
线性回归分析的思路梳理

作者：夏雨骄阳 来源：SPSS学堂

本文原地址：<https://www.iikx.com/news/statistics/9693.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

线性回归分析的思路梳理：

简单线性回归

1 根据研究目的确定因变量和自变量。

2 判断有无异常值。

通过绘制散点图直观观察;亦可通过线性回归的【统计】 【个案诊断】 【所有个案】进行分析，若标准残差超过[-3,3]，则可视为异常值。

如果发现异常值，则首先应该检查是否是数据收集或录入方面的错误，如是则应及时纠正。如果不是数据收集或录入方面的错误，则需根据实际情况，选择剔除或者保留异常值。

3 判断数据是否满足简单线性回归假设条件。

第一，线性(linear)，因变量与自变量呈线性关系，通过绘制散点图判断。

第二，独立性(independent)，任意两个观察值之间相互独立，通过线性回归的【统计】 【德宾-沃森】进行分析，一般来说Durbin-Waston检验值分布在0-4之间，越接近2，观察值相互独立的可能性越大。

第三，残差正态性(normal)，随机误差近似正态性，可通过直方图或者P-P图判断残差是否符合正态分布。

第四，通过线性回归的【图】 【产生所有部分图】，即可得到残差随着估计值的变化趋势，若所有点均匀分布于直线 $Y=0$ 的两侧，则可认为方差齐性。

4 估计回归模型参数，建立模型。

5 对模型进行假设检验。

对回归模型进行假设检验一般使用方差分析法，对回归系数进行假设检验一般使用t检验方法。

多重线性回归

1 根据研究目的确定因变量和自变量。

2 判断有无异常值。

通过线性回归-统计-个案诊断，线性回归-保存-勾选学生化删除、库克距离、杠杆值，根据新生成的学生化删除残差、库克距离、杠杆值来判断。

学生化删除残差的值在-3至3的范围内，库克距离均小于1，杠杆值均均小于0.2，不存在异常值。

如果发现异常值，则首先应该检查是否是数据收集或录入方面的错误，如是则应及时纠正。如果不是数据收集或录入方面的错误，则需根据实际情况，选择去除异常值、转换异常值的变量，或者选用非参数分析法、最小一乘法来处理。

3 判断数据是否满足多重线性回归假设条件。

第一，因变量与所有自变量之间是否存在线性关系。

通过建立未标化预测值(PRE_1)和学生化残差(SRE_1)的散点图判断，未标化预测值(PRE_1)和学生化残差(SRE_1)的散点图呈水平带状，则满足因变量与所有自变量之间存在线性关系的假设。

第二，因变量与每一个自变量之间是否存在线性关系。

通过线性回归的【图】 【产生所有部分图】中的散点图判断。

第三，方差齐性。通过线性回归的【ANOVA】表的Sig值判断，小于0.05为方差齐性，大于0.05为方差不齐。

第四，各自变量之间是相互独立的。

通过线性回归的 【统计】 【共线性诊断】的结果，容许度越接近1，多重共线性越弱，膨胀因子越接近1，多重共线性越弱;膨胀因子小于10为弱多重共线性，大于10则存在严重共线性。

如果各自变量之间存在严重共线性，则可选用岭回归或者主成分分析法来处理。

第五，残差近似正态性。可通过直方图或者P-P图判断残差是否符合正态分布。

4 估计回归模型参数，建立模型。

可根据实际情况，选用强迫引入法、逐步引入法、强迫剔除法、向后剔除法、向前引入法来进行模型估计。

方差不齐时，则需要用加权最小二乘法来进行模型估计。

5 对模型进行假设检验。

对回归模型进行假设检验一般使用方差分析法，对回归系数进行假设检验一般使用t检验方法。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发