

基于因子分析的高校学生竞争力模型构建

中国工程类高校毕业生的数量近年来逐渐升高，而随着产业转型升级，人才素质不适应市场需求的问题加剧，“就业难”“招工难”的结构性矛盾愈发突出，高校毕业生就业形势更加严峻。而高校毕业生作为国家在就业问题上高度关注的重点群体，其就业质量成为了衡量国家就业布局和经济发展的指标。分析研究工程类专业学生就业率和就业质量的影响因素，提出科学有力的解决方案，对我国就业压力的缓解和工业化的推进具有重大意义。

对于问题一，通过对提供数据的描述性分析和初步假设检验，锁定了对薪水有显著影响的变量，并将其构造为二级**竞争力模型**指标体系。通过验证性**因子分析**，对指标体系的合理性进行验证和构造，认为竞争力模型符合要求。通过对正态化规约处理后的数据计算出了指标得分，并利用**熵权法**获得了**竞争力模型**指标的权重，分析出指标的影响力大小排序为通用技能>专业技能>自身因素>学习情况>人格特质。

对于问题二，为量化研究竞争力指标与薪水之间的关系，首先构建了**OLS回归模型**，模型基本显著。但为了对人格特质这一变量对其他变量影响薪水过程中的**调节效应**进行建模，对原有回归模型进行了修正，考虑了人格特质对其他变量的**调节效应**，发现人格特质在自身因素、学习情况和通用技能中产生了显著的调节效应，但对专业技能的交互影响不大。最终也获得了具体的量化方程关系。

对于问题三，通过综合分析当下国内高校毕业生尤其是工程类毕业生的学习、就业情况，提出了一系列合理的意见。

整体来看，模型方法综合了统计学和计量经济学手段，在保持简洁直观的同时排除了一系列干扰，能够合理对问题进行分析与解释，且量化结果与现实社会热点问题相印证，所提出的建议对当下热点问题有一定参考价值。

关键词：竞争力模型，因子分析，熵权法，OLS 回归，调节效应

一. 问题重述

1.1 问题背景

“十三五”时期作为全面建成小康社会的决胜阶段，在国内外严峻的经济形势和全球新冠肺炎疫情的双重冲击下，国家稳步推进供给侧结构性改革，实现了从“高速发展”到“高质量发展”的历史性跨越。在 2021 年的国务院印发《“十四五”就业促进规划》中指出，随着产业转型升级，人才素质不适应市场需求的问题加剧，“就业难”“招工难”的结构性矛盾愈发突出，而根据教育局公布的官方数据显示，2021 年高校应届毕业生人数突破 909 万，首次突破 900 万大关，同比增加 35 万人。

在内外因素的共同作用下，高校毕业生就业形势更加严峻。而高校毕业生作为国家在就业问题上高度关注的重点群体，其就业质量成为了衡量国家就业布局和经济发展的指标。值得注意的是，基于我国目前大力推进工业化、城镇化的发展策略，工程类专业学生是国家工业化进程的主力军。因此，分析研究工程类专业学生就业率和就业质量的影响因素，提出科学有力的解决方案，对我国就业压力的缓解和工业化的推进具有重大意义。

1.2 问题要求

附件给出了印度工程类专业毕业生就业的工资水平和各因素情况表。基于上述背景和附件信息，我们需要建立数学模型解决以下问题：

- (1) 结合国家政策性文件，毕业生就业分析质量报告、雇主满意度调查等数据分析类报告，综合分析得出工程类专业学生就业率和就业质量中的关键因子。
- (2) 基于附件一提供的数据，建立严谨的数学模型，获得工程类专业大学生薪水与各指标的关系
- (3) 综合上述分析，结合自己所在高校的实际情况，对所在高校的工程类专业毕业生的综合培养方案建言献策，提供可落实、有成效的参考建议。

二. 问题分析

2.1 问题 1 的分析

问题 1 要求我们分析高校各专业毕业生的就业影响因素，为了定量研究毕业生就业相关的影响因素，本文考虑用验证性因子分析的方法考察毕业生的薪资水平与附表中各变量的相关数结构，用以解释这些变量与毕业生就业情况的关联。为了研究性别与样本的关系，本文先对样本数据进行 t 检验，得出不同薪资下的性别占比，接着为了处理较多维度的变量影响因素，又将不同专业、地区与竞争力等各类变量抽象为几类宏观影响因素变量，通过验证性因子分析得出各变量与

就业情况的关联性，并结合熵权法得出各影响因素的权重与排名。具体实现流程操作如图 1 所示。

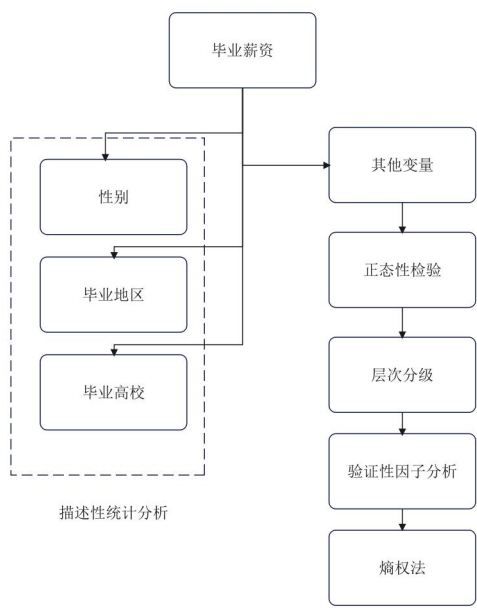


图 1 问题 1 流程图

2.2 问题 2 的分析

问题 2 要求对问题一抽象的因素进行量化建模分析其如何影响薪水，可以对问题一中抽象出的因子进行 OLS 回归，通过多元线性回归可以初步探究其影响关系。而为了进一步对模型进行修正，以防止内生变量过多或对更复杂的作用关系进行建模可以考虑使用调节效应和结构方程等方法。

2.3 问题 3 的分析

针对问题三，基于问题一和二的分析结果提出一系列有价值的意见。

三. 模型假设

- 基于上述分析，我们对模型进行如下假设：
- 1. 假设所有数据真实可靠，受试者在回答问卷时说没有说谎。这一假设可以从问卷的信效度检验中得到，但即使满足信效度检验也不能保证问卷问题设置足够合理或受试者存在说谎现象，故作此假设。
 - 2. 假设印度数据与中国数据有一定共通性，考虑到不同国家国情不同教育背景和水平不同，做这样一个假设的目的是为了消除不同意识形态下的背景差异。
 - 3. 假设不知道受试者的应聘岗位对问题没有太大影响。

四. 符号说明

正文中出现的符号及其释义如下表所示：

符号	说明
\bar{x}	平均值
S^2	方差
n	数值个数
μ	数学期望
t	t 检验统计量
e_i	熵权法权重
y	薪酬
F_1	自身因素
F_2	学习情况
F_3	通用技能
F_4	专业技能
F_5	人格特质

五. 模型建立与求解

5.1 问题一的模型建立与求解

可雇佣能力（employability）是在 20 世纪初的英国最早出现，随后一些外国学者逐步确定了其内涵及外延[1]。可雇佣能力内涵的历史演变、概念及结构变化，与可雇佣能力本身伴随劳动力市场特征不断变化有很大的关系，在适龄劳动力短缺与劳动力市场不活跃时期，对可雇佣能力的研究主要针对社会不利人群和促进全民就业，而随着无边界职业生涯的到来，可雇佣能力则用以培养员工的终身可雇佣能力。对于员工个人来说，为了最大限度地实现自身价值，谋求更好的发展，员工首先需要对自己做出准确评价。岗位胜任力评价就是解决这些问题的关键所在[2,3]。可以说，如何对员工的岗位胜任力进行评价，对每个企业和员工个人都具有极其重要的意义。

我们在这里基于这一数据提出竞争力模型。随着科技的进步和社会的发展，系统工程在国内外呈现出社会化、特色化、专业化、集成化、精细化、实用化等特点。作为社会的重要组成部分，各个机构是一个典型的复杂系统，内部要素之间有机联系、相互作用，并呈现出要素及其层次多、结构复杂和社会性突出等一系列特点。因此，如何通过统筹协调内部因素间关系，达到改善效率的目的，成为越来越多的机构必然选择。具体到教育领域，本科院校承担着人才培养、科学研究、社会服务和文化传承等重要使命，而人才竞争力又是本科院校竞争力的核心。因此，如何有效地对本科院校人才竞争力进行客观公正的评价，显得至关重要[4, 5]。

5.1.1 统计性描述分析

对于所研究的 3000 条样本数据，我们首先进行一些描述性统计分析。

首先可以注意到，在 3000 位调查对象中有 716 名女性，2282 名男性。我们想要探究性别对薪水的影响，故而我们将女性样本和男性样本进行了双样本 t 检验。t 检验也用来判断样本均值和总体均值的显著性差异。很多地方 t 检验和 z 检验类似，但是最大的区别在于总体的理论方差是未知的，t 分布只能用样本数据估计。独立样本 t 检验分析定类数据与定量数据之间的差异，配对样本 t 检验用来揭示定量数据的对比关系，样本先后的顺序要一一对应[1]。配对 t 检验的统计量定义为：

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (1)$$

t 检验的原假设为统计量在两样本中无显著差异，备择假设为两样本中这一统计量有显著性差异。通常若概率小于 0.05，则接受备择假设，显著度 0.05。

经莱文检验（方差齐性检验），检验的概率值为 0.03 拒绝原假设，认为样本不具备方差齐性。故进行方差不齐下的双样本 T 检验，得到 T 检验概率 0.033，满足 0.05 显著度，拒绝原假设，从而认为性别对薪水的影响是显著的，男性应聘者在市场上获得的薪水会相比与女性平均水平高出 10 万美金。性别占比和双样本薪水的箱线图如图 2 所示：

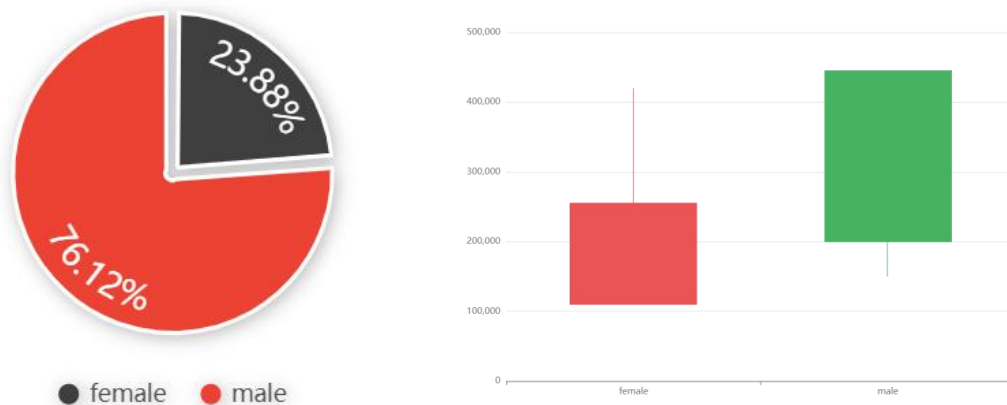


图 2. 性别占比和双样本薪水箱线图

而就受试者的学历和学校而言，如图 2，2757 位受试者均取得本科学历，而 243 位受试者为硕士以上学历，占比不到 10%。这说明社会对于硕士以上学历的高层次人才需求缺口仍然存在。受试者中有 226 名受试者的学校高于平均 AMCAT 分数，也就是常说的“名校”；而受试者的学校所在地中有 888 名受试者的学校是核心城市。同样对受试者是否为名校毕业、是否在核心城市学习抽象为两类定类变量，对其是否影响薪水进行双样本 T 检验，检验结果发现也均满足 0.05 显著度，认为名校学历和学校地区确实对薪水存在显著影响。

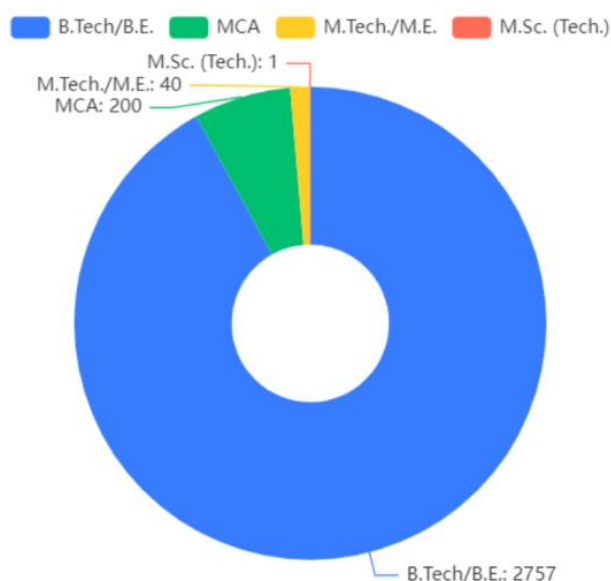


图 3. 受试者的学历占比

对于受试者的专业，我们绘制如图 4 所示的饼状图。可以看到，电类和信息大类学生占比超过 75%。其中包括当下的热门专业电气工程与自动化、通信工程、计算机科学与技术、计算机工程、软件工程、电子信息工程等，而类似于自动控制、管理工程、生物医学工程、航空航天等专业人才相对要少的多。

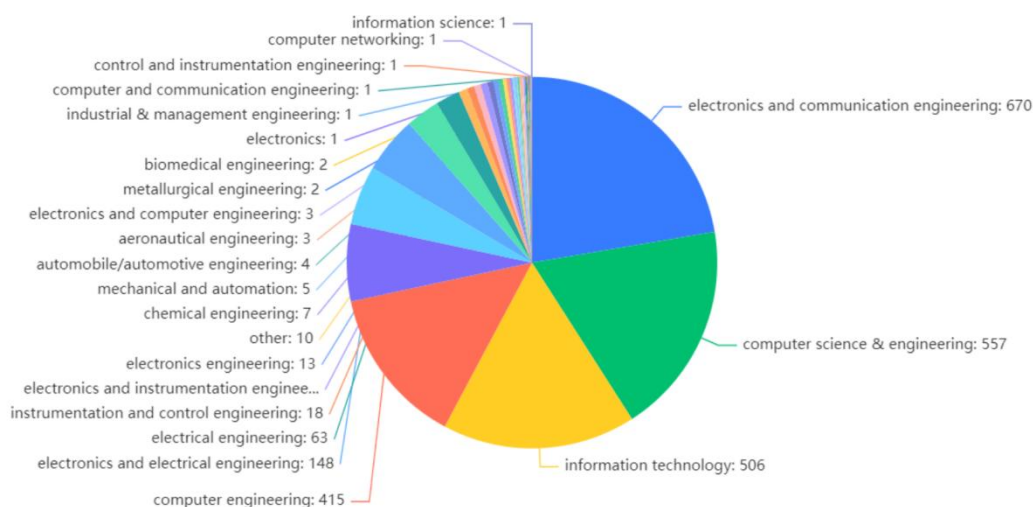


图 4. 受试者的专业分布

受试者学校的所在地也可以从图 5 中的词云图中抽象出一些重要信息。学生最多的犹他州在美国，说明赴美留学进修是此类信息大类人才一条获取更深知识技能的途径。另外，例如孟加拉、新德里等地的学生人数也居高不下，这些地区也是印度电子信息与软件类产业兴盛的地区，说明学生的学习工作与区域的产业布局密切相关。



图 5. 受试者入学的地区

而梅加拉亚邦、比哈尔等印度地区经济发展与相关产业不占优势的地区的毕业生则相对获得不那么高的就业水平，这也反映高校学生的毕业地区会对就业情况产生一定的影响。

5. 1. 2 验证性因子分析

特殊的成长背景和学习环境使当代大学生呈现出与以往数代大学生明显的区别,其就业能力的表象特征,有正面优势也有不可忽视的负面缺陷,要引导大学生重视自身的能力发展。当代大学生就业能力发展的影响因子,主要体现在教育、环境和个性三个方面。可以基于因子分析对学生的竞争力模型进行抽象构造[6,7]。

根据上述分析，结合附表中对受试者各因素所做调查，我们将这些变量进行图 5 所示的层次分级图，将其抽象为 5 类影响因素变量。



图 6. 变量的层次分级图

从图 6 可以看出，附表中的各类参数被我们分为自身因素、学习情况、通用技能、专业技能与性格特征 5 个大类。其中，性别、中学毕业年份、本科毕业年份与专业被归为自身因素；是否名校出身、是否核心城市、中学成绩、本科成绩、GPA 与学历被归为学习情况；量化计算、逻辑推演、英语和领域知识被归为停

用技能；程序设计、电子和半导体工程、计算机科学、机械工程、电气工程、电子信息类和土木工程被归为专业技能；责任心、外向性、认同感、神经质和开放性被归为性格特征。

为验证这一层次分级是否合理，我们对数据进行了验证性因子分析。首先我们对数据进行正态性检验。具体分析结果如表 1 所示。

表 1 正态性检验分析表格

变量	Jarque-Bera 检验				Kolmogorov-Smirnov 检验			
	h	p	sta	cv	h	p	sta	cv
10percentage	1	0.0006	167.21	5.9735	1	0.0021	1	0.0247
12percentage	1	0.0023	50.9	5.9735	1	0.0004	1	0.0247
collegeGPA	1	0.000423	271	5.9735	1	0.0013	0.8	0.0247
English	1	0.0004	32.1	5.9735	1	0.0002	1	0.0247
conscientiousness	1	0.000012	33.4	5.9735	1	0.0021	1	0.0247
agreeableness	1	0.00045	231	5.9735	1	0.0002	0.9	0.0247
extraversion	1	0.00065	33.1	5.9735	1	0.0054	0.7	0.0247
neuroticism	1	0.00054	89.3	5.9735	1	0.0009	1	0.0247
openess_to_experience	1	0.00001	31.6	5.9735	1	0.0032	0.9	0.0247

上表显示了部分变量正态性检验的结果，其中 h 为 1 表示拒绝原假设，接受原始变量符合正态分布的假设。表中无论是通过 Jarque-Bera 检验还是利用 Kolmogorov-Smirnov 检验的方法，各变量的显著性水平 p 值都十分小，例如 openess_to_experience 两种方法的显著性水平 p 达到了 0.0001 与 0.032，与 0.05 相比十分的小，且其他各变量的显著性水平显著都在 0.05 以下。两种检验方法的 sta 都远大于 cv，这也表明上述变量基本符合正态分布的条件。

下图展示了部分变量的正态性拟合效果图。

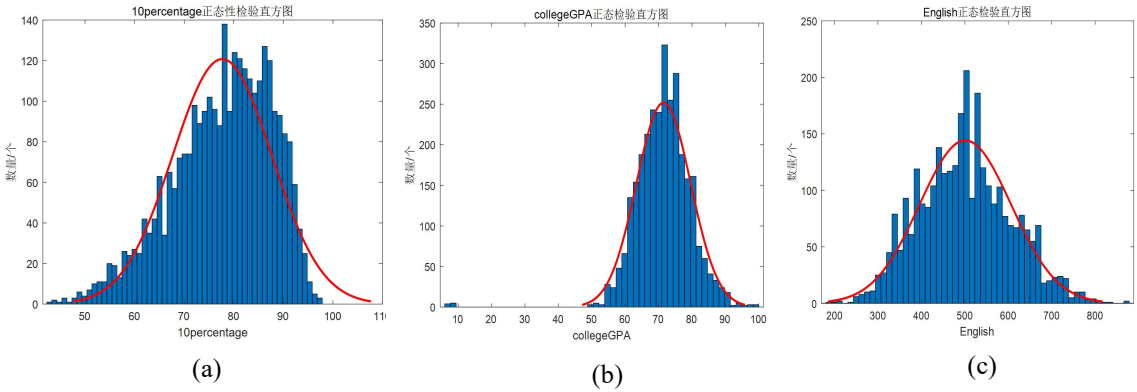


图 7 部分变量正态性拟合效果图

图 7 的(a)、(b)、(c)分别表示 10percentage、collegeGPA、English 变量的正态分布检验直方图，从图可以看出三个变量之间都呈现出中间高两头低的钟状曲线，大体符合正态分布的条件，这与上文的正态性检验结果相一致，表明可以通过表中数据进行验证性因子分析。

本文结合有关理论知识，提出假设，利用附表中的数据拟合特定的因子模型，并进行检验，验证因子分析模型[1]是否成立并估计各个因变量之间的因子载荷值。

验证性因子分析的数学模型表示如下：

$$X = A_x \xi + \delta \quad (1)$$

其中：X 表示 $p \times 1$ 的可观测变量矩阵； A_x 表示 $p \times n$ 阶的待估计因子载荷矩阵； ξ 表示 $n \times 1$ 的潜在公因子向量； δ 为 $p \times 1$ 的测量误差向量。

首先，本文选取合适的公因子数与模型中的相关参数，设立一个确定性模型；接着，利用数据进行模型估计，从上文的正态性检验中可以看出，数据基本符合多元正态的条件，可通过最大似然法进行模型估计；然后，选择因子载荷使模型的相关阵与实际的观测阵最为接近并得到每个指标的相应评价指标；最后，根据模型的检验结果与相关的统计学知识，得到合理的检验模型。

经过分析，我们对五个因子与变量之间的关系进行了相关检验，并得出因子载荷系数表如表 2 所示。

表 2. 因子载荷系数表

因子	变量	非标准载荷系数	标准化载荷系数	Z	S.E.	P
自身因素	Gender	1	0.021	-	-	-
	12graduation	-115.812	-0.64	-6.073	19.069	0.000***
	GraduationYear	-462.582	-0.089	-8.727	53.006	0.000***
	Specialization	119.592	0.223	5.325	22.458	0.000***
学习情况	10percentage	1	0.836	-15.966	3.258	0.000***
	12percentage	1.026	0.771	35.809	0.029	0.000***
	collegeGPA	0.365	0.377	18.862	0.019	0.000***
	Degree	-0.02	-0.328	-16.415	0.001	0.000***
	CollegeCityTier	0.007	0.131	6.503	0.001	0.000***
	CollegeTier	-0.006	-0.186	-9.269	0.001	0.000***
专业技能	ComputerProgramming	1	0.406	-32.697	0.075	0.000***
	ElectronicsAndSemicon	-1.519	-0.659	-35.867	0.042	0.000***
	ComputerScience	1.524	0.615	32.552	0.047	0.000***
	MechanicalEngg	-0.124	-0.093	-4.186	0.03	0.000***
	ElectricalEngg	-0.094	-0.074	-3.33	0.028	0.001***
	TelecomEngg	-0.744	-0.519	-26.089	0.029	0.000***
	CivilEngg	-0.031	-0.067	-3.02	0.01	0.003***
人格技能	conscientiousness	1	0.606	29.001	0.037	0.000***
	agreeableness	1.278	0.83	29.861	0.043	0.000***
	extraversion	0.901	0.581	25.016	0.036	0.000***
	nueroticism	-0.412	-0.253	-12.162	0.034	0.000***
	openess_to_experience	1.14	0.703	28.493	0.04	0.000***
通用技能	English	1	0.685	45.925	0.049	0.000***
	Logical	0.85	0.73	44.927	0.019	0.000***
	Quant	1.153	0.72	43.994	0.026	0.000***
	Domain	0.002	0.258	12.752	0	0.000***
薪水	Salary	1	0	-	-	-

注：***、**、*分别代表 1%、5%、10%的显著性水平

从表 2 中可以看到 12graduation 与自身因素的标准化载荷因子为-0.64 其绝对值大于 0.6；10percentage、12percentage 与学习情况的标准化在和因子的绝对值也都在 0.6 以上；ElectronicsAndSemicon、ComputerScience 与专业技能相关的标准化载荷系数绝对值也在 0.6 以上；conscientiousness、agreeableness、openess_to_experience 与人格技能相关的标准化载荷因子的绝对值分别为 0.606、0.83、0.703，都在 0.6 以上；而 English、Logical、Quant 与通用技能的标准化载荷因子分别为 0.685、0.73、0.72 都近似等于 0.7。

另外，我们可以将因子之间的关联性绘制为图 8 所示的热力图。从图 8 中可以看到，因子之间的关联性普遍不强，稍强一些的是通用技能与学习情况。这表明学校中的学习更多的是影响一些通识基础能力例如数学、英语，而对于专业能力的培养更多还是依靠学生自身。



图 8. 因子之间的关联性热力图

上述变量与各自对应的公因子的标准化载荷系数的绝对值都在 0.6 以上，表明上述变量对公因子贡献了较多的信息，具有更大的贡献率。在保留上述标准化系数较高的变量，删去标准化载荷系数较低的变量后，重新进行验证性因子分析，得到因子载荷系数表如下所示。

表 3 处理后因子载荷系数

因子	因子载荷系数					共同度
	人格技能	通用技能	学习情况	专业技能	自身因素	
conscientiousness	0.761	-0.093	0.121	0.066	0.019	0.607
agreeableness	0.852	0.174	0.039	0.008	-0.007	0.75
openess_to_experience	0.825	0.051	-0.069	-0.017	-0.018	0.688
English	0.087	0.721	0.159	0.118	0.099	0.576
Quant	-0.011	0.772	0.157	-0.158	-0.069	0.651
Logical	0.039	0.82	0.099	0.070	-0.009	0.690
12graduation	0.028	-0.059	0.619	0.462	0.298	0.690
10percentage	0.058	0.335	0.778	-0.092	-0.023	0.730
12percentage	0.028	0.211	0.821	-0.171	-0.072	0.754

ElectronicsAndSemicon	-0.014	-0.025	0.229	-0.666	0.064	0.501
ComputerScience	0.032	0.021	0.090	0.826	-0.077	0.699
Specialization	-0.009	0.026	0.004	-0.123	0.958	0.934

上表展示了经过删除变换处理后各因子与潜变量（各类影响因素）的因子载荷系数，表中字体为蓝色的表明因子载荷系数在 0.6 以上。表中的共同度显示了模型对数据的利用程度，其值越高，表明模型依赖的原始数据越多，表 3 中的共同度值都在 0.5 以上，这说明了模型对原始数据的依赖性较高，具有一定可靠性。而表中各因子与各类影响因素的因子载荷系数在 0.6 以上的，我们则将该因子归为此种影响因素。

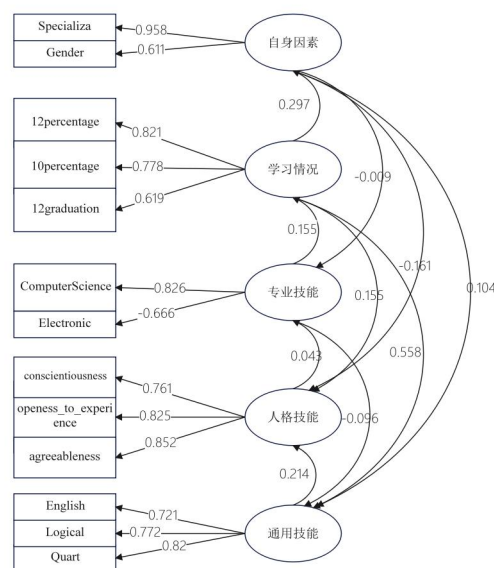


图 9 一阶因子模型

图 9 展示了该竞争力模型的一阶因子模型图，从图可以看出除学习情况与通用技能的相关性在 0.558 以上外，其余各影响因素之间的相关性均不高，都在 0.25 以下，这反映了该模型可以从不同的角度共同测量毕业生的就业竞争力，也侧面反映该模型的一阶因子已经可以对不同因素的影响力进行评价，无需再进行高阶因子模型的讨论。

5.1.3 熵权法构造竞争力模型

为将五个因子构造为整体的竞争力模型，我们采取熵权法为因子赋权。

熵权法是一种为多指标评价提供依据的方法[2]，可以为在此竞争力模型中的多种指标进行赋权。为了客观评价各类指标因素对竞争力模型的权重，本文选取熵权法这一具备多指标综合评价的模型，试图建立各因素与工程类毕业生就业情况的关系，具有更好的客观性。

熵权法的主要计算步骤如下：

Step1:构建 m 个事物 n 个评价指标的判断矩阵 $R=(x_{ij})$, ($i=1, 2, 3, \dots, n$; $j=1, 2, \dots, m$)。此处我们的评价指标包括专业技能、人格特征、自身因素、学习情况和通用技能五个方面。

Step2: 将判断矩阵进行归一化处理，得到新的归一化判断矩阵 B。

$$B = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (4)$$

Step3:熵权法可利用信息熵计算出各指标的权重，从而为多指标评价提供依据。根据信息论中对熵的定义[3]，熵值 e 的计算如下所示

$$e_j = - \frac{\sum_{i=1}^n p_{ij} \ln p_{ij}}{\ln n} \quad (4)$$

通过上式的熵值，可以评价不同指标的离散程度，一般情况下，信息熵越小，离散程度越大，因子对综合评价的权重就越大。

Step4:计算权重系数，式子(5)中代表对于某一个属性 j，第 i 类占样本的比例。n 为属性 j 的取值数量。所以权重系数 w 定义为：

$$w_j = \frac{1 - e_j}{\sum_{i=1}^m (1 - e_i)} \quad (5)$$

最终得到的权重分别为：自身因素 0.253；学习情况 0.164；通用技能 0.265；专业技能 0.261；人格特质 0.057。将权重和因子载荷归一化后，我们可以得到如图 10 所示的竞争力模型图：



图 10. 竞争力模型图

图 10 显示竞争力模型中权重排名前三的因素分别是通用技能、专业技能、自身因素，三者的权重相差不大，都在 0.26 左右，表明这三种因素都是企业招聘时所看重的对象。权重排名第四的因素为学习情况，其权重为 0.164，这与排名前三的因素的权重系数相比有较大的差距，反映了印度企业在选拔员工时对于学习情况并没有过多的看重。人格特征则在竞争力模型中权重占比最小，仅为 0.057，显示出印度企业对于个人的个性要求并不高，反而会更看重毕业生的其

他方面。

对于自身因素，性别与专业在工程类毕业生的就职时占据较大的因素，这反映出部分企业对于就职者的性别拥有一定偏见，而这则可能导致男女的薪资水平的差异；专业在自身因素中的较大占比则表明企业会更加青睐于专业与职位要求更加匹配的毕业生。

对于学习情况，高中表现与本科表现更受到印度企业的重视，反而 GPA 的占比相对较少，这与国内企业的招聘现状有较大的区别；而毕业生的就读院校与所在学校的城市地区也并没有受到更多的关注，但我国的大部分企业却十分看重毕业生的学校出身，这也反映出印度企业更加看重毕业生的个人能力而非学校品牌。

对于通用技能，毕业生的量化计算、逻辑能力与英语水平占比很大，表明印度企业对于工程类专业的项目解决能力与个人交流水平十分看重，体现出印度企业对个人能力的重视。但毕业生的领域掌握能力则占比并不高，这体现出企业面对的多元化与对时代变化的适应度。

对于专业技能，程序设计、计算机科学、电子与半导体和通信工程专业的要求比较高，在 21 世纪的背景下，新型科技产业越发受到企业的重视，而传统的工科类专业诸如机械工程、土木工程与电气工程则到达饱和，企业对其的重视程度也相对较低。

而对于占比最小的人格特质，企业则更加偏爱外向、负责有认同感和开放性的求职者，具备上述人格特质的毕业生在求职时会具备更大的优势。

5.2 问题二的模型建立与求解

5.2.1 基于 OLS 的回归模型构建

为量化衡量五项因子评分与薪酬之间的关系，可以对其进行基于最小二乘法的 OLS 多元线性回归，得到五项归一化的因子评分对薪水的 OLS 回归模型检验如表 4：

表 4 OLS 回归的统计检验

	非标准化系数		标准 化系 数	t	p	R²	调整 R²	F
	B	标准误	Beta					
常数	-331079.38 4	38364.299	-	-8.63	0.000***	0.12	0.118	F=81.58 P=0.000 ***
自身因素	498569.212	44610.384	0.203	11.176	0.000***			
学习情况	106459.669	12313.594	0.17	8.646	0.000***			
通用技能	165190.542	14658.161	0.214	11.27	0.000***			
专业技能	29432.613	10400.969	0.05	2.83	0.005***			
人格特质	-3477.203	12286.986	-0.005	-0.283	0.777			
因变量：薪酬								
注：***、**、*分别代表 1%、5%、10%的显著性水平								

表 4 中显示最终得到的 F 统计量是 0.05 显著的，而自身因素、学习情况、

通用技能和专业技能也都在方程中显著。水平上呈现显著性,拒绝回归系数为 0 的原假设,因此模型基本满足要求。对于变量共线性表现,VIF 全部小于 10,因此模型没有多重共线性问题,模型构建良好。只有人格特质是不太显著的。我们认为,这一方程基本上取得了成功,而人格特质不够显著的原因是因为忽略了一项内生变量,也就是应聘者所面试的岗位。不同的岗位对人格特质的要求是不同的,需要何种人格特质也是不同的,对于本数据而言我们并不知道应聘者所面试的岗位,但从专业来看,受试者大多以信息类行业为主,但也有机械工程、电气工程等传统领域。这也是导致人格特质这一因素没有被很好地解析的原因。

得到方程:

$$y = -331079.384 + 498569.212F_1 + 106459.66F_2 + 165190.542F_3 + 29432.613F_4 - 3477.203F_5 \quad (5)$$

5.2.2 人格特质在竞争力模型中的调节效应

我们认为,人格特质在竞争力模型中可能并不仅仅是直接回归,如果将其抽象为一个调节效应模型可能会更具有说服力。于是,我们以人格特质这一变量为调节变量,以其他四项因子为自变量,薪水为因变量构造调节效应的回归模型。

调节效应的基本步骤如下:

Step 1. 基于调节效应分析表,首先对三个模型的变量进行分析阐述:

针对模型 1:其目的在于研究在不考虑调节变量的干扰时,自变量对于因变量的影响情况;

针对模型 2:模型 2 主要是在模型 1 的基础上加入调节变量;

针对模型 3:模型 3 在模型 2 的基础上加入自变量与调节变量的交互乘积项。

Step 2. 查看模型 2 到模型 3 时,F 值是否呈现显著性变化或者查看模型 3 中交互项的是否呈现显著性(以该项为准)。

Step 3. 若变化呈现出显著性,则意味着存在调节效应。

Step 4. 若以上模型存在调节效应,接着对简单斜率图进行分析,通过观察调节变量在不同水平时,自变量对因变量产生的影响幅度(斜率)差异情况,进而得到具体的调节作用情况。

Step 5. 对分析进行总结。

首先探究的是人格特质对学习情况的调节效应。经检验,调节变量与学习情况的交互项对斜率的影响是显著的,概率值小于 0.05,说明调节变量人格特质对于学习情况对薪酬的影响会产生显著干扰。在学校学习情况不同的人往往性格上有一定程度的差异,学习成绩更好的同学往往更加自信,另外,学校教育也承担了一定立德树人的作用,使学生的人格更加健全。

其次是人格特质对通用技能的调节效应。经检验,人格特质与通用技能的交互项对斜率的影响也同样显著,说明调节变量人格特质对于通用技能对薪酬的影响会产生显著干扰。事实上,良好的人格特质与职场“软技能”同样是一种形式上的通用技能。

而人格特质对专业技能的调节效应,经检验,人格特质与专业技能的交互项并没有太强的显著性,也就是说人格特质并不会显著影响专业技能。这说明 HR 在考察受试者的专业素养时是相对客观公正的,也并不是性格上内向一些的学生就难以掌握专业技能知识。二者之间并没有显著的影响关系。

最后是人格特质对自身因素的调节效应，经检验二者同样存在显著的调节效应。而对于这一层调节效应，是因为自身的性别、学历、年龄等能够对自己在市场上的自信心和活力产生显著影响，从而造成这一调节效应。

四项调节效应对薪水的斜率图如图 11 所示：

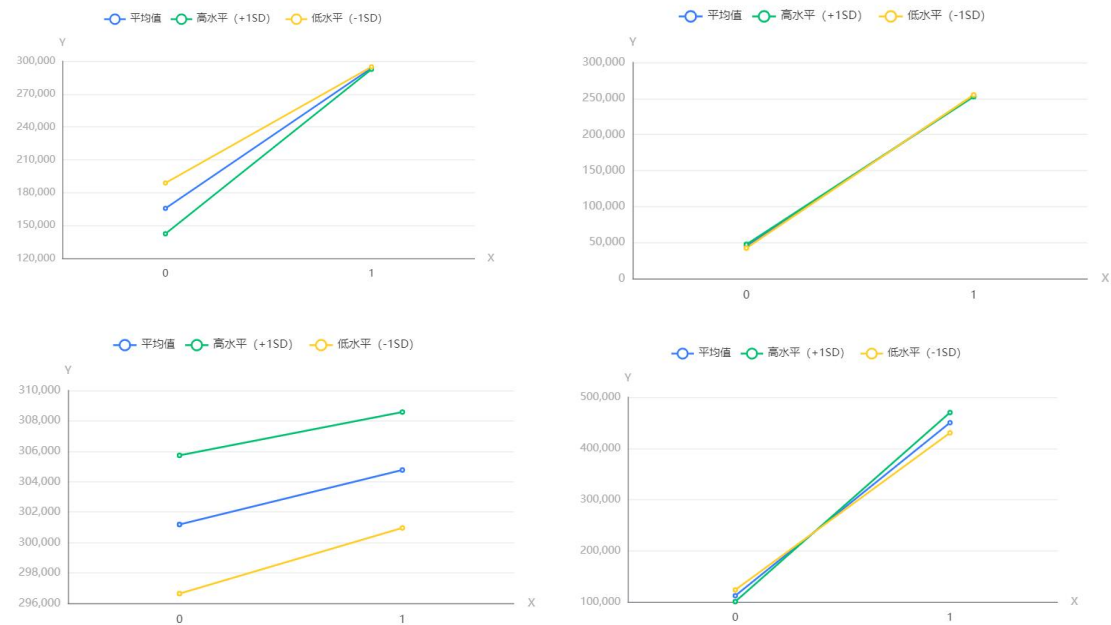


图 11. 四项调节效应对薪水的影响斜率图

图 11 中的四幅图分别表示学习情况、通用技能、专业技能、自身因素。可以看到，人格特质对学习情况的影响最深远，第二是自身因素的调节，第三反倒是通用技能，不显著的是专业技能。

最终我们绘制出如图 12 所示的调节模型架构：

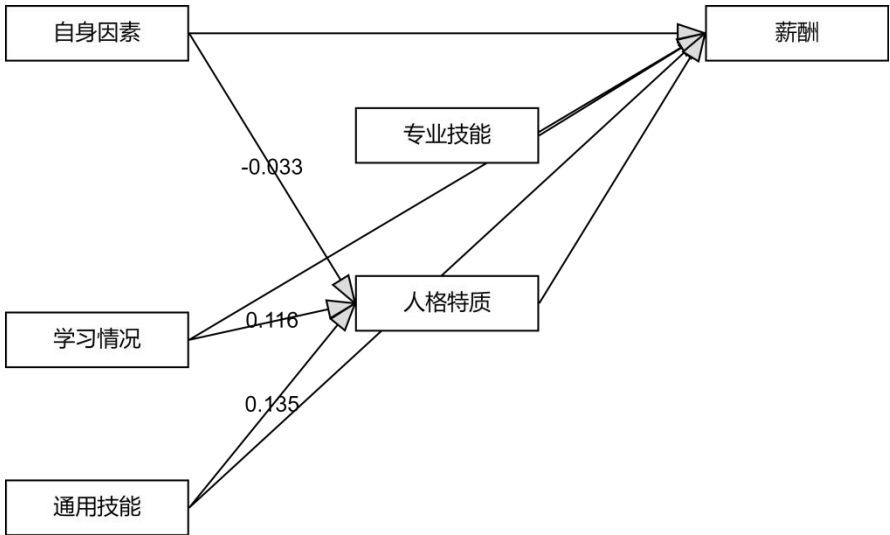


图 12. 调节模型架构

5.3 问题三 模型建立与求解

中国的高等教育学科门类齐全，学生体量庞大，每年向各行业输送大量优秀的工程类专业学生，构建出色的工程类学生培养体系，不断完善工程类学生的培

养策略，是高校建设层面最关键的 task 之一。近年来，高校教育致力于推进科教融合、产教融合的双线培养模式，基于不同培养目标，针对性地进行课程设计，提供对应平台，以达到培养研究型人才或工程型人才的目的。但我们当下在研究型人才培养上的实际投入与输出，仍然远高于工程型人才的培养，造成两类人才的不平衡输出。

在实际的就业市场中，不乏各高校培养出的大量研究型人才。而对于以产品为导向企业，在推动技术革新的过程中，各高校毕业生的工程能力与素养成为一大制约因素。这就意味着，传统的科教融合体系下培养出的高校毕业生，与企业的需求存在不同程度的错位。一方面，各大高校输出大量研究型人才，造成部分行业人才溢出甚至饱和；另一方面，产教融合体系下培养出的卓越工程师，却供不应求。因此，调整我国对于工程类人才的培养体系，是推动我国工程类毕业生就业的一大重要手段，更是推动产业发展的重要策略。而基于我们在 5.1，5.2 中对各因子与就业能力的关联程度，分析，我们诚恳地提出以下几方面建议：

(1) 制定特色课程体系，打造优质工程人才

当下除个别院校和各学院内设立的少量实验班外，对大部分工程类学生的课程设计均采用通识课程到专业课程的设计，课程结构与内容类似，没有在后后期进行研究型人才和工程型人才的分化课程体系设置。因此，我们建议，在保留通识课程和重要专业课程的基础上，对于倾向于选择就业的学生，提高该类学生课程体系中实验课程、实践课程的占比，以课程为导向，培养学生解决实际工程问题的意识和能力。

(2) 引进工程专业导师，加强学校企业联合

在目前的导师体系中，我们以研究型导师居多，致力于应用方面的工程硕士，博士数量较少，对此，我们建议，我国应加大对工程方面高素质人才导师的引进，调整研究型人才和应用型人才的比例，实现对工程类学生的有效培养。同时，学校应加强与各科技企业的合作，探索实行高校和企业联合培养高素质复合型工科人才的有效机制，实现产学研深度融合，从而解决工程技术人才培养与实际生产过程脱节的问题。

(3) 调整学生评价体系，合理引导学生选择

近年来大部分理工科院校存在评价体系“唯绩点，唯论文，唯奖项”的风气，针对评价体系趋同化的问题，我国应积极营造多元评价的氛围，鼓励学生进行多方面的尝试，引导学生动手探索，解决实际问题，真正落实科研人才与工程人才并重，从而增加工程技术人才的输入与输出，既丰富了人才体系，又能对口企业实际需求，从一定程度上提高了本校学生的就业率与就业质量。

(4) 提升学生综合能力，树立正确价值观念

除专业素养，工程能力两方面，学生个性与社会的适配度，作为隐性能力，同样在就业问题上占据重要地位。因此，我们建议学校加强对学生人格体系的培养：一方面，学校可开设相应课程，进行相应的人格测试等，针对部分可能不适应社会环境的学生提供咨询和帮助；另一方面，每一学院的各层级老师，应紧密关注学生状况，积极引导学生完善健全自身人格。

六. 模型优缺点分析

6.1 模型优点分析

我们的模型形式简洁方法科学，在经过合理评估后，我们认为我们的模型具有如下几个方面的优势：

1.模型形式简洁，通过基本的计量手段揭示了抽象的社会科学变量之间存在的统计建模关系，并通过量化方法对问题提出合理的建议。

2.模型策略合理，充分排除了无关变量干扰。这里经过多重对比，使用的调节变量法相比于普通 OLS 回归不仅拟合效果上得到了提升，更多变量产生了显著效应并且有效消除了内生性因素。

3.模型具有良好的可解释性。从模型的统计结果与显著性可以清晰看出变量之间的作用关系以及作用程度，能够比较直观地揭示其作用机理并映证当下社会热点问题，能够给出一些参考意见。

4.模型形式直观，无论是路径图还是竞争力模型等都能直观反映研究结论。

6.2 模型缺点分析

但就客观而言，我们认为我们的模型还存在以下不足：

1.数据体量不够，样本单一，这一问题主要是受限于印度问卷调查难以铺开且专业过于集中在信息类。并且本文暂时还没有分区域进行实证，没有考虑到地缘因素与经济发展水平差异对问题的影响。

2.模型策略简单，内生性问题未能充分消除。这一问题主要是没有获得受试者应聘的岗位导致。

6.3 模型的扩展

由于目前的数据量不足，且中国和印度国情不同，无法得到足够的统计数据，在充分调研各不同地区的相关变量与发展水平以及获取更长时间的面板以后，可以通过上述手段对问题进行更为合理的建模，对应聘者竞争力模型做出更加精准的预测，从而更好地刻画我国青年就业竞争力并给出相应对策。

参考文献

- [1]王晔,隗媛.可雇佣能力内涵及影响因素研究[J].吉林省经济管理干部学院学报,2017,31(02):51-55.DOI:10.19448/j.cnki.1009-0657.2017.0050.
- [2]李明斐,卢小君.胜任力与胜任力模型构建方法研究[J].大连理工大学学报(社会科学版),2004(01):28-32.DOI:10.19525/j.issn1008-407x.2004.01.006.
- [3]冯明,尹明鑫.胜任力模型构建方法综述[J].科技管理研究,2007(09):229-230+233.
- [4]马明.基于AHP-网络DEA模型的本科院校人才竞争力评价研究[D].南京理工大学,2018.
- [5]曹丽娜.企业人才竞争力的评价模型构建[J].商场现代化,2010(26):128-129.
- [6]焦静,冯明.大学生就业胜任力模型的构建[J].科学决策,2010(10):59-67+87.
- [7]王世通,谢爱国.基于因子分析的大学生就业能力影响因素研究[J].当代经济,2013(21):140-142.
- [8]刘亭亭,于晓辉,吕大刚,基于偏最小二乘的地震动负荷强度参数构造与统计性分析[J],建筑结构学报,2020年S1期:23-25,2020.
- [9]朱晓霞,肖群,国际物流服务供应链合作风险的熵权法评价模型研究[J],太原师范学院学报,20-1:36-40,2021.
- [10]刘红雨,刘友存,孟丽红,熵权法在水资源与水环境评价中的研究进展[J],冰川冻土,44-01:300-304,2022.
- [11]文化创新比较研究所,关于计量资料中的t检验[J],34:130,2022.
- [12]毕向阳,基于多水平验证性因子分析的城市社区社会资本测量[J],社会学研究,19934:213-217,2019.
- [13]胡良,姚泽清,基于相关系数的总体熵权法及其应用[J],解放军理工大学学报,2018:2-3,2018.

附录

工具: SPSSPRO, echart.js

文件列表: 因子.xlsx, 处理.py, 承诺书.pdf

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
data=pd.read_excel("附件.xlsx")
data['Gender']=le.fit_transform(data['Gender'])
data['Degree']=le.fit_transform(data['Degree'])
data['Specialization']=le.fit_transform(data['Specialization'])
data=(data-data.min())/(data.max()-data.min())
factor=pd.DataFrame({})
factor['自身因素']=0.021*data['Gender']-0.64*data['12graduation']-\
    0.089*data['GraduationYear']+0.223*data['Specialization']+1
factor['学习情况']=data['10percentage']*0.836+data['12percentage']*0.771+\
    data['collegeGPA']*0.377-data['Degree']*0.328+data['CollegeCityTier']*0.131-\
    data['CollegeTier']*0.186
factor['通用技能']
']=data['English']*0.685+data['Logical']*0.73+data['Quant']*0.72+data['Domain']*0.258
factor['专业技能']=data['ComputerProgramming']*0.406-data['ElectronicsAndSemicon']*0.659+\
    data['ComputerScience']*0.615-data['MechanicalEngg']*0.093-data['ElectricalEngg']*0.074-\
    data['TelecomEngg']*0.519-data['CivilEngg']*0.067+1
factor['人格特质']=data['conscientiousness']*0.606+data['agreeableness']*0.83+\
    data['extraversion']*0.581-data['nueroticism']*0.253+data['openess_to_experience']*0.703-1
```