一元线性回归分析和
多元线性回归分析

信171 李金哲 201707010119

CONTENT

01

回归分析一般步骤

0 1 确定回归分析中的解释变量和被解释变量

由于回归分析用于分析一个事物如何随其他事物的变化而变化，因此回归分析的第一步应确定哪个事物是需要解释的，即哪个变量是被解释变量(记为y);哪些事物是用于解释其他变量的、即哪些变量是解释变量(记为x)。回归分析正是要建立y关于x的回归方程，并在给定x的条件下，通过回归方程预测y的平均值。这点是有别于相关分析的。

0 2 确定回归模型

根据函数拟合方式，通过观察散点图确定应通过哪种数学模型来概括回归线。如果被解释变量和解释变量之间存在线性关系，则应进行线性回归分析，建立线性回归模型;反之，如果被解释变量和解释变量之间存在非线性关系，则应进行非线性回归分析，建立非线性回归模型

0 3 建立回归方程

根据收集到的样本数据以及上一步所确定的回归模型，在一定的统计拟合准则下估计出模型中的各个参数，得到一个确定的回归方程。

0 4 对回归方程进行各种检验

前面已经提到，由于回归方程是在样本数据基础上得到的，回归方程是否真实地反映了事物总体间的统计关系以及回归方程能否用于预测等都需要进行检验。

0 5 利用回归方程进行预测

建立回归方程的目的之一是根据回归方程对新数据的未知被解释变量取值进行预测。利用SPSS进行回归分析时，应重点关注上述过程中的第一步和最后一步，至于中间各步，SPSS会自动进行计算并给出最佳的模型，。

02

一元线性回归操纵及其分析

壹

一元线性回归实验操作

1 选择菜单 【分析】->【回归】->【线性】

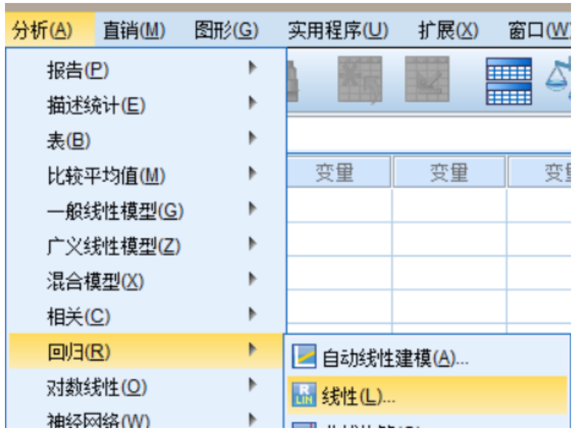

图2-1-1 选择菜单

2 将腰围拖入因变量；体重拖入自变量

图2-1-2 线性回归分析窗口

3 点击选项，将方程中包含常量勾选

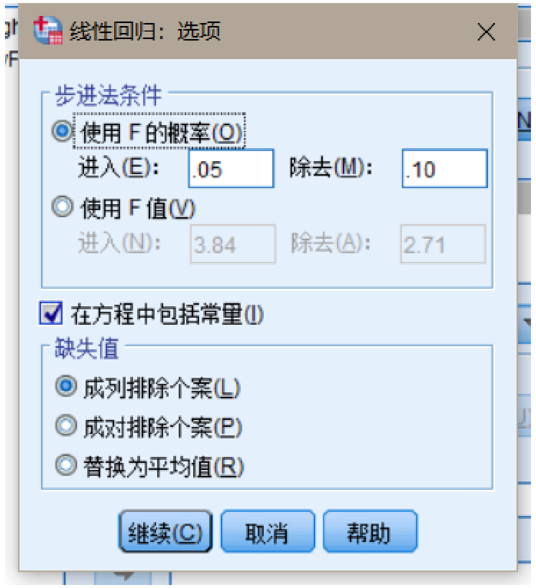

图2-1-3 绘图窗口

贰

一元线性回归结果分析

表2-2-1 一元线性回归方程结果分析

ANOVA[a]

模型		平方和	自由度	均方	F	显著性
1	回归	201.419	1	201.419	48.000	.000[b]
	残差	75.531	18	4.196		
	总计	276.950	19			

a. 因变量：腰围（英寸）

b. 预测变量：(常量), 体重（磅）

由表2-2-1可知，回归方程显著性检验的F统计量的观测值为48.000，其对应的概率P-值近似为0。若显著性水平a为0.05，因概率P值小于a，**拒绝回归方程显著性检验的原假设，即回归系数不同时为0，解释变量与被解释变量间存在显著的线性关系，选择线性模型具有合理性。**

表2-2-2 一元线性回归方程结果分析

系数[a]

模型		未标准化系数		标准化系数	t	显著性	B 的 95.0% 置信区间	
		B	标准误差	Beta			下限	上限
1	(常量)	14.019	3.356		4.178	.001	6.969	21.069
	体重（磅）	.122	.018	.853	6.928	.000	.085	.159

a. 因变量：腰围（英寸）

由表2-2-2可知，腰围和体重的回归方程近似等于

腰围 = 14.019 + 0.122 * 体重

其中常数的取值范围为[6.969, 21.069]，体重的取值范围为[0.085, 0.159]。

03

多元线性回归操纵及其分析

多元线性回归实验操作

1 选择菜单 【分析】->【回归】->【线性】

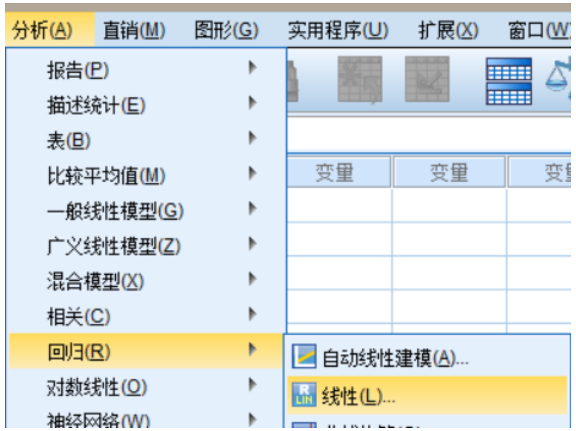

图3-1-1 选择菜单

2 将腰围拖入因变量；体重和脂肪比重拖入自变量，其中方法选择前进

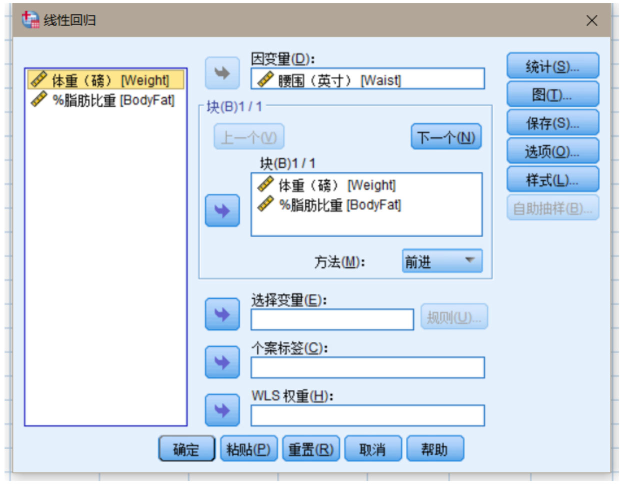

图3-1-2 线性回归分析窗口

3 点击图，如图3-1-3设置

图3-1-3 绘图窗口

4 点击保存，将预测值中的未标准化，预测区间的平均值勾选上

图3-1-4 保存窗口

贰

多元线性回归结果分析

表3-2-1关于腰围的分析结果

模型摘要[c]

模型	R	R 方	调整后 R 方	标准估算的误差
1	.887[a]	.787	.775	1.812
2	.945[b]	.894	.881	1.315

a. 预测变量：(常量), %脂肪比重

b. 预测变量：(常量), %脂肪比重, 体重（磅）

c. 因变量：腰围（英寸）

由表3-2-1可知。第一个模型是以脂肪比重为解释变量的一元线性回归方程。该模型的判定系数为0.787，回归方程的估计标准误为1.812。

第二个模型是包含脂肪比重和体重的二元线性回归方程，其判定系数增加至0.894，且调整的判定系数也有所增加，回归方程的估计标准误减小。从拟合优度角度看，**第二个模型的拟合效果更佳。**

表3-2-2关于腰围的分析结果

ANOVA[a]

模型		平方和	自由度	均方	F	显著性
1	回归	217.829	1	217.829	66.320	.000[b]
	残差	59.121	18	3.284		
	总计	276.950	19			
2	回归	247.541	2	123.770	71.545	.000[c]
	残差	29.409	17	1.730		
	总计	276.950	19			

a. 因变量：腰围（英寸）

b. 预测变量：(常量), %脂肪比重

c. 预测变量：(常量), %脂肪比重, 体重（磅）

由表3-2-2可知，被解释变量(腰围)的总离差平方和SST为276.950。一元模型(第一个模型)的回归平方和(SSR)为217.829，剩余平方和(SSE)为59.121;二元模型(第二个模型)增加了一个解释变量，剩余平方和减少为29.409，回归平方和增大为247.541。

对于二元模型，回归方程显著性检验的F统计量的观测值为71.545，其对应的概率P-值近似为0。若显著性水平a为0.05，因概率P值小于a，**拒绝回归方程显著性检验的原假设，即回归系数不同时为0，解释变量全体与被解释变量间存在显著的线性关系，选择线性模型具有合理性。**

叁

多元线性回归结果分析

表3-2-3 关于腰围的分析结果 系数[a]

模型		未标准化系数 B	标准误差	标准化系数 Beta	t	显著性
1	(常量)	30.058	.949		31.657	.000
	%脂肪比重	.354	.043	.887	8.144	.000
2	(常量)	20.236	2.468		8.199	.000
	%脂肪比重	.227	.044	.569	5.163	.000
	体重（磅）	.065	.016	.457	4.144	.001

a. 因变量：腰围（英寸）

表3-2-4 关于腰围的分析结果

排除的变量[a]

模型		输入 Beta	t	显著性	偏相关	共线性统计 容差
1	体重（磅）	.457[b]	4.144	.001	.709	.515

a. 因变量：腰围（英寸）

b. 模型中的预测变量：(常量), %脂肪比重

表3-2-3和表3-2-4是回归系数显著性检验的结果。

对于第一个模型，因脂肪比重与腰围的相关性高于体重，所以首先进入模型得到一元线性回归方程。此时，脂肪比重的回归系数显若性检验的t统计量的观测值为8.14，概率P值近似为零。当显著性水平a为0.05时，**应拒绝回归系数检验的原假设，认为脂肪比重与腰围有显著的线性关系，应保留在模型中。**

此时，按照回归策略，体重尚未进入回归模型，被列在表3-2-4中。表3-2-4表明如果体重被加入到第一个模型中(即建立二元模型)，其回归系数显著性检验的t统计量的观测值和概率P-值将为4.14和0.001(即概率P值近似为零。当显著性水平a为0.05时，**应拒绝回归系数检验的原假设，认为体重与腰围有显著的线性关系，应保留在模型中**)。在考虑脂肪比重的条件下，体重与腰围的偏相关系数为0.709，且线性关系显著，可以引入到第二个模型中。

对于第二个模型，表3-2-3表明，脂肪比重的偏回归系数以及回归系数显著性检验结果均因体重进入回归模型而发生了变化。两者的回归系数显著性检验均显著，无应被别除的解释变量，此时建模过程结束。

肆

多元线性回归结果分析

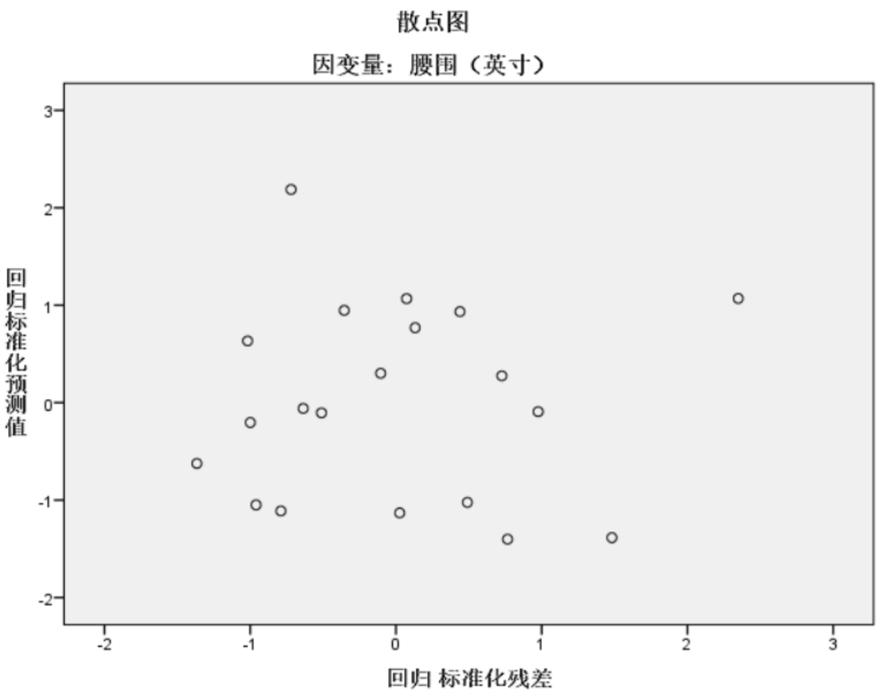

图3-2-5 关于腰围的结果图

图3-2-5是回归方程标准化预测值与标准化残差散点图，由于图3-2-5中的散点没有明显的趋势，故可以认为不存在明显的异方现象。

根据表3-2-3可得到最终的回归方程为：

腰围 = 20.236 + 0.227 * 脂肪比重 + 0.065 * 体重

04

曲线估计操纵及其分析

壹

曲线估计实验操作

1 选择菜单 【分析】->【回归】->【曲线估计】

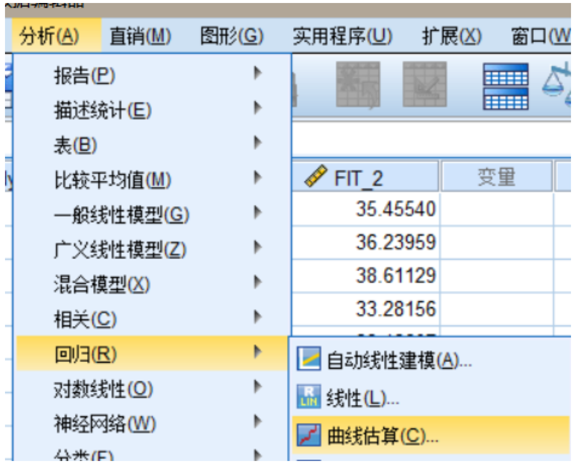

图4-1-1 选择菜单

2 将腰围拖入因变量；体重拖入自变量

图4-1-2 曲线估计分析窗口

3 勾选线性、二次和显示ANOVA表

图4-1-2 曲线估计分析窗口

贰

曲线估计结果分析

表4-2-1 曲线估计结果（一次）

ANOVA

	平方和	自由度	均方	F	显著性
回归	201.419	1	201.419	48.000	.000
残差	75.531	18	4.196		
总计	276.950	19			

自变量为 体重（磅）。

表4-2-2 曲线估计结果（一次）

系数

	未标准化系数		标准化系数		
	B	标准误差	Beta	t	显著性
体重（磅）	.122	.018	.853	6.928	.000
(常量)	14.019	3.356		4.178	.001

表4-2-3 曲线估计结果（二次）

ANOVA

	平方和	自由度	均方	F	显著性
回归	202.123	2	101.062	22.960	.000
残差	74.827	17	4.402		
总计	276.950	19			

自变量为 体重（磅）。

表4-2-4 曲线估计结果（二次）

系数

	未标准化系数		标准化系数		
	B	标准误差	Beta	t	显著性
体重（磅）	.214	.231	1.497	.927	.367
体重（磅）** 2	.000	.001	-.646	-.400	.694
(常量)	5.144	22.443		.229	.821

由表4-2-1和4-2-3可知，对于一次模型，回归方程显著性检验的F统计量的观测值为48.000；对于二次模型，回归方程显著性检验的F统计量的观测值为22.960，两个模型的概率P均值近似为零。当显著性水平a为0.05时，**应拒绝回归系数检验的原假设，在一次或二次模型腰围和体重均呈现明显相关性，应保留在模型中。**

但是由表4-2-2和4-24可知，二次项的拟合的优度明显低于线性方程，**当显著性水平a为0.05时，二次项回归方程的显著性检验呈现不显著，线性方程的显著检验呈现显著性，表明选择线性方程更合理**，符合拟合曲线的趋势（图4-2-5）。最终的回归方程为：

腰围 = 0.122 * 体重 + 14.019

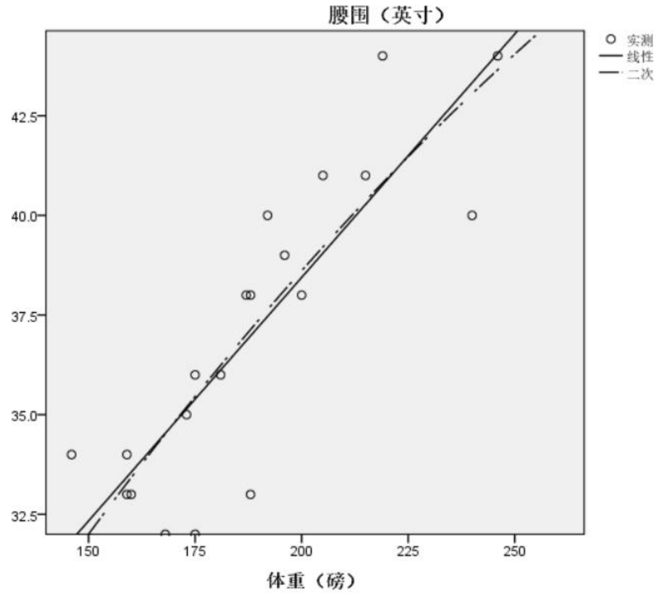

表4-2-5 拟合曲线