 **卡方检验**就是统计样本的实际观测值与理论推断值之间的偏离程度，实际观测值与理论推断值之间的偏离程度就决定卡方值的大小，卡方值越大，越不符合；卡方值越小，偏差越小，越趋于符合，若两个值完全相等时，卡方值就为0，表明理论值完全符合。

为什么还要使用ROC和AUC呢？因为ROC曲线有个很好的特性：**当测试集中的正负样本的分布变化的时候，ROC曲线能够保持不变。**

在实际的数据集中经常会出现类不平衡（class imbalance）现象，即负样本比正样本多很多（或者相反），而且测试数据中的 正负样本的分布也可能随着时间变化。

**准确率(accuracy)**,其定义是: 对于给定的测试数据集，分类器正确分类的样本数与总样本数之比。

是最常见的evaluation metric但在正反例不均衡的情况下尤其是我们对少数的类更感兴趣的时候，如fraud detection(欺诈监测)，癌症检测。

在测试集里，有100个sample，99个反例，只有1个正例。如果我的模型不分青红皂白对任意一个sample都预测是反例，那么我的模型的accuracy是 正确的个数／总个数 = 99/100 = 99%。你拿着这个accuracy高达99%的模型屁颠儿屁颠儿的去预测新sample了，而它一个正例都分不出来，有意思么。。。

也有人管这叫accuracy paradox。

**精确率(precision)**的公式是P = {TP}/{TP+FP} 它计算的是所有"正确被检索的item(TP)"占所有"实际被检索到的(TP+FP)"的比例

**召回率(recall)**的公式是 R = TP}/{TP+FN},它计算的是所有"正确被检索的item(TP)"占所有"应该检索到的item(TP+FN)"的比例。

**F1值:**是精确率和召回率的调和均值，即F1=2PR/(P+R)，相当于精确率和召回率的综合评价指标。

**Fα值**:为F1值的变体， (α²+1）PR/(α²P+R)，利用α给P和R赋予不同的权重，若α=1则为F1值。

也有人会用Gini系数来评价模型，其实Gini系数与AUC所表示的意义相同，只是计算方式不同。Gini系数指ROC曲线与中线（上图红线）围成的面积和中线（上图红线）之上的面积（0.5）的比例，两者之间换算公式为Gini=2\*AUC-1。

在评价模型时还会用到KS（Kolmogorov-Smirnov）值，KS=max(TPR-FPR)，即为TPR与FPR的差的最大值，KS值可以反映模型的最优区分效果，此时所取的阈值一般作为定义好坏用户的最优阈值。上图ROC曲线的KS值为0.45，此时TPR=0.79，FPR=0.34。

当然，阈值的选取还要考虑应用场景及业务要求，对于FPR不敏感而对TPR敏感的场景，可以适当减少阈值以增加TPR。

如精准营销领域的商品推荐模型，模型目的是尽量将商品推荐给感兴趣的用户，若用户对推荐的商品不感兴趣，也不会有很大损失，因此此时TPR相对FPR更重要。

再比如反欺诈领域的欺诈预测模型，由于模型结果会对识别的坏人进行一定的处置措施，FPR过高会对好人有一定干扰，造成误杀，影响客户体验，因此模型需保证在低于一定FPR的基础上尽量增加TPR。

**PRC， precision recall curve:**

PR曲线（Precision-Recall curve）和ROC曲线类似，ROC曲线是FPR和TPR的点连成的线，PR曲线是准确率和召回率的点连成的线，如下图所示。

我们又知道，Recall=TPR，因此PRC的横坐标为ROC的纵坐标。

