尊敬的老师、亲爱的同学们：

大家好！

我的课程设计题目是——多元统计分析揭示省市发展差异（即对全国 31 省市 2022 年十四项综合指标的多元统计分析）。

汇报共分为6个部分

01

首先是第一部分，背景，本次课程设计中，我的研究对象是中国的31个省份。通过对大量数据的收集和分析，试示出各个地区之间的发展差异，以及这些差异对社会经济的影响。

02

第二部分数据集选择及部分可视化

本次课设题目的数据来源于国家统计局2023年《中国统计年鉴》

十四项指标如下，其中包括了经济、人口、金融与计算机两大热门行业工资、科技、教育、社会服务等可指标，这些指标能够较全面反映一个地区的发展水平和综合实力，为区域发展评估和政策制定提供全面的数据支持。

在数据处理部分，我使用EXCEL用图表将数据集可视化，使用SPSS对数据进行了标准化，在论文附录中有完整数据的展示。

03

回归分析是一种用于确定两个或多个变量之间关系的统计分析方法

在回归分析中，我们使用模型拟合度来检测模型预测的准确性和可靠性、回归分析得到的结果如表所示，R²和调整后R²值都非常高，分别为0.969和0.945，R值接近1，说明自变量与因变量之间存在很强的线性关系。意味着所选择的自变量对因变量有很强的解释力。

04

聚类分析，我分别使用系统聚类法中的最小距离法、最长距离法以及离差平方和法（WARD）进行数据分析。通过聚类分析得到的谱系图。可以发现，三种聚类方法的结果有着较大的差别，而结合我们生活中对于各个省市发展结果的认知，离差平方和层次聚类法应该是较为准确的，这是因为不同省市的不同指标通常存在较大差异，而单纯使用最长距离和最短距离法进行判断通常会导致判别的使用过于注重某个指标，无法综合考虑多个指标的影响结果，因此最终我选择了离差平方和法。

根据聚类分析的结果我们可以得知道，沿海发达省份和直辖市因经济、科技和教育优势，以及较早的发展历史而领先，而西部和东北省份由于经济发展较晚和历史发展缓慢，整体社会发展水平较低。

05

在指标的选择过程中，会有部分指标和其他指标实际上信息会存在重复，这会增加数据的复杂度，而数据的复杂度会引起分析的难度加大，并且会影响分析的结果，通过使用主成分分析的方法，可以达到降维的目的，降低数据的复杂度。

累计贡献率是主成分分析中可以帮助我们量化地衡量主成分的重要性的指标，这个比例越高，就意味着该主成分对总差异的解释程度越大。当保留 2 个主成分时，累计贡献率达到 84.411%，超过了 80%，因此在后续分析过程中选取了前两个特征值 *λ*1 = 6*.*418*, λ*2 = 5*.*4 及其对应的主成分。

成分矩阵通过特征值与特征向量组成的排序矩阵，有效地进行数据降维和提取重要特征，我们可以看出第一个主成分（PC1）：对地区生产总值（0.95）、年末常住人口（0.783）、经营单位所在地进出口总额（0.942）、城市绿地面积（0.913）、规模以上工业企业专利申请数（0.906）等指标有较高的正负荷。这表明PC1主要代表了经济规模和工业活动的综合情况

第二个主成分（PC2）：对城镇单位就业人员平均工资（0.897）、金融业城镇单位就业人员平均工资（0.806）、信息城镇单位就业人员平均工资（0.776）、住宅商品房平均销售价格（0.818）等指标有较高的正负荷。这表明PC2主要代表了收入水平和生活质量相关的指标。

对成分矩阵表中每一列除以可得到得分系数矩阵。得分系数，是衡量原始变量对主成分影响的重要工具，这个表格数据可以进一步验证成分矩阵的结果，即PC1代表了经济和工业规模。PC2代表了收入水平和生活质量。

通过主成分分析我们可以看出在经济发展上，地区差异是明显的，差异的产生受到地理位置、资源环境和人力资本等多种因素的影响，共同决定了各地区的经济发展水平和速度

通过本次研究，我们对区域的发展差异有了较深的认知，这些差异提醒我们在制定区域发展策略时，我们应该考虑到该地区的产业特点、人力资源和地理优势等因素。通过合理配置资源，优化产业结构，希望通过这次演讲，能够为大家提供一些新的思路和启发，更好地理解省市发展的动态变化和潜在影响因素，促进经济的可持续增长和社会的进步。谢谢大家的聆听！