**QUANTITATIVE METHODS IN SOCIO-POLITICAL-ECONOMIC RESEARCH – Session 1**

**IPS SEMINAR FOR THE PHD STUDENTS Vladimír Benáček, 28.3.2019**

**There will be 4 seminars starting on 28 March – all at distance via Moodle and e-mails**

**Original strategy of the course:**

*Instead of one-sided lecturing there will be a partnership of all participants where they collectively discuss the topics outlined by the instructor. At the same time the participants elaborate the technical exercises delivered before each meeting. It is expected that the participants come with their own laptops where there will be installed all required add-ins of excel, i.e. Data Analysis and Statistical functions. Work at home before the seminars are a part of the exercise. At the meeting the participants solve again the practical exercises.*

**REMARK V.B.: Sorry that this grand idea cannot be honoured once I am on a sick leave since 22.3.2019. Instead, there will be reading assignments in form of my review of the topics from statistics, written as a cook-book). I will try to be as clear as possible (sorry if I do not succeed in full). Then there will be homework assignments, which should be delivered to** **vladimir.benacek@fsv.cuni.cz****. We will communicate by e-mails. Sorry if this will be rather awkward. I do not know on which level of statistics you perform. THIS IS A TURBO INTRODUCTION TO PRACTICAL ASPECTS OF STATISTICS AND REGRESSIONS. In case you are more advanced and experienced, you can have a restive approach (but still performing in assignments). Any textbook of statistics will help you as a feedback. There are numerous webpages on each topic on internet.**

**All the best, Vladimír Benáček**

**Review of the course topics:**

28.3.2018: Methodological pre-conditions for a meaningful quantitative analysis. Reading 1 on shortcomings of “soft sciences”. Homework: Databases and the organisation of data.

4.4. Univariate descriptive statistics. Averages, correlations and data analysis by graphs. Distribution functions. Homework: practicing with the data.

11.4. The techniques and the art of regressions. Non-linearities and data transformations. Properties of data samples. Heteroskedasticity and elasticity. Homework.

18.4. Omitted variables syndrome, autocorrelation of residuals, multicollinearity. Estimation of trends. Work with dummy variables. Homework.

All this will be illustrated on 12 practical exercises to be solved by Excel statistical instruments.

**SESSION 1:**

Learning statistics by doing: The core of statistics relies on logic, common sense and ability to experiment quantitatively. That is neither the crunching of formulas and proofs, nor following blindly some technical routines. It is a process or creative description of reality by using numbers and causal links.

**Objectives of this introductory course:**

1. To be able to read and understand quantitative papers (construction of models, functional forms, correlations (= inter-dependencies vs causes & effects), interpretation of coefficients statistical significances and knowledge of limits of statistics (e.g. biases and spurious dependencies, missing explanatory variables).
2. To be able to conduct elementary quantitative exercises (e.g. work with databases, data collection and processing, editing graphs, running simple & multiple regressions, test for bias, working with hypothesis testing.
3. This course **does not** aspire to build knowledge that will lead the users to achieve such a technical skill that would allow publishing papers in top international journals. Further study is required or a cooperation with a skilled statistician in order to publish. However, quantitative analyses for policy recommendations could lead to successful publications even by working with simple data and simple inferential models.

**Weaknesses of the quantitative research in political science:**

* the lack of robust descriptive theoretical models guiding the empirical research
* reliance on the intuition in causalities
* limited number of data in cardinal scales and too many dummies (nominal data)
* non-stationarity of behavioural networks in political processes (endogeneity and circular causality).

**Political science can build methodologically on the advances in economics and econometrics.**

Example: Gravity models of human exchanges (of goods, services, labour, finance, knowledge, **COOPERATION and POWER**) as taught at IES FSV.

Structure:

1. Fundamentals: JEB039-Theory of Trade; JEB135-Theory of Business.
2. Methodology: JEM179-History and Methodology of Economics.
3. Technical courses:

ca] JEM 062: Introductory Econometrics. Required textbook: J. Wooldridge (Introductory Econometrics: A Modern Approach, 5th edition). See its Appendix B, D, and C1-C3. Those chapters are summarising basic mathematics and statistics necessary to understand the class.

cb] If your competences do not emulate these appendices you must attend JEB 117: Mathematics for Economists. cc] There are also three introductory courses of statistics at IES: JEB 142: Introductory Statistics; JEB105: Statistics.

An independence in working with advanced econometrics requires attending intensive courses for 2-3 years .

cc] Simple introductory statistics can be learned at JSM406 Statistics in SPSS: Soukup/Petrúšek (applied in sociology) and **JEB111: Advanced Data Analysis in MS Excel**. This

**CLASH OF THE METHODOLOGY: Rationalism embodied in generalised axiomatic theories versus metaphysics of intuition ending in particularism *ad hoc*.**

**DISCUSSION BLOCK (1):**

The post-modernist dogmas (in Rand E. 1957. Atlas Shrugged. Penguin-Plume Publ., N. York, p. 556):

**“Everything is relative; nobody can know anything; reason is an illusion; there is not any reality”.**

Please analyse the logical implications of this statement:

1. All is a play of convention (compare with the treatment of the “paradigm shift” by T. Kuhn).
2. Perception is subjective and we create the reality by our images (= constructivism).
3. There is not only the head (brain-rationality, logic) but also the heart (intuition, emotions, fake news, lies, …). 🡺 The decision-making is made more by heart then by brain.
4. There are logical inconsistences in that statement since this is an absolute statement subject to logical self-denial and resignation to empirical testing as a proof. It also claims implicitly that any lie is both truth as well as non-truth.
5. The pragmatic value of knowledge has no meaning because all is particular, specific, non-repeatable and thus **generalisations have no meaning**.
6. There is a moment of resignation to individual freedom (= to being a master of one’ own destiny). 🡺„The liberty is in being a slave (an ignoramus, an object of manipulation) voluntarily“. 🡺 „The liberty is in having the capacity to produce happy slaves”. 🡺 Machiavellist aggression.
7. This leads to the denial of universal (absolute, generalised) meaning of science, mathematics, logics, quantification, statistics, ranking and prediction. 🡺 Escape from rationality in objective processes to the domains of speculation, intuition, agnosticism and voluntarism, where random chances and subjectivism dominate over reality.

**Pragmatic objectivism:**

M. Friedman’s positivism related to the statistics of probability:

If a prediction derived from some statement (hypothesis, model) is highly probable in reality then the statement can be accepted as „objectively valid“ (i.e. pragmatically, operationally valid). If the probability is converging to infinity, we can talk about objective „laws of science“.

**DISCUSSION BLOCK (2):**

***Consider a model of gravitation (gravity G) where G = Φ (Mass 1, Mass2, Distance, error term ε) and apply it to physics, economics and politics.***

***T. Kuhn*** *– evolution of paradigms, ranking of knowledge/science (Structure of Scientific Revolutions).*

***K. Popper*** *– falsifying by empiricism, critical rationalism (Poverty of Historicism).*

*All non-trivial (i.e. non-tautological) statements in form of* ***implications (relationships)*** *are subject to* ***formalisation*** *(e.g. by symbols, oriented graphs, hypotheses, models, laws …) and quantification by data. Is this a sufficient base for building a science?*

**Acquisition and generation of data:**

a] Controlled experiments (field or laboratory experiments are rather rare in political science)

b] Surveys by using own questionnaires (simulation of reality)

c) Observational studies by using statistics describing real processes

d) Availability of professional data for empirical studies that can be repeated and compared.

**Data generation processes (quantifying observations):**

a] Natural laws – universal causalities that are non-dependent on particular time or space of measurement. Ideally considered as close systems (with a fixed number of causes) where just the measuring errors distort the data.

b] Social „laws“ as **behavioural** inter-dependences (“regularities”) – dependent on time and space. In reality socio-political systems are open systems where the number of causal factors is unlimited and subject to changes.

c] Phenomenon of “omitted variables” and “non-stationarity of stochastic processes” versus “white noise random disturbances” + ”entropy”.

🡺 Different role of theories (hypotheses) in natural/technical and social sciences.

**Measurement of measurable phenomena (scales):**

**Nominal** (binary, dichotomous) data describing **qualities** (yes / no) – male x female, won x lost, present x absent.

**Ordinal** (non-equidistant intensity of performance; rank scale) data – grades in school, preferences, deciles of quantity ranking.

**Cardinal** (metric) data:

a] With natural zero – **ratio scale**. Election participation, budget of parties, number of seats in parliament …). Multiplication and divisibility can be applied.

b] With arbitrary zero (or arbitrary 100% mark) – **interval scale**. Temperature, IQ coefficients, average popularity of a party in a population cohort, gains of votes between elections. Multiplication and divisibility lead to biased ratios here – just additions and subtractions can be used.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Characteristics | Nominal | Ordinal | Cardinal: Interval scale | Cardinal: Ratio scale |
| Ordering of values (ranking) | - | Y | Y | Y |
| Assignment to the data set | Y | Y | Y | Y |
|  Calculating the Mode (frequency of distribution) | Y | Y | Y | Y |
|  Calculating the Median | - | Y | Y | Y |
|  Calculating the Mean | - | ? | Y | Y |
|  Quantifying the difference between values | - | - | Y | Y |
|  Adding and subtracting of values | - | - | Y | Y |
| Has an absolute („true“) zero | - | - | - | Y |
|  Multiplying and dividing of values | - | - | - | Y |

**Working with cardinal values – ratio scales**: The original data are in “absolute levels” – i.e. in nominal values (e.g. number of votes gained in elections). They can be transferred into “relative levels”, e.g. to number of voters per 1 thousand CZK spent on elections. Or to “differences”, e.g. to changes of levels relative to past elections. Or to “percentages” (growth rates in two different time observations), in order to compare things that are not similar in size.

**Transformation of levels to percentages of growth can be either computed directly or by transforming the data by taking natural logarithms LN (x):**

**Model Ln(Y) = b \* Ln(X) is approximately interpreted as: %Δ Y = b \* %Δ X , where b is the coefficient of elasticity.**

**A model approach to the description of human behaviour by a list of causal factors:**

Whenever there is more than one causal factor (e.g. in multiple regressions) the model is valid only if the assumption of ceteris paribus (= other factors are unchanged) has a real interpretation.

E. g. Quantity of meat purchased = a - b\*Price + c\*Income + ε. **Q = 6,05 – 0,027 P + 0,0001\* I**

 Q of meat c\*Δincome = 0,0001\*10 000 = 1

 5 kg

 4 kg

 **2 kg for income 30 000 CZK per consumer 🡺 Q =**a - b\*P + 0,0001\* 30000

 **for income 20 000 CZK per consumer 🡺 Q =**a - b\*P + 0,0001\* 20000

 **Q =**a - b\*P

 **0 150 Kč Price of meat**

 **-4,05** -4,05 = -0,027\*150

**Review of methodology in social sciences:**

a] Positive (descriptive) versus normative (prescriptive, policy) statements. This is mostly used in economics where axiomatic (mathematical) theories dominate the research.

b] Realism versus liberalism (idealism), plus perhaps constructivism and structuralism in political studies. E.g. see the Morgenthau’s Classical Realism vs Waltz’s Structural Realism or the neo-liberal institutionalism <https://en.wikipedia.org/wiki/International_relations_theory> .

**Any quantitative analytical study must start with the review of theories** dealing with the object of writing and the present state of knowledge (references to literature). The next step is to declare your **method of analysis** including a list of **HYPOTHESES** to be tested for validity. **Empirical tests** are then quantitative (by statistics) based on inferences from assumptions and data to conclusions.

Any quantitative analysis should be underpinned by two-pronged views:

a] DESCRIPTIVE Statistics – i.e. the properties of data samples (bundles): average (mean & median), minimum x maximum, quartiles, standard deviation, skewness, correlation, graphics. Provided by ANOVA basic tests. They use univariate or bivariate analytical methods.

b] INFERENCIAL Statistics – analyses of the validity of hypothesis and causal relations in the real world: regressions, confidence intervals and significances. Implications and conclusions (e.g. policy recommendations, predictions). They use multivariate analytical methods.

**Recommended databases:**

|  |  |
| --- | --- |
| <http://knihovna.fsv.cuni.cz/Elektronicke-zdroje>/%2A | IES Library databases |
| <http://ies.fsv.cuni.cz/cs/node/188> | Access to databases (ISI, Reuters, Magnus, Bluenomics, etc) |
| <http://www.czso.cz/>, <http://data.mfcr.cz/> <http://www.cnb.cz/cs/index.html> | Sources of data on the Czech economy |
| http://ec.europa.eu/eurostat/data/browse-statistics-by-theme | Eurostat: databases |
| <http://stats.oecd.org/> | OECD statistics: development, science, etc |
| <http://www.oecdbetterlifeindex.org/topics/life-satisfaction/> | OECD studies on life quality |
| <http://data.worldbank.org/indicator> | The World Bank – human development indicators |
| <http://info.worldbank.org/governance/wgi/#home>  | The World Bank – governance indicators |
| <https://www.heritage.org/index/heatmap>  | Heritage Foundation: Economic Freedom Indicators |
| <https://www.fraserinstitute.org/research>  | Frazer Institute: Vancouver, CND – Economic and Human Freedom indicators and indices |
| <https://comtrade.un.org/>  | COMTRADE UnNations, see the visualization tools.  |
| <http://guides.library.umass.edu/c.php?g=672403&p=4735643>  | Review of databases for political science |
| <http://guides.library.umass.edu/c.php?g=672403&p=4735644>  | Review of databases for political science |
| <https://www.sipri.org/databases>  | Resources on global security |

**PRACTICAL EXERCISE 1 – HOMEWORK 1:**

SKIM THROUGH THE DATABASES AND SELECT TWO SAMPLES OF DATA THAT ARE RELATED TO YOUR FIELD OF RESEARCH. The first data set should be in **cross-sections** observed in one single year for at least 10 objects (e.g. districts, individuals) reporting at least three behavioural characteristics: e.g. 10 persons (in rows) reporting income, fear of foreigners (scaled 1, 2, ..., 5) and age stored in columns. The second data set should be in **time series** where in rows there will be years (at least 15) and observed data series 1, 2, 3 in columns B, C and D. (E.g. data for 2001-2015 reporting the number of immigrants for three different countries.) Calculate the averages of data by columns. Bring the output of your exercise to the first session.

**HOMEWORK 2:** Read the appended two short articles and prepare an argument defending a position that political science can be considered a mature part of the world science family.

Disprove the devil advocates claims that: Political science has no real quantifiable and refutable theory; It has no real methodology; It has no objective behavioural regularities in causality and effects; Its arguments are just ad hoc claims and/or subjective opinions. The proof of its academic analysis is in the accord of the academic authorities (“peers”) and not in empirical falsifying evidence; Its predictability is just verbal, free of quantifying elements. Etc.

PLEASE SEND THE FILE REQUIRED IN EXERCISE 1 PLUS THE ESSAY (of 1600 words at least) to

vladimir.benacek@fsv.cuni.cz

Deadline: 2.4.2019, 16:00 at the latest.