

## 1. 论述True Positive Rate、False Positive Rate与Precision、Recall的联系。

首先需要了解TP, FP, TN, FN

TP: **true positive**, 正确的阳性, 说明预测是阳性, 而且预测对了, 那么实际也是正例。

FP: **false positive**, 假阳性, 说明预测是阳性, 预测错了, 所以实际是负例。

TN: **true negative**, 正确的阴性, 说明预测是阴性, 而且预测对了, 那么实际也是负例。

FN: **false negative**, 假阴性, 说明预测是阴性, 预测错了, 所以实际是正例。

由以上四个可以得到TPR和FPR:

**True Positive Rate**: 真正率, 表示当前分到正样本中真实的正样本所占所有正样本的比例

$$TPR = TP / (TP + FN)$$

**False Positive Rate**: 假正率, 表示当前被错误分到正样本类别中真实的负样本所占所有负样本总数的比例

$$FPR = FP / (FP + TN)$$

由以上四个可以得出Precision 和 Recall

查准率 (**precision**)

精度是精确性的度量, 表示被分为正例的示例中实际为正例的比例, **precision**= $TP / (TP + FP)$

查全率 (**recall**)

查全率(也称召回率)是覆盖面的度量, 度量有多个正例被分为正例, **recall**= $TP / (TP + FN)$

## 2. 假阴性在近期新冠肺炎检测中出现的主要原因?

假阴性在这次疫情中指的是检测报告上写的是阴性(没有感染), 但其实这个结果是错的, 其实应该是阳性(已经得病)称假阴性。

而在这次新型冠状病毒肺炎的检测中, 最常用的方式是核酸检测法。核酸检测是金标准, 根据国家卫健委临床检验中心介绍, 新型冠状病毒与2003年的SARS冠状病毒一样, 由RNA核酸和蛋白等组成。RNA核酸由四种不同碱基排列而成, 不同病毒的核酸碱基排列方式, 就像发报的密电码, 各有不同。因此可以通过采集人的特定部位细胞样本, 检测其中是否含有病毒RNA核酸, 确认机体是否受到感染。核酸检测就像分析密电码, 新型冠状病毒也有其特异的密电码, 但与其它冠状病毒也有相似性, 根据2月19日国家卫健委发布的《新型冠状病毒肺炎诊疗方案(试行第六版)》, 目前研究显示, 新型冠状病毒与蝙蝠SARS样冠状病毒同源性达85%以上。就像同一地区身份证号有几位数字相同, 那么就通过其他区域的数字对个人进行区分。

但近日, 国家卫健委印发《新型冠状病毒肺炎防控方案(第五版)》, 在实验室检测技术指南中明确指出, 核酸检测结果阴性不能排除新冠病毒感染, 需要排除可能产生假阴性的因素, 原因可能有:

(1) 患者可能病毒含量太少: 不同病程、不同病情患者机体中的病毒存在量可能不同, 已有病毒感染, 但由于在相关部位采集不到病毒或采集到的病毒量太少, 以致于现有方法检测不到, 也称窗口期

(2) 采集部位不当, 采集深度不够, 采集到的细胞可能大部分都不含病毒: 呼吸道采集的痰、肺泡灌洗液标本阳性率高于上呼吸道的口鼻咽拭子采集的标本; 鼻咽拭子标本检测阳性率又高于口咽拭子

(3) 检测试剂质量或参差不齐

**3.数据集包含1000个样本，其中500个正例，500个负例，将其划分为70%训练，30%测试的留出法，估算多少种划分方式。**

可以将数据集划分成两个对立的集合，占70%元素的为训练集，占30%元素的为测试集，可以用分层抽样的方式。

其结果为  $C_{500}^{350} C_{500}^{150} \approx 1.4 \times 10^{298}$