

Анализ и представление данных психологического исследования

Лекция 4.
Изучение различий между группами

Решаемые задачи

ВОПРОС: Можно ли объяснить имеющиеся различия между группами статистическими колебаниями, или между ними имеют место реальные значимые различия?

- Поиск различий между естественными группами (гендерные, межкультурные, возрастные и т.д.)
- Сравнение контрастных групп при разработке и валидации теста
- Доказательство различий между экспериментальной и контрольной группой

Методы изучения различий между группами

- Параметрические тесты
- Непараметрические тесты
- Однофакторный дисперсионный анализ
- Многофакторный дисперсионный анализ
- Дискриминантный анализ

Независимые и зависимые выборки

- Выборки называются *независимыми*, если процедура отбора единиц в первую выборку никак не связана с процедурой отбора единиц во вторую выборку
- Выборки называются *зависимыми*, если каждая единица одной выборки «привязывается» к определенной единице второй выборки

Параметрические тесты

(параметрические шкалы, нормальное распределение)

Ситуация	Статистический тест
Сравнение среднего в выборке с заданной величиной	t-тест для одной выборки
Сравнение двух независимых выборок	t-тест для независимых выборок (тест Стьюдента)
Сравнение двух зависимых выборок	t-тест для зависимых выборок
Сравнение более 2 независимых выборок	Однофакторный дисперсионный анализ
Сравнение более двух зависимых выборок	Однофакторный дисперсионный анализ с повторными измерениями

Процедуры в SPSS

Статистический тест	
t-тест для одной выборки	Analyze – Compare means – One-sample T-test
t-тест для независимых выборок (тест Стьюдента)	Analyze – Compare means – Independent-sample T-test
t-тест для зависимых выборок	Analyze – Compare means – Paired-sample T-test
Однофакторный дисперсионный анализ	Analyze – Compare means – One-Way ANOVA
Однофакторный дисперсионный анализ с повторными измерениями	Analyze – General Linear Model – Repeated Measures

Пример

Изменится ли содержание холестерина через месяц после приема нового лекарства?

Paired Samples Test (Тест для парных выборок)

		Paired Differences (Парные разницы)					T	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95 % Confidence Interval of the Difference				
					Lower	Upper			
Pair (Пары)	Холестерин, исходный - Холестерин, через 1 мес.	-1,93	26,09	1,98	-5,83	1,98	-,974	173	,332

Вывод

Гипотеза o : Наблюдаемые различия между средними значениями выборок находятся в пределах случайных отклонений (*различий нет*)

Если $p < 0,05$, то нулевая гипотеза о равенстве средних значений выборок **отвергается** при 95% уровне доверия

Вывод

Изменится ли содержание холестерина через месяц после приема нового лекарства?

Paired Samples Test (Тест для парных выборок)

		Paired Differences (Парные разницы)					T	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95 % Confidence Interval of the Difference				
					Lower	Upper			
Pair (Пары)	Холестерин, исходный - Холестерин, через 1 мес.	-1,93	26,09	1,98	-5,83	1,98	-,974	173	,332

Вывод: Значимого изменения содержания холестерина за один месяц после начала приема лекарства не наблюдается

Непараметрические тесты

(порядковые шкалы, нет нормального распределения)

Ситуация	Статистический тест
Сравнение двух независимых выборок	U-тест Манна-Уитни, тест Колмогорова-Смирнова
Сравнение двух зависимых выборок	Тест Уилкоксона, тест знаков
Сравнение более 2 независимых выборок	H-тест Крускала-Уоллиса
Сравнение более двух зависимых выборок	Тест Фридмана

Процедуры в SPSS

Статистический тест	SPSS
U-тест Манна-Уитни, тест Колмогорова-Смирнова	Analyze – Nonparametric Tests – 2 Independent Samples
Тест Уилкоксона, тест знаков	Analyze – Nonparametric Tests – 2 Related Samples
H-тест Крускала-Уоллиса	Analyze – Nonparametric Tests – K Independent Samples
Тест Фридмана	Analyze – Nonparametric Tests – K Related Samples

Пример

Отличаются ли показатели сахара в крови для мужчин и женщин?

Статистика теста ^a

	Blutzucker, Ausgangswert (Сахар, исходное значение)
Mann-Whitney U (U-тест по Манну и Уитни)	3048,000
W Уилкоксона	4818,000
Z	-1,096
Asymp. Sig. (2-tailed) Асимптотическая значимость (2-сторонняя)	,273

a. Grouping Variable: Geschlecht (Групповая переменная: пол).

Вывод

Гипотеза H_0 : Наблюдаемые различия между средними значениями выборок находятся в пределах случайных отклонений (*различий нет*)

Если $p < 0,05$, то нулевая гипотеза о равенстве средних значений выборок **отвергается** при 95% уровне доверия

Пример

Отличаются ли показатели сахара в крови для мужчин и женщин?

Статистика теста ^a

	Blutzucker, Ausgangswert (Сахар, исходное значение)
Mann-Whitney U (U-тест по Манну и Уитни)	3048,000
W Уилкоксона	4818,000
Z	-1,096
Asymp. Sig. (2-tailed) Асимптотическая значимость (2-сторонняя)	,273

a. Grouping Variable: Geschlecht (Групповая переменная: пол).

Вывод: Значимого различия в показателях сахара в крови между мужчинами и женщинами нет.

Дисперсионный анализ (ANOVA)-1

- Группа методов, применяемых для определения того, значимо ли отличаются друг от друга несколько выборок по признаку, измеренному в метрической шкале (зависимая переменная, отклик).
- Деление на выборки происходит по какому-либо основанию (независимая переменная, фактор).

Дисперсионный анализ (ANOVA)-2

- Зависимая переменная (отклик) должна быть измерена в параметрической шкале и иметь нормальное распределение
- Независимые переменные могут быть категориальными (дискретными)
- Число респондентов, приходящееся на каждое сочетание значений категорий по разным факторам, должно быть не менее 20

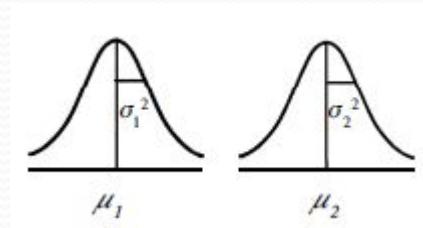
В случае 3-факторного анализа, где два фактора имеют по 2 категории, а один – 3, всего возможно 12 групп, □ необходимо не менее 240 респондентов

Дисперсионный анализ (ANOVA)-3

- Нулевая гипотеза: влияние независимой переменной (фактора) на зависимую переменную (отклик) отсутствует
- Альтернативная гипотеза: межгрупповая дисперсия значимо больше внутригрупповой, а значит, среди множества наборов средних значений (это число равно числу уровней фактора) по крайней мере два средних значения различаются

Математические допущения ДА

- Значения признаков, соответствующих каждому уровню фактора, должны быть нормально распределены



- Равенство дисперсий выборочных распределений на каждом уровне фактора
- Независимость полученных наблюдений

Виды ДА

По количеству факторов

- Однофакторный анализ – исследуется влияние одного фактора
- Многофакторный – изучается одновременное воздействие двух или более факторов

По наличию связи между выборками

- Анализ несвязанный выборок
- Анализ связанных выборок

По количеству зависимых переменных

- Одномерный анализ (ANOVA) – одна зависимая переменная
- Многомерный анализ (MANOVA) – несколько зависимых переменных

Идея ДА

Основная идея ДА состоит в сравнении общей дисперсии и внутри групповой дисперсии в каждой из подвыборок, соответствующих разным уровням фактора

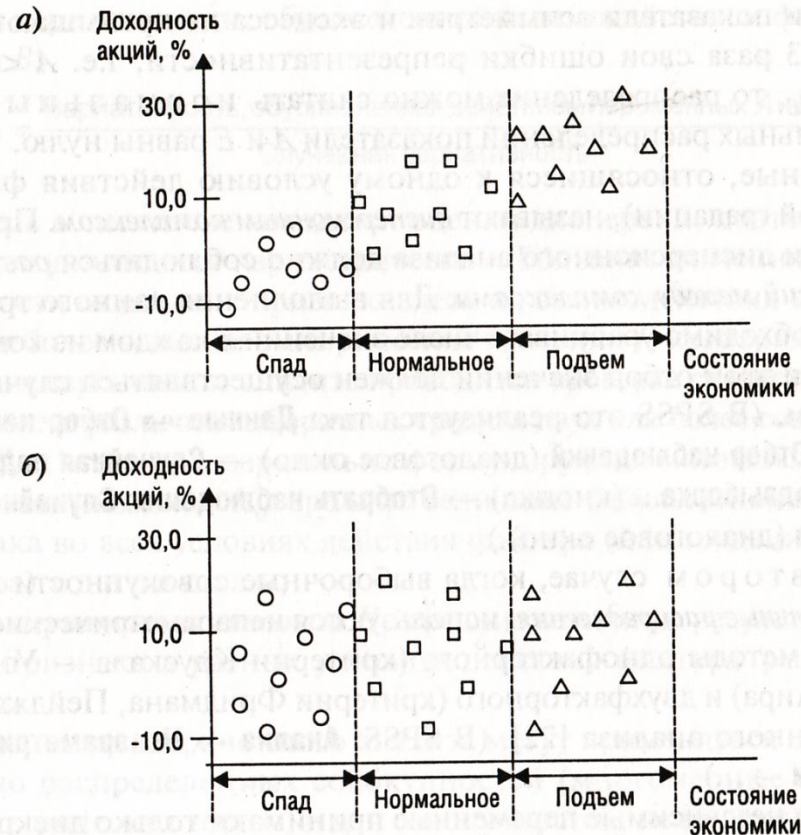
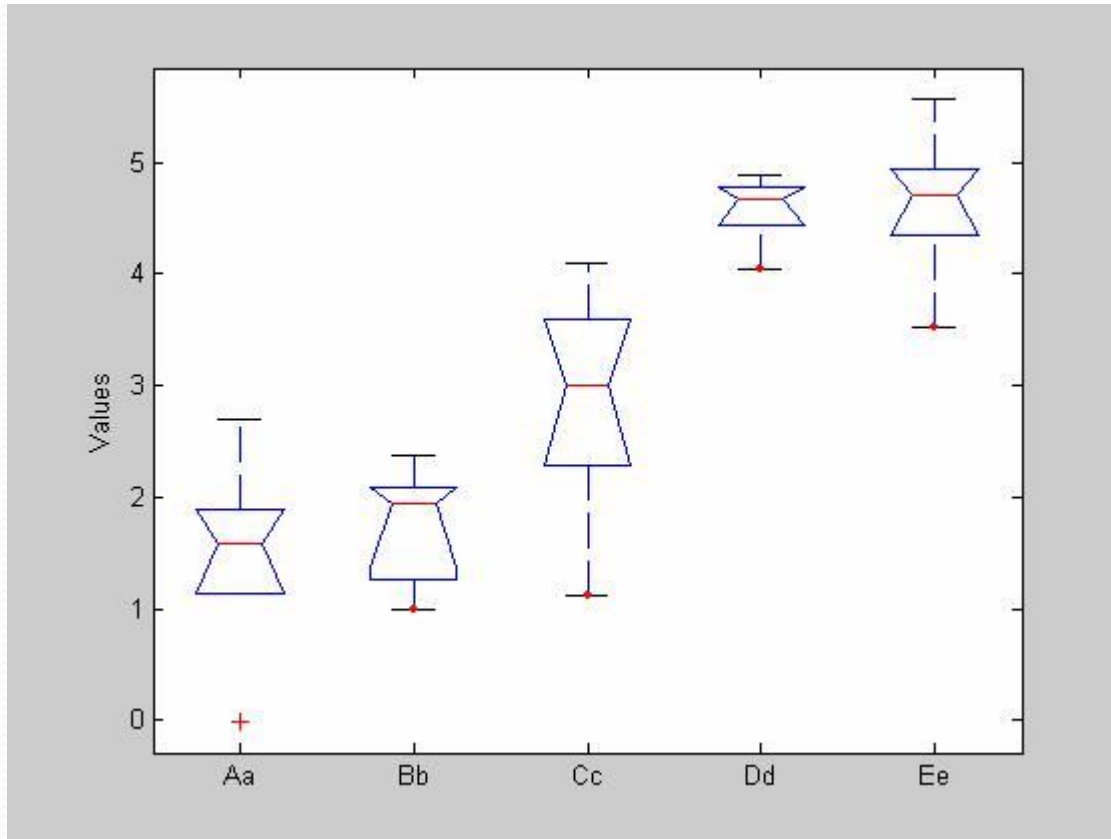


Рис. 6.1. Соотношение влияния фактора и внутригруппового разброса
а — существенное влияние фактора; б — незначимое влияние фактора (значителен внутригрупповой разброс)

Графическое представление



Графическое представление

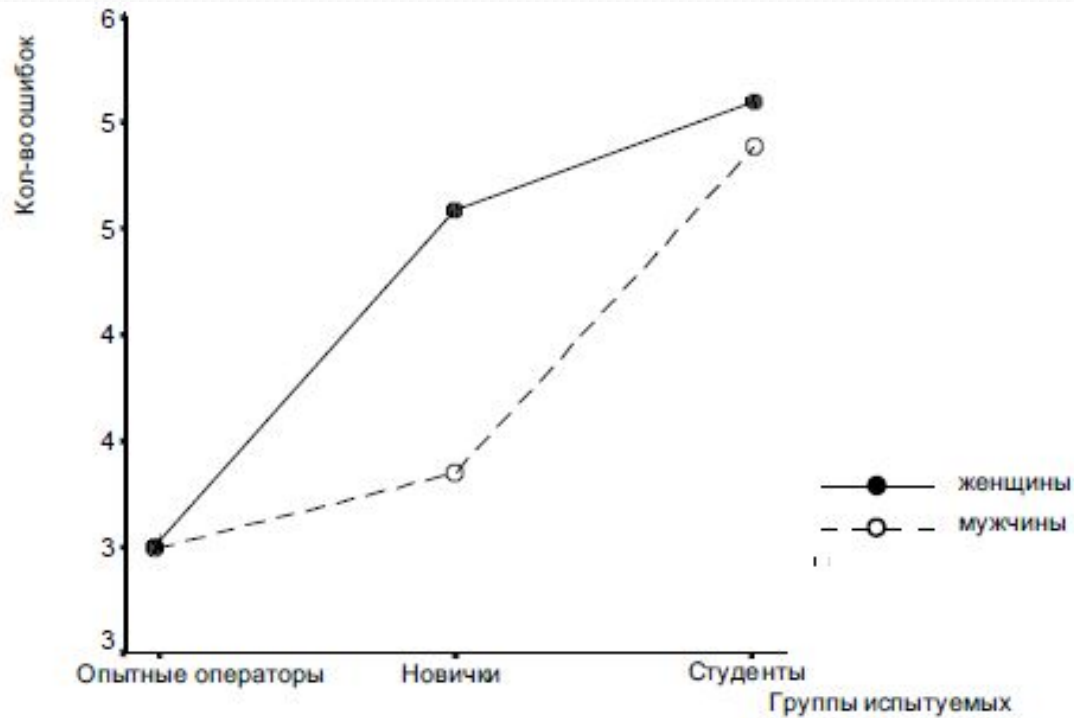


Рис. 8. Результаты двухфакторного эксперимента по слежению за движущимся объектом:
ось X — группа испытуемых; ось Y — количество допущенных ошибок

Графическое представление

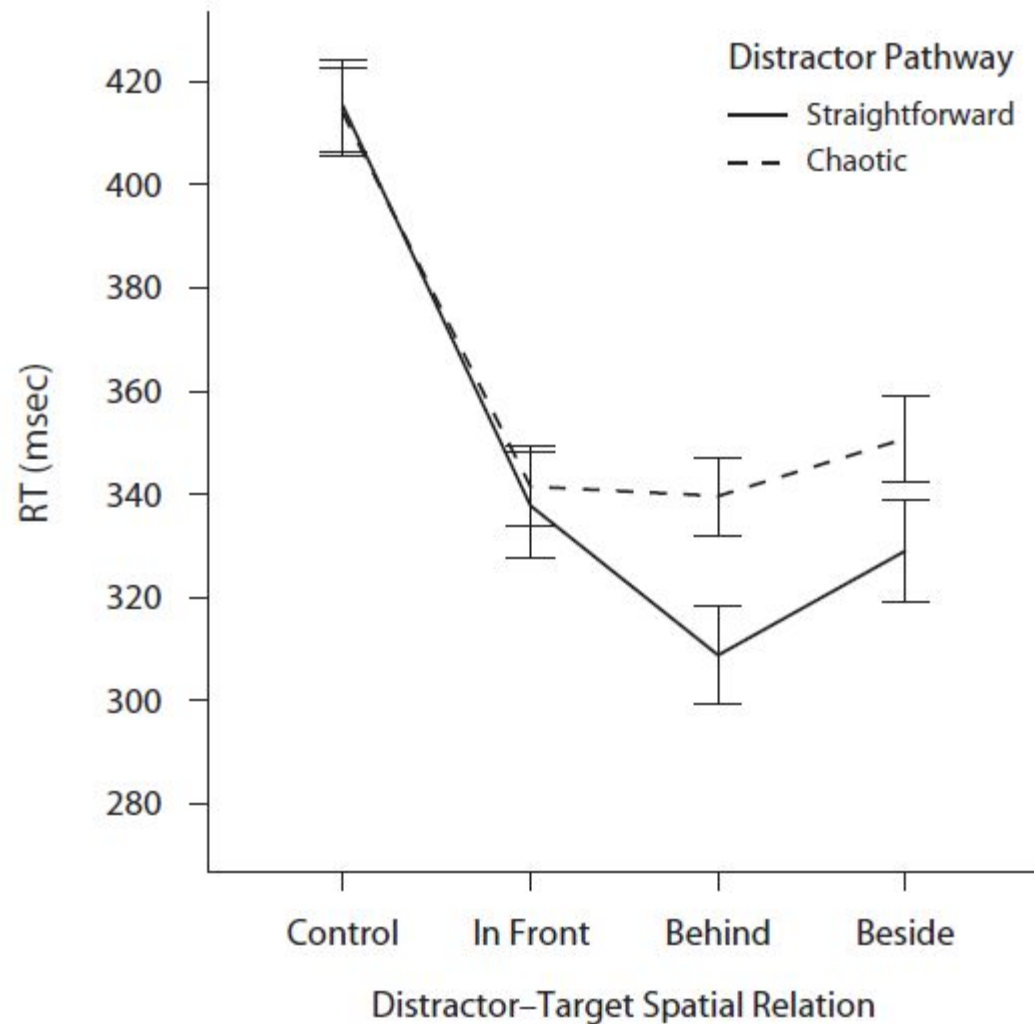


Figure 2. Effects of a moving distractor on reaction time (RT). Error bars denote $\pm SE$.

Графическое представление

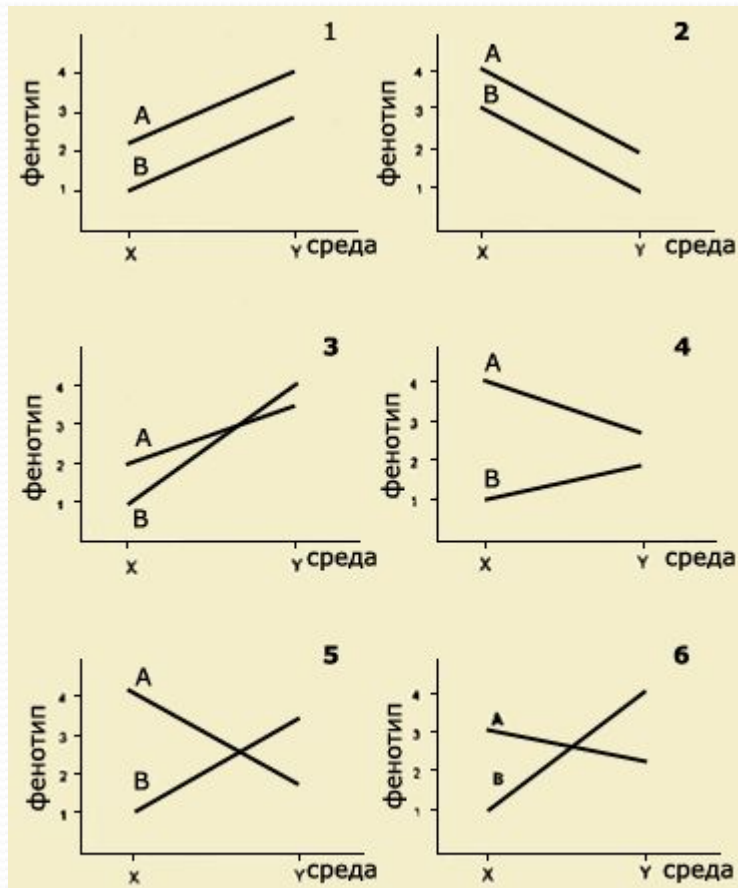


Рис. 4.10
Примеры генотип-средового взаимодействия