

Using of statistical methods in oral historical studies of life satisfaction in urban space

Начну свой доклад с высказывания известного английского государственного деятеля Бенджамина Дизраэли (1804- 1881) (слайд 2): «Существуют три вида лжи: ложь, наглая ложь и статистика». Чтобы результаты статистического анализа не ставить в один ряд с ложью, необходимо адекватно использовать статистические методы анализа данных, в частности, проводя научные исследования в историческом аспекте. Необходимо овладеть не столько всевозможными математическими тонкостями, хотя это тоже важная составляющая, сколько вопросами логики и корректности применения тех или иных статистических методов, умением выбирать верные способы решения конкретных исследовательских проблем. Современная наука активно использует статистические данные, а, следовательно, и математико-статистические методы анализа данных. Открыв любой научный отечественный и зарубежный журнал, можно встретить многообразие цифр и таблиц с анализом статистических данных. Как же не потеряться в этом море статистической информации и таблиц с данными? В своем докладе расскажем о некоторых методах статистического анализа, часто используемых при изучении удовлетворенности жизнью, на примере научных статей. В частности, остановимся на проверке «парадокса Истерлина»: как влияет доход на удовлетворенность жизнью, используя различные статистические методы.

Дескриптивный анализ. Любой анализ данных начинается с дескриптивного анализа: необходимы обобщающие количественные показатели, раскрывающие общие свойства статистической совокупности и дающие общую картину, показывающие тенденцию развития процесса или явления, нивелируя случайные индивидуальные отклонения, позволяющие сравнивать различные совокупности и т.д. Все показатели, используемые в дескриптивном анализе можно разделить на две основные группы: меры среднего уровня (среднее, мода, медиана) и меры рассеяния (среднее квадратическое отклонение, дисперсия, коэффициент вариации).

В качестве иллюстрации рассмотрим статью Гуриева и Журавской¹ (2009), в которой используются различные виды статистического анализа (слайд 3-4). В частности, на рис.1 и в табл.1 представлен как раз дескриптивный анализ - средние значения ВВП на душу населения (рис.1) и некоторые индикаторы потребления для сравнения в странах с переходной экономикой, страны со средним доходом и США.

¹ Guriev S., Zhuravskaya E. (Un)Happiness in Transition. *Journal of Economic Perspectives*, 2009, vol. 23(2), pp.143-168.

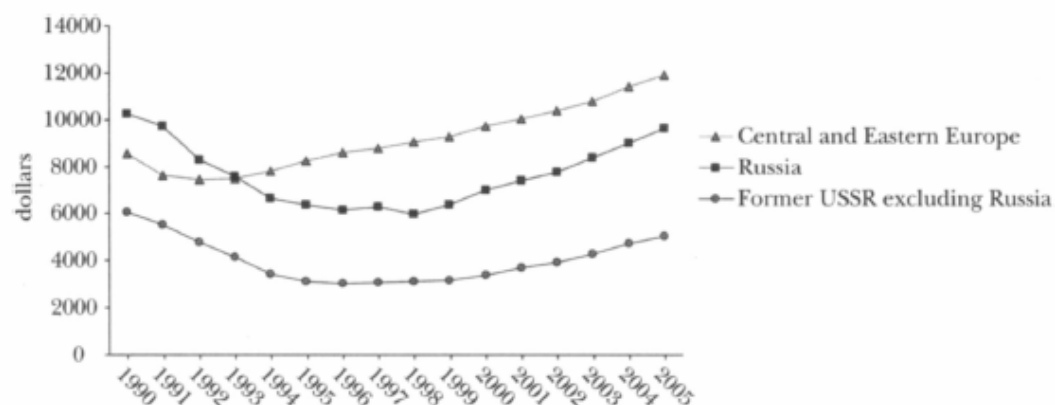
*Table 1 Selected Indicators of Consumption
(cross-country average)*

	1985	1990	1995	2000	2004 (unless otherwise noted)
Household consumption expenditure per capita (constant 2000 US\$)					
Transition countries	-	1154	1009	1155	1543
Middle-income countries	774	813	925	1044	1174
United States	17081	19110	20405	23880	25841
Housing (square feet per person)					
CIS	-	172	-	183	215*
United States	-	-	694	720	752
Cars (per 1,000 people)					
Transition countries	-	110	134	187	223**
Middle-income countries	-	37	50	69	54**
United States	-	758	756	785	-
Telephone mainlines (per 1,000 people)					
Transition countries	94	125	159	216	264
Middle-income countries	29	40	68	127	195
United States	487	545	600	682	606
Personal computers (per 1,000 people)					
Transition countries	-	4.1	19.2	57.3	109.8
Middle-income countries	-	2.2	9.5	29.1	58.3
United States	106	217	324	570	762

Figure 1

The Dynamics of Real GDP Per Capita

(in constant international 2000 U.S. dollars adjusted for purchasing power parity)



Source: World Development Indicators database, World Bank.

Гуриев, 2009.

Авторы заключают, что переход от плановой к рыночной экономике сопровождался спадом в 90-е г.г. уровня жизни населения в России и странах бывшего СССР.

Вторым часто встречающимся статистическим методом является *корреляционный анализ*. Корреляционный анализ - метод обработки статистических данных, с помощью которого измеряется теснота связи между двумя или более переменными. При проведении корреляционного анализа исследователь должен ставить перед собой несколько вопросов:

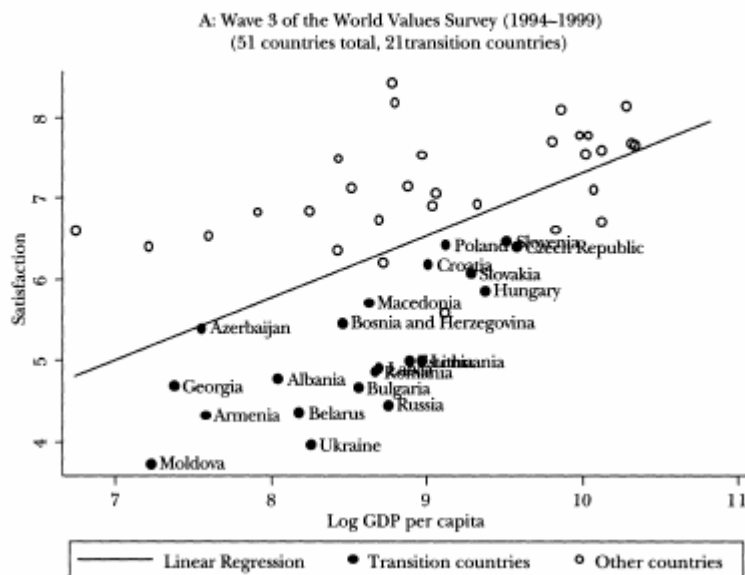
- как выбрать (с учетом специфики и природы анализируемых переменных) подходящий измеритель статистической связи (параметрический или непараметрический коэффициент корреляции)?

- как оценить (с помощью точечной и интервальной оценок) его числовое значение по имеющимся выборочным данным?
- как проверить гипотезу о том, что полученное числовое значение анализируемого измерителя связи действительно свидетельствует о наличии статистической связи?
- как определить структуру связей между компонентами исследуемого признака («связь есть» или «связи нет»)?

Простейшим (визуальным) способом выявить наличие взаимосвязи между количественными переменными является **построение диаграммы рассеяния**. Это график, на котором по горизонтальной оси (X) откладывается одна переменная, по вертикальной (Y) другая. Каждому объекту на диаграмме соответствует точка, координаты которой равняются значениям пары выбранных для анализа переменных.

Вернемся к примеру (слайд 5). На рис.2 (Гуриев, 2009) приведена диаграмма рассеяния ВВП на душу населения и среднее значение удовлетворенности жизнью в разных странах, которые разделены на 2 группы: страны с переходной экономикой и другие страны. Каждая точка – это страна. Помимо точек мы видим прямую линию – это усредненная линия линейной зависимости (регрессии) между удовлетворенностью жизнью и доходом в рассматриваемых странах. На основании графика можно сделать вывод, что связь существует (с ростом дохода удовлетворенность жизнью увеличивается), причем удовлетворенность жизнью в 1994-1999 гг. в странах с переходной экономикой ниже, чем во всех остальных.

Figure 2
Life Satisfaction and per Capita GDP



При проведении корреляционного анализа исследователи рассчитывают и интерпретируют различные виды коэффициентов корреляции, которые характеризуют тесноту линейной зависимости между двумя переменными. Например, наиболее часто используемым является выборочный парный коэффициент **корреляции (Пирсона)** между показателями x_i и x_l , который рассчитывается по формуле (слайд 6):

$$r_{jl} = \frac{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{il} - \bar{x}_l)}{s_j s_l}.$$

Коэффициент корреляции изменяется в пределах от -1 до +1, причем в статистике принято считать (слайд 7), что если $|r|$ изменяется в пределах от 0 до 0,3, то связь очень слабая, от 0,3 до 0,5 – слабая, от 0,5 до 0,7 – умеренная, от 0,7 до 0,9 - сильная, свыше 0,9 - очень сильная. Отметим, чтобы оценить зависимость между переменными, нужно знать как «величину» корреляции, так и ее значимость, которая представляет собой главный источник информации о надежности корреляции. Для проверки значимости используют различные критерии, например, критерий Стьюдента. Рассчитывая коэффициенты корреляции для нескольких показателей, в статистике принято строить **корреляционные матрицы**. Пример корреляционной матрицы приведен в работе (Rafael, 2003) (слайд 8).

CORRELATION COEFFICIENTS, 12 EUROPEAN NATIONS: 1975 TO 1992

	Reported Life Satisfaction	GDP per Capita (1985 U.S.\$)	ΔGDP per Capita	Benefit Replacement Rate	Inflation Rate
Reported life satisfaction	1				
GDP per capita (1985 U.S.\$)	0.209	1			
ΔGDP per capita	0.056	0.278	1		
Benefit replacement rate	0.281	0.471	0.111	1	
Inflation rate	-0.161	-0.659	-0.379	-0.521	1
Unemployment rate	-0.023	-0.151	0.062	-0.016	-0.230

Из таблицы видно, что корреляция между удовлетворенностью жизнью и ВВП является одной из самых высоких среди рассматриваемых показателей, однако взаимосвязь очень слабая и встает вопрос о значимости этого показателя. Однако данных, приведенных в таблице недостаточно, чтобы определить значимость коэффициента корреляции.

В заключении, хотелось бы еще упомянуть **регрессионный анализ** - статистический метод исследования влияния одной или нескольких независимых переменных на зависимую переменную. В зависимости от задач исследователи используют модели линейной и нелинейной множественной регрессии, модели бинарного и множественного выбора и др. В качестве примера регрессионного анализа рассмотрим модель из работы Гуриева (слайд 8).

Table 2
Is Life Satisfaction Lower in Transition?

	Dependent variable: life satisfaction (1–10)					
	(1)	(2)	(3)	(4)	(5)	(6)
					Absolute household income	Absolute household income per member
Transition country dummy	-1.40*** [0.33]	-1.13*** [0.33]				
Transition country dummy * wave 2			-0.72*** [0.22]	-0.67*** [0.24]		
Transition country dummy * wave 3			-1.44*** [0.28]	-1.56*** [0.27]		
Transition country dummy * wave 4			-0.87*** [0.29]	-0.90*** [0.32]		
Log GDP pc (PPP \$)	0.47*** [0.17]	0.35 [0.24]	0.42*** [0.12]	0.44*** [0.12]		
Relative HH income (1–10)	0.14*** [0.02]	0.20*** [0.02]	0.14*** [0.02]	0.13*** [0.02]		
Log absolute HH income					0.41*** [0.06]	0.23*** [0.04]
Transition country dummy * (Log GDP pc – mean)			0.38* [0.23]			
Transition country dummy * (Relative HH income – mean)				0.07** [0.03]		
Transition country dummy * Log absolute HH income					0.26*** [0.07]	0.21*** [0.06]
Wave dummies			yes	yes		
Country-level controls	yes	yes	yes	yes		
Country dummies					yes	yes
Sample: wave	3	4	all	all	4	4
Observations	51,516	56,903	161,508	161,508	63,237	27,290
R ² adjusted	0.25	0.18	0.19	0.19	0.23	0.28
Countries	39	44	55	55	53	26
No. of transition countries	14	16	17	17	16	11

В таблице приведены результаты оценивания модели удовлетворенности жизнью в зависимости от различных социально-экономических факторов, приведенных в первом столбце. Традиционно в таблицу включают сами оценки коэффициентов, их стандартные ошибки, звездочки указывают на «значимость» («надежность») коэффициентов. Также мы видим в таблице число наблюдений для каждой модели и коэффициент детерминации – показатель качества модели, описывающий долю объясненной дисперсии зависимой переменной за счет факторов модели. Так например, ВВП является значимым фактором, положительно влияющим на удовлетворенность жизнью: с ростом ВВП удовлетворенность жизнью возрастает.

Литература

1. Белова Е.Б., Бородкин Л.И., Гарскова И.М., Измestьева Т.Ф., Лазарев В.В., Тихонов А.И. Компьютеризованный статистический анализ для историков. Учебное пособие / Под ред. Л.И.Бородкина и И.М.Гарсковой. - М., 1999. - 187 с.; <http://www.hist.msu.ru/Labs/HisLab/Stud/textbook.htm>
2. Эконометрика. Под ред. Мхитаряна В.С. М., Проспект, 2010.
3. Guriev S., Zhuravskaya E. (Un)Happiness in Transition// *Journal of Economic Perspectives*, 2009, vol. 23(2), pp.143-168.
4. Rafael Di Tella & Robert J. MacCulloch & Andrew J. Oswald, 2003. The Macroeconomics of Happiness// *The Review of Economics and Statistics*, MIT Press, vol. 85(4), pages 809-827