Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4267 Linear statistical models
Recommended exercises 8 – solutions

Problem 1 One- and two-way ANOVA – and the linear model

```
## a

income <- c(300, 350, 370, 360, 400, 370, 420, 390,
            400, 430, 420, 410, 300, 320, 310, 305,
            350, 370, 340, 355, 370, 380, 360, 365)
gender <- c(rep("Male", 12), rep("Female",12))
place <- rep(c(rep("A",4), rep("B",4), rep("C",4)),2)
data <- data.frame(income, gender, place)
data

pairs(data)
plot(income~place, data=data)
plot(income~gender, data=data)
interaction.plot(data$gender, data$place, data$income)
plot.design(income~place+gender, data = data)

## b

X <- cbind(rep(1,length(data$income)),data$place=="A",
          data$place=="B",data$place=="C")
X
XtX <- t(X)%*%X
qr(XtX)$rank

# We need full rank to invert XtX. Find a design matrix of full rank having
# the same column space:

## c

model <- lm(income~place-1, data=data, x=TRUE)
model$x # design matrix
summary(model)
anova(model)
```

```
# This is a parametrization without intercept, and with three estimated
# effects for place.

## d

options(contrasts=c("contr.treatment", "contr.poly"))
model1 <- lm(income~place, data=data, x=TRUE)
model1$x
summary(model1)
anova(model1)

# Treatment contrast parametrization codes the factor at the lowest level
# (which is A here) as 0, so that the value of the intercept will be the
# estimate for the level \texttt{A}. Compare this with the model above.
model$coeff
model1$coeff

options(contrasts=c("contr.sum", "contr.poly"))
model2 <- lm(income~place, data=data, x=TRUE)
model2$x
summary(model2)
model2$coeff
data$place
# Sum-to-zero contrast parametrization puts the coefficient of C as minus the
# sum of the coefficients for A and B, so that the sum of the coefficients
# for A, B and C is zero.

## e

# Using linear hypothesis - starting with model 1:
r <- 2
C <- cbind(rep(0,r), diag(r))
d <- matrix(rep(0,r), ncol=1)
n <- length (data$income)

betahat <- matrix(model1$coefficients, ncol=1)
sigma2hat <- summary(model1)$sigma^2
X <- model.matrix(model1)

F1 <- (t(C%*%betahat-d)%*%solve(C%*%solve(t(X)%*%X)%*%t(C))%*%
  (C%*%betahat-d))/(r*sigma2hat)
```

```
F1
1-pf(F1,r,n-length(betahat))
# Same as anova(model1) above

betahat <- matrix(model2$coefficients, ncol=1)
sigma2hat <- summary(model2)$sigma^2
X <- model.matrix(model2)

F2 <- (t(C%*%betahat-d)%*%solve(C%*%solve(t(X)%*%X)%*%t(C))%*%
  (C%*%betahat-d))/(r*sigma2hat)
F2
1-pf(F2,r,n-length(betahat))
# Same result of hypothesis test.

# What about the no intercept that was in b) (not asked for)?
r <- 2
C <- matrix(c(1,-1,0,0,1,-1), ncol=3, byrow=TRUE)
C
d <- matrix(rep(0,r), ncol=1)

betahat <- matrix(model$coefficients, ncol=1)
sigma2hat <- summary(model)$sigma^2
X <- model.matrix(model)

F <- (t(C%*%betahat-d)%*%solve(C%*%solve(t(X)%*%X)%*%t(C))%*%
  (C%*%betahat-d))/(r*sigma2hat)
F
1-pf(F,r,n-length(betahat))
# This also gives the same result.

## f

options(contrasts=c("contr.treatment", "contr.poly"))
model3 <- lm(income~place+gender, data=data, x=TRUE)
model3$x
anova(model3)
summary(model3)

options(contrasts=c("contr.sum", "contr.poly"))
model4 <- lm(income~place+gender, data=data, x=TRUE)
model4$x
```

```
summary(model4)
anova(model4)

# Testing the place effect in model 4, and then the gender effect:
betahat <- matrix(model4$coefficients,ncol=1)
sigma2hat <- summary(model4)$sigma^2
X <- model.matrix(model4)

r <- 2
Cplace <- cbind(rep(0,r), diag(r), rep(0,r)) # gender coeff. last column
d <- matrix(rep(0,r), ncol=1)

Fplace <- (t(Cplace%*%betahat-d)%*%
  solve(Cplace%*%solve(t(X)%*%X)%*%t(Cplace))%*%
  (Cplace%*%betahat-d))/(r*sigma2hat)
Fplace
1-pf(Fplace,r,n-length(betahat))

# There's no need to test the significance of gender, since only one
# parameter can be read off of the summary. This gives the same result as
# using anova(model4).
options(contrasts=c("contr.sum", "contr.poly"))
model5 <- lm(income~place*gender, data=data, x=TRUE)
summary(model5)
X <- model5$x
anova(model5)
# The interaction is not significant. Now perform the same test (significance
# of place:gender interaction, given that all main effects are in the model)
# using the C beta = d approach:
r <- 2
Cinteract <- cbind(rep(0,r),rep(0,r),rep(0,r),rep(0,r),diag(r))
d <- matrix(rep(0,r),ncol=1)

betahat <- model5$coefficients
betahat
Cinteract%*%betahat
sigma2hat <- summary(model5)$sigma^2
Finteract <- (t(Cinteract%*%betahat-d)%*%solve(Cinteract%*%
  solve(t(X)%*%X)%*%t(Cinteract))%*%
  (Cinteract%*%betahat-d))/(r*sigma2hat)
Finteract
```

```
1-pf(Finteract,r,n-length(betahat))
# This gives the same result as above. Finally, repeat the same test using
# dummy variable coding (contr.treatment).
options(contrasts=c("contr.treatment", "contr.poly"))
model5 <- lm(income~place*gender, data=data, x=TRUE)
summary(model5)
X <- model5$x
anova(model5)
r <- 2
Cinteract <- cbind(rep(0,r),rep(0,r),rep(0,r),rep(0,r),diag(r))
d <- matrix(rep(0,r),ncol=1)

betahat <- model5$coefficients
betahat
Cinteract%*%betahat
sigma2hat <- summary(model5)$sigma^2
Finteract <- (t(Cinteract%*%betahat-d)%*%
  solve(Cinteract%*%solve(t(X)%*%X)%*%t(Cinteract))%*%
  (Cinteract%*%betahat-d))/(r*sigma2hat)
Finteract
1-pf(Finteract,r,n-length(betahat))
# This also gives the same result.
```

Problem 2 Teaching reading

a) Let μ_{A}, μ_{B} and μ_{C} be the expected reading scores for each of the three methods. Then the null hypothesis is H_0: $\mu_{\mathrm{A}} = \mu_{\mathrm{B}} = \mu_{\mathrm{C}}$, and the alternative H_1: at least one differs from the others.

For one-way ANOVA, the *treatment sum of squares* (what we have called the *regression sum of squares* in general) is in the present case (with obvious notation) $\mathrm{SSR} = n_{\mathrm{A}}(\bar{x}_{\mathrm{A}} - \bar{x})^2 + n_{\mathrm{B}}(\bar{x}_{\mathrm{B}} - \bar{x})^2 + n_{\mathrm{C}}(\bar{x}_{\mathrm{C}} - \bar{x})^2 = 22 \cdot (41.05 - 44.02)^2 + 22 \cdot (46.73 - 44.02)^2 + 22 \cdot (44.27 - 44.02)^2 = 357.005$, and the error sum of squares is $\mathrm{SSE} = (n_{\mathrm{A}} - 1)s_{\mathrm{A}}^2 + (n_{\mathrm{B}} - 1)s_{\mathrm{B}}^2 + (n_{\mathrm{C}} - 1)s_{\mathrm{C}}^2 = 21 \cdot 5.636^2 + 21 \cdot 7.388^2 + 21 \cdot 5.767^2 = 2511.712$.

The model under the null hypothesis is a model with only an intercept. The error sum of squares in this model, SSE_0, is the same as the total sum of squares, SST, so $\mathrm{SSE}_0 - \mathrm{SSE} = \mathrm{SST} - \mathrm{SSE} = \mathrm{SSR}$, and the usual test statistic for testing H_0 in terms of the restricted model defined by H_0 becomes

$$F = \frac{(\mathrm{SSE}_0 - \mathrm{SSE})/r}{\mathrm{SSE}/(n-p)} = \frac{\mathrm{SSR}/r}{\mathrm{SSE}/(n-p)},$$

which is F-distributed with r and $n-p$ degrees of freedom under H_0. We get the value

$$\frac{357.005/2}{2511.712/(66-3)} = \frac{178.50}{39.87} = 4.477,$$

since we have $n = 66$ observations, and $p = 3$ covariates in a full-rank design matrix, and H_0 can be described in the form $C\boldsymbol{\beta} = \mathbf{0}$ with $r = \operatorname{rank} C = 2$. The p-value is $P(F \geq 4.477) = 0.015$ (df = 2 and 63), so at the 0.05 level we reject the null hypothesis and conclude that the teaching method matters (the expected reading score is not the same for all the methods).

Note: Traditionally, the computations for such a test is summarized in a so-called ANOVA table:

Source of variation	Sum of squares	Degrees of freedom	Mean square	Computed f
Treatments	357.005	2	178.50	4.477
Error	2511.712	63	39.87	
Total	2868.717	65		

The assumptions made is that $X_{ij} = \mu + \alpha_i + \epsilon_{ij}$ for $i = 1, 2, 3$ and $j = 1, 2, \ldots, 22$, where X_{ij} is the reading score for subject j receiving teaching method i (where $i = 1, 2, 3$ correspond to method A, B, C, respectively), where μ, α_1, α_2 and α_3 are parameters and the ϵ_{ij} are independent and normally distributed with mean zero and the same variance for all observations.

b) A natural estimator for γ is $\hat{\gamma} = \bar{X}_{\mathrm{B}}/\bar{X}_{\mathrm{C}}$, where $\bar{X}_{\mathrm{B}}$ is the mean of the sample receiving method B and $\bar{X}_{\mathrm{C}}$ the mean of the sample receiving method C.

Let h be the function defined by $h(x, y) = x/y$. The first-order Taylor approximation of h at $(\mu_{\mathrm{B}}, \mu_{\mathrm{C}})$ is then given by

$$\begin{aligned}\frac{x}{y} = h(x,y) &\approx h(\mu_{\mathrm{B}}, \mu_{\mathrm{C}}) + D_1 h(\mu_{\mathrm{B}}, \mu_{\mathrm{C}})(x - \mu_{\mathrm{B}}) + D_2 h(\mu_{\mathrm{B}}, \mu_{\mathrm{C}})(y - \mu_{\mathrm{C}}) \\ &= \frac{\mu_{\mathrm{B}}}{\mu_{\mathrm{C}}} + \frac{1}{\mu_{\mathrm{C}}}(x - \mu_{\mathrm{B}}) - \frac{\mu_{\mathrm{B}}}{\mu_{\mathrm{C}}^2}(y - \mu_{\mathrm{C}}),\end{aligned}$$

where D_1 and D_2 denote partial differentiation with respect to the first and the second variable, respectively.

At $(\bar{X}_{\mathrm{B}}, \bar{X}_{\mathrm{C}})$ this gives

$$\hat{\gamma} = \frac{\bar{X}_{\mathrm{B}}}{\bar{X}_{\mathrm{C}}} \approx \frac{\mu_{\mathrm{B}}}{\mu_{\mathrm{C}}} + \frac{1}{\mu_{\mathrm{C}}}(\bar{X}_{\mathrm{B}} - \mu_{\mathrm{B}}) - \frac{\mu_{\mathrm{B}}}{\mu_{\mathrm{C}}^2}(\bar{X}_{\mathrm{C}} - \mu_{\mathrm{C}}),$$

so that

$$\begin{aligned} E\hat{\gamma} &\approx \frac{\mu_{\mathrm{B}}}{\mu_{\mathrm{C}}} + \frac{1}{\mu_{\mathrm{C}}}(E\bar{X}_{\mathrm{B}} - \mu_{\mathrm{B}}) - \frac{\mu_{\mathrm{B}}}{\mu_{\mathrm{C}}^2}(E\bar{X}_{\mathrm{C}} - \mu_{\mathrm{C}}) = \frac{\mu_{\mathrm{B}}}{\mu_{\mathrm{C}}}, \\ \operatorname{Var}\hat{\gamma} &\approx \frac{1}{\mu_{\mathrm{C}}^2}\operatorname{Var}\bar{X}_{\mathrm{B}} + \frac{\mu_{\mathrm{B}}^2}{\mu_{\mathrm{C}}^4}\operatorname{Var}\bar{X}_{\mathrm{C}} = \frac{1}{\mu_{\mathrm{C}}^2}\frac{\sigma_{\mathrm{B}}^2}{n_{\mathrm{B}}} + \frac{\mu_{\mathrm{B}}^2}{\mu_{\mathrm{C}}^4}\frac{\sigma_{\mathrm{C}}^2}{n_{\mathrm{C}}}, \\ \operatorname{SD}\hat{\gamma} &= \frac{1}{\mu_{\mathrm{C}}}\sqrt{\frac{\sigma_{\mathrm{B}}^2}{n_{\mathrm{B}}} + \frac{\mu_{\mathrm{B}}^2}{\mu_{\mathrm{C}}^2}\frac{\sigma_{\mathrm{C}}^2}{n_{\mathrm{C}}}}. \end{aligned}$$

(We still assume that the two random samples are independent, but we allow the corresponding teaching methods to have different variances.) We have estimates

$$\widehat{E\hat{\gamma}} = \frac{\bar{x}_{\mathrm{B}}}{\bar{x}_{\mathrm{C}}} \qquad \text{and} \qquad \widehat{\operatorname{SD}\hat{\gamma}} = \frac{1}{\bar{x}_{\mathrm{C}}}\sqrt{\frac{s_{\mathrm{B}}^2}{n_{\mathrm{B}}} + \frac{\bar{x}_{\mathrm{B}}^2}{\bar{x}_{\mathrm{C}}^2}\frac{s_{\mathrm{C}}^2}{n_{\mathrm{C}}}}.$$

With numbers from the table we get

$$\hat{\gamma} = \widehat{E\hat{\gamma}} = \frac{46.73}{44.27} = 1.06 \qquad \text{and} \qquad \widehat{\operatorname{SD}\hat{\gamma}} = \frac{1}{44.27}\sqrt{\frac{7.388^2}{22} + \frac{46.73^2}{44.27^2}\cdot\frac{5.767^2}{22}} = 0.046.$$