

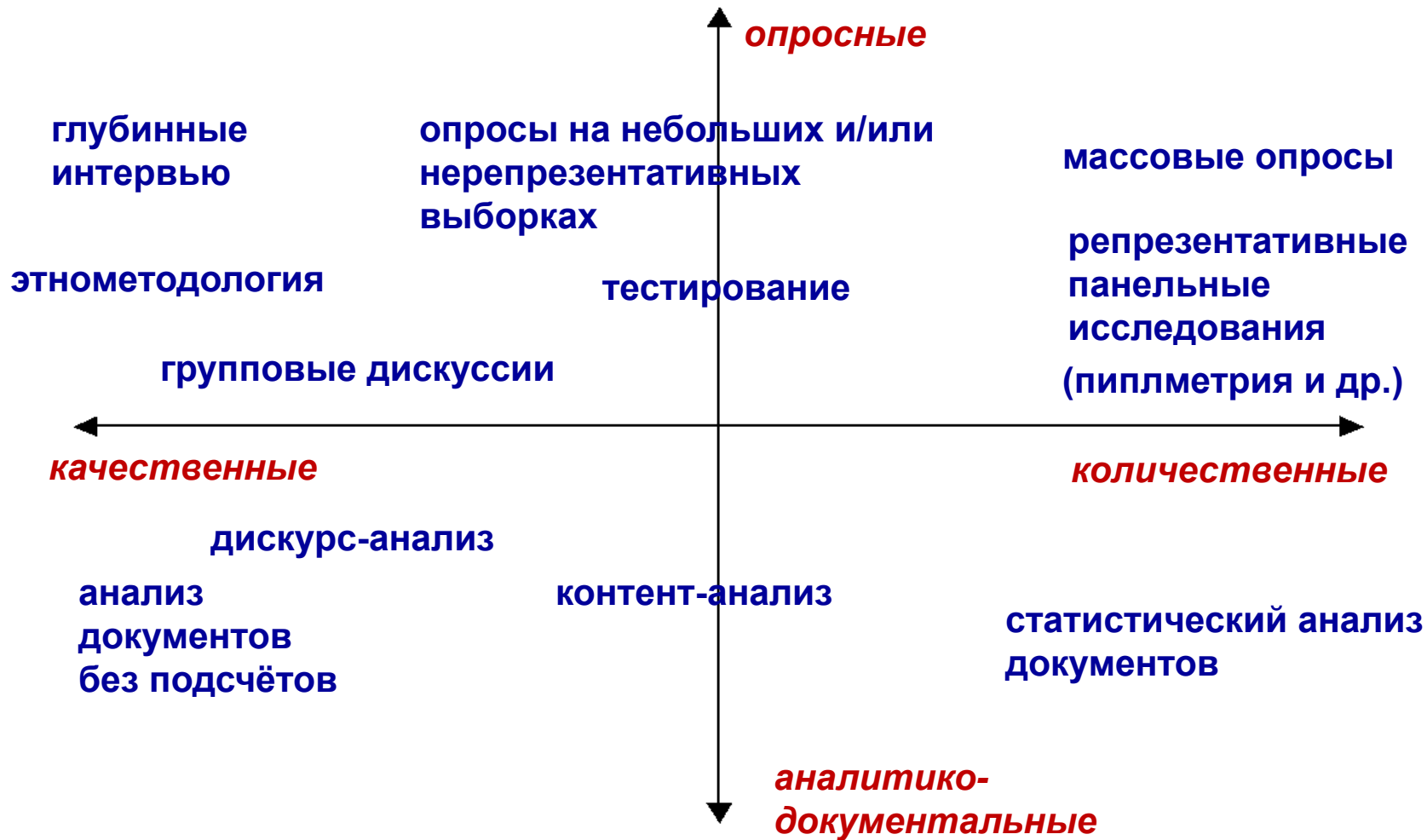
*Национальный исследовательский университет
«Высшая школа экономики»
Факультет коммуникаций, медиа и дизайна
Дисциплина «Анализ медиа-аудиторий»*

КЛЮЧЕВЫЕ ПОНЯТИЯ СТАТИСТИКИ



*А.В.Шариков,
профессор факультета
коммуникаций, медиа и
дизайна,
НИУ «Высшая школа
экономики»*

Условное пространство эмпирических методов, широко используемых в медиаисследованиях



Количественные методы исследования основаны на статистической теории

Статистические исследования (обследования) предполагают, что имеется большое число данных - сотни, тысячи, миллионы и т.п.

Например, мы хотим знать преобладающее мнение россиян о каком-нибудь событии или персоне. Здесь возникает сложность – а как опросить всех россиян? Их более 146 млн. чел. Не все пользуются интернетом и даже телефонами.

*Статистическая теория говорит следующее:
не надо опрашивать всех, а надо выбрать достаточно большое, но достижимое число людей, опросить их и перенести результаты на все население, соблюдая определенные правила.*

На примере приведенного суждения можно проследить основные понятия, на которых зиждется теория массовых опросов. Это:

- генеральная совокупность**
- выборочная совокупность или выборка**
- респонденты или опрашиваемые**
- экстраполяция**
- репрезентативность**
- погрешность измерения**

Генеральная совокупность (universe)

Генеральная совокупность – это множество объектов, относительно которого мы хотим получить результат.

Пример: население России



Выборочная совокупность или выборка (sample)

Это часть генеральной совокупности, на которой проводятся статистические наблюдения.

Пример: 1600 человек из России



Респонденты или опрашиваемые (respondents)

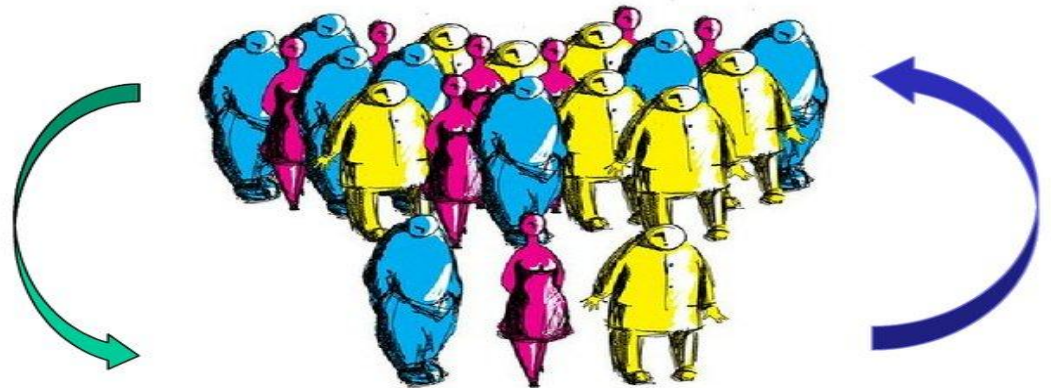
*Респонденты – люди, которых
опрашивают.*

Все респонденты формируют выборку.

Генеральная совокупность и выборка

- 150 тыс. человек
- Генеральная совокупность

- 250 человек
- Выборка

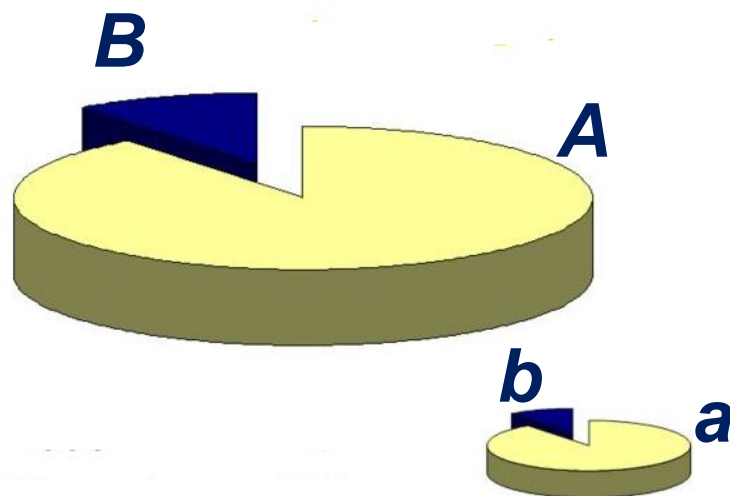


Респонденты

Экстраполяция (extrapolation)

Экстраполяция – это перенос некоторых результатов, полученных в выборке, на всю генеральную совокупность. Чаще всего применяется для процентных и средних величин.

$$\frac{B}{A} = \frac{b}{a}$$



Экстраполяция (extrapolation)

Пример.

Генеральная совокупность - 100 000 чел.

Выборка – 1000 чел.

На вопрос о том, нравится ли «Первый канал» 400 чел. ответили «да».

Принцип экстраполяции предполагает, что в генеральной совокупности такого же мнения придерживается 40 000 чел.

$$\frac{400}{1000} = \frac{40000}{100000}$$

Но!

Экстраполяция (extrapolation)

Но!

Экстраполяция возможна, если социальная структура выборки и генеральной совокупности совпадают.

Т.е. все социальные пропорции генеральной совокупности должны присутствовать и в выборке.

Например, в России 54% населения – женщины. Значит, и в выборке их должно быть 54%.

Репрезентативность или представительность выборки (representativeness of the sample)

Свойство сохранения в выборке структурных пропорций, присущих генеральной совокупности, называют репрезентативностью.

Такую выборку называют репрезентативной.

И только в случае репрезентативной выборке возможна экстраполяция – перенос процентных и средних величин, полученных в исследовании, из выборки на генеральную совокупность.

Репрезентативность или представительность выборки (representativeness of the sample)

*Это понятие ввел в
научный обиход
знаменитый
американский социолог
Джордж Гэллуп (George
Gallup) в 1936 году.*



Репрезентативность или представительность выборки (representativeness of the sample)

*В 1936 году в США проходили выборы президента.
Баллотировались Франклин Делано Рузвельт
от Демократической партии и Альф Лэндон от
Республиканской партии.*

*Франклин
Делано
Рузвельт*



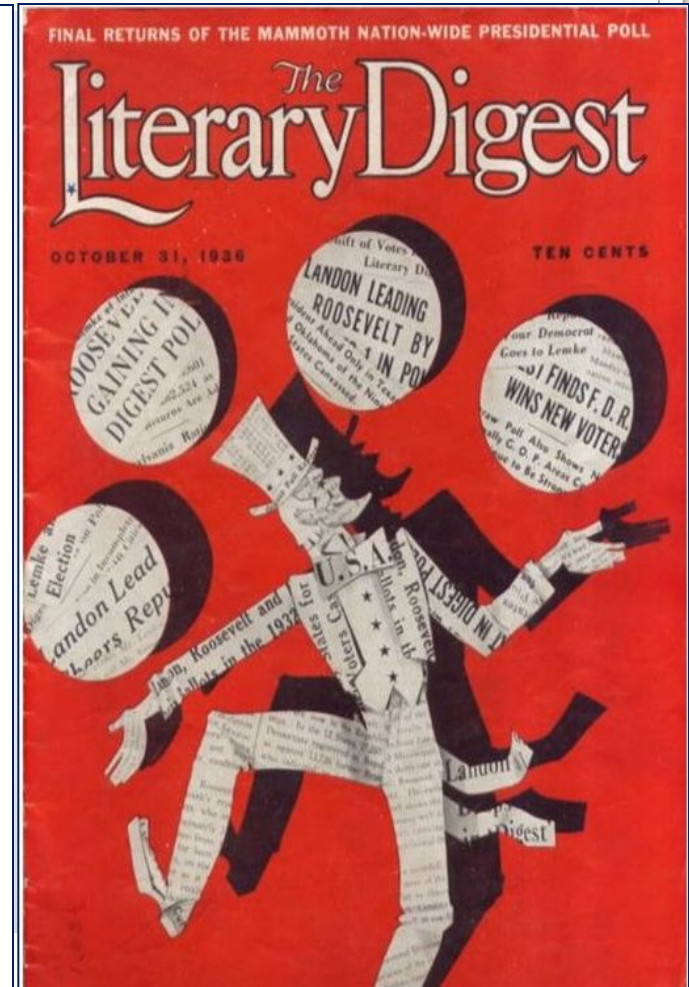
*Альф
Лэндон*

Репрезентативность или представительность выборки (representativeness of the sample)

*Популярный в то время журнал
«Literary Digest» организовал
беспрецедентную акцию.*

*Редакция разослала **10 млн. писем**
американцам, в которых была
анкета с одним вопросом:
«Кто, по-Вашему, победит на
выборах?»*

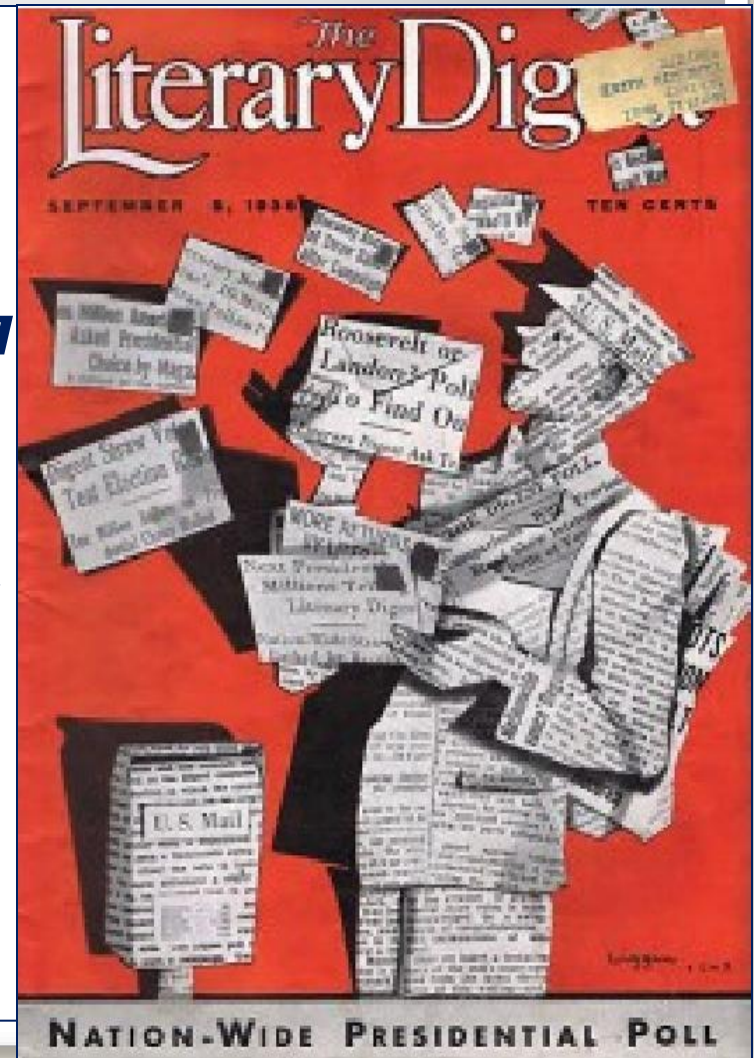
*Анкета была напечатана на
открытке с маркой и адресом
редакции. Респондент должен
был ответить и отправить
открытку – расходы оплачивал
журнал.*



Репрезентативность или представительность выборки (representativeness of the sample)

*Из 10 млн. разосланных
анкет в редакцию
вернулись 2,3 млн.
заполненных. Огромная
выборка!!!*

*Подсчет ответов привел
к выводу: победит
республиканец Альф
Лэндон.*



Репрезентативность или представительность выборки (representativeness of the sample)

*Однако Джордж Гэллап
раскритиковал журнал «Literary
Digest» и представил
результаты своего опроса на
выборке **50 тыс.** чел.
По его данным должен был победить
Рузвельт, что и произошло на
самом деле.
Гэллап доказал, что большая
выборка сама по себе не
гарантирует корректный
результат, если она не отвечает
требованию
репрезентативности.*



Статистическая ошибка или погрешность (statistical error)

Часто спрашивают: мы взяли выборку и получили некоторые результаты. А если мы выберем других людей, то получим ли мы такие же результаты?

Математическая статистика дала ответ на этот вопрос.

В самом простом виде он звучит так: если выборка достаточно большая, то результаты в общем случае будут разными, но отличия будут невелики. Чем больше выборка, тем меньше эти отличия.

Статистическая ошибка или погрешность (statistical error)

Все возможные вариации укладываются в сравнительно небольшой интервал значений измеряемой величины. Его называют «доверительный интервал».

Ориентируются на среднюю в этом интервале величину и рассчитывают отклонения от нее – плюс или минус.

Величина отклонения и называется статистической ошибкой.



Статистическая ошибка или погрешность (statistical error)

Пример.

Пусть измеренная величина составляет 3, а погрешность равна 1.

Тогда нижняя граница доверительного интервала равна $3 - 1 = 2$, а верхняя граница доверительного интервала равна $3 + 1 = 4$.

*Следовательно, доверительный интервал: {2;4}.
Другая форма записи: 3 ± 1 .*



Статистическая ошибка или погрешность (statistical error)

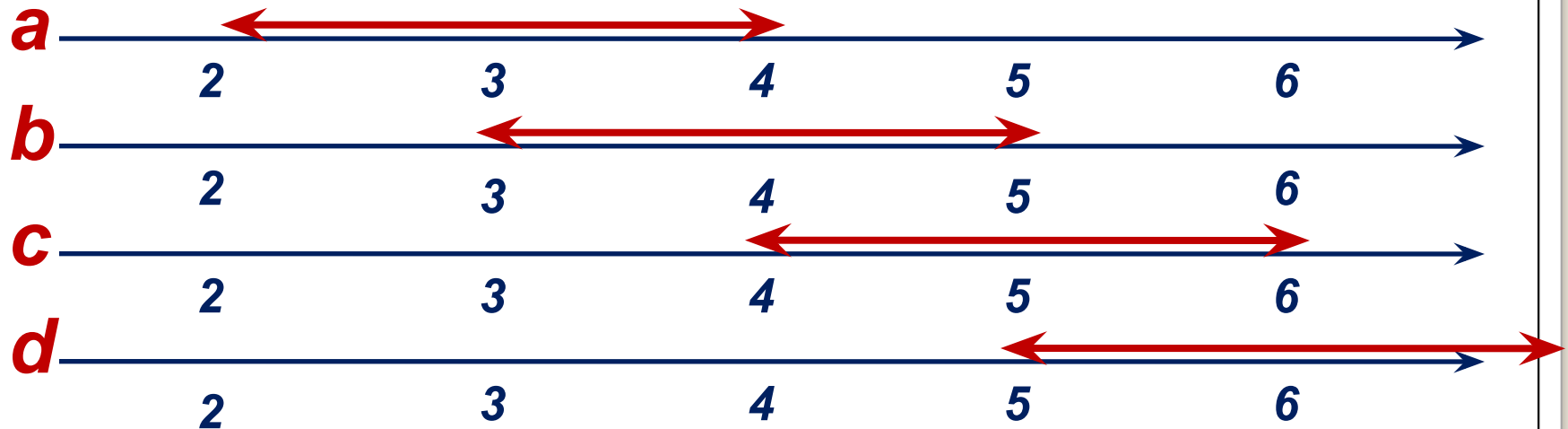
Когда сравнивают две величины, полученные в статистическом исследовании (например, массовом опросе), то при интерпретации учитывают доверительные интервалы.

Если интервалы пересекаются, то считается, что величины примерно одинаковые «с точностью до погрешности измерения».

Если интервалы не пересекаются, то величины не равны.



Статистическая ошибка или погрешность (statistical error)



$$a = 3 \pm 1$$

$$b = 4 \pm 1$$

$$c = 5 \pm 1$$

$$d = 6 \pm 1$$

Статистическая интерпретация:

$$a \approx b$$

$$a \approx c$$

$$a \neq d$$

$$b \approx c$$

$$b \approx d$$

$$c \approx d$$