ДАННЫЕ СТАТИСТИКА АНАЛИТИКА

Виды аналитики

- Описательная
  "Что случилось?"
  Графики, сводные таблицы
- Диагностическая
  "Почему это случилось?"
  Анализ данных
- Прогнозная
  "Что может случиться?"
  Статистика и машинное обучение
- Предписательная
  "Что делать?"
  Статистика и машинное обучение. В отличие от прогнозной носит декларативный характер.

Этапы стат. анализа

1. Описательная статистика - описание полученного в ходе исследования массива данных
2. Аналитическая статистика - анализ данных и проверка различных статистических гипотез

Задачи и статметоды для их решения

Гипотезы

- Гипотеза - это догадка, предположение, недоказанное утверждение.
- На основе экспериментов или данных гипотеза либо ...
  - ... опровергается
    - Тогда это ложное утверждение
  - ... подтверждается
- Открытая проблема - это неопровергнутая и недоказанная гипотеза
- Уровень значимости - это величина для оценки гипотезы. Вероятность отклонить гипотезу, если она на самом деле истинна. Популярные значения: 5%, 1% и 0,1%
  - Большой уровень значимости => Меньше уверенности в H0. Больше уверенности в H1 => Большой риск не отвергнуть ложную гипотезу. Риск ошибки второго рода.
  - Меньший уровень значимости = > Больше уверенности в H0. Меньше уверенности в H1 => Большой риск принять ложную гипотезу. Риск ошибки первого рода.
- Основная/Нулевая гипотеза (H0) - это гипотеза, которой мы придерживаемся, пока наблюдения не заставят признать обратное.
  - Альтернативная гипотеза (H1)
- Ошибки
  - Первого рода - ситуация когда H0 опровергается, хотя она на самом деле верна
  - Второго рода - ситуация когда H0 принимается, хотя он на самом деле неверна.

Случайное событие, случайная величина и вероятность

- Испытание - это любой опыт, при проведении которого задана совокупность условий, неоднозначно определяющих его исход
- Случайное событие - одно из возможных явлений, которое может осуществиться в результате испытания.
  - по другому: событие о наступлении которого мы не можем сказать заранее.
- Вероятность события A - это безразмерная величина от 0 до 1:
  P(A) = M/N, где
  M - благоприятсвующие исходы
  N - всего возможных исходов
  - Пример. Монетка. Вероятность выпадени орла = 1/2 = 0,5.
- Частота события.
  Можно многократно повторить опыт (смоделировать на компе) и вычислить частоту события A - долю "успешных" опытов. Это есть метод Монте-Карло.
  При большом кол-ве повторов опыта Частота ~ Вероятности.
  P(A) = кол-во успехов/кол-во опытов
- Случайная величина — численное выражение случайного события
  - Пространство вариантов — диапазон возможных значений случайной величины.

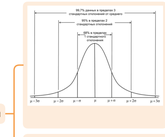
Выборка и генеральная совокупность

- Выборка - это подмножество данных из более крупного набора данных - популяции (или генеральной совокупности)
  - Репрезентативная выборка - имеющая такие же характеристики, как и генеральная совокупность.
  - Смещенная выборка (sample bias)
    Выборка, которая представляет популяцию в искаженном виде.
    - Смещение возникает, когда данные измерений, или наблюдения, систематически ошибочны
  - Зависимые и независимые выборки
    - Если можно установить гомоморфную пару (то есть, когда одному случаю из выборки X соответствует один и только один случай из выборки Y и наоборот) для каждого случая в двух выборках (и это основание взаимосвязи является важным для измеряемого на выборках признака), такие выборки называются зависимыми. Примеры зависимых выборок: пары близнецов, два измерения какого-либо признака до и после экспериментального воздействия, мужья и жёны и т. п. Зависимые выборки всегда имеют одинаковый объём
    - В случае, если такая взаимосвязь между выборками отсутствует, то эти выборки считаются независимыми, например: мужчины и женщины, психологи и математики. Объёмы независимых выборок могут отличаться.
- Отборы
  - Случайный отбор (random sampling) - процесс, в котором каждый доступный член популяции, подвергаемой отбору, имеет равную возможность попасть в выборку при каждой выемке.
    - Такая выборка называется простой случайной выборкой
  - Стратифицированный отбор (stratified sampling)
    Разделение популяции на страты и случайный отбор элементов из каждой страты
  - Возвраты при отборе
    - С возвратом - после каждой выемки наблюдения кладутся назад в популяцию для возможно повторного отбора в будущем.
    - Без возврата - однажды выбранные наблюдения недоступны для будущих выемок.
- Распределение данных (data distribution)
  Частотное распределение индивидуальных значений в наборе данных
- Выборочная статистика (sample statistic) - показатель статистики, вычисленный для для выборки из популяции.
- Выборочное распределение (sampling distribution) - распределение выборочной статистики на большом числе выборок, вынимаемой из одной и той же популяции.
- Центральная предельная теорема (ЦПТ, (central limit theorem) - тенденция выборочного распределения принимать нормальную форму по мере увеличения размера выборки, даже если исходная популяция не является нормально распределенной.
  - По другому: средние значения, вынутые из многочисленных выборок, будут принимать нормальную форму, при условии, что размер выборок достаточно крупный.
- Выборочное распределение статистики
  - Стандартная ошибка (standard error, se). Стандартное отклонение выборочной статистики на многочисленных выборках.
    - se = s / sqrt(n), где
      s - стандартное отклонение значений выборки
      n - размер выборки
    - По мере увеличения размера выборки se уменьшается. Для сокращения стандартной ошибки в 2 раза, размер выборки должен быть увеличен в 4 раза.
    - se показывает вариабельность выборочного метрического показателя. Не путать со стандартным отклонением, которое показывает вариабельность отдельных точек данных.
  - Доверительный интервал
    - Еще один способ понять потенциальную ошибку в оценке выборки
    - ДИ показывает насколько параметры из выборки могут отличаться от реально существующих данных в ГС. Насколько мы ошибаемся при формировании той или иной выборки, мы закладываем в так называемую ошибку репрезентативности, в ошибку средней и вокруг нее собственно и строим доверительный интервал.
    - ДИ задается самостоятельно исследователем. Чаще всего он равен 95 %.
      - Это значат, что с 95%-ной вероятностью в нашем интервале лежит реальное значение, и лишь в 5 % случаев мы ошибаемся.
  - Бутстрап
    - Один из простых и эффективных способов оценки выборочного распределения статистики состоит в том, чтобы вынимать дополнительные выборки с возвратом из самой выборки и повторно вычислять статистику для каждой повторной выборки.
- Статистический анализ - использование различных методов для того, чтобы определить свойства генеральной совокупности по выборке.
  - Весь статистический анализ строится на основе предположений о свойствах ГС по некоторой выборки из этой ГС. Так как мы не можем взять всю ГС и оценить ее параметр (напр., средний рост), то мы берем случайную выборку из ГС и оцениваем параметр выборки и делаем предположения о том, как параметр может быть устроен в ГС.

Подготовка данных

- Нормализация.
  Это преобразование данных к неким безразмерным единицам.
  - Цели:
    - приведение различных данных в самых разных единицах измерения и диапазонах значений к единому виду, который позволит сравнивать их между собой или использовать для расчёта схожести объектов
    - ускорение получения желаемого результата за счет того, что машине приходится обрабатывать меньший диапазон данных
  - Аналитически любая нормализация сводится к формуле
    Хнорм = (Xi - Хсмещ)/Хед,
    где Xi — текущее значение,
    Хсмещ — величина смещения значений,
    Хед — величина интервала, который будет преобразован к "единице"
    По сути всё сводится к тому, что исходный набор значений сперва смещается, а потом масштабируется.
  - Методы нормализации числовых данных
    - sklearn
      - sklearn.preprocessing.normalize()
        Получим данные от 0 до 1
      - sklearn.preprocessing.MinMaxScaler()
        Есть возможность настроить разные параметры
        - устанавливает наименьшее наблюдаемое значение равным 0, а наибольшее наблюдаемое значение — 1
        - $z = \frac{x - \min(x)}{\max(x) - \min(x)}$
    - pandas
      - Использование максимального абсолютного масштабирования
        Все значения в столбце делим на макс. число по модулю.
        df[column] = df[column] / df[column].abs().max()
        Таким образом все данные нормализуются между 0 и 1.
      - z-стандартизация.
        Вычесть из данных среднее значение и поделить на стандартное отклонение.
        (X - X.mean)/X.std
        Получим новый со средним значением равным 0 и стандартным отклонением равным 1.
        - Изменяет форму распределения данных (приводится к нормальному распределению).
- Обработка пропусков
  - Удаление строк/столбцов с большим количеством пропусков
  - Восстановление данных
    - Подстановка среднего, медианы или моды
    - Взять значение из наиболее близкой строки (используя меры близости объектов)
      - При этом нужно отнормировать данные
    - С помощью коэф. корреляции
- Обработка дубликатов
- Обнаружение выбросов (аномальных значений)
  - Удаление
- Кодирование данных
  - Многие алгоритмы ML ожидают числовые входные данные, поэтому категориальные данные нужно представить в виде чисел.
- Качество
  - Отсутствие дублей и противоречий
  - Отсутствие пропущенных значение
  - Адекватность

Виды распределения

- Нормальные (гауссовое) распределения
  - 
    - 68% данных находятся в пределах одного стандартного отклонения от среднего и 95% — в двух стандартных отклонениях.
  - Стандартное нормальное распределение (z-распределение) - это нормальное распределение со средним = 0, и стандартным отклонением = 1.
    - Процедура нормализация (или стандартизации) .
      Вычесть из данных среднее значение и разделить на стандартное отклонение.
      - z-Оценка (z-score)
        Результат стандартизации отдельной точки данных.
- Длинноxвостые распределения
  - Это распределения с длинными хвостами - длинная узкая часть частотного распределения, где относительно предельные значения встречаются с низкой частотой.
    - Такие распределения могут быть ассиметричными (skew) - когда один хвост длиннее другого.
  - Ключевая идея: Принятие нормального распределения может привести к недооценке предельных событий (" черных лебедей").
    - Нассим Талеб. Теорию черного лебедя - аномальные события, такие как обвал фондового рынка, будут возникать с намного большей вероятностью, чем их предсказание нормальным распределением.
- t-Распределение Стьюдента
  - это распределение нормальной формы, но немного толще и длиннее в хвостах
- Биномиальное распределение
- Распределение Пуассона

Анализ данных. Инструменты оценки

Виды данных

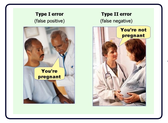
- Структурированные (Данные в табличной форме)
  - Количественные (quantative)
    - Непрерывные значения (continuous)
      - Рост, вес, продажи, время
    - Дискретные данные (discrete). Могут принимать только целочисленные значения.
      - например количество возникновений события
  - Категориальные/номинальные/качественные (categorical/nominal/qualitive).
    Только определенный набор значений, в частности набор возможных категорий. Не имеют количественного отношения друг к другу.
    - Местонахождение: Америка, Европа, Азия.
    - Классификация сотрудников: руководитель, менеджер и пр.
    - Двоичные/бинарные/дихотомические. Особый случай категориальных данных всего с двумя категориями значений (0/1, истина/ложь).
  - Порядковые. Для ранжирования данных по отношению друг к другу
    - Размеры футболок: S, M, L, XL
    - Оценки: плохо, хорошо, отлично
- Неструкутрированные
  - Данные в произвольной форме (картинка, статья и пр.)

Структура данных

- Data frame (по сути электронная таблица)
  - Столбец (признак, feature, атрибут, вход, предиктор, переменная)
  - Столбец Исход (outcome, зависимая переменная, отклик, цель, выход)
  - Строки (записи, records, экземпляр, наблюдение, шаблон, паттерн)
- Временно́й (динамический) ряд (последовательные данные измерений одной и той же переменной)
- Графовые (или сетевые). Пример, граф соцсети может представлять связи между людьми в сети.

Тестирование гипотез

Гипотезы

- Уровень значимости - это величина для оценки гипотезы
  P-значение (P-value)
  - Вероятность отклонить гипотезу, если она на самом деле истинна. Популярные значения: 5%, 1% и 0,1%
    - В большинстве случаев это 0.05, и если p-value меньше данного заветного значения, считается, что тогда разница является статистически значимой.
  - Вероятность, что результат является случайным
  - Большой уровень значимости => Меньше уверенности в H0. Больше уверенности в H1 => Большой риск не отвергнуть ложную гипотезу. Риск ошибки второго рода.
  - Меньший уровень значимости = > Больше уверенности в H0. Меньше уверенности в H1 => Большой риск принять ложную гипотезу. Риск ошибки первого рода.
- Основная/Нулевая гипотеза (H0)
  - это гипотеза, которой мы придерживаемся, пока наблюдения не заставят признать обратное.
  - это гипотеза, которую каждый исследователь хочет отвергнуть, и принять альтернативную.
  - разница не является статистически значимой, т. е. наблюдаемые данные могут происходят из ожидаемого распределения.
- Альтернативная гипотеза (H1)
  - разница является статистически значимой, т. е. наблюдаемые данные не могут происходят из ожидаемого распределения.
- 

  - Ошибка 1-го рода (type I error) Ошибочный вывод о том, что эффект — реальный (в то время, как он случайный)
  - Ошибка 2-го рода (type 2 error) Ошибочный вывод о том, что эффект — случайный (в то время, как он реальный)

Идея проверок гипотез

- Цель: проверить является ли наблюдаемый эффект случайностью или нет.
- Принимаем нулевую гипотезу H0 — всему виной является случайность.
- Альтернативная гипотеза H1 — Обратная нулевой (то, что надеемся доказать)
  - Взятые вместе нулевая и альтернативная гипотезы охватывают абсолютно все имеющиеся возможности.
- Находим вероятность того, что найденные различия были случайными (p-уровень) — p-уровень значимости - вероятность получить различия при условии, что в ГС различий нет
- Если p-значение < уровня значимости, то отвергаем H0.
  Если p-значение > уровня значимости, то оснований для отклонения H0 нет. — уровень значимости (альфа) = 0,05 - это самая распространенная оценка

"Хитрости"

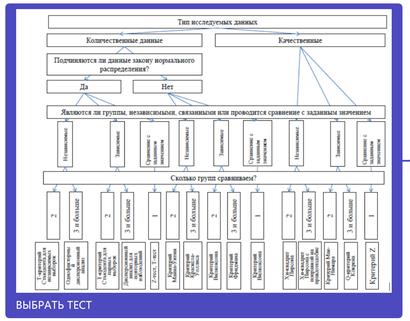
Можно воспользоваться тестами и критериями, работающими с нормальным распределение, если:

- Нет выбросов в данных (строгое правило)
- Нет явной асимметрии плотности распределения (иногда можно исправить с помощью логарифмирования или преобразования Бокса-Кокса)
- Нет сильного отклонения от колоколообразной формы. Для бимодального распределния - попробовать разделить на выборки.

Алгоритм

1. Формулируем нулевую гипотезу — Что изучаем и какие есть предположения относительно результата?
2. Определяем тип данных — Какие данные имеются?
3. Соотносим гипотезу и тип данных с критериями проверки — Какой критерий выбрать?
4. Подставляем данные в формулы — Какие результаты получили?
5. Отклоняем или не отклоняем нулевую гипотезу — Какие выводы можно сделать?

ВЫБРАТЬ ТЕСТ



A/B тестирование

- Это эксперимент с двумя группами для определения того, какой из двух сравниваемых объектов лучше.
- Требования
  - В некоторых случаях - Единовременность (для исключения влияния случайных факторов)
  - Однородность выборок
  - Достаточный объем выборки (исходя из стандартной ошибки)
  - Независимость выборок
- Разбиение на группы
  - Контрольная
  - Тестовая
- Проверочная статистика — метрический показатель, который используется для сравнения группы A с группой B (как правило, двоичная переменная: нажатие/отсутствие нажатия, купить/воздержаться от покупки, мошенничество/не мошенничество и т. д.)
- A/B-тест, как правило, конструируется с учетом гипотезы. Например, гипотеза может заключаться в том, что цена B приносит более высокую прибыль.
- Статистическая проверка гипотез является дальнейшим анализом A/B-теста с целью установить, является ли случайная возможность разумным объяснением наблюдаемой разницы между группами A и B.

ВЫБРАТЬ ТЕСТ:

Одновыборочные

- Критерий Шапиро-Уилка
  - Проверка на принадлежность выборки нормальной генеральной совокупности
- Нормальное распределение
  - Одновыборочный t-test (равенство средних)
    H0: μ = μ0
    H1: μ != μ0
    - Есть мнение, что требование нормальности данных относится только для выборок меньше 30.
      При больших выборках нормальность необязательна.
      Нормальным должно быть распределение средних (проверка методом бутстрапа)!
- Ненормальное распределение
  - критерий Уилкоксона
- Категориальные
  - биномиальный тест
  - Критерий Хи-квадрат Пирсона (принадлежность выборки теоретическому закону распределения)

Двухвыборочные

- Независимые
  - Нормальное распределение
    - Двухвыборочный t-test для независимых выборок (равенство средних)
      H0: μ1 = μ2
      H1: μ1 != μ2
      - Есть мнение, что требование нормальности данных относится только для выборок меньше 30.
        При больших выборках нормальность необязательна.
        Нормальным должно быть распределение средних (проверка методом бутстрапа)!
    - F-тест Фишера (равенство дисперсий)
  - Ненормальное распределение
    - Ранговый U-критерий Манна-Уитни
      Для оценки различий двух выборок по ранжируемому признаку
    - Критерий Флингера-Килина
  - Категориальные
    - Точный критерий Фишера
      Для таблиц сопряженности
      Для выборок маленьких размеров
      (насколько случайно распределены значения между группами)
    - G-test (LL-score)
    - Критерий Хи-квадрат Пирсона
      Для таблиц сопряженности
      (насколько случайно распределены значения между группами)
- Зависимые
  - Нормальное распределение
    - Парный t-test (равенство средних)
      H0: μ1 = μ2
      H1: μ1 != μ2
      - Есть мнение, что требование нормальности данных относится только для выборок меньше 30.
        При больших выборках нормальность необязательна.
        Нормальным должно быть распределение средних (проверка методом бутстрапа)!
  - Ненормальное распределение
    - Критерий Уилкоксона
      H0: выборки однородны
      H1: выборки неоднородны
  - Категориальные
    - критерий Мак-Немара

Многовыборочные

- Дисперсионный анализ ANOVA (analysis of variance)

Анализ данных. Инструменты оценки

Описательная статистика. Ключевая идея в том, что первый и самый важный шаг - посмотреть на данные. Одномерный анализ (одна переменная за раз)

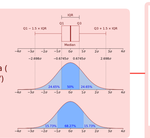
Числовые данные

Оценки центрального положения (где расположено большинство данных, центральная тенденция, меры среднего)

- Среднее (mean)
  Сумма всех значений, деленная на количество значений.
  Синоним: среднее арифметическое.
- Среднее взвешенное (weighted mean)
  Сумма произведений всех значений на их веса, деленная на сумму весов.
  Синоним: среднее арифметическое взвешенное.
- Медиана (median)
  Такое значение, при котором половина сортированных данных находится выше и ниже данного значения.
  Синоним: 50-й процентиль.
- Медиана взвешенная (weighted median)
  Такое значение, при котором половина суммы весов находится выше и ниже сортированных данных.
- Среднее усеченное (trimmed mean)
  Среднее число всех значений после отбрасывания фиксированного числа предельных значений.
  Синоним: обрезанное среднее.
- Выброс — это любое значение, которое сильно удалено от других значений в наборе данных.
- Робастный, устойчивый (robust)
  Не чувствительный к предельным значениям. Медианные характеристики.
- Мода - наиболее часто встречающееся значение.

Визуализация данных

- Коробчатая диаграмма (boxplot, "ящик с усами")



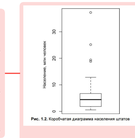
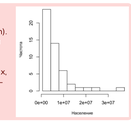
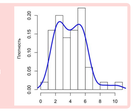


Рис. 1.3. Коробчати диаграма населения штатов

- Гистограмма (histogram). Диапазон переменной делятся на равные сегменты и откладываются на оси x, а количество данных — на оси y



- Частотная таблица. Табличная версия гистограммы.
- График плотности (density plot)
  Сглаженная версия гистограммы, часто на основе ядерной оценки плотности.



Оценки вариабельности, дисперсии, меры разброса

- Отклонения (deviations)
  Разница между наблюдаемыми значениями и оценкой центрального положения.
  Синонимы: ошибки, остатки
- Дисперсия (variance)
  Сумма квадратических отклонений от среднего, деленная на n − 1, где n — число значений данных.
  Синонимы: среднеквадратическое отклонение, среднеквадратическая ошибка.

$$\frac{\sum (x - \bar{x})^2}{n-1}$$

- Среднее абсолютное отклонение (mean absolute deviation)
  Среднее абсолютных значений отклонений от среднего.
  Синонимы: манхэттенская норма.

$$\frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

- Стандартное отклонение (standard deviation)
  Квадратный корень из дисперсии.
- Размах (range) -
  разница между самым большим и самым малым числом
- Порядковые статистики (order statistics)
  Метрические показатели на основе значений данных, отсортированных от самых малых до самых больших.
  Синоним: ранг.
- Процентиль (percentile)
  Такое значение, что P процентов значений принимает данное значение или меньшее и (100 − P) процентов значений принимает данное значение или большее.
  Синоним: квантиль.
  Медиана = 50-й процентиль.
  Квантиль 0,8 = 80-й процентиль.

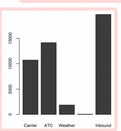
- Межквартильный размах, МКР (interquartile range, IQR)
  Разница между 75-м и 25-м процентилями.

Категориальные данные

- Мода (mode)
  Наиболее часто встречающаяся категория в наборе данных.
- Математическое ожидание, ожидаемое значение (expected value)
  Это сумма значений, умноженных на вероятность их возникновения. По сути является формой среднего взвешенного: оно привносит понятие будущих ожиданий и весов вероятности, которые зачастую основаны на субъективном суждении.
  1. Умножить каждый исход на вероятность его наступления.
  2. Просуммировать эти значения.
  Пример. Маркетолог новой "облачной" технологии, например, продает два уровня веб-служб: один по цене 300 долларов в месяц, а другой по цене 50 долларов в месяц. Фирма полагает, что 5% посетителей подпишутся на веб-службы за 300 долларов, 15% на веб-службы за 50 долларов и 80% не подпишутся совсем.
  EV = 0,05*300+0,15*50+0,80*0 = 22,5.
  Таким образом, матожидание = 22,5 доллара.

Визуализация

- Столбчатые диаграммы (bar charts)
  Категории - на оси x, частоты (доли) - на оси y



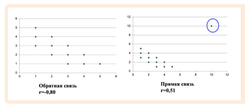
- Круговые диаграммы (pie charts).
  Альтернативой столбчатым диаграммам, визуально менее информативные

Двумерный анализ (две переменные за раз)

Корреляционный анализ.

X и Y коррелируют положительно = рост X сопровождается ростом Y.
X и Y коррелируют отрицательно = рост X сопровождается уменьшением Y.

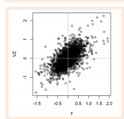
- Коэффициент корреляции (correlation coefficient). Измеряет степень линейной взаимосвязи друг с другом (от −1 до +1) и его направление (прямая или обратная)



  - Коэф. корреляции очень чувствителен к выбросам
  - Работает только с непрерывными данными
  - могут быть проблемы при ненормальном распределении данных
  - Проблемы
  - Корреляция - это просто число
  - Пирсена
    Спирмена
    Кенделла
- Корреляционная матрица (correlation matrix)
  Таблица, в которой строки и столбцы — это переменные, и значения ячеек — корреляции между этими переменными.

	T	CT_L	PTR	VZ	LVL^2
T	1.000	0.475	0.328	0.878	0.279
CT_L	0.475	1.000	0.400	0.417	0.287
PTR	0.328	0.400	1.000	0.287	0.280
VZ	0.878	0.417	0.287	1.000	0.242
LVL^2	0.279	0.287	0.280	0.242	1.000

- Диаграмма рассеяния (scatterplot)
  График, в котором ось x является значением одной переменной, а ось y — значением другой.



- Корреляция не подразумевает причинно-следственных связей
- ссылка

Регрессионный анализ – инструмент для количественного предсказания значения одной переменной на основании другой

- По кол-ву независимых переменных
  - Простой - между двумя переменными
  - Множественный. Между зависимой переменной Y и несколькими независимыми X
- По типу зависиммости
  - Линейный
  - Нелинейный
- Общий подход к решению:
  1.Определение формы зависимости (напр., y = ax + b)
  2.Построение модели регрессии (подобрать наилучшие параметры a и b, которые макс. точно описывают нашу модель)
  3.Оценка неизвестных значений зависимой переменной
- ссылка

Корреляция VS Регрессия

- РЕГРЕССИЯ – предсказание одной переменной на основании другой. Одна переменная – независимая, а другая – зависимая.
  Пример: чем больше студент занимается перед экзаменом, тем выше его оценка
- КОРРЕЛЯЦИЯ показывает, в какой степени две переменные СОВМЕСТНО ИЗМЕНЯЮТСЯ. Нет зависимой и независимой переменных, они эквивалентны.
  Пример: рост человека положительно связан с его массой

Многомерный анализ (более двух переменных за раз)

Числовые-Числовые

- Шестиугольная сетка и контуры

Две категориальных переменных

- таблица сопряженности

Категориальные и числовые

- Коробчатые диаграммы

Временные ряды

Цели анализа

- выявление закономерности изменения изучаемого явления во времени
- прогнозирование (экстраполирование) полученных данных на последующие периоды

Методика аналитики предусматривает выполнение последовательных действий:

- Представить полученные данные графически и выявить форму зависимости изучаемого явления от времени.
- Оценить наличие и силу корреляции изучаемого явления от времени. В зависимости от выраженности изменений явления во времени (определяется по коэффициенту корреляции между временем и изучаемым явлением).
  - а) с выраженной тенденцией (r =0,7 — 1,0);
  - б) с неустойчивой тенденцией (r =0,3 — 0, 69);
  - в) с отсутствием тенденции (r = 0 — 0,29).
- Если установлено, что ряд обладает выраженной тенденцией, проводят анализ компонентов динамики ряда: основной тенденции (эволюции, тренда), кратковременных систематических движений и случайных колебаний. Основная задача анализа — разделить эти компоненты и выявить основную закономерность изменения явления во времени. Для выявления и описания тренда динамический ряд подвергают обработке — выравниванию.