Introduction:

Empiricism - Positivism - Logical positivism - Verificationism - Pragmatism - Falsificationism

How to make a decision:

• The classical statistics: Significance testing.

a tool for decision making under uncertainty

**Null hypothesis (H0) (no effect) and its complement (H1).**

**Significance level (α, p value).**

Statistical power (1- β), probability of rejecting H0 given that it should be rejected.

Sample size n / Effect size.

**Power curve**: central limit theorem, whether under the curve,

(Assume complete ignorance about the world (“anything possible”).

Formulate a hypothesis H0 and compare P(data|H0) to a pre-defined probability threshold (significance level) α.

**If P(data|H0)<α, reject H0 (the data are unlikely/surprising if H0 is true).**

(If the observed mean falls outside the 95% range, we reject H0 ))

risks

**Type I errors (reject H0 when true).  
Type II errors (accept H0 when false).**

Whether the “correct” significance level? α = 0.05

Decision making: we need to consider uncertainty and preferences.

Probability is a measure of uncertainty

Utility is a measure of preference that combines with probability as mathematical expectation.

The main problem (of course, after determining what the value of the outcomes are) is to determine the **prior probability of the hypothesis**. To see that, start with Pr(H0|D) and the derive everything using Bayes theorem in terms of Pr(H0), Pr(D|H0), and Pr(D|H1).

Recall the possible errors are: (type I and type II) - α and β are probabilities of these errors.

**Decision-theoretic (Bayesian) view allows to explore the exact relation between the significance level and the decision.**

**Confidence Intervals**

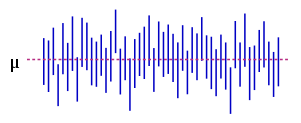
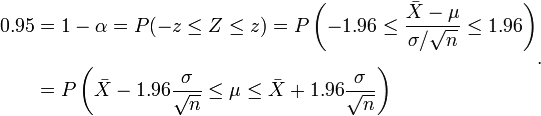
from the point of view of the **classical hypothesis testing**

***I'm 95% sure that the true mean is inside this interval***

标准差：standard deviation

page34image3894944

Standard error:

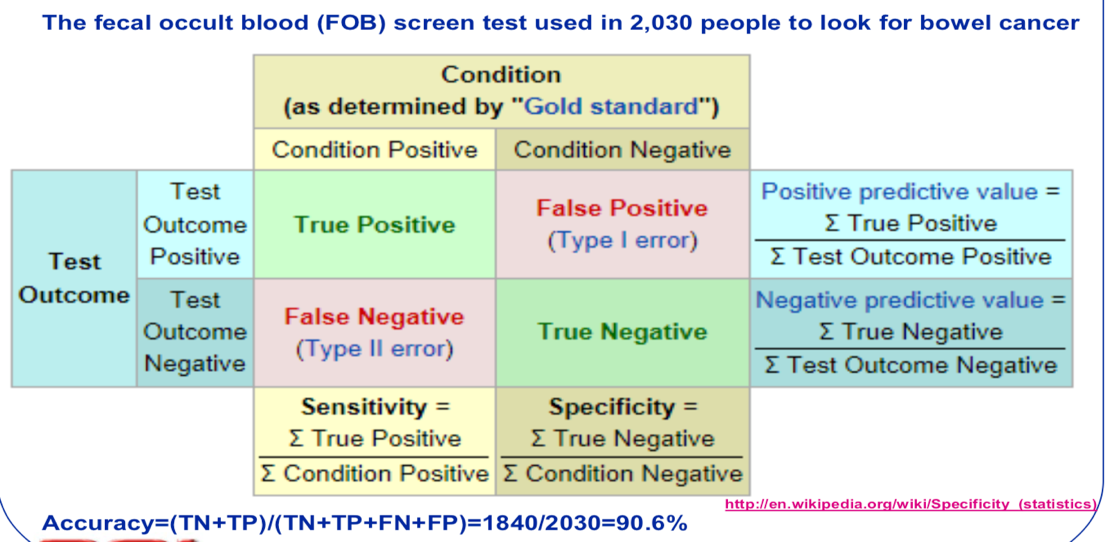


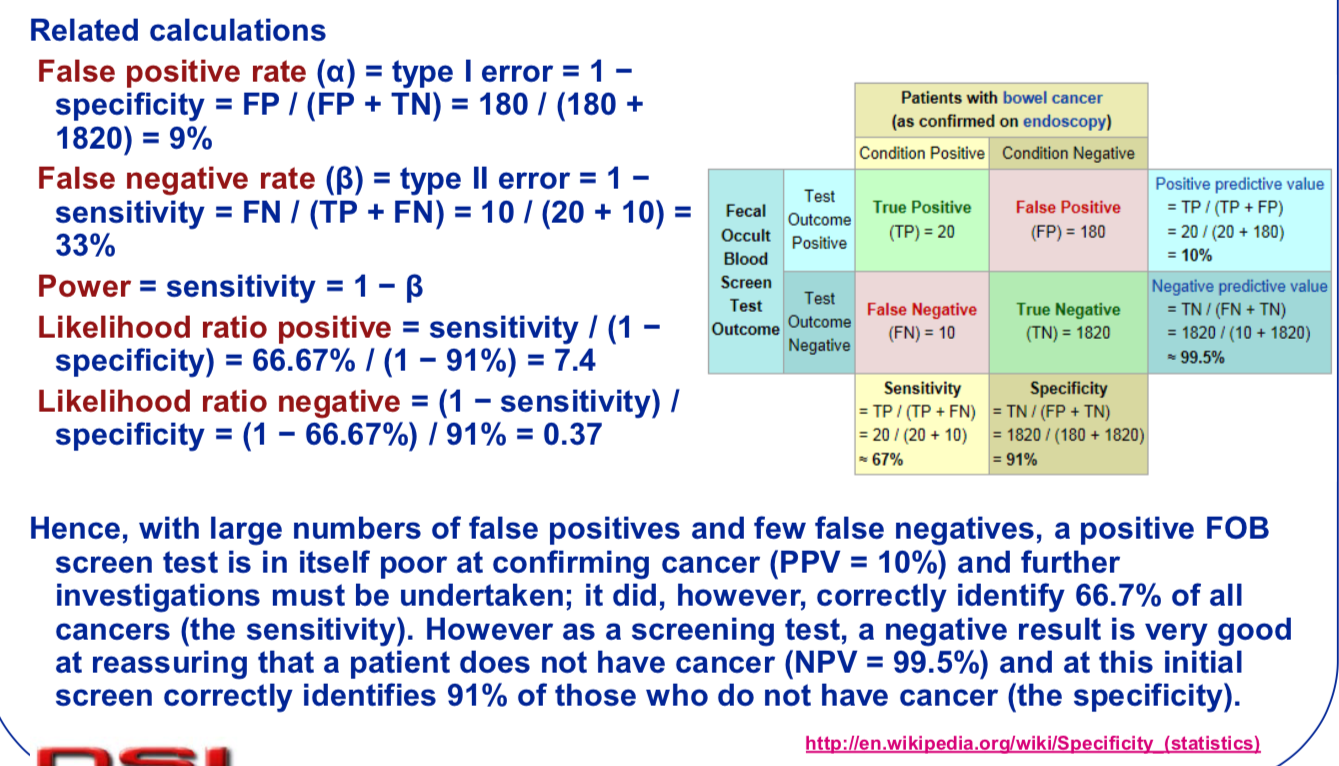
Confusion matrix: 混淆矩阵

**Sensitivity**: recall rate, measures the proportion of actual positives which are correctly identified

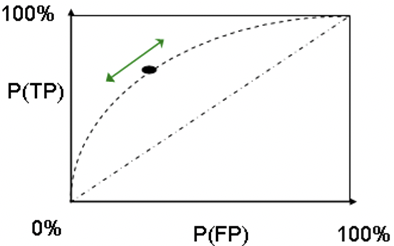
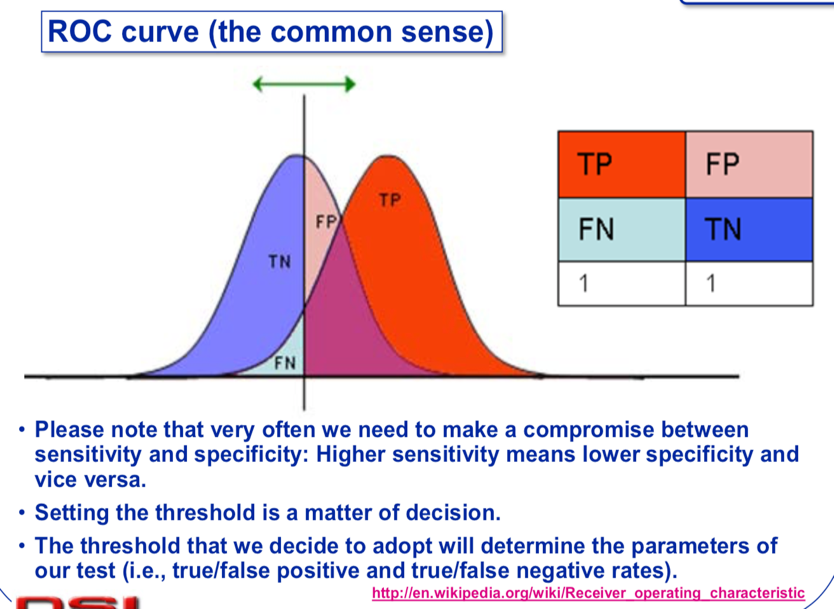
**Specificity:** measures the proportion of negatives which are correctly identified

A perfect predictor would be described as 100% senitivity and 100% specificity.

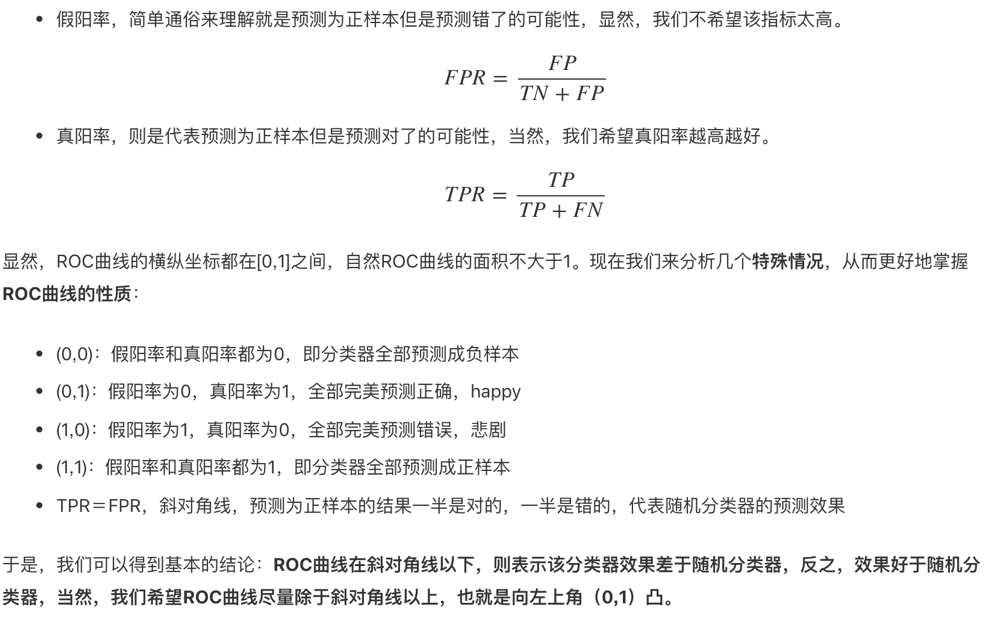




**ROC (Receiver Operating Characteristic) Curves**



对于某个**二分类**分类器来说，输出结果标签（0还是1）往往取决于输出的概率以及预定的概率阈值，比如常见的阈值就是0.5，大于0.5的认为是正样本，小于0.5的认为是负样本。如果**增大**这个阈值，预测错误（针对正样本而言，即指预测是正样本但是预测错误，下同）的概率就会降低但是随之而来的就是预测正确的概率也降低；如果**减小**这个阈值，那么预测正确的概率会升高但是同时预测错误的概率也会升高。实际上，这种阈值的选取也一定程度上反映了分类器的**分类能力**。我们当然希望无论选取多大的阈值，分类都能尽可能地正确，也就是希望该分类器的分类能力越强越好，一定程度上可以理解成一种**鲁棒能力**吧。

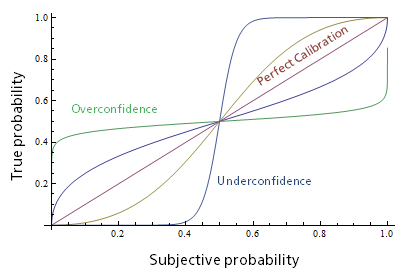


**AUC: Area Under the (ROC) Curve**

* AUC ＝ 1，代表完美分类器
* 0.5 < AUC < 1，优于随机分类器
* 0 < AUC < 0.5，差于随机分类器

**Calibration**

We plot the frequencies observed in the data (y axis) against the probabilities calculated by the system (x axis).



**Cross-validation:**

Testing a model on the same data that we used for training it does not seem fair.

It will favor most complex models that fit the data best.

Simpler model may actually fit future instances of data better than complex models.

**Cross-validation prevents over-fitting**

Cross-validation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set.

It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice.

Divide the data into two disjoint sets: (1) training set and (2) test set (a.k.a. validation set).

Perform the analysis on the training set and validate the results on the test set.

Simple and effective.

**Disad**: It wastes data that could have been used for learning.

**k-Fold cross-validation**

**“Leave-One-Out” cross-validation**

Uses effectively n-1 instances for training and tests the model on all n instances, one at a time (an extreme case of k-fold, k=n).

**Bootstrap cross-validation**

Repetitive test set method, each repetition involves selecting a new test set from among all records.