> #独立性检验

> x = c(52,64,24,60,59,52,50,65,74)

> M = matrix(x,nr=3,nc=3,byrow = TRUE,dimnames = list(c("甲","乙","丙"),c("一级","二级","三级")))

> chisq.test(M)

Pearson's Chi-squared test

data: M

X-squared = 19.822, df = 4, p-value = 0.0005414

>

> #方差分析

> example10\_1 = read.csv("C:/Users/K49/Desktop/2022\_2023\_2/stat/example/chap08/example10\_1.csv",fileEncoding="GBK")

> example10\_2 = melt(example10\_1,variable.name = "行业",value.name = "投诉次数")

No id variables; using all as measure variables

>

> #拟合单因子方差分析模型并输出结果

> model1\_1w = aov(投诉次数~行业,data = example10\_2)

> summary(model1\_1w)

Df Sum Sq Mean Sq F value Pr(>F)

行业 3 1457 485.5 3.407 0.0388 \*

Residuals 19 2708 142.5

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

因为不存在，5个观察量被删除了

>

> #方差分析模型参数估计

> model1\_1w$coefficients

(Intercept) 行业旅游业 行业航空公司 行业家电制造业

49 -1 -14 10

>

> #输出基本结果

> library(agricolae)

> model1\_1w = aov(投诉次数~行业 ,data = example10\_2)

> LSD = LSD.test(model1\_1w,"行业");LSD

$statistics

MSerror Df Mean CV

142.5263 19 47.86957 24.93952

$parameters

test p.ajusted name.t ntr alpha

Fisher-LSD none 行业 4 0.05

$means

投诉次数 std r LCL UCL Min Max Q25 Q50 Q75

航空公司 35 10.41633 5 23.82528 46.17472 21 49 31.0 34 40.00

家电制造业 59 12.74755 5 47.82528 70.17472 44 77 51.0 58 65.00

零售业 49 10.80123 7 39.55564 58.44436 34 66 42.0 49 55.00

旅游业 48 13.59412 6 37.79892 58.20108 29 68 40.5 48 54.75

$comparison

NULL

$groups

投诉次数 groups

家电制造业 59 a

零售业 49 ab

旅游业 48 ab

航空公司 35 b

attr(,"class")

[1] "group"

>

> #输出更多信息

> library(DescTools)

> PostHocTest(model1\_1w,method = "lsd")

Posthoc multiple comparisons of means : Fisher LSD

95% family-wise confidence level

$行业

diff lwr.ci upr.ci pval

旅游业-零售业 -1 -14.901728 12.9017278 0.8819

航空公司-零售业 -14 -28.631146 0.6311462 0.0597 .

家电制造业-零售业 10 -4.631146 24.6311462 0.1688

航空公司-旅游业 -13 -28.130646 2.1306458 0.0880 .

家电制造业-旅游业 11 -4.130646 26.1306458 0.1446

家电制造业-航空公司 24 8.196556 39.8034441 0.0049 \*\*

---

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

>

>

> #回归

> example9\_1 = read.csv("C:/Users/K49/Desktop/2022\_2023\_2/stat/example/chap09/example9\_1.csv",fileEncoding="GBK")

> model = lm(不良贷款 ~ 各项贷款余额,data = example9\_1)

> summary(model)

Call:

lm(formula = 不良贷款 ~ 各项贷款余额, data = example9\_1)

Residuals:

Min 1Q Median 3Q Max

-2.2882 -1.5233 -0.1802 0.8935 6.3038

Coefficients:

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

(Intercept) -0.82952 0.72304 -1.147 0.263

各项贷款余额 0.03789 0.00503 7.534 1.18e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.98 on 23 degrees of freedom

Multiple R-squared: 0.7116, Adjusted R-squared: 0.6991

F-statistic: 56.75 on 1 and 23 DF, p-value: 1.183e-07

>

> #计算回归系数的置信区间

> confint(model,level=0.95)

2.5 % 97.5 %

(Intercept) -2.32524963 0.66620840

各项贷款余额 0.02748905 0.04830036

>

> #输出方差分析表

> anova(model)

Analysis of Variance Table

Response: 不良贷款

Df Sum Sq Mean Sq F value Pr(>F)

各项贷款余额 1 222.486 222.49 56.754 1.183e-07 \*\*\*

Residuals 23 90.164 3.92

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1