

## 预测方程和标准值(Z)

此模块是根据一个指标(Y)的分布,定义该指标的正常值范围。

如果Y符合正态分布,可以直接计算均值为 $\mu$  (“mu”)和标准差 $\sigma$  (“sigma”),然后计算Z值: $Z_i=(Y_i-\mu)/\sigma$

Z值为标准正态分布,其均数为0,标准差为1。根据Z值可以判断某个体偏离人群均值的大小,从而判断是否在正常范围内。

如果Y不符合正态分布,可以尝试通过函数转换,如对数转换,倒数转换,平方根转换等,使得转换后的值符合正态分布。

如果所观测的Y的值取决于另一个指标X,如儿童的身高受年龄影响,这时判断一个儿童的身高是否正常,需要结合年龄与身高两个指标来判断。通常我们调查一批人后,首先建立X(如年龄)对Y(如身高)的预测方程,如:

$$f(Y) = a + b(X) + e$$

方程中e称为残差,须符合正态分布,可以转换成Z值。每一个体的X与Y值代入方程后就可以计算其残差的Z值,从而判断其偏离人群均值的程度。

X对Y的预测方程可以有多种方式,如单纯直线性(只含X的一次项)的,含X的二次项等。易侓统计软件提供了多个选择,帮助用户比较。在评估方程拟合前,需要了解几个概念。如总离均差平方和(SST)、残差离均差平方和(SSE)、平均总离均差平方和(MST)、平均残差离均差平方和(MSE)、R平方、调整的R平方、Akaike信息(AIC)、对数似然值(logLikelihood)等概念。

总离均差平方和,即SST,即原始数据和均值之差的平方和;平均总离均差平方和:  $MST = SST / (n-1)$ ,均用来反映变异大小。

残差离均差平方和,即SSE,该统计参数计算的是原始数据与其对应的拟合数据的差的平方和。SSE越接近于0,说明模型拟合越好,数据预测也越成功。MSE和SSE效果一样。

平均残差离均差平方和:即  $MSE = SSE/n$ ,和SSE的意义相似。

R平方 =  $1 - SSE/SST$ ,也称为确定系数,表示拟合的好坏。“确定系数”的正常取值范围为[0, 1],越接近1,表明方程对Y的解释能力越强,模型对数据拟合得越好。调整的R平方是在R平方的基础上考虑了模型中的参数个数,其作用和功能与R平方相似。

AIC信息用来评价模型的简洁性和精确性。它的假设条件是模型的误差服从正态分布。优先考虑的模型应是AIC值最小的那一个。

似然值(likelihood):回归方程的似然值越大越好。似然值是个很小的小数,取对数后总是负数。所以对数似然值(likelihood)的绝对值则是越小,说明模型拟合的较好。

易侓统计预测方程与Z值模块提供了(1) multivariable fractional polynomials(MFP)方法筛选分数多项式,(2)对多个模型进行比较,(3)计算Z值曲线,判断正常值范围。

1) 多个预测方程比较:

例1:根据数据 demo.xls 中的 SBP 与 AGE,建立 AGE 对 SBP 的预测方程。

**单因素预测方程与Z值** ?

标题:

选择分析对象:

应变量Y(连续性变量)

变量  
Systolic BP, mmhg

自变量X(连续性变量)

变量  
Age, years

选择预测方程

$Y = \beta_0 + \beta_1 X$   
  $Y = \beta_0 + \beta_1 X + \beta_2 X^2$   
  $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$   
  $Y = \beta_0 + \beta_1 \ln(X)$   
  $Y = \beta_0 + \beta_1 (1/X)$   
  $Y = \beta_0 + \beta_1 (X^{0.5})$   
  $Y = \text{Smooth}(X)$   
  $Y = \text{mfp}(X)$

计算Z值

分层变量

这里比较 6 个可能的预测模型。

输出结果：

比较预测模型

Model	SST	SSE	DFT	DFE	MST	MSE	R2	Adj.R2	AIC	LLK
101.7622 + 0.7534*X	403639.76	314002.2	792	791	509.65	396.97	0.222	0.221	6999.6	-3496.8
128.0191 - 0.6677*X + 0.0169*X <sup>2</sup>	403639.76	303377.1	792	790	509.65	384.02	0.248	0.246	6974.3	-3483.2
38.8153 + 25.6623*X.log	403639.76	329929.74	792	791	509.65	417.1	0.183	0.182	7038.9	-3516.4
75.7587 + 9.0155*X.sqrt	403639.76	321175.96	792	791	509.65	406.04	0.204	0.203	7017.5	-3505.8
s(X, 4.84)	403639.76	295598.65	792	787.16	509.65	375.52	0.268	0.263	6959.4	-3472.9
335.7233 - 130.2259*(X/100) <sup>0.5</sup> + 215.7281*(X/100) <sup>0.5</sup> * log((X/100))	403639.76	301913.65	792	790	509.65	382.17	0.252	0.25	6970.5	-3481.3

### 结果解释

SST: 总离均差平方和

SSE: 残差离均差平方和

DFT: 总自由度; DFE: 残差自由度

MST: 平均总离均差平方和;

MSE: 平均残差离均差平方和

R2: R 平方= 1-SSE/SST;

Adj. R2: 调整的 R 平方= 1-MSE/MST

AIC: Akaike 信息;

LLK: 对数似然值: logLikelihood

fp(X): fractional polynomials

平滑曲线拟合结果 AIC 最小，对数似然值最大。但曲线拟合没有参数，无法通过公式计算预测值。最接近的是  $f_p(X)$  即分数多项式方程。

输出各模型 Z 值曲线图，及各 AGE 分别对应 SBP、DBP 的 Z 值，存放在 excel 文件中，右击输出文件即可查看该 excel 文件。

