Solutions (TMA4267 2024 August)

1. If $x \geq y$, then

$$F_Y(x,y) = P(X_{\min} \leq x, X_{\max} \leq y) = P(X_{\max} \leq y) =$$

$$= P(X_1 \leq y, ..., X_n \leq y) = \prod_{i=1}^{n} P(X_i \leq y) = (F(y))^n.$$

If $x < y$, then

$$F_Y(x,y) = P(X_{\min} \leq x, X_{\max} \leq y) = P(X_{\max} \leq y) - P(X_{\min} > x, X_{\max} \leq y) =$$

(we used that $P(A \cap B) = P(B) - P(A^c \cap B)$)

$$= P(X_1 \leq y, ..., X_n \leq y) - P(x < X_1 \leq y, ..., x < X_n \leq y) =$$

$$= \prod_{i=1}^{n} P(X_i \leq y) - \prod_{i=1}^{n} P(x < X_i \leq y) = (F(y))^n - (F(y) - F(x))^n.$$

2. We have

$$a_1X_1 + a_2X_2 + a_3X_3 = c.$$

Multiplying both sides of this equality by X_1 and taking the expectation, we obtain

$$a_1 + \rho a_2 + \rho a_3 = 0.$$

In the same way (multiplying by X_2 and X_3)we obtain equations

$$\rho a_1 + a_2 + \rho a_3 = 0,$$

$$\rho a_1 + \rho a_2 + a_3 = 0.$$

Subtracting the second equation from the first, obtain $a_1 = a_2$. In the same way $a_1 = a_3$ i.e. $a_1 = a_2 = a_3$. Then

$$a_i(1 + 2\rho) = 0$$

Since $a_i \neq 0$ (otherwise $a^T X = 0$), this implies that $\rho = -1/2$

3. Since

$$Var(X_1) = Var(X_2) = Cov(X_1, X_2) = 1,$$

the equality $X_1 = X_2$ holds with probability one.

a)

$$P(X_1 > 2X_2) = P(X_1 > 2X_1) = P(X_1 < 0) = \frac{1}{2}.$$

Problen (a) can also be solved by noting that $P(X_1 > 2X_2) = P(X_1 - 2X_2 > 0)$ and $Y = X_1 - 2X_2$ is normal with mean 0 and variance 1. Then $P(Y > 0) = 0.5$.

b)

$$F_X(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2) = P(X_1 \le x_1, X_1 \le x_2) =$$

$$= P(X_1 \le \min\{x_1, x_2\}) = \Phi(\min\{x_1, x_2\}),$$

where $\Phi(z)$ is the univariate standard normal distribution function.

4.

a) Since

```
lm(formula = Y ~ x1 + x2),
```

there are two explanatory variables, therefore $p = 3$. The number ofdegrees of freedom for

```
Residual standard error
```

is $n - p$, therefore $n - p = 17$, $n = 20$.

The relationship between

```
R-squared
```

and

```
Adjusted R-squared
```

is

$$R^2 = 1 - \frac{n-p}{n-1}(1 - R^2_{\text{adj}}),$$

therefore for the first question mark we obtain

$$R^2 = 1 - \frac{17}{19}(1 - 0.5435) = 0.5916.$$

The numbers of degrees os freedom of

```
F-statistic
```

are $p - 1$ and $n - p$, therefore the second the third question marks are 2 and 17.

b) Consider the following matrix: $A = [0\ 1\ 1]$. The null hypothesis and alternative can be written as

$$H_0 : A\beta = 0 \text{ vs. } H_1 : A\beta \ne 0.$$

We use the general F-test. Test statistic is

$$F = \frac{(A\hat{\beta})^T (A(X^T X)^{-1} A^T)^{-1} A\hat{\beta}}{\hat{\sigma}^2}.$$

Under H_0 this statistic has F-distribution with 1 and $n - p$ degrees of freedom. In our case $n = 20$, $p = 3$, $\hat{\beta}_1 = 1.3058$, $\hat{\beta}_2 = 2.1532$, $\hat{\sigma} = 0.8248$,

$$A(X^T X)^{-1} A^T = 0.8,$$

$$(A(X^T X)^{-1} A^T)^{-1} = \frac{10}{8} = 1.25,$$

$$A\hat{\beta} = \hat{\beta}_1 + \hat{\beta}_2 = 3.459,$$
$$(A\hat{\beta})^T(A(X^TX)^{-1}A^T)^{-1}A\hat{\beta} = 3.459^2 \cdot 1.25 = 14.96,$$
$$F = \frac{14.96}{0.8248^2} = 22.$$

Thus the observed value of the test statistic is $F = 22$. Since $f_{0.05,1,17} = 4.45$, the null hypothesis is rejected.

c) H_0 is rejected if the corresponding p-value is smaller than α_{loc}. For the Bonferroni method

$$\alpha_{\text{loc}} = \frac{1}{m}\text{FWER}$$

, where m is the number of hypotheses. For the Šidák method

$$\alpha_{\text{loc}} = 1 - (1 - \text{FWER})^(1/m).$$

Since $m = 3$, FWER $= 0.01$, we have respectively $\alpha_{\text{loc}} = 0.0033$ and $\alpha_{\text{loc}} = 0.0033$

The p-values are in the column

```
Pr(>|t|)
```

They are 0.405281, 0.010165, 0.000179. Therefore for both methods only the third hypothesis is rejected.

d) Test statistic is $(n-p)\hat{\sigma}^2$. Under H_0 this test statistic has chi-square distribution with $n-p$ degrees of freedom. Therefor H_0 is rejected if $(n-p)\hat{\sigma}^2 > c$, where c is such a number that

$$P(\chi^2_{n-p} > c) = 0.05.$$

In our case, $n - p = 17$, $\hat{\sigma}^2 = 0.8248^2$, $c = 27.587$. H_0 is not rejected.

5. It is known from the course that

$$g(y) = c\int \frac{dy}{\sqrt{h(y)}},$$

where c is an arbitrary constant, and $h(y)$ is such a function that

$$\text{Var}Y = h(EY).$$

In our case $h(y) = y^2$, and

$$g(y) = c\int \frac{dy}{y} = c\ln|y|.$$

Take for simplicity $c = 1$. Then (since $Y > 0$) $g(y) = \ln y$.

6. Matrices R and S are symmetric, idempotent, rank$(R) = 1$, rank$(S) = 1$, and $RS = 0$, therefore (this is known from our course, see FKLM theorem B.8) the ratio

$$\frac{Z^TRZ}{Z^TSZ}$$

has the Fisher distribution with 1 and 1 degrees of freedom. Using statistical tables we obtain

$$P\left(\frac{Z^TRZ}{Z^TSZ} \geq 161.45\right) = 0.05.$$