**多重插补方法介绍**

**1、简介**

多重插补（MI）是一种基于重复模拟的处理缺失值的方法。在面对复杂的缺失值问题时，MI是最常选用的方法，它将从一个包含缺失值的数据框中生成n组插补后的完整数据集（n通常是3到10），每一组插补后的数据集为一种插补可能，n组数据集即为n个插补可能。每个插补后的数据集中，缺失数据将用蒙特卡洛方法来填补。之后，with（）函数利用统计模型，对每个插补后的数据集进行统计分析，得到n组统计结果。n组统计结果通过组合函数pool()输出最终结果。利用summary()函数查看针对最终结果的评价指标，通过这些评价直接评估插值拟合后的插值质量。本课题分别以前12小时和后12小时的有创收缩压、舒张压、平均压插值填补前12小时和后12小时的无创收缩压、舒张压、平均压。

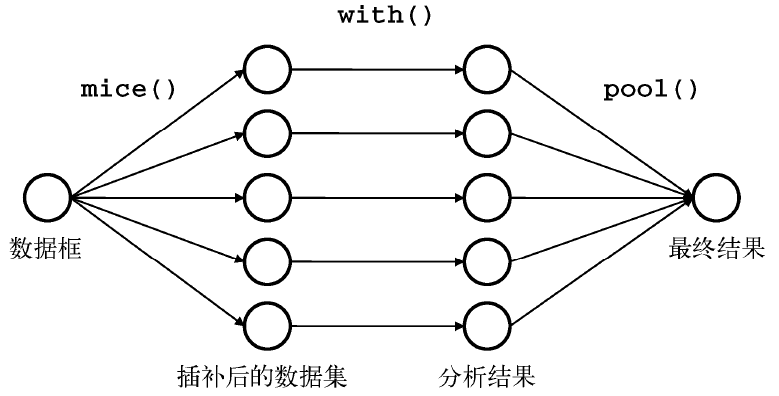


图1 多重插补步骤

**2、mice（）插补返回多个完整数据集及相关信息**

imp<-mice(data,m=5,method='norm.predict')

在mice函数中输入包含缺失数据的数据框data，该函数的输入参数包括插补后的完整数据集个数m，插补方法method。每个完整数据集都是通过对原始数据框中的缺失数据进行插补而生成的，每一个插补后的完整数据集代表一种插补可能性。由于插补有随机的成分，因此每个完整数据集都略有不同。插补方法如图2所示，返回imp信息如图3所示，插补的数据集在图3红框中标出。

在本例中，data为前12小时的有创收缩压和无创收缩压，m为2（因为经过试验，m>2时插补后的若干完整数据集差别不大，因此为了简化计算量，这里的m=2），方法为norm.predict（一种具有预测功能的线性回归插值方法）。

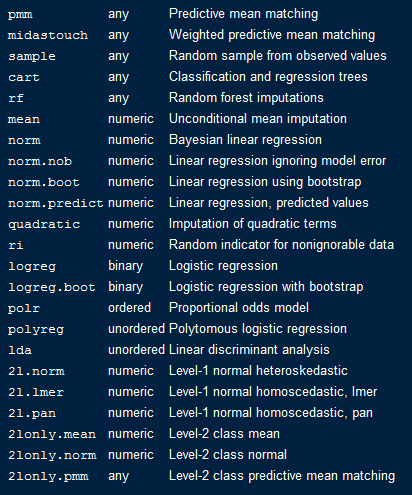


图2 插补方法汇总

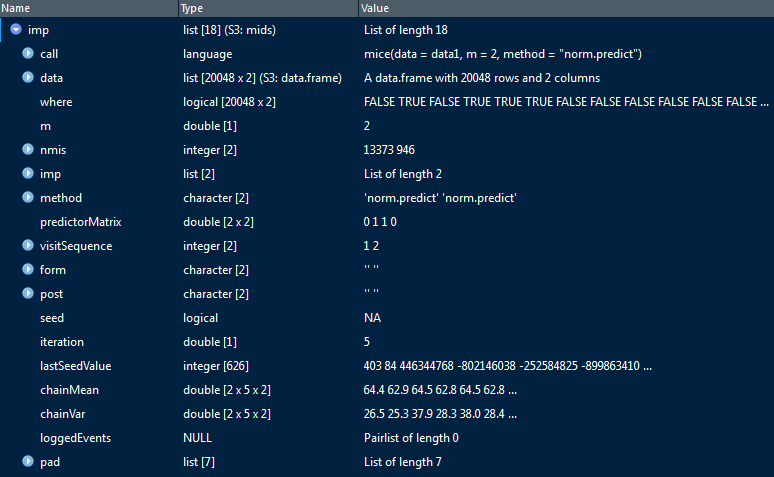


图3 mice（）插补结果

**3、with（）返回包含单独统计分析结果的列表对象**

fit<-with(imp,lm(qnimeanbp\_min~qmeanbp\_min))

利用with()函数依次对每个插值后的数据集应用统计模型，进行统计分析。这里的统计模型包括：线性模型lm()、广义线性模型glm()等模型。这里我们使用的是线性模型。Imp是上一步返回值，qnimeanbp\_min是前12小时无创收缩压，qmeanbp\_min是前12小时有创收缩压。输出结果如图4所示，相应分析的信息包含在fit中，在analyses中查看两个数据集的相关信息（[[1]]和[[2]]）。

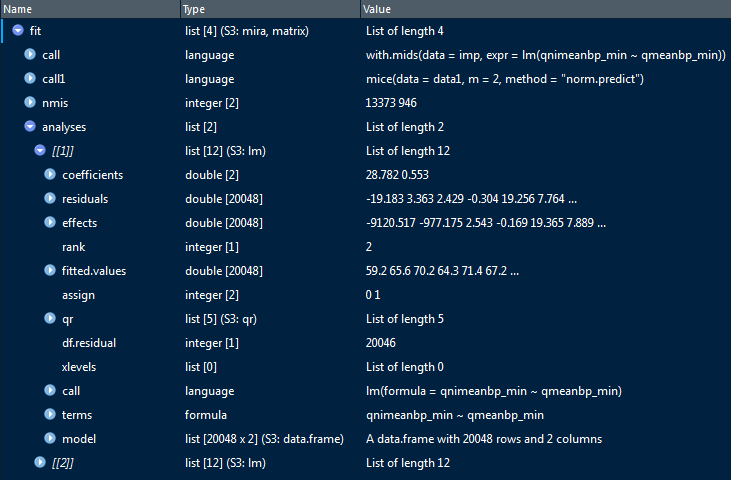


图4 with（）函数结果

**4、pool（）返回统计分析平均结果的列表对象**

pooled<-pool(fit)

利用pool()函数将n组单独的分析结果fit,整合为一组结果。检验整个方法是否合格。输出结果如图5所示。fmi即由于引入了缺失数据而引起的变异所占整体不确定性的比例，详细统计学信息由下一步得到。

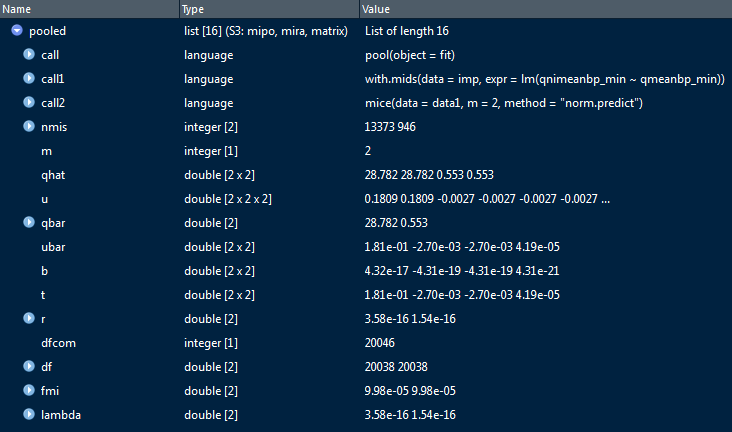


图5 pool（）函数结果

**5、评价指标**

summary(pooled)

该函数将整合后的统计结果进行综合分析，得到最终的评价指标。这些评价指标具有代表意义的为: Pr与fmi。查看回归模型系数t检验是否显著，Pr<0.05即显著，同时查看fmi,该值越小越好。图6是最终评价指标值。从图中可见，本例中，Pr为0，fmi为,9.978052e-05。



图6 summary（）函数结果