目录

[第一次作业 1](#_Toc97829096)

[1.1题目1 1](#_Toc97829097)

[1.2 题目1解答 2](#_Toc97829098)

[1.3题目2 6](#_Toc97829099)

[1.4 题目2解答 7](#_Toc97829100)

[第二次作业 7](#_Toc97829101)

[2.1 题目1题干 7](#_Toc97829102)

[2.2 题目1解答 7](#_Toc97829103)

# 第一次作业

统计分析与R软件，P389，6.1，6.3

## 1.1题目1

表格

描述已自动生成

## 1.2 题目1解答

先画出散点图，感觉像是线性关系；

图表, 散点图

描述已自动生成

然后开始线性回归，使用代码如下：

depth <- c(5.1, 3.5, 7.1, 6.2, 8.8, 7.8,4.5,5.6,8.0,6.4)

migration <- c(1907,1287,2700,2373,3260,3000,1947,2273,3113,2493)

data.frame(name=depth,val=migration)

#plot(depth,migration,main="21901200405Lzx")

lm.reg <- lm(migration~depth+1)

print(summary(lm.reg))

结果如下图：

文本

描述已自动生成

发现截距项不显著，而migration关于depth的系数很显著，所以去掉截距项

（方程改为y~x-1）

文本

描述已自动生成

发现R^2 更接近于1了，虽然SSE变大了，但是幅度很小，所以采用无截距项的方案。

进行回归诊断，做出残差图，发现不对劲：

图表, 折线图

描述已自动生成

所以再看一下QQ图：

图表, 散点图

描述已自动生成

那么去掉7,8点（去掉5号点数据就太少了），结果 如下：

文本

描述已自动生成

R^2增加至0.9997, 更为显著了； 残差范围也显著的减小了；

图表, 折线图

描述已自动生成

所以最终的拟合方程为：migration=377.561\*depth；

将上述结果用于预测：

使用语句：

new <- data.frame(depth=7)

print("Predict value is:")

print(predict(lm.reg, new, interval="prediction", level=0.95))

结果如下图所示：

文本

描述已自动生成

预测值为[2518.581,2767,267]

至此，题目一解答完成。

## 1.3题目2



## 1.4 题目2解答

先画出散点图如下：

图表, 散点图

描述已自动生成

可以看出有同一个x值可能对应不同的y值，并且随着x的增大，y的变化幅度越来越大。

所以猜测logY与x成线性；同时还注意到在8~10附近有一个明显大小不对的点（24号），所以也将其去掉；做出如下

文本, 信件

描述已自动生成

可以看出截距项依旧不明显，将之去掉；

使用的代码是：

x <- c(1,1,1,1,2,2,2,3,3,3,4,4,4,5,6,6,6,7,7,7,8,8,8,9,11,12,12,12)

y<-c(0.6,1.6,0.5,1.2,2.0,1.3,2.5,2.2,2.4,1.2,3.5,4.1,5.1,5.7,3.4,9.7,8.6,4.0,5.5,10.5,17.5,13.4,4.5,30.4,12.4,13.4,26.2,7.4)

i<-1:28

x<-x[-24]

y<-y[-24]

plot(x,y,main="21901200405Lzx")

lm.reg=lm(formula = log(y)~x-1)

print(summary(lm.reg))

结果如下图，可以看出R^2显著提升；

文本, 信件

描述已自动生成

将回归曲线画在散点图上，如下图：

图表, 散点图

描述已自动生成

分析Summary的内容：T检验：t值较大（显著>1），Pr值显著<1，认为通过t检验；

F检验：p值很小，认为通过f检验；

(3)标准化残差图如下图：

图表, 散点图

描述已自动生成

普通残差图如下图：（下图和上图其实横纵坐标一样，只不过叫法不一样）

图表, 散点图

描述已自动生成

等方差性可以使用：white检验，代码如下

# load lmtest package

library(lmtest)

# perform White's test on the model

bptest(model, ~ formula, data)

但是从图形上直观地判断是否存在异方差是更为快捷的方法，

我们从残差图：Residuals-Fitted values(上图)上数据点的分布就可以看出是否存在异方差；本模型中的残差图数据点分布较为均匀，且在0附近，故认为其等方差性较好。

(4)

# 第二次作业

## 2.1 题目1题干

电脑屏幕的照片上有文字

描述已自动生成

对法国进口数据进行回归分析：判断是否存在多重共线性，如果存在，试着给出你的处理方案。

## 2.2 题目1解答