

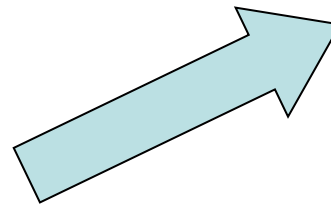
Основные проблемные аспекты методологии научных исследований: как сделать хорошее исследование?

Евгений Осин, НИУ ВШЭ, 2013

eosin@hse.ru

С чего начинается исследование?

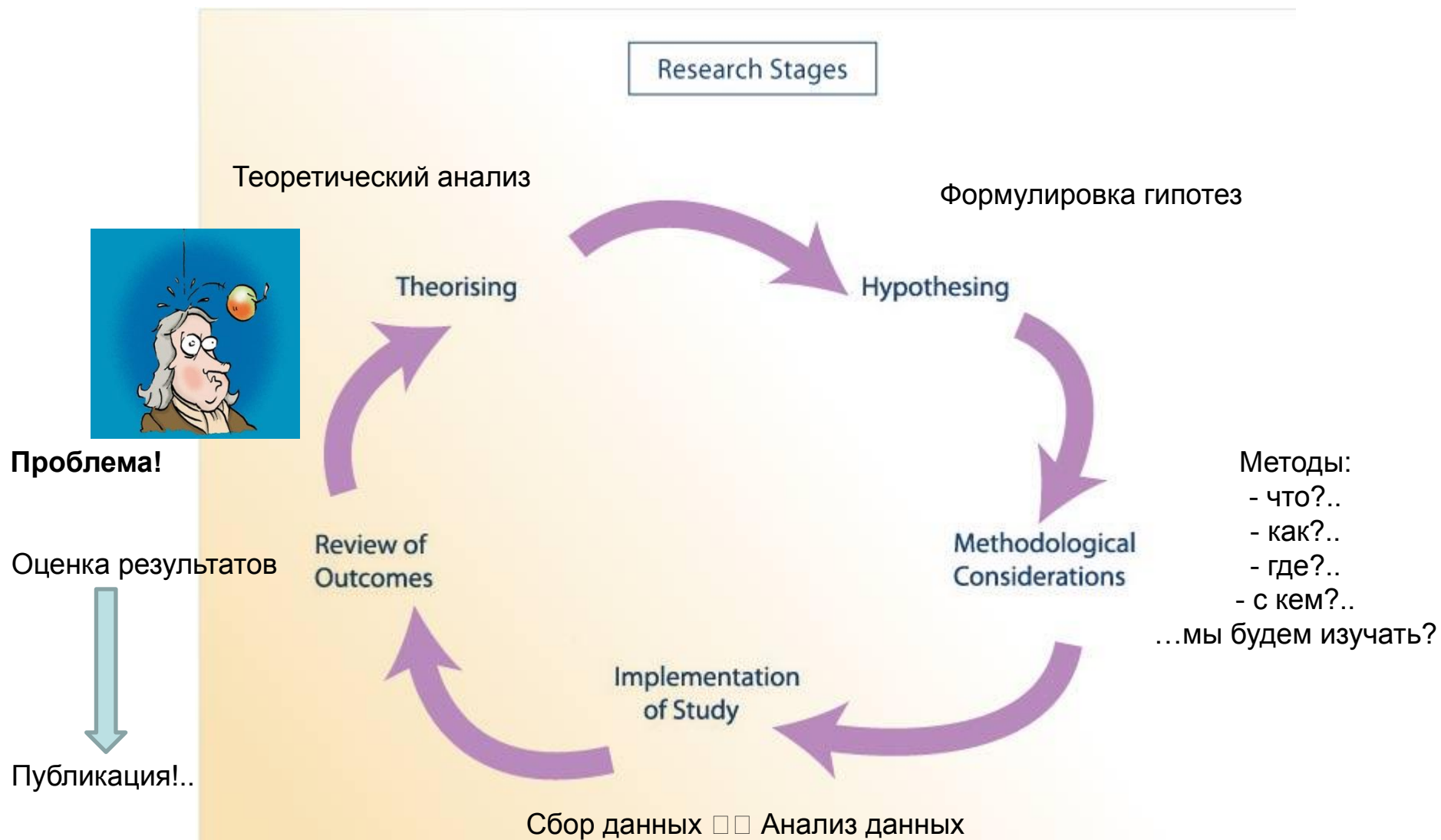
- **Научная проблема, или исследовательский вопрос.**



Любой вопрос (даже бредовый), касающийся некоторого психического (социального, экономического...) явления.

Этапы научного исследования

(самая простая из сотен тысяч разных схем)



Научная проблема

- Исследовательский вопрос = проблема исследования □ Является ли проблема исследования *научной проблемой*?
- Шансы на успех зависят от того, как она поставлена:
 - используются ли научные понятия, опирается ли она на научную картину реальности?
 - соотносится ли она с существующими теориями, релевантна ли актуальному научному дискурсу? (однако Вы можете поставить новую научную проблему, преодолев инерцию научного сообщества)
 - актуальна ли она с точки зрения социума? (готов ли кто-то дать на это денег?)

Вопросы и типы исследования

Некоторые варианты постановки вопросов сразу предполагают определённый тип исследования

«Количественные»

- Есть ли связь...?
- Есть ли различия в... в зависимости от ...?

«Качественные»

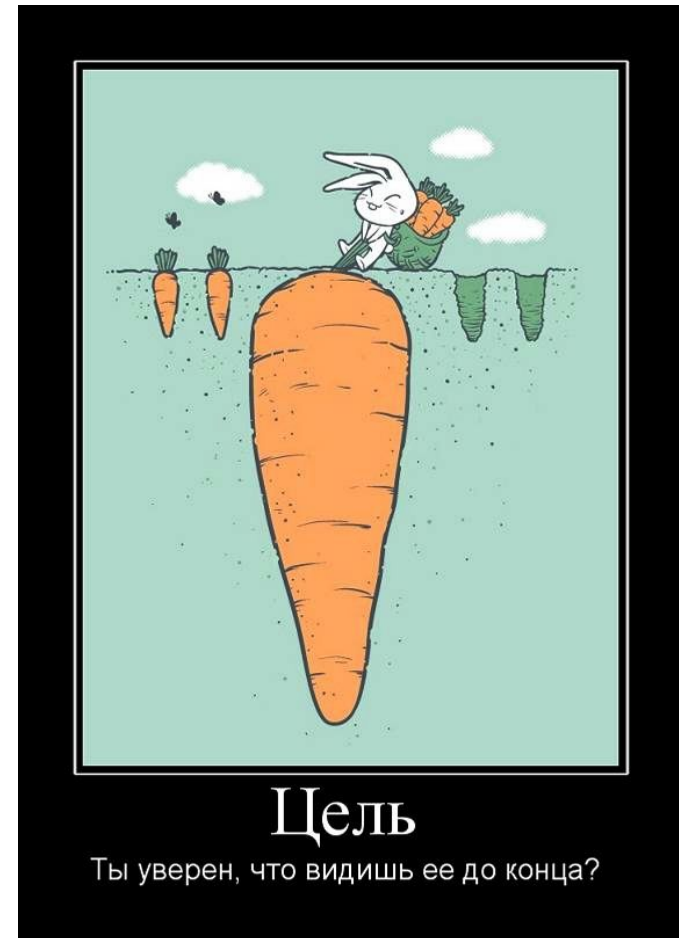
- Как...? (□ описать процесс, ситуацию)
- Зачем...? (цели людей)



Иногда просто нет смысла использовать методы определённого типа, если проблема (вопрос) исследования не сформулирована «под них».

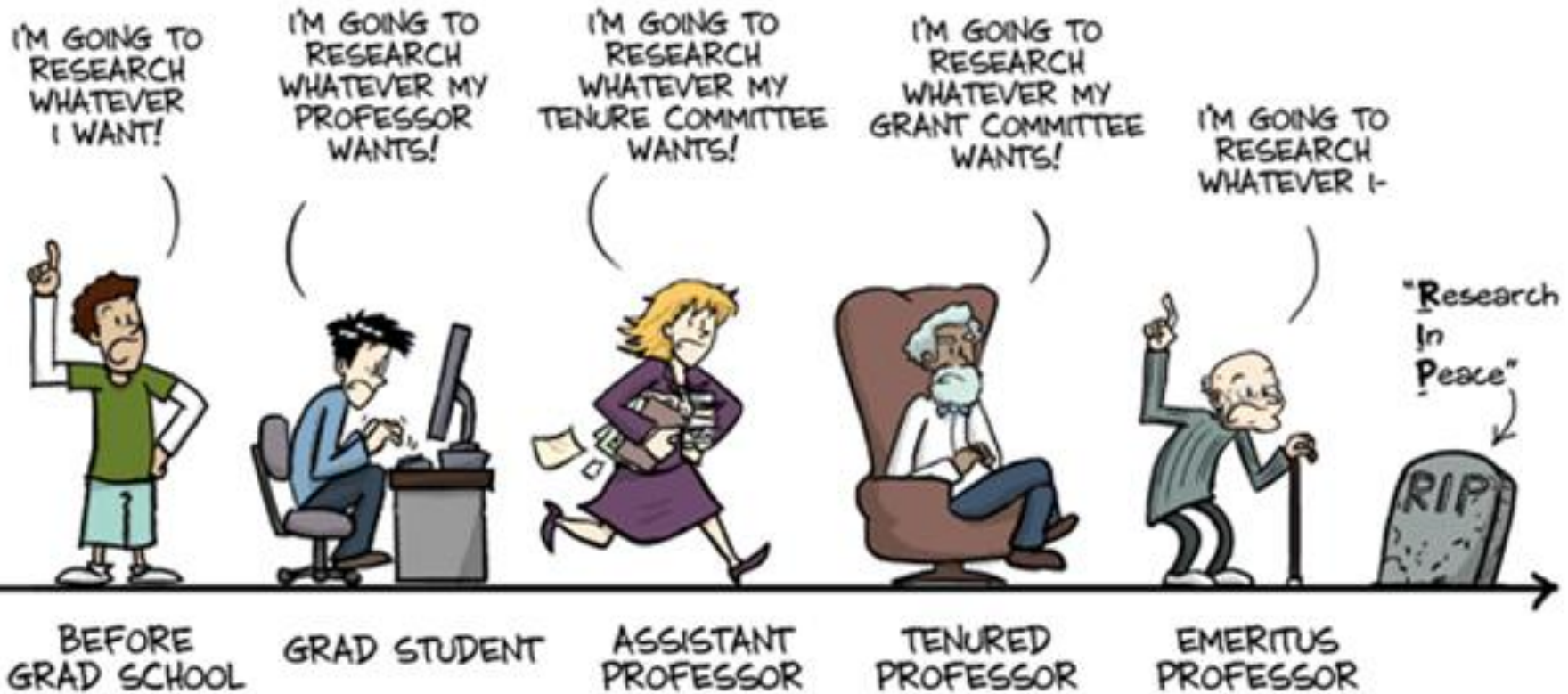
По цели исследования бывают...

- **Поисковое** (изучение неизученного процесса, поиск закономерностей)
- **Подтверждающее** (воспроизведение ранее полученного факта)
- **Критическое** (спланированная проверка определённой модели, положения теории)



Выбор проблемы исследования

THE EVOLUTION OF INTELLECTUAL FREEDOM



JORGE CHAM © 2011

ТЕОРИЯ

Зачем нам теория?

- Два варианта развития научного поиска:
 - **Теория** □ Проблема □ Выбор феномена □
Исследование феномена □ Интерпретация фактов
 - Феномен □ Проблема □ Исследование феномена □
Интерпретация фактов □ **Теория**

Зачем нам нужен теоретический обзор?

- Убедиться в том, что наше исследование **актуально** и мы не **«изобретаем велосипед»**.
- Рассмотреть различные **варианты постановки** интересующей нас **проблемы**.
- Рассмотреть **различные подходы и методы** изучения интересующей нас проблемы.
- **Обобщить** существующие теоретические **знания** и **накопленные эмпирические данные**.

Виды теоретических обзоров

- Теоретический обзор как **необходимая основа любого эмпирического** исследования:
 - из него вытекает, зачем Вы проводите своё исследование и почему делаете это именно так;
- Теоретический обзор как **особый вид аналитической работы**:
 - **прояснение** того, как ставится та или иная проблема в науке;
 - **обобщение** предыдущих исследований в форме «дайджеста» для читателей;
 - **выявление** связей, противоречий, «белых пятен» и несоответствий в имеющейся литературе;
 - наметить **следующие шаги** в решении проблемы (Eisenberg, 2000).

Характеристики хорошего обзора

- **Широта** круга источников
- **Глубина** анализа источников
- **Релевантность** источников
- **Аккуратность** в интерпретациях
- **Критика** существующих позиций
- Качественное **обобщение**
- **Логичная структура** (A->B->C)
- **Эффективность**: качество/объём

Варианты организации обзора

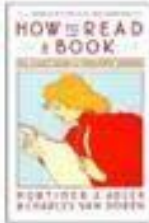
- По **логике теории**: основные идеи теории □ частные теории, модели...
- По **исторической логике**:
Платон □ ... □ Вундт □ ... □ Пупкин
- От **феноменов**: есть А, есть Б □ сопоставление, проблема
- «**Как получится**»: Nancy Eisenberg: нет единого «правильного» способа структурировать обзор литературы.

Этапы создания обзора

- **Определение проблемного поля**
- **Постановка вопросов (целей)**
- **Поиск и определение круга источников**
- **«Путешествие по ссылкам»**
- **Структурирование**
- **Анализ и обобщение**

Ресурсы для поиска литературы

- **Поисковые системы:**
 - Google Scholar
- **Базы научных статей:**
 - базы издательств (hse.ru □ Электронные ресурсы библиотеки): APA, Elsevier, Wiley, etc.
 - базы-«агрегаторы», такие как EBSCO
- **Индексы цитирования:**
 - Мировых два: Scopus и ISI Web of Science
 - Российский индекс научного цитирования (РИНЦ): ELibrary.ru
- **Разнообразные источники:**
 - Google □ Wikipedia, библиотека!



HOW TO READ A BOOK

MORTIMER J. ADLER & CHARLES VAN DOREN

INSPECTIONAL READING

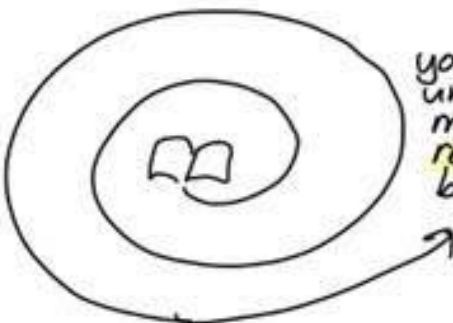
systematic skimming

Does the book deserve careful reading?

Title, table of contents, index, blurb, key chapters, skimmed pages

Superficial reading:

In tackling a book for the first time, read it through without ever stopping to look up or ponder the things you do not understand right away.

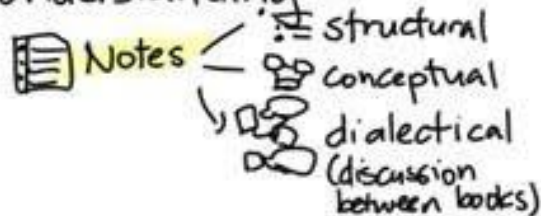


you will understand more as you reread the book.

Read at the speed appropriate for the book.

ANALYTICAL READING

understanding



1. Classify the book.

2. Summarize the book in one sentence or a short paragraph.

3. Show the book's organization/structure.

might not be the same as chapter structures!

4. Define the problem the author tries to solve.

Understand:

Terms Message

Is it true? So what?

Sentences Arguments Propositions

SYNOPTICAL READING

many books & their relationships



1. Find the relevant passages (relevant to you) ← important? not important?

2. Reconcile authors' terms.

3. Clarify the key questions in the discussion between books.

4. Define the issues.

5. Analyze the discussion.

Lots of tips on how to read different types of material. (including philosophy!)

Exercises for different levels of reading!

Как быстро сориентироваться в теме?

- Электронные ресурсы библиотеки Scopus
- Вводим ключевые слова
- Сортируем найденные статьи по убыванию кол-ва цитирований
- Просматриваем первые 10-20-... (в зависимости от наличия времени) статей; в первую очередь, читаем обзоры и мета-анализы

Рекомендуемый алгоритм поиска

- 1) Ищем по ключевым словам нужные статьи в **Scopus / ISI Web of Science**.
- 2) По кнопке **HSE_FullText** переходим к статьям (если ссылка работает).
- 3) Либо проверяем, есть ли журнал в нашей подписке, через «**A-to-Z сводный каталог**» (если есть – там же выходим на нужную БД).
- 4) Ищем по ключевым словам в **Google Scholar** (более широкий охват + ссылки на бесплатные ресурсы).
- 5) Ищем в **РИНЦ** (elibrary.ru) и русскоязычных поисковых системах.

Создание обзора

- Даёт ли обзор исчерпывающую информацию о состоянии проблемы в науке, учитывает ли разные основные имеющиеся подходы и методы её решения?
- Является ли обзор достаточным обоснованием исследования: вытекает ли из него, что нужно провести именно такое исследование и именно так?
- Является ли текст обзора достаточно экономным (кратким), хорошо структурированным и читаемым?

ГИПОТЕЗЫ

От вопроса к гипотезе

- Гипотеза – это **конкретное** предположение об изучаемой реальности:
 - сформулированное на языке научных понятий (а не житейских терминов), предполагающих тот или иной вариант понимания изучаемой реальности;
 - истинность которого между проверить теоретически, либо соответствие которого реальности можно проверить с помощью эмпирической процедуры.
- **Хорошую гипотезу можно проверить.**
Плохую гипотезу проверить невозможно.
- (Хорошая гипотеза – когда неочевидно, подтвердится ли она...)

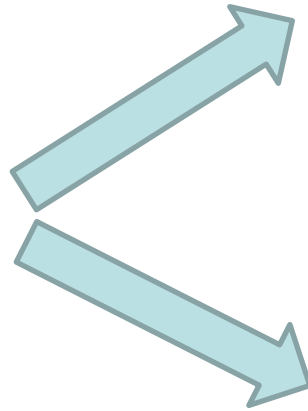
Определения понятий

- Чтобы сформулировать гипотезы, нам нужно дать **определения понятий**, опираясь на существующие теории или описания феноменов.
- **Операциональное определение** понятия (описывающее то, с чем мы можем/будем работать в исследовании) может не совпадать с теоретическим определением (описывающим понятие в целом):
 - например, агрессивность можно операционально определить как наличие установок, выражающих враждебное отношение к другим.

От понятия к определению



Содержание понятия
(в реальности)



Операциональное определение
(в исследовании; зависит от задачи)

Гипотезы

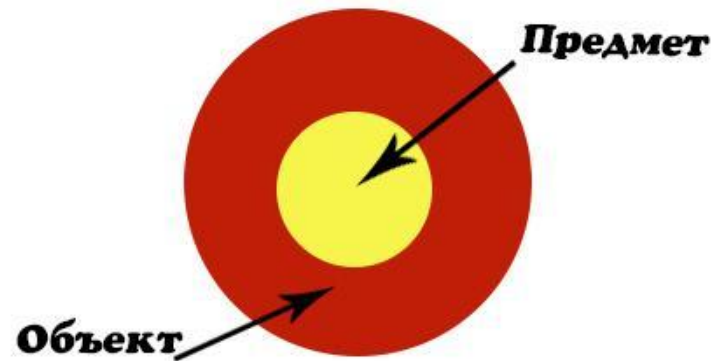
- Теоретические гипотезы (противоречие в теории(ях) □ проверка теоретически)
- Эмпирические гипотезы (проверка опытом):
 - о наличии явления («А был ли мальчик?»);
 - о связях между явлениями;
 - о причинно-следственной связи между явлениями.
- Статистические гипотезы (в терминах измеряемых переменных, проверка статистическая):
 - Нулевая (основная) гипотеза (H_0): «Мальчика не было».
 - Альтернативная гипотеза (H_1): нулевая гипотеза неверна.

Нужны ли нам гипотезы?

- Они **обязательны**, если исследование подтверждающее или критическое.
- В поисковом исследовании гипотезы **по сути могут отсутствовать** (но по форме, как правило, должны быть □ формулируем исследовательский вопрос в виде гипотезы).
- (Иногда руководители рекомендуют формулировать гипотезу после исследования... Хорошо ли это?)

Предмет и объект

- Предмет = что нас реально интересует?
- Объект = с чем мы работаем в исследовании, чтобы это узнать?
- Как правило, П и О соотносятся как общее и частное:
- «Объект – люди...» – это слишком общо. Берите одну из областей феноменов психического.
- В западных (англоязычных) исследованиях П и О, как правило, никто не выделяет.



Гипотезы

- Ясны ли гипотезы?
- Проверяемы ли они?
- На какой теоретический контекст они опираются (и почему именно на этот)?
- Какие возможности операционализации данных гипотез существуют и почему выбрана именно эта?

МЕТОДЫ

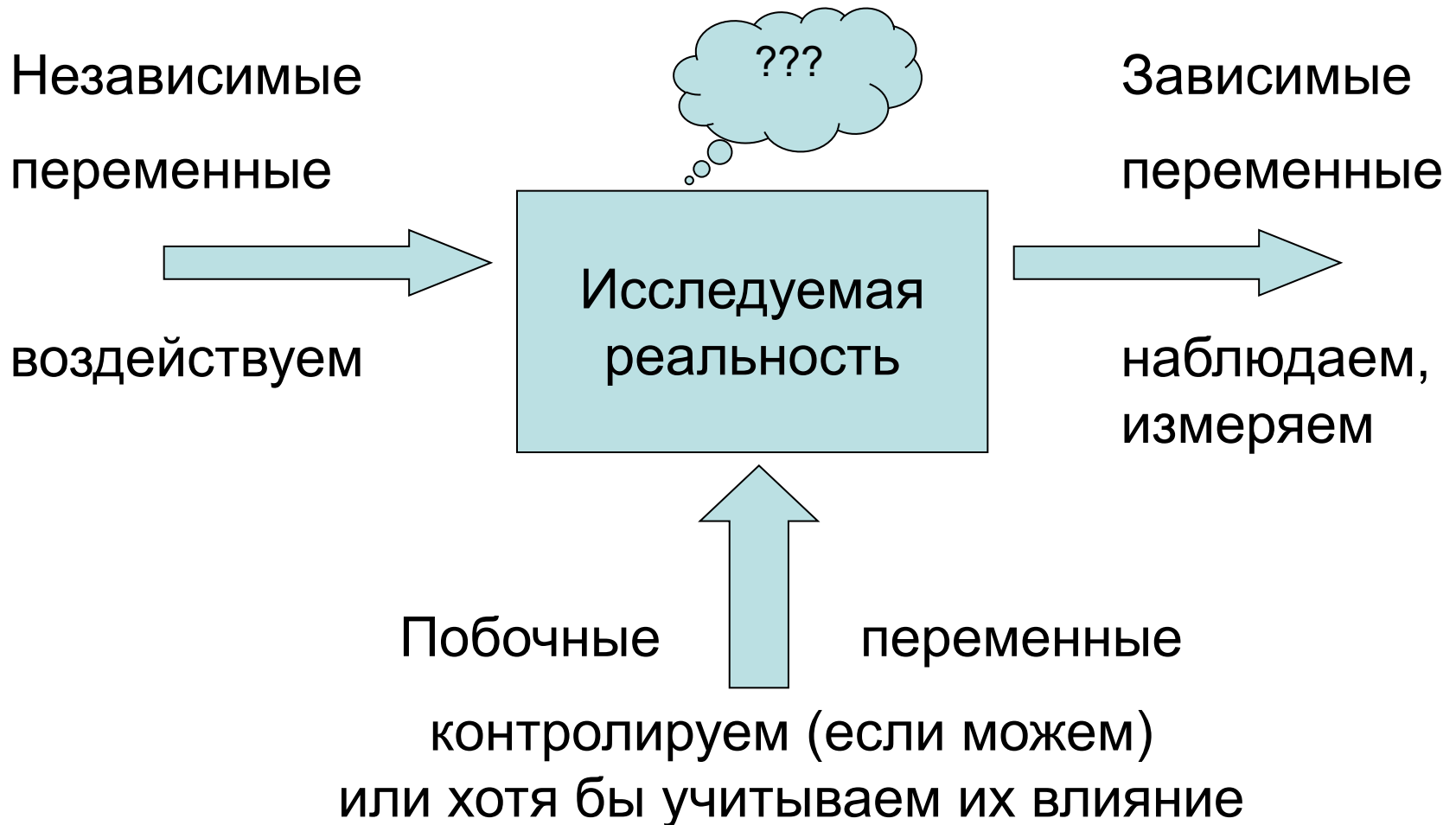
Вопросы о методе

- **Что и где** мы будем изучать?
 - Какие феномены? (сознания, поведения, ...)
 - Какие процедуры измерения? (□ тип данных)
 - В каких условиях?
 - На какой выборке?
- **Как** мы это будем изучать?
 - Каков общий план исследования?
 - Какие методы анализа данных?
- **Что** конкретно для этого **будем делать**?
 - Процедура исследования



ПЛАН ИССЛЕДОВАНИЯ

Виды переменных – по отношению к гипотезам



Случайная ошибка, систематическая ошибка, смешение

Типы исследований

- **Эксперимент:** есть воздействие, есть контроль побочных переменных □ можно делать выводы о причинно-следственных связях.
- **Квазиэксперимент:** есть воздействие, но нет полного контроля побочных переменных (в первую очередь, рандомизации испытуемых).
- **Неэкспериментальное эмпирическое исследование:** нет ни воздействия, ни контроля побочных переменных (коррел. исследование, лонгитюд и т.д.) □ причинно-следственный характер связей можно лишь предполагать.

Виды экспериментальных планов

- **Доэкспериментальные планы**
(для 1 независимой переменной)
- **Экспериментальные планы:**
 - *кросс-индивидуальные планы*, когда разные уровни НП предъявляются разным группам испытуемых:
 - для 1 НП;
 - для 2 и более НП: факторные планы $N \times M$;
 - *интраиндивидуальные планы*, когда разные уровни НП предъявляются одним и тем же испытуемым последовательно:
 - для 1 испытуемого.

Экспериментальные планы для одной независимой переменной (Д. Кэмпбелл)

- Обозначения (по Д. Кэмпбеллу):
- R – рандомизация
- O – измерение (observation) ЗП
- X – воздействие (наличие) НП

Взаимодействие
тестирования и
воздействия

Планы истинных экспериментов

План с предварительным и итоговым тестированием и контрольной группой

R O X O
R O O

+ + + + + + + +



- ? ?

План Соломона для 4-х групп

R O X O
R O O
R X O
R O

+ + + + + + + +

+ ? ?

План с контрольной группой и тестированием только после воздействия

R X O
R O

+ + + + + + + +

+ ? ?

Кросс- и интра-

- Кросс-индивидуальные планы включают несколько групп испытуемых (например, экспериментальная и контрольная).
- Группы должны быть **эквивалентными**. Для их формирования из выборки могут использоваться разные стратегии:
 - рандомизация (бросаем монету, в какую группу);
 - попарный отбор (сортируем людей по важному свойству и последовательно делим: 1-й – ЭГ, 2-й – КГ, 3-й – ЭГ и т.д.)
 - стратометрический отбор (подбираем испытуемых в группы так, чтобы уравнивать их по полу, возрасту и т.д.)
- Интраиндивидуальные планы включают одну группу испытуемых. Могут быть сложные смешанные планы.

Восемь угроз внутренней валидности по Д. Кэмпбеллу

- **Эффект истории (фона):** события между воздействием и измерением
- **Эффект созревания (естественного развития):** испытуемые со временем развиваются
- **Эффект тестирования** (первое измерение влияет на результаты последующих)
- **Погрешность измерения** зависимой переменной
- **Регрессия к среднему:** группы, отобранные как контрастные по какому-то показателю, при повторном измерении покажут результат близкий к среднему
- **Неэквивалентность групп:** в силу индивидуальных различий группы испытуемых изначально различаются
- **Эффект отсева или вымирания:** со временем часть испытуемых выбывает из исследования
- **Взаимодействие перечисленных факторов**

Четыре угрозы внешней валидности (репрезентативности) по Д. Кэмпбеллу

- **Влияние тестирования:** результаты эксперимента будут применимы лишь к людям, которые были протестированы
- **Взаимодействие отбора и экспериментального воздействия:** результаты применимы лишь к людям с определёнными характеристиками
- **Условия организации эксперимента** определяют реакцию испытуемых на эксперимент: результаты применимы лишь к людям в определённых условиях
- **Взаимная интерференция (взаимовлияние) экспериментальных воздействий**, если их несколько: результаты будут применимы лишь к людям, на которых воздействовали именно так (в таком порядке...)

Квазиэксперимент

- **Не полностью реализован экспериментальный контроль:** исследователь не полностью контролирует, когда и/или кому и/или в каком порядке предъявляются воздействия.
- **Оправдан в ситуациях, когда применение более совершенного (экспериментального) плана невозможно.**

Корреляционный план исследования

- **Задача – поиск взаимосвязей** (и проверка статистических гипотез о взаимосвязях) **между переменными** (двумя и более).
- **Нет воздействия** □ **нет «независимых» и «зависимых» переменных**, но эти слова могут употребляться для обозначения логики анализа (зависимость чего от чего мы предполагаем).
- **Нет возможности проверять гипотезы о причинно-следственном характере** наблюдаемых взаимосвязей □ **говорим не о «влиянии», а только о «взаимосвязи»!**

Планы корреляционных исследований

- **Межгрупповые:**
 - сравнение выраженности признака(ов) или их взаимосвязей в двух (и более) группах;
- **Внутригрупповые (repeated measures):**
 - план «одна группа в разных условиях» или в разные моменты времени.

Анализ динамических процессов в рамках корреляционного плана

Метод [поперечных] срезов: замер в один и тот же момент времени у разных когорт (возрастных групп, классов и пр.). Основная проблема: неэквивалентность групп (когортные различия).

+

Метод последовательных срезов (лонгитюд): одна и та же группа, замеры в разные моменты времени. Основная проблема: эффект тестирования и проблемы с внешней валидностью.

=

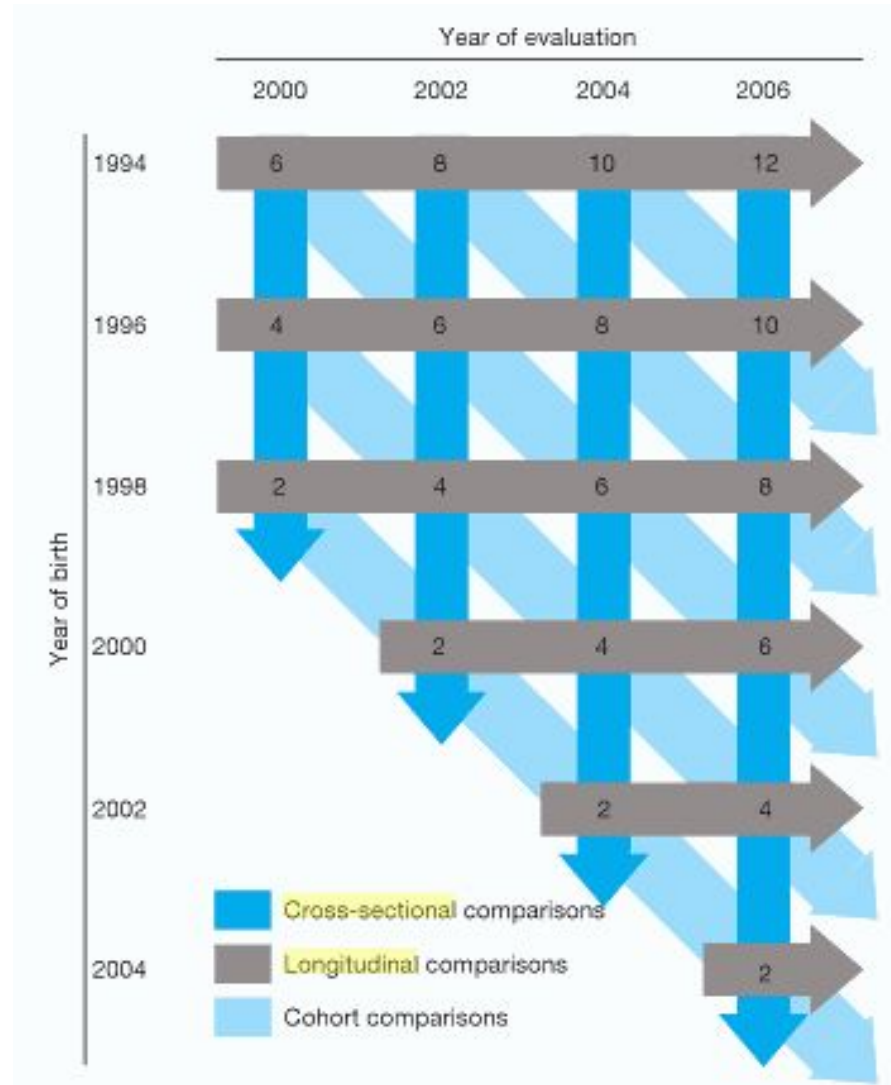
Метод поперечно-последовательных срезов: несколько разных когорт отслеживается в разные моменты времени.

Метод поперечно-последовательных срезов

Соединение поперечной и последовательной стратегий позволяет исследователям выявлять как различия, связанные с возрастом, так и когортные различия.

Цифры на пересечении года замера и года рождения обозначают возраст детей.

(из Parke & Clarke-Stewart, 2011)

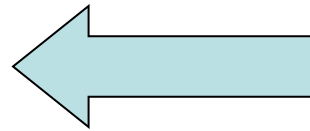


Выбор плана

- Выделены ли в гипотезах независимые и зависимые переменные? Говорят ли гипотезы о причинно-следственной связи явлений?
- Возможен ли истинный эксперимент и реализуем ли он на практике, с учётом ограниченных ресурсов?
- Какие угрозы внешней и внутренней валидности выводов связаны с выбранным вами планом? Есть ли возможности их контролировать?
- Какого рода исследования «ценятся» научным сообществом в данной области знаний?

ПЕРЕМЕННЫЕ

Измерение



И. – процедура приписывания психологическим объектам чисел таким образом, чтобы отношения между числами соответствовали отношениям между психологическими объектами.

Специфика психологических измерений



Не всякая операция, применимая к числам, имеет смысл по отношению к исходным психологическим объектам нужны некоторые ограничения.

Виды шкал

- **номинативная**
(=)
- **порядковая**
(=, >)
- **интервальная**
(=, <>, 1)
- **отношений**
(=, <>, 1, 0)



Особенности психологических измерений

- Измеряем ненаблюдаемые объекты озабочены валидностью наших измерений: то ли мы меряем?

инструмент
физика:



инструмент
психолога:



- В отличие от физических измерений, в психологии:
 - субъективные шкалы оценки ненадежны;
 - переменные сложны, их трудно операционализировать.
- нужно иметь много линеек = пунктов в шкале
- как понять, что вместе они вообще что-то меряют?

Надёжность измерения

- **Надёжность измерения – это доля дисперсии, связанной с измеряемым свойством, в дисперсии полученных значений (баллов по тесту).**
- **Если измерение ненадёжно, значит, шкала теста или опросника не позволяет измерять точно.**
- **Классическая тестовая теория альфа Кронбаха как показатель надёжности (недооценивает надёжность).**

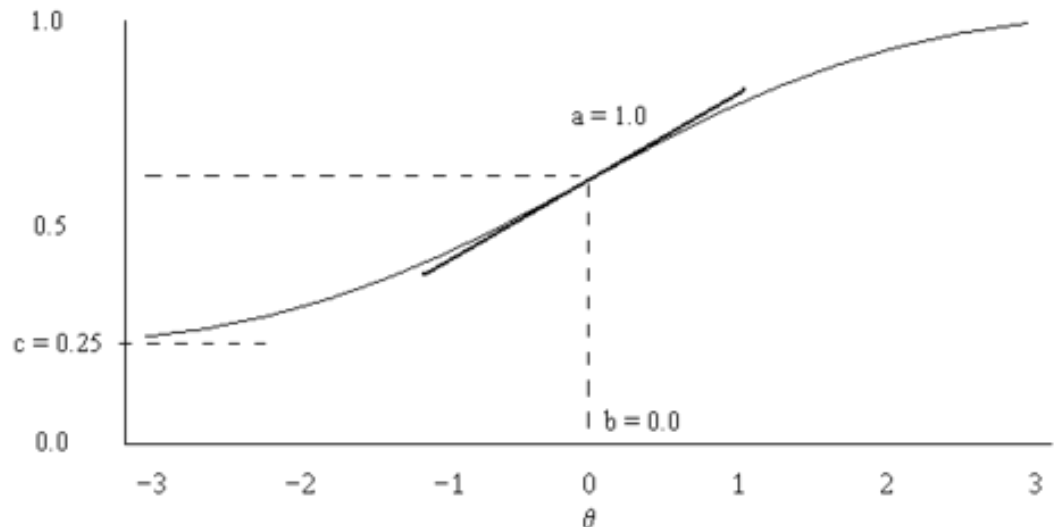
Связь надёжности теста с его
стандартной ошибкой измерения

Некоторые проблемы классической тестовой теории (СТТ)

- СТТ исходит из того, что ошибка не связана с тестовым баллом. Но на самом деле **точность измерения с использованием полученной с помощью СТТ шкалы неравномерна**: она максимальна для средних баллов и снижается по мере удаления от них (т.е., связана с тестовым баллом).
- На основе СТТ мы просто складываем баллы по пунктам теста, но на самом деле **у разных пунктов разная дисперсия ошибки**: умножая их перед сложением на некоторый коэффициент качества, можно повысить общую надежность шкалы.
- СТТ (и альфа коэффициент) исходит из того, что вся общая дисперсия пунктов – это true score. Но на самом деле **не вся общая дисперсия пунктов связана с измеряемым свойством** (например: человек на все вопросы, независимо от их содержания, отвечает «да») => необходимо различать разные виды ошибки.

Item Response Theory: современная альтернатива СТТ

- «Неклассическая тестовая теория» (IRT) опирается на более сложные регрессионные модели и преодолевает ряд ограничений СТТ.
- Моделируются различные параметры пункта, определяющие ответ респондента на пункт (сложность, дискриминативность, «прозрачность» к угадыванию).



Достоинства IRT по сравнению с СТТ

- Можно отбирать в шкалу пункты, дающие высокую **точность измерения в различных диапазонах измеряемого конструкта** (СТТ: точность максимальна возле среднего и сильно снижается по мере отдаления от него).
- Можно **подбирать для каждого респондента пункты**, дающие наиболее высокую точность в его диапазоне конструкта (например, для его уровня способностей), не утрачивая сопоставимости с баллами других респондентов (СТТ: для сопоставления баллов каждый должен выполнить один и тот же набор заданий) => можно создавать компьютеризованные **адаптивные тесты**.
- => IRT даёт **повышение точности и эффективности** измерения (уменьшение необходимого количества заданий). Но при этом у IRT более высокие требования к объёму выборки.

Методы оценки надёжности

Одномоментная надёжность:

- split-half reliability: делим тест пополам и считаем корреляцию между половинами теста = устаревший метод;
- классическая тестовая теория => внутренняя согласованность теста, альфа-коэффициент Кронбаха;
- более современные коэффициенты на основе данных Item Response Theory, конфирматорного факторного анализа (по Райкова).

Ретестовая надёжность:

- корреляция между результатами 2 замеров с помощью теста (с интервалом от 2 недель до года): метод пригоден, если измеряемое свойство не меняется (бесполезен, например, для тестов эмоций).

Разновидности ошибки

- **Случайная ошибка** (random error): доля балла, не связанная с измеряемым свойством, различная для каждого пункта теста.
- **Систематическая ошибка** (bias): доля балла, не связанная с измеряемым свойством, но не случайная (действующая сходным образом на разные пункты).
- Случайная ошибка угрожает **надёжности** измерения, систематическая ошибка – скорее, его **валидности**:
 - надёжность измерения: насколько точно мы что-то померили?
 - валидность измерения: насколько мы померили то, что нам нужно – или что-то другое?

Виды валидности теста – 1

- **Конструктивная в.:** тот ли конструкт, который был теоретически заявлен и описан, измеряет наш тест?
 - можно рассматривать этот вид валидности как наиболее общий, а остальные виды валидности – как его подвиды.
- **Операциональная в.:** измеряем ли мы конструкт тем способом, который позволяет его измерить?
- **Конвергентная и дискриминантная (дивергентная) в.:** как результаты нашего метода измерения соотносятся с другими данными измерений того же конструкта и других конструктов:
 - для оценки этих видов в. Д. Кэмпбелл и Д. Фиске (1959) предложили тип исследования «*multitrait-multimethod*»: несколько конструктов x несколько методов измерения = матрица Multi-Trait-Multi-Method (MTMM).

Виды валидности теста – 2

- **Структурная в.:** соответствует ли структура связей между пунктами теста теоретически ожидаемой?
- **Критериальная в.:** как результаты теста соотносятся с объективными (или поведенческими) данными?
 - **Прогностическая (предиктивная) в.:** предсказывает ли наш показатель будущие объективные (или поведенческие) данные?
Иногда говорят о «дискриминативной» валидности: насколько хорошо тест различает представителей разных подгрупп. В нашей классификации это критериальная валидность.
- **Очевидная в. (face validity):** очевидно ли из пунктов теста, что они измеряют?
 - **Экспертная в.:** согласны ли эксперты в данной области с тем, что эти пункты позволяют измерить конструктор?

Валидность теста?

- Л. Кронбах: валидность – свойство не самого теста, а решений, принимаемых на его основе в определённом контексте (ситуации, задач и т.д.)
- Как правило, под «коэффициентом валидности теста» имеют в виду его критериальную валидность (корреляцию с каким-то принципиально важным критерием):
 - например, корреляция балла по тесту при отборе персонала с последующей успешностью человека в качестве сотрудника.
- **Даже тест с низкой валидностью может быть практически**

полезным:

пример: т.н. таблицы
Расселла-Тэйлора для
принятия решений при
отборе персонала
(1939) .

	Доля кандидатов, отбираемых на основе теста			
r (крит. вал.)	.10	.30	.50	.90
.00	.50	.50	.50	.50
.25	.67	.62	.58	.52
.50	.84	.74	.67	.54

В ячейках – доля отобранных кандидатов, которые окажутся успешными, при условии, что по умолчанию успешны 50% кандидатов.

Что угрожает валидности решений?

- [Случайная ошибка □] **низкая надёжность теста** (А.Г. Шмелёв: ненадёжный тест не может быть валидным)
- [Систематическая ошибка (смещение измеряемого эффекта с другими) □] **низкая валидность показателя**
- Ошибки при
 - применении (выборе) теста (не то!);
 - предъявлении теста (не так!);
 - интерпретации баллов по тесту (не об этом!).

Чтобы снизить случайную ошибку

(Guilford, 1959, Клайн, 1994):

- Вопросы (утверждения) должны быть понятными респондентам, не использовать сложной лексики;
- Утверждения должны быть недвусмысленными (содержать только одну мысль, а не две связанные);
- Утверждения должны быть одинаково применимы ко всем респондентам, независимо от гендера, соц.статуса и пр.;
- Вопросы должны быть конкретными (связанными с конкретным примером, а не заданными в общем);
- Вопросы не должны провоцировать у подавляющего большинства респондентов однозначный ответ (напр., «да»);
- Из вопросов (утверждений) не должно быть ясно, что они измеряют [ЕО: в наши дни от этой идеи отказались];
- Вопросы не стоит формулировать в относительных (субъективных) терминах (часто...редко, нравится...не нравится);
- + Стоит обратить внимание на порядок и количество вопросов для снижения позиционных эффектов и усталости респондентов.

Виды систематической ошибки (bias) в пунктах субъективной оценки

- Ошибка центрации (mid-point responding)
- Ошибка сгущения (extreme responding, ER)
- Склонность к согласию (acquiescence, AR)
- Случайные ответы (random responding, pattern responding, RR)
- Социальная желательность (social desirability, SD):
 - сознательное или неосознанное искажение ответов;
 - мотивация: быть хорошим (аффилиация) или быть крутым (достижение).

Борьба с bias в опросниках, (методиках самоотчета):

- **Балансирование** шкалы (прямые и обратные пункты – борьба с AR)
- **Оптимальный подбор числа категорий** для ответа (снижение ER)
- **Пункты с очевидными ответами**, напр. «Я умею читать» (выявление RR)
- **Меры** выявления, снижения и статистического контроля **эффектов социальной желательности**

Межгрупповая ошибка (bias)

- Если в исследовании используется несколько групп, причиной межгрупповой ошибки может быть:
 - неэквивалентность конструкта (в одной из культур такого понятия нет или оно связано с другими проявлениями);
 - неэквивалентность выборок (по демографии...);
 - неэквивалентность пунктов (отдельные пункты понимаются респондентами конкретной группы иначе).
- При использовании переводных инструментов в нескольких культурах необходимо оценить уровень их эквивалентности и вытекающие возможности и ограничения сопоставления данных.

Установление эквивалентности инструментов

- Уровни эквивалентности:
 - **эквивалентность конструкта** (структурная эквивалентность): одна и та же картина связей пунктов с латентными переменными (факторами) в двух культурах □ можно содержательно сопоставлять результаты;
 - **эквивалентность единицы измерения** (эквивалентность шкалы): добавляется требование равенства нагрузок пунктов на факторы в двух культурах □ можно сопоставлять корреляции шкалы с другими и стандартизованные баллы;
 - **эквивалентность сырых баллов** (полная эквивалентность): добавляется требование равенства остатков (и иногда дисперсий ошибки) пунктов в разных культурах □ можно сопоставлять сырые баллы.

Методология: конфирматорный факторный анализ, IRT.

Переменные

- Являются ли измерения **надёжными** и **валидными** (в данных условиях)?
- Если используется несколько групп, являются ли измерительные инструменты **эквивалентными**?

ВЫБОРКА

Создание выборки

- Определитесь, насколько важна репрезентативность.
- Выберите оптимальную стратегию подбора:
 - полностью случайный отбор (по списку);
 - стратифицированный отбор;
 - кластерный отбор;
 - смешанные (двухуровневые стратегии);
 - слабые: «снежный ком» / удобная выборка / добровольцы.
- Оцените потребный объем выборки с учётом:
 - количества переменных, характера гипотез и планируемых методов анализа данных;
 - размера интересующих вас эффектов и статистической мощности анализа.
- Рассмотрите возможность проведения исследования онлайн.

Статистическая мощность анализа

- **Уровень значимости:** вероятность того, что мы приняли гипотезу H_1 , которая на самом деле неверна.
- **Статистическая мощность анализа ($1 - \beta$):** вероятность того, что мы на выборке примем гипотезу H_1 , если на самом деле она верна (= шанс обнаружить эффект, если он на самом деле есть).

Уровень значимости

- **Достоверность взаимосвязи зависит:**
 - **от её силы** (чем больше r по модулю, тем больше шансов, что он будет значим);
 - **от объёма выборки** (чем больше выборка, тем больше шансов, что r будет значим).
- на очень маленьких выборках даже для сильных взаимосвязей значимость может не достичь приемлемого уровня;
- на очень больших выборках даже очень слабые (и потому практически бессмысленные) взаимосвязи могут оказаться значимыми.

Статистическая мощность анализа

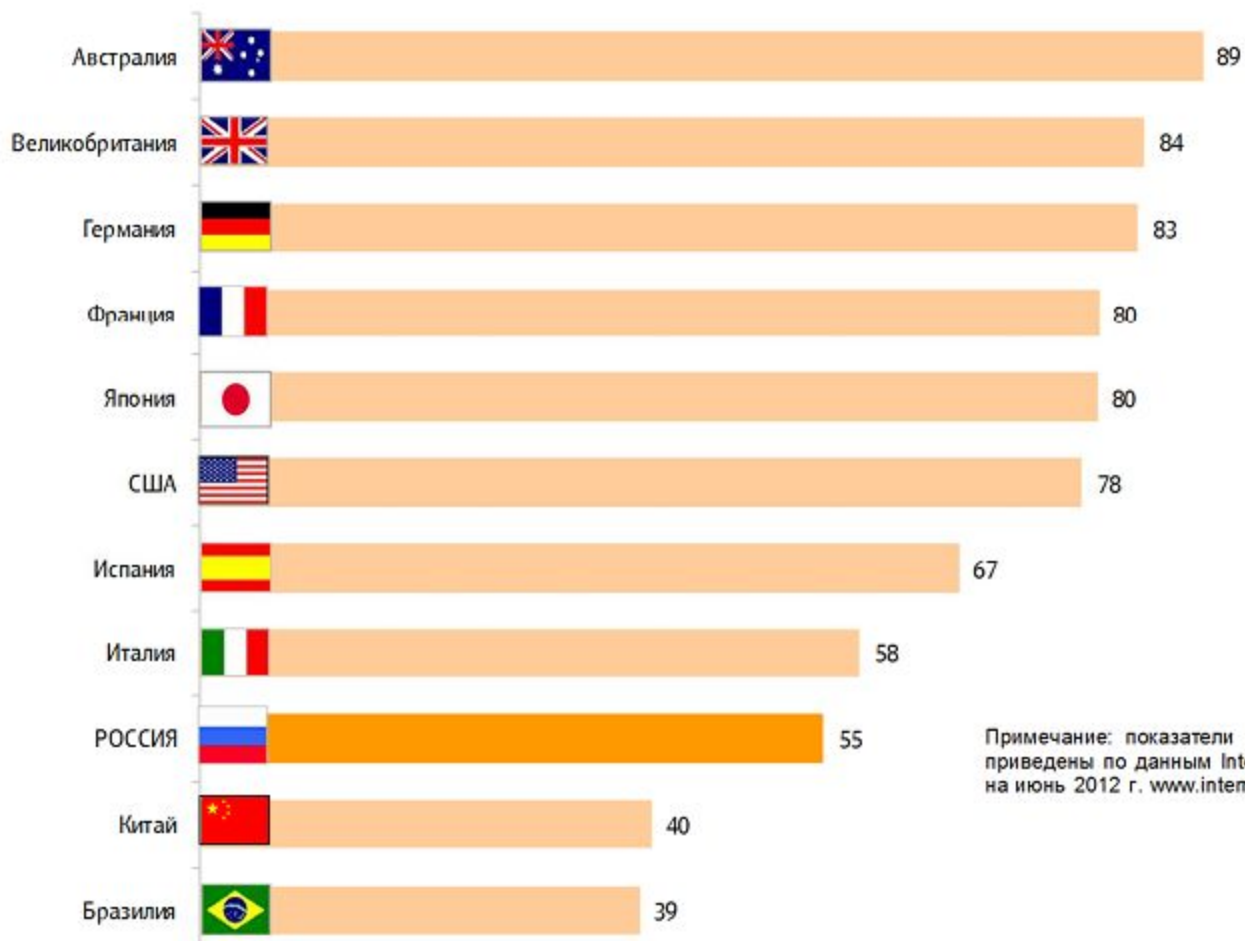
- Зависит от...
 - **объёма выборки**: чем он больше, тем она выше;
 - **размера эффекта**: чем он сильнее, тем она выше;
 - от выбранного **критерия принятия решений** о значимости: чем строже требование к уровню значимости, тем она ниже;
 - от используемого **статистического метода** (для разных способов проверки одной и той же гипотезы она м.б. разной).
- Является критерием для определения объёма выборки с учётом размера ожидаемого эффекта.
- Важно! Только высокая мощность (0,95 и выше) даёт нам возможность делать **достоверный** вывод о том, что **искомый эффект отсутствует** (верна H_0).
- При недостаточной статистической мощности подобный вывод является необоснованным (правильный вывод: мы не обнаружили эффект, но не можем сказать, есть он или нет).

Зависимость статистической мощности от размера выборки

Зависимость статистической МОЩНОСТИ ОТ СИЛЫ ВЗАИМОСВЯЗИ

Исследования в Интернет

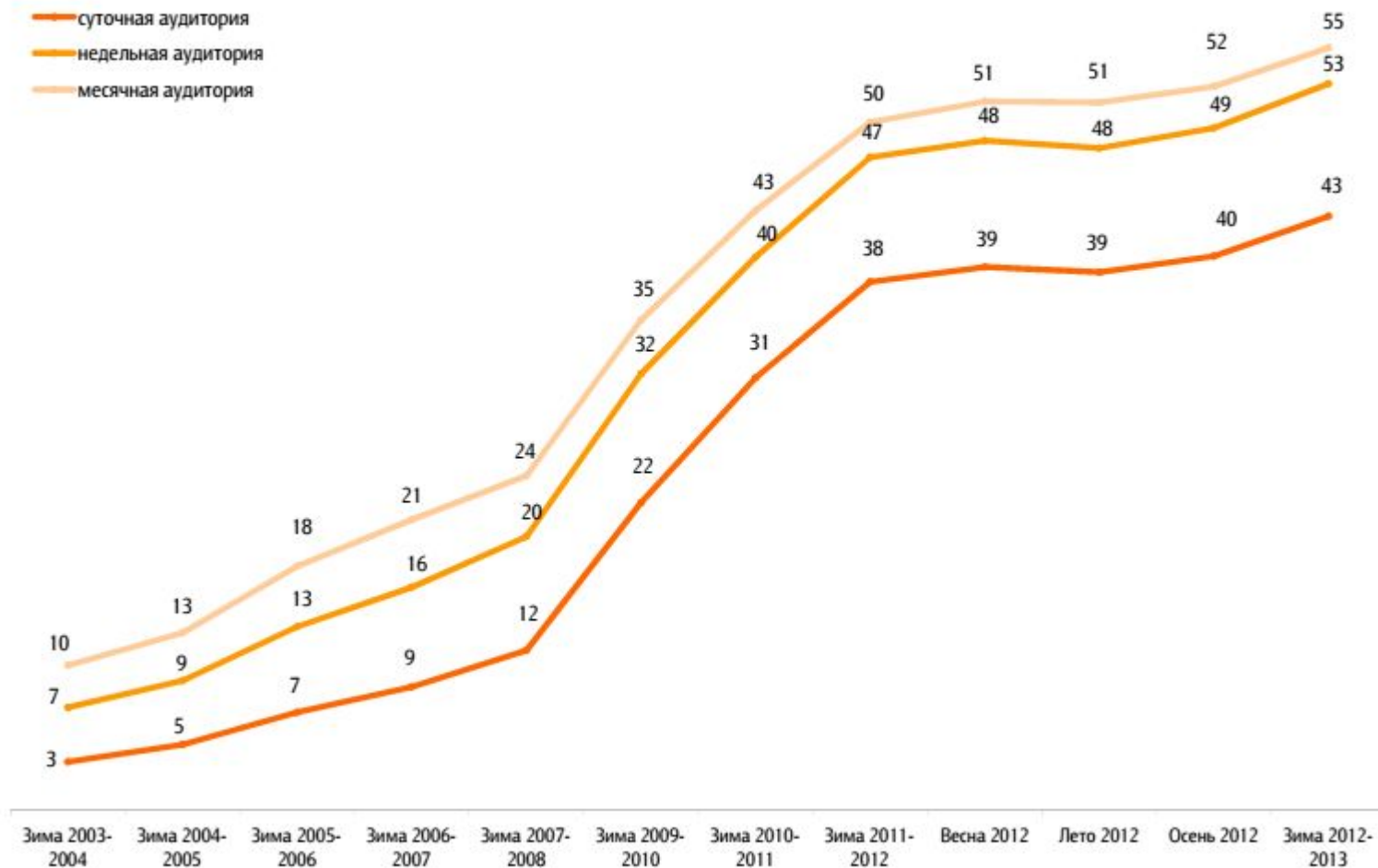
1.6. Проникновение Интернета в странах, % от населения



Примечание: показатели зарубежных стран приведены по данным Internet World Stats на июнь 2012 г. www.internetworldstats.com

Интернет в России

1.1. Численность интернет - пользователей. Динамика. Россия, 18 лет и старше, %



Данные ФОМ, весна 2013

Проблема репрезентативности

- Выборки из Интернет имеют ограниченную репрезентативность по отношению к некоторым социальным группам, в первую очередь, людям:
 - старших возрастов;
 - с невысоким уровнем образования;
 - с низким социально-экономическим статусом.
- Но для многих исследовательских задач это не является проблемой. Интернет-выборки принимаются в ведущих мировых журналах.
- Борьба с проблемой репрезентативности:
 - балансировка выборки путём взвешивания наблюдений;
 - проведение отдельного анализа на разных подвыборках (например, младшей и старшей сравнить результаты).

Выборка

- Достаточен ли **объём выборки**?
 - Каков ожидаемый размер эффекта, какой будет статистическая мощность?
- Достаточно ли выборка **репрезентативна** относительно группы, на которую обобщаются результаты?
- **Позволяет ли состав выборки** (подгруппы) проверять нужные гипотезы планируемыми методами?

АНАЛИЗ ДАННЫХ

Шаги выбора метода

- Определение набора исходных переменных и их места в анализе (есть ли независимые и зависимые).
- Определение шкалы, по которой они измерены и характера распределения (допустимы ли параметрические методы).
- Выбор оптимального метода из допустимых, с учётом объёма выборки.

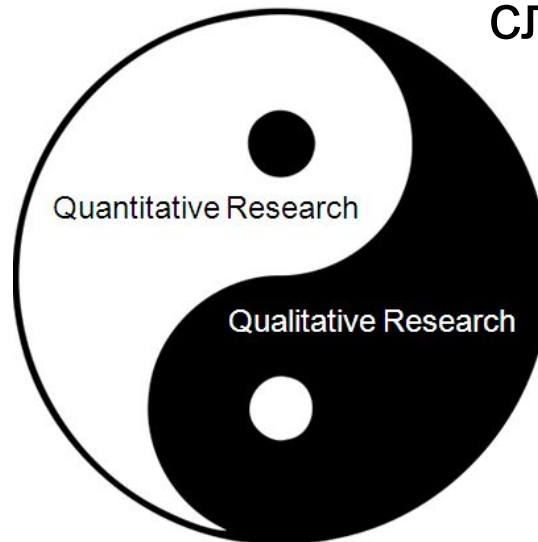
Два вида методов

- Количественные:

- выявление общих закономерностей и статистическая оценка достоверности обобщений;
- познание частного случая на основе общих закономерностей.

- Качественные:

- описание и анализ частного случая;
- познание общих закономерностей на основе частного случая.



Смешанная методология: зачем?

(Greene, Caracelli, Graham)

- **Триангуляция:** соответствие, подтверждение результатов одних методов другими.
- **Дополнение:** расширение, прояснение, иллюстрация результатов одних методов другими.
- **Развитие:** использование результатов одних методов для принятия решений о развитии дальнейших исследований другими методами (напр., решений о выборке, операционализации).
- **Инициация:** поиск парадоксов или противоречий, новых точек зрения благодаря пониманию результатов одних методов исходя из других.
- **Экспансия:** большая широта исследования благодаря использованию разных методов для решения разных отдельных вопросов.

Способы объединения качественных и количественных данных (Creswell)

Merge the data:



Connect the data:



Embed the data:

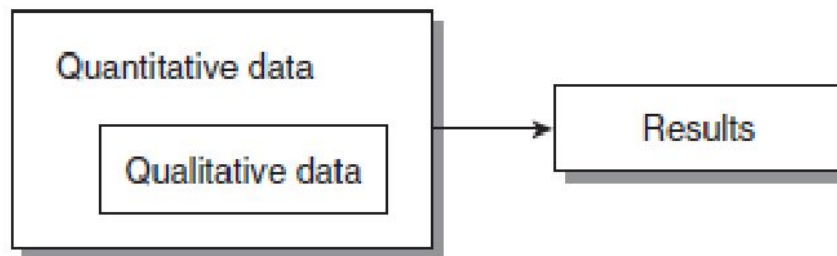


Figure 1.2 Three Ways of Mixing Quantitative and Qualitative Data

Примеры качественных методов

- Дескриптивный феноменологический анализ
- Интерпретативный феноменологический анализ
- «Тематический анализ», качественный контент-анализ
- Конверсационный анализ (conversation analysis)
- Дискурс-анализ (discourse analysis)
- Нарративный анализ и нарративное интервью
- Фокус-группы (focus groups)
- Обоснованная теория (grounded theory)
- Исследование действием (action research)

Процедуры повышения валидности качественных данных (Yardley)

- **Триангуляция:** сопоставление данных, полученных разными методами или от разных источников.
- **Экспертное согласие:** сравнение данных (картин реальности), полученных разными исследователями (экспертами).
- **Обращение к респондентам:** респонденты выступают в качестве экспертов, которые оценивают валидность полученной картины реальности.
- **Анализ выпадающих случаев:** выявление наблюдений (респондентов), не согласующихся с общей картиной.
- **«Бумажный след» (paper trail):** документированы все этапы анализа □ можно проследить его ход.
- **Рефлексия** собственных ценностей, ожиданий, представлений исследователя об изучаемой реальности.

Виды количественных методов

- Описательные статистики
- Методы проверки гипотез:
 - гипотезы о соответствии теоретическому распределению;
 - гипотезы о различиях эмпирических распределений критерии сравнения выборок;
 - гипотезы о связях переменных корреляционный анализ, регрессионный анализ, ...
- Методы поиска взаимосвязей:
 - на наблюдаемых переменных: путь анализ, многомерное шкалирование, эксплораторный факторный анализ, кластерный анализ, ...;
 - моделирование латентных переменных: структурное моделирование, IRT, анализ латентных классов, ...

Проблемы статистических методов

- Статистические методы исходят из того, что наша **выборка является случайной**. Если она не такова, статистические методы могут **вводить нас в заблуждение**.
- Содержательная валидность выводов всех корреляционных методов зиждется на **отсутствии неизмеренной третьей переменной**, определяющей наши связи.
- Возможность **артефактов**.

Проблема множественных сравнений

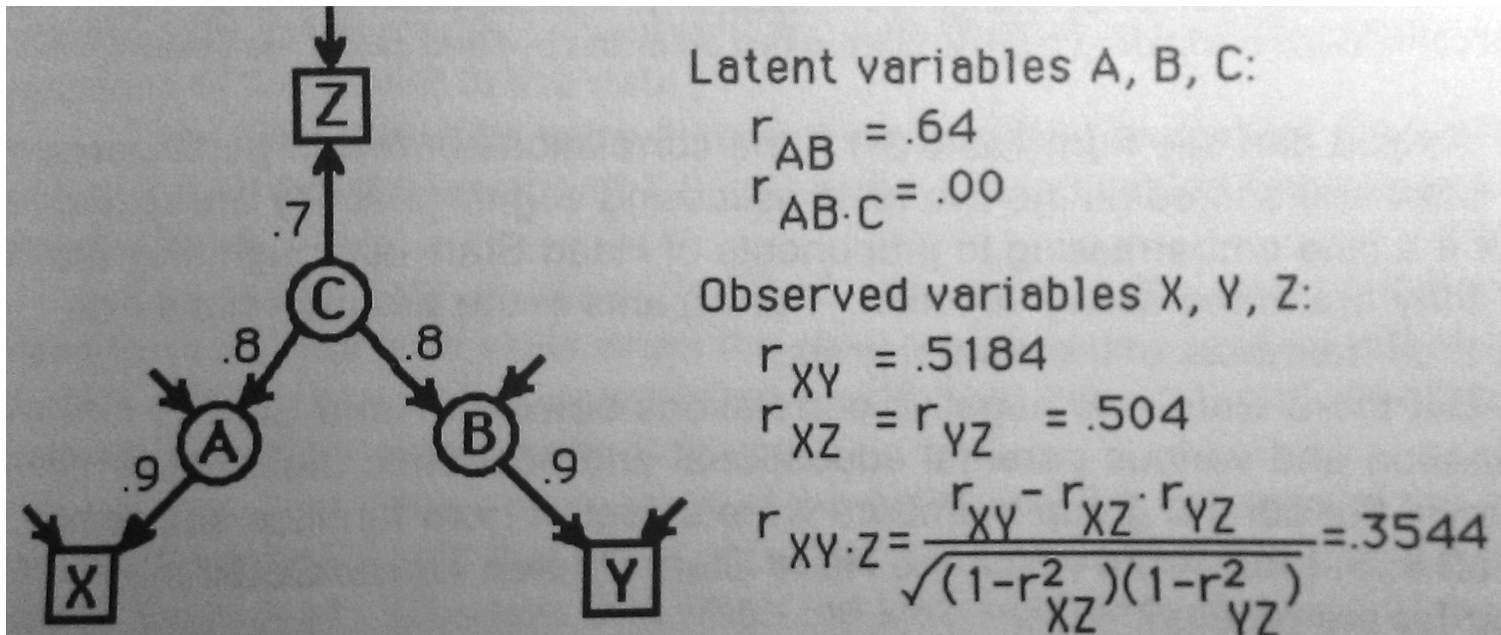
- Уровень значимости = 0,05: шанс того, что такая корреляция получена случайно, составляет 1 из 20.
- Если у вас в матрице 20 корреляций, значимых на уровне 0,05, то в среднем одна из них окажется случайной.
- Чем больше статистических гипотез вы проверяете, тем более жёсткий критерий принятия решения (уровень значимости 0,01 или даже 0,001) имеет смысл брать:
 - существуют поправки для множественных сравнений, например, поправка Holme-Bonferroni;
 - при простом выборе более жёсткого критерия снижается статистическая мощность анализа.

Другие возможные артефакты

- Корреляция двух переменных может объясняться **влиянием третьей переменной** (пример: у детей размер ноги и IQ связаны друг с другом, а реальная причина – связь обеих переменных с возрастом) □ расчёт частной корреляции.
- **Объединение двух выборок** с различающимися средними по обеим переменным может приводить к возникновению ложных корреляций на объединённой выборке □ проверять наличие различий в средних и в характере и степени взаимосвязи перед объединением.
- **Наличие выбросов** (наблюдений, существенно отклоняющихся от общих закономерностей) в данных на небольших выборках может приводить к появлению ложных корреляций или к снижению значений коэффициента корреляции □ проверка на нормальность, анализ графика разброса.
- **Низкая надёжность** наблюдаемых переменных приводит к снижению коэффициента корреляции □ коррекция аттенюации.

Источники артефактов

- В исследовании не измерены (или не включены в модель) переменные, являющиеся общими причинами моделируемых переменных (□ «spurious correlations»).
- Низкая надёжность измерений (Loehlin, 1998) – пример с частной корреляцией:



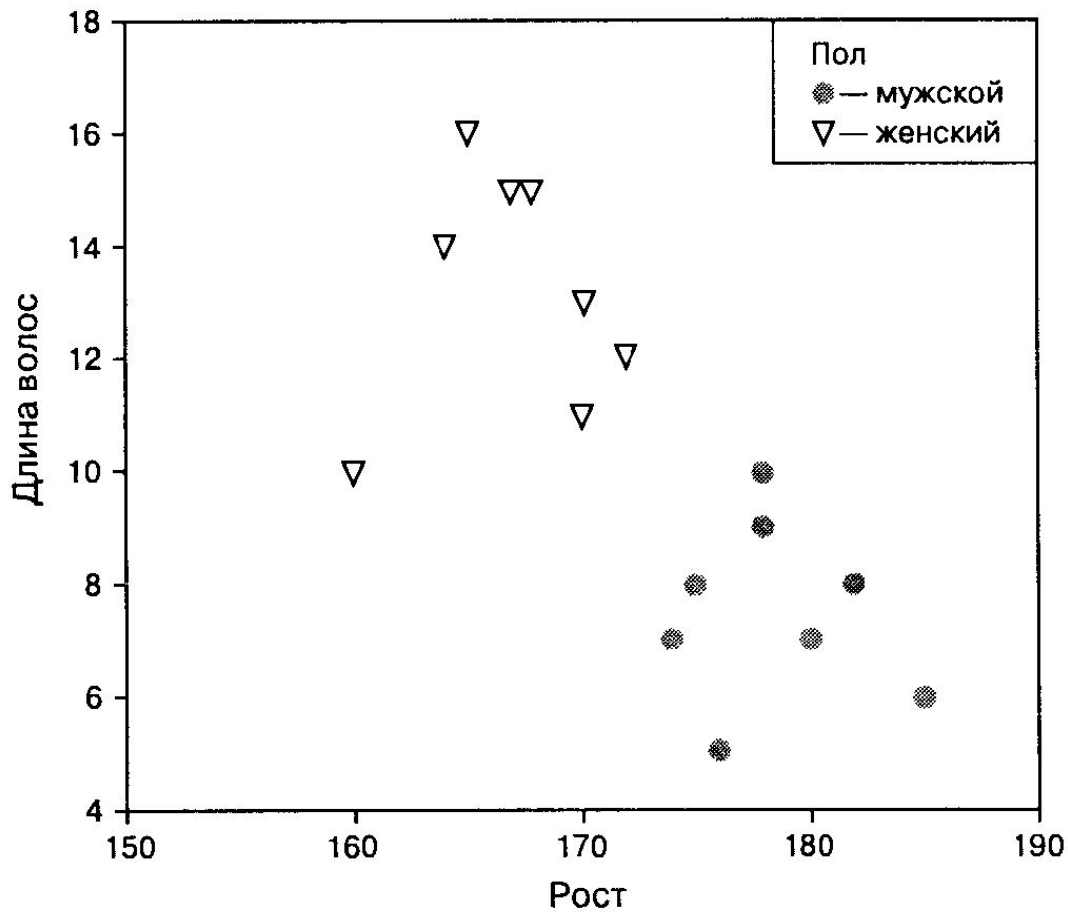
Источники артефактов: N выборки

df	N			N needed for power of .80	N needed for power of .90
	100	300	500		
1	.07	.15	.26	2475	3427
2	.10	.27	.43	1289	1763
3	.13	.36	.56	891	1206
4	.15	.44	.66	690	927
5	.18	.51	.75	569	758
6	.20	.58	.81	487	646
7	.22	.64	.86	428	564
8	.24	.69	.90	384	503
9	.26	.73	.93	349	455
10	.28	.77	.95	320	417
20	.46	.96	~1.00	188	239
30	.61	.99	~1.00	141	177
40	.72	~1.00	~1.00	115	144
50	.81	~1.00	~1.00	99	123
60	.87	~1.00	~1.00	88	109
70	.91	~1.00	~1.00	80	98
80	.94	~1.00	~1.00	73	90
90	.96	~1.00	~1.00	68	83
100	.97	~1.00	~1.00	64	78

Note: Lefthand columns: the power to reject the hypothesis of poor fit (RMSEA \geq .10), given population RMSEA of .05, for indicated sample size (top) and df (side). Assumes .05 significance level. Righthand columns: Ns required for powers of .80 and .90 for this test. Follows method of MacCallum, Browne, & Sugawara (1996).

- На малой выборке может не хватить статист. мощности для отвержения плохой модели.
- На малой выборке оценки параметров неточные, значимость параметров низкая.
- Стоит учитывать доверительные интервалы для статистик.

Артефакт объединения двух выборок (Наследов, 2004)



В каждой из выборок по отдельности корреляция отсутствует.

В объединённой выборке она наблюдается.

Многоуровневый анализ

- Кросс-культурное исследование □ пример многоуровневых данных, когда в рамках общей выборки есть группы наблюдений (индивидов), свойства которых связаны друг с другом.
- **Обычные статистические методы исходят из того, что наблюдения (= чаще всего индивиды) попарно независимы, и применение этих методов к подобным выборкам будет приводить к артефактам.**
- В таких случаях необходим многоуровневый анализ.

Виды пропущенных данных

- **MCAR (Missing Completely At Random):** пропущенность ни от чего не зависит (на практике это бывает редко).
- **MAR (Missing At Random):** пропущенность зависит от значений других измеренных переменных (x-variables).
- **NMAR (Not Missing At Random):** пропущенность зависит от значений неизмеренных переменных, значения которых мы пытаемся оценить (y-variables).

Примеры:

- MCAR – некоторые люди случайно отвлеклись при заполнении теста или не пришли на один из замеров;
- MAR – тест не заполнили некоторые люди с низкой Conscientiousness, которую мы померили, или на пост-тест не пришли люди с высокими показателями претеста;
- NMAR – на пост-тест не пришли люди, которым не помог тренинг, эффективность которого нас интересует.

Работа с пропущенными данными: 1. Выявление

- Подсчитать долю пропущенных данных: чем она больше, тем важнее с ними правильно обойтись при анализе.
- Проверка допущения MCAR:
Little's test в SPSS (значим => не MCAR).
- Проверка MAR:
 - 1) создаём для каждой переменной dummy variable, обозначающую пропущенность (напр., 1=значение пропущено, 0=значение есть);
 - 2) смотрим, есть ли значимые корреляции между переменными пропущенности (если да, то какова их факторная структура);
 - 3) смотрим, коррелирует ли пропущенность со значениями измеренных переменных (если да, то имеет место как минимум MAR, или есть какая-то 3-я переменная, которая может определять это всё – надо это теоретически осмыслить).

Работа с пропущенными данными: 2. Коррекция

- Слабые методы, пригодные при условии MCAR и/или небольшом количестве пропущенных значений:
 - **casewise/listwise deletion** □ теряется много данных;
 - **pairwise matrix** □ хорош только если пропущенных очень мало;
 - **mean imputation**: замена пропущенных значений средними □ уменьшает дисперсию, крайне не рекомендуется;
 - **regression imputation (факультативно: +error term)**, или предсказание пропущенных с введением ошибки (факультативно) □ не так уж плохо, если очень нужно.
- Сильные методы, пригодные при MAR:
 - **EM imputation**: итеративная регрессия (пропущенные значения каждой переменной восстанавливаются на основе всех остальных по кругу, пока алгоритм не сойдётся на стабильных значениях) – можно делать в SPSS;
 - более сложные:
 - **Multiple Imputation** (генерируется несколько наборов данных, где пропущенные заменены ожидаемыми случайными значениями из распределений, результаты оценки моделей сравниваются).
 - **Full-Information Maximum Likelihood** (в структурном моделировании): метод максимального правдоподобия с использованием только имеющейся информации, при этом пропущенные данные не заменяются как таковые:
- Если NMAR: использовать сильные методы и думать содержательно о том, почему данные пропущены.

Работа с пропущенными данными

- Сильные методы, пригодные при MAR:
 - **EM imputation**: итеративная регрессия (пропущенные значения каждой переменной восстанавливаются на основе всех остальных по кругу, пока алгоритм не сойдётся на стабильных значениях) – можно делать в SPSS;
 - **MI imputation**: генерируется несколько наборов данных, где пропущенные заменены ожидаемыми случайными значениями из распределений, результаты проверки модели сравниваются (Mplus с Estimator=BAYES – годится для *любых* распределений).
 - **Full-Information Maximum Likelihood**: метод максимального правдоподобия с использованием только имеющейся информации, при этом пропущенные данные не заменяются как таковые:
 - EQS: надо создать dummy variable V999 с нагрузками на все переменные
 - Mplus: используется по умолчанию при методах оценки ML/MLR (WLSMV всегда исходит из допущения MCAR!)
- Если NMAR: использовать сильные методы и думать содержательно о том, почему данные пропущены.

Кросс-валидизация

- Если модель эксплораторная (построена отталкиваясь от данных), **её кросс-валидизация необходима** (иначе модель может включать артефакты = особенности конкретной выборки).
- Самый простой способ – делим выборку случайно пополам, на одной половине строим модель, на другой – проверяем.
- Для некоторых методов возможен bootstrapping (из выборки выбирается много случайных подвыборок более точные оценки ошибок и эффектов).

«Попытка **подогнать цифру** под то или другое **предвзятое** мнение есть **преступление** уголовного характера» -- И. В. Сталин



(А попытка представить модель, полученную в результате долгих мучений с данными, как исходную теоретически обоснованную гипотезу – всего лишь мелкое научное жульничество) – ЕО

Советы В. М. Аллаhverдова

(2005: «Блеск и нищета эмпирической психологии»)

- «Осмысленность вычислений статистических параметров не определяется используемыми математическими методами, **правомерность применения математического аппарата должна специально содержательно обосновываться и проверяться**».
- **Любое обобщение** полученных эмпирических результатов является **внеэмпирической** интерпретацией и должно **независимо проверяться**.
- **Фиксируйте алгоритм обработки данных до начала анализа.** Если в ходе анализа выяснится, что другой алгоритм работает лучше, применяйте его ко всем данным последовательно.
- Из всех **способов обработки данных начинайте с простых** (например, описательных статистик, анализа распределений, корреляций, сравнения средних), чтобы лучше понять ваши данные, и только потом переходите к более сложным.

Проблема факторного анализа

«Проблема заключается в том, что ФА часто используется в ходе попыток “спасти” плохо спланированное исследование. В ситуациях, когда иные статистические процедуры не применимы, данные по крайней мере можно подвергнуть факторному анализу. Таким образом, в сознании многих многочисленные варианты ФА ассоциируются с сырыми исследованиями.

Способность ФА и АГК создавать видимость порядка там, где реально имеет место хаос, вносит свой вклад в их подмоченную репутацию в качестве инструментов научного исследования»

(Tabachnik & Fidell, 2007, p. 608-609)



Барбара Табачник

Бывший профессиональный танцор живота, выдающийся художник и писатель, почётный профессор California State University, автор учебника Using Multivariate Statistics (в соавторстве с Линдой Фиделл).

Выбор метода

- Чем обоснован именно такой выбор методов?
- Соответствуют ли друг другу результаты разных методов?
- Как теоретически обоснована выбранная математическая модель?
- Соответствуют ли результаты полученным на других выборках?
- Каковы возможные источники артефактов, какие шаги предприняты для борьбы с ними, каковы возможные погрешности?

НАПИСАНИЕ ТЕКСТА

Виды текстов по содержанию

- Эмпирические исследования
 - Количественные
 - Качественные
- Обзоры литературы
- Теоретические статьи
- Методологические статьи
- Case study

Структура статьи с эмпирическим исследованием (по АРА)

- Название (title)
- Информация об авторах, аффилиация
- Аннотация (abstract) 150-250 слов
- Введение (introduction)
- Методы (methods)
 - Aim, Design, Instruments, Procedure, Sample
- Результаты (results)
- Обсуждение (discussion)
- Литература (references)
- Приложения (appendices)

Представление эмпирических данных по стандартам APA

- Стандартный вид таблиц данных для наиболее популярных матметодов = легко разобратся.
- Давать не только результат проверки значимости, но и точную информацию о размере эффекта (для мета-анализа):
 - например, для корреляции $Z = r * \sqrt{N}$, где Z – значение нормального распределения, соответствующее уровню значимости, N – количество наблюдений в выборке; соответственно, в публикации должны быть представлены как минимум 2 из 3 членов уравнения, чтобы при мета-анализе можно было рассчитать недостающий.

Советы по представлению данных эмпирического исследования

- Давать в тексте полную информацию обо всех действиях по сбору и обработке данных, чтобы процедуру можно было повторить
- Давать в тексте полную информацию о результатах: уровень значимости и размер эффекта
- Представлять количественные данные в общепринятой форме (таблицы по стандартам APA)
- Не перегружать текст результатами: лучше выбрать самое важное и представить это как следует
- Можно разбить большое исследование на несколько публикаций, но нужно чётко указать, как они соотносятся друг с другом (не должны повторять)

Этические стандарты

- Используются ли процедуры защиты прав респондентов? Рекомендуется:
 - право на добровольное участие информированное согласие;
 - право на информацию, благополучие дебрифинг;
 - право частной жизни конфиденциальность, защита личных данных.
- Нет ли нарушений научной этики? Необходимо:
 - убедиться, что не допущен плагиат;
 - верное указание авторства, аффилиаций.
- Нет ли нарушений авторских прав? Необходимо:
 - убедиться, что есть разрешение на использование чужих инструментов, примеров и пр.
- Нужно/есть ли одобрение этической комиссии (IRB)?

Только один вопрос

- Выглядит ли моя статья (по структуре, оформлению) так же, как действительно хорошие статьи в действительно хороших журналах?

ГДЕ ПУБЛИКОВАТЬ?

Куда податься?

- Статьи в рецензируемых журналах
- Статьи в нерецензируемых журналах, в сборниках статей
- Авторские монографии
- Тезисы в сборниках тезисов
- Отчёты о научно-исследовательской работе
- Препринты, неопубликованные рукописи



Качество публикаций

- Определяется не статусом и авторитетом автора, а качеством экспертизы (рецензирования) материала:
 - статью в рецензируемом журнале или диссертацию оценивают как минимум 3 человека (автор + 2 и более рецензента);
 - статью в нерецензируемом журнале оценивает 1 человек (редактор);
 - научную монографию в хорошем издательстве оценивают несколько человек (рецензенты), в плохом – никто;
 - учебник – как повезёт.

Рецензирование

- Западные журналы peer-review: 2-4 рецензента + редактор. От 2 недель до 3 месяцев, 1-3 варианта.
- Российские журналы ВАК: чаще всего 1 рецензент + редактор (или даже только редактор). Около 3 месяцев, чаще всего 1 итерация.

Оценка

- Accept
- Accept with Minor Revisions
- R & R (Revise and Resubmit)
- Major Revision
- Reject

Если 2 рецензента не согласны, 3-ю рецензию даёт редактор.

Редактор может также сразу отвергнуть статью без рецензирования.



- Reject – не приговор статье, а всего лишь пессимистическое **мнение** редактора относительно **возможности** и **желания** автора привести статью в приемлемый **для данного журнала** вид за **относительно небольшой** промежуток времени.
- Даже безнадежно плохую статью часто можно «спасти», вопрос в том, ценой каких усилий.

Возражения рецензентов

- Рецензенты и редактор хотят помочь вам улучшить вашу статью.
- У них может быть своя, узкая точка зрения на вашу предметную область.
- Вы можете внести в статью изменения или мягко поспорить с возражениями по отдельным пунктам, хорошо аргументируя свою позицию ссылками.

Выбор журнала

- Поиск подходящих журналов:
 - Индексы цитирования: Impact Factors по Web of Science / Scopus, РИНЦ
- Определить круг возможных журналов, примерно ознакомиться с их стандартами, почитать типичные статьи, оценить свои шансы с учётом их rejection rate
- Выбрать приоритетный журнал, подробно ознакомиться со стандартами оформления

Цитирование

- Чем больше людей прочтут вашу статью, тем больше шансов, что её процитируют.
- Для этого она должна быть в достаточно хорошем и достаточно доступном журнале.
- Легче всего найти статью в журнале открытого доступа (open-access).
- Главы в монографиях цитируются хуже.

Типичные ошибки

- Статья подаётся в журнал, для которого она не подходит по тематике или типу исследований (обзоры, эксперименты, серии экспериментов, репликации, корреляционные исследования...)
- Статья подаётся в журнал слишком высокого уровня (но часто ничего не теряем, т.к. в таких журналах обычно быстрый отказ).
- Статья не оформлена по стандартам конкретного журнала.

Оценка своего исследования

- Публикуются ли исследования с таким планом, выборкой, результатами в журнале такого уровня?
- Убедиться, что ваш математический анализ проведён эффективно и валидно.
- Чётко определить те исследования и результаты, которые вы хотите представить.
- Можно определить общую идею и выстроить статью вокруг неё.



Типичные ошибки авторов



Типичные ошибки авторов

- Автор «изобрёл велосипед»: решает проблему 1960-х годов методами 1930-х
- Автор не рефлексирует сущностный характер того, что он изучает, или множит сущности
- Автор строит теорию, которая становится для него самоцелью
- Автор не утруждает себя идентификацией своего места в контексте науки и не задумывается о том, понятен ли его текст



Типичные ошибки авторов

- Автор не осознаёт ограничения выбранного плана исследования и не думает о других возможностях
- Автор не задумывается о качестве и адаптации измерительных инструментов
- Автор не задумывается о необходимости валидации своих экспертных процедур
- Автор использует заведомо неадекватную целям или нерепрезентативную выборку



Типичные ошибки авторов

- Автор не описывает процедуру сбора данных
- Автор недостаточно полно описывает данные и/или свои шаги по их анализу
- Автор ненавидит статистику и использует одни лишь корреляции или вообще ничего
- Автор не задумывается о статистической мощности своего анализа
- Автор забывает об условиях применимости матметодов



Типичные ошибки авторов

- Автор думает, что хорошая статистика гарантирует достоверные выводы
- Автор не осознаёт или умышленно не обсуждает ограничений своего исследования
- Автор не стремится получить обратную связь о своих результатах от коллег до публикации

Sternberg: Критерии качества теорий

- Clarity and Detail
- Original Substantive Contribution (8 видов)
- Relation to Past Work
- Falsifiability
- Generalizability
- Discriminability
- Internal Consistency
- Correspondence to Past Data
- Prediction
- Parsimony
- Excitement

Sternberg & Grigorenko: Методы

- Выборка
 - Достаточный ли объём?
 - Подходящая ли популяция использована?
 - Сбалансирована ли выборка по демографии?
 - Для лонгитюдов: есть ли выпадение, сторонние эффекты смешения?
 - Для срезов: сравнимы ли когорты? Подходят ли методики для разных возрастных групп?

Sternberg & Grigorenko: Методы

- Материалы
 - Подходят ли материалы для респондентов?
 - Понятны ли материалы респондентам так, как понимает их экспериментатор?
 - Мотивированы ли респонденты?
 - Адекватно ли материалы операционализируют изучаемые теоретические конструкции?
 - Удовлетворительно ли описание материалов?

Sternberg & Grigorenko: Методы

- План исследования
 - Каков принцип разбиения на группы?
 - Есть ли контрольные группы?
 - Продуман ли в целом дизайн?
 - Подходит ли зависимая переменная к гипотезам и соответствует ли выводам?
 - Подходят ли независимые переменные гипотезам и выводам?

Sternberg & Grigorenko: Методы

- Процедура
 - Чётко ли показана процедура?
 - Нет ли пропущенных шагов?
 - Подходит ли процедура для гипотез?
 - Описано ли оборудование?
 - Выполнены ли этические нормы (информированное согласие, дебрифинг).
 - Было ли что-то неэтичное в исследовании?
 - Подходит ли процедура респондентам?
 - Если были ошибки, влияют ли они на валидность?

Sternberg & Grigorenko: Методы

- Гипотезы
 - Ясно ли, как гипотезы следуют из теории?
 - Соответствуют ли гипотезы имеющимся данным? Если нет, объясняется ли это?
 - Правдоподобны ли гипотезы?
 - Интересны ли гипотезы?
 - Согласуются ли гипотезы друг с другом?
 - Проверяемы ли они?

Спасибо за внимание!



JORGE CHAM © 2006