DTU

Advanced Data Analysis and Statistical Modelling

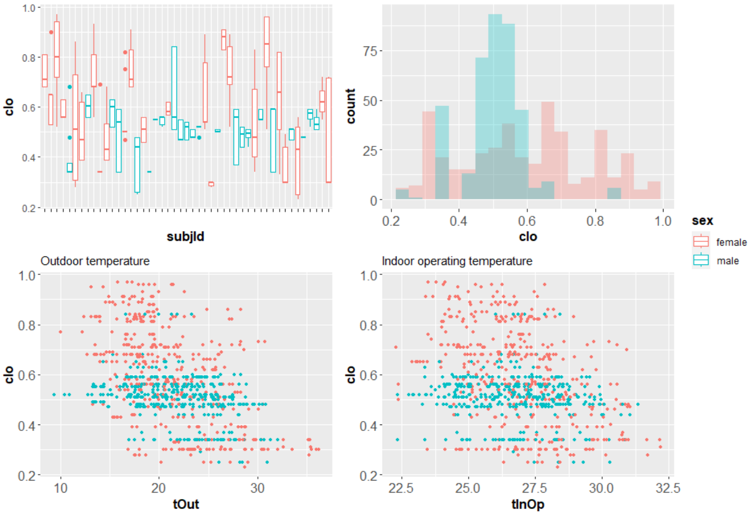
Exam - Part 1

Søren Skjernaa - s223316
Tymoteusz Barcinski - s221937

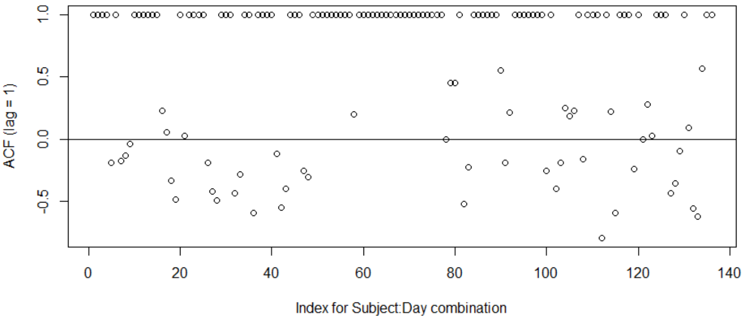
Overview of the experiment

Variable	Type	Explanation
clo	Clothing insulation level	Positive variable, with higher values implying higher insulation.
tOut	Outdoors air temperature	Measured in C^o
tInOp	Indoor operating temperature	Measured in C^o
sex	Sex	Female/male.
subjId	Subject ID	Uniqiue ID for each subject.
time	Time	Time difference since last observation for the subject (continues variable, but unit not given).
day	Day	Number of experimentation day for the subject.
subDay	Subject $\times$ Day ID	Unique ID for each combination of subject and day.

Overview of the experiment



Overview of the experiment



Modelling choices

Linear mixed effects model $Y = \log(clo) = X\beta + ZU + \epsilon$,
where $\epsilon \sim N(0, \Sigma)$ and $U \sim N(0, \Psi)$.
For observation $i = 1, 2, \dots, 803$ we have $Y_i = \log(clo_i) = \mu_i + \phi_i + \epsilon_i$.

Modelling choices

Linear mixed effects model $Y = \log(clo) = X\beta + ZU + \epsilon$,
where $\epsilon \sim N(0, \Sigma)$ and $U \sim N(0, \Psi)$.
For observation $i = 1, 2, \ldots, 803$ we have $Y_i = \log(clo_i) = \mu_i + \phi_i + \epsilon_i$.

Fixed effects structure:

$$\begin{aligned}
\mu_i = \beta_0 &+ \beta_1(\text{sex}_i) \\
&+ \beta_2(\text{sex}_i) \cdot \text{tOut}_i + \beta_3(\text{sex}_i) \cdot \text{tOut}_i^2 \\
&+ \beta_4(\text{sex}_i) \cdot \text{tInOp}_i + \beta_5(\text{sex}_i) \cdot \text{tInOp}_i^2 \\
&+ \beta_6(\text{sex}_i) \cdot \text{tOut}_i \cdot \text{tInOp}_i \\
&+ \beta_7(\text{sex}_i) \cdot \text{tOut}_i \cdot \text{tInOp}_i^2 + \beta_8(\text{sex}_i) \cdot \text{tOut}_i^2 \cdot \text{tInOp}_i \\
&+ \beta_9(\text{sex}_i) \cdot \text{tOut}_i^2 \cdot \text{tInOp}_i^2.
\end{aligned}$$

Modelling choices

Linear mixed effects model $Y = \log(clo) = X\beta + ZU + \epsilon$,
where $\epsilon \sim N(0, \Sigma)$ and $U \sim N(0, \Psi)$.
For observation $i = 1, 2, \ldots, 803$ we have $Y_i = \log(clo_i) = \mu_i + \phi_i + \epsilon_i$.

Random effects structure - Model 1:

$$\begin{aligned}\phi_i = &u_1(\text{subjId}_i) + \\ &u_2(\text{subjId}_i) \cdot \text{tOut}_i + u_3(\text{subjId}_i) \cdot \text{tOut}_i^2 + \\ &u_4(\text{subjId}_i) \cdot \text{tInOp}_i + u_5(\text{subjId}_i) \cdot \text{tInOp}_i^2\end{aligned}$$

$$\begin{pmatrix} u_1(\text{subjId}_i) \\ u_2(\text{subjId}_i) \\ u_3(\text{subjId}_i) \\ u_4(\text{subjId}_i) \\ u_5(\text{subjId}_i) \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_{u_1}^2 & \sigma_{u_1u_2} & \sigma_{u_1u_3} & \sigma_{u_1u_4} & \sigma_{u_1u_5} \\ \sigma_{u_1u_2} & \sigma_{u_2}^2 & \sigma_{u_2u_3} & \sigma_{u_2u_4} & \sigma_{u_2u_5} \\ \sigma_{u_1u_3} & \sigma_{u_2u_3} & \sigma_{u_3}^2 & \sigma_{u_3u_4} & \sigma_{u_3u_5} \\ \sigma_{u_1u_4} & \sigma_{u_2u_4} & \sigma_{u_3u_4} & \sigma_{u_4}^2 & \sigma_{u_4u_5} \\ \sigma_{u_1u_5} & \sigma_{u_2u_5} & \sigma_{u_3u_5} & \sigma_{u_4u_5} & \sigma_{u_5}^2 \end{pmatrix}\right)$$

Modelling choices

Linear mixed effects model $Y = \log(clo) = X\beta + ZU + \epsilon$,
where $\epsilon \sim N(0, \Sigma)$ and $U \sim N(0, \Psi)$.
For observation $i = 1, 2, \ldots, 803$ we have $Y_i = \log(clo_i) = \mu_i + \phi_i + \epsilon_i$.

Random effects structure - Model 2:

$$\begin{aligned}\phi_i =\,& u_1(\text{subjId}_i) + u_2(\text{subjId}_i)\cdot \text{tOut}_i + u_3(\text{subjId}_i)\cdot \text{tInOp}_i \\ &+ v_1(\text{subDay}_i) + v_2(\text{subDay}_i)\cdot \text{tOut}_i + v_3(\text{subDay}_i)\cdot \text{tInOp}_i.\end{aligned}$$

$$U_i = \begin{pmatrix} u_1(\text{subjId}_i) \\ u_2(\text{subjId}_i) \\ u_3(\text{subjId}_i) \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_{u_1}^2 & \sigma_{u_1u_2} & \sigma_{u_1u_3} \\ \sigma_{u_1u_2} & \sigma_{u_2}^2 & \sigma_{u_2u_3} \\ \sigma_{u_1u_3} & \sigma_{u_2u_3} & \sigma_{u_3}^2 \end{pmatrix}\right)$$

$$V_i = \begin{pmatrix} v_1(\text{subDay}_i) \\ v_2(\text{subDay}_i) \\ v_3(\text{subDay}_i) \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_{v_1}^2 & \sigma_{v_1v_2} & \sigma_{v_1v_3} \\ \sigma_{v_1v_2} & \sigma_{v_2}^2 & \sigma_{v_2v_3} \\ \sigma_{v_1v_3} & \sigma_{v_2v_3} & \sigma_{v_3}^2 \end{pmatrix}\right)$$

Modelling choices

Linear mixed effects model $Y = \log(clo) = X\beta + ZU + \epsilon$,
where $\epsilon \sim N(0, \Sigma)$ and $U \sim N(0, \Psi)$.
For observation $i = 1, 2, \ldots, 803$ we have $Y_i = \log(clo_i) = \mu_i + \phi_i + \epsilon_i$.

Random effects structure - Model 3:

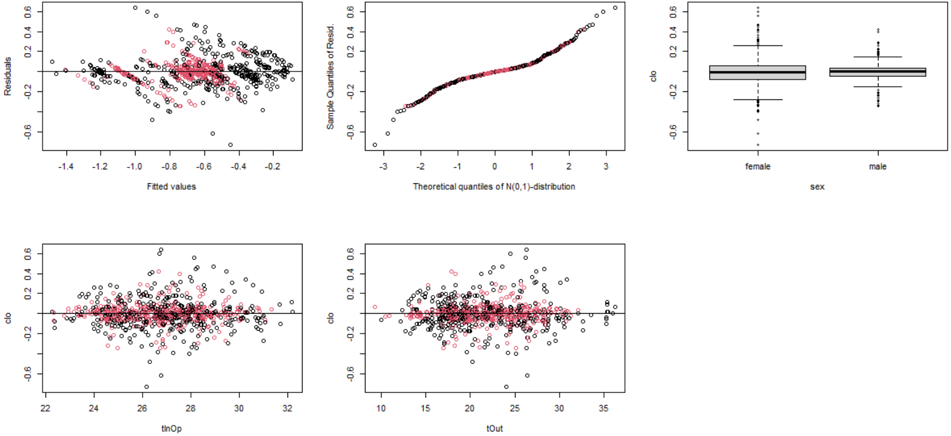
$$Y_i = \log(\text{clo}_i) \sim N(\mu_i, V)$$

and

$$\begin{aligned}
\text{subDay}_{i_1} \neq \text{subDay}_{i_2} &\Rightarrow V_{i_1 i_2} = 0 \\
\text{subDay}_{i_1} = \text{subDay}_{i_2} &\Rightarrow \\
V_{i_1 i_2} &= \sigma^2_{\text{subDay}} + \sigma^2 \cdot \lambda(t_{i_1} - t_{i_2})
\end{aligned}$$

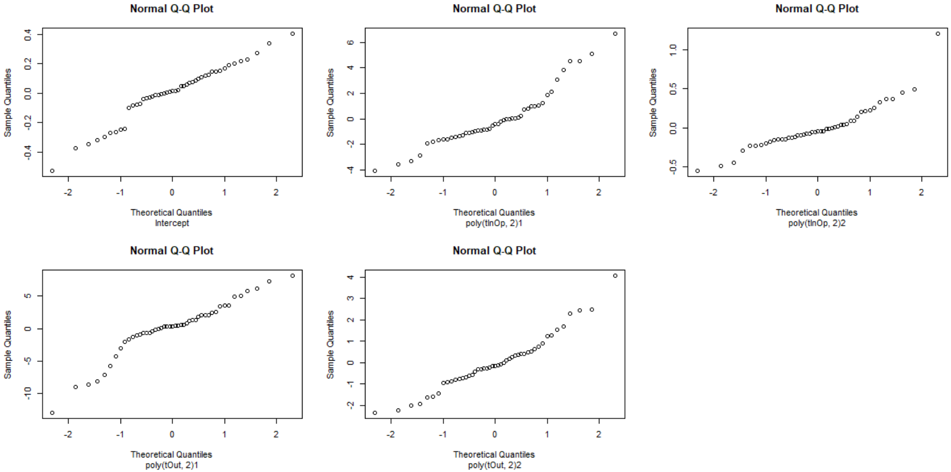
Model 1 - Only random subject effect

Model diagnostics for initial model



Model 1 - Only random subject effect

Model diagnostics for initial model



Model 1 - Only random subject effect

Final model equation:

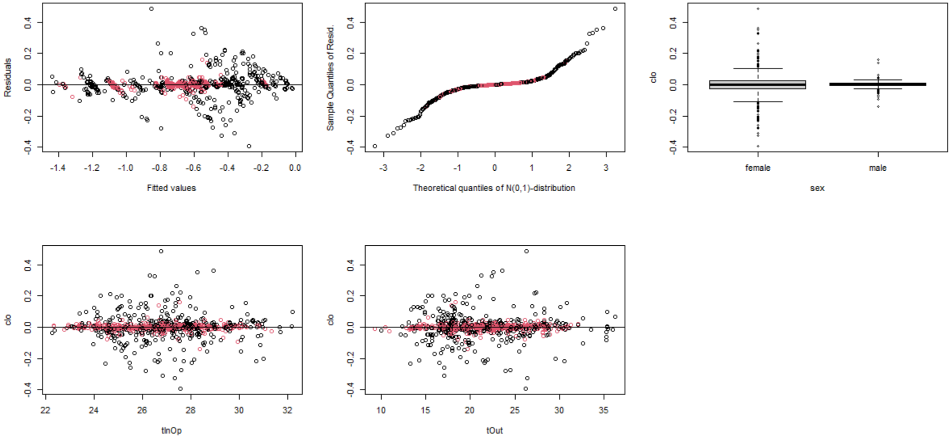
$$
\begin{aligned}
Y_i = \log(\text{clo}_i) = &\beta_0 + \beta_1(\text{sex}_i) + \beta_2(\text{sex}_i) \cdot \text{tOut}_i + \beta_3(\text{sex}_i) \cdot \text{tOut}_i^2 + \\
&\beta_4(\text{sex}_i) \cdot \text{tInOp}_i + \\
&u_1(\text{subjId}_i) + u_2(\text{subjId}_i) \cdot \text{tOut}_i + u_3(\text{subjId}_i) \cdot \text{tOut}_i^2 + \\
&u_4(\text{subjId}_i) \cdot \text{tInOp}_i + \\
&\epsilon_i
\end{aligned}
$$

$$
\begin{pmatrix} u_1(\text{subjId}_i) \\ u_2(\text{subjId}_i) \\ u_3(\text{subjId}_i) \\ u_4(\text{subjId}_i) \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_{u_1}^2 & \sigma_{u_1 u_2} & \sigma_{u_1 u_3} & \sigma_{u_1 u_4} \\ \sigma_{u_1 u_2} & \sigma_{u_2}^2 & \sigma_{u_2 u_3} & \sigma_{u_2 u_4} \\ \sigma_{u_1 u_3} & \sigma_{u_2 u_3} & \sigma_{u_3}^2 & \sigma_{u_3 u_4} \\ \sigma_{u_1 u_4} & \sigma_{u_2 u_4} & \sigma_{u_3 u_4} & \sigma_{u_4}^2 \end{pmatrix}\right) \quad , \quad \epsilon_i \sim N(0, \sigma^2)
$$

Model AIC = -557.84.

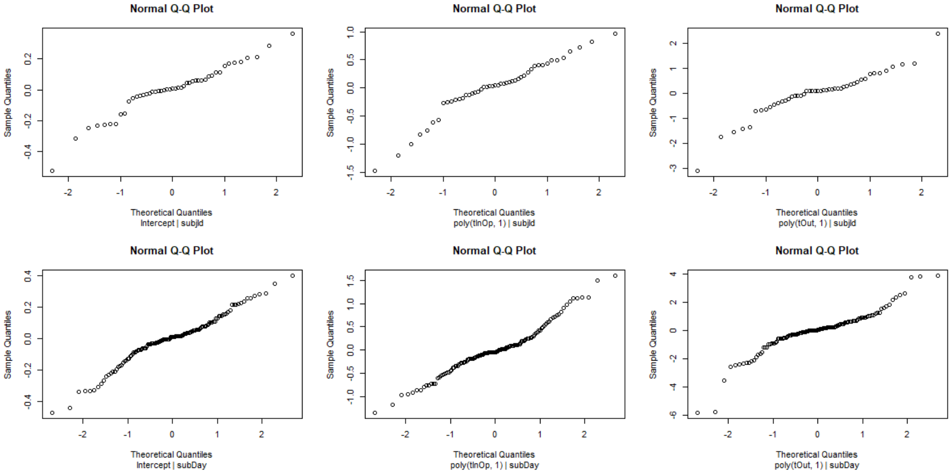
Model 2 - Random subject and random day effect

Model diagnostics for initial model



Model 2 - Random subject and random day effect

Model diagnostics for initial model



Model 2 - Random subject and random day effect

Final model equation:

$$
\begin{aligned}
Y_i = \log(\text{clo}_i) = &\beta_0 + \beta_1(\text{sex}_i) + \beta_2 \cdot \text{tOut}_i + \\
&\beta_4(\text{sex}_i) \cdot \text{tInOp}_i + \beta_5(\text{sex}_i) \cdot \text{tInOp}_i^2 + \\
&u_1(\text{subjId}_i) + u_2(\text{subjId}_i) \cdot \text{tOut}_i + u_3(\text{subjId}_i) \cdot \text{tInOp}_i + \\
&v_1(\text{subDay}_i) + \\
&\epsilon_i
\end{aligned}
$$

$$
U_i = \begin{pmatrix} u_1(\text{subjId}_i) \\ u_2(\text{subjId}_i) \\ u_3(\text{subjId}_i) \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_{u_1}^2 & \sigma_{u_1 u_2} & \sigma_{u_1 u_3} \\ \sigma_{u_1 u_2} & \sigma_{u_2}^2 & \sigma_{u_2 u_3} \\ \sigma_{u_1 u_3} & \sigma_{u_2 u_3} & \sigma_{u_3}^2 \end{pmatrix}\right)
$$

$$
v_1(\text{subDay}_i) \sim N\left(0, \sigma_{v_1}^2\right)
$$

$$
\epsilon_i \sim N\left(0, \sigma^2\right).
$$

Model AIC = -1127.73.

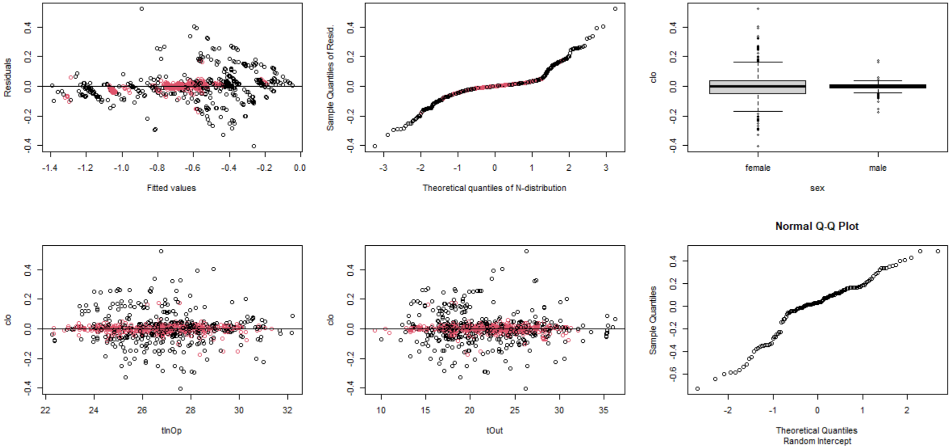
Model 3 - Repeated measurements model

Determining the time correlation structure:

Covariance structure	Correlation term $\lambda(t_{i_1} - t_{i_2})$	AIC$_{re}$	BIC$_{re}$
Gaussian	$\exp\left(\frac{-(t_{i_1} - t_{i_2})^2}{\rho^2}\right)$	-1194.65	-1092.00
Exponential	$\exp\left(\frac{-\lvert t_{i_1} - t_{i_2}\rvert}{\rho}\right)$	-1196.23	-1093.59
Autoregressive(1)	$\rho^{\lvert i_1 - i_2\rvert}$	-1204.14	-1106.16

Model 3 - Repeated measurements model

Model diagnostics for the initial AR(1) model



Model 3 - Repeated measurements model

Reduction of the fixed effects structure using the Likelihood Ratio Test

$$-2\log\left(\frac{\sup_{\theta\in\Theta_0} L(\theta; y)}{\sup_{\theta\in\Theta} L(\theta; y)}\right) \to \chi^2(k-m)$$

under $\mathcal{H}_0$.

Model	N Parameters	AIC	BIC	l	Df	p-value
Initial model	21	-1178.49	-1080.04	610.25		
All three-way-interactions removed	17	-1185.45	-1105.74	609.72	4	0.903
All tInOp $\times$ tOut interactions removed	13	-1193.26	-1132.31	609.63	4	0.996
All sex $\times$ tOut interactions removed	11	-1196.12	-1144.55	609.06	2	0.566
All tOut terms removed	9	-1194.95	-1152.75	606.47	2	0.075
$\beta_5(\text{sex}_i)\cdot\text{tInOp}_i^2$ removed	7	-1197.93	-1165.11	605.97	2	0.602

Model 3 - Repeated measurements model

Final model equation

$$\log(\mathrm{clo}_i) \sim N(\mu_i, V), \qquad \text{where}$$

$$\mu_i = \beta_1(\mathrm{sex}_i) + \beta_4(\mathrm{sex}_i) \cdot \mathrm{tInOp}_i$$

$$V_{i_1,i_2} = \begin{cases} 0 & \text{, if } \mathrm{subDay}_{i_1} \neq \mathrm{subDay}_{i_2} \\ \sigma^2_{\mathrm{subDay}} + \sigma^2 \cdot \rho^{|i_1 - i_2|} & \text{, if } \mathrm{subDay}_{i_1} = \mathrm{subDay}_{i_2} \end{cases}$$

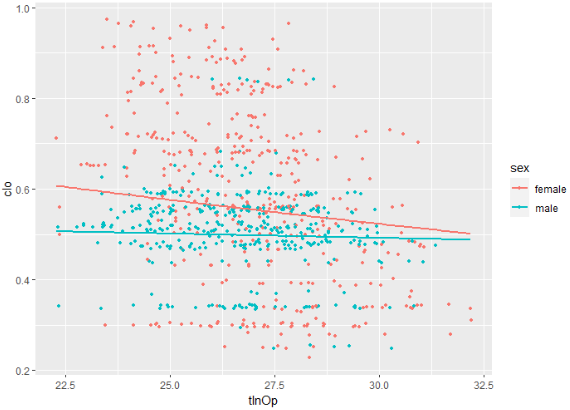
Model AIC = -1197.93.

Model 3 - Repeated measurements model

Parameter	Estimate	Std. Error	Lower CI (2.5 %)	Upper CI (97.5 %)
Intercept β_1 (female)	-0.070	0.109	-0.286	0.146
Intercept β_1 (male)	-0.596	0.113	-0.819	-0.372
Slope β_4 (female)	-0.019	0.004	-0.027	-0.012
Slope β_4 (male)	-0.004	0.004	-0.012	0.004

Std. deviations and correlations	Lower CI (2.5 %)	Estimate	Upper CI (97.5 %)
$\tilde{\sigma}$	0.091	0.101	0.112
$\tilde{\sigma}_{\text{subDay}}$	0.235	0.267	0.303
$\tilde{\rho}$	0.403	0.512	0.606

Model 3 - Repeated measurements model



- **Males:** Intercept = 0.55 (0.44, 0.69), Slope = -0.38% (-1.17%, +0.42%)
- **Females:** Intercept = 0.93 (0.75, 1.16), Slope = -1.9% (-2.7%, -1.2%)

Model 3 - Repeated measurements model