

Association Rules

X->Y	X associated to Y
Supp(X): Antecedent support	P(contains X)
Supp(X and Y): Rule Support	P(contains X and Y)
Conf(X->Y): Rule Confidence	P(contains Y contains X)
Left(X->Y): Rule Lift	P(contains Y contains X) / P(contains Y)

Confidence Interval

总体方差已知时

CI : $[(X_n - z \cdot \sigma / \sqrt{n}, X_n + z \cdot \sigma / \sqrt{n})]$
 X_n 为样本均值, σ 表示总体标准差, z 依据置信水平而定

$z = \text{qnorm}(0.9 - 0.1, 0, 1)$ 置信水平为90%时, z 的计算

总体方差未知

CI : $[(X_n - t \cdot S / \sqrt{n}, X_n + t \cdot S / \sqrt{n})]$
 用样本方差 S 代替 σ , 用 t 分布代替整体分布

$t_value = \text{qt}(0.975, df = n - 1)$ 置信水平为95%时, t 的计算

$t = (x - \mu) / (s / \sqrt{n})$ T统计量的计算

$t.test()$ 会得出T值, p 值, 这里的 p 值检测的是T统计量的极端性, 当 p 非常小时, 拒绝原假设

$t.test(\text{ChickWeight}\$weight[\text{ChickWeight}\$Diet == 1], \text{ChickWeight}\$weight[\text{ChickWeight}\$Diet == 2], \text{conf.level} = 0.9)$ 检验两个总体差值的置信区间

proportion test 比例检验用于比较两个二项分布 (二项分布表示成功和失败的次数, 或者说比例) 的比例是否相等或满足某种假设。比例检验通常用于判断两个不同组群的比例是否有显著差异。

Hypothesis test

type 1 error reject H_0 while it's true (often denoted by " α ")

type 2 error not reject H_0 even though it's false (denoted by " β ")

significance level maximum on Type-1 error probability

p-value minimum α needed to reject H_0

