

课后题题目要改的地方：

1. P265 10.9

改成：对例 10.7，根据输出结果 10.8，手工算出以下两个样品的预测概率：

样品号	mode	income	样品号	mode	income
1	charter	7083.332	2	boat	3750.000

2. P266 10.11

改成：根据输出 10.9，手工计算例 10.8 数据中两个样品的预测概率。2 个样品的取值如下：

样品号	Infl	Type	Cont
1	Low	Tower	Low
2	Low	Atrium	Low

3. P266 10.12 删掉

答案：

第二章

$$2.2 \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

2.7 提示：

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^n e_i (\hat{\beta}_1 x_i - \hat{\beta}_1 \bar{x}) \\ &= \hat{\beta}_1 \sum_{i=1}^n e_i x_i - \hat{\beta}_1 \bar{x} \sum_{i=1}^n e_i = 0 \end{aligned}$$

2.9 提示：

$$\begin{aligned} \text{var}(e_i) &= \text{var}(y_i - \hat{y}_i) \\ &= \text{var}(y_i) + \text{var}(\hat{y}_i) - 2 \text{cov}(y_i, \hat{y}_i) \\ &= \text{var}(y_i) + \text{var}(\hat{\beta}_0 + \hat{\beta}_1 x_i) - 2 \text{cov}(y_i, \bar{y} + \hat{\beta}_1 (x_i - \bar{x})) \\ &= \sigma^2 + \sigma^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{L_{xx}} \right] - 2\sigma^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{L_{xx}} \right] \\ &= \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{L_{xx}} \right] \sigma^2 \end{aligned}$$

2.14

R 代码及部分输出：

```
> x=c(1:5)
> y=c(10,10,20,20,40)
```

```

> #散点图
> plot(x,y,main="销售收入与广告费用关系图",xlab="广告费用",ylab="销售收入")
> fit=lm(y~x) #最小二乘估计
> summary(fit) #参数估计，参数检验，回归标准误及参数显著性检验 p 值，决定系数
Call:
lm(formula = y ~ x)
Residuals:
    1      2      3      4      5 
4.000e+00 -3.000e+00  5.004e-16 -7.000e+00  6.000e+00 
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.000      6.351  -0.157   0.8849
x              7.000      1.915   3.656   0.0354 *
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.055 on 3 degrees of freedom
Multiple R-squared:  0.8167,    Adjusted R-squared:  0.7556 
F-statistic: 13.36 on 1 and 3 DF,  p-value: 0.03535

> confint(fit) #参数的区间估计
                2.5 %    97.5 %
(Intercept) -21.2112485 19.21125
x              0.9060793 13.09392

> anova(fit) #方差分析
Analysis of Variance Table

Response: y
            Df Sum Sq Mean Sq F value    Pr(>F)
x              1     490   490.00  13.364 0.03535 *
Residuals    3      110    36.67
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> cor.test(x, y,alternative = "two.side") #相关系数显著性检验
Pearson's product-moment correlation

data:  x and y
t = 3.6556, df = 3, p-value = 0.03535
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:

```

```

0.1057216 0.9936915
sample estimates:
      cor
0.9036961
> plot(x,fit$residuals,main = "残差图")    #残差图
> #prediction 是对 y 的实际值的预测，参数 confidence 是对 y 的期望值的预测
> new=data.frame(x=4.2)
> pred_y=predict(fit, new)
> pred_plim=predict(fit, new, interval = "prediction")
> pred_clim=predict(fit, new, interval = "confidence")

```

由输出可知，估计方程为 $\hat{y} = -1 + 7x$ ，回归标准误差 $\hat{\sigma} = 6.06$ ， $\hat{\beta}_0$ 置信水平为 95% 的置信区间为 $(-21.21, 19.21)$ ， $\hat{\beta}_1$ 置信水平为 95% 的置信区间为 $(0.91, 13.09)$ ，决定系数为 $r^2=0.82$ ，调整后 $r^2=0.76$ ， β_1 在 0.05 的显著性水平下显著不为 0， x 与 y 在 0.05 的显著性水平下有高度显著的线性依赖关系， $x_0 = 4.20$ 时 $y_0 = 28.40$ ， y_0 置信水平为 95% 的区间估计为 $(6.06, 50.74)$ ， $E(y_0)$ 置信水平为 95% 的区间估计为 $(17.10, 39.70)$ 。本例样本量 $n = 5$ 很小，所以区间估计的误差很大。

2.15

R 代码与 2.14 类似。由输出可知，估计方程为 $\hat{y} = 0.12 + 0.0036x$ ，回归标准误差 $\hat{\sigma} = 0.48$ ， $\hat{\beta}_0$ 置信水平为 95% 的置信区间为 $(-0.70, 0.94)$ ， $\hat{\beta}_1$ 置信水平为 95% 的置信区间为 $(0.0026, 0.0046)$ ，决定系数为 $r^2=0.90$ ，调整后 $r^2=0.89$ ， β_1 显著不为 0， x 与 y 有高度显著的线性依赖关系， $x_0 = 1000.00$ 时 $y_0 = 3.70$ ， y_0 置信水平为 95% 的区间估计为 $(2.52, 4.89)$ ， y_0 置信水平为 95% 的近似区间估计为 $(2.74, 4.66)$ ， $E(y_0)$ 置信水平为 95% 的区间估计为 $(3.28, 4.12)$ 。

2.16

R 代码：

```

fit=lm(y~x)    #最小二乘估计
hist(fit$residuals)
qqnorm(fit$residuals,main="残差 qq 图")
qqline(fit$residuals)

```

- (1) 散点图（略），可以用直线反映两变量之间的关系。
- (2) $\hat{y} = 12112 + 3.31x$
- (3) 从残差的直方图看，略呈右偏分布；从正态概率图看，散点基本呈直线趋势，可以认为残差服从正态分布。

第三章

3.11

R 代码及部分输出：

```
> x1=c(70,75,65,74,72,68,78,66,70,65)
> x2=c(35,40,40,42,38,45,42,36,44,42)
> x3=c(1.0,2.4,2.0,3.0,1.2,1.5,4.0,2.0,3.2,3.0)
> y=c(160,260,210,265,240,220,275,160,275,250)
> data1=data.frame(x1,x2,x3,y)
> cor(data1) #相关系数矩阵
> fit1=lm(y~x1+x2+x3) #最小二乘估计
> summary(fit1) #参数估计，参数检验，回归标准误及参数显著性检验 p 值，决定系数
```

Call:

```
lm(formula = y ~ x1 + x2 + x3)
```

Residuals:

Min	1Q	Median	3Q	Max
-25.198	-17.035	2.627	11.677	33.225

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-348.280	176.459	-1.974	0.0959 .
x1	3.754	1.933	1.942	0.1002
x2	7.101	2.880	2.465	0.0488 *
x3	12.447	10.569	1.178	0.2835

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.44 on 6 degrees of freedom

Multiple R-squared: 0.8055, Adjusted R-squared: 0.7083

F-statistic: 8.283 on 3 and 6 DF, p-value: 0.01487

```
> fit2=lm(y~x1+x2) #最小二乘估计
> summary(fit2) #参数估计，参数检验，回归标准误及参数显著性检验 p 值，决定系数
```

Call:

```
lm(formula = y ~ x1 + x2)
```

Residuals:

Min	1Q	Median	3Q	Max
-42.012	-10.656	4.358	11.984	28.927

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
--	----------	------------	---------	----------

```

(Intercept) -459.624    153.058   -3.003   0.01986 *
x1           4.676      1.816    2.575   0.03676 *
x2           8.971      2.468    3.634   0.00835 **
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 24.08 on 7 degrees of freedom
Multiple R-squared:  0.7605,    Adjusted R-squared:  0.6921
F-statistic: 11.12 on 2 and 7 DF,  p-value: 0.006718
> confint(fit1) #参数的区间估计
                2.5 %    97.5 %
(Intercept) -780.0603287 83.499990
x1           -0.9766149  8.484688
x2            0.0529179 14.148507
x3          -13.4147488 38.309689
> data2=scale(data1)/sqrt(length(y)-1)
> data2=data.frame(data2)
> names(data2)=c("x1","x2","x3","y")
> fit3=lm(y~x1+x2+x3,data=data2) #最小二乘估计
> summary(fit3) #参数估计, 参数检验, 回归标准误及参数显著性检验 p 值, 决定系数
Call:
lm(formula = y ~ x1 + x2 + x3, data = data2)
Residuals:
    Min       1Q   Median       3Q      Max
-0.19353 -0.13083  0.02017  0.08969  0.25518
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.928e-16  5.693e-02   0.000   1.0000
x1           3.848e-01  1.982e-01   1.942   0.1002
x2           5.355e-01  2.172e-01   2.465   0.0488 *
x3           2.771e-01  2.353e-01   1.178   0.2835
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.18 on 6 degrees of freedom
Multiple R-squared:  0.8055,    Adjusted R-squared:  0.7083
F-statistic: 8.283 on 3 and 6 DF,  p-value: 0.01487
> newdata=data.frame(75,42,3.1)

```

```

> names(newdata)=c("x1","x2","x3")
> pred_y=predict(fit2, newdata)
> pred_plim=predict(fit2, newdata, interval = "prediction")
> pred_clim=predict(fit2, newdata, interval = "confidence")
> pred_y+2*24.08
> pred_y-2*24.08

```

由 R 输出, 可见 x_3 的 $P = 0.28$, 最不显著, 因此予以剔除。 $x_{01} = 75$, $x_{02} = 42$ 时 $\hat{y}_0 = 267.83$, y_0 置信水平为 95% 的区间估计是 $(204.44, 331.22)$, y_0 置信水平 95% 的近似区间估计是 $(219.67, 315.99)$, 本例样本量 $n = 10$ 较小, 所以区间估计的误差较大。 $E(y_0)$ 置信水平为 95% 的区间估计为 $(239.97, 295.69)$ 。

3.12

R 输出如下:

```

Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min       1Q   Median       3Q      Max
-6589.0 -2504.9  123.3  2110.1 10961.5

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4490.3824   3797.2636  -1.183    0.251
x1              0.7983     0.6475   1.233    0.232
x2              2.0014     0.1254  15.963 7.6e-13 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4055 on 20 degrees of freedom

Multiple R-squared:  0.9993,    Adjusted R-squared:  0.9993

F-statistic: 1.468e+04 on 2 and 20 DF,  p-value: < 2.2e-16

```

x_1 系数 0.80 明显不合理。

第四章

4.9

R 代码及部分输出:

```

> fit1=lm(y~x) #最小二乘估计
> plot(x,fit1$residuals,main="残差图") #残差图
> cor.test(x,abs(fit1$residuals),alternative="two.side",method="spearman") #相关系数显著性检验

```

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/497022166124006060>