**GOplan 使用手册**

GOplan 是一个用于制定动物育种方案的R 软件包，包括核心种群的育种计划和整个杂交繁育体系的育种策略。此外，它还能实现繁育体系的优化。

该软件包有三个主要功能。参数 “prm\_path ”是需要输入的参数文件的完整路径：

runCore(prm\_path)是用于评估核种群各种育种方案的函数。可以比较不同核心群结构下的遗传进展和经济效益，如种畜使用年限、公母比例、交配方法、母本数、淘汰率、育种值估算方法、表型或基因分型选择的个体数等。该函数共返回五个方面的结果，即实现的遗传进展、性状的最终群体平均值、遗传变异的下降程度、近交增量和经济利润。该函数支持使用常见的遗传评估方法，包括 BLUP、GBLUP 和 ssGBLUP（单步 GBLUP），均通过调用 DMU 软件执行。此外，为了提高软件的灵活性，用户还可以通过指定路径调用外部选配相关程序。

runWhole(prm\_path)旨在通过内置繁育体系框架实现轻松评估杂交繁育体系育种效果的目的。它可以评估二元、三元以及四元杂交繁育体系，并通过返回终端杂种的表型预测值和经济收益来评估不同种杂交繁育体系方案的有效性。

runOpt(prm\_path)用于寻找能使杂交繁育体系获得最大经济效益的最佳参数值组合。利用贝叶斯优化方法，指定待优化参数参数的变化范围，在固定的迭代时间内得出一个最优解。

**参数文件说明：**

育种方案中大部分参数的指定都可以通过编辑参数文件来修改。值得注意的是，文件中不需要的参数应以 0 指定。除了空格外，不要使用其他分隔符分隔不同的数字，否则可能会导致未知错误。不要更改每个参数的顺序。

参数文件包含六个部分，每个部分的说明如下：

**### Analyse Parameters ###**

分析参数。这部分用于确定运行所有三个函数时的一些基本信息。

|  |  |  |
| --- | --- | --- |
| **参数** | **值** | **说明或参数的可选值** |
| QUICK | FALSE | #FALSE, TRUE |
| Method | blup | #blup, ssgblup, gblup |
| EarlySelect | FALSE | #FALSE, TRUE |
| Lit\_pro | 0 | #终选前窝选的比例（即选排名前%的窝）, 范围为0 - 1 |
| Mate | rand | #选配方法: rand, MC 或是自定义选配程序的路径 |
| maxF | 999 | #父母对允许的最大亲缘系数 |
| Ctype | 33 | #繁育体系类型: 1, 21, 22, 31, 32, 33, 41, 42 |
| Nrep | 50 | #重复次数 |
| Time | 20 | #育种规划期长度 |
| Ncores | 10 | #可用线程数 |
| nChr | 18 | #染色体数目 |
| nSnpPerChr | 3000 | #每条染色体上SNP数 |
| nQtlPerChr | 100 | #每条染色体上QTL数 |
| out\_path |  | #结果输出路径（必须以“/”结尾） |

**QUICK：**设置为 “ＴＲＵＥ ”时，程序将用模拟代替育种值估计过程。程序将只计算第一代的估计育种值（EBV）（用 DMU 估计育种值，然后用 EBV 和 TBV 的相关性作为育种值估计的准确性），并根据估计的准确性用 Cholesky 方法模拟其他各代的育种值。我们假设TBV和EBV ~ N (,),，因此模拟 EBV 的计算方式为：。

其中，为所有个体的模拟 EBV 向量；为所有个体 TBV 的平均值；为准确度（TBV 与 EBV 之间的皮尔逊相关系数）；为真实育种值向量；为随机生成的向量，遵循正态分布 ~ N (0,1)；为 TBV 的标准偏差。

QUICK 选项可以节省时间并快速得到结果，如果您只想定性地了解哪个育种方案更好，而不关心精确的利润，尤其是当方法设置为 “gblup ”或 “ssgblup ”时，我们建议将 QUICK 设为 “TRUE”。

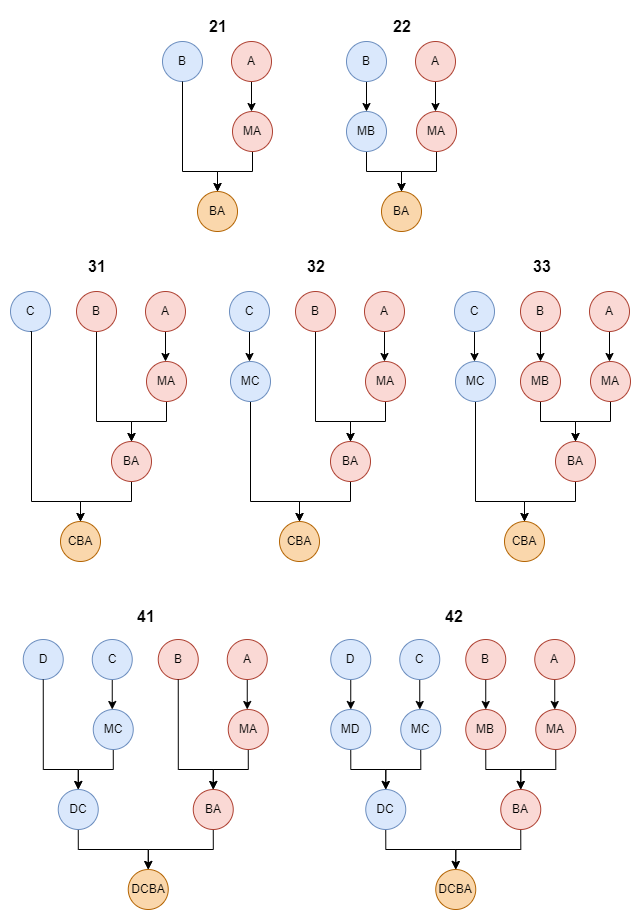
**Method:** 目前软件共支持三种育种值估计方法，并使用DMU进行估计。blup：最佳无偏预测，仅使用血统信息；ssgblup：单步gblup，同时使用基因组和血统信息；gblup：基因组blup，仅使用基因组信息。

**EarlySelect：**早期选择，是否对幼仔进行预选，默认为FALSE。如果为 “true”，GOplan 将根据窝指数（即父亲的育种值和母亲的育种值的平均值）选择一些窝指数高的窝。

**Mate**：有四个选项供用户选择：rand（随机交配）、HOMO（同质选配）、HETER（异质同配）、MC（最小共祖系数交配，使用退火算法求解）；或者您也可以指定自己的选配程序的完整路径，软件包会写出一个名为 “plans.txt ”的文件供自定义的选配程序使用，其中包含四列内容：母系 ID、父系 ID、亲缘系数和后代的预期育种值，要求返回的文件为（母畜-公畜）两列的配对文件“reMate.txt”。

**maxF：**父母间的最大亲缘系数。亲缘系数大于此值的公母畜个体不会交配。当设置为 999 时，为随机选配。

**Ctype：**杂交类型，详见图 1，1 表示单一种群，2\* 表示二元杂交（B 为父系，A 为母系），3\* 表示三元杂交（C 为父系，B 为母系的父系，A 为母系的母系），4\* 表示四元杂交（D 为父系的父系，C 为母系的父系，B 为母系的父系，A 为母系的母系）。

图 1：不同杂交繁育体系类型简图

**Ncores：**线程数。我们的软件包支持多线程运行。但请确保 Ncores 不超过系统可用线程的最大值。

**nChr：**模拟染色体数。更改该参数以适应不同的品种。

**### VARIABLES ###**

变量的设置，这一部分可以指定或忽略。

变量数量没有限制，但变量越多，组合越多，运行时间越长。我们建议每次设置少于三个变量，一个变量一行，每个变量列出的所有水平用空格隔开。

runCore(prm\_path) 函数可设置变量： Ys、Yd、Sor、nfam、nfam\_F、nfam\_M、ngeno、nGeno\_F、nGeno\_M。使用 nfam 或 ngeno 表示同时调整 nfam\_F 和 nfam\_M 或 nGeno\_F 和 nGen\_M。

runWhole(prm\_path) 函数可设置变量： Asec、Bsec、Csec、Dsec、YABd、YCDs、SorAB、SorCD、SorP。

**### Optimize Parameters ###**

该部分用于设置各优化参数的范围。如果不需要某些参数，只需删除其中的一行即可。使用 runCore(prm\_path) 和 runWhole(prm\_path) 时，请忽略此部分。

|  |  |  |  |
| --- | --- | --- | --- |
| **参数** | **值** | **值** | **说明** |
| Pars | lower | upper |  |
| Asec | 2 | 6 | #扩繁群A的使用年限 |
| Bsec | 1 | 4 | #扩繁群 B的使用年限 |
| Csec | 2 | 6 | #扩繁群 C的使用年限 |
| Dsec | 1 | 4 | #扩繁群 D的使用年限 |
| YABd | 2 | 6 | #二元杂种母畜AB的使用年限 |
| YCDs | 1 | 4 | #二元杂种公畜CD的使用年限 |
| NAd | 600 | 3000 | #核心群A的母畜数目 |
| NBd | 300 | 1000 | #核心群 B的母畜数目 |
| NCd | 300 | 1000 | #核心群C的母畜数目 |
| NDd | 300 | 1000 | #核心群D的母畜数目 |

**### Population Structure ###**

这部分用于设置育种方案中的基础种群结构。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **参数** | **值** | **值** | **值** | **值** | **说明** |
| dam0 | 2 |  |  |  | #第一个后代出生时母畜的年龄 |
| sire0 | 2 |  |  |  | #第一个后代出生时公畜的年龄 |
| YABd | 6 |  |  |  | #二元杂种母畜AB的使用年限 |
| cull\_AB | 0 | 0 | 0 |  | #二元杂种母畜AB各年龄的淘汰比例 |
| YCDs | 4 |  |  |  | #二元杂种公畜CD的使用年限 |
| cull\_CD | 0 | 0 | 0 |  | #二元杂种公畜CD各年龄的淘汰比例 |
| SorAB | 100 |  |  |  | #AB 杂交时公母比例 |
| SorCD | 100 |  |  |  | # CD 杂交时公母比例 |
| SorP | 100 |  |  |  | #终端父母本杂交时公母比例 |
| ABSR | 0.85 |  |  |  | # AB种畜的存活率 |
| CDSR | 0.85 |  |  |  | #CD种畜的存活率 |
| PSR | 0.85 |  |  |  | #商品畜的存活率 |
| N\_P | 0 |  |  |  | #期望生产的商品畜数目（runOpt时需要） |
| nPoints | 0 |  |  |  | #每次迭代时选择的采样点数（runOpt时需要） |
| nIter | 0 |  |  |  | #迭代次数（runOpt时需要） |
|  | | | | | |
| Breed | A | B | C | D |  |
| n\_female | 500 | 500 | 500 | 0 | #母畜数目 |
| Sor | 50 | 50 | 50 | 0 | #公母比例 |
| Ys | 2 | 2 | 2 | 0 | #公畜使用年限 |
| Yd | 4 | 4 | 4 | 0 | #母畜使用年限 |
| n\_progeny | 12 | 12 | 12 | 0 | #窝产仔数 |
| SR | 0.85 | 0.85 | 0.85 | 0 | #后代存活率 |
| nSec | 3000 | 0 | 0 | 0 | #扩繁群规模 |
| Ysec | 6 | 2 | 2 | 0 | #扩繁群使用年限 |
| nPedG | 3 | 0 | 0 | 0 | #追溯的系谱代数 |
| nGenoG | 3 | 0 | 0 | 0 | #上溯使用的基因型代数 |
| nSel\_F | 2 | 0 | 0 | 0 | #终选时窝选母畜数目 |
| nSel\_M | 2 | 0 | 0 | 0 | #终选时窝选公畜数目 |
| nfam\_F | 2 | 2 | 2 | 0 | #每窝表型测定母畜数目 |
| nfam\_M | 2 | 2 | 2 | 0 | #每窝表型测定公畜数目 |
| nGeno\_F | 0 | 0 | 0 | 0 | #每窝基因型测定母畜数目 |
| nGeno\_M | 0 | 0 | 0 | 0 | #每窝基因型测定公畜数目 |

**### Breeds Details ###**

本部分用于指定每个品种的基本信息。您必须列出所有品种的信息。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **参数** | **值** | **值** | **值** | **值** | **值** | **说明** |
| Breedname | A |  |  |  |  |  |
| cull\_s | 0 |  |  |  |  | #公畜各年龄的淘汰比例 |
| cull\_d | 0 | 0 | 0 |  |  | #母畜各年龄的淘汰比例 |
| cull\_sec | 0 | 0 | 0 | 0 | 0 | #扩繁群各年龄的淘汰比例 |
| trait | JZRL |  |  |  |  | #性状名称 |
| SexLimit | 0 |  |  |  |  | #0: 非限性性状, 1: 限性性状（公）, 2: 限性性状（母） |
| mean | 180 |  |  |  |  | #性状的群体均值 |
| var | 100 |  |  |  |  | #性状的表型方差 |
| heri | 0.3 |  |  |  |  | #性状的遗传力 |
| weigh | -1 |  |  |  |  | #性状的的权重 |
| analyse | 1 |  |  |  |  | #分析方法 |
| ## phenotype covariance ## | | | | | | |
| JZRL | 1 |  |  |  |  |  |
|  | | | | | | |
| Breedname | B |  |  |  |  |  |
| cull\_s | 0 |  |  |  |  |  |
| cull\_d | 0 | 0 | 0 |  |  |  |
| cull\_sec | 0 | 0 | 0 | 0 | 0 |  |
| trait | JZRL |  |  |  |  |  |
| SexLimit | 0 |  |  |  |  |  |
| mean | 170 |  |  |  |  |  |
| var | 100 |  |  |  |  |  |
| heri | 0.3 |  |  |  |  |  |
| weigh | -1 |  |  |  |  |  |
| analyse | 1 |  |  |  |  |  |
| ## phenotype covariance ## | | | | | | |
| JZRL | 1 |  |  |  |  |  |
|  | | | | | | |
| Breedname | C |  |  |  |  |  |
| cull\_s | 0 |  |  |  |  |  |
| cull\_d | 0 | 0 | 0 |  |  |  |
| cull\_sec | 0 | 0 | 0 | 0 | 0 |  |
| trait | JZRL |  |  |  |  |  |
| SexLimit | 0 |  |  |  |  |  |
| mean | 160 |  |  |  |  |  |
| var | 100 |  |  |  |  |  |
| heri | 0.3 |  |  |  |  |  |
| weigh | -1 |  |  |  |  |  |
| analyse | 1 |  |  |  |  |  |
| ## phenotype covariance ## | | | | | | |
| JZRL | 1 |  |  |  |  |  |

**analyse:** “1”表示估计育种值时单性状估计，其他相同数字表示多个性状联合估计。例如，如果有三个性状 T1、T2 和 T3，T1 用单性状估计，而 T2 和 T3 则需要用双性状模型估计。那么我们设置**analyse**值为：1，2，2。

## phenotype covariance ##:：这一部分定义了性状间的表型协方差。例如，当有三个性状 T1、T2 和 T3 时。则定义如下：

|  |  |  |  |
| --- | --- | --- | --- |
| T1 | 1 | 0 | 0 |
| T2 | 0 | 1 | 0 |
| T3 | 0 | 0 | 1 |

注意：默认列名与行名相同，但我们省略了列名。

所有定义淘汰率的参数的数目应始终比其相应的使用年限少1。例如，如果品种 C 的 Yd 为 4，那么 C 的 cull\_d 必须有三个值（0、0、0）。因为在第一个使用年限时没有淘汰，而当父母代超过其寿命时，它们将被全部淘汰。

**### Basic** **Economic Parameters ###**

这一部分定义了育种方案中涉及的经济参数。

|  |  |  |  |
| --- | --- | --- | --- |
| **参数** | **值** | **值** | **说明** |
| trait | 0 | JZRL |  |
| jb\_cost | 906 | 1.5 | #个体的基本成本 |
| dam\_cost | 1500 | 0 | #每个母畜所需成本 |
| sire\_cost | 1000 | 0 | #每个公畜所需成本 |
| other\_cost | 10 | 0 | #每繁殖季节的其它成本 |
| ind\_sale | 2040 | 0 | #商品畜的售卖收入/头 |
| cull\_sale | 1500 | 0 | #淘汰种畜的收入/头 |
| sire\_sale | 3000 | 0 | #售卖种公畜的收入/头 |
| dam\_sale | 2000 | 0 | #售卖种母畜的收入/头 |
| meas\_cost | 20 |  | #表型测定成本/头 |
| geno\_cost | 200 |  | #基因型测定成本/头 |

trait：第一列固定为 0，表示所有经济参数的初始值。然后列出相关性状，每个性状对应一列，表示每个性状变化一个单位时的货币变化量。该值可以为负。

sire\_sale：售卖种公畜的收入/头，已测定但未被选的公畜可作为种公畜出售。

dam\_sale：售卖种母畜的收入/头，已测定但未被选的母畜可作为种母畜出售。

**安装：install.packages("GOplan\_0.1.0.tar.gz",repos = NULL)**

**运行示例 & 输出文件**

* **示例1：**

***runCore("example\_prm/prm\_Core.txt")***

***runCore()*** 将返回三个文件：coreOut.xlsx、coreOut\_detail.xlsx 和 detailInfo.txt。

coreOut.xlsx 有以下 9 列，是在育种规划期结束后的结果。

|  |  |
| --- | --- |
| mean\_G / sd\_G | 相对遗传进展的平均值和标准误差 |
| mean\_Inb / sd\_Inb | 近交增量的平均值和标准误差 |
| Profit | 经济效益 |
| meanPhenotype / sdPhenotype | 预测表型均值的平均值和标准误差 |
| Vg\_meanPhenotype / Vg\_sdPhenotype | 群体遗传多样性减少的平均值和标准误差 |
| var | 变量组合的名字 |

coreOut\_detail.xlsx 记录了每个繁殖季节的结果，每种信息都保存在单独的表单中。

|  |  |
| --- | --- |
| varg | 每个性状的遗传方差 |
| acc | 每个性状的估计育种值准确性均值 |
| acc\_index | 综合选择指数的选择准确性 |
| index | 群体的综合选择指数均值 |
| dfInb | 群体近交系数均值 |
| pop0\_G | 预估的群体表型均值 |

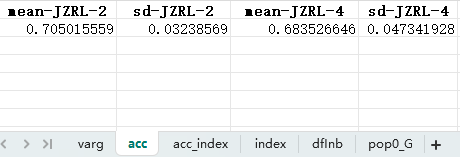
detailInfo.txt 显示了不同变量组合下的种群结构和育种方案细节信息。

在此，我们举例说明如何比较以Yd为变量下核心群的育种效果。参数文件为 example\_prm 文件夹中的 prm\_Core.txt： ，结果如下：

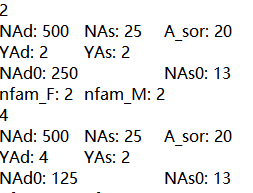
coreOut.xlsx ：

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **mean\_G** | **sd\_G** | **mean\_Inb** | **sd\_Inb** | **Profit** | **meanJZRL** | **sdJZRL** | **Vg\_meanJZRL** | **Vg\_sdJZRL** | **var** |
| 0.12 | 0.01 | 0.03 | 0.01 | 5020030 | 158.40 | 1.11 | -8.55 | -1.76 | 2 |
| 0.11 | 0.01 | 0.02 | 0.00 | 5075439 | 160.42 | 1.64 | -5.84 | -1.54 | 4 |

**coreOut\_detail.xlsx :**



detailInfo.txt:



* **示例2：**

runWhole() 将输出四个结果文件。

WholeOut.xlsx包含不同育种方案下最终杂交个体的预测表型和经济收益。

\*out.xlsx 包含每个繁殖季节不同亚群的表型信息。

detailInfo.txt与之前描述的相同。

p.csv 是基因流动方法中使用的 P 矩阵。

在此，我们给出一个以 SorP 为变量进行杂交育种方案比较的示例。参数文件是 example\_prm 文件夹中的 prm\_Whole.txt：

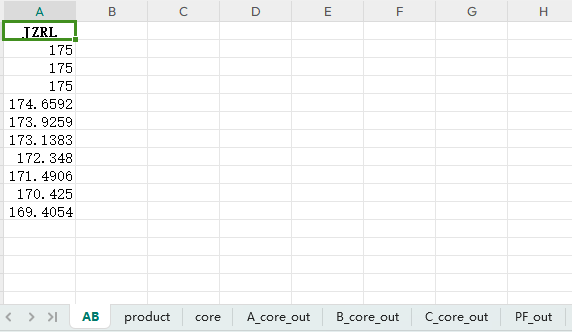
***runWhole(“example\_prm/prm\_Whole.txt”)***

结果如下：

**WholeOut.xlsx:**



**\*out.xlsx:**



* **示例3：**

***res = runOpt(prm\_path = "prm\_Opt.txt")***

***getOptRes(res = res, out\_path = out\_path) #将结果输出到csv文件中***

runOpt() 的结果保存在输出日志文件中，我们提供了一个函数 getOptRes(log\_name, out\_path)，用于从日志中提取结果。 log\_name 是日志的全名，out\_path 是写出结果的路径。该函数将返回一个名为 mboRes.csv 的文件。

**附录1：利用退火算法实现最小共祖交配（MC）**

1. 随机生成初始交配方案，并计算总体近交系数： E0，设 T=1.0，Nre=0，Neva=0，其中 T 为退火温度，Nre 为计划的替换次数，Neva 为评估次数；
2. 通过随机选择两对公母对并交换其母本来调整计划，新计划的 Ei 等于 E0+δ，其中，，是公畜1和母畜1的亲缘系数（图 2）；

c)如果δ<0，则接受新计划，或以概率 接受新计划。

d)当计划已被替换 10\*nmax 次或已调整计划 100\*nmax 次（nmax 为母本数和父本数的最大值）时，更新 T 值，每次减少 10%；

e)如果最后一次和倒数第二次改变 T 之间结果没有改善，则停止，或重复上述步骤。

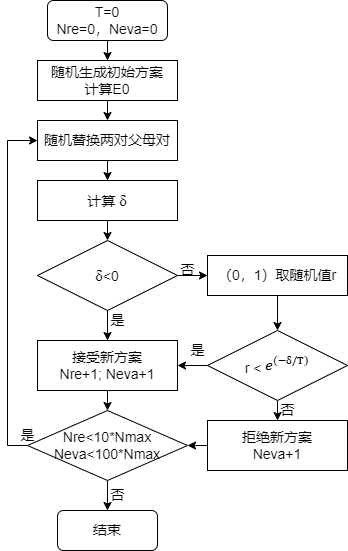
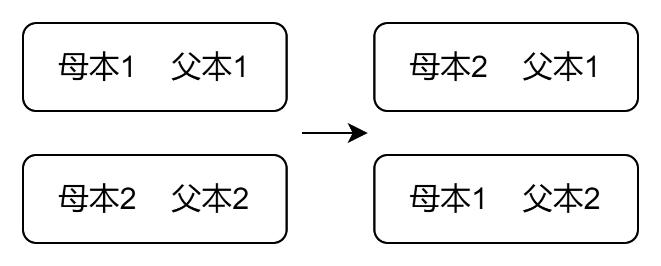


图 2 父母对交换示意图

图 1 MC选配算法逻辑示意图

**附录2：经济效益的计算**

本软件包中，我们分别计算每个季节的利润，不考虑贴现。

主要步骤：

1. 计算第 i 繁殖季节的表型进展，并计算更新每个经济参数的值；
2. 计算不同类型个体的数量：

|  |  |
| --- | --- |
| 个体类型 | 计算方法 |
| 断奶个体数目 | 母畜总数\*窝产仔数 |
| 淘汰数目 | 所有淘汰的个体数目 |
| 表型测定数目 | 所有表型测定的个体数目 |
| 基因型测定数目 | 所有基因型测定的个体数目 |
| 售卖的母畜数目 | 测定母畜数 - 淘汰的母畜数目 |
| 售卖的公畜数目 | 测定公畜数 - 淘汰的公畜数目 |
| 商品畜（仅只终端群体数目） | 终端母本的数目 \*窝产仔数\* 存活率 |
| 商品总数 | 断奶数目 \*存活率– max(淘汰数目, (测定数目 + 淘汰终端父母本的数目)) |

1. 计算更新各经济参数:

|  |  |
| --- | --- |
| 经济参数 | 计算方法 |
| 所有基本成本 | 断奶数目 \*基本成本 |
| 所有公畜成本 | 公畜数目\* 公畜成本 |
| 所有母畜成本 | 母畜数目\* 母畜成本 |
| 所有表型测定成本 | 表型测定个体数\* 表型测定成本 |
| 所有基因分型成本 | 基因分型个体数\* 基因分型成本 |
| 销售产品收入 | 产品数目\*产品售卖收入 |
| 销售种母畜收入 | 售卖母畜数目 \*母畜售卖收入 |
| 销售种公畜收入 | 售卖公畜数目 \*公畜售卖收入 |
| 淘汰个体收入 | 淘汰个体数目\*个体淘汰收入 |

1. 利润 = （销售产品收入 + 销售种公畜收入 + 销售种母畜收入 + 淘汰个体收入）- （所有基本成本 + 所有公畜成本 + 所有母畜成本 + 所有表型测定成本 + 所有基因分型成本 + 其他成本）