**11.8 R 语言应用**

**# 陈文贤 着 《大话统计学》 清华大学出版社 2022年**

# R例11.1

if(!require(ggplot2)){install.packages("ggplot2")} ; library(ggplot2)

if(!require(magrittr)){install.packages("magrittr")} ; library(magrittr)

if(!require(UsingR)){install.packages("UsingR")} ; library(UsingR)

data <- read.csv("C:/大话统计学 网络资源/StatData/Chap11\_1.csv",header=TRUE)

# 读入 Chap11\_1.csv

lm <- simple.lm(data$X, data$Y, show.ci=TRUE, conf.level=0.9)

cor(data$X, data$Y) # 相关系数

model <- lm(Y ~ X, data = data) ; model # 一元线性回归模型

model$coefficients ; model$residuals # 回归系数与残差

b1 <- model$coefficients[1] ; b2 <- model$coefficients[2]

data %>% ggplot(aes(x = X, y = Y)) + geom\_point() +

geom\_abline(intercept = b1, slope = b2, col = "red") # 模型绘图 + 置信区间

predict(model, newdata = data.frame(X = 125), interval = "confidence") # 置信区间估计

predict(model, newdata = data.frame(X = 125), interval = "predict") # 预测区间估计

range <- data.frame(X = seq(85, 122, 1)) # 绘图 : 置信区间估计与预测区间估计

pred <- predict(model, newdata = range, interval = "p")

colnames(pred) <- c("p\_fit", "p\_lwr", "p\_upr")

conf <- predict(model, newdata = range, interval = "c")

colnames(conf) <- c("c\_fit", "c\_lwr", "c\_upr")

bounds <- cbind(range, conf, pred)

plot1 <- data %>% ggplot(aes(x = X, y = Y)) + geom\_point() +

geom\_abline(intercept = b1, slope = b2) + geom\_point() + geom\_smooth(method = "lm")

plot1 <- plot1 + geom\_line(data = bounds, aes(y = c\_lwr), linetype = "dashed") +

geom\_line(data = bounds, aes(y = c\_upr), linetype = "dashed")

plot1 <- plot1 + geom\_line(data = bounds, aes(y = p\_lwr), color = " red") +

geom\_line(data = bounds, aes(y = p\_upr), color = " red") ; plot1

model$fit ; plot(model, which = 1) # 预测值与残图差

summary(model) ; aov(model) # 模型方差分析

## 补充数据 ISwR::thuesen 置信区间与预测区间

if(!require(ISwR)){install.packages("ISwR")} ; library(ISwR)

attach(thuesen) ; str(thuesen)

y <- short.velocity ; x <- blood.glucose

# lm(short.velocity~blood.glucose)

lm(y~x)

plot(x,y)

abline(lm(y~x))

options(na.action=na.exclude)

lm <- lm(y~x)

summary(lm)

anova(lm)

fitted(lm)

segments(x,fitted(lm), x,y, col="red") # 回归直线与残差

plot(fitted(lm),resid(lm))

qqnorm(resid(lm))

pred <- data.frame(x=4:20)

pp <- predict(lm, int="p", newdata=pred)

pc <- predict(lm, int="c", newdata=pred)

plot(x,y, ylim=range(y, pp, na.rm=T))

pred.gluc <- pred$x

matlines(pred.gluc, pc, lty=c(1,2,2), col="red")

matlines(pred.gluc, pp, lty=c(1,3,3), col="blue")

cor(x, y, use="complete.obs")

cor(thuesen, use="complete.obs")

cor.test(x, y, method="pearson")

cor.test(x, y)

cor.test(x, y, method="pearson") # 皮尔逊 Pearson 相关系数检验

cor.test(x, y, method="spearman") # 15.8 节 spearman 相关系数检验

cor.test(x, y, method="kendall") # kendall 相关系数检验