

## 应用回归分析第 10 章

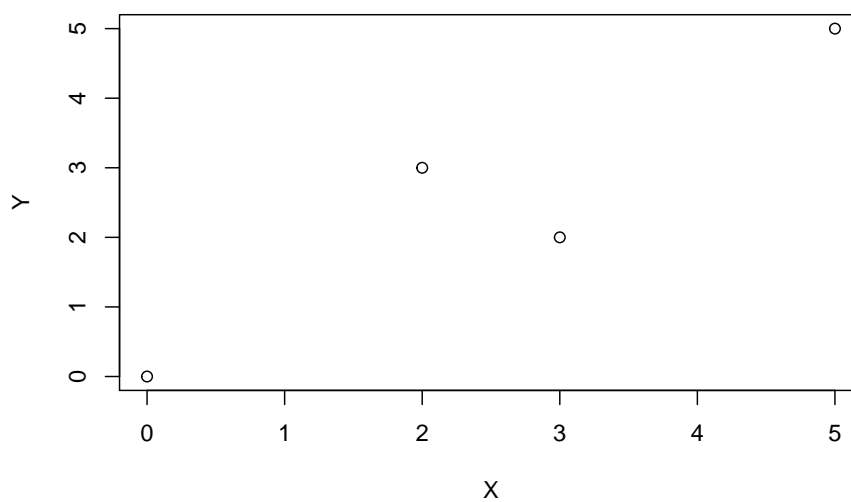
邵智轩

1400012141

物理学院

### 习题

```
DATA<-data.frame('X'=c(0,5,2,3,2,NA),  
                  'Y'=c(0,5,3,2,NA,2))  
plot(DATA)
```



删除不完全观测的案例后进行拟合

```
summary(lm.CC<-lm(Y~X,data = DATA,na.action = na.omit))
```

```
##
```

```
## Call:
## lm(formula = Y ~ X, data = DATA, na.action = na.omit)
##
## Residuals:
##      1      2      3      4
## -0.1923  0.1923  0.9615 -0.9615
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.1923     0.8382   0.229   0.8399
## X              0.9231     0.2720   3.394   0.0769 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9806 on 2 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.8521, Adjusted R-squared:  0.7781
## F-statistic: 11.52 on 1 and 2 DF, p-value: 0.07692
```

根据类似案例填补遗失数据方法填补遗失数据后进行拟合

```
DATA.sameImputed<-data.frame('X'=c(0,5,2,3,2,3),
                             'Y'=c(0,5,3,2,3,2))
lm.sameImputed<-lm(Y~X,data = DATA.sameImputed)
lm.sameImputed$df.residual<-lm.sameImputed$df.residual-2# 须扣除填补数据的自由度
summary(lm.sameImputed)
```

```
## Warning in summary.lm(lm.sameImputed): residual degrees of freedom in
## object suggest this is not an "lm" fit
```

```
##
## Call:
## lm(formula = Y ~ X, data = DATA.sameImputed)
##
```

```
## Residuals:
##      1      2      3      4      5      6
## -0.3704  0.3704  0.9259 -0.9259  0.9259 -0.9259
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.3704     1.0798   0.343   0.764
## X             0.8519     0.3704   2.300   0.148
##
## Residual standard error: 1.361 on 2 degrees of freedom
## Multiple R-squared:  0.7257, Adjusted R-squared:  0.3141
## F-statistic:  5.29 on 1 and 2 DF,  p-value: 0.1481
```

根据已观测数据建立回归模型填补数据后进行拟合

用完全案例 (complete cases) 的数据, 用  $Y \sim X$  的 OLS 填补  $Y$  的缺失值, 用  $X \sim Y$  的 OLS 填补  $X$  的缺失值。

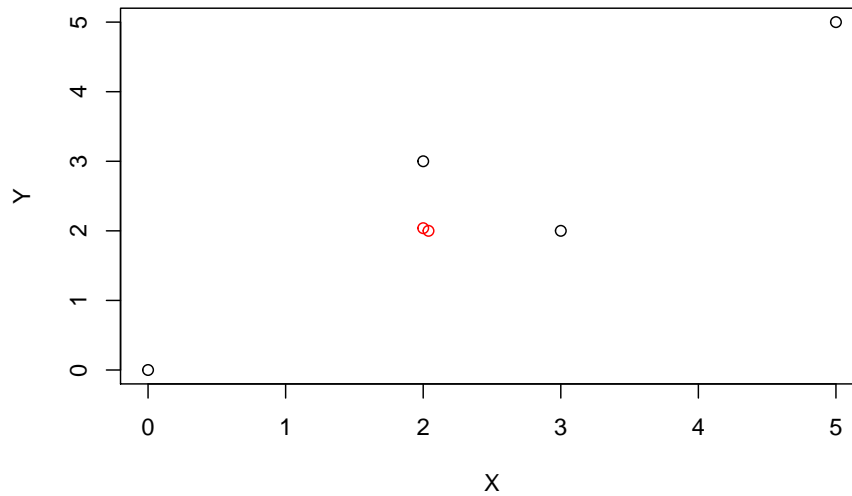
```
Y_5<-predict(lm.CC,data.frame('X'=2));Y_5
```

```
##      1
## 2.038462
```

```
X_Y.CC<-lm(X~Y,data = DATA,na.action = na.omit)
X_6<-predict(X_Y.CC,data.frame('Y'=2));X_6
```

```
##      1
## 2.038462
```

```
DATA.regImputed<-data.frame('X'=c(0,5,2,3,2,X_6),
                             'Y'=c(0,5,3,2,Y_5,2))
plot(DATA.regImputed,col=c(rep('black',4),rep('red',2)))
```



```
lm.regImputed<-lm(Y~X,data=DATA.regImputed)
lm.regImputed$df.residual<-lm.regImputed$df.residual-2
summary(lm.regImputed)
```

```
## Warning in summary.lm(lm.regImputed): residual degrees of freedom in object
## suggest this is not an "lm" fit
```

```
##
```

```
## Call:
```

```
## lm(formula = Y ~ X, data = DATA.regImputed)
```

```
##
```

```
## Residuals:
```

```
##      1      2      3      4      5      6
## -0.17606  0.20018  0.97443 -0.95032  0.01290 -0.06113
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.1761     0.7464   0.236  0.8355
## X             0.9248     0.2691   3.436  0.0752 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9817 on 2 degrees of freedom
## Multiple R-squared:  0.8552, Adjusted R-squared:  0.6379
## F-statistic: 11.81 on 1 and 2 DF,  p-value: 0.07525
```

### 迭代填补法

```
beta=beta_new=alpha=alpha_new=c(0,1)
epsilon<-1e-8
Delta<-Inf
X<-DATA$X
Y<-DATA$Y
while (Delta>=epsilon){
  X[6]<-alpha[1]+alpha[2]*Y[6]
  Y[5]<-beta[1]+beta[2]*X[5]
  beta_new<-lm(Y~X)$coef
  alpha_new<-lm(X~Y)$coef
  Delta<-max(abs(c(beta_new-beta,alpha_new-alpha)))
  beta<-beta_new; alpha<-alpha_new
}
cat('beta:', beta_new,'\n', 'alpha:',alpha_new,'\n',
    'X[6]:', X[6], 'Y[5]:',Y[5])

## beta: 0.1756455 0.9248595
## alpha: 0.1756455 0.9248595
## X[6]: 2.025364 Y[5]: 2.025364
```