

## 似然比检验 (log likelihood ratio test) :

似然, 英文 likelihood, 实际上也可以翻译为可能性。一个回归模型的似然值, 可以理解为给定模型参数后出现所观察样本的可能性。如模型 I:  $Y = \beta_0 + \beta_1 * X1 + \beta_2 * X2$ , 这个模型, 给定的模型参数是  $\beta_0$ 、 $\beta_1$ 、 $\beta_2$ , 抽样得出所观察到的样本的可能性为  $L1$ 。

似然比检验 (LRT) 用来比较两个模型, 看是否可以用一个简单模型来替代一个复杂模型。如上例模型 I, 若除去自变量  $X2$ , 模型 II:  $Y = \beta_0 + \beta_1 * X1$ , 这个模型参数少一个, 相对简单, 由这个模型抽样得出所观察到的样本的可能性为  $L2$ 。因为模型 II 少一个参数, 自然  $L2$  要小于  $L1$ 。现在要问的问题是: 是否  $L2$  比  $L1$  小得太多, 以至于我们不应该剔除  $X2$  呢?

判断是否可以简化模型 I 为模型 II 的标准, 是用似然比检验 (LRT):  $LR = 2 * (\ln L1 - \ln L2)$ , 其中  $\ln L1$  为复杂模型 I 的似然值对数,  $\ln L2$  为简单模型 II 的似然值对数。

$LR$  近似符合卡方分布, 自由度等于复杂模型 I 中的模型参数个数与简单模型 II 的参数个数差, 这里等于 1。这样根据卡方分布临界值表, 我们就可以判断模型差异是否显著。如果差异显著, 表示不能用简单模型 II 替代复杂模型 I。一般来说, 似然比检验得出结论与模型 I 中对  $X2$  回归系数的 Wald 检验结论基本一致。

在做似然比检验比较两个模型时, 要注意: (1) 简单模型是从复杂模型简化得来的, 即嵌套关系。(2) 两个模型所用的观察对象完全相同 (当所剔除的变量有缺失值需注意两模型所用的样本量是否相同)。