



СРС

**Применение ППП SPSS. Statistica и SAS для
статистических анализов медико-биологических
данных**

Специальность: стоматология

Дисциплина: Биологическая
статистика

Кафедра: Общественное
здравоохранение и информатики

Курс: 2, 201 группа

Выполнила: Сакенова М.Е.

Проверил: Базарбек Женисбек
Базарбекулы

План

Введение

1. Теоретические сведения

1.1 Статистика. Виды статистического анализа

1.2 Статистический пакет STATISTICA

2. Статистический анализ

экономических данных в STATISTICA

2.1 Корреляционно-регрессионный анализ в STATISTICA

2.2 Кластерный анализ в STATISTICA

Список литературы

Введение

Для решения многих других задач в области социологических и маркетинговых исследований, прогнозирования и управления качеством необходимы знания математической и экономической статистики. Принятие любого технического, финансового и производственного решения немыслимо без статистического анализа информации; выделять закономерности из случайностей, сравнивать вероятные альтернативы выбора, строить прогнозы развития процессов, обнаруживать связи и различия множества объектов возможно только и исключительно средствами математической статистики.

Статистика позволяет адекватно оценивать складывающуюся ситуацию и выявлять тенденции, принимать оперативные и стратегические решения

Статистика. Виды статистического анализа

Статистика – наука о сборе, измерении и анализе массовых количественных данных. Статистические данные всегда являются приближенными, усредненными. Поэтому они носят оценочный характер и для достоверности результатов необходимо большое число исходных данных.

Существует несколько видов статистического анализа данных: корреляционный, регрессионный, дисперсионный, факторный, кластерный и др. Рассмотрим некоторые из них.

Корреляционный анализ

Иногда корреляцию и регрессию рассматривают как совокупный процесс статистического исследования. **Корреляционно-регрессионный** анализ является одним из значимых методов построения математических моделей в экономике и считается одним из главных методов в маркетинге.

Корреляция в широком смысле слова означает связь между объективно существующими явлениями.

Корреляционный анализ – вид статистического анализа, который состоит в количественной оценке **силы** и **направления связи** между двумя (**парная корреляция**) или несколькими (**множественная корреляция**) наборами данных. Для количественной оценки силы связи используются коэффициенты парной корреляции r и множественной корреляции R .

Коэффициент корреляции (безразмерная величина) – количественный показатель **линейной связи** между двумя или более наборами данных, значение которого лежит в интервале от -1 до 1. Если коэффициент равен ± 1 , то связь функциональная, если равен 0, то связь отсутствует.

Для качественной оценки **силы связи** используются специальные табличные соотношения (например, шкала Чеддока, табл. 1)

Таблица 1 – Шкала Чеддока

Значения коэффициента корреляции	Характер связи
$0 < r < 0,3$	Очень слабая
$0,3 \leq r < 0,5$	Слабая
$0,5 \leq r < 0,7$	Заметная
$0,7 \leq r < 0,9$	Сильная
$0,9 \leq r < 1$	Очень сильная

Направление связи определяется знаками \pm : близость к +1 означает, что возрастанию одного набора значений соответствует возрастание другого набора, близость к -1 означает обратное.

Для наглядности измерения всех связей в случае множественной корреляции целесообразно использовать **корреляционную матрицу** – матрицу из попарных коэффициентов корреляции.

Регрессионный анализ – вид статистического анализа, который состоит в представлении зависимости одних факторов от других в виде некоторой функции (**уравнения регрессии**) с помощью которой осуществляется прогнозирование и поиск ответа на вопросы «Что будет через какое-то время?» или «Что будет, если...?».

В случае **парной регрессии** уравнение определяется по двум наборам данных, один из которых представляет значения зависимой переменной y , а другой – независимой переменной x . В случае **множественной регрессии** уравнение определяется по нескольким наборам данных, один из которых представляет значения зависимой переменной y , а другие независимыми переменными x_1, x_2, \dots, x_m .

Получение уравнения регрессии происходит в два этапа: подбор вида функции и вычисление параметров функции.

Выбор функции, в большинстве случаев, производится среди линейной, квадратичной, степенной и др. видов функций (табл. 2). К функции предъявляются следующие требования: она должна быть достаточно простой для использования ее в дальнейших вычислениях и график этой функции должен проходить вблизи экспериментальных точек так, чтобы сумма квадратов отклонений y -координаты всех экспериментальных точек от y -координат графика функции была бы минимальной (метод наименьших квадратов).

Таблица 2 – Виды функций, применяемых в регрессионных моделях

Парная (простая) регрессия	Множественная регрессия
Линейная регрессия	
$y=ax+b,$	$y = a_0 + a_1x_1+ \dots + a_mx_m$
Квадратичная (параболическая)	
$y=ax^2+bx+c$	$y= a_0 + a_1x_1^2 + \dots +a_m x_m^2$
Степенная	
$y=ax^b$	$y = a_0 x_1^{a_1} x_2^{a_2} \dots x_m^{a_m}$
Логарифмическая $y=a\ln x+b,$	Гиперболическая $y = a_0 + a_1 (1/x_1) + \dots +a_m(1/x_m)$
Экспоненциальная $y=aeb^x$	
где a, b, c – коэффициенты парной регрессии.	где $a_0, a_1, a_2, \dots, a_m$ – коэффициенты множественной регрессии, n – объем совокупности, m – количество факторных признаков.



Для количественной оценки точности построения уравнения регрессии предназначен **коэффициент детерминации R^2** , равный квадрату коэффициента корреляции и указывающий, какой процент изменения функции y объясняется воздействием факторов x_k . Чем его значение ближе к 1, тем уравнение точнее описывает исследуемую зависимость.

Значимое уравнение (с R^2 близким к 1) используется, как правило, для **прогнозирования** изучаемого явления.

Прогноз – это вероятностное суждение о будущем, полученное путем использования совокупности научных методов. Например, прогнозирование финансового состояния выполняется для того, чтобы получить ответы на два вопроса: «как это может быть (какими могут стать финансовые показатели, если не будут приняты меры по их изменению)» и «как это должно быть (какими должны стать финансовые показатели фирмы для того, чтобы ее финансовое состояние обеспечивало высокий уровень конкурентоспособности)». Прогнозирование с целью получения ответа на первый вопрос принято называть исследовательским, на второй – нормативным.

Существует два способа прогнозов по уравнению регрессии: в пределах экспериментальных значений (**интерполяция**) и за пределами (**экстраполяция**). Применимость всякой регрессионной модели ограничена, особенно за пределами экспериментальной области, т.к. характер зависимости может существенно измениться. Поэтому достоверность исследовательского прогноза может быть невысокой. Однако его выполнение полностью обосновано.

Статистический пакет STATISTICA

Так как статистические методы находят широкое применение во всех сферах производства, то рынок компьютерных технологий предлагает большое количество прикладных программ, которые позволяют проводить такой анализ. Обилие систем, создатели которых утверждают, что их программа является наилучшей для обработки данных, а также отсутствие у большинства специалистов достаточного времени для освоения нескольких пакетов приводит к усложнению процесса выбора. Однако, по данным statsoft.ru, лидером статистических пакетов является STATISTICA.

ЗАКЛЮЧЕНИЕ

STATISTICA (американской компании StatSoft, <http://www.statsoft.com>, StatSoft RUSSIA – российское представительство StatSoft) – система, реализующая известные методы статистической обработки и визуализации данных, управления базами данных и разработки пользовательских приложений при помощи встроенного языка программирования Statistica Basic.

Пакет разработан в 1984 г., и первоначально он был представлен в виде модуля для самой популярной в то время электронной таблицы Lotus. Как самостоятельный продукт Statistica впервые заявила о себе в 1991 г. и с тех пор постоянно занимает лидирующее место среди специализированных пакетов по статистике.

Литература

<https://ru.wikipedia.org/wiki/SPSS>

<http://allbest.ru/o-2coa65625b3bc78a5d43b88521216d37.html>