短学期作业五

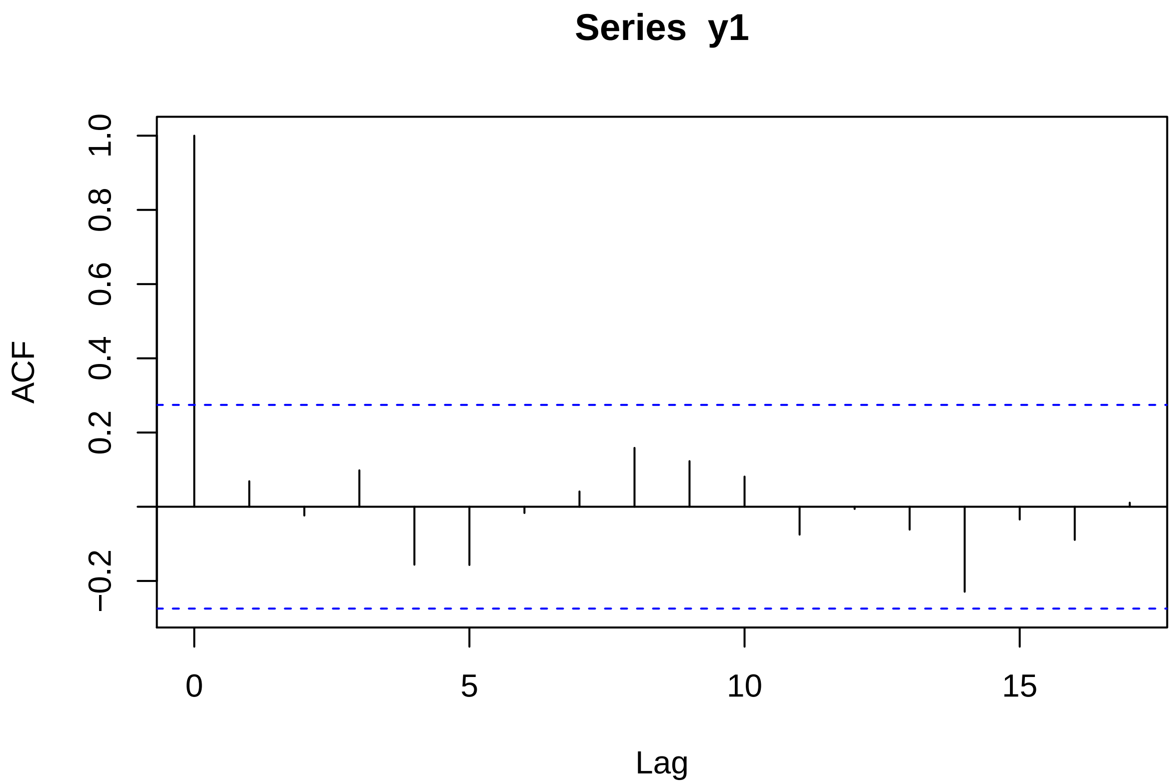
汪利军 *3140105707*

July 10, 2017

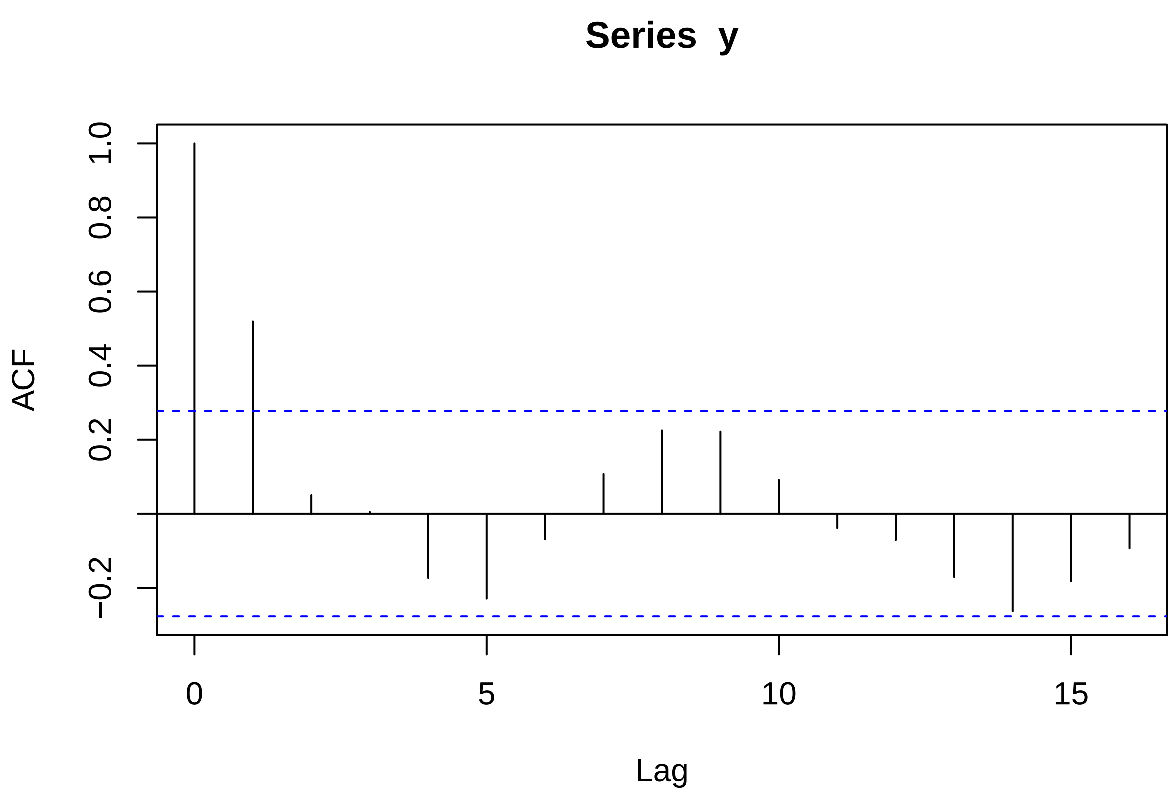
Contents

1 MB 4.6

```
y1 <- rnorm(51)
y <- y1[-1] + y1[-51]
acf(y1)
```


Series y1
ACF
1.0
0.8
0.6
0.4
0.2
−0.2
0
5
10
15
Lag

```
acf(y)
```


Series y
ACF
1.0
0.8
0.6
0.4
0.2
−0.2
0
5
10
15
Lag

2 MB 4.7

产生 y 的函数如下

```
GenY <- function()
{
  y1 <- rnorm(51)
  y <- y1[-1] + y1[-51]
  return(y)
}
```

重复 25 次

```
data <- sapply(1:25, function(i) GenY())
```

计算均值

```
av = colMeans(data)
av
```

```
##  [1]  0.09916692 -0.12000797 -0.28236536  0.08654361  0.07530796
##  [6]  0.29097112 -0.02933765  0.04520298 -0.04543752  0.12052568
## [11] -0.43090709  0.33511485 -0.26602769  0.21887430 -0.39813619
## [16] -0.04764569  0.63294186 -0.48010572  0.19442445  0.14598963
## [21] -0.00387911 -0.30356651 -0.17144732  0.16341417 -0.17916579
```

计算方差

```
v = apply(data, 2, var)
v
```

```
##  [1] 2.131564 2.054967 1.521866 1.734443 2.393482 2.807121 2.746444
##  [8] 1.073084 1.781633 2.451884 1.422889 2.579296 2.113654 1.568889
## [15] 2.554014 2.285997 1.866159 1.802057 1.816088 2.364262 2.730328
## [22] 2.390861 2.759917 2.162368 1.672024
```

3 MB 4.9

3.1 (a)

构造表格如下

```
tab1 <- as.table(rbind(c(74, 43, 11), c(71, 38, 65)))
dimnames(tab1) <- list(reduction = c(">= 50%", "<50%"),
                       treatment = c("Acupuncture", "Sham acupuncture", "Waiting list"))
tab1
```

```
##          treatment
## reduction Acupuncture Sham acupuncture Waiting list
##    >= 50%          74               43           11
##    <50%            71               38           65
```

为检验不同的 treatment 对于 reduction 是否独立，我们采用卡方检验

```r
chisq.test(tab1)
```

```
##
##  Pearson's Chi-squared test
##
## data:  tab1
## X-squared = 32.486, df = 2, p-value = 8.825e-08
```

由上述的 p 值 (p-value << 0.05) 可以看出，不同的治疗方案 (treatment) 与疗效 (reduction) 不独立。

3.2 (b)

构造表格如下

```r
tab2 <- as.table(cbind(c(82, 17, 30), c(30, 26, 16)))
dimnames(tab2) <- list(guess = c("Chinese", "Other", "Don't know"),
                       treatment = c("Acupuncture", "Sham acupuncture"))
tab2
```

```
##             treatment
## guess         Acupuncture Sham acupuncture
##   Chinese              82               30
##   Other                17               26
##   Don't know           30               16
```

为检验病人的猜测与接受的 treatment 直接的独立性，采用卡方检验。

```r
chisq.test(tab2)
```

```
##
##  Pearson's Chi-squared test
##
## data:  tab2
## X-squared = 15.358, df = 2, p-value = 0.0004624
```

由上述结果的 p 值（p-value << 0.05），可以看出不同的治疗方案与病人的猜测是不独立的。

4 MB 4.12

```
admissions.A <- array(c(30,30,10,10,15,5,30,10),dim=c(2,2,2))
admissions.B <- array(c(30,30,20,10,10,5,20,25),dim=c(2,2,2))
## 从 mantelhaen.test() 的帮助文档中得到 woolf 函数
woolf <- function(x) {
  x <- x + 1 / 2
  k <- dim(x)[3]
  or <- apply(x, 3, function(x) (x[1,1]*x[2,2])/(x[1,2]*x[2,1]))
  w <-  apply(x, 3, function(x) 1 / sum(1 / x))
  1 - pchisq(sum(w * (log(or) - weighted.mean(log(or), w)) ^ 2), k - 1)
}
```

对于 Table4.10A

```
woolf(admissions.A)
```

```
## [1] 0.9695992
```

表明 p-value $= 0.9696$，表面不能拒绝原假设，也就是不同的类别（如性别，专业和录取情况）之间没有显著的差异。

对于 Table4.10B

```
woolf(admissions.B)
```

```
## [1] 0.04302033
```

表面 p-value $= 0.043 < 0.05$，表面在 0.05 的置信水平下拒绝原假设，也就是不同的类别（如性别，专业和录取情况）之间存在显著的差异。

下面进行 Mantel–Haenzel 检验，对于 Table4.10A，

```
mantelhaen.test(admissions.A)
```

```
##
##  Mantel-Haenszel chi-squared test without continuity correction
##
## data:  admissions.A
## Mantel-Haenszel X-squared = 0, df = 1, p-value = 1
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.4565826 2.1901841
## sample estimates:
## common odds ratio
##                 1
```

由上述结果可以看出，因 p-value $= 1$，所以不能拒绝原假设，也就是真实的比率等于 1，这意味着不同的类别之间不存在差异；估计的 common odds ratio $= 1$，且 95% 的置信区

间为 $[0.4565826, 2.1901841]$

对于 Table4.10B

```r
mantelhaen.test(admissions.B)
```

```
##
##  Mantel-Haenszel chi-squared test with continuity correction
##
## data:  admissions.B
## Mantel-Haenszel X-squared = 0.014147, df = 1, p-value = 0.9053
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.448071 1.807749
## sample estimates:
## common odds ratio
##               0.9
```

由上述结果可以看出，因 p-value = 0.9053，所以不能拒绝原假设，也就是真实的比率等于 1，这意味着不同的类别之间不存在差异；估计的 common odds ratio = 0.9，且 95% 的置信区间为 $[0.448071, 1.807749]$

5 MDL Chapter 13 Worksheet C: East German athletes

5.1 Problem 13.1

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu > \mu_0$$

则在 H_0 下，

$$T = \sqrt{n}(\frac{\bar{X} - \mu_0}{\hat{\sigma}}) \sim t(n-1)$$

采用 t 检验

```r
data = c(3.22, 3.07, 3.17, 2.91, 3.40, 3.58, 3.23, 3.11, 3.62)
t.test(data, mu = 3.1, alternative = "greater")
```

```
##
##  One Sample t-test
##
## data:  data
## t = 1.9968, df = 8, p-value = 0.04046
## alternative hypothesis: true mean is greater than 3.1
## 95 percent confidence interval:
```

```
##  3.110771      Inf
## sample estimates:
## mean of x
##  3.256667
```

5.2 Problem 13.2

由上述 t 检验结果可以看出，p 值为 0.04046<0.05，故在 0.05 的水平下拒绝原假设，也就是 $\mu > 3.1$，于是可以认定 East German 运动员服用了 performance-enhancing drugs。

6 MDL Chapter 13 Worksheet C: Drinking and driving question

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

在 H_0 下，

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}\sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2)$$

其中

$$\hat{\sigma}^2 = \frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2}$$

```r
before = c(57, 54, 62, 64, 71, 65, 70, 75, 68, 70, 77, 74, 80, 83)
after = c(55, 60, 68, 69, 70, 73, 74, 74, 75, 76, 78, 81, 90)
t.test(before, after)
```

```
##
##  Welch Two Sample t-test
##
## data:  before and after
## t = -0.98231, df = 24.577, p-value = 0.3355
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.07854   3.57305
## sample estimates:
## mean of x mean of y
##  69.28571  72.53846
```

由上述 t 检验结果可以看出，p 值为 0.3355>0.05，也就是不能拒绝原假设，故 alcohol 对 reflexes 没有效果。

7 MDL Chapter 15 Worksheet B: Study of batteries

7.1 Problem 15.1

存在两个因子，temperature 和 type of battery；temperature 有 3 个层次，分别为 $15^oC, 70^oC, 125^oC$；type of battery 有三个层次，分别为 Type I，Type II，Type III。响应变量为 lifetime。

7.2 Problem 15.2

令 A，B 都是含三个层次的因子，对于每个数据对 $(i, j), i = 1, 2, 3; j = 1, 2, 3$，都观察了 4 次 lifetime 的值，并且假设

$$Y_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma^2)$$

得到模型

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \qquad \text{for } k = 1, 2, 3, 4; i = 1, 2, 3; j = 1, 2, 3$$

其中 ε_{ijk} 为独立同分布的随机变量，$\mathcal{N}(0, \sigma^2)$

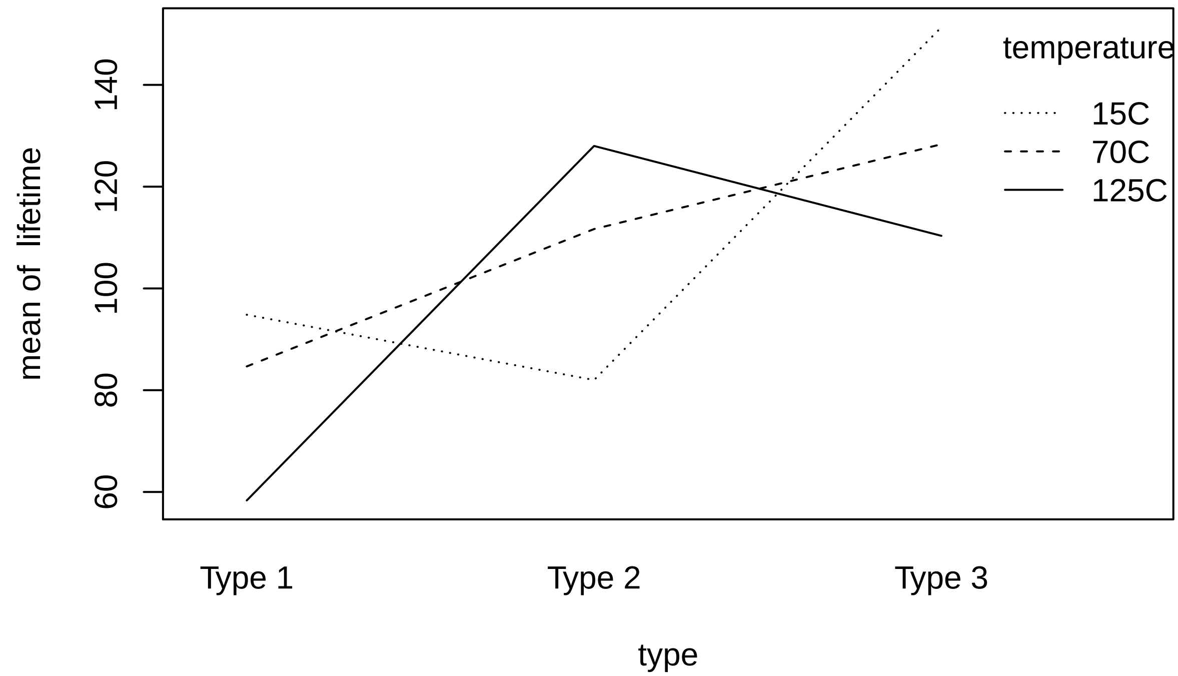
7.3 Problem 15.3

构造数据集

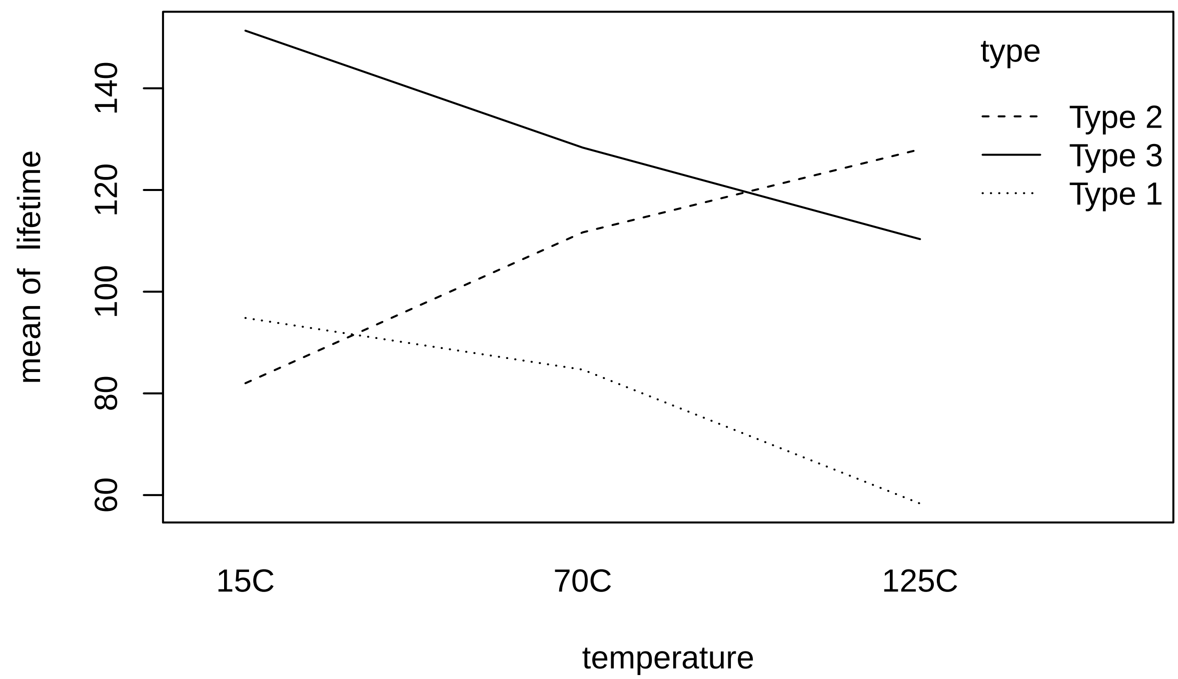
```r
lifetime <- c(130, 155, 74, 180, 34, 40, 80, 75, 20, 70, 82, 58,
              150, 188, 159, 126, 136, 122, 106, 115, 25, 70, 58, 45,
              138, 110, 168, 160, 174, 120, 150, 139, 96, 104, 82, 60)
type <- gl(3, 12, 36, labels = paste0("Type ", 1:3))
temperature <- gl(3, 3, 36, labels = paste0(c(15, 70, 125), "C"))
battery <- data.frame(lifetime, type, temperature)
```

interaction 图象如下

```r
interaction.plot(type, temperature, lifetime)
```




```r
interaction.plot(temperature, type, lifetime)
```



7.4 Problem 15.4

参数估计为

```r
summary(lm(lifetime~type*temperature))
```

```
##
## Call:
## lm(formula = lifetime ~ type * temperature)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -66.667 -33.458  -1.167  24.917  95.333
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    94.83      18.85   5.030 2.81e-05 ***
## typeType 2                    -12.83      32.65  -0.393    0.697
## typeType 3                     56.50      32.65   1.730    0.095 .
## temperature70C                -10.17      32.65  -0.311    0.758
## temperature125C               -36.50      32.65  -1.118    0.274
## typeType 2:temperature70C      39.83      46.18   0.863    0.396
## typeType 3:temperature70C     -12.83      49.88  -0.257    0.799
## typeType 2:temperature125C     82.50      49.88   1.654    0.110
## typeType 3:temperature125C     -4.50      46.18  -0.097    0.923
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46.18 on 27 degrees of freedom
## Multiple R-squared:  0.2585, Adjusted R-squared:  0.03875
## F-statistic: 1.176 on 8 and 27 DF,  p-value: 0.3488
```

7.5 Problem 15.5

ANOVA Table 如下

```
summary(aov(lifetime~type*temperature, data = battery))
```

```
##                  Df Sum Sq Mean Sq F value Pr(>F)
## type              2  10684    5342   2.505  0.100
## temperature       2   1498     749   0.351  0.707
## type:temperature  4   7887    1972   0.925  0.464
## Residuals        27  57578    2133
```

7.6 Problem 15.6

为进行相关性检验，我们采用卡方检验，对 $Y_{ijk}, k = 1, 2, 3, 4$ 取均值来表示 Type i，第 j 个温度的 lifetime，从而构造列联表进行相关性检验

```r
lifetime2 <- array(lifetime, dim = c(4, 3, 3))
tab = margin.table(lifetime2, margin = c(3,2))/4
dimnames(tab) = list(type = c("Type 1", "Type 2", "Type 3"),
                     temperature = c("15", "70", "125"))
tab
```

```
##         temperature
## type          15     70  125
##   Type 1 134.75  57.25 57.5
##   Type 2 155.75 119.75 49.5
##   Type 3 144.00 145.75 85.5
```

进行卡方检验

```r
chisq.test(tab)
```

```
## 
##  Pearson's Chi-squared test
## 
## data:  tab
## X-squared = 27.042, df = 4, p-value = 1.95e-05
```

由上述 p 值可以看出，电池类型与温度之间不独立，也就是存在一定的相关性。