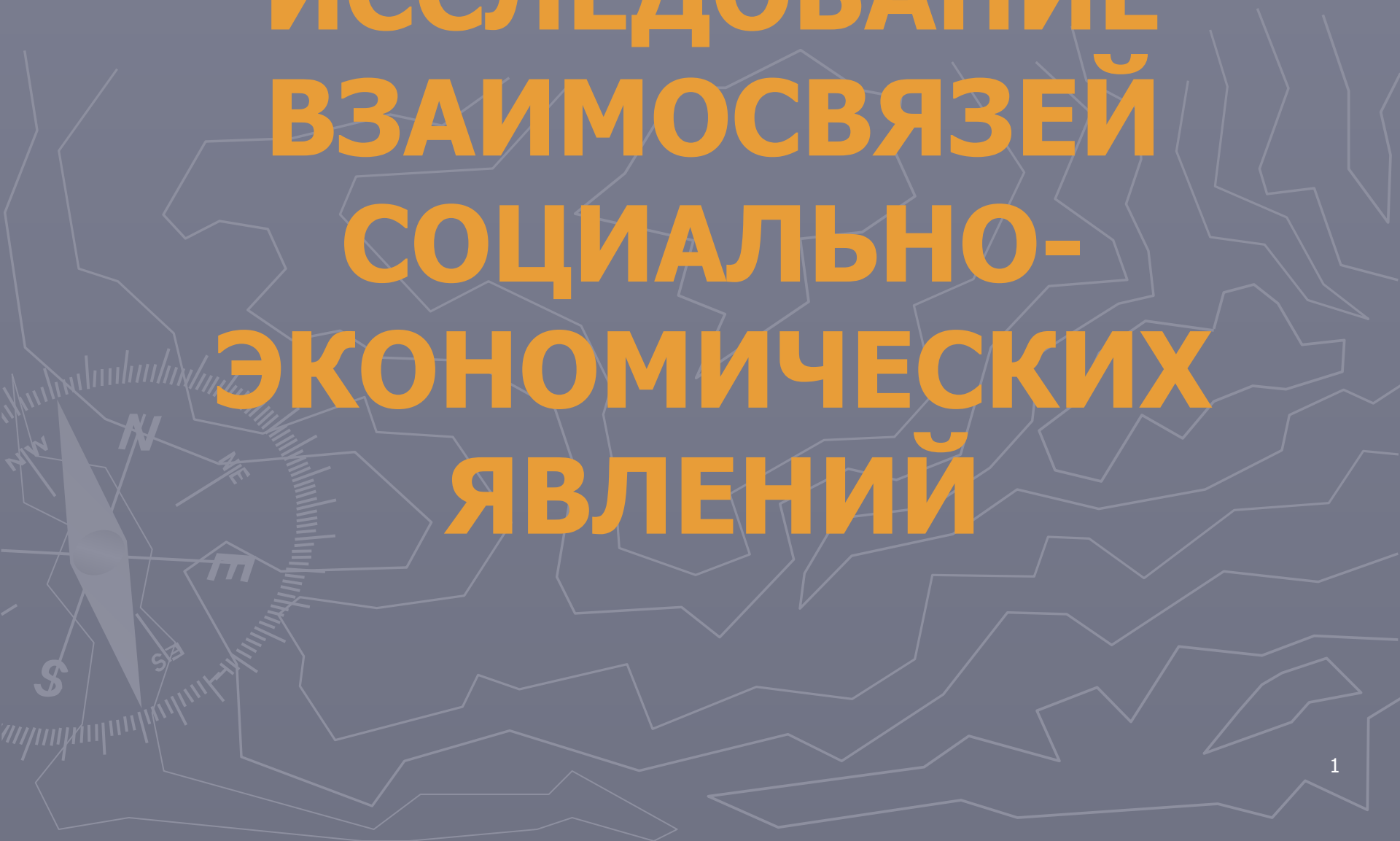


ИССЛЕДОВАНИЕ ВЗАИМОСВЯЗЕЙ СОЦИАЛЬНО- ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ



Виды и формы связи между явлениями

- ▶ Взаимосвязанные признаки подразделяются на **факторные** (под их воздействием изменяются другие, зависящие от них признаки) и **результативные**.

Функциональная связь

- ▶ При **функциональной** связи каждому значению факторного признака соответствует строго определенное одно или несколько значений результативного признака; для выявления такой связи достаточно одного наблюдения. Она полностью сохраняет свою силу и проявляется во всех случаях наблюдения и для всех единиц наблюдения. Функциональные связи иначе называются полными

Стохастическая связь

- ▶ Связь между случайными величинами называется **стохастической**. Частный случай стохастической связи — статистические связи. **Статистические связи** подразделяются на **корреляционные и регрессионные**. Корреляционные связи отражают зависимость между случайными величинами, а регрессионные — между случайными и неслучайными.

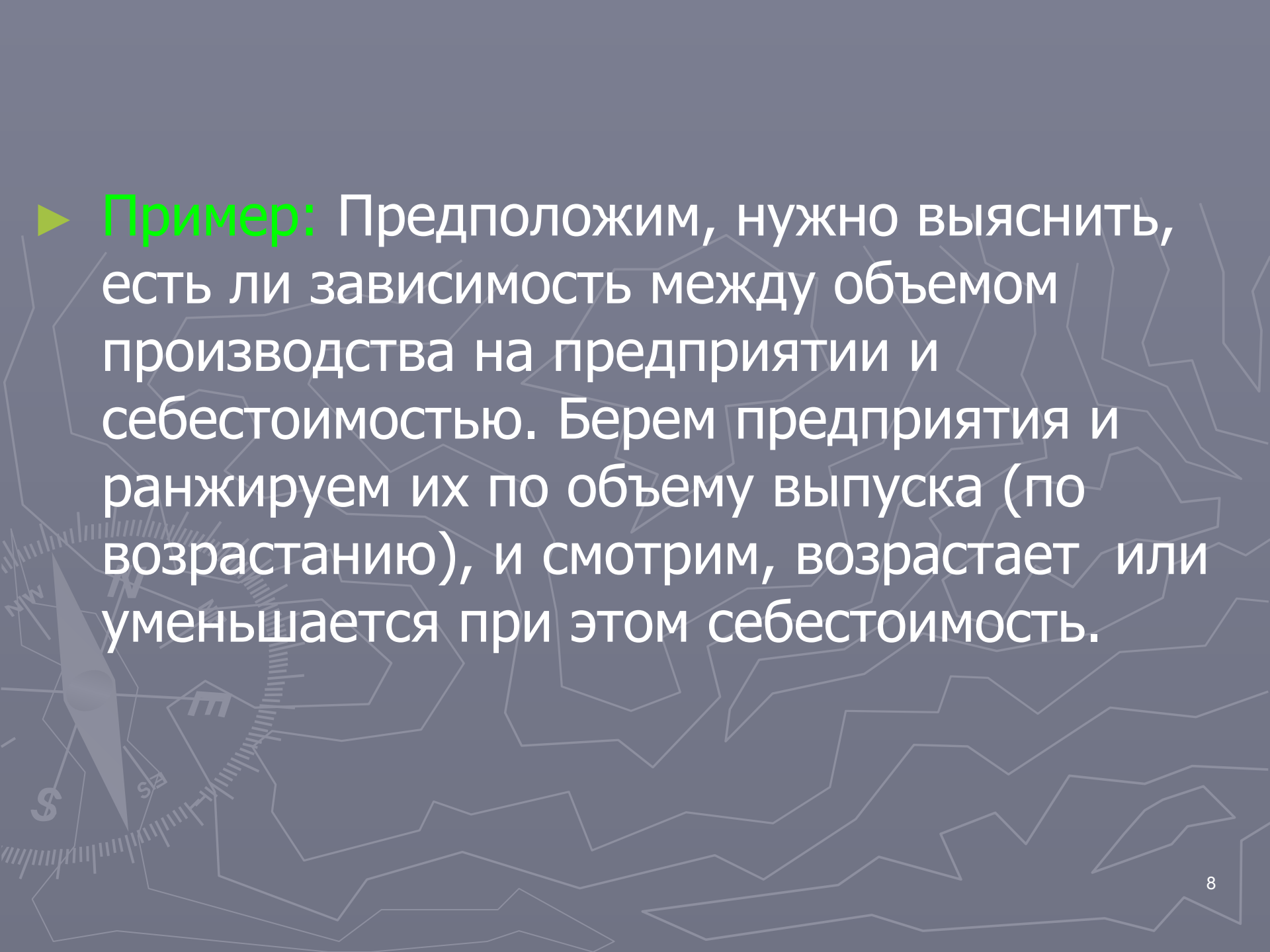
- При *статистической* связи одному и тому же значению факторного признака может соответствовать несколько значений результативного признака. Особенность статистических закономерностей состоит в том, что строгую функциональную зависимость между явлениями выявить нельзя. Статистические закономерности проявляются только для большого числа единиц совокупности, а динамические проявляются у каждой отдельной единицы.

Основные приемы изучения взаимосвязей

- ▶ Для изучения, измерения и количественного выражения взаимосвязей между явлениями применяются различные методы. Важнейшие из них: метод сопоставления параллельных рядов, балансовый, графический, методы аналитических группировок, дисперсионного и корреляционного анализа. Наибольший эффект достигается при комбинировании нескольких методов.

Метод сравнения параллельных рядов

- ▶ Приводится ряд данных по одному признаку и параллельно с ним по другому признаку, связь с которым предполагается. По вариации признака в первом и втором ряду судят о наличии связи признаков. Такой метод позволяет вывести только направление связи, но не измерить ее.

- 
- **Пример:** Предположим, нужно выяснить, есть ли зависимость между объемом производства на предприятии и себестоимостью. Берем предприятия и ранжируем их по объему выпуска (по возрастанию), и смотрим, возрастает или уменьшается при этом себестоимость.

► **Балансовый метод** состоит в том, что данные взаимосвязанных показателей изображаются в виде таблицы и располагаются таким образом, чтобы итоги между отдельными ее частями были равны, т. е., чтобы был баланс. Балансовый метод используется для характеристики взаимосвязи между производством и распределением продуктов, денежными доходами и расходами населения. Почти все внутренние и внешние хозяйственные связи выражаются в виде балансов.

Регионы отправле ния	Регионы прибытия				Итого отправлено
	А	Б	В	Г	
А	25	75	80	100	280
Б	100	20	95	30	245
В	80	70	30	65	245
Г	70	45	55	35	205
Итого прибыло	275	210	260	230	975

- **Графический метод** может использоваться как самостоятельно, так и совместно с другими методами.

Если конкретные данные перенести на график, то получим поле точек. На оси абсцисс откладываются значения факторного признака x , а на оси ординат – значения результативного признака y . Каждая единица, обладающая определенным значением факторного и результативного признака, обозначается точкой. По совместному расположению точек на графике делают вывод о направлении и наличии зависимости

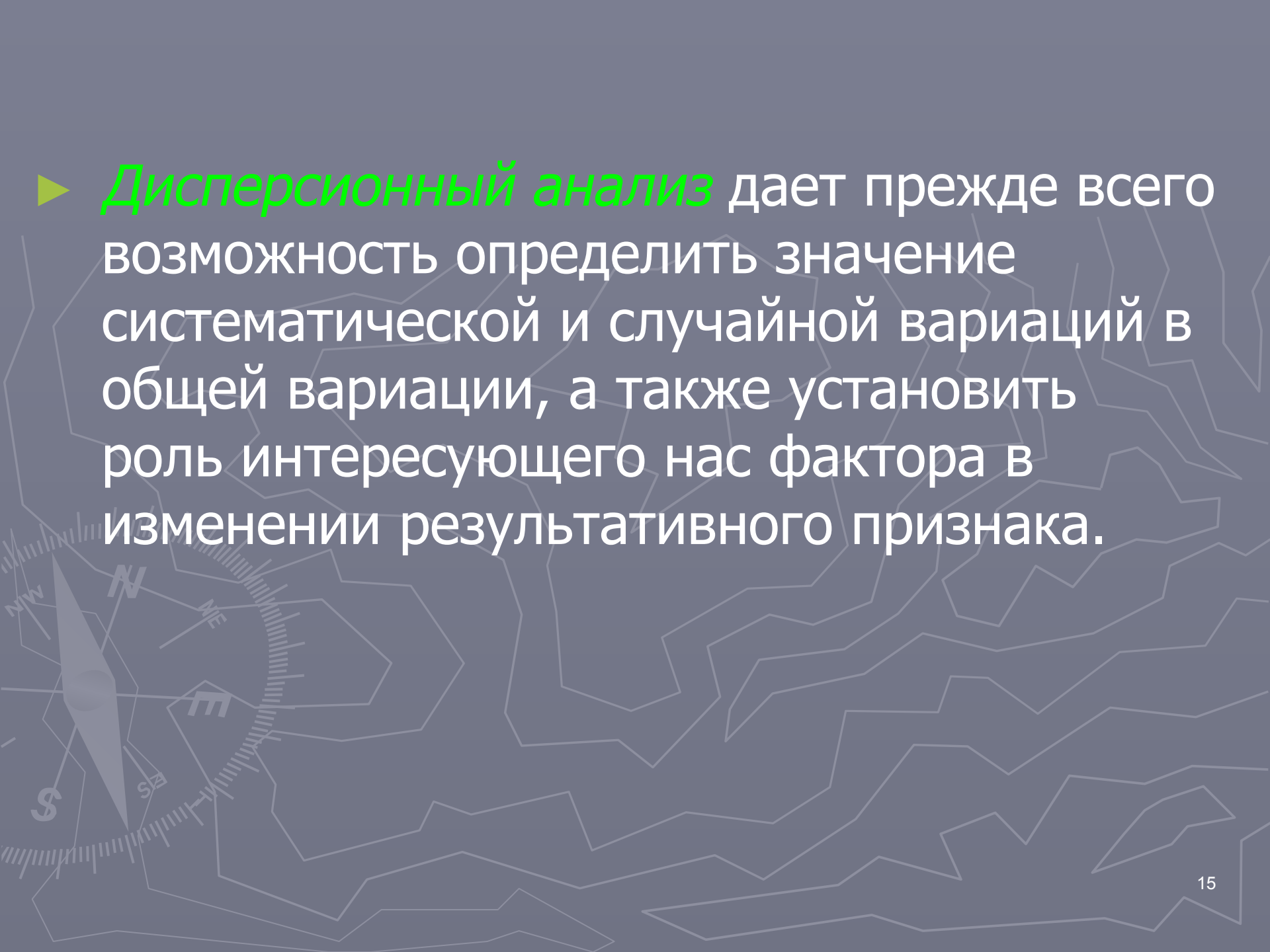
- ▶ Беспорядочное расположение говорит об отсутствии связи. Напротив, чем сильнее связь, тем теснее точки группируются вокруг определенной линии. Если точки на графике концентрируются вокруг прямой, идущей снизу вверх, зависимость между признаками **прямая**. Если точки концентрируются вокруг прямой, спускающейся сверху вниз, то это свидетельствует о наличии **обратной зависимости**.

Метод аналитической группировки

- ▶ Единицы статистической совокупности группируются, как правило, по факторному признаку и для каждой группы рассчитывается средняя или относительная величина по результативному признаку. Затем изменения средних или относительных значений результативного признака сопоставляются с изменениями факторного признака для выявления характера связи между ними. Сначала выбираются два признака: факторный и результативный. По факторному признаку производится группировка, а по результативному – подсчет средних или относительных величин. Путем сопоставления характера изменений значений факторного и результативного признака можно сделать вывод о наличии связи и ее направлении. При помощи метода аналитической группировки можно сделать вывод и о тесноте связи.

Пример : Количество вкладчиков и средний остаток вклада по трем филиалам Сбербанка (тыс. руб.)*.

Число вкладчиков, человек	Средний остаток по вкладу, тыс. руб.
13 500	1,50
1 290	1,81
2 205	2,05

- 
- *Дисперсионный анализ* дает прежде всего возможность определить значение систематической и случайной вариаций в общей вариации, а также установить роль интересующего нас фактора в изменении результативного признака.

Анализ проходит четыре итерации.

1. Определяем, какой признак факторный, какой – результативный.
2. Производим группировку по факторному признаку.
3. Считаем среднее значение факторного и результативного признака в группах.
4. Выясняем взаимосвязь между этими средними.

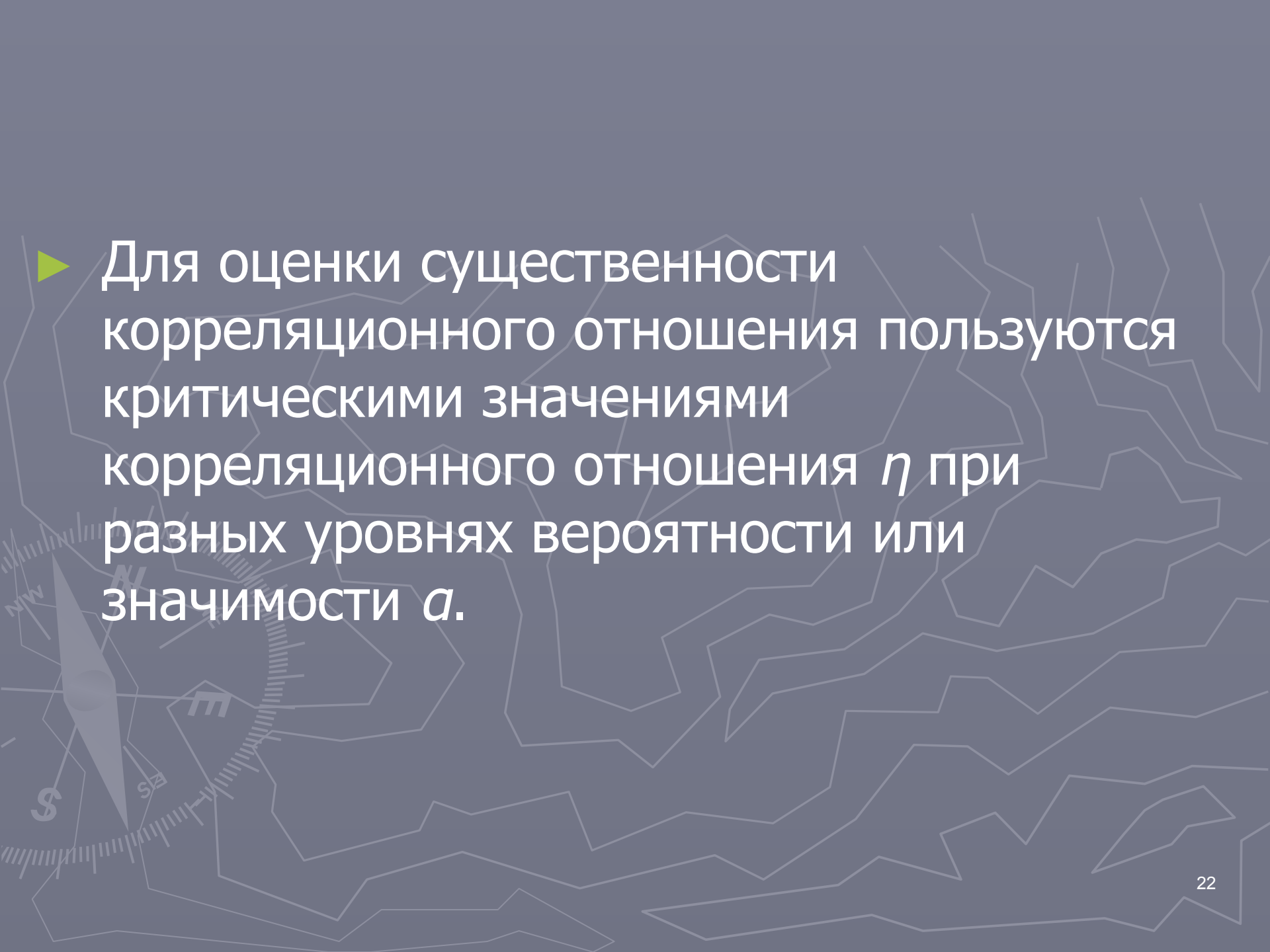
- Для оценки тесноты связи по результатам факторной группировки используется **межгрупповая дисперсия**. Напомним, что межгрупповая дисперсия характеризует колеблемость групповых средних вокруг общей средней, которая возникает под действием факторов, положенных в основу группировки. Если при этом групповая средняя равна средней общей, значит, фактор, положенный в основу группировки, на результативный признак не влияет.

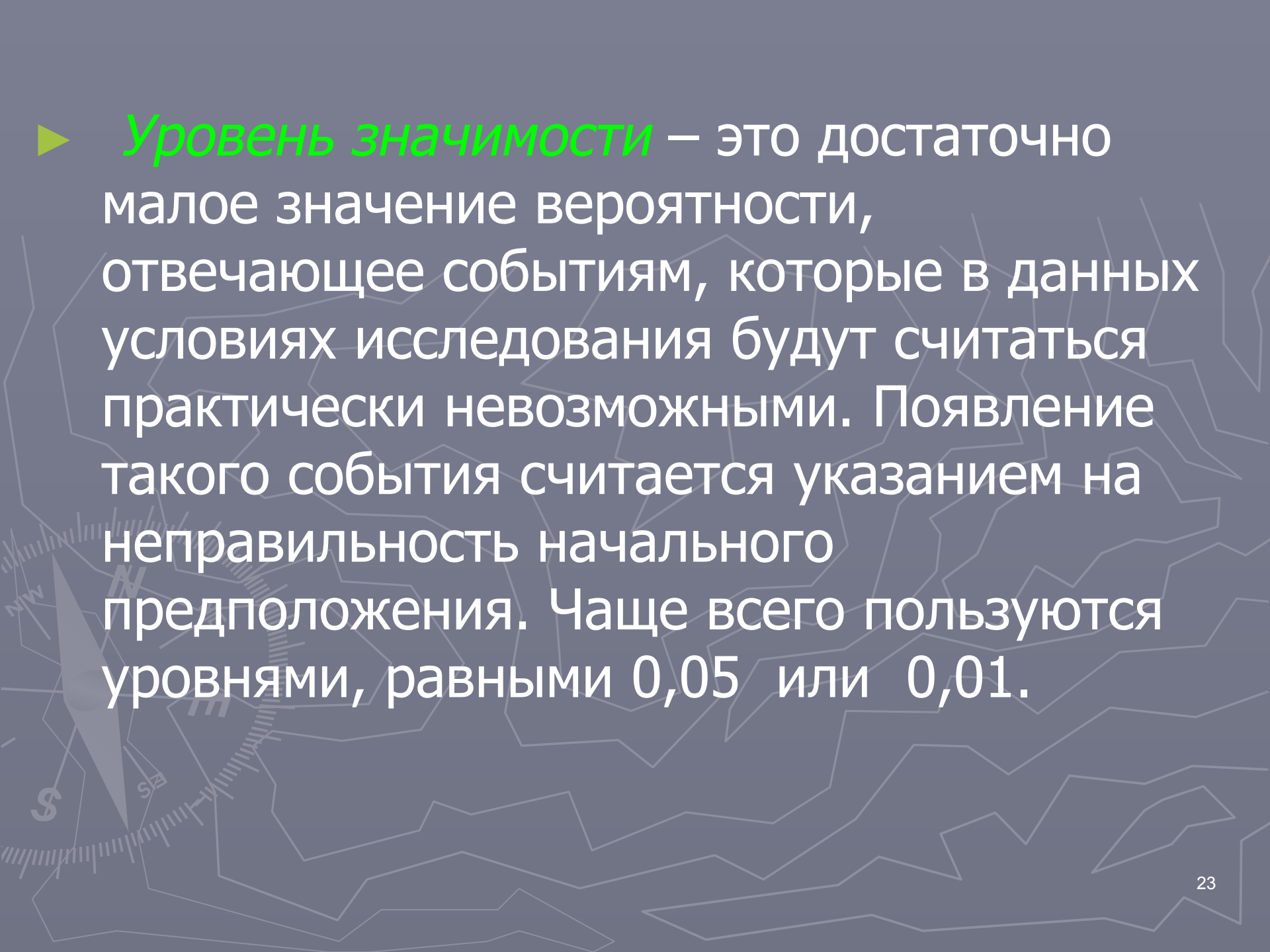
- Для характеристики тесноты корреляционной связи между признаками в аналитических группировках межгрупповую дисперсию сопоставляют с общей. Это сопоставление называется **коэффициентом детерминации**. Он показывает, какая часть общей дисперсии связана с факторами, положенными в основу группировки.

- ▶ Подобные операции выполняют и с **корреляционным отношением**. Оно характеризует долю вариации результативного признака, вызванной действием факторного признака, положенного в основание группировки. Корреляционное отношение по своему абсолютному значению колеблется в пределах **от 0 до 1**. Чем ближе корреляционное отношение к 1, тем большее влияние оказывает факторный признак на результативный.

- ▶ Если факторный признак не влияет на результативный, то вариация, обусловленная им, будет равна нулю ($\delta^2 = 0$); корреляционное отношение также равно нулю ($\eta = 0$), что говорит о полном отсутствии связи. Напротив, если результативный признак изменяется только под воздействием одного факторного признака, то вариация, обусловленная этим признаком, будет равна общей вариации ($\delta^2 = \sigma^2$) и корреляционное отношение будет равно единице ($\eta = 1$), что говорит о существовании полной связи.

- **Дисперсионный анализ** позволяет не только определить роль случайной и систематической вариаций в общей вариации, но и оценить достоверность вариации, обнаруженной методом аналитических группировок. Определение достоверности вариации дает возможность с заданной степенью вероятности установить, вызвана ли межгрупповая вариация признаком, положенным в основание группировки, или она является результатом действия случайных причин.

- 
- Для оценки существенности корреляционного отношения пользуются критическими значениями корреляционного отношения η при разных уровнях вероятности или значимости α .

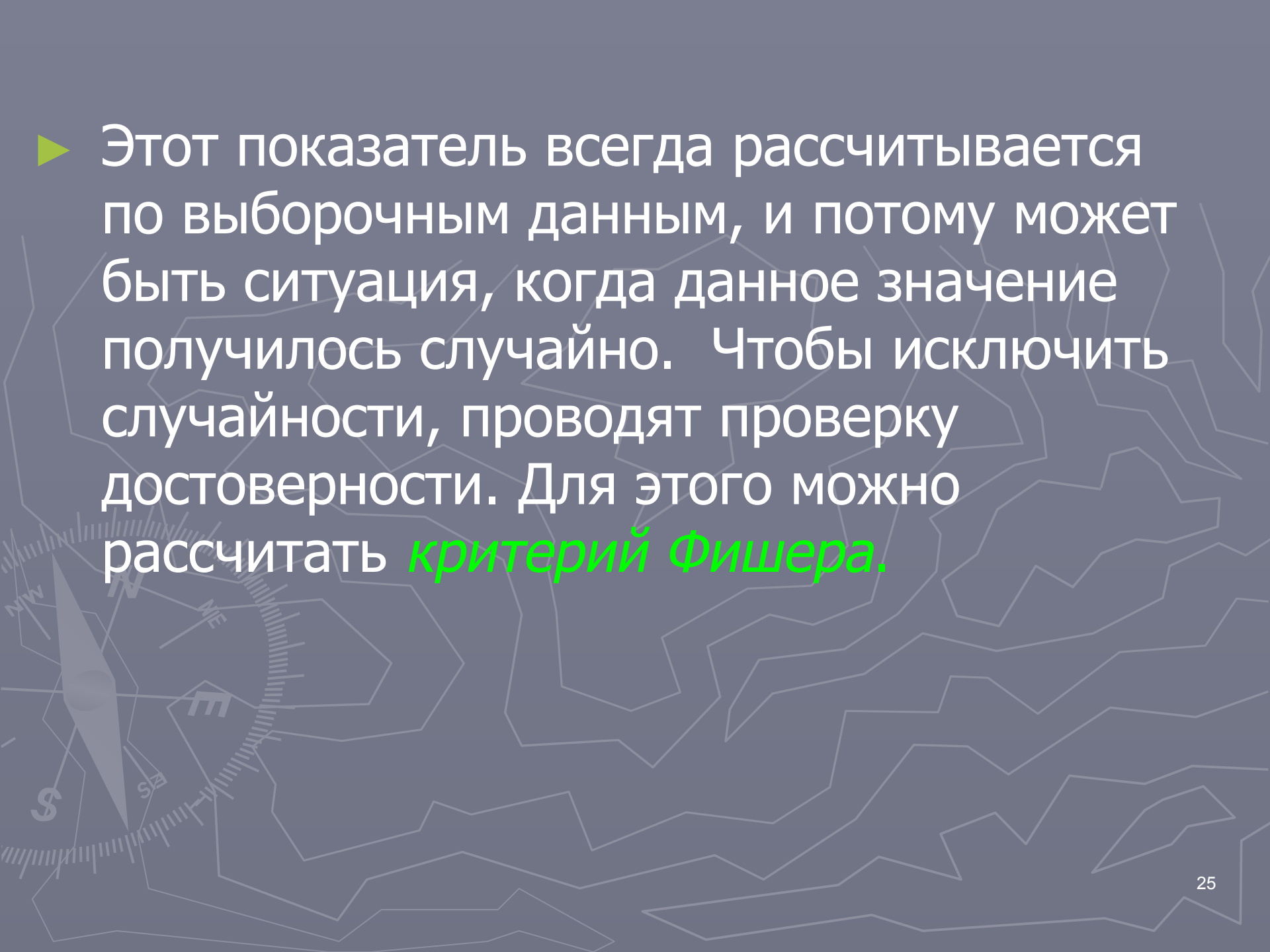
- 
- **Уровень значимости** – это достаточно малое значение вероятности, отвечающее событиям, которые в данных условиях исследования будут считаться практически невозможными. Появление такого события считается указанием на неправильность начального предположения. Чаще всего пользуются уровнями, равными 0,05 или 0,01.

$$k1 = m - 1,$$

где m — число групп,
для случайной дисперсии

$$k2 = n - m,$$

где n — число вариантов,
 m — число групп.

- 
- ▶ Этот показатель всегда рассчитывается по выборочным данным, и потому может быть ситуация, когда данное значение получилось случайно. Чтобы исключить случайности, проводят проверку достоверности. Для этого можно рассчитать **критерий Фишера**.

Критерий Фишера

$$F = \frac{\delta^2}{n-1} : \frac{\overline{\sigma_i^2}}{n-m}.$$

Фишер составил таблицы, по которым можно определить значения критерия. Если мы получаем значение F , большее, чем значение в таблице, то рассчитываемое значение получилось неслучайно (с вероятностью 99% или 95%, в зависимости от того, какую таблицу мы взяли). Критерий Фишера можно использовать только тогда, когда распределение близко к нормальному

Особенности корреляционно-регрессионных связей

- ▶ **Корреляционная связь** проявляется, когда одному и тому же значению факторного признака соответствует ряд значений признака-результата, причем связь обнаруживается в виде тенденции изменения среднего значения результативного признака в зависимости от изменения факторного признака. Это свободная и неполная связь.

- ▶ При **корреляционной связи** имеет место не изменение функции в зависимости от изменения аргумента, а имеет место вариация результативного признака вокруг его среднего значения в зависимости от изменения факторного признака. Вот почему корреляционная связь является не строгой. Кроме того, как правило, корреляционные зависимости являются не полными: мы не знаем всех факторов и их воздействия.

Корреляционно-регрессионный позволяет решить две задачи:

- ▶ определение формы связи;
- ▶ измерение тесноты связи.



► *Первая задача* заключается в определении формы связи, так как от этого зависит конечный результат изучения взаимосвязи между признаками. Первая задача решается путем нахождения уравнения регрессии.

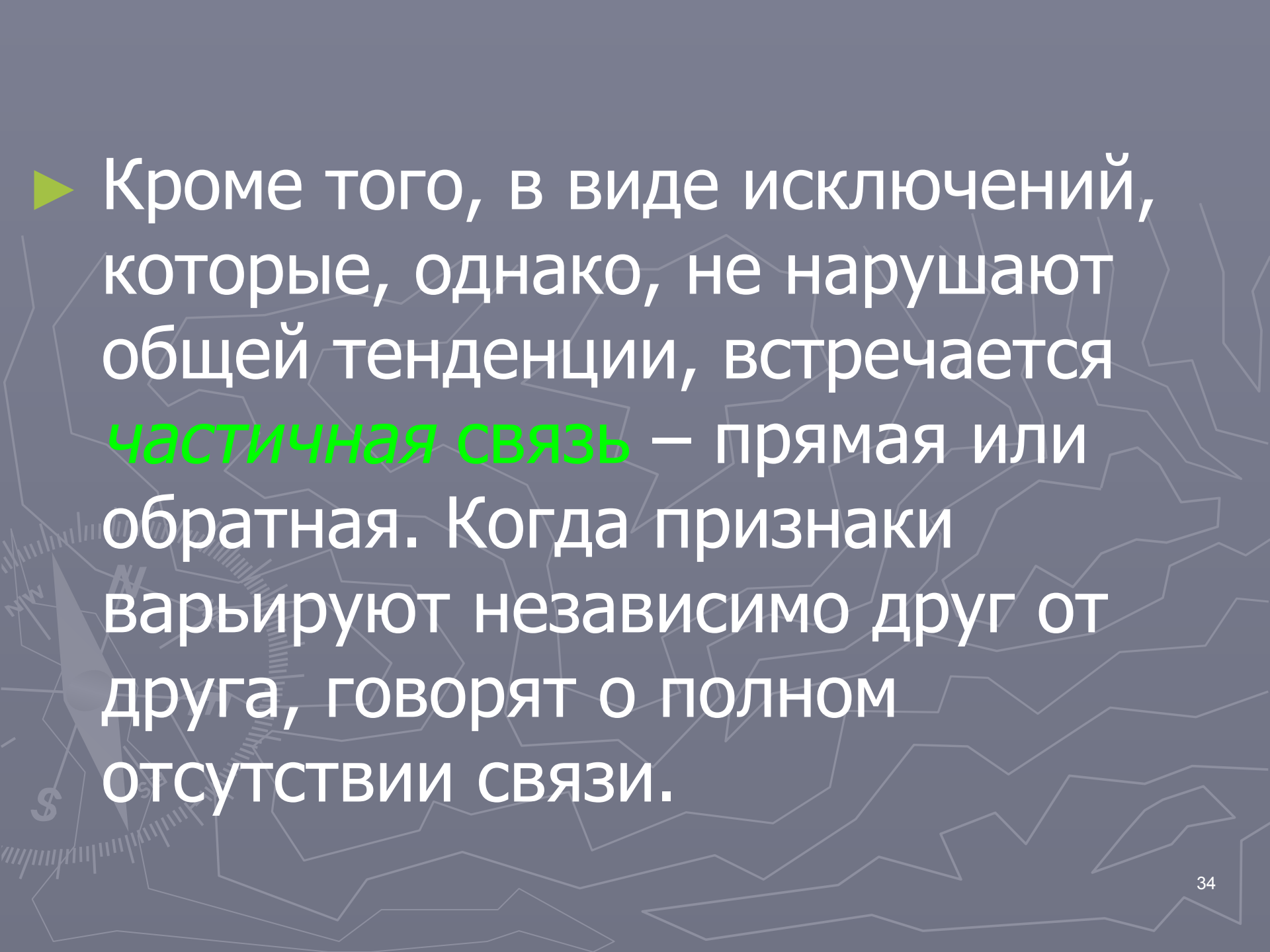
► *Вторая задача* состоит в измерении тесноты, т. е. меры связи между признаками с целью установить степень влияния данного фактора на результат и оценки степени влияния неучтенных факторов.

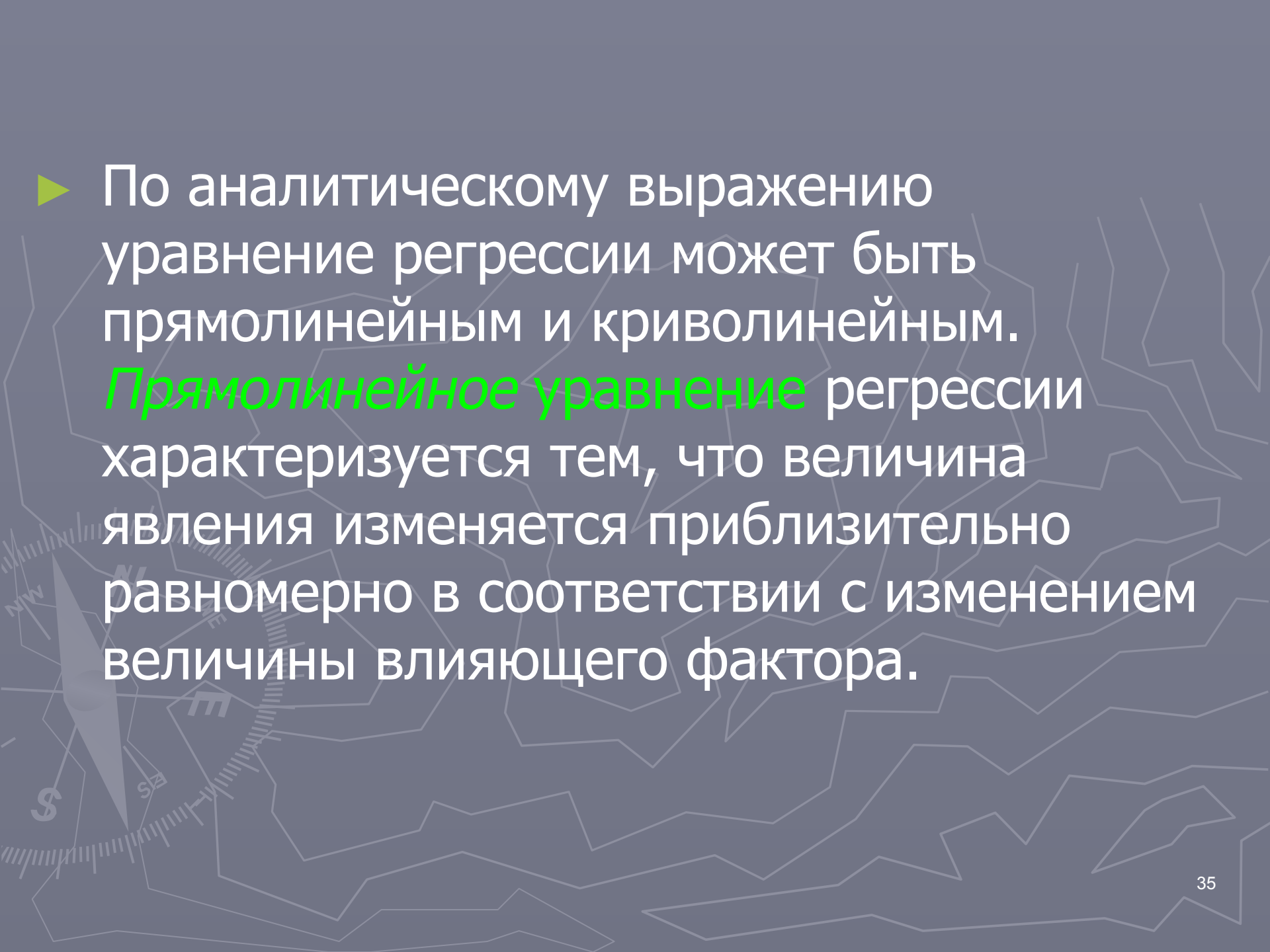
Этапы корреляционного анализа:

- 1) предварительный анализ объекта исследования;
- 2) сбор и первичная обработка информации;
- 3) построение уравнения регрессии и определение его параметров;
- 4) проверка адекватности полученной модели.

Виды корреляционно-регрессионных связей

- ▶ **прямой связи** с увеличением аргумента x функция y также увеличивается без всяких единичных исключений.
- ▶ **обратной связи** увеличение факторного признака x сопровождается уменьшением результативного признака y без всяких единичных исключений.

- 
- ▶ Кроме того, в виде исключений, которые, однако, не нарушают общей тенденции, встречается **частичная связь** – прямая или обратная. Когда признаки варьируют независимо друг от друга, говорят о полном отсутствии связи.

- 
- По аналитическому выражению уравнение регрессии может быть прямолинейным и криволинейным. *Прямолинейное уравнение* регрессии характеризуется тем, что величина явления изменяется приблизительно равномерно в соответствии с изменением величины влияющего фактора.

- ▶ **Прямолинейная зависимость** в этом случае может быть выражена уравнением прямой:

$$y = a_0 + a_1 x,$$

- ▶ Параметр называется **коэффициентом регрессии** и показывает, насколько в среднем отклоняется величина результативного признака y при отклонении величины факторного признака x на одну единицу.

- ▶ Если происходит неравномерное изменение явления в связи с изменением величины влияющего фактора, то такая связь называется *криволинейной*. Математически криволинейная зависимость может быть выражена уравнением криволинейной связи. В экономическом анализе для ее выражения часто пользуются *уравнением параболы второго порядка*:

$$y = a_0 + a_1x + a_2x^2$$

a_2 характеризует степень ускорения или замедления кривизны параболы

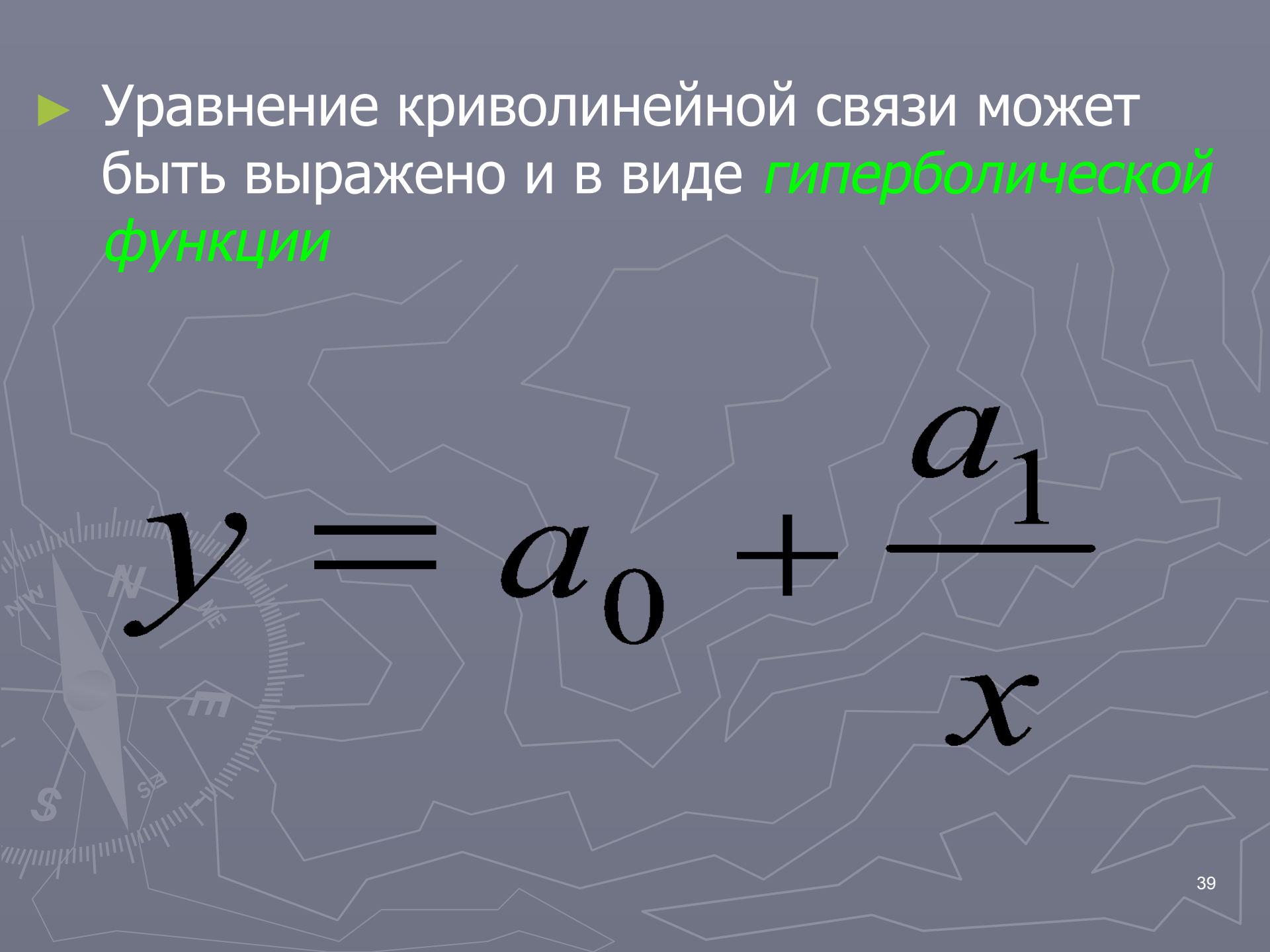
$a_2 > 0$ парабола имеет минимум,

$a_2 < 0$ — максимум.

a_1 характеризует угол наклона кривой,

a_0 — начало кривой.

- Уравнение криволинейной связи может быть выражено и в виде *гиперболической функции*

The background of the slide is a dark gray map with white contour lines. In the bottom-left corner, there is a compass rose with a needle pointing towards the top-left. The compass rose has labels for 'N' (North), 'NW' (Northwest), 'SE' (Southeast), and 'S' (South).
$$y = a_0 + \frac{a_1}{x}$$

Аналитическое выражение связи.

- ▶ Аналитические методы – основной способ изучения связи. Они делятся на **непараметрические и параметрические**. Применение методов корреляционного анализа дает возможность выражать связь между признаками аналитически (в виде уравнения) и придавать ей количественное выражение.

Измерение тесноты связи

- Для измерения тесноты прямолинейной связи между двумя признаками используется линейный коэффициент корреляции, который обозначается r_{xy}

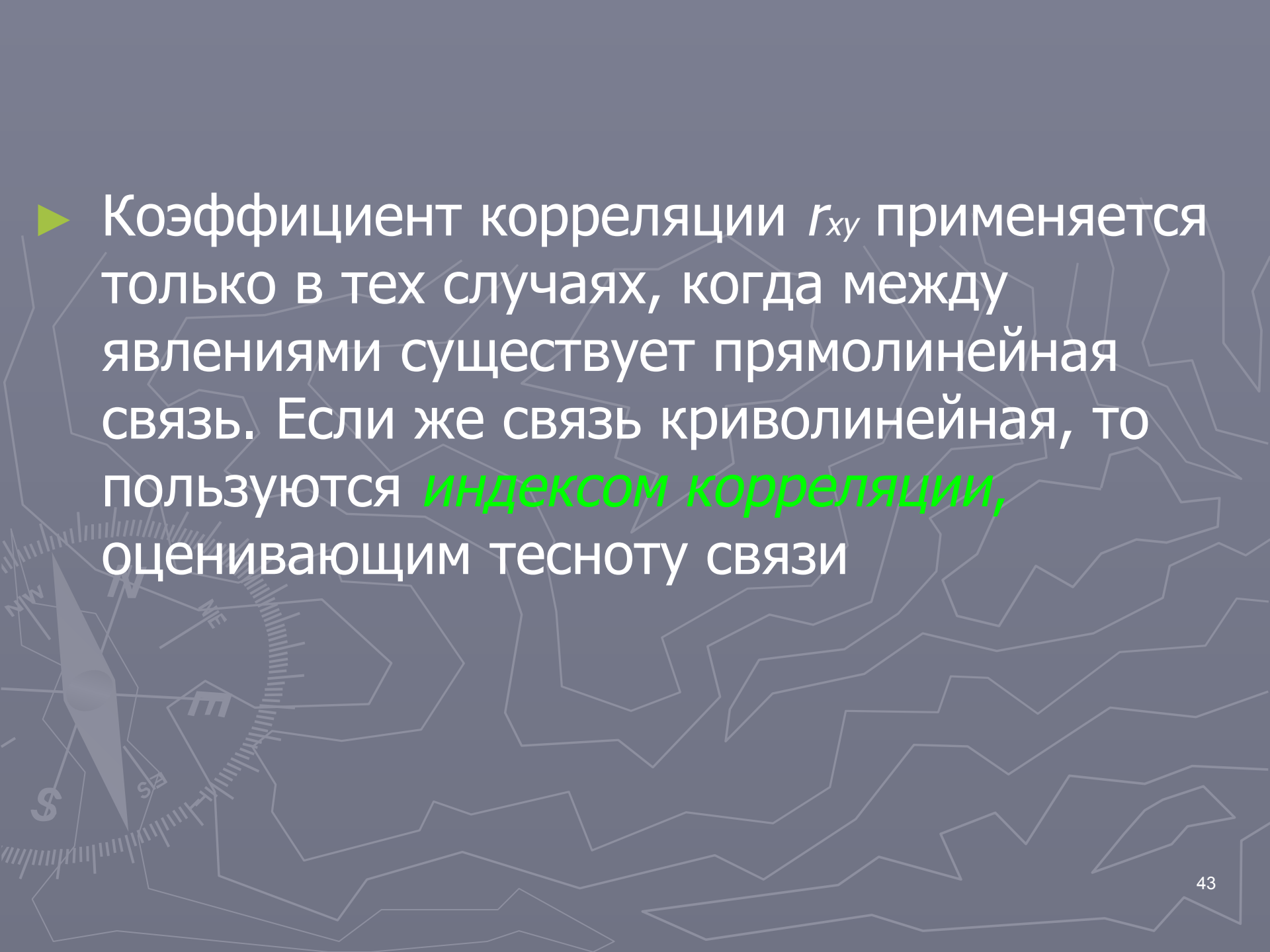
Пределы изменения парного коэффициента корреляции

$$-1 \leq r_{xy} \leq 1.$$

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}}$$

ИЛИ

$$r_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \cdot \sum_i (y_i - \bar{y})^2}}.$$

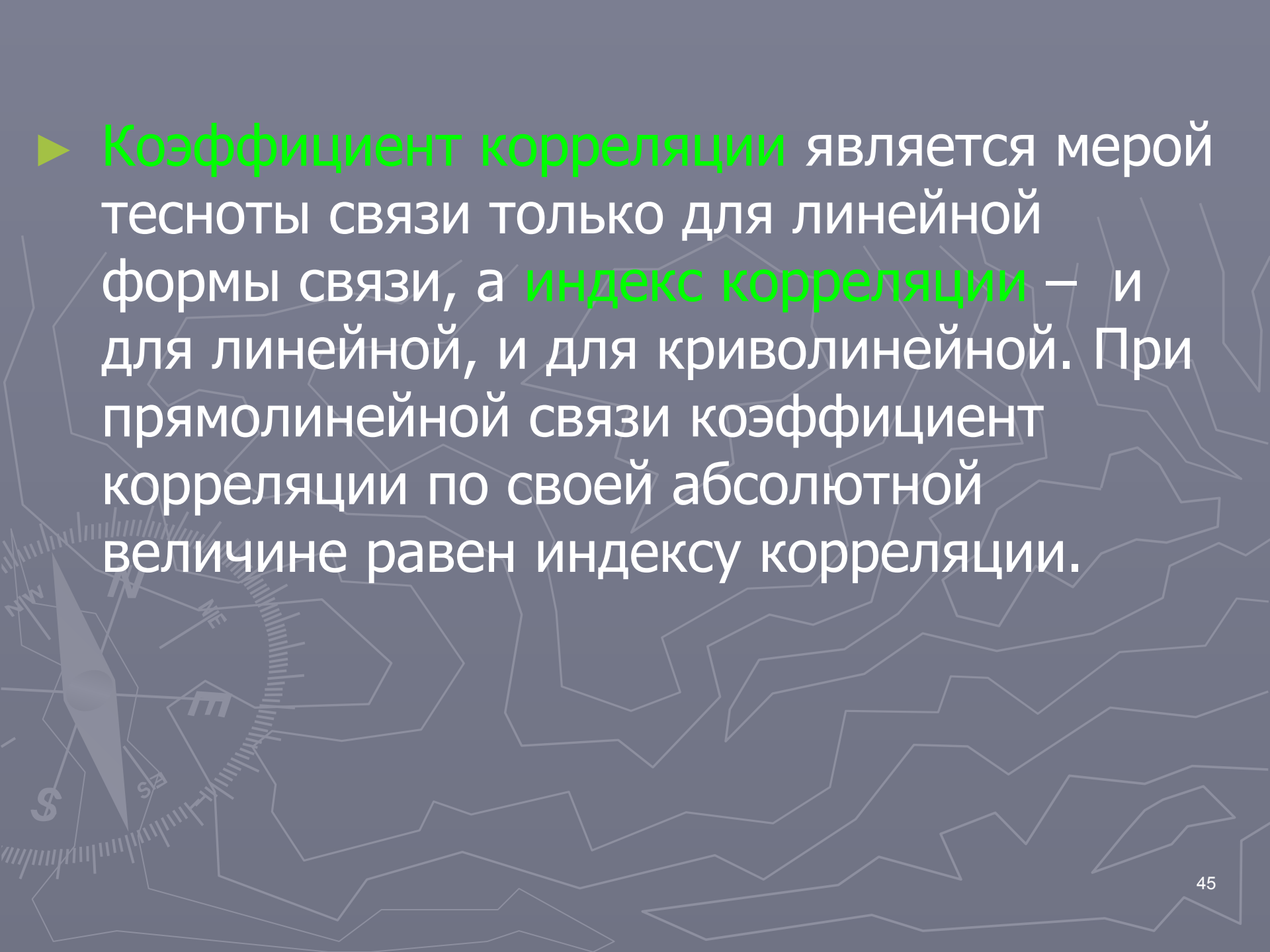
- 
- Коэффициент корреляции r_{xy} применяется только в тех случаях, когда между явлениями существует прямолинейная связь. Если же связь криволинейная, то пользуются *индексом корреляции*, оценивающим тесноту связи

Индекс корреляции

$$R = \sqrt{\frac{\sigma_T^2}{\sigma_{\text{общ}}^2}}$$

где σ_T^2 — теоретическая дисперсия;

$\sigma_{\text{общ}}^2$ — общая дисперсия.

- 
- Коэффициент корреляции является мерой тесноты связи только для линейной формы связи, а индекс корреляции — и для линейной, и для криволинейной. При прямолинейной связи коэффициент корреляции по своей абсолютной величине равен индексу корреляции.

Коэффициент (индекс) детерминации

$$R^2 = \frac{\sigma_T^2}{\sigma_{\text{общ}}^2}.$$

Этот показатель универсален: может использоваться при любом количестве факторных признаков, при любой форме связи. Он показывает, какая часть общей дисперсии результативного признака определяется факторами, включенными в уравнение регрессии.

Множественная корреляция

- ▶ Множественная корреляция занимается изучением, измерении связи между результативным признаком, двумя и более факторными.

► Множественная корреляция определяет:

1) форму связи;

2) тесноту связи;

3) влияние отдельных факторов на общий результат.

- *Определение формы связи* сводится обычно к отысканию уравнения связи y с факторами $x, z, \omega, \dots, v$. Так, линейное уравнение зависимости результативного признака от двух факторных определяется по формуле:

$$Y_{xz} = a_0 + a_1X + a_2Z$$

- Для определения параметров a_0 , a_1 и a_2 по способу наименьших квадратов необходимо решить следующую систему трех нормальных уравнений:

$$\begin{cases} \sum y = na_0 + a_1 \sum x + a_2 \sum z; \\ \sum yx = a_0 \sum x + a_1 \sum x^2 + a_2 \sum xz; \\ \sum yz = a_0 \sum z + a_1 \sum xz + a_2 \sum z^2. \end{cases}$$

Измерение тесноты связи

- ▶ После получения коэффициентов регрессии нужно измерить тесноту связи между факторными и результативным признаками для полученной модели. Измерение тесноты производится на основе вариации результативного признака и **правила сложения дисперсий**.

Правило сложения дисперсий

$$\sigma_{\text{общ}}^2 = \sigma_T^2 + \sigma_{\text{ост}}^2$$



Теоретическая дисперсия

$$\sigma_T^2 = \frac{\sum_{i=1}^n (Y_i^T - \bar{y})^2}{n}.$$

- ▶ теоретическая дисперсия – это вариация теоретического признака вокруг общей средней

Остаточная дисперсия

$$\sigma_{ост}^2 = \frac{\sum_{i=1}^n (y_i - Y_i^T)^2}{n}.$$

- ▶ Остаточная дисперсия – это среднее квадратическое отклонение теоретического признака от фактического.

Коэффициент множественной корреляции

$$R_{y;x,z} = \sqrt{\frac{r_{yx}^2 + r_{yz}^2 - 2r_{yx} \cdot r_{yz} \cdot r_{xz}}{1 - r_{xz}^2}},$$

► где r_{xy} , r_{zy} , r_{xz} — парные коэффициенты корреляции

Частных коэффициентов корреляции

$$r_{xy(z)} = \frac{r_{xy} - r_{zy} \cdot r_{xz}}{\sqrt{(1 - r_{zy}^2)(1 - r_{xz}^2)}}.$$