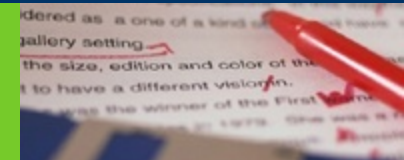




# 多元回归分析模型

主讲人：泰山教育 小石老师

# 背景

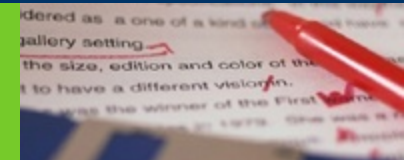


回归分析定义：

回归分析是一种统计学上分析数据的方法，目的在于了解两个或多个变量间是否相关、相关方向与强度，并建立数学模型以便观察特定变量来预测研究者感兴趣的变量。

回归分析思想：

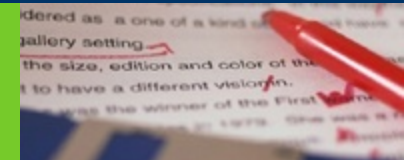
回归分析的基本思想是：虽然自变量和因变量之间没有严格的、确定性的函数关系，但可以设法找出最能代表它们之间关系的数学表达形式。



多元回归分析的由来：

在自变量很多时，其中有的因素可能对应变量的影响不是很大，而且 $x$ 之间可能不完全相互独立的，可能有种种互相作用的关系。在这种情况下可用逐步回归分析，进行 $x$ 因子的筛选，这样建立的多元回归模型预测效果会更好。

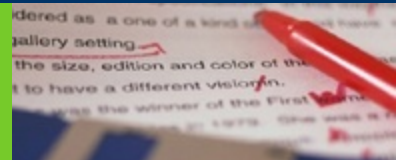
# 应用范围举例



收入水平与受教育程度、所在行业、工作年限、工作种类的关系。

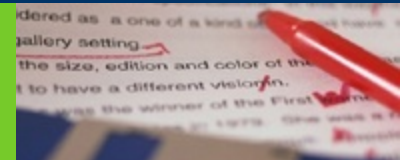
公路客运量与人口增长量、私家车保有量、国民生产总值、国民收入、工农业总产值、基本建设投资额、城乡居民储蓄额、铁路和水运客运量等因素的关系。

# 多元回归分析模型应用举例



以陕西省长武地区1984~1995年的烟蚜传毒病情资料、相关虫情和气象资料为例，建立蚜传病毒病情指数的逐步回归模型，说明逐步回归分析的具体步骤。影响蚜传病毒病情指数的虫情因子和气象因子一共有21个，通过逐步回归，从中选出对病情指数影响显著的因子，从而建立相应的模型。

# 多元回归分析模型应用举例



$y$ : 历年病情指数

$x_1$ : 前年冬季油菜越冬时的蚜量(头/株)

$x_2$ : 前年冬季极端气温

$x_3$ : 5月份最高气温

$x_4$ : 5月份最低气温

$x_5$ : 3~5月份降水量

$x_6$ : 4~6月份降水量

$x_7$ : 3~5月份均温

$x_8$ : 4~6月份均温

$x_9$ : 4月份降水量

$x_{10}$ : 4月份均温

$x_{11}$ : 5月份均温

$x_{12}$ : 5月份降水量

$x_{13}$ : 6月份均温

$x_{14}$ : 6月份降水量

$x_{15}$ : 第一次蚜迁高峰期百株烟草有翅蚜量

$x_{16}$ : 5月份油菜百株蚜量

$x_{17}$ : 7月份降水量

$x_{18}$ : 8月份降水量

$x_{19}$ : 7月份均温

$x_{20}$ : 8月份均温

$x_{21}$ : 元月均温

# 结果分析

输入／移去的变量<sup>a</sup>

模型	输入的变量	移去的变量	方法
1	x15	.	步进（准则：F-to-enter 的概率 $\leq .150$ ， F-to-remove 的概率 $\geq .200$ ）。
2	x4	.	步进（准则：F-to-enter 的概率 $\leq .150$ ， F-to-remove 的概率 $\geq .200$ ）。
3	x7	.	步进（准则：F-to-enter 的概率 $\leq .150$ ， F-to-remove 的概率 $\geq .200$ ）。
4	x5	.	步进（准则：F-to-enter 的概率 $\leq .150$ ， F-to-remove 的概率 $\geq .200$ ）。

# 结果分析

系数

模型		非标准化系数		标准系数		
		B	标准 误差	试用版	t	Sig.
1	(常量)	-.047	.114		-.414	.687
	x15	.009	.000	.993	27.061	.000
2	(常量)	-.163	.118		-1.386	.199
	x15	.009	.000	1.008	30.058	.000
	x4	.036	.019	.064	1.916	.088
3	(常量)	.966	.577		1.674	.133
	x15	.009	.000	1.016	34.600	.000
	x4	.034	.016	.060	2.070	.072
	x7	-.118	.059	-.057	-1.988	.082
4	(常量)	2.492	.652		3.821	.007
	x15	.009	.000	.979	40.466	.000
	x4	.034	.011	.062	3.005	.020
	x7	-.235	.057	-.114	-4.101	.005
	x5	-.003	.001	-.094	-3.000	.020



# 结果分析

在统计假设检验中，公认的小概率事件的概率值被称为统计假设检验的显著性水平，对同一量，进行多次计量，然后算出平均值。对于偏离平均值的正负差值，就是其不确定度。其差值越大，则计量的不确定性越大

Anova

模型		平方和	df	均方	F	Sig.
1	回归	18.555	1	18.555	732.287	.000 <sup>a</sup>
	残差	.253	10	.025		
	总计	18.808	11			
2	回归	18.628	2	9.314	465.793	.000 <sup>b</sup>
	残差	.180	9	.020		
	总计	18.808	11			
3	回归	18.688	3	6.229	413.702	.000 <sup>c</sup>
	残差	.120	8	.015		
	总计	18.808	11			
4	回归	18.755	4	4.689	622.720	.000 <sup>d</sup>
	残差	.053	7	.008		
	总计	18.808	11			

# 结果分析

系数

		非标准化系数		标准系数		
		B	标准 误差	试用版	t	Sig.
模型1	(常量)	-.047	.114		-.414	.687
	x15	.009	.000	.993	27.061	.000
模型2	(常量)	-.163	.118		-1.386	.199
	x15	.009	.000	1.008	30.058	.000
	x4	.036	.019	.064	1.916	.088
模型3	(常量)	.966	.577		1.674	.133
	x15	.009	.000	1.016	34.600	.000
	x4	.034	.016	.060	2.070	.072
	x7	-.118	.059	-.057	-1.988	.082
模型4	(常量)	2.492	.652		3.821	.007
	x15	.009	.000	.979	40.466	.000
	x4	.034	.011	.062	3.005	.020
	x7	-.235	.057	-.114	-4.101	.005
	x5	-.003	.001	-.094	-3.000	.020

$$y = 2.492 + 0.009x_{15} + 0.034x_4 - 0.235x_7 - 0.003x_5$$

# 结果分析

由回归方程式可以看出，在陕西长武烟草蚜传病毒病8月份的病情指数( $y$ )与 $x_4$ (5月份最低气温)、 $x_{15}$ (第一次蚜迁高峰期百株烟草有翅蚜量)呈显著正相关，而与 $x_5$ (3~5月份降水量)和 $x_7$  (3~5月份均温)呈显著负相关。



Thank You !