Answer sheet 9

Assignment 1. Define the random vector Y as suggested. Then the covariance matrix of Y is XX^T : indeed, the covariance of Y_k and Y_j is $\mathbb{E}Y_kY_j$ (since $\mathbb{E}Y = p^{-1}\sum x_i = 0$) and this equals $p^{-1}\sum_{i=1}^d X_{ki}X_{ji} = (XX^T)_{kj}$. Let $H : \mathbb{R}^d \to \mathbb{R}^d$ be the projection onto the span of the top k eigenvectors of $XX^T = cov(Y)$ and let Q be another rank k projection. Then by the optimal reduction theorem (slide 295) we know that

$$\frac{1}{p}\sum_{i=1}^p \|x_i - Hx_i\|^2 = \mathbb{E}\|Y - HY\|^2 \le \mathbb{E}\|Y - QY\|^2 = \frac{1}{p}\sum_{i=1}^p \|x_i - Qx_i\|^2.$$

Assignment 2.

a) The assumptions imply that A is injective. If $v \in \mathbb{R}^p \setminus \{0\}$ then

$$v^T Bv = v^T A^T \Omega Av = (Av)^T \Omega(Av) > 0$$

since $Av \neq 0$ and Ω is positive definite. Thus B is positive definite and in particular invertible. The special case $\Omega = I_n$ shows that $A^T A$ is strictly positive definite.

b) Choose $A = (1, 1)^T$ and

$$\Omega = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

then $A^T \Omega A = 0$.

Remark. If Ω has one positive and one negative eigenvalues, we can always find an injective A such that $A^T \Omega A = 0$.

Assignment 3. We shall use the following fact. If X_1 and X_2 are indepndent, and Y_1 and X_2 are independent, and X_1 and Y_1 have the same distribution, the for any (measurable) function g, $g(X_1, X_2)$ and $g(Y_1, X_2)$ have the same distribution.

(i) Take $c^T = (1, 0)$

(ii) Take $c^T = (0, 1)$ and use (i).

(iii) Take $c^T = (-1, 0)$, to get $-X \sim X$, so that $\mathbb{E} - X = \mathbb{E}X$, and then $\mathbb{E}X = 0$.

(iv) Take $c^T = (1, 1)/\sqrt{2}$. Then

$$X \sim (X + Y)/\sqrt{2} \sim (X_1 + X_2)/\sqrt{2}.$$

(v) We know that this is true for $n = 1, 2$. Suppose that this is true for n and write

$$(X_1 + \cdots + X_{n+1})/\sqrt{n+1} = \sqrt{n/(n+1)}[(X_1 + \cdots + X_n)/\sqrt{n}] + \sqrt{1/(n+1)}X_{n+1}.$$

This has the same distribution as $\sqrt{n/(n+1)}X + \sqrt{1/(n+1)}Y$ by the induction hypothesis. Now choose $c^T = (\sqrt{n}, 1)/\sqrt{n+1}$

(vi) Since X has zero mean and finite variance σ^2, by the central limit theorem

$$(X_1 + \cdots + X_n)/\sqrt{n} \xrightarrow{d} N(0, \sigma^2).$$

By (v) this gives $X \sim N(0, \sigma^2)$, and by (ii) $Y \sim N(0, \sigma^2)$.

If $U \sim U[0,1]$ and

$$(X,Y) = (\cos(2\pi U), \sin(2\pi U)),$$

then by symmetry $c^T(X,Y)$ has the same distribution for all $c \in S^1$ but X and Y are not Gaussian. This is the uniform distribution on the unit circle.

Assignment 4.
Here an example of solutions (but they are not unique). The parameters that have been fixed are underlined (e.g. $\underline{\beta_0}$).

a) $y = (1 \quad \frac{1}{x} \quad \frac{1}{x^2})(\beta_0 \quad \beta_1 \quad \beta_2)^{\mathrm{T}} + \varepsilon$. It must be $x \neq 0$.

b) $y = (\frac{1}{1+\underline{\beta_1}x})(\beta_0) + \varepsilon$. It must be $\beta_1 x \neq -1$.

c) $y = (1/x)(\gamma) + \varepsilon$ with $\gamma = \beta_0/\beta_1$ where $y = (\frac{1}{x\underline{\beta_1}})(\beta_0) + \varepsilon$. It must be $x \neq 0$ and that $\beta_1 \neq 0$.

d) $1/y = (1 \quad x)(\beta_0 \quad \beta_1)^{\mathrm{T}} + \varepsilon$. It must be $y \neq 0$.

e) $y = (1 \quad x_1^{\underline{\beta_2}} \quad x_2^{\underline{\beta_4}})(\beta_0 \quad \beta_1 \quad \beta_3)^{\mathrm{T}} + \varepsilon$. It must be that $x_1^{\beta_2}$ is well defined, which will happen for example if
   - $x_1 > 0$, $\underline{\beta_2} \in \mathbb{R}$ (possibly the most natural condition)
   - $x_1 \geq 0$, $\underline{\beta_2} > 0$
   - $x_1 \in \mathbb{R}$, $\underline{\beta_2} = 1, 2, \ldots$
   - $x_1 \neq 0$, $\underline{\beta_2} \in \mathbb{Z}$

   Similar conditions are necessary for $x_2^{\beta_4}$ to be well defined.

f) $\log(y) = (1 \quad \log(x_1) \quad \log[\cos(x_2)])(\log(\beta_1) \quad \beta_2 \quad \beta_3)^{\mathrm{T}} + \log(\varepsilon)$ with $x_1, \beta_1, \varepsilon > 0$ et $\cos(x_2) > 0$.

g) $\log(y - \underline{\beta_1}) = (\log(x_1) \quad \log[2 + \cos(x_2)])(\beta_2 \quad \beta_3)^{\mathrm{T}} + \log(\varepsilon^2 + 1)$ with $x_1 > 0$ et $y > \underline{\beta_1}$.

h) $y = (1 \quad \cos(x + \underline{\beta_2}))(\beta_0 \quad \beta_1)^{\mathrm{T}} + \varepsilon$ ou $y = (1 \quad \cos x \quad \sin x)(\beta_0 \quad \beta_1 \cos\beta_2 \quad -\beta_1 \sin\beta_2)^{\mathrm{T}} + \varepsilon$, where we have used $\cos(x+\beta_2) = \cos x \cos\beta_2 - \sin x \sin\beta_2$.

Assignment 5.

$$X_{y\sim a-1} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \end{pmatrix}; \quad X_{y\sim a+b} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & 3 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$$

Assignment 6.

a) The parameters α_1 and β_1 are only influenced by the rats in the first group, while α_2 and β_2 are only influenced by the second group. Thus

$$\begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \end{bmatrix}, \qquad \text{for group 1,}$$

$$\begin{bmatrix} y_{12} \\ y_{22} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & x_{12} \\ 1 & x_{22} \\ 1 & x_{32} \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \end{bmatrix}, \qquad \text{for group 2.}$$

The model for the two groups toghether is obtained by combining the two previous models into

$$y = \begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{12} \\ y_{22} \\ y_{32} \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & 0 & x_{11} & 0 \\ 1 & 0 & x_{21} & 0 \\ 1 & 0 & x_{31} & 0 \\ 0 & 1 & 0 & x_{12} \\ 0 & 1 & 0 & x_{22} \\ 0 & 1 & 0 & x_{32} \end{bmatrix}, \qquad \beta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \qquad \varepsilon = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \end{bmatrix}.$$

b) The models assume that (i) $\beta_1 = \beta_2$, (ii) $\alpha_1 = \alpha_2$ et (iii) $\alpha_1 = \alpha_2$ et $\beta_1 = \beta_2$. In order to fulfill these assumptions we need to fix some parameters to 0, hence we should re-write the model using $\alpha_2 - \alpha_1$ and $\beta_2 - \beta_1$ as parameters. We can for example write the model for group 2 in terms of the difference w.r.t. group 1 :

$$\begin{bmatrix} y_{12} \\ y_{22} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & x_{12} \\ 1 & x_{22} \\ 1 & x_{32} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} 1 & x_{12} \\ 1 & x_{22} \\ 1 & x_{32} \end{bmatrix} \begin{bmatrix} \alpha_2 - \alpha_1 \\ \beta_2 - \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \end{bmatrix}.$$

With this formulation, the parameters α_1 and β_1 are now common to the two groups, and the new parameters $\alpha_2 - \alpha_1$ and $\beta_2 - \beta_1$ represent the difference between the groups.

The model for the two groups combined is written as :

$$y_{jg} = \mu + \mu_d \delta_{2g} + (\gamma + \gamma_d \delta_{2g}) x_{jg} + \varepsilon_{jg}, \quad j = 1, 2, 3 \quad g = 1, 2,$$

with design matrix and parameters vector

$$X = \begin{bmatrix} 1 & 0 & x_{11} & 0 \\ 1 & 0 & x_{21} & 0 \\ 1 & 0 & x_{31} & 0 \\ 1 & 1 & x_{12} & x_{12} \\ 1 & 1 & x_{22} & x_{22} \\ 1 & 1 & x_{32} & x_{32} \end{bmatrix}, \qquad \beta = \begin{bmatrix} \mu \\ \mu_d \\ \gamma \\ \gamma_d \end{bmatrix} \quad \left(= \begin{bmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \beta_1 \\ \beta_2 - \beta_1 \end{bmatrix} \right).$$

We have used the indicator function δ_{2g} that takes value 1 for the group 2 and 0 otherwise. It is represented by the second column of X above.

The submodels assume (i) $\gamma_d = 0$, (ii) $\mu_d = 0$, (iii) $\mu_d = \gamma_d = 0$ and thus we suppress the following columns X : (i) 4, (ii) 2, (iii) 2 et 4.

Assignment 7.

a) If $Y \in \mathbb{R}^n$ follows a multivariate normal $N_n(X\beta, \sigma^2 I)$, then $Z = Q^{\mathrm{T}} Y$ follows a multivariate Normal distribution with expected value

$$\mathbb{E}(Z) = Q^{\mathrm{T}} X\beta = Q^{\mathrm{T}} QR\beta = R\beta$$

and covariance

$$\operatorname{Var} Z = \sigma^2 Q^{\mathrm{T}} Q = \sigma^2 I.$$

b) By direct computation and using the fact that Q is orthogonal we find

$$\begin{aligned} u = Q^T\hat{y} &= Q^T H y \\ &= Q^T X(X^T X)^{-1} X^T y \\ &= Q^T QR((QR)^T QR)^{-1}(QR)^T y \\ &= Q^T QR(R^T Q^T QR)^{-1} R^T Q^T y \\ &= R(R^T R)^{-1} R^T Q^T y \\ &= \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \left(\begin{bmatrix} R_1^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} R_1^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix} y \\ &= \begin{bmatrix} R_1 \\ 0 \end{bmatrix} (R_1^{\mathrm{T}} R_1)^{-1} R_1^{\mathrm{T}} Q_1^{\mathrm{T}} y \\ &= \begin{bmatrix} R_1 \\ 0 \end{bmatrix} R_1^{-1} R_1^{-\mathrm{T}} R_1^{\mathrm{T}} Q_1^{\mathrm{T}} y \\ &= \begin{bmatrix} R_1 R_1^{-1} Q_1^{\mathrm{T}} y \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} Q_1^{\mathrm{T}} y \\ 0 \end{bmatrix}, \end{aligned}$$

where R_1 is invertible since it's upper triangular with positive diagonal elements.

c) Since $e = y - \widehat{y}$, we find

$$\begin{aligned} v &= Q^{\mathrm{T}} e = Q^{\mathrm{T}}(y - \hat{y}) = Q^{\mathrm{T}} y - Q^{\mathrm{T}} \hat{y} \\ &= \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix} y - \begin{bmatrix} Q_1^{\mathrm{T}} y \\ 0 \end{bmatrix} = \begin{bmatrix} Q_1^{\mathrm{T}} y \\ Q_2^{\mathrm{T}} y \end{bmatrix} - \begin{bmatrix} Q_1^{\mathrm{T}} y \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ Q_2^{\mathrm{T}} y \end{bmatrix}, \end{aligned}$$

where we have used part (b).

d) By (a) it follows $\operatorname{Var} Z = \sigma^2 I$ is diagonal, $z_1 = Q_1^{\mathrm{T}} y \in \mathbb{R}^p$ and $z_2 = Q_2^{\mathrm{T}} y \in \mathbb{R}^{n-p}$ are independent since they are marginals of z

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} Q_1^{\mathrm{T}} y \\ Q_2^{\mathrm{T}} y \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

where $u_1 \in \mathbb{R}^p$ and $v_2 \in \mathbb{R}^{n-p}$ are the non zero components of u and v :

$$u = \begin{bmatrix} Q_1^{\mathrm{T}} y \\ 0 \end{bmatrix} = \begin{bmatrix} u_1 \\ 0 \end{bmatrix}, \qquad v = \begin{bmatrix} 0 \\ Q_2^{\mathrm{T}} y \end{bmatrix} = \begin{bmatrix} 0 \\ v_2 \end{bmatrix}.$$

We are going to show that S^2 is a function of z_2 and $\widehat{\beta}$ is a function of z_1. This will conclude the proof.
Since $v = Q^{\mathrm{T}} e$, we have $e = Qv$. Then

$$\begin{aligned} S^2 &= \frac{1}{n-p} e^{\mathrm{T}} e = \frac{1}{n-p} v^{\mathrm{T}} Q^{\mathrm{T}} Q v = \frac{1}{n-p} v^{\mathrm{T}} v \\ &= \frac{1}{n-p} \begin{bmatrix} 0 & v_2^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} 0 \\ v_2 \end{bmatrix} = \frac{1}{n-p} v_2^{\mathrm{T}} v_2 = \frac{1}{n-p} z_2^{\mathrm{T}} z_2, \end{aligned}$$

is a function of z_2. A similar computation to the one in part (b) yields that

$$\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y = R_1^{-1}Q_1^{\mathrm{T}}y = R_1^{-1}u_1 = R_1^{-1}z_1,$$

is a function of z_1. The proof is done since

$$z_1 \text{ indep } z_2 \quad \Longrightarrow \quad u_1 \text{ indep } v_2 \quad \Longrightarrow \quad \hat{\beta} \text{ indep } S^2.$$

Indeed, measurable functions of independent variables stay independent : here $S^2 = f(v_2)$ and $\widehat{\beta} = g(u_1)$, hence for every $B, C \subseteq \mathbb{R}$ (mesurable),

$$\begin{aligned}
&\mathbb{P}(S^2 \in B, \widehat{\beta} \in C) \\
=&\mathbb{P}(v_2 \in f^{-1}(B), u_1 \in g^{-1}(C)) \\
=&\mathbb{P}(v_2 \in f^{-1}(B))\mathbb{P}(u_1 \in g^{-1}(C)) \\
=&\mathbb{P}(S^2 \in B)\mathbb{P}(\widehat{\beta} \in C).
\end{aligned}$$