

中国社会科学院大学《社会统计软件应用》课程大纲

课程基本信息 (Course Information)					
课程编号 (Course ID)	1252021026	*学时 (Credit Hours)	32	*学分 (Credits)	2
*课程名称 (Course Name)	社会统计软件应用				
	Application of Social Statistical Software				
先修课程 (Prerequisite Courses)	社会统计学				
*课程简介 (Description)	<p>1. 课程概述与教学目标 随着当代数据和信息的大量涌现,迫切需要研究者具备从海量数据中挖掘有价值信息的能力。定量研究是社会科学研究中基本技能,而社会统计软件是进行定量研究的主要工具。本课程的主要内容涉及当前主流社会统计软件 STATA 和 SPSS 的介绍和基本使用,是社会统计软件的基础课程,是针对社会统计软件零基础或基础较为薄弱的学生而设计,目的是为学生能够独立运用相关统计软件开展定量研究奠定扎实的基础,也为后续学习社会统计软件高级分析方法进行知识和技能的储备。</p> <p>2. 课程的重点难点 本课程将重点介绍数据的录入和读入软件,变量的基础处理、数据的描述统计、数据的相关分析、单因素和双因素分析,假设检验和数据的模型分析。难点在于学生如何将自己的研究目标转化为软件语言,让软件辅助研究,再将软件输出的结果转换为规范的学术表达,回答自己的研究问题。</p> <p>3. 课程的基本要求 本课程平时成绩占 30%,包括 5 次以上的课后作业和平时课堂表现;期末成绩占 70%,学生必须完成一篇 5000 字以上课程论文,并在论文中用所学软件进行数据分析和解释分析结果。</p>				
*课程简介 (Description)	<p>1. Course overview and teaching objectives With the emergence of massive data and information, it is urgent for researchers to have the ability to mine valuable information from massive data. Quantitative research is a basic skill in social sciences research, and social statistical software is the main tool for quantitative research. The major contents of this course involve the introduction and basic use of current mainstream social statistical software Stata and SPSS. This is a basic course of social statistical software. It is designed for students with zero or weak foundation of social statistical software. The purpose is to establish a solid foundation for students to independently use relevant statistical software to carry out quantitative research, and also to reserve knowledge and skills for subsequent study of advanced analysis methods of social statistical software.</p> <p>2. Keys and difficulties of the course This course will focus on data entry and reading in software, basic processing of variables, descriptive statistical of data, correlation analysis of data, single factor and double factor analysis, hypothesis test and model analysis of data. The difficulties lie in how students</p>				

		<p>translate their research objectives into software language, let the software assist the research, and then convert the outputs of the software into standard academic expression to support their research problems.</p> <p>3. Basic requirements of the course</p> <p>The usual performance of this course accounts for 30%, including more than 5 assignments and usual classroom performance; The final score accounts for 70%. Students must complete a course paper with more than 5000 words, and use the software they have learned to analyze the data and explain the results in their paper.</p>						
*教材 (Textbooks)		Lawrence, C., Hamilton 著, 巫锡炜、焦开山、李丁等译,《应用 STATA 做统计分析》, 清华大学出版社, 2017 年						
参考资料 (Other References)		许琪著,《Stata 数据管理教程》, 北京大学出版社, 2021 年						
*课程类别 (Course Category)		<input type="checkbox"/> 公共基础课/全校公共必修课 <input type="checkbox"/> 通识教育课 <input type="checkbox"/> 专业基础课 <input type="checkbox"/> 专业核心课/专业必修课 <input checked="" type="checkbox"/> 专业拓展课/专业选修课 <input type="checkbox"/> 其他_____						
*授课对象 (Target Students)		本科生	*授课模式 (Mode of Instruction)		<input type="checkbox"/> 线上, 教学平台_____ <input checked="" type="checkbox"/> 线下 <input type="checkbox"/> 混合式 <input type="checkbox"/> 其他_____ <input type="checkbox"/> 实践类 (70%以上学时深入基层)_____			
*开课院系 (School)		社会与民族学院	*授课语言 (Language of Instruction)		<input checked="" type="checkbox"/> 中文 <input type="checkbox"/> 全外语_____ <input type="checkbox"/> 双语: 中文+_____(外语讲授不低于 50%)			
*授课教师信息 (Teacher Information)		课程负责人 姓名及简介	马妍, 中国社会科学院社会学研究所社会调查与方法研究室, 副研究员, 研究方向为人口社会学和社会学研究方法。					
		团队成员 姓名及简介						
学习目标 (Learning Outcomes)		1. 牢记 Stata 软件的命令格式, 能够编写基本的操作命令 2. 掌握数据基础处理的主要方法 3. 牢固掌握描述统计和绘制图表的主要命令, 并能够准确的解释软件输出的结果 4. 了解和熟悉线性回归和逻辑斯蒂回归的基础命令和结果解读 5. 通过课程学习让学生能够让软件为自己的研究服务, 验证自己的研究假设, 回答自己的研究问题。						
*考核方式 (Grading)		平时成绩占 30%, 包含平时作业和课堂表现; 期末成绩占 70%, 提交一篇 5000 字左右的定量研究论文。						
*课程教学计划 (Teaching Plan)								
周次	周 学 时	其中			教学内容摘要 (必含章节名称、讲述的内容提要、实验的名称、教学方法、课堂讨论的题目、阅读文献参考书目及作业等)			
		讲 授	实 验 课	习 题 课				课 程 讨 论

第一周	2	2				第一讲 社会统计软件概述
						<p>一、课程简介</p> <ol style="list-style-type: none"> 1. 课程主要内容 2. 课程要求 3. 课程的参考书目 <p>二、本课程主要使用的统计软件</p> <ol style="list-style-type: none"> 1. 软件的安装、界面 2. 软件的主要功能介绍
第二周	2	2				<p>教学方法：课堂讲授</p> <p>作业：完成课堂小调查</p>
						<p>第二讲 社会统计中的测量</p> <p>一、测量概述</p> <ol style="list-style-type: none"> 1. 测量层次 2. 测量误差 3. 评估测量质量的主要标准 <p>二、测量层次与统计方法的选择</p> <ol style="list-style-type: none"> 1. 测量层次与统计方法的关系 2. 根据测量层次选择不同的统计分析方法
第三周	2	2				<p>教学方法：课堂讲授</p>
						<p>第三讲 软件的基础操作</p> <p>一、STATA 软件命令格式</p> <p>二、数据的读入</p> <ol style="list-style-type: none"> 1. 用窗口操作读入数据 2. 用命令读入数据 3. 其他格式数据的导入 <p>三、数据的存储、浏览和导出</p> <ol style="list-style-type: none"> 1. 数据的数据 2. 数据的浏览 3. 数据的导出 <p>四、命令编辑器：do 文件</p> <ol style="list-style-type: none"> 1. 编写 do 文件 2. 执行 do 文件 3. 调用 do 文件 <p>五、日志编辑：Log 文件</p> <p>Log 文件的新建、读取、暂停和关闭</p>
第四周	2	2				<p>教学方法：课堂讲授</p>
						<p>第四讲 变量的基础内容</p> <p>一、变量名与变量标签</p> <ol style="list-style-type: none"> 1. 变量的命名规则 2. 变量的命名与标签 <p>二、变量值与值标签</p> <ol style="list-style-type: none"> 1. 变量的取值类型 2. 缺失值

					<p>3. 变量的值标签</p> <p>三、变量的显示格式</p> <ol style="list-style-type: none"> 1. 数值变量的显示格式 2. 字符变量的显示格式 <p>四、创建或修改变量</p> <ol style="list-style-type: none"> 1. 用 <code>gen</code> 命令生成新变量 2. 用 <code>egen</code> 命令生成新变量 <p>教学方法：课堂讲授</p> <p>作业：请大家用 2015cgss 数据用 <code>gen</code> 命令生成 4 个新变量：</p> <ol style="list-style-type: none"> (1) 月收入变量 (2) 被访者的年龄 (3) 用不同于课堂讲义的方法对受教育程度重新分组 (4) 用民族 (a4) 或宗教信仰 (a5) 或户口类型 (a18) 其中一个变量进行重新分组。 <p>要求：四个新变量都要有标签，第三和第四个新变量还需要给变量值贴标签所有结果以 <code>log</code> 日志形式保存。</p>
第五周	2	2			<p>第五讲 数据的基础处理（一）</p> <p>一、数据的概要统计</p> <ol style="list-style-type: none"> 1. 描述数据的常用命令 2. 选择数据的常用命令 3. 数据的逻辑检验 <p>二、检查重复个案</p> <p>三、数据的保留与删除</p> <ol style="list-style-type: none"> 1. 保留变量 2. 删除变量 <p>四、数据的排序</p> <p>数据排序的两个命令及其区别</p> <p>教学方法：课堂讲授</p> <p>作业：</p> <ol style="list-style-type: none"> 1. 完成最高学历的年龄（参考 a7c）的平均值、标准差、中位数和四分位数（答案用注释写入 <code>log</code>） 2. 是否存在完成最高学历的年龄小于出生年龄的被访者？ 3. 按照出生年份和样本编号（<code>token</code> 变量）对数据进行降序排列，并列出倒数 100 个个案的出生年份和样本编号
第六周	2	2			<p>第六讲 数据的基础处理（二）</p> <p>一、变量类型的转换</p> <ol style="list-style-type: none"> 1. 字符型变量转换为数值型变量 2. 数值型变量转化为字符型变量 <p>二、变量的重新编码</p> <ol style="list-style-type: none"> 1. <code>gen+replace</code> 2. <code>recode</code> 命令 3. 函数 <p>教学方法：课堂讲授</p>

第七周	2	2				第七讲 数据的基础处理（三） 一、虚拟变量的生成 1. gen+replece 2. recode 3. tab, gen() 4. 使用逻辑表达式 二、频数统计命令的使用与拓展 1. tab 命令的基本使用 2. tab 与 tab1 和 tab2 的比较 三、相似命令的使用与比较 1. tabstat 命令 2. table 命令 3. tab 与 tabstata 和 table 的比较
						教学方法：课堂讲授 作业： 1. 对年龄进行分组，15岁为一组，给年龄组变量贴标签和值标签，并用描述其频数分布，哪一个年龄组比例最高？ 2. 对婚姻状态（参考变量 a69）进行重新编码，贴标签和值标签，描述频数分布，并生成虚拟变量 3. 生成关于户口类型的虚拟变量（参考 a18），贴标签和值标签 4. 不同性别和婚姻状况被访者的频数和比例分别是什么？不同性别和户口类型被访者的频数是什么？
第八周	2	2				第八讲 数据的基础处理（四） 一、数据的追加与合并 1. 数据的追加 2. 数据的合并 二、数据的重组：长宽变换 1. 长数据变宽数据 2. 宽数据变长数据
						教学方法：课堂讲授
第九周	2	2				第九讲 数据的展示（一） 一、直方图及其绘制 1. 用 Excel 绘制直方图 2. 用 STATA 绘制直方图 二、柱形图和条形图的绘制 1. 用 Excel 绘制柱形图和条形图 2. 用 STATA 绘制柱形图和条形图 三、饼图及其绘制 1. 用 Excel 绘制饼图 2. 用 STATA 绘制饼图
						教学方法：课堂讲授
第十周	2	2				第十讲 数据的展示（二） 一、箱线图

					<p>1. 用 Excel 绘制箱线图 2. 用 STATA 绘制箱线图</p> <p>二、散点图</p> <p>1. 用 Excel 绘制散点图 2. 用 STATA 绘制散点图</p> <p>三、折线图</p> <p>1. 用 Excel 绘制折线图 2. 用 STATA 绘制折线图</p> <p>教学方法：课堂讲授 作业：</p> <p>1. 用 2015CGSS 数据，自选变量用 stata 画一个条形图和一个饼图，显示数据标签、设定坐标轴标签、标注坐标轴标题。 2. 用分发的 excel 数据，导入 stata 画一个折线图，设置线条类型宽度、设定图例、坐标轴标题、注释和资料来源。</p>
第十一周	2	2			<p>第十一讲 数据的分布与检验</p> <p>一、正态分布与 Z 检验</p> <p>1. 正态分布与标准正态分布 2. 单样本、双样本的 Z 检验</p> <p>二、T 分布与 T 检验</p> <p>1. T 分布的概况 2. 单样本、双样本的 t 检验 3. 回归系数的 t 检验</p> <p>三、F 分布与 F 检验</p> <p>1. F 分布的概况 2. 方差分析 3. 拟合优度检验</p> <p>四、卡方分布与卡方检验</p> <p>五、相关分析</p> <p>1. 线性相关和非线性相关</p> <p>教学方法：课堂讲授</p>
第十二周	2	2			<p>第十二讲 线性回归（一）</p> <p>一、简单线性回归</p> <p>1. 线性回归基本概念 2. 方程拟合优度 3. 显著性检验和置信区间</p> <p>二、多元线性回归的建立</p> <p>1. 回归模型的设定 2. 虚拟变量及其生成 3. 回归预测值</p> <p>教学方法：课堂讲授</p>

第十三周	2	2					第十三讲 线性回归（二） 一、多元线性回归（续） 1. 回归方程的稳健估计 2. 回归诊断 3. 稳健回归 教学方法：课堂讲授
第十四周	2	2					第十四讲 Logistic 回归（一） 一、logistic 回归概述 1. 回归模型的建立 2. 回归系数的意义和统计检验 3. 回归模型的整体评价和检验 4. 回归诊断 二、两分类 Logistic 回归 教学方法：课堂讲授
第十五周	2	2					第十五讲 Logistic 回归（二） 一、多项 logistic 回归 二、序次 logistic 回归 三、定量研究论文写作与期末论文的要求 1. 定量研究论文写作简述 2. 期末论文的要求 教学方法：课堂讲授
第十六周	2	2					第十六讲 SPSS 软件操作简介 一、数据的基础处理 二、描述统计 三、假设检验 四、回归分析 教学方法：课堂讲授
总计							
备注 (Notes)							