

山东省粮食产量的影响因素与预测模型研究

——基于 2000—2025 年数据的统计分析

北京大学附属中学预科部 高泽宇

摘要：本研究以山东省 2000—2025 年粮食产量数据为样本，选取粮食播种面积、有效灌溉面积、农作物受灾面积三个变量，分别代表土地、水利与自然灾害，用最小二乘法建立三个预测模型，并从预测精度与稳健性两方面加以检验。结果表明：全省粮食总产量由 3835 万吨增至 5710.6 万吨，增长 48.9%，其中单位面积产量提高贡献 61.7%，播种面积扩大仅贡献 29.4%，增产路径已由扩面积转向提单产。三因素多元回归模型效果最优， R^2 为 0.9760，经 2023—2025 年三轮滚动检验平均误差率 2.085%，据此预测 2026 年产量约 5635 万吨，与 2025 年基本持平。本研究还发现：受灾面积与产量的相关系数虽达-0.818，但控制时间趋势后即失去显著性（ $p=0.345$ ），说明该相关主要源于二者共享同一时间趋势，而非因果联系。

关键词：粮食产量；播种面积；线性回归；预测模型；滚动检验

1. 研究背景与意义

1.1 研究背景

粮食安全是国家安全的基础。山东省位于黄河下游、我国东部沿海，属暖温带季风气候，地形以平原和丘陵为主，冬小麦和夏玉米一年两熟，是全国重要的粮食主产省之一。2025 年全省粮食总产量达到 5710.6 万吨，位居全国前列。

但山东的粮食生产也受到不少限制。春季降水少、蒸发强，容易发生干旱；夏季降水集中，又可能出现涝灾。同时，随着城镇化推进，耕地资源越来越紧张。在这种情况下，弄清楚粮食产量到底受哪些因素影响、影响有多大、能不能据此预测未来产量，就成了一个值得研究的地理问题。

1.2 研究意义

从地理学习的角度看，农业生产是自然条件和社会经济条件共同作用的结果。本研究把课本上抽象的“农业区位因素”变成可以测量的数据，用统计方法检验

土地、水利、灾害这三类因素与产量的数量关系，是对“农业区位”和“粮食安全”这两个知识点的一次实际应用。

从方法上看，本研究主要采用的相关系数和线性回归分析方法。但研究并未停留在“算出回归系数、看 p 值是否小于 0.05”这一步，而是进一步追问：换一种模型设定，这个结论是否依然成立？正是这一追问，修正了本研究的部分初步结论。

2. 研究设计

2.1 研究地区概况

山东省位于北纬 34° 22′ 至 38° 23′ 、东经 114° 47′ 至 122° 42′ 之间，地处黄河下游、黄淮海平原东部，东临渤海与黄海。全省属暖温带季风气候，年均气温 11℃至 14℃，年降水量 550 至 950 毫米，且降水的七成以上集中于夏季。地形以平原和丘陵为主，鲁西北、鲁西南为黄河冲积平原，土层深厚、地势平坦，是粮食生产的主体区域。

从农业地域类型看，山东属于典型的商品谷物农业与混合农业交错区，冬小麦—夏玉米一年两熟是最主要的种植制度。这一制度对水分条件高度依赖：冬小麦的返青、拔节、灌浆期恰逢春季少雨时段，灌溉能否及时保证，直接关系到当季产量；而夏玉米生长期与雨季重合，又面临涝渍和风雹的风险。土地、水利、灾害三者对山东粮食生产的制约，正是本研究选取这三个变量的地理依据。

2.2 数据来源

研究数据中粮食总产量与播种面积的数据采集自国家统计局历年粮食产量数据的公告；有效灌溉面积采集自聚汇数据网；受灾面积数据采集自山东省应急管理厅新闻报道的数据。表 1 给出了 2000-2025 年山东省粮食产量数据分布，作为后续预测模型分析的数据基础。

表 1 2000—2025 年山东省粮食产量分布

年份	粮食总产量 (万吨)	播种面积 (万公顷)	有效灌溉面积 (万公顷)	受灾面积 (万公顷)	单位面积产量 (公斤/公顷)
2000	3835.0	736.32	482.49	369.70	5208
2001	3721.0	715.35	483.60	359.60	5202
2002	3747.0	691.26	479.74	494.19	5421
2003	3420.0	641.54	476.08	263.20	5331

2004	3509.0	617.63	476.68	211.91	5681
2005	3917.0	671.17	479.00	177.98	5836
2006	4048.0	699.91	481.82	188.01	5784
2007	4145.0	696.72	483.68	186.54	5949
2008	4264.0	715.43	485.75	67.23	5960
2009	4316.0	729.77	489.69	234.19	5914
2010	4502.8	745.11	495.53	258.23	6043
2011	4701.3	761.67	498.69	211.72	6172
2012	4815.8	775.04	499.70	182.26	6214
2013	4883.4	800.76	472.90	146.17	6099
2014	5038.3	827.77	490.19	88.61	6087
2015	5153.1	840.70	496.44	137.90	6130
2016	5332.3	851.73	516.12	55.22	6261
2017	5374.3	845.56	519.11	85.69	6356
2018	5319.5	840.48	523.60	98.38	6329
2019	5357.0	831.28	527.14	134.07	6444
2020	5447.0	828.20	529.36	38.22	6577
2021	5500.7	835.51	506.99	10.87	6584
2022	5543.8	838.51	520.91	4.54	6611
2023	5655.3	838.79	524.17	1.74	6742
2024	5710.2	841.26	527.14	12.15	6788
2025	5710.6	842.26	529.00	8.46	6780

（注：数据采集国家统计局官网、聚汇数据网、山东省应急管理厅数据；前五列为原始数据，第六列为计算所得：单位面积产量=粮食总产量×1000÷播种面积。）

2.3 变量设定及基本统计特征

本研究以粮食总产量为因变量，选取三个自变量，其地理含义与理论预期如表 2 所示。三个自变量的选择遵循“一个维度一个代表变量”的原则：X₁代表土地要素，X₂代表水利要素，X₃代表自然灾害。辅助变量 T 的设定需要特别说明——时间本身不可能“导致”粮食增产，T 在这里只是一个综合趋势指标，用来概括良种培育、化肥农药、农业机械、高标准农田建设、农业政策等一系列随时间稳步推进却难以逐项量化的因素。

表 2 变量设定

类型	变量	符号	单位	代表的地理含义	预期方向
因变量	粮食总产量	Y	万吨	粮食生产的最终结果	—

自变量	粮食播种面积	X_1	万公顷	土地投入，种植规模	正
自变量	有效灌溉面积	X_2	万公顷	水利条件，抗旱能力	正
自变量	农作物受灾面积	X_3	万公顷	旱涝风雹等自然灾害	负
辅助变量	时间序号	T	—	技术进步等长期因素的总代表	正

26 年样本的基本统计特征如下表 3。表中最值得注意的是变异系数（数据标准方差/均值）这一列，它反映一个变量波动的剧烈程度。受灾面积的变异系数高达 81.1%，远高于播种面积的 9.5%和灌溉面积的 3.9%。这说明土地和水利是人可以长期规划、慢慢改善的“慢变量”，年际变化很小；而自然灾害是“快变量”，好年份和坏年份能差出几百倍。

表 3 主要变量的基本情况

变量	单位	均值	最小值（年份）	最大值（年份）	变异系数(%)
粮食总产量	万吨	4729.5	3420.0 (2003)	5710.6 (2025)	16.0
粮食播种面积	万公顷	771.5	617.6 (2004)	851.7 (2016)	9.5
有效灌溉面积	万公顷	499.8	472.9 (2013)	529.4 (2020)	3.9
农作物受灾面积	万公顷	154.9	1.7 (2023)	494.2 (2002)	81.1

（注：变异系数 = 标准差 ÷ 均值 × 100%）

2.4 研究方案设计

本研究按“提出问题—数据与变量—建立模型—检验验证—结论与应用”五个环节依次展开（见图 1）。第一环节从山东粮食生产面临的地理约束入手，提出三个层层递进的问题：产量究竟受哪些因素影响，各因素的作用有多大，以及能否据此预测未来产量。第二环节确定研究对象，以 2000—2025 年共 26 个年度样本为基础，取粮食总产量为因变量，取粮食播种面积、有效灌溉面积和农作物受灾面积为自变量，分别代表土地、水利和自然灾害三个方面，并计算单位面积产量、受灾率等派生指标。第三环节用最小二乘法由简到繁建立三个模型：模型一只用时间序号，模型二加入播种面积，模型三综合三个因素。第四环节从拟合优度 R^2 、残差分析和 2023 至 2025 年的滚动事后检验衡量模型的预测能力，另一方面用去趋势偏相关检验和加入时间趋势项的模型四，检验各系数是否稳健。第五环节为结论。

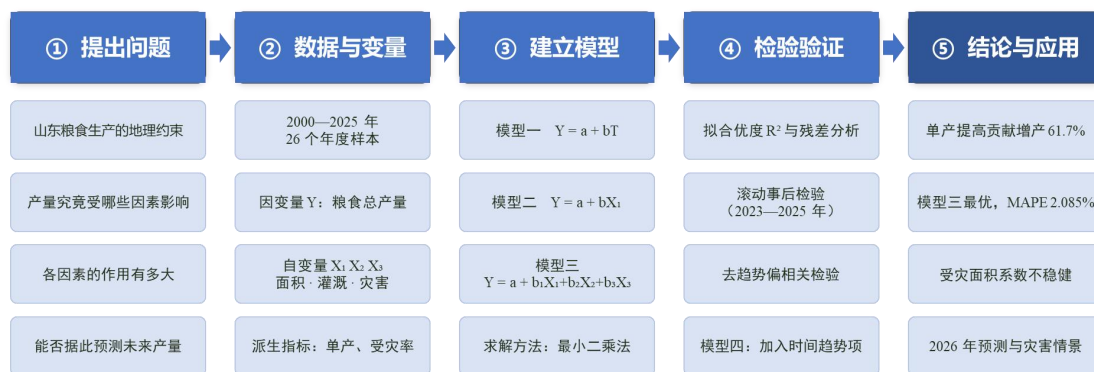


图 1 研究框架图

本研究主要利用相关性分析与线性回归分析;统计分析工具利用大模型(豆包)与 excel 完成。**相关分析**主要用 Pearson 相关系数 r 衡量两个变量的线性关系, r 越接近 1 或-1 关系越强, 正负号表示方向。Excel 函数为=CORREL(产量区域, 自变量区域)。**线性回归**中主要用一个线性方程来描述自变量对产量的影响, 并据此预测。一元回归用=SLOPE()、=INTERCEPT()、=RSQ() 三个函数;多元回归利用大模型辅助完成分析, 并进行了交叉印证。

3. 预测模型的建立与检验

本研究按由简到繁的顺序建立三个模型: 模型一只使用时间, 模型二加入土地因素, 模型三综合土地、水利、灾害三个方面。三个模型的建立都遵循同样的五个步骤: 确定自变量、写出模型形式、用最小二乘法求系数、计算 R^2 判断拟合效果、解释系数的实际含义。

3.1 模型一: 时间趋势模型

模型一主要考察产量随时间变化的趋势, 即用年份来预测产量。第一步, 确定自变量, 确定模型公式。设 2000 年为 $T=1$, 2001 年为 $T=2$, 依此类推, 具体公式如下:

$$Y = a + bT$$

式中 Y 为粮食总产量 (万吨), a 为常数项, b 为斜率, 即时间序号每增加 1 时产量的平均变化量。

第二步, 利用最小二乘法求出 a 和 b 。任何一条直线都会与 26 个实际数据点存在偏差, 把每年的实际值减去直线上的对应值, 称为残差。残差有正有负,

直接相加会互相抵消，因此先平方再求和，得到残差平方和。最小二乘法的原则，就是选取使残差平方和最小的那条直线。由此可以推出两个系数的计算公式：

$$b = \Sigma(T-\bar{T})(Y-\bar{Y}) \div \Sigma(T-\bar{T})^2 \quad a = \bar{Y} - b \cdot \bar{T}$$

代入本研究的数据。先求两个平均值：T 的平均值 $\bar{T} = 13.5$ ，产量的平均值 $\bar{Y} = 4729.52$ 万吨。再分别求出分子与分母：

$$\Sigma(T-\bar{T})(Y-\bar{Y}) = 140625.22 \quad \Sigma(T-\bar{T})^2 = 1462.50$$

于是：

$$b = 140625.22 \div 1462.50 = 96.1540$$

$$a = 4729.52 - 96.1540 \times 13.5 = 3431.4384$$

模型一的回归方程为：

$$Y = 3431.44 + 96.15T$$

实际操作时可在 Excel 中用 =SLOPE(产量区域, T 区域) 求 b，用 =INTERCEPT(产量区域, T 区域) 求 a。

第三步，计算 R^2 ，判断拟合效果。 R^2 衡量的是：使用 T 来预测，比不使用任何自变量、直接用平均值预测，误差减少了多少。若不使用 T，最简单的预测是每年都取平均值 4729.5 万吨，由此产生的误差平方和记为 SST；改用回归线之后，剩余的误差平方和记为 SSE。 R^2 即为误差减少的比例：

$$R^2 = 1 - SSE \div SST = 1 - 752355 \div 14274031 = 0.9473$$

即引入时间变量后，原有误差减少了 94.73%，只剩下 5.27%。 R^2 越接近 1，表示直线与实际数据贴合得越好。图 2 用图形表示了这一含义：左图灰线为只用平均值时的误差，中图蓝线为使用回归线后剩余的误差，右图对比两者，绿色箭头所示的部分即为 R^2 。

模型一中的时间本身并不能生产粮食，这个系数概括的是良种、化肥、农机、水利建设等随年份稳步推进的因素的共同作用。模型一的不足也比较明显：它假定产量始终按同一速度增长，而山东近年的增速实际上已明显放缓。

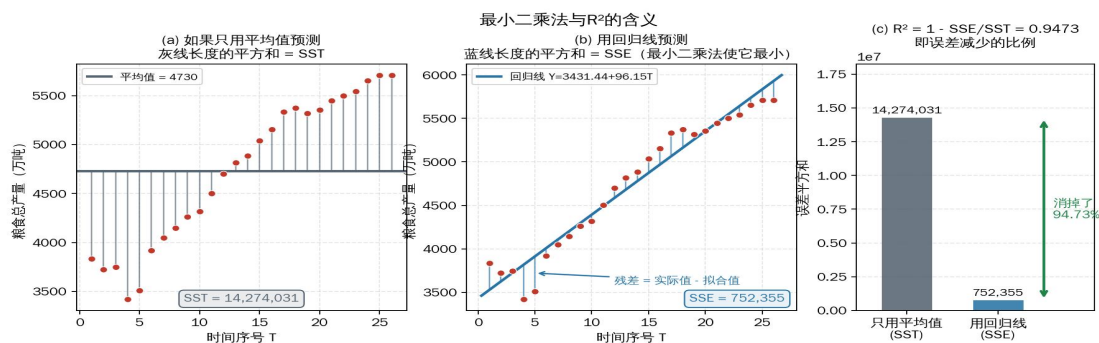


图2 最小二乘法与 R^2 的含义

补充说明：一元回归中， R^2 恰好等于相关系数 r 的平方。Excel 中用 `=RSQ(产量区域, T 区域)` 可直接求 R^2 。

3.2 模型二：播种面积回归模型

模型二用粮食播种面积预测产量。它与模型一使用的是完全相同的算法，同样是一元线性回归，同样用最小二乘法求系数，区别只在于把自变量由时间序号 T 换成播种面积 X_1 。模型二形式为：

$$Y = a + bX_1$$

按照与 3.1 节相同的步骤计算，得到回归方程：

模型二的回归方程为：

$$Y = -2914.73 + 9.9079X_1 \quad (R^2 = 0.9156)$$

为便于对照，表 4 把两个模型的计算过程逐步并列出来。

表 4 模型一与模型二的计算过程对比

计算步骤	模型一（自变量 T ）	模型二（自变量 X_1 ）
自变量均值 $\bar{x}$	13.50	771.53
$\Sigma(x-\bar{x})(Y-\bar{Y})$	140,625.22	1,319,116.76
$\Sigma(x-\bar{x})^2$	1,462.50	133,137.50
$b = \text{分子} \div \text{分母}$	96.1540	9.9079
$a = \bar{Y} - b \cdot \bar{x}$	3,431.4384	-2,914.7341
总误差平方和 SST	14,274,031	14,274,031
剩余误差平方和 SSE	752,355	1,204,316
$R^2 = 1 - SSE/SST$	0.9473	0.9156

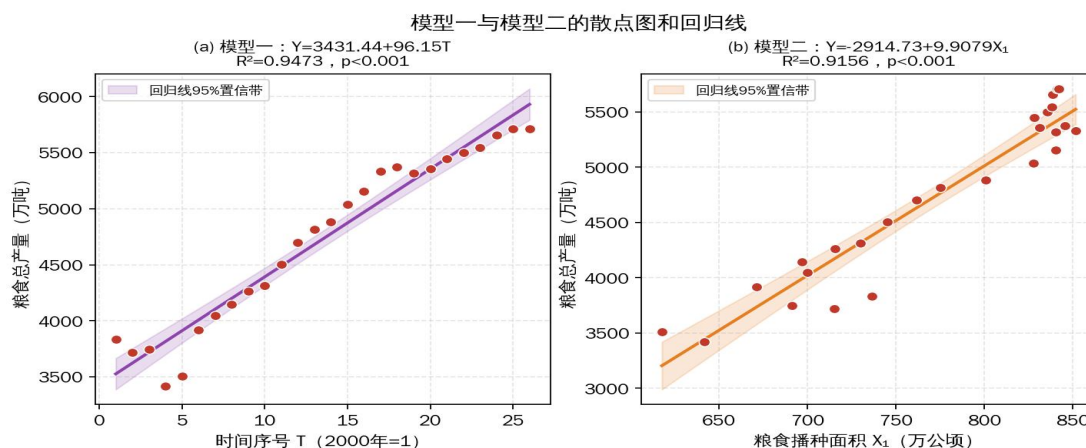


图3 模型一与模型二的散点图和回归线

3.3 模型三：三因素多元回归模型

模型三把土地、水利、灾害三个因素同时纳入，属于多元线性回归：

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3$$

多元回归的求解原则与一元回归相同，仍是使残差平方和最小，只是自变量由一个增加到三个，几何上不再是一条直线而是一个平面，用 SLOPE 与 INTERCEPT 这类函数已无法逐个求解。

本研究改用大语言模型完成这一步计算。做法是把 26 年的四列数据（Y、 X_1 、 X_2 、 X_3 ）连同计算要求一并输入：以 Y 为因变量， X_1 、 X_2 、 X_3 为自变量作多元线性回归，输出常数项、三个回归系数、 R^2 与残差标准误。模型随即给出完整结果：

模型三的回归方程为：

$$Y = -3830.48 + 6.4228X_1 + 7.6811X_2 - 1.5149X_3$$

三个系数的方向完全符合预期：播种面积为正、灌溉面积为正、受灾面积为负，而且都通过了显著性检验。 R^2 提高到 0.9760，残差标准误从模型二的 224.0 万吨降到 124.7 万吨，降幅超过四成。

3.4 模型的有效性检验

R^2 高只能说明模型对历史数据解释得好，不能说明它对未来预测得准。为检验预测能力，研究设计了滚动检验，基本原则是：预测某一年时，不使用该年及其以后的任何数据。

检验分三轮进行。第一轮，只用 2000—2022 年的数据重新估计三个模型的系数，再把 2023 年的自变量代入所得方程，算出 2023 年的预测产量，与实际值比较。第二轮，改用 2000—2023 年的数据重估系数，预测 2024 年；第三轮，用 2000—2024 年的数据重估系数，预测 2025 年。每一轮的系数都要重新计算（见表 5），不能沿用全部 26 年数据的系数。误差率 =（预测值-实际值）÷ 实际值 ×100%。表 6 给出了三轮滚动检验的预测误差率。

表 5 三个模型在不同训练窗口下的系数

模型	系数	全样本 (n=26)	2000—2022 (n=23)	2000—2023 (n=24)	2000—2024 (n=25)
模型一	常数项 a	3431.4384	3382.2509	3396.4503	3411.5626
模型一	斜率 b	96.1540	101.8101	100.1061	98.3624
模型二	常数项 a	-2914.7341	-2526.9361	-2667.3048	-2802.0630
模型二	斜率 b	9.9079	9.3520	9.5538	9.7468
模型三	常数项 a	-3830.4780	-3336.9225	-3490.2191	-3695.0641
模型三	b ₁ (X ₁)	6.4228	6.5841	6.5299	6.4674
模型三	b ₂ (X ₂)	7.6811	6.3881	6.8031	7.3297
模型三	b ₃ (X ₃)	-1.5149	-1.4359	-1.4749	-1.5006

（注：3.1 至 3.4 节报告的方程均由全部 26 年数据估计；滚动检验中重新估计的系数仅用于计算误差率，不用于后续的预测与解释。）

表 6 三轮滚动检验的预测误差率

检验年份	实际产量(万吨)	模型一误差率(%)	模型二误差率(%)	模型三误差率(%)
2023	5655.3	3.01	5.97	2.18
2024	5710.2	3.31	5.96	2.43
2025	5710.6	4.52	5.31	1.64
平均	—	3.615	5.748	2.085

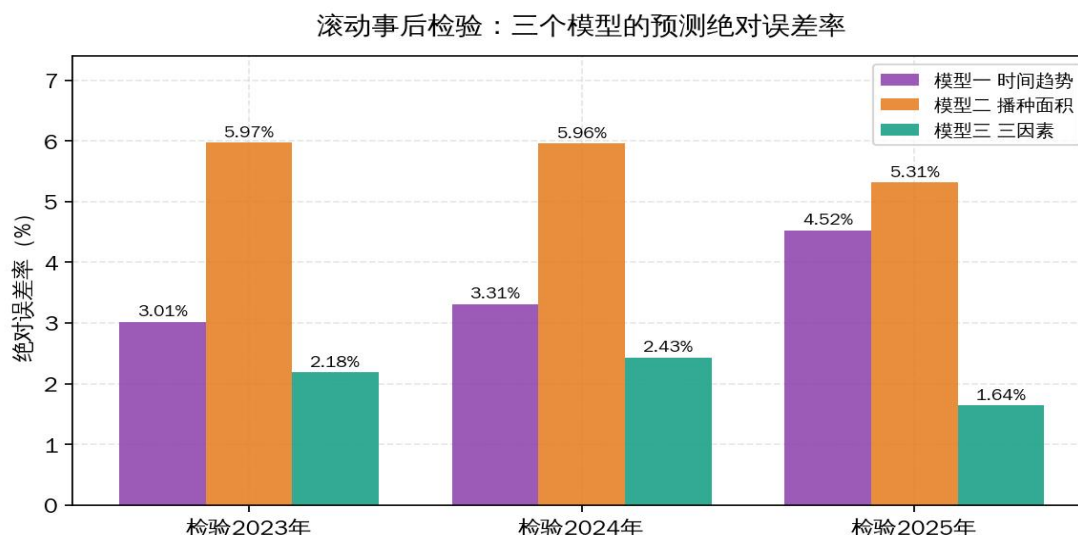


图 4 三个模型的预测误差率对比

检验结果为：模型三的平均误差率最低，为 2.085%；模型一次之，为 3.615%；模型二最高，为 5.748%。这说明在土地因素之外再加入水利与灾害两个因素，确实提高了预测能力。

另外，九次预测的结果全部低于实际值。模型二的低估最为明显，原因是近年播种面积基本不再增加，而单位面积产量仍在提高，只含面积一个变量的模型无法反映这一变化。这从另一个角度说明，播种面积已不是当前增产的主要来源。

3.5 2026 年产量预测

利用模型预测 2026 年产量，第一步先确定 2026 年各自变量的取值。本研究以近七年数据作线性趋势外推获得 2026 年三个变量的取值，所用方法与模型一相同，仍然是一元最小二乘回归法。最后得到播种面积 845.45 万公顷、灌溉面积 526.77 万公顷。受灾面积的外推值为-32.58 万公顷，但受灾面积不可能为负，说明其下降已近极限，因此改用近三年平均值 7.45 万公顷。第二步，把上述取值分别代入三个模型，算出 2026 年的预测产量，并给出 95%预测区间。

$$95\% \text{ 预测区间} = \text{预测值} \pm 2S_e$$

式中 S_e 为残差标准误，即 $\sqrt{(SSE \div \text{自由度})}$ ，表示模型的典型误差。即若误差近似服从正态分布，约 95% 的观测值落在上下两倍标准差之内。

表 6 2026 年山东省粮食产量预测

模型	2026 年预测值(万吨)	95%预测区间	较 2025 年变化(万吨)
模型一 时间趋势	6027.6	5673 ~ 6382	317.0
模型二 播种面积	5461.9	5014 ~ 5910	-248.7
模型三 三因素	5634.6	5385 ~ 5884	-76.0
加权组合预测	5719.2	—	8.6

模型一预测 6027.6 万吨，比 2025 年高出 317 万吨（+5.5%），明显偏高，因为它假定产量还会按前些年的速度增长，而实际上 2017 年以来年均增速只有 0.76%。模型三预测为 5634.6 万吨，比 2025 年略低 76.0 万吨，与近年“高位稳定”的实际情况最吻合。

考虑到三个模型各有长短，研究过程中还按误差率（三年算术平均值）的倒数给三个模型分配权重，得到加权组合预测 5719.2 万吨。综合来看，2026 年山东粮食产量大概率与 2025 年基本持平。

四、影响因素分析

上一章建立并检验了预测模型，本章进一步分析各因素对产量的影响。分析按三步进行：先考察产量的变化过程与增产来源，再用相关系数判断各因素与产量的关系，最后检验这些关系是否可靠。

4.1 产量变化与增产来源

2000 至 2025 年，山东粮食总产量由 3835 万吨增至 5710.6 万吨，增长 48.9%，但过程并不平稳，可分为四个阶段：2000—2003 年调减波动，产量连续下降；2004—2009 年恢复回升，年均增长 4.23%；2010—2016 年快速增长；2017—2025 年高位稳定，年均增速降到 0.76%。

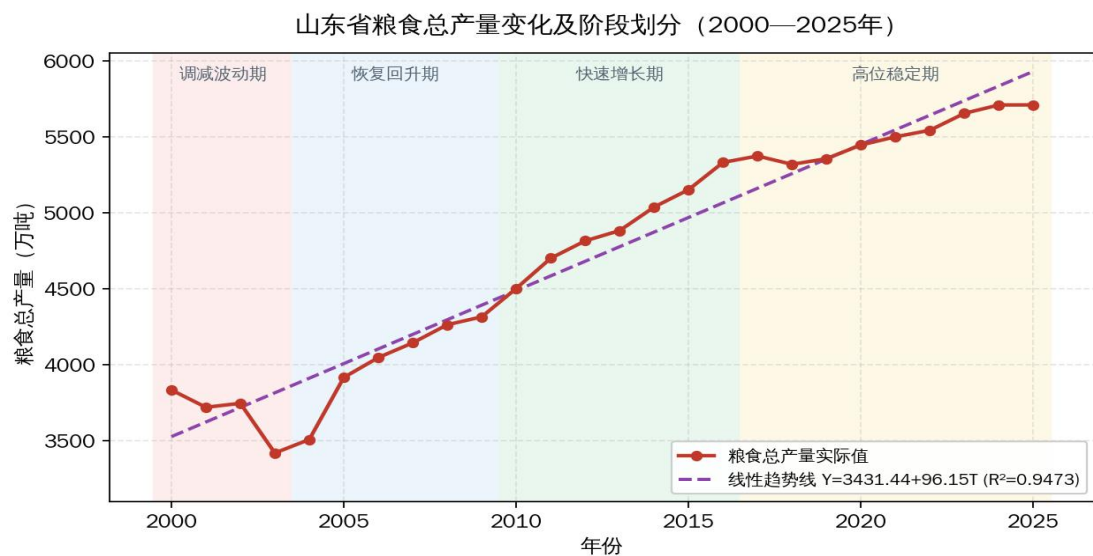


图5 山东省粮食总产量变化及阶段划分

增产究竟来自何处？由于“总产量=播种面积×单位面积产量”，可以把增产量拆成面积、单产和二者交互三部分。设A为播种面积，U为单位面积产量：

$$\text{面积效应} = (A_{2025} - A_{2000}) \times U_{2000}$$

$$\text{单产效应} = A_{2000} \times (U_{2025} - U_{2000})$$

$$\text{交互效应} = (A_{2025} - A_{2000}) \times (U_{2025} - U_{2000})$$

代入数据（ $A_{2000} = 736.32$ 万公顷， $A_{2025} = 842.26$ 万公顷， $U_{2000} = 5208$ 公斤/公顷， $U_{2025} = 6780$ 公斤/公顷），算得三项效应如表7。三者之和应等于2025年与2000年的产量之差，可用作校验。

表7 2000—2025年增产来源分解

增产来源	贡献量(万吨)	占总增产比重(%)
播种面积扩大效应	551.8	29.4
单位面积产量提高效应	1157.3	61.7
面积—单产交互效应	166.5	8.9
合计(2025年较2000年)	1875.6	100.0

表7结果表明，单位面积产量提高贡献了61.7%，播种面积扩大只贡献29.4%。也就是说，山东这26年增产主要靠“种好地”而不是“多种地”。单产由5208公斤/公顷提高到6780公斤/公顷，提高了30.2%。

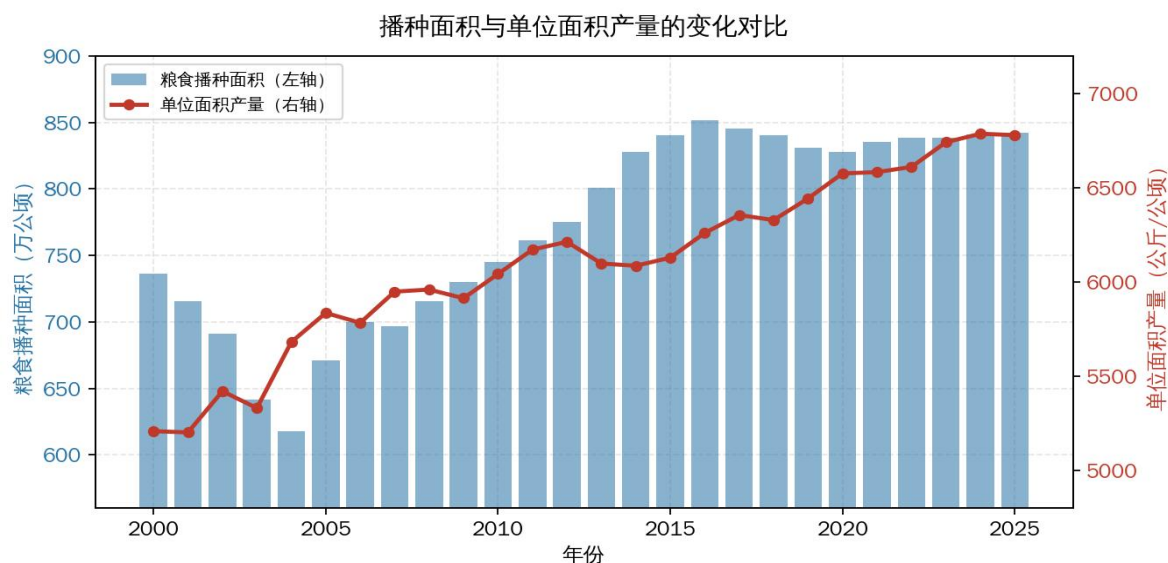


图6 播种面积与单位面积产量的变化对比

图6表明播种面积（蓝色）在2016年以后基本走平，而红线（单产）一直在往上走。这说明在耕地越来越紧张的情况下，未来增产的空间主要在提高单产。

4.2 三个因素与产量的相关性

利用豆包和 excel 分别计算三个因素与产量的相关性分析，见表8。三个因素的相关方向全部符合预期（表8）：播种面积（ $r=0.957$ ）和灌溉面积（ $r=0.880$ ）为正，受灾面积（ $r=-0.818$ ）为负，且都非常显著。

表8 各变量与粮食总产量的相关分析

自变量	相关系数 r	p 值	方向	相关强度
播种面积	0.9569	<0.001	正相关	极强
有效灌溉面积	0.8797	<0.001	正相关	强
受灾面积	-0.8175	<0.001	负相关	强
时间序列 T（辅助变量）	0.9733	<0.001	正相关	极强

（注：T 为辅助变量，4.3 节专门讨论）

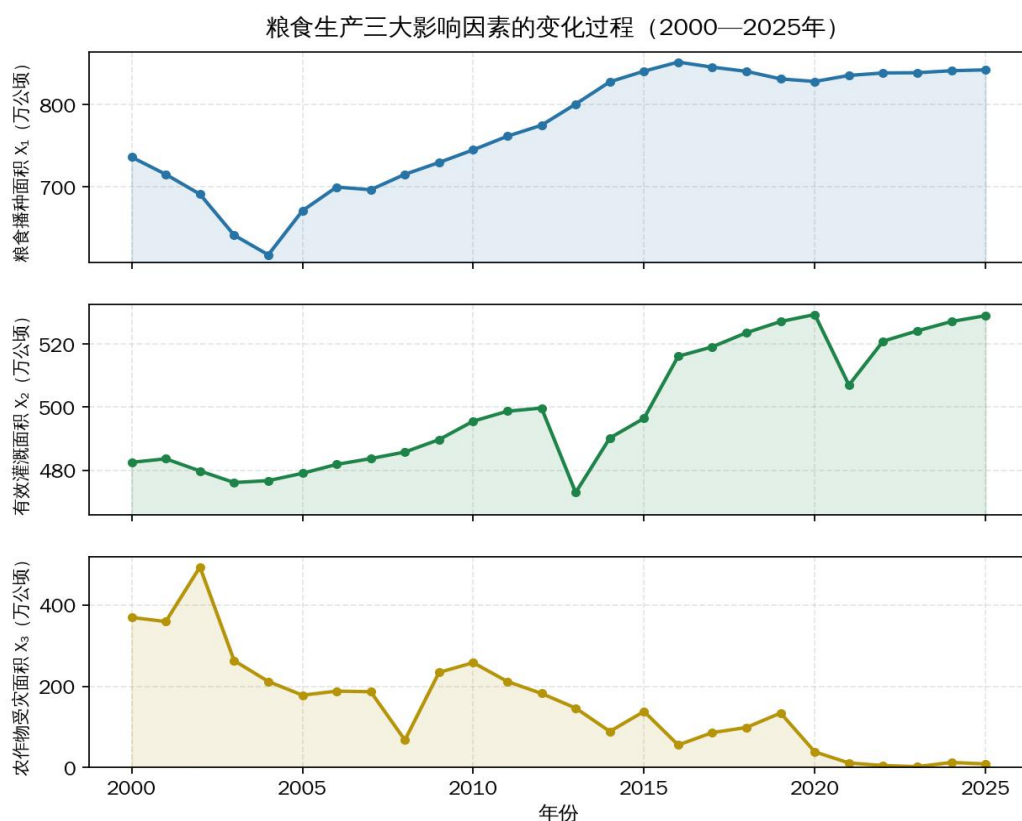


图 7 2000-2025 年三个影响因素的变化过程

播种面积代表土地基础。在单产大体稳定时，种得多产得多。2000—2003 年因为面积调减了近 13%，产量就连续三年下降，是最直接的例子。

灌溉面积代表水利保障。山东春季少雨，而冬小麦返青、拔节、灌浆都需要水，灌溉能不能跟上往往决定收成。扩大灌溉面积，实质是把“靠天吃饭”的不确定性变成“人力可控”的确定性。

受灾面积代表灾害冲击。旱涝风雹会直接毁坏庄稼。2003 年受灾 263.2 万公顷，当年减产 8.73%，是 26 年里最严重的一次。不过 2018 年受灾面积只有 98.38 万公顷却也减产了，说明“受灾面积”这个指标只记范围、不分轻重，也不看灾害发生在哪个时段。

4.3 剔除时间趋势后的稳健性检验

上节的三个相关系数都很高，但在采信之前还需回答一个问题：它们究竟反映真实的关联，还是仅因各变量都随时间变化？本节先说明疑问的由来，再用两种独立的方法检验。

受灾面积与产量的相关系数为-0.818，据此似乎可得出“灾害严重制约粮食生产”的结论。但表 8 中时间序号 T 与产量的相关系数达 0.973，高于三个因素

中的任何一个，而时间本身并不能生产粮食。结合图 5 和图 7 可以看出原因：这 26 年中产量持续上升，受灾面积则由 369.7 万公顷降至 8.46 万公顷，降幅超过 97%。一个持续上升，一个持续下降，两者的相关系数必然接近-1。这就是统计上说的“虚假相关”。为了弄清楚，本研究做了两项检验。

4.3.2 检验一：去除时间趋势后的相关分析

这一检验先把时间的影响扣除，再看两个变量是否仍然相关，分三步。第一步，把产量对时间序号 T 作一元回归，用实际值减去回归值得到残差，它代表产量偏离长期趋势的部分；第二步，对每个自变量作同样处理，得到各自的趋势偏离；第三步，计算两组残差之间的相关系数，即去除趋势后的偏相关系数（见表 9）。

表 9 去掉时间趋势前后的相关系数对比

自变量	原来的相关系数 r	去掉趋势后的 r	去掉趋势后是否显著
播种面积	0.9569	0.9393	是($\alpha=0.05$)
有效灌溉面积	0.8797	0.1660	否
受灾面积	-0.8175	0.2094	否

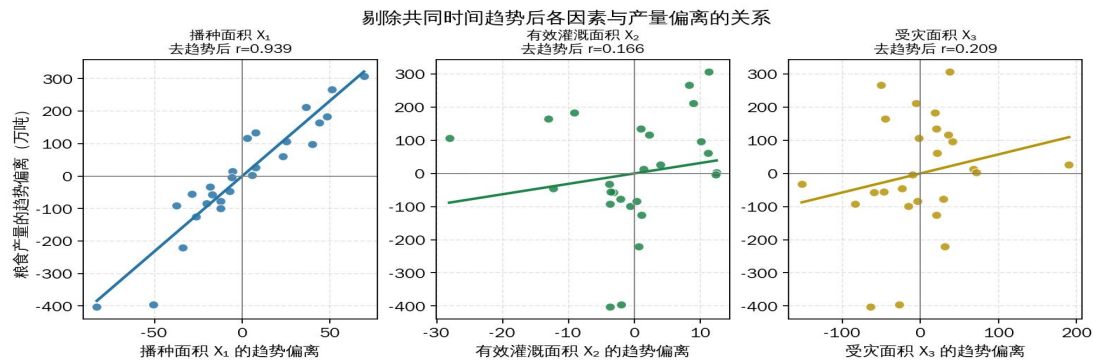


图 8 去掉时间趋势后各因素与产量偏离的关系

图 8 表明，播种面积几乎没受影响（r 从 0.957 降到 0.939，仍然非常显著），说明它和产量之间存在真实的年际关联。灌溉面积则由 0.880 降至 0.166，受灾面积由-0.818 变为 0.209，两者都不再显著，受灾面积的符号甚至发生反转。

4.3.2 检验二：在模型中加入时间趋势项

第二种方法更直接：在模型三的基础上加入时间序号 T 构成模型四，观察三个因素的系数是否依然显著。若某因素的作用真实，即使控制时间趋势，其系数也应保持稳定（见表 10）。

表 10 受灾面积系数在三种算法下的对比

模型设定	受灾面积 X_3 的系数	p 值	是否显著
模型三（不含时间趋势项）	-1.5149	0.000	是
模型四（加入时间趋势项 T）	-0.2118	0.345	否
去趋势双残差回归（ X_3 、T）	0.5758	0.304	否

加入时间项后，受灾面积的系数从-1.5149 缩小到-0.2118，p 值升到 0.345，不再显著；灌溉面积也是同样情况。

4.6 灾害情景分析

受灾面积的系数虽不稳健，但用模型三作情景推演仍有参考价值。做法分两步：固定播种面积 845.45 万公顷（按照播种面积与时间关系，趋势外推获得）、灌溉面积 526.77 万公顷（趋势外推不显著，选用近三年算是平方值）；再设定五种受灾水平代入模型三，算出 2026 年产量（见表 11）。

表 11 受灾面积系数在三种算法下的对比

情景	假定受灾面积(万公顷)	预测产量(万吨)	较 2025 年变化(万吨)
情景 A 无明显灾害	5	5638.3	-72.3
情景 B 轻度灾害	20	5615.6	-95.0
情景 C 中度灾害	60	5555.0	-155.6
情景 D 重度灾害	150	5418.6	-292.0
情景 E 特重灾害	300	5191.4	-519.2

基于模型三的2026年自然灾害情景预测

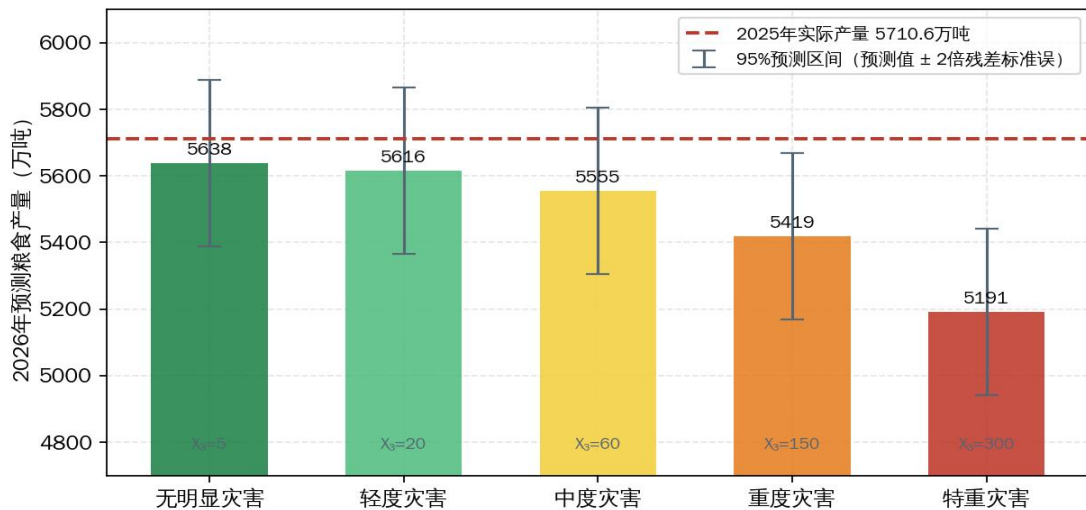


图 9 不同灾害情景下的 2026 年产量预测

从无明显灾害到特重灾害，预测产量由 5638.3 万吨降至 5191.4 万吨，相差 446.9 万吨，约占 2025 年产量的 7.8%。可见土地与水利条件稳定时，自然灾害

是产量年际波动的重要原因。

五、主要结论与建议

5.1 主要结论

山东粮食生产已进入高位稳定期。2000—2025 年产量由 3835 万吨增至 5710.6 万吨，增长 48.9%，经历调减波动、恢复回升、快速增长、高位稳定四个阶段。最近一阶段年均增速仅 0.76%，产量波动明显变小，呈现增速放缓而趋于稳定的特点。

增产主要靠单产提高，贡献率 61.7%。单位面积产量提高贡献 1157.3 万吨，播种面积扩大贡献 551.8 万吨。单产从 5208 公斤/公顷提高到 6780 公斤/公顷。在耕地紧张的背景下，增产路径已经从“扩面积”转向“提单产”。

三因素模型预测最准，平均误差率 2.085%。三轮滚动检验中，三因素多元回归模型的平均误差率为 2.085%，优于时间趋势模型的 3.615%和播种面积模型的 5.748%。据此预测 2026 年产量约 5635 万吨，加权组合预测 5719 万吨，均指向与 2025 年基本持平。

受灾面积的负向作用并不稳健。这是本研究最重要的发现。受灾面积与产量的相关系数高达-0.818，模型三中的系数-1.5149 也非常显著，但加入时间因素后系数缩至-0.2118 并失去显著性（ $p=0.345$ ），去趋势后更变为 0.209。灌溉面积同样如此，而播种面积在所有算法下都保持稳健。

相关不等于因果。这是本研究在方法上最大的收获。-0.818 这个数很容易让人直接写下“灾害严重制约粮食生产”，但两项稳健性检验表明，该相关在控制时间趋势后即不再显著，说明它主要源于共同趋势而非因果联系。凡由时间序列得出的相关，都应先作去趋势检验。

5.2 研究的不足

本研究的局限有四点：一是样本仅 26 年，个别异常年份的影响较大；二是时间与各因素高度重叠，无法彻底分离各自的作用；三是“受灾面积”不区分轻重与发生时段，2018 年受灾不多却减产的现象无法解释；四是化肥、农机、良种与价格政策等因素未能纳入。

今后若继续研究，最有效的改进是采用面板数据，纳入山东 16 个地市的资料。这样既可大幅增加样本量，又能利用地区差异分离时间趋势的影响，正可解决本研究遇到的最大困难。