

**Пермский филиал федерального государственного автономного  
образовательного учреждения высшего образования  
"Национальный исследовательский университет  
"Высшая школа экономики"**

**Программа учебной дисциплины  
«Анализ данных в Spreadsheets»  
(Data analysis with Spreadsheets)**

Утверждена

Академическим советом основных образовательных программ по направлениям подготовки  
38.03.01 Экономика, 38.04.08 Финансы и кредит

Протокол № 8.2.2.1-32-09/04 от 30 августа 2019

Академический руководитель ОП

\_\_\_\_\_ Белых С.А.  
Подпись                      ФИО

Разработчик	Паклина София Николаевна, преподаватель, департамент экономики и финансов
Число кредитов	4
Контактная работа (час.)	2
Самостоятельная работа (час.)	150
Образовательная программа, курс	Экономика, 4 курс
Формат изучения дисциплины	На английском языке, с использованием онлайн курса

Syllabus  
Data analysis with Spreadsheets  
(4 ECTS)

Sofia Paklina ([snpaklina@hse.ru](mailto:snpaklina@hse.ru), <https://www.hse.ru/en/staff/snpaklina>)

School of Economics and Finance

Meeting Minute # \_\_\_ dated \_\_\_\_\_ 20\_\_

## 1. Course Description

### a) Pre-requisites:

- The theory of probability;
- Statistics;
- Students are expected to be an experienced PC user.

### b) Abstract

The course “Data analysis with Spreadsheets” is designed to provide students with basic knowledge and skills of data analysis with Spreadsheets, in particular Microsoft Excel. The course begins with an introduction to basics of Spreadsheets, data types and predefined and conditional functions. Then students will learn how pivot tables are created and analyzed and how statistical analysis is performed. The final part of the course is devoted to techniques of data visualization using Spreadsheets.

The course is supported by online platform for education DataCamp ([www.datacamp.com](http://www.datacamp.com)). Students are expected to watch online lectures and complete assignments using the platform. Some lectures and final examination are provided by lecturers of National Research University Higher School of Economics.

## 2. Learning Objectives

The course is designed to make students be able to:

- Know basic principles of working with Spreadsheets;
- Import data, make basic manipulation with it to prepare data for analysis, apply basic methods of preliminary data analysis and visualize data;
- Have skills of interpreting the result of data analysis and understanding limitation and relevance of applied methods.

## 3. Learning Outcomes

## 4. Course Plan

№	Topics	All hours	Lecture hours	Self-study hours
<b>Topic 1. Spreadsheet Basics</b>				
1	Basics of Spreadsheets and data types	51	1	50
2	Predefined and conditional functions			
<b>Topic 2. Data Analysis and Visualization with Spreadsheets</b>				
3	Pivot tables	101	1	100
4	Statistics in Spreadsheets			
5	Data visualization			
<b>Sum</b>		<b>152</b>	<b>2</b>	<b>150</b>

## 5. Reading List

### c) Required

Goldmeier, J. (2014). *Advanced Excel Essentials*. Apress (online access: <https://link.springer.com/book/10.1007/978-1-4842-0734-5>).

Slager, D. (2016). *Essential Excel 2016: A Step-by-Step Guide*. Apress (online access: <https://link.springer.com/book/10.1007/978-1-4842-2161-7>).

### d) Optional

Quirk, T. J. (2016). *Excel 2016 for social science statistics: a guide to solving practical problems*. Springer (online access: <https://link.springer.com/book/10.1007/978-3-319-38959-2>).

Schmuller, J. (2013). *Statistical analysis with Excel for dummies*. John Wiley & Sons (online access: <https://proxylibrary.hse.ru:2137/toc.aspx?bookid=52869>).

## 6. Grading System

The lecture assesses a self-study students work using completion of DataCamp assignments. This grade ( $O_{self}$ ) is calculated as a percentage of maximum possible grade and transferred to a 10-point scale. The assignments completed after deadline are treated as 0.

**The resulting course grade** is calculated as follows:

$$O_{result} = 0,4 * O_{self} + 0,6 * O_{exam},$$

where  $O_{exam}$  – a grade for the exam on a 10-point scale.

## 7. Examination Type

Final student assessment (exam) is a project, that is performed in a team of no more than 2 people. Each team uses provided dataset of collects their own data, define research question and apply one or a combination of the learnt methods of data analysis with Spreadsheets. As a result of the project each team write down the report and prepare working file. The grade for the exam includes the grade for the report, grade for the working file and the grade for answering questions.

Example of exam tasks:

1. Import the file “data\_for\_exam.xlsx”
2. Let’s clean the dataset:
  - The variable Company\_Year contains the information about company name and year of observation. Separate this column to two columns: Company and Year.
  - Are there any missings? If so, delete such observations.
3. Let’s explore the dataset:
  - Do all variables have appropriate data type? If not, then convert them.
  - What is the unit of observation of the dataset?
  - How many variables and observations does it contain?
  - You have the information about year of company foundation. Create new column entitled “age” with the age of each company.
4. Let’s analyze the dataset:
  - What percentage of companies has corporate university by years?
  - What percentage of companies has state owners by region?
  - How many companies have site quality equals to 3?
  - Provide descriptive analysis (in words) of summary statistics of variables age and number of employees.
5. Let’s do graphical analysis of the dataset:
  - Plot the histograms of variables about company age and number of subsidiaries. Describe the distribution of these variables. Does the distributions of these variables are different for public and non-public companies?
6. Let’s provide correlation analysis:
  - Create correlation matrix between book value, sales, number of employees, number of subsidiaries and age.
  - Analyze correlation between these numeric variables in the dataset. What variables have the highest and lowest correlation? How it can be explained?

## **8. Methods of Instruction**

Students are expected to watch online video lectures and complete tasks assigned by a lecturer on the platform for online education DataCamp. During lecture hours, students are able to answer questions and discuss self-studied material with a lecturer.

## **9. Special Equipment and Software Support**

Students are able to watch online lectures and complete assignments using university computers with internet access. Software that is essential for the course is Microsoft Excel.