**Контрольные вопросы по регрессионному анализу**

**1. Виды зависимостей между двумя СВ**.

Между двумя случайными величинами (СВ) может существовать зависимость нескольких типов:

* Функциональная зависимость - это такая зависимость, при которой каждому значению одной СВ соответствует единственное значение другой СВ. Например, зависимость между площадью круга и его радиусом является функциональной зависимостью.
* Статистическая зависимость - это такая зависимость, при которой изменение одной СВ сопровождается изменением другой СВ с некоторой вероятностью. Например, зависимость между ростом человека и его весом является статистической зависимостью.
* Независимость - это такая зависимость, при которой изменение одной СВ никак не влияет на изменение другой СВ.

**2. В чем различие между статистической и функциональной зависимостями двух СВ?**

Основное различие между статистической и функциональной зависимостями заключается в том, что при функциональной зависимости каждому значению одной СВ соответствует единственное значение другой СВ, а при статистической зависимости такому значению может соответствовать несколько значений или ни одного значения.

Например, рассмотрим зависимость между ростом человека и его весом. Рост человека является функциональной зависимостью от его веса, поскольку каждому значению веса соответствует единственное значение роста. Например, если рост человека 180 см, то его вес должен быть не менее 70 кг и не более 120 кг.

Рассмотрим также зависимость между размером обуви человека и его возрастом. Эта зависимость является статистической, поскольку одному и тому же возрасту могут соответствовать несколько значений размера обуви. Например, возраст 20 лет может соответствовать размеру обуви 42, 43 или 44.

**3. Что такое регрессионная зависимость между двумя СВ?**

Регрессионная зависимость - это статистическая зависимость между двумя СВ, описываемая математической моделью. Регрессионная модель позволяет предсказать значение одной СВ по значению другой СВ.

**4. Основные задачи корреляционного анализа.**

Основные задачи корреляционного анализа:

* Оценка тесноты связи между двумя СВ. Для этого используется коэффициент корреляции.
* Проверка статистической значимости связи между двумя СВ. Для этого используется критерий Пирсона.
* Определение формы связи между двумя СВ. Для этого используются различные графические методы.

**5. Основные задачи регрессионного анализа.**

Основные задачи регрессионного анализа:

* Оценка параметров регрессионной модели. Для этого используются различные методы, например, метод наименьших квадратов.
* Проверка значимости параметров регрессионной модели. Для этого используются различные критерии, например, критерий Фишера.
* Предсказание значения одной СВ по значению другой СВ.

**6. На основании чего осуществляется выбор вида функции регрессии?**

Выбор вида функции регрессии осуществляется на основании следующих факторов:

* Форма зависимости между двумя СВ. Если связь между двумя СВ является линейной, то целесообразно использовать линейную регрессионную модель. Если связь между двумя СВ является нелинейной, то целесообразно использовать нелинейную регрессионную модель.
* Наличие данных. Если количество данных ограничено, то целесообразно использовать простую регрессионную модель. Если количество данных достаточно велико, то целесообразно использовать более сложную регрессионную модель.
* Цели исследования. Если цель исследования заключается в прогнозировании значений одной СВ, то целесообразно использовать линейную регрессионную модель. Если цель исследования заключается в изучении формы связи между двумя СВ, то целесообразно использовать нелинейную регрессионную модель.

**7. Что называется корреляционным полем?**

Корреляционное поле - это графическое изображение зависимости между двумя СВ. Корреляционное поле строится на основе выборочных данных.

**8. Почему наиболее часто используется модель линейной регрессии?**

Модель линейной регрессии используется наиболее часто по следующим причинам:

* Линейная регрессия является простой и легко интерпретируемой моделью.
* Метод наименьших квадратов, используемый для оценки параметров линейной регрессии, является эффективным и устойчивым.
* Линейная регрессия имеет широкое применение в различных областях науки и практики.

**9. Какой статистический показатель используется в качестве количественной меры линейной связи между двумя наблюдаемыми величинами?**

В качестве количественной меры линейной связи между двумя наблюдаемыми величинами используется коэффициент корреляции.

**10. Свойства выборочного коэффициента корреляции:**

* Значение выборочного коэффициента корреляции всегда лежит в пределах от -1 до 1.

Это свойство объясняется тем, что коэффициент корреляции измеряет степень линейной связи между двумя переменными. Если связь между переменными является идеальной линейной, то коэффициент корреляции равен 1 или -1. Если связь между переменными отсутствует, то коэффициент корреляции равен 0.

* Если выборочный коэффициент корреляции равен 0, то между двумя переменными нет линейной связи.

Это свойство является прямым следствием предыдущего. Если коэффициент корреляции равен 0, то связь между переменными не является линейной.

* Если выборочный коэффициент корреляции равен 1 или -1, то между двумя переменными существует идеальная линейная связь.

Это свойство также является прямым следствием предыдущего. Если коэффициент корреляции равен 1 или -1, то связь между переменными является идеальной линейной.

* Значение выборочного коэффициента корреляции не зависит от масштаба переменных.

Это свойство объясняется тем, что коэффициент корреляции измеряет только направление и силу линейной связи между переменными. Масштаб переменных влияет только на значение коэффициентов регрессии, но не на значение коэффициента корреляции.

* Значение выборочного коэффициента корреляции не зависит от порядка переменных.

Это свойство объясняется тем, что коэффициент корреляции измеряет только направление и силу линейной связи между переменными. Порядок переменных влияет только на знак коэффициента корреляции, но не на его величину.

**11. Основные этапы регрессионного анализа:**

1. Формулирование задачи исследования.

На этом этапе необходимо сформулировать цель исследования и определить, какие данные необходимы для ее достижения.

1. Подготовка данных.

На этом этапе необходимо проверить данные на наличие ошибок и пропусков. Также необходимо нормализовать данные, если это необходимо.

1. Выбор вида регрессионной модели.

На этом этапе необходимо выбрать вид регрессионной модели, который наилучшим образом описывает взаимосвязь между переменными.

1. Оценка параметров регрессионной модели.

На этом этапе необходимо оценить параметры регрессионной модели с помощью одного из методов оценки параметров регрессионной модели.

1. Проверка значимости параметров регрессионной модели.

На этом этапе необходимо проверить значимость параметров регрессионной модели с помощью одного из методов проверки значимости параметров регрессионной модели.

1. Оценка точности модели.

На этом этапе необходимо оценить точность модели с помощью одного из методов оценки точности регрессионной модели.

1. Прогнозирование.

На этом этапе необходимо использовать регрессионную модель для прогнозирования значений зависимой переменной по значениям независимых переменных.

**12. Практическая задача регрессионного анализа:**

Рассмотрим следующую практическую задачу регрессионного анализа:

* Задача: предсказать стоимость автомобиля по его возрасту и пробегу.
* Данные: выборка из 100 автомобилей, включающая в себя информацию о возрасте, пробеге и стоимости автомобиля.
* Вид модели: линейная модель.

Решение:

1. Формулирование задачи исследования:

Целью исследования является предсказание стоимости автомобиля по его возрасту и пробегу.

1. Подготовка данных:

Данные были предварительно проверены на наличие ошибок и пропусков.

1. Выбор вида регрессионной модели:

На основании анализа данных было выбрано линейную модель.

1. Оценка параметров регрессионной модели:

Параметры регрессионной модели были оценены методом наименьших квадратов.

1. Проверка значимости параметров регрессионной модели:

Значимость параметров регрессионной модели была проверена с помощью критерия Фишера.

1. Оценка точности модели:

Точность модели была оценена с помощью коэффициента детерминации.

1. Прогнозирование:

На основе полученной модели можно прогнозировать стоимость автомобиля по его возрасту и пробегу.

Пример вывода:

Согласно результатам исследования, стоимость автомобиля можно предсказать по следующей формуле:

Стоимость = 10000 \* Возраст + 5000 \* Пробег

Значение коэффициента детерминации составляет 0,8, что означает, что 80% вариации стоимости автомобиля объясняются его возрастом и пробегом.

На основании полученной модели можно сделать вывод, что стоимость автомобиля увеличивается с возрастом и пробегом

**13. Какие значения принимает выборочный коэффициент корреляции, если наблюдаемые величины независимы?**

Если наблюдаемые величины независимы, то выборочный коэффициент корреляции принимает значение 0. Это объясняется тем, что независимые величины не имеют никакой линейной связи между собой, поэтому коэффициент корреляции, измеряющий степень линейной связи, должен принимать значение 0.

**14. Что показывает знак выборочного коэффициента корреляции?**

Знак выборочного коэффициента корреляции показывает направление связи между наблюдаемыми величинами. Если знак коэффициента корреляции положительный, то связь между величинами положительная, то есть при увеличении одной величины увеличивается и другая. Если знак коэффициента корреляции отрицательный, то связь между величинами отрицательная, то есть при увеличении одной величины уменьшается другая.

**15. Для чего проводится проверка значимости коэффициента корреляции?**

Проверка значимости коэффициента корреляции проводится для того, чтобы определить, является ли связь между наблюдаемыми величинами статистически значимой. Если связь между величинами не является статистически значимой, то она может быть вызвана случайными факторами, а не общей закономерностью.

**16. Как проводится проверка значимости коэффициента корреляции в случае, если наблюдаемые величины имеют совместное нормальное распределение?**

В случае, если наблюдаемые величины имеют совместное нормальное распределение, проверка значимости коэффициента корреляции проводится с помощью критерия Пирсона. Критерий Пирсона позволяет определить, с какой вероятностью связь между величинами могла возникнуть случайно.

**17. В чем суть метода наименьших квадратов?**

Метод наименьших квадратов является методом оценки параметров регрессионной модели, который минимизирует сумму квадратов отклонений фактических значений зависимой переменной от значений, предсказанных регрессионной моделью.

**18. Система нормальных уравнений метода наименьших квадратов.**

Система нормальных уравнений метода наименьших квадратов представляет собой систему линейных уравнений, которые необходимо решить для получения оценки параметров регрессионной модели. Система нормальных уравнений имеет следующий вид:

*X^TXb = X^Ty*

где:

* X - матрица независимых переменных;
* y - вектор зависимой переменной;
* b - вектор параметров регрессионной модели.

Решение системы нормальных уравнений метода наименьших квадратов можно найти с помощью различных методов, например, методом Гаусса-Жордана или методом обратной матрицы.

**19. Для чего проводится проверка значимости параметров регрессионной модели?**

Проверка значимости параметров регрессионной модели проводится для того, чтобы определить, являются ли полученные оценки параметров регрессионной модели статистически значимыми. Если оценка параметра регрессионной модели не является статистически значимой, то она может быть вызвана случайными факторами, а не общей закономерностью.

Проверка значимости параметров регрессионной модели проводится с помощью различных критериев, например, критерия Стьюдента или критерия Фишера.

**20. Какие существуют методы оценки точности регрессионной модели?**

Существует множество методов оценки точности регрессионной модели. Наиболее распространенными методами являются следующие:

* Коэффициент детерминации - это статистическая характеристика, которая показывает, какая часть вариации зависимой переменной объясняется регрессионной моделью.
* Стандартная ошибка оценки - это статистическая характеристика, которая показывает, насколько близко фактическое значение зависимой переменной может находиться к предсказанному значению регрессионной модели.
* Критерий t - это статистический критерий, который используется для проверки гипотезы о равенстве нулю параметра регрессионной модели.
* Критерий F - это статистический критерий, который используется для проверки гипотезы о незначимости всех параметров регрессионной модели.

Выбор метода оценки точности регрессионной модели зависит от конкретной задачи и целей исследования.