单位点检测用于评估单个SNP的遗传效应，其主要目标是确定该SNP与复杂性状之间的潜在关联。通过分析SNP的基因型与目标性状的关系，可以揭示特定基因位点对性状的贡献。下表汇总了常用的单位点检测方法，说明了每种方法的主要优点及其局限性。

**表 主要的单位点检测方法**

**Table Major single-marker detection methods**

|  |  |  |
| --- | --- | --- |
| 方法 | 主要优点 | 局限性 |
| 卡方检验 | 广泛的适用性，不依赖参数假设 | 无法有效控制混杂因素 |
| Logistic回归 | 能控制协变量，可解释性好 | 计算复杂度高，参数估计的不稳定性 |
| Fisher精确检验 | 小样本数据精确性较高 | 仅适合用于分析2×2列联表 |
| Armitage趋势检验 | 较高的计算效率 | 局限于线性效应 |

卡方检验通过检测单个SNP基因型在不同群体中的频率差异，来评估SNP与目标性状之间的关联性。该方法通过比较观察到的基因型分布与期望分布之间的差异，以确定基因型频率是否显著偏离随机分布。其统计量基于列联表，并通过计算卡方统计值来判断关联性。其核心公式为：

其中O 是每个基因型类别的观察频数。Ei是每个基因型类别的期望频数。k 是基因型的类别数量（例如0、1、2）。卡方检验由于计算速度快且不依赖参数假设，因而具有较广泛的适用性。然而，该方法无法有效控制潜在的混杂因素，因此分析结果容易受到种群结构的影响。

Fisher精确检验可用于评估SNP与复杂性状之间的关联关系，该检验通过构建SNP与性状的2×2列联表来进行计算和分析，旨在通过精确的概率计算检验变量之间的统计独立性。其核心公式为：

其中a, b, c, d 是列联表中的观察频数，n是样本总数。Fisher精确检验不依赖于正态分布的假设，因而在处理小样本数据时具有较高的精确性。该方法能够提供精确的概率计算结果，适合用于分析2×2列联表中变量之间的关联。然而，由于其较高的计算复杂度以及对2×2列联表的局限性，Fisher精确检验在处理更复杂的数据结构时的适用性受到一定限制。

Logistic回归是常用的单位点检测方法，该方法通过最大化对数似然函数来计算每个SNP的效应，以评估SNP与性状之间的关联关系，其核心公式为：

其中是表型发生的概率，是截距，代表没有SNP效应时的基础风险，**β1​** 是SNP的回归系数，表示SNP对目标性状的影响。**X1**是SNP的基因型值，通常编码为0、1、2，分别表示不同等位基因组合。Logistic回归具有良好的可解释性，并且通过引入多个协变量，可以有效提高模型的预测准确性。然而，其计算成本较高，尤其在样本量较小或数据特征较为复杂的情况下，可能会导致参数估计的不稳定性。

Armitage趋势检验用于评估SNP基因型与目标性状之间是否存在显著的线性趋势。该检验通过构建2×3列联表进行分析，其检验统计量依据不同基因型的频率和目标性状的分布，旨在检测基因型与性状之间的线性关联。该方法的主要公式为：

其中ni表示每个基因型（0、1、2）的病例数，是期望的病例数。是权重，通常为0、1、2（分别表示基因型的递增）。p 是病例组中事件发生的总体比例。由于其较高的计算效率，Armitage趋势检验被广泛应用于GWAS的初步筛选阶段。然而，该检验的适用范围主要局限于检测基因型与性状之间的线性效应，因此在分析稀有等位基因或复杂的非线性关联时表现不足。

全基因组关联研究（GWAS）中的单位点检测存在以下主要问题：

（1）多重检验问题：单位点检测涉及成千上万的SNP，随着检验次数的增加，假阳性结果的风险显著增加。为了解决这一问题，通常采用严格的多重检验校正方法（如Bonferroni校正或FDR控制）。然而，过于严格的校正方法会显著降低检验功效，增加假阴性结果的风险，导致潜在的真实关联被忽略。

（2）稀有等位基因的分析困难：单位点检测中大部分统计检验方法对常见等位基因的表现较好，但在分析稀有等位基因时，往往由于样本量不足，难以获得有力的统计支持。

（3）忽视SNPs间的相互作用：单位点检测假设每个SNP独立地影响目标性状，未能充分考虑SNP之间的相互作用（上位性）。已有研究表明这种忽略可能导致对复杂性状的解释力不足，进而引发遗传力缺失（Missing Heritability）的问题。