



图 6-6 残差诊断图

6.2 分层线性回归

通常一个数据集里包含研究中收集到的多个变量。探索两个变量在第三个分类变量的不同类别之间的关系是很有意义的，这本质上是对回归方程进行比较的问题。这里首先要检验两条（或多条）回归直线是否平行；若平行，再检验其截距是否相等。在 6.1 节的分析中我们把所有儿童看作一个整体，并没有考虑儿童是否患病。实际上，在数据集里还有一个因子型变量 `group`，用于区分正常儿童和大骨节病患病儿童。

在 6.1 节我们已经建立一个简单线性回归模型 `mod`，拟合了一条回归直线。模型的结果表明儿童的年龄对尿肌酐含量有显著影响。接下来探索不同类型儿童的年龄对尿肌酐含量的影响是否有显著差异。

如果假定年龄对正常儿童和患病儿童尿肌酐含量的影响是一致的，那么可以通过在模型里加入变量 `group` 拟合两条平行的回归直线。

```
> mod1 <- lm(ucr ~ age + group, data = UCR)
> summary(mod1)
Call:
lm(formula = ucr ~ age + group, data = UCR)
Residuals:
    Min       1Q   Median       3Q      Max
-0.29885 -0.15905  0.01675  0.14186  0.34023
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
            -----
(Intercept)  0.01675    0.01675    1.000  0.317
age          0.14186    0.01419   10.000 <0.0001
group1      -0.29885    0.02989  -10.000 <0.0001
```