Homework 12

Pascal Pilz, k12111234

2024-01-15

Exercise 30

Partial correlation

```r
n_crimes   <- c(210, 234, 279, 179, 160, 136)
n_police   <- c(20, 21, 23, 19, 18, 15)
population <- c(30000, 40000, 50000, 20000, 10000, 5000)

r_xy <- cor(n_crimes, n_police)
r_xu <- cor(n_crimes, population)
r_yu <- cor(n_police, population)

r_xy.u <- (r_xy - r_xu * r_yu) / sqrt((1 - r_xu^2) * (1 - r_yu^2))
cat(sprintf("r_xy: %.3f, r_xu: %.3f, r_yu: %.3f, r_xy.u: %.3f",
            r_xy, r_xu, r_yu, r_xy.u))
```

```
## r_xy: 0.968, r_xu: 0.993, r_yu: 0.958, r_xy.u: 0.490
```

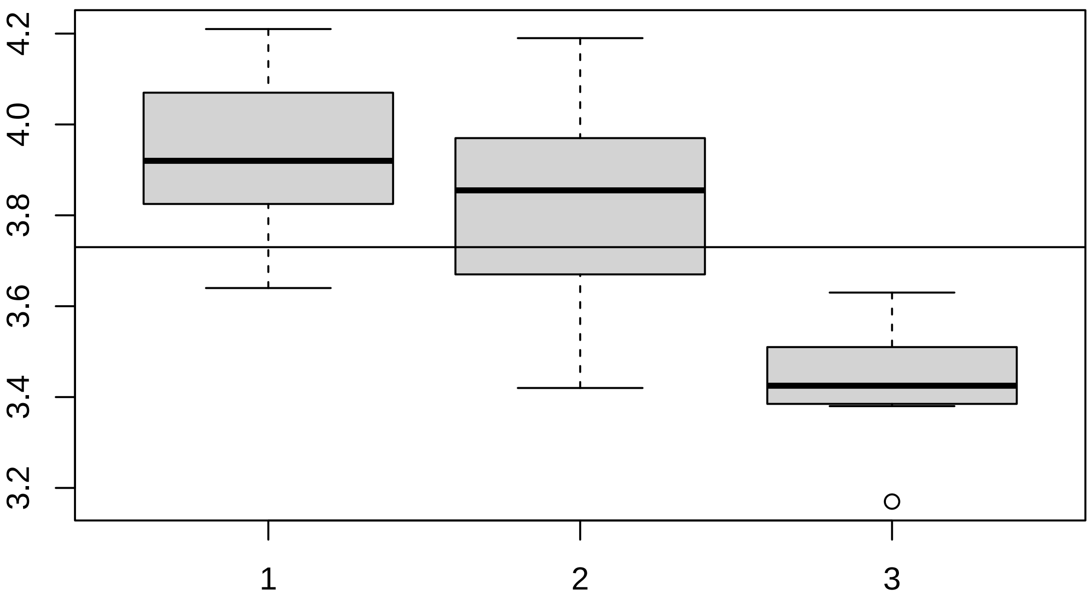
Statistical test

```r
n <- length(n_crimes)
test_statistic <- r_xy.u * sqrt((n - 3) / (1 - r_xy.u^2))
critical_value <- qt(0.975, n-3)

cat(sprintf("test statistic: %.3f; critical value: %.3f"
            , test_statistic, critical_value), "\n",
    sprintf("We can see that we %s reject the null hypothesis.",
            if (abs(test_statistic) > critical_value) "can" else "cannot"))
```

```
## test statistic: 0.974; critical value: 3.182
##  We can see that we cannot reject the null hypothesis.
```

Exercise 31

```r
A <- c(3.64, 3.77, 4.18, 4.21, 3.88, 3.93, 3.91, 3.96)
B <- c(3.42, 3.96, 3.87, 4.19, 3.58, 3.76, 3.84, 3.98)
C <- c(3.17, 3.63, 3.38, 3.47, 3.39, 3.41, 3.55, 3.44)

boxplot(A, B, C)
abline(h=mean(c(A, B, C)))
```



Formulation of a statistical hypothesis

Since we are comparing three different groups, we choose the following:

- H_0: $\mu_A = \mu_B = \mu_C$
- H_1: $\mu_A \neq \mu_B$ or $\mu_A \neq \mu_C$ or $\mu_B \neq \mu_C$

Statistical test

The obvious choice would be an analysis of variance (ANOVA). But first, we check the assumptions.

Normality

```r
library(nortest)
lillie.test(A)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  A
## D = 0.19781, p-value = 0.4714
```

```r
lillie.test(B)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
```

```
## data:  B
## D = 0.14983, p-value = 0.861
```

```r
lillie.test(C)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  C
## D = 0.23089, p-value = 0.238
```

As we can see, the data is normally distributed.

Homoscedasticity

For this we perform a Levene test.

```r
library(car)
```

```
## Loading required package: carData
```

```r
dv <- c(A, B, C)
iv <- as.factor(c(rep("A", length(A)), rep("B", length(B)), rep("C", length(C))))

leveneTest(dv~iv, center="mean")
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##       Df F value Pr(>F)
## group  2  0.9894 0.3885
##       21
```

As we can see, the data can be assumed to be homoscedatic. Thus, we can proceed with ANOVA.

ANOVA

```r
m_grand <- mean(c(A, B, C))

m_A <- mean(A)
l_A <- length(A)

m_B <- mean(B)
l_B <- length(B)

m_C <- mean(C)
l_C <- length(C)

SS_between <- l_A*(m_A-m_grand)^2 + l_B*(m_B-m_grand)^2 + l_C*(m_C-m_grand)^2
SS_within  <- sum((A - m_A)^2 + (B - m_B)^2 + (C - m_C)^2)
SS_total   <- SS_between + SS_within

cat(sprintf("SS_between: %.3f, SS_within: %.3f, SS_total: %.3f",
            SS_between, SS_within, SS_total), "\n")
```

```
## SS_between: 1.128, SS_within: 0.788, SS_total: 1.917
```

```r
k <- 3
N <- length(c(A, B, C))
```

```r
MS_between <- SS_between / (k-1)
MS_within  <- SS_within / (N-k)
MS_total   <- SS_total / (N-1)

cat(sprintf("MS_between: %.3f, MS_within: %.3f, MS_total: %.3f",
            MS_between, MS_within, MS_total), "\n")
```

```
## MS_between: 0.564, MS_within: 0.038, MS_total: 0.083
```

```r
test_statistic <- MS_between / MS_within
critical_value <- qf(0.95, k-1, N-k)

cat(sprintf("test statistic: %.3f; critical value: %.3f"
            , test_statistic, critical_value), "\n",
    sprintf("We can see that we %s reject the null hypothesis.",
            if (abs(test_statistic) > critical_value) "can" else "cannot"))
```

```
## test statistic: 15.028; critical value: 3.467
##  We can see that we can reject the null hypothesis.
```

Verifying

```r
model <- aov(dv~iv)
summary(model)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## iv           2 1.1284  0.5642   15.03 8.89e-05 ***
## Residuals   21 0.7884  0.0375                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```