

判断正误（除非特别说明，所有模型都带有截距项，所有拟合都是 LS）

- (1) 线性回归模型 $y = a + bx + cz + \epsilon$ (满足通常假设) 蕴含了如下事实: 给定 z 值时, x, y 之间的相关性不依赖于 z 的特定取值.
- (2) 最小二乘拟合值的样本均值等于响应变量的样本均值.
- (3) 最小二乘拟合值的样本方差不大于响应变量的样本方差.
- ✗(4) 简单线性回归中自变量的方差越小, 斜率的最小二乘估计的方差越小.
- ✗(5) 随机变量 y 与 x 不相关, z 与 x, y 都正相关, 则回归方程 $y = a + bx + cz + \epsilon$ 中的回归系数 $b \geq 0$.
- (6) 复共线性 (collinearity) 指的是自变量之间的相关性. 复共线性并不会导致 OLS 估计出现偏差.
- (7) 复共线性会导致某些回归系数的 LS 估计的标准差变大.
- (8) 基于观察研究数据进行因果推断, 可使用回归模型对干扰因素进行控制.
- ✗(9) 如果数据不符合线性回归模型的 Gauss-Markov 假设, 则计算机统计软件就不能运用该模型拟合数据.
- ✗(10) 若响应变量是 0-1 变量, 则不能拟合线性回归模型.
 - (11) 若回归方程中某个变量与其它变量正交, 则剔除该变量不会使其它系数的 LS 估计出现偏差.
- ✗(12) 如果应用线性回归模型拟合数据得到的决定系数 R^2 接近于 1, 则表明该模型是正确的模型.
 - (13) 线性回归模型中增加一个自变量, 则决定系数会变大或保持不变.
 - (14) 残差向量与拟合值向量一定正交.
- ✗(15) 杠杆值 h_{ii} 最小为 0.
- (16) 在多重线性回归模型中删除一些自变量有可能会降低预测误差.
- (17) 如果一个样本点恰好落在回归直线上, 则其 Cook 距离值为 0.
- (18) 如果一个样本点的自变量恰好等于自变量的样本均值, 则其杠杆值 h_{ii} 最小.
- ✗(19) 如果一个样本点的自变量恰好等于自变量的样本均值, 则这个样本点一定落在回归直线上.
 - (20) 回归方程中某个变量与其它变量正交, 剔除该变量不改变其它变量回归系数的 LS 估计.
 - (21) 残差之和一定为 0.
 - (22) 最小二乘法得到的残差与所有自变量的样本相关系数都为 0.
- ✗(23) 如果 $Y = X\beta + \epsilon$ 中的 $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ 各分量独立, 均值为 0, 与 X 独立, 但方差不同, 则最小二乘估计可能是有偏的.
 - (24) 两样本 t -检验等价于简单线性模型的 F 检验.
 - (25) 决定系数越大, 回归方程的显著性检验统计量的值越大.
 - (26) 投影矩阵一定是半正定矩阵.
- ✗(27) 删除一个样本点, 残差平方和会增大或不变.
- ✗(28) 均方误差等于方差与偏差之和.
 - (29) C_p 准则是子模型 (或变量) 选择的准则, 它代表了所选子模型拟合值的均方误差.
 - (30) 待预测随机变量的方差越小, 则预测误差越小.

b与偏相关系数
 $\rho(x, y, z)$ 符号相同

1/n

偏差平方