2-1、数据集包含 1000 个样本，其中 500 个正例、500 个反例，将其划分为包含 70%样本的训练集和 30%样本的测试集用于留出法评估，试估算共有多少种划分方式？

答：采用分层采样方式，训练集包含 350 个正例和 350 个反例；测试集包含 150 正例和 150 反例

$$N = C_{500}^{350} \times C_{500}^{350} \times C_{500}^{350} \times C_{500}^{350}$$

2-2、试述真正例率（TPR）、假正例率（FPR）与查准率（P）、查全率（R）之间的关系。

答：$TPR = \frac{TP}{TP + FN}$　　$P = \frac{TP}{TP + FP}$

$FPR = \frac{FP}{TN + FP}$　　$R = \frac{TP}{TP + FN}$

2-3、试述错误率与 ROC 曲线之间的关系。

答：错误率 = 1 − 正确率 = $1 - \frac{TP + TN}{TP + TN + FP + FN} = \frac{FN + FP}{TP + TN + FP + FN}$

真正例率：$TPR = \frac{TP}{TP + FN}$

假正例率：$FPR = \frac{FP}{TN + FP}$

ROC 图像上的每个点，对应一个错误率

因为样本中，正反例比例确定，理想模型在 (0,1) 点上

故离 (0,1) 点越近，错误率越低