

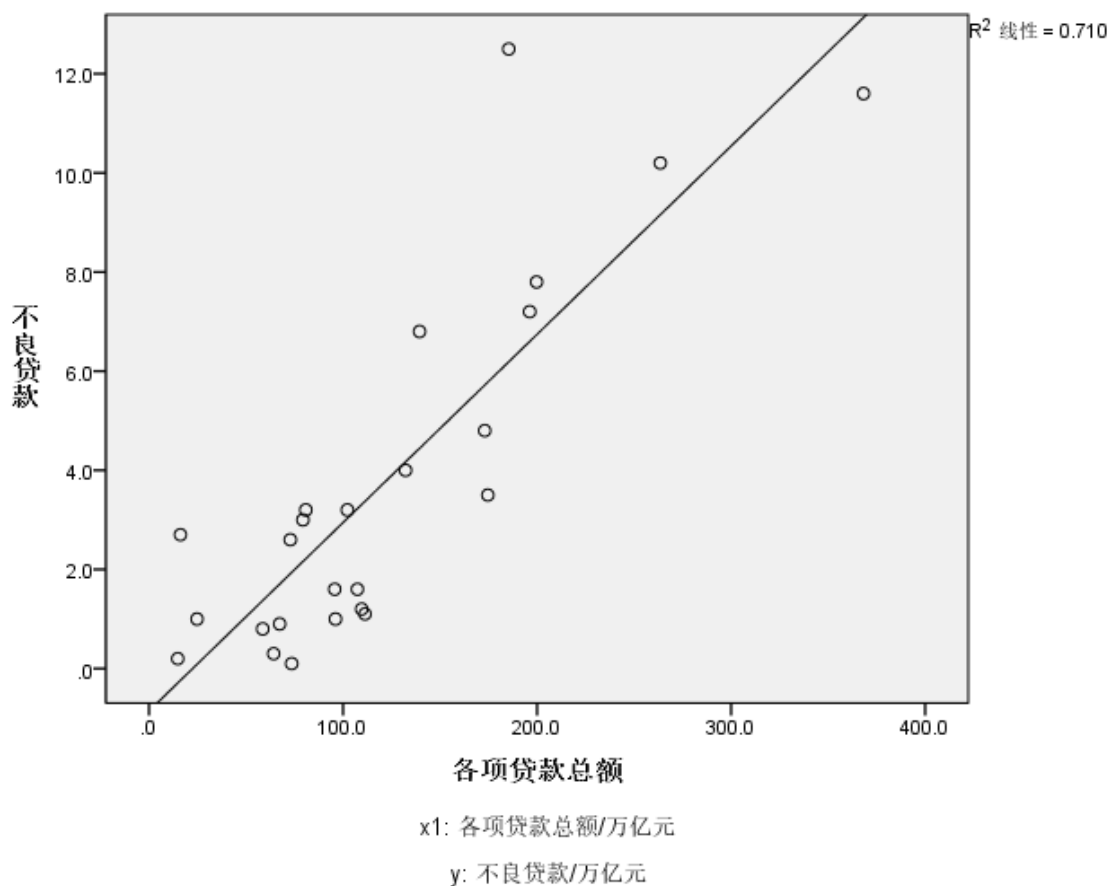
回归分析

银行贷款样例

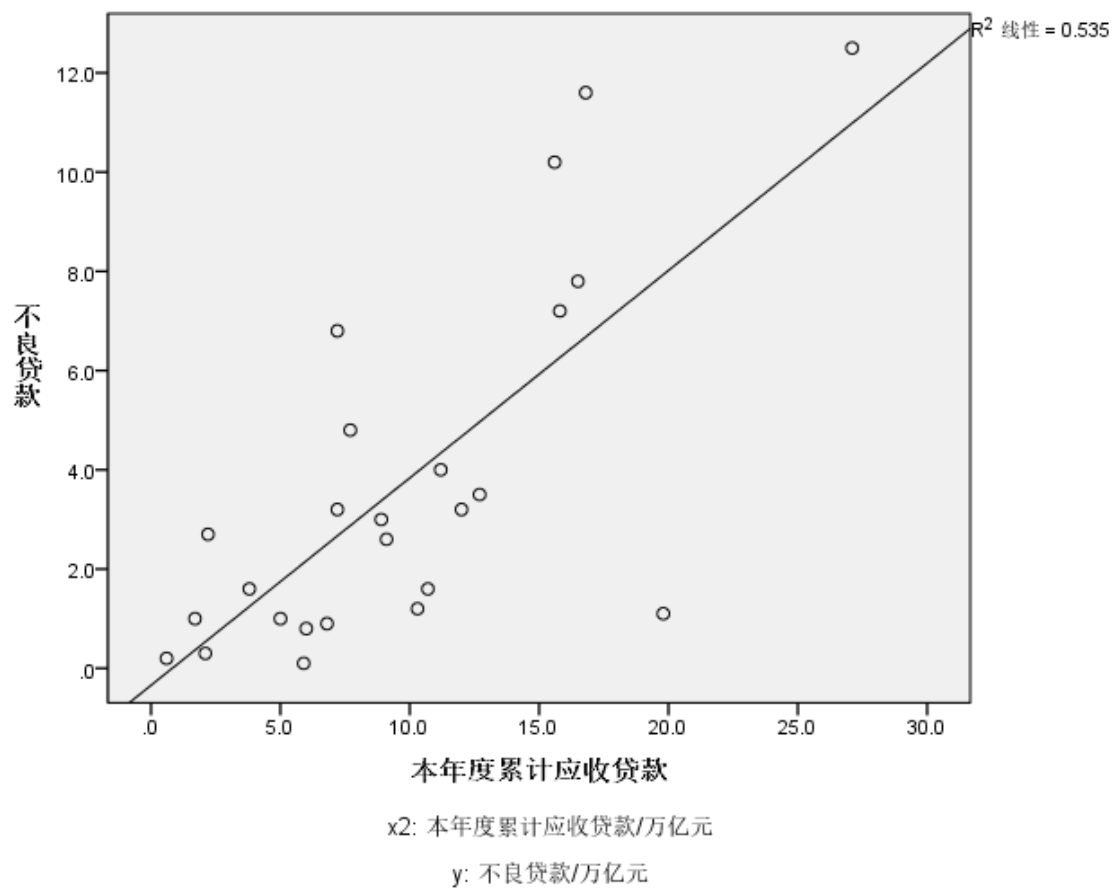
一家大型商业银行为了弄清不良贷款形成的原因，以便找出控制不良贷款的办法，现利用相关业务数据进行分析。

1. 利用SPSS分别作出分行不良贷款与其他四个变量(依此记为x1,x2,x3,x4)的散点图

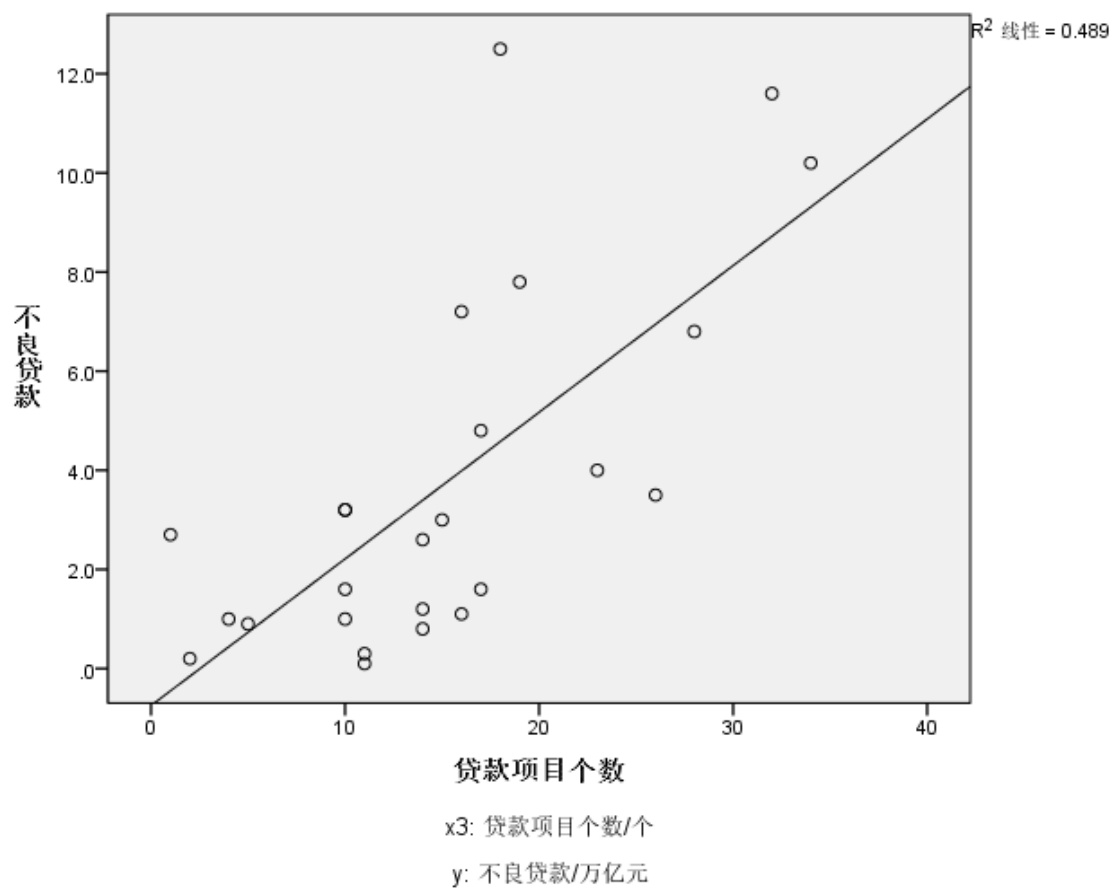
- 不良贷款和各项贷款总额近似于线性关系



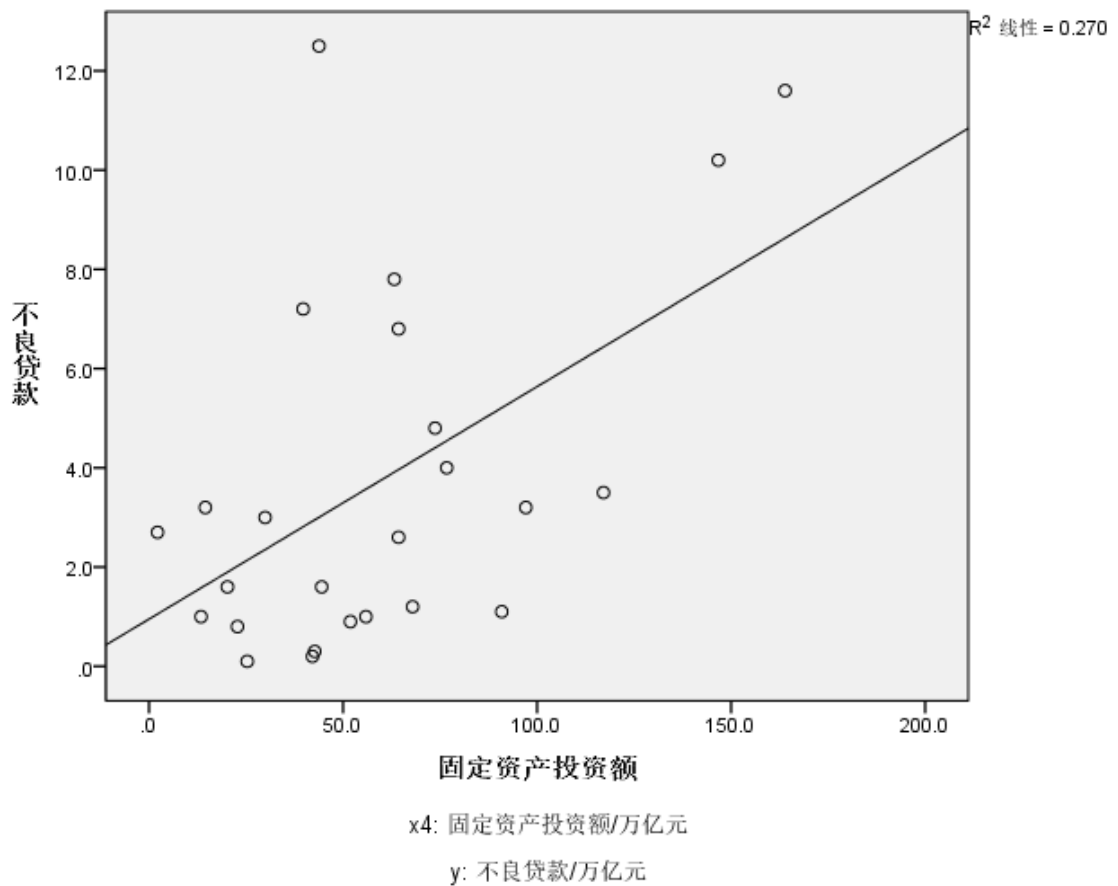
- 本年度累计营收贷款和不良贷款近似于线性关系



- 贷款项目个数和不良贷款近似于线性关系



- 固定资产投资额和不良贷款率不近似线性关系



2. 利用SPSS计算分行不良贷款与其他四个变量的皮尔逊相关系数;

相关性		不良贷款	固定资产投资额	分行编号	各项贷款总额	贷款项目个数	本年度累计应收贷款
不良贷款	Pearson 相关性	1	.520**	.086	.843**	.700**	.731**
	显著性 (双侧)		.008	.682	.000	.000	.000
	N	25	25	25	25	25	25
固定资产投资额	Pearson 相关性	.520**	1	.196	.780**	.747**	.472*
	显著性 (双侧)	.008		.349	.000	.000	.017
	N	25	25	25	25	25	25
分行编号	Pearson 相关性	.086	.196	1	.151	.152	-.059
	显著性 (双侧)	.682	.349		.471	.468	.781
	N	25	25	25	25	25	25
各项贷款总额	Pearson 相关性	.843**	.780**	.151	1	.848**	.679**
	显著性 (双侧)	.000	.000	.471		.000	.000
	N	25	25	25	25	25	25
贷款项目个数	Pearson 相关性	.700**	.747**	.152	.848**	1	.586**
	显著性 (双侧)	.000	.000	.468	.000		.002
	N	25	25	25	25	25	25
本年度累计应收贷款	Pearson 相关性	.731**	.472*	-.059	.679**	.586**	1
	显著性 (双侧)	.000	.017	.781	.000	.002	
	N	25	25	25	25	25	25

** 在 .01 水平 (双侧) 上显著相关。

* 在 0.05 水平 (双侧) 上显著相关。

通过观察表格，我们可以得知各变量与不良贷款的皮尔逊相关系数，各项贷款总额和不良贷款的皮尔逊系数为0.843，在四个变量中和不良贷款的皮尔逊系数最高，线性关系最强，再看本年度累计应收贷款和不良贷款的皮尔森相关系数为0.731，贷款项目个数和不良贷款的皮尔森相关系数为0.7，可以得出这两个变量和不良贷款的线性关系也较强，最后看固定资产投资额的皮尔森相关系数只有0.52，所以线性关系相对较弱

3. 分析不良资产的形成与哪些变量有关，并建立回归方程。

进入选择法

输入/移去的变量^b

模型	输入的变量	移去的变量	方法
1	固定资产投资额, 本年度累计应收贷款, 贷款项目个数, 各项贷款总额	.	输入

a. 已输入所有请求的变量。

b. 因变量: 不良贷款

模型汇总^b

模型	R	R 方	调整 R 方	标准估计的误差
1	.892 ^a	.795	.754	1.7952

a. 预测变量: (常量), 固定资产投资额, 本年度累计应收贷款, 贷款项目个数, 各项贷款总额。

b. 因变量: 不良贷款

Anova^b

模型		平方和	df	均方	F	Sig.
1	回归	250.282	4	62.570	19.416	.000 ^a
	残差	64.452	20	3.223		
	总计	314.734	24			

a. 预测变量: (常量), 固定资产投资额, 本年度累计应收贷款, 贷款项目个数, 各项贷款总额。

b. 因变量: 不良贷款

系数^a

模型		非标准化系数		标准系数	t	Sig.
		B	标准误差	试用版		
1	(常量)	-1.050	.790		-1.329	.199
	各项贷款总额	.040	.011	.887	3.798	.001
	本年度累计应收贷款	.149	.080	.261	1.874	.076
	贷款项目个数	.014	.084	.033	.166	.870
	固定资产投资额	-.029	.015	-.320	-1.896	.073

a. 因变量: 不良贷款

共线性诊断^a

模型	维数	特征值	条件索引	方差比例				
				(常量)	各项贷款总额	本年度累计应收贷款	贷款项目个数	固定资产投资额
1	1	4.538	1.000	.01	.00	.01	.00	.00
	2	.203	4.733	.68	.03	.02	.01	.09
	3	.157	5.378	.16	.00	.66	.01	.13
	4	.066	8.287	.00	.09	.20	.36	.72
	5	.036	11.215	.15	.87	.12	.63	.05

a. 因变量: 不良贷款

由上列表可得

- 模型汇总表中，多元决定系数为0.795，修正的多元决定系数为0.754，说明选择的变量对于因变量的解释度还算较高

- Anova表给出了方程检验的方差分析，从中可见F检验的P值近似为0，故可以判断模型十分显著
- 系数表中，变量各项贷款总额和固定资产投资额通过了检验，贷款项目个数的p值为0.870，对因变量而言不够显著，可以考虑剔除

逐步选择法

输入／移去的变量^a

模型	输入的变量	移去的变量	方法
1	各项贷款总额	.	步进（准则：F-to-enter的概率 ≤ .050，F-to-remove的概率 ≥ .100）。
2	固定资产投资额	.	步进（准则：F-to-enter的概率 ≤ .050，F-to-remove的概率 ≥ .100）。

a. 因变量: 不良贷款

模型汇总^a

模型	R	R 方	调整 R 方	标准估计的误差
1	.843 ^a	.710	.698	1.9911
2	.871 ^b	.758	.736	1.8589

a. 预测变量: (常量), 各项贷款总额。

b. 预测变量: (常量), 各项贷款总额, 固定资产投资额。

c. 因变量: 不良贷款

Anova^a

模型		平方和	df	均方	F	Sig.
1	回归	223.551	1	223.551	56.388	.000 ^a
	残差	91.183	23	3.964		
	总计	314.734	24			
2	回归	238.712	2	119.356	34.540	.000 ^b
	残差	76.022	22	3.456		
	总计	314.734	24			

a. 预测变量: (常量), 各项贷款总额。

b. 预测变量: (常量), 各项贷款总额, 固定资产投资额。

c. 因变量: 不良贷款

系数^a

模型		非标准化系数		标准系数	t	Sig.	共线性统计量	
		B	标准误差	试用版			容差	VIF
1	(常量)	-.852	.727		-1.172	.253		
	各项贷款总额	.038	.005	.843	7.509	.000	1.000	1.000
2	(常量)	-.470	.703		-.669	.511		
	各项贷款总额	.050	.008	1.116	6.669	.000	.392	2.551
	固定资产投资额	-.032	.015	-.351	-2.095	.048	.392	2.551

a. 因变量: 不良贷款

已排除的变量^a

模型		Beta In	t	Sig.	偏相关	共线性统计量		
						容差	VIF	最小容差
1	本年度累计应收贷款	.295 ^a	2.062	.051	.403	.539	1.854	.539
	固定资产投资额	-.351 ^a	-2.095	.048	-.408	.392	2.551	.392
	贷款项目个数	-.055 ^a	-.254	.802	-.054	.280	3.569	.280
2	本年度累计应收贷款	.262 ^b	1.933	.067	.389	.531	1.883	.268
	贷款项目个数	.055 ^b	.263	.795	.057	.262	3.821	.232

a. 模型中的预测变量: (常量), 各项贷款总额。

b. 模型中的预测变量: (常量), 各项贷款总额, 固定资产投资额。

c. 因变量: 不良贷款

共线性诊断^a

模型	维数	特征值	条件索引	方差比例		
				(常量)	各项贷款总额	固定资产投资额
1	1	1.837	1.000	.08	.08	
2	2	.400	2.224	.92	.92	

	2	.163	3.354	.92	.92	
2	1	2.735	1.000	.03	.01	.02
	2	.198	3.718	.97	.09	.10
	3	.067	6.392	.00	.90	.88

a. 因变量: 不良贷款

残差统计量^a

	极小值	极大值	均值	标准 偏差	N
预测值	-1.056	12.873	3.716	3.1538	25
标准 预测值	-1.513	2.903	.000	1.000	25
预测值的标准误差	.384	1.235	.613	.201	25
调整的预测值	-1.278	13.878	3.701	3.2596	25
残差	-2.6943	5.0277	.0000	1.7798	25
标准 残差	-1.449	2.705	.000	.957	25
Student 化 残差	-1.542	2.979	.005	1.035	25
已删除的残差	-3.0484	6.1003	.0150	2.0912	25
Student 化 已删除的残差	-1.595	3.769	.038	1.142	25
Mahal. 距离	.064	9.630	1.920	2.099	25
Cook 的距离	.000	.631	.062	.131	25
居中杠杆值	.003	.401	.080	.087	25

a. 因变量: 不良贷款

- 剔除本年度累计应收贷款和贷款项目个数后， $R^2 = 0.758$ ，调整后的 $R^2 = 0.736$ ，相对原理变化较小，说明本年度累计应收贷款和贷款项目个数对因变量的变化小
- 系数表中最后两列入选变量各项贷款总额和固定资产投资的容差均为 $0.392 > 0.1$ ， $VIF = 2.551 < 10$ ，说明各项贷款总额和固定资产投资不存在多重共线性
- 已排除的变量表中最后两列本年度累计应收贷款和贷款项目个数的容差为， 0.531 和 0.262 ， VIF 为 1.883 和 3.821 ，虽然没有达到多重共线的标准，但剔除这两个变量后多元决定系数变化很小，所以剔除这两个变量是合理的
- 利用系数表格得到回归方程：

$$y = -0.47 + 0.05 * \text{各项贷款总额} - 0.032 * \text{固定资产投资额}$$

粮食产量样例

根据理论和经验分析，影响粮食生产(y)的主要因素大致有：化肥施用量(x1)、播种面积(x2)、成灾面积(x3)、农业机械总动力(x4)、劳动力(x5)

1. 将所有变量选入线性回归模型，得到参数估计并完成参数的显著性检验;

输入/移去的变量^b

模型	输入的变量	移去的变量	方法
1	农业劳动人口, 粮食播种面积, 农业化肥施用量, 成灾面积, 农业机械总动力	.	输入

a. 已输入所有请求的变量。

b. 因变量: 粮食产量

模型汇总

模型	R	R 方	调整 R 方	标准估计的误差
1	.994 ^a	.989	.986	778.969

a. 预测变量: (常量), 农业劳动人口, 粮食播种面积, 农业化肥施用量, 成灾面积, 农业机械总动力。

Anova^b

模型		平方和	df	均方	F	Sig.
1	回归	1.409E9	5	2.817E8	464.299	.000 ^a
	残差	16383397.52	27	606792.501		
	总计	1.425E9	32			

a. 预测变量: (常量), 农业劳动人口, 粮食播种面积, 农业化肥施用量, 成灾面积, 农业机械总动力。

b. 因变量: 粮食产量

系数^a

模型		非标准化系数		标准系数	t	Sig.
		B	标准误差	试用版		
1	(常量)	-40608.989	6135.853		-6.618	.000
	农业化肥施用量	4.570	.284	.966	16.086	.000
	粮食播种面积	.639	.045	.354	14.173	.000
	成灾面积	-.112	.035	-.105	-3.249	.003
	农业机械总动力	.010	.016	.046	.634	.532
	农业劳动人口	.072	.057	.049	1.255	.220

a. 因变量: 粮食产量

- 模型汇总表中, 多元决定系数为0.989, 修正的多元决定系数为0.986, 说明选择的变量对于因变量的解释度极高
- Anova表给出了方程检验的方差分析, 从中可见F检验的P值近似为0, 故可以判断模型十分显著

2. 利用向前选择法, 逐步增加变量找到最优的线性模型。

设显著性水平 α 为0.05

引入变量化肥施用量(x1)

输入/移去的变量^a

模型	输入的变量	移去的变量	方法
1	农业化肥施用量	.	向前 (准则: F-to-enter 的概率 <= .050)

a. 因变量: 粮食产量

模型汇总

模型	R	R 方	调整 R 方	标准估计的误差
1	.919 ^a	.844	.839	2679.938

a. 预测变量: (常量), 农业化肥施用量。

Anova^b

模型		平方和	df	均方	F	Sig.
1	回归	1.202E9	1	1.202E9	167.418	.000 ^a
	残差	2.226E8	31	7182067.287		
	总计	1.425E9	32			

a. 预测变量: (常量), 农业化肥施用量。

b. 因变量: 粮食产量

系数^a

模型		非标准化系数		标准系数	t	Sig.
		B	标准误差	试用版		
1	(常量)	30769.547	1406.872		21.871	.000
	农业化肥施用量	4.347	.336	.919	12.939	.000

a. 因变量: 粮食产量

多元决定系数 R^2 为0.844，变化较小，判断 x_1 是显著型变量，所以保留
引入变量播种面积(x_2)

输入/移去的变量^a

模型	输入的变量	移去的变量	方法
1	农业化肥施用里	.	向前 (准则: F-to-enter 的概率 <= .050)
2	粮食播种面积	.	向前 (准则: F-to-enter 的概率 <= .050)

a. 因变量: 粮食产量

模型汇总

模型	R	R 方	调整 R 方	标准估计的误差
1	.919 ^a	.844	.839	2679.938
2	.991 ^b	.982	.980	935.132

a. 预测变量: (常量), 农业化肥施用里。

b. 预测变量: (常量), 农业化肥施用里, 粮食播种面积。

Anova^c

模型		平方和	df	均方	F	Sig.
1	回归	1.202E9	1	1.202E9	167.418	.000 ^a
	残差	2.226E8	31	7182067.287		
	总计	1.425E9	32			
2	回归	1.399E9	2	6.994E8	799.806	.000 ^b
	残差	26234168.12	30	874472.271		
	总计	1.425E9	32			

a. 预测变量: (常量), 农业化肥施用里。

b. 预测变量: (常量), 农业化肥施用里, 粮食播种面积。

c. 因变量: 粮食产量

系数^a

模型		非标准化系数		标准系数	t	Sig.
		B	标准误差	试用版		
1	(常量)	30769.547	1406.872		21.871	.000
	农业化肥施用里	4.347	.336	.919	12.939	.000
2	(常量)	-47212.814	5226.512		-9.033	.000
	农业化肥施用里	4.822	.121	1.019	39.708	.000
	粮食播种面积	.694	.046	.385	14.987	.000

a. 因变量: 粮食产量

已排除的变量^b

模型	Beta In	t	Sig.	偏相关	共线性统计量
					容差
1 粮食播种面积	.385 ^a	14.987	.000	.939	.932

a. 模型中的预测变量: (常量), 农业化肥施用里。

b. 因变量: 粮食产量

引入x2后, 多元决定系数R²增长为0.982, 所以x2也是一个显著性变量, 所以保留
引入变量成灾面积(x3)

输入/移去的变量^a

模型	输入的变量	移去的变量	方法
1	农业化肥施用量	.	向前 (准则: F-to-enter 的概率 <= .050)
2	粮食播种面积	.	向前 (准则: F-to-enter 的概率 <= .050)
3	成灾面积	.	向前 (准则: F-to-enter 的概率 <= .050)

a. 因变量: 粮食产量

模型汇总

模型	R	R 方	调整 R 方	标准估计的误差
1	.919 ^a	.844	.839	2679.938
2	.991 ^b	.982	.980	935.132
3	.994 ^c	.988	.987	773.306

a. 预测变量: (常量), 农业化肥施用量。

b. 预测变量: (常量), 农业化肥施用量, 粮食播种面积。

c. 预测变量: (常量), 农业化肥施用量, 粮食播种面积, 成灾面积。

Anova^d

模型		平方和	df	均方	F	Sig.
1	回归	1.202E9	1	1.202E9	167.418	.000 ^a
	残差	2.226E8	31	7182067.287		
	总计	1.425E9	32			
2	回归	1.399E9	2	6.994E8	799.806	.000 ^b
	残差	26234168.12	30	874472.271		
	总计	1.425E9	32			
3	回归	1.408E9	3	4.692E8	784.672	.000 ^c
	残差	17342075.84	29	598002.615		
	总计	1.425E9	32			

a. 预测变量: (常量), 农业化肥施用量。

b. 预测变量: (常量), 农业化肥施用量, 粮食播种面积。

c. 预测变量: (常量), 农业化肥施用量, 粮食播种面积, 成灾面积。

d. 因变量: 粮食产量

系数^a

模型		非标准化系数		标准系数	t	Sig.
		B	标准误差	试用版		
1	(常量)	30769.547	1406.872		21.871	.000
	农业化肥施用量	4.347	.336	.919	12.939	.000
2	(常量)	-47212.814	5226.512		-9.033	.000
	农业化肥施用量	4.822	.121	1.019	39.708	.000
	粮食播种面积	.694	.046	.385	14.987	.000
3	(常量)	-36475.097	5141.413		-7.094	.000
	农业化肥施用量	4.655	.109	.984	42.569	.000
	粮食播种面积	.622	.043	.345	14.598	.000
	成灾面积	-.098	.025	-.091	-3.856	.001

a. 因变量: 粮食产量

已排除的变量^c

模型	Beta In	t	Sig.	偏相关	共线性统计量
					容差

1	粮食播种面积	.385 ^a	14.987	.000	.939	.932
	成灾面积	-.243 ^a	-4.017	.000	-.591	.926
2	成灾面积	-.091 ^b	-3.856	.001	-.582	.748

a. 模型中的预测变量: (常量), 农业化肥施用量。

b. 模型中的预测变量: (常量), 农业化肥施用量, 粮食播种面积。

c. 因变量: 粮食产量

引入x3后，多元决定系数R²增长为0.988，所以x3也是一个显著性变量，所以保留
引入变量劳动力(x5)

输入/移去的变量^a

模型	输入的变量	移去的变量	方法
1	农业化肥施用量	.	向前 (准则: F-to-enter 的概率 <= .050)
2	粮食播种面积	.	向前 (准则: F-to-enter 的概率 <= .050)
3	成灾面积	.	向前 (准则: F-to-enter 的概率 <= .050)

a. 因变量: 粮食产量

模型汇总

模型	R	R 方	调整 R 方	标准估计的误差
1	.919 ^a	.844	.839	2679.938
2	.991 ^b	.982	.980	935.132
3	.994 ^c	.988	.987	773.306

a. 预测变量: (常量), 农业化肥施用量。

b. 预测变量: (常量), 农业化肥施用量, 粮食播种面积。

c. 预测变量: (常量), 农业化肥施用量, 粮食播种面积, 成灾面积。

Anova^d

模型		平方和	df	均方	F	Sig.
1	回归	1.202E9	1	1.202E9	167.418	.000 ^a
	残差	2.226E8	31	7182067.287		
	总计	1.425E9	32			
2	回归	1.399E9	2	6.994E8	799.806	.000 ^b
	残差	26234168.12	30	874472.271		
	总计	1.425E9	32			
3	回归	1.408E9	3	4.692E8	784.672	.000 ^c
	残差	17342075.84	29	598002.615		
	总计	1.425E9	32			

a. 预测变量: (常量), 农业化肥施用量。

b. 预测变量: (常量), 农业化肥施用量, 粮食播种面积。

c. 预测变量: (常量), 农业化肥施用量, 粮食播种面积, 成灾面积。

d. 因变量: 粮食产量

系数^a

模型		非标准化系数		标准系数	t	Sig.
		B	标准误差	试用版		
1	(常量)	30769.547	1406.872		21.871	.000
	农业化肥施用量	4.347	.336	.919	12.939	.000
2	(常量)	-47212.814	5226.512		-9.033	.000
	农业化肥施用量	4.822	.121	1.019	39.708	.000
	粮食播种面积	.694	.046	.385	14.987	.000
3	(常量)	-36475.097	5141.413		-7.094	.000
	农业化肥施用量	4.655	.109	.984	42.569	.000
	粮食播种面积	.622	.043	.345	14.598	.000
	成灾面积	-.098	.025	-.091	-3.856	.001

a. 因变量: 粮食产量

已排除的变量^d

模型	Beta In	t	Sig.	偏相关	共线性统计量
					容差

1	粮食播种面积	.385 ^a	14.987	.000	.939	.932
	成灾面积	-.243 ^a	-4.017	.000	-.591	.926
	农业劳动人口	-.239 ^a	-3.321	.002	-.519	.735
2	成灾面积	-.091 ^b	-3.856	.001	-.582	.748
	农业劳动人口	-.042 ^b	-1.279	.211	-.231	.567
3	农业劳动人口	.038 ^c	1.097	.282	.203	.349

a. 模型中的预测变量: (常量), 农业化肥施用量。

b. 模型中的预测变量: (常量), 农业化肥施用量, 粮食播种面积。

c. 模型中的预测变量: (常量), 农业化肥施用量, 粮食播种面积, 成灾面积。

d. 因变量: 粮食产量

引入x5后, p值大于0.1, 不显著, 所以排除

引入变量农业机械总动力(x4)

输入/移去的变量^a

模型	输入的变量	移去的变量	方法
1	农业化肥施用量	.	向前 (准则: F-to-enter 的概率 <= .050)
2	粮食播种面积	.	向前 (准则: F-to-enter 的概率 <= .050)
3	成灾面积	.	向前 (准则: F-to-enter 的概率 <= .050)

a. 因变量: 粮食产量

模型汇总

模型	R	R 方	调整 R 方	标准估计的误差
1	.919 ^a	.844	.839	2679.938
2	.991 ^b	.982	.980	935.132
3	.994 ^c	.988	.987	773.306

a. 预测变量: (常量), 农业化肥施用量。

b. 预测变量: (常量), 农业化肥施用量, 粮食播种面积。

c. 预测变量: (常量), 农业化肥施用量, 粮食播种面积, 成灾面积。

Anova^d

模型		平方和	df	均方	F	Sig.
1	回归	1.202E9	1	1.202E9	167.418	.000 ^a
	残差	2.226E8	31	7182067.287		
	总计	1.425E9	32			
2	回归	1.399E9	2	6.994E8	799.806	.000 ^b
	残差	26234168.12	30	874472.271		
	总计	1.425E9	32			
3	回归	1.408E9	3	4.692E8	784.672	.000 ^c
	残差	17342075.84	29	598002.615		
	总计	1.425E9	32			

a. 预测变量: (常量), 农业化肥施用量。

b. 预测变量: (常量), 农业化肥施用量, 粮食播种面积。

c. 预测变量: (常量), 农业化肥施用量, 粮食播种面积, 成灾面积。

d. 因变量: 粮食产量

系数^a

模型		非标准化系数		标准系数	t	Sig.
		B	标准误差	试用版		
1	(常量)	30769.547	1406.872		21.871	.000
	农业化肥施用量	4.347	.336	.919	12.939	.000
2	(常量)	-47212.814	5226.512		-9.033	.000
	农业化肥施用量	4.822	.121	1.019	39.708	.000
	粮食播种面积	.694	.046	.385	14.987	.000
3	(常量)	-36475.097	5141.413		-7.094	.000
	农业化肥施用量	4.655	.109	.984	42.569	.000
	粮食播种面积	.622	.043	.345	14.598	.000
	成灾面积	-.098	.025	-.091	-3.856	.001

a. 因变量: 粮食产量

已排除的变量^d

模型	Beta In	t	Sig.	偏相关	共线性统计量
					容差

1	粮食播种面积	.385 ^a	14.987	.000	.939	.932
	成灾面积	-.243 ^a	-4.017	.000	-.591	.926
	农业劳动人口	-.239 ^a	-3.321	.002	-.519	.735
	农业机械总动力	.351 ^a	2.161	.039	.367	.170
2	成灾面积	-.091 ^b	-3.856	.001	-.582	.748
	农业劳动人口	-.042 ^b	-1.279	.211	-.231	.567
	农业机械总动力	.123 ^b	2.093	.045	.362	.159
3	农业劳动人口	.038 ^c	1.097	.282	.203	.349
	农业机械总动力	.005 ^c	.077	.940	.014	.100

a. 模型中的预测变量: (常量), 农业化肥施用量。

b. 模型中的预测变量: (常量), 农业化肥施用量, 粮食播种面积。

c. 模型中的预测变量: (常量), 农业化肥施用量, 粮食播种面积, 成灾面积。

d. 因变量: 粮食产量

引入x4后, p值仍大于0.1, 不显著, 所以x4和x5均排除

回归方程

$$y = -36475.097 + 4.655 * X1 + 0.622 * X2 - 0.098 * X3$$