление дает ломаную линию, о называется полигоном распределения
- при увеличении количества класс-интервалов и следовательно при уменьшении числа элементов в каждом из них, полигон распределения сглаживается; при бесконечном числе интервалов полигон превращается в кривую распределения
- кривая распределения - это функция плотности распределения
- интеграл от нее по области изменения варианты - это функция распределения

Каждое распределение характеризуется 2 типами параметров:

- параметры положения или средние:
  - среднее арифметическое
  - медиана
  - мода
- меры рассеивания:
  - дисперсия
  - среднее квадратическое отклонение

Вопрос 2

Измерение связи между количественными переменными

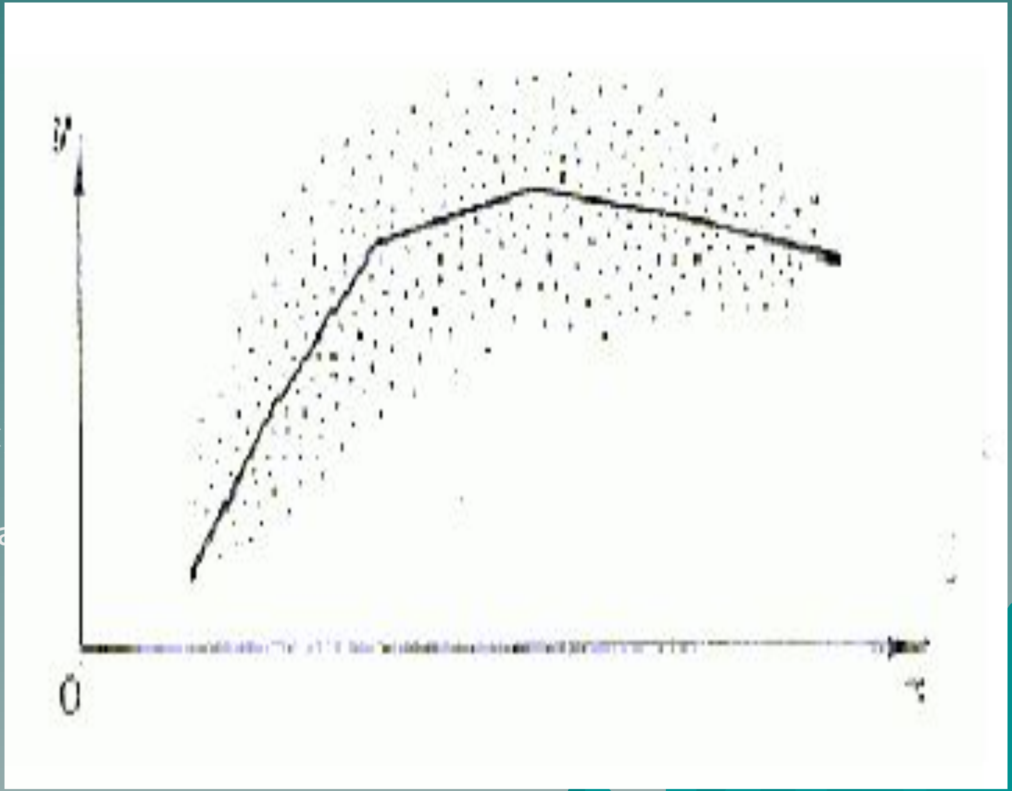
Типы связи

- связь между количественными переменными может быть:
    - функциональной
    - Нефункциональной
- функциональная – такая связь, при которой каждому значению независимой переменной (x) ставится определенное значение зависимой переменной (y); она бывает:
    - однозначной
    - многозначной
- нефункциональная – такая связь, при которой каждому значению одной переменной (x) ставится распределение значений другой переменной (y); она бывает:
    - регрессионной
    - корреляционной

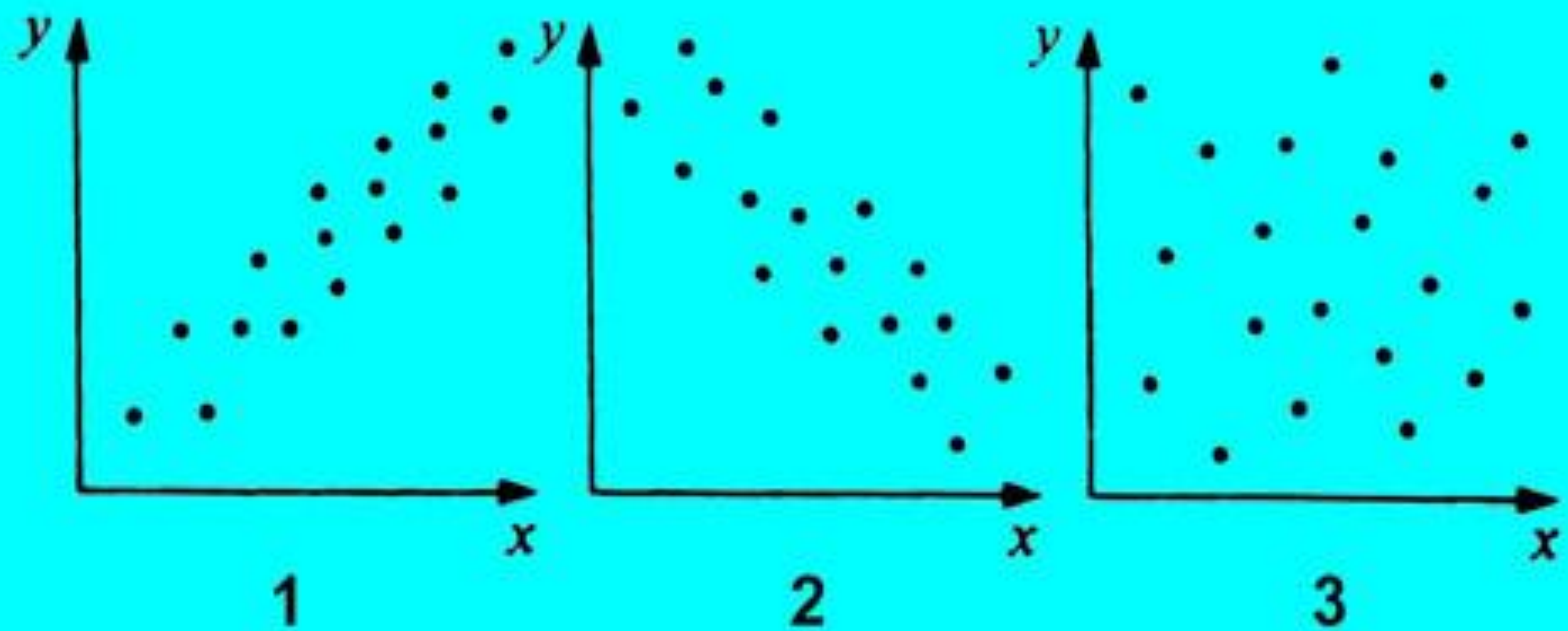
Построение регрессионной связи

Регрессионная связь – связь, характеризующая изменение среднего (y) от (x)

- ***например***, связь между ростом мужа и жены (N = 100):
- по оси (x) – рост мужа
- по оси (y) – рост жены
- точка на плоскости – супружеская пара
- полученное графическое изображение – корреляционное поле
- разбиваем (x) на класс-интервалы
- находим среднее значение (y) на каждом класс-интервале и эту точку наносим на график
- соединяем все полученные точки ломаной линией = эмпирическая линия регрессии (x по (y)
- ломаная линия выражает зависимость среднего роста жены в зависимости от роста мужа
- взяв другие 100 супружеских пар, получим несколько другую эмпирическую линию, которая будет все же близка к первой --- обе эти линии лежат около некоторой плавной линии = теоретической линии регрессии



Корреляционное поле и наличие статистической связи



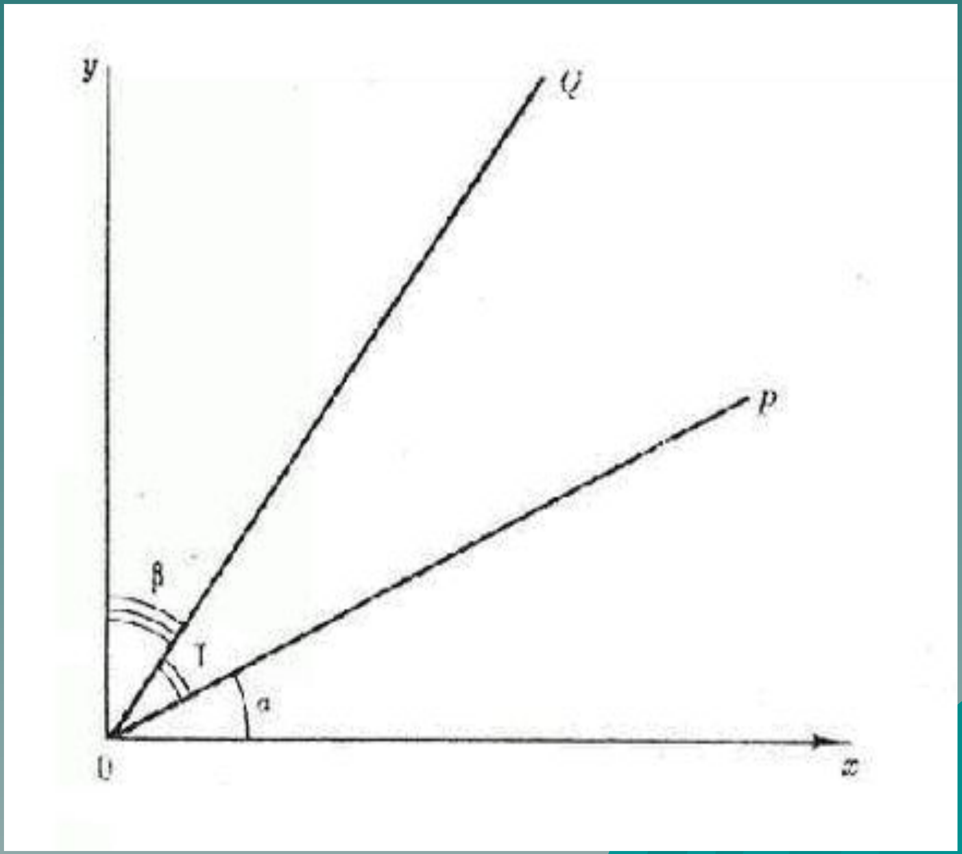
1 - высокая положительная корреляция;
2 - высокая отрицательная корреляция;
3 - корреляция отсутствует (близка к 0, признаки не связаны)

Корреляционная связь и ее геометрическая интерпретация

Корреляционная связь – связь между признаками (x) и (y), определяемая как среднее геометрическое из коэффициентов регрессии (x) по (y) и (y) по (x)

- графическое представление: две линии регрессии (x) по (y) и (y) по (x); чем они ближе, тем больше корреляция между (x) и (y)
- аналитическое выражение для случая линейной регрессии:

$$r_{xy} = \sqrt{tg\alpha tg\beta}.$$



Вопрос 3

Измерение связи между качественными переменными

Качественные переменные

Качественные – переменные, полученные при измерении в рамках 2 шкал:

- номинальной
- ординальной

Измерение связи между номинальными переменными

- имеются признаки А и В
- они принимают значения А1, А2 …, Аm и В1, В2, … , Вn
- nji – количество лиц с образованием Aj и доходом Bi
- вместо nji вводится относительная частота P ji
- тогда коэффициент связи признаков А и В выражается коэффициентом Пирсона:

$$\varphi^2 = \sum \frac{\left(P_{ij} - P_{\cdot i} P_{\cdot j}\right)^2}{P_{\cdot i} P_{\cdot j}},$$

	B_1	B_2	…	B_i	B_n
A_1	n_{11}	n_{12}		n_{1i}	n_{1n}
A_2	n_{21}	n_{22}		n_{2i}	n_{2n}
…					
A_j	n_{j1}	n_{j2}		n_{ji}	n_{jn}
A_m	n_{m1}	n_{m2}		n_{mi}	n_{mn}

Измерение связи между ординальными переменными

- строится таблица сопряженности

Объекты	Ранг объекта		Разность между рангами по 1 и 2 признакам	
	по 1 признаку	по 2 признаку	Значение разности	Квадрат разности

- связь рассчитывается с помощью коэффициента ранговой корреляции Спирмена:

$$\rho = 1 - \frac{6\sum D_i^2}{n\left(n^2 - 1\right)},$$

Выводы по теме:

- Признаком (вариантой) называется переменная величина, которой характеризуется каждый элемент генеральной совокупности. Признаки могут быть дискретными и непрерывными

- Для измерения связи между признаками статистической совокупности необходимо построить частотное распределение значений каждого признака, а также представить его набором статистик – средних и мер рассеивания

- Связь между признаками может быть функциональной и статистической. Связь признаков в социологии чаще всего имеет статистический характер и может быть выражена в форме регрессионной и корреляционной связи

- Измерению подлежат корреляционная связь не только между непрерывными признаками, но также и между дискретными. В последнем случае используются коэффициент номинальной корреляции Пирсона и коэффициент ранговой корреляции Спирмена.