**KAPPA一致性检验**

（一）用途：1)可当做一个多分类评估指标，用于表征该模型预测和实际的一致性。如下图所示。



2)也可以用于检验两个人对20个房屋进行评价的意见是否一致。如下图所示。



（二）取值范围[-1,1]，但是经常取值范围[0,1]。以下是取值的含义：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 区间 | [0, 0.2] | [0.21, 0.4] | [0.41, 0.6] | [0.61, 0.8] | [0.81, 1] |
| 含义 | 极低一致性 | 一般一致性 | 中等一致性 | 高度一致性 | 几乎完全一致 |

（三）计算公式

k=(p0-pe)/(1-pe)，p0=预测值和实际值相等的个数/总数，pe=sum(行和\*列和)/(总数\*总数)，在上图混淆矩阵中计算可得

p0=(7+2+4)/20=0.65

pe=(13\*7+2\*8+5\*5)/(20\*20)=0.33

k=(0.65-0.33)/(1-0.33)=0.478

结论：这两个分析师的建议基本一致。

**相对风险(relative risk)和优势比(odds)**

如下图所示，泰坦尼克号人员生存统计。

****

相对风险Female=154/462=0.3333；Male=709/851=0.8331

Male relative risk/Female relative risk =0.8331 /0.3333=2.5

这说明男性比女性死亡率高2.5倍。

优势比 Female=154/308=0.5；Male=709/142=4.993

Odds ratio=4.993/0.5=9.986

这说明男性比女性死亡发生的可能性高9.986倍。

在临床研究中,以及在其他一些设置,通常是最受关注的参数相对风险而不是优势比。相对风险最好使用人口样本估计,但如果罕见疾病的假设持有的优势比是一个很好的近似相对风险。

参考链接<http://www.pmean.com/01/oddsratio.html>

**简单相关系数和偏相关系数**

在[多元回归分析](https://baike.baidu.com/item/%E5%A4%9A%E5%85%83%E5%9B%9E%E5%BD%92%E5%88%86%E6%9E%90" \t "_blank)中，在消除其他变量影响的条件下，所计算的某两变量之间的相关系数。在多元[相关分析](https://baike.baidu.com/item/%E7%9B%B8%E5%85%B3%E5%88%86%E6%9E%90" \t "_blank)中，[简单相关系数](https://baike.baidu.com/item/%E7%AE%80%E5%8D%95%E7%9B%B8%E5%85%B3%E7%B3%BB%E6%95%B0" \t "_blank)可能不能够真实的反映出变量X和Y之间的相关性，因为变量之间的关系很复杂，它们可能受到不止一个变量的影响。这个时候偏相关系数是一个更好的选择。

假设我们需要计算X和Y之间的相关性，Z代表其他所有的变量，X和Y的偏相关系数可以认为是X和Z[线性回归](https://baike.baidu.com/item/%E7%BA%BF%E6%80%A7%E5%9B%9E%E5%BD%92)得到的[残差](https://baike.baidu.com/item/%E6%AE%8B%E5%B7%AE" \t "_blank)Rx与Y和Z线性回归得到的残差Ry之间的简单相关系数，即pearson相关系数。

可以利用偏相关系数进行变量间的净相关分析：

1. 计算样本的偏相关系数

method1：从线性回归的角度计算变量间的偏相关系数。（不知怎么做？）

Method2：可以认为简单相关系数是0阶偏相关系数，任意N阶偏相关可以通过3个（N-1）阶偏相关系数计算出来。

Method3：首先计算出所有变量的相关性矩阵，然后求它的逆矩阵，这样可以求出任何两两变量之间的偏相关系数。

1. 对样本来自的两个总体是否存在显著的净相关进行推断

t-test：原假设—两总体的偏相关系数与0无差异。（可以理解）

fisher转化法