**КРАТКИЙ КОНСПЕКТ ЛЕКЦИЙ**

**Тема 1. Методы обработки статистических данных**

**1. Типы данных**

* **Количественные (Числовые):** Измеряются численно. Делятся на:
  + *Дискретные:* Принимают целые значения (например, количество детей в семье, число ошибок).
  + *Непрерывные:* Принимают любые значения в интервале (например, рост, вес, температура).
* **Качественные (Категориальные):** Описывают категории или группы. Делятся на:
  + *Номинальные:* Порядок не важен (например, пол, цвет глаз, марка автомобиля).
  + *Порядковые:* Порядок имеет значение (например, уровень образования, оценка по шкале «удовл./хорошо/отлично «).

**2. Описательная статистика**

Цель: описать и суммировать основные особенности совокупности данных.

* **Меры центральной тенденции:**
  + **Среднее арифметическое:** Сумма всех значений, деленная на их количество. Чувствительно к выбросам.
  + **Медиана:** Значение, которое делит упорядоченный набор данных пополам. Устойчива к выбросам.
  + **Мода:** Наиболее часто встречающееся значение в наборе данных.
* **Меры разброса (вариабельности):**
  + **Дисперсия:** Средний квадрат отклонений отдельных значений от среднего.
  + **Стандартное отклонение:** Квадратный корень из дисперсии. Показывает среднее отклонение значений от среднего в исходных единицах измерения.
  + **Размах:** Разница между максимальным и минимальным значениями.
  + **Интерквартильный размах:** Разница между 75-м и 25-м процентилями (Q3 - Q1). Показывает разброс средних 50% данных.

**3. Инференциальная (статистическая) статистика**

Цель: сделать выводы о генеральной совокупности на основе анализа выборки.

* **Генеральная совокупность:** Вся группа объектов, которую изучают.
* **Выборка:** Подмножество объектов, отобранных из генеральной совокупности. Должна быть репрезентативной.
* **Статистическая гипотеза:**
  + **Нулевая гипотеза (H₀):** Гипотеза об отсутствии эффекта, различий или связи.
  + **Альтернативная гипотеза (H₁):** Гипотеза о наличии эффекта, различий или связи.
* **p-уровень значимости:** Вероятность получить наблюдаемые результаты при условии, что верна H₀. Малое p-value (< 0.05) свидетельствует против H₀.

**4. Основные методы анализа данных**

* **Корреляционный анализ:** Оценка силы и направления линейной связи между двумя количественными переменными.
  + **Коэффициент корреляции Пирсона (r):** От -1 до +1. Не доказывает причинно-следственную связь.
* **Регрессионный анализ:** Изучение зависимости одной переменной (зависимой) от одной или нескольких других (независимых).
  + Позволяет предсказывать значения зависимой переменной.
* **Дисперсионный анализ (ANOVA):** Используется для сравнения средних значений трех и более групп.
  + Проверяет гипотезу о том, что средние всех групп равны.

**5. Программное обеспечение**

* **SPSS:** Популярен в социальных науках, интуитивно понятный интерфейс.
* **R:** Мощный язык с открытым исходным кодом, огромное количество пакетов для любого анализа. Требует знания программирования.
* **Python (библиотеки Pandas, NumPy, SciPy, statsmodels, scikit-learn):** Универсальный язык, широко используется в научных исследованиях.
* **Excel:** подходит для базового описательного анализа и визуализации.

**6. Этика в статистическом анализе**

* **Честность и прозрачность:** Не манипулируйте данными для получения желаемого результата.
* **Корректная интерпретация:**  **Корреляция не означает причинность**. Избегайте преувеличений.
* **Отчетность:** Обязательно сообщайте обо всех проведенных тестах, размерах выборки, значениях p-value и эффекта. Не скрывайте неудобные или незначимые результаты.
* **Конфиденциальность:** Обеспечивайте анонимность и защиту персональных данных респондентов.

**Тема 2. Регрессионный анализ. Метод наименьших квадратов**

**1. Основная концепция и применение**

* Регрессионный анализ – это статистический метод, позволяющий **смоделировать и проанализировать** зависимости между переменными.
* **Цель:** Оценить, как изменения одной или нескольких переменных влияют на изменение другой переменной. Это позволяет **предсказывать** значения одной переменной на основе значений других.
* **Применение в психологии:**
  + Предсказание уровня стресса на основе количества рабочих часов и социальной поддержки.
  + Изучение влияния возраста и уровня образования на когнитивные способности.
  + Оценка связи между временем практики и успешностью выполнения задачи.

**2. Ключевые понятия**

* **Зависимая переменная (Y):** Переменная, которую мы пытаемся предсказать или объяснить (например, уровень тревожности).
* **Независимая переменная (X):** Переменная, которую мы используем для предсказания или объяснения (например, уровень стресса на работе).
* **Уравнение регрессии:** Математическая формула, описывающая зависимость.
  + **Простая линейная регрессия:** Y = a + bX + e
    - Y – зависимая переменная.
    - X – независимая переменная.
    - a (константа) –значение Y, когда X = 0.
    - b (коэффициент регрессии) – показывает, насколько изменится Y при изменении X на одну единицу.
    - e – ошибка предсказания (разница между реальным и предсказанным Y).

**3. Метод наименьших квадратов (МНК)**

* **Цель МНК:** Найти такие параметры модели ( a и b ), которые минимизируют сумму квадратов отклонений (ошибок) между реально наблюдаемыми значениями зависимой переменной и значениями, предсказанными моделью.
* **Проще говоря:** МНК проводит линию (или плоскость в многомерном случае) через данные так, чтобы общая «ошибка» предсказания была наименьшей.

**4. Типы регрессии**

* **Простая (одномерная) линейная регрессия:**
  + Использует одну независимую переменную (X) для предсказания одной зависимой (Y).
  + **Область применения:** Изучение связи между двумя переменными.
* **Множественная (многомерная) регрессия:**
  + Использует **две и более** независимых переменных (X₁, X₂, ...) для предсказания одной зависимой (Y).
  + **Уравнение:** Y = a + b₁X₁ + b₂X₂ + ... + e
  + **Область применения:** Более сложные исследования, где на результат влияет множество факторов (например, предсказание успеваемости на основе IQ, мотивации и социоэкономического статуса).

**5. Оценка качества модели**

* **Коэффициент детерминации (R²):**
  + Покажает, **какую долю дисперсии** (изменчивости) зависимой переменной можно объяснить с помощью независимых переменных модели.
  + Принимает значения от 0 до 1 (или 0% до 100%). Чем ближе к 1, тем лучше модель описывает данные.
* **Тестирование гипотез о коэффициентах (t-тест и p-value):**
  + **Нулевая гипотеза (H₀):** Коэффициент регрессии b равен нулю (т.е. переменная X не оказывает значимого влияния на Y).
  + **Альтернативная гипотеза (H₁):** Коэффициент регрессии b статистически значимо отличается от нуля.
  + **Интерпретация:** Если p-value < 0.05, мы отвергаем H₀ и делаем вывод, что данная независимая переменная является значимым предиктором зависимой переменной.

**6. Интерпретация результатов в психологии**

* **Акцент на значимости:** Недостаточно просто найти связь. Важно, является ли она статистически значимой (p-value) и практически значимой (сила эффекта, величина коэффициента).
* **Корреляция и Причинность:** Регрессия показывает предсказание, а не причинно-следственную связь. Возможное влияние скрытых (латентных) переменных всегда нужно учитывать.
* **Пример интерпретации:** «Результаты регрессионного анализа показали, что уровень стресса (β = 0.45, p < .001) является значимым предиктором уровня тревожности, в то время как возраст (β = -0.08, p = .15) – нет. Модель объясняет 30% дисперсии зависимой переменной (R² = .30)».

**7. Практическое применение в ПО**

* **Программы:** SPSS, R, Python (библиотеки statsmodels, scikit-learn), Jamovi.
* **Типовые шаги:**
  1. Ввод и подготовка данных.
  2. Выбор процедуры регрессии (например, «Linear Regression « в SPSS).
  3. Назначение зависимой и независимых переменных.
  4. Запуск анализа и интерпретация:
     + Значения коэффициентов (B).
     + Стандартизированные коэффициенты (Beta) для сравнения силы влияния предикторов.
     + p-value (часто в колонке Sig.).
     + Коэффициент детерминации (R Square).

**Тема 3.** **Назначение и возможности статистических пакетов**

**1. Назначение и основные функции**

Статистические пакеты – это специализированное программное обеспечение, предназначенное для комплексной работы с данными. Их основные функции:

* **Управление и обработка данных:**
  + Импорт данных из различных источников (Excel, текстовые файлы, базы данных).
  + Очистка данных: обработка пропущенных значений, выбросов, преобразование переменных.
  + Создание новых переменных и реструктуризация наборов данных.
* **Проведение статистического анализа:**
  + **Описательная статистика:** расчет средних, медиан, моды, стандартного отклонения и др.
  + **Проверка статистических гипотез:** проведение t-тестов, дисперсионного анализа (ANOVA), хи-квадрат тестов, регрессионного анализа, факторного анализа и многих других методов.
* **Визуализация результатов:**
  + Построение графиков и диаграмм (гистограммы, box-plot, scatter plot, линейные графики) для наглядного представления данных и результатов анализа.

**2. Обзор и сравнение популярных пакетов**

| **Программа** | **Уровень сложности** | **Основные преимущества** | **Недостатки** | **Идеально для** |
| --- | --- | --- | --- | --- |
| **SPSS** | Начальный/Средний | Очень простой и интуитивный интерфейс (меню и кнопки). Легок в освоении. | Дорогой коммерческий лицензия. Ограниченная гибкость для сложных и нестандартных анализов. | Студентов и исследователей в социальных науках, начинающих свой путь в статистике. |
| **R (+ RStudio)** | Средний/Продвинутый | Бесплатный и с открытым исходным кодом. Огромное сообщество и тысячи пакетов для любых задач. Максимальная гибкость и возможности. | Сложный порог входа, требует изучения синтаксиса и основ программирования. | Исследователей, которым нужна мощь, гибкость и современные методы анализа. |
| **Python (Pandas, NumPy, SciPy, statsmodels, scikit-learn)** | Средний/Продвинутый | Бесплатный и открытый. Мощь и гибкость, сравнимая с R. Лидер в области машинного обучения и работы с большими данными. | Требует глубоких знаний программирования. | Исследователей на стыке дисциплин,, тех, кто работает с ИИ и машинным обучением. |
| **Stata** | Средний | Мощный для работы с панельными данными и сложными экономико-статистическими моделями. Удобное управление данными. | Коммерческий, узкоспециализированный. Менее популярен в психологии по сравнению с SPSS и R. | Экономистов, социологов, политологов. |

**3. Применение в психологических исследованиях**

Статистические пакеты используются на всех этапах анализа данных:

1. **Подготовительный этап:** Очистка и подготовка сырых данных к анализу (SPSS, R, Python).
2. **Первичный анализ:** Расчет описательной статистики и построение графиков для общего понимания данных.
3. **Проверка гипотез:**
   * **Сравнение групп:** t-тест (для 2-х групп) или ANOVA (для 3-х и более групп) в SPSS или R.
   * **Анализ связей:** Корреляционный анализ (Pearson, Spearman).
   * **Предсказание:** Регрессионный анализ (линейный, логистический) для моделирования влияния переменных.
4. **Сложное моделирование:** Проведение факторного анализа, структурного моделирования (SEM), многоуровневого моделирования (HLM) в R или специализированных пакетах (Mplus).

**4. Критерии выбора программного обеспечения**

Правильный выбор зависит от двух ключевых факторов:

* **Задачи исследования:**
  + Стандартный анализ (t-тесты, ANOVA, регрессия) – SPSS.
  + Сложные, нестандартные или современные методы анализа – R или Python.
  + Работа с большими данными или машинное обучение – Python.
* **Уровень подготовки исследователя:**
  + Новичок: SPSS или Jamovi (бесплатный аналог SPSS с открытым кодом).
  + Готов инвестировать время в обучение для больших возможностей: R.
  + Программист/Data scientist: Python.

**5. Пример практического использования**

* **Исследование:** Влияние метода терапии на уровень тревожности.
* **Действия в SPSS:**
  1. Импорт файла Excel с данными (группа испытуемых и показатели тревожности до и после терапии).
  2. **Описательный анализ:** Analyze > Descriptive Statistics > Descriptives для расчета средних и стандартных отклонений.
  3. **Визуализация:** Graphs > Chart Builder для построения диаграмм размаха (box-plot) по группам.
  4. **Проверка гипотезы:** Analyze > Compare Means > Paired-Samples T-Test для сравнения показателей «до « и «после « в одной группе или Independent-Samples T-Test для сравнения экспериментальной и контрольной групп.
  5. Интерпретация: значения t, степеней свободы (df) и p-value (Sig.).

**Тема 4. Примеры работы в математических пакетах**

1. Обзор основных математических пакетов и их возможностей

Математические пакеты – это специализированные программные среды для сложных вычислений, моделирования и анализа данных. В психологии они применяются для решения задач, требующих расширенной статистики, математического моделирования и обработки больших данных.

Ключевые пакеты:

* **MATLAB**: Основные возможности включают матричные вычисления, разработку алгоритмов, симуляцию моделей и нейросетевой анализ. В психологии используется для обработки сигналов ЭЭГ/фМРТ, моделирования когнитивных процессов и психофизических данных.
* **Mathematica**: Подходящая для аналитических вычислений, символьной математики и создания интерактивных визуализаций. Применяется для теоретического моделирования в психологии (например, модели принятия решений) и сложного статистического анализа.

2. Практические примеры использования

* **Обработка данных ЭЭГ в MATLAB**: Фильтрация сигналов, выделение паттернов (например, P300), анализ когерентности (Signal Processing Toolbox). Пример: исследование нейронных коррелятов внимания.
* **Моделирование поведения в Mathematica**: Анализ уравнений дифференциальных моделей обучения, симуляция траекторий принятия решений. Пример: прогнозирование динамики научения в экспериментах.
* **Статистический анализ**: Реализация методов, выходящих за рамки стандартных пакетов (например, байесовская статистика, непараметрические методы), расчёт сложных метрик (энтропия данных, фрактальный анализ).

3. Алгоритмы и методы

Математические пакеты используют:

* Численные методы (оптимизация, Монте-Карло симуляции) для подбора параметров моделей.
* Символьные вычисления для аналитического решения уравнений (например, в психофизических законах).
* Встроенные алгоритмы машинного обучения (кластеризация, классификация) для выявления паттернов в данных.
* Специализированные пакеты (e.g., Statistics and Machine Learning Toolbox в MATLAB) для реализации дисперсионного анализа, регрессии, PCA.

4. Визуализация данных

* **MATLAB**: Построение 2D/3D графиков, тепловых карт, анимированных диаграмм для представления динамики процессов (например, изменение активации мозга во времени).
* **Mathematica**: Интерактивные графики, возможность создания диаграмм с высокой степенью кастомизации, визуализация многомерных данных.
* Примеры: графики распределения данных, траектории моделей, карты активации в нейроисследованиях.

5. Примеры успешного применения в психологии

* **Когнитивное моделирование**: Использование MATLAB для создания и тестирования вычислительных моделей памяти (например, ACT-R) и восприятия.
* **Психофизика**: Расчёт порогов восприятия с помощью психометрических функций в Mathematica.
* **Нейронаука**: Анализ фМРТ-данных с подключением кастомных скриптов в MATLAB для определения активации зон мозга.  
  Влияние на интерпретацию: точные количественные оценки, возможность проверить сложные гипотезы, выявить скрытые зависимости.

6. Преимущества и недостатки

* **Преимущества**:
  + Гибкость и мощь вычислений.
  + Возможность реализации кастомных алгоритмов.
  + Интеграция с оборудованием (e.g., нейроинтерфейсы).
  + Высокая точность и воспроизводимость результатов.
* **Недостатки**:
  + Высокий порог входа (требуются знания математики и программирования).
  + Дорогие лицензии (особенно MATLAB).
  + Избыточны и неоправданно сложныдля стандартных статистических задач.
  + Меньше готовых решений «в один клик « по сравнению сSPSS/R.

Ключевой вывод

Математические пакеты — это инструмент для расширенных исследований в психологии, где требуются сложные вычисления, моделирование или работа со специализированными данными (нейро-, психофизиология). Их использование оправдано при необходимости выйти за рамки стандартных статистических методов.