

决定收入高低的因素都有哪些？

摘要与引言

本研究分析中国居民税后工资性收入的决定因素，意义在于揭示收入不平等的根源，为政策制定（如教育公平和区域平衡）提供依据。数据源自中国家庭追踪调查（CFPS）2022年个人问卷，样本量8168。方法采用OLS回归，控制年龄、健康状况和省份固定效应。核心发现显示，教育年限每增加一年收入显著提升约4406元（ $p < 0.001$ ），工作时间每增加一小时收入上升149元（ $p < 0.001$ ），同时性别（男性高20270元）和城乡分类（城镇高3700元）凸显结构性差距。结果表明，人力资本投资和劳动供给是收入增长的关键，但性别与城乡不平等需政策干预。（字数：198）

研究计划

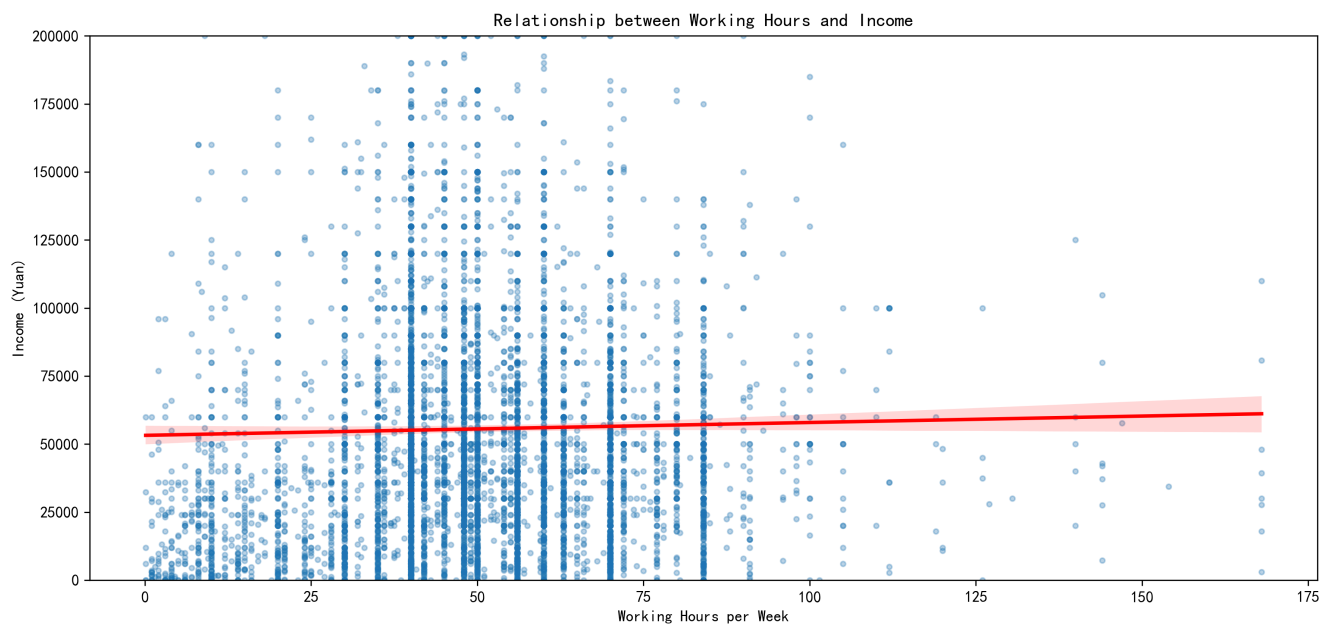
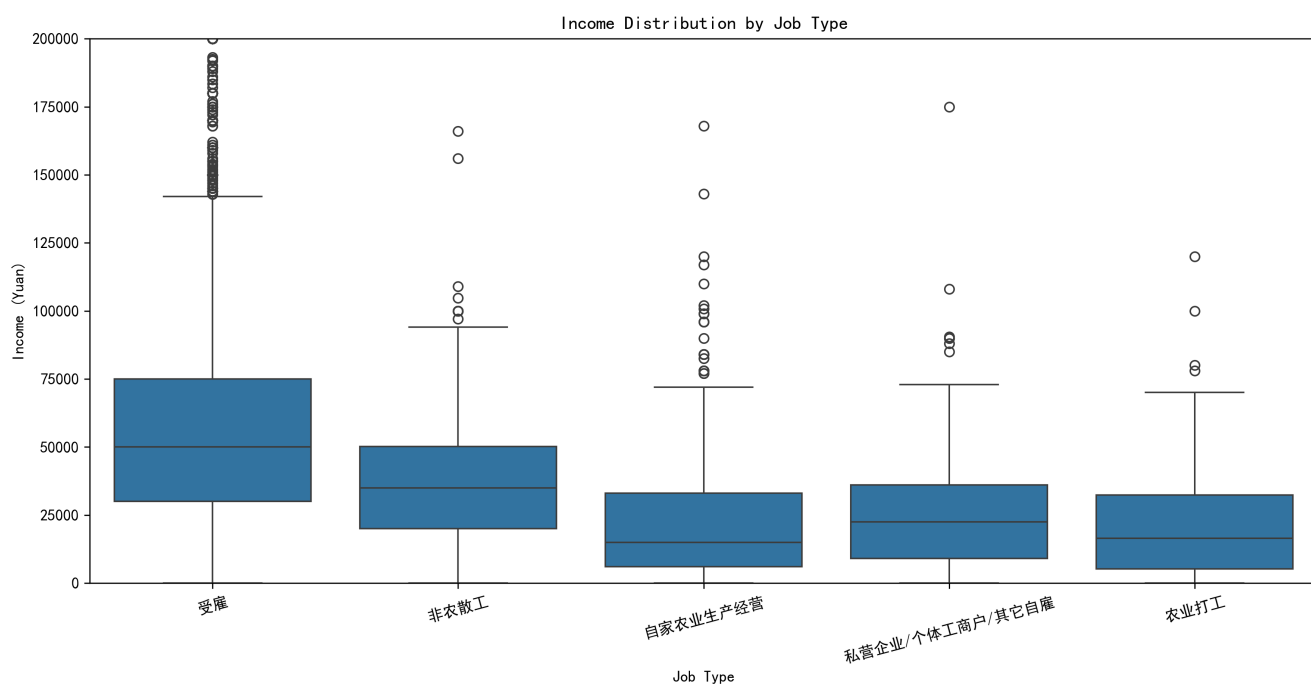
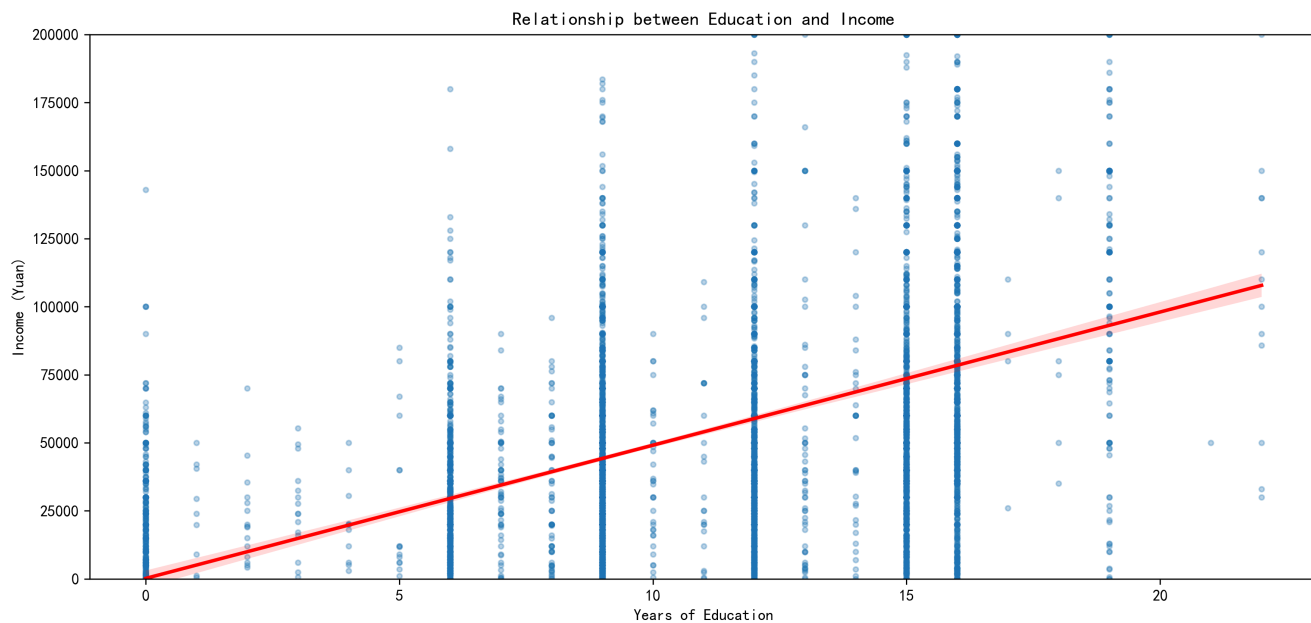
核心研究问题：本研究旨在识别影响中国居民收入水平的关键个人特征和职业因素，具体探究教育、工作经验、工作类型等变量如何显著决定收入高低。基于人力资本理论和社会经济框架，问题可操作化为：在中国背景下，个人教育水平、职业特征及人口特征对工资性收入的边际贡献如何？

计量经济模型：采用普通最小二乘（OLS）回归模型，因收入（'emp_income'）为连续变量。模型形式为：收入 = $\beta_0 + \beta_1 \text{教育年限} + \beta_2 \text{工作类型} + \beta_3 \text{工作时间} + \sum \beta_i \text{控制变量} + \epsilon$ 。若教育存在内生性（如能力偏误），可扩展为两阶段最小二乘（IV-2SLS）模型，使用父母教育作为工具变量。

关键变量：

- 因变量：收入（'emp_income'），因其直接测量税后工资性收入，且为收入高低的代理变量。
- 核心自变量：教育年限（'cfps2022eduy'），预期正相关（人力资本理论）；工作类型（'jobclass'），预期受雇类（如代码4）收入更高；每周工作时间（'qg6'），预期正相关（劳动供给理论）。
- 控制变量：年龄（'age'）控制生命周期效应；性别（'gender'）控制性别差距（预期男性收入更高）；健康状况（'qp201'）缓解健康遗漏偏误；城乡分类（'urban22'）控制地域经济差异；省份（'provcd22'）作为固定效应控制区域异质性。

识别策略：初步使用OLS，但针对教育内生性，建议以父母教育（'qv102', 'qv202'）为工具变量进行稳健性检验（2SLS），确保因果推断的可靠性。数据处理中将排除缺失值（如-8、-9）。



回归结果

OLS Regression Results

=====						
Dep. Variable:	emp_income	R-squared:	0.239			
Model:	OLS	Adj. R-squared:	0.235			
Method:	Least Squares	F-statistic:	63.64			
Date:	Thu, 29 May 2025	Prob (F-statistic):	0.00			
Time:	11:25:57	Log-Likelihood:	-99688.			
No. Observations:	8168	AIC:	1.995e+05			
Df Residuals:	8127	BIC:	1.997e+05			
Df Model:	40					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	4534.2542	6087.467	0.745	0.456	-7398.740	1.65e+04
cfps2022eduy	4405.7257	166.092	26.526	0.000	4080.144	4731.308
qg6	148.8826	30.499	4.882	0.000	89.097	208.668
age	-11.2803	52.507	-0.215	0.830	-114.208	91.648
gender	2.027e+04	1107.850	18.297	0.000	1.81e+04	2.24e+04
qp201	748.2083	527.252	1.419	0.156	-285.341	1781.758
urban22	3699.8047	1224.057	3.023	0.003	1300.339	6099.270
job_2.0	-2.108e+04	4925.790	-4.279	0.000	-3.07e+04	-1.14e+04
job_3.0	-713.5133	5337.331	-0.134	0.894	-1.12e+04	9749.021
job_4.0	5858.2066	2261.487	2.590	0.010	1425.114	1.03e+04
job_5.0	1774.6133	3423.295	0.518	0.604	-4935.922	8485.148
prov_12.0	-2.12e+04	6639.626	-3.193	0.001	-3.42e+04	-8185.976
prov_13.0	-3.764e+04	4724.491	-7.967	0.000	-4.69e+04	-2.84e+04
prov_14.0	-3.704e+04	4961.944	-7.465	0.000	-4.68e+04	-2.73e+04
prov_15.0	-5795.3642	1.52e+04	-0.380	0.704	-3.57e+04	2.41e+04
prov_21.0	-3.638e+04	4677.189	-7.779	0.000	-4.56e+04	-2.72e+04
prov_22.0	-3.845e+04	5832.004	-6.592	0.000	-4.99e+04	-2.7e+04
prov_23.0	-4.246e+04	5406.179	-7.854	0.000	-5.31e+04	-3.19e+04
prov_31.0	1.506e+04	4868.265	3.093	0.002	5512.530	2.46e+04
prov_32.0	-8241.4338	5353.562	-1.539	0.124	-1.87e+04	2252.918
prov_33.0	-8637.5703	5272.351	-1.638	0.101	-1.9e+04	1697.588
prov_34.0	-1.387e+04	5652.028	-2.453	0.014	-2.49e+04	-2785.687
prov_35.0	-1.088e+04	6695.868	-1.625	0.104	-2.4e+04	2242.920
prov_36.0	-2.823e+04	5815.232	-4.855	0.000	-3.96e+04	-1.68e+04
prov_37.0	-3.024e+04	4807.879	-6.289	0.000	-3.97e+04	-2.08e+04
prov_41.0	-3.388e+04	4565.458	-7.422	0.000	-4.28e+04	-2.49e+04
prov_42.0	-2.056e+04	6182.502	-3.325	0.001	-3.27e+04	-8439.937
prov_43.0	-2.539e+04	5335.769	-4.758	0.000	-3.58e+04	-1.49e+04
prov_44.0	-1.507e+04	4587.946	-3.285	0.001	-2.41e+04	-6078.126
prov_45.0	-3.699e+04	5899.240	-6.271	0.000	-4.86e+04	-2.54e+04
prov_46.0	1.195e+04	2.02e+04	0.590	0.555	-2.77e+04	5.16e+04
prov_50.0	-1.858e+04	7967.230	-2.333	0.020	-3.42e+04	-2966.668
prov_51.0	-2.967e+04	5104.117	-5.812	0.000	-3.97e+04	-1.97e+04
prov_52.0	-3.534e+04	5637.694	-6.269	0.000	-4.64e+04	-2.43e+04
prov_53.0	-3.782e+04	5498.275	-6.878	0.000	-4.86e+04	-2.7e+04
prov_54.0	2.222e+04	2.21e+04	1.005	0.315	-2.11e+04	6.56e+04

prov_61.0	-3.039e+04	5515.530	-5.511	0.000	-4.12e+04	-1.96e+04
prov_62.0	-3.581e+04	4617.835	-7.754	0.000	-4.49e+04	-2.68e+04
prov_63.0	-2.977e+04	1.46e+04	-2.033	0.042	-5.85e+04	-1070.877
prov_64.0	-2.413e+04	1.42e+04	-1.705	0.088	-5.19e+04	3619.485
prov_65.0	-3.302e+04	7506.694	-4.399	0.000	-4.77e+04	-1.83e+04
=====						
Omnibus:	9733.798	Durbin-Watson:	1.774			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2887998.587			
Skew:	5.980	Prob(JB):	0.00			
Kurtosis:	94.339	Cond. No.	3.26e+03			
=====						
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[2] The condition number is large, 3.26e+03. This might indicate that there are strong multicollinearity or other numerical problems.						

结果解读

任务1：自变量对因变量的经济含义（300字内）

基于OLS回归结果，自变量对因变量“过去12个月所有工作（主要工作+一般工作）的税后工资性收入”的经济含义如下（引用中文标签）：

- **CFPS2022个人问卷受访者已完成的受教育年限**：系数为4405.73 ($p<0.001$)，显著正相关，表明每增加一年教育，收入平均提升约4406元，符合人力资本理论，教育投资通过提升技能增加收入。
- **每周工作时间(小时/周)**：系数为148.88 ($p<0.001$)，显著正相关，每增加一小时工作，收入上升约149元，反映劳动供给效应，工作时间延长直接贡献收入。
- **受访者性别**：系数为20270 ($p<0.001$)，显著正相关（性别=1为男性），男性收入平均比女性高20270元，凸显性别收入差距，可能源于劳动力市场歧视或职业隔离。
- **基于国家统计局资料的城乡分类**：系数为3699.80 ($p=0.003$)，显著正相关（分类=1为城镇），城镇居民收入比乡村高约3700元，体现城乡经济鸿沟，城镇就业机会和工资水平更高。
- **当前最主要工作/最近结束的工作类型**：以“自家农业生产经营”(参考类)为基准，系数显示“受雇”工作收入显著高5858元 ($p=0.010$)，表明正规雇佣提升收入；“私营企业/个体工商户/其它自雇”收入显著低21080元 ($p<0.001$)，反映自雇风险或收入不稳定；其他类型不显著。
- **年龄和健康状况**：系数不显著 ($p>0.05$)，年龄每增加一岁收入变化不显著 (-11.28元)，健康对收入影响未验证，可能被其他因素抵消。
- **2022年省国标码**：作为固定效应，区域差异显著（如上海收入高、河北收入低），但不单独解释。

任务2：研究发现总结（400字内）

本研究通过OLS回归分析中国居民收入影响因素，样本量8168， R^2 为0.239，模型部分解释收入变异。核心发现如下：

教育和工作时间是收入的关键驱动因素。教育年限每增加一年，收入显著提升约4406元 ($p<0.001$)，验证人力资本理论，表明教育投资是收入增长的核心途径。每周工作时间每增加一小时，收入上升约149元 ($p<0.001$)，符合劳动供给理论，强调劳动投入的直接回报。

社会人口因素显示显著不平等：男性收入平均比女性高20270元 ($p<0.001$)，反映持续性别差距；城镇居民收入比乡村高约3700元 ($p=0.003$)，凸显城乡二元结构的经济影响。工作类型分析揭示，受雇工作比自家农业收入高5858元

($p=0.010$)，但私营企业收入低21080元 ($p<0.001$)，说明正规就业更有利，而自雇风险较高。

控制变量中，年龄和健康对收入影响不显著，可能因样本偏差或生命周期效应未捕获。省份固定效应确认区域差异（如东部沿海收入较高），但未改变核心结论。模型存在潜在多重共线性（条件数 $3.26e+03$ ），建议未来使用工具变量（如父母教育）缓解内生性问题。

总之，教育、劳动供给、性别和城乡因素是收入决定的核心，政策应聚焦教育公平、劳动保护和区域平衡发展以缩小收入差距。