This compulsory assignment consists of 4 exercises. Exercises must be solved in groups (assigned in Canvas) and uploaded to Canvas by the submission deadline. Please provide a single `.html` or `.pdf`-file, entitled `dat320_comp2_groupX.html` (or `.pdf`), where `X` should be replaced by your group number (1-16). The file should be structured into sections (one section per exercise) and subsections (one subsection per task). The usage of R markdown (R package `knitr`) is strongly recommended.

Exercise 1 (Univariate forecasting using Exponential Smoothing & ARIMA)

The dataset `MeatConsumption` describes the yearly meat consumption in Norway between 1990 and 2019. In specific, the meat types *pig*, *poultry*, *beef*, and *sheep* are provided. Data are given in thousand tonnes of carcass weight.

Material provided:

- `MeatConsumption.csv`: dataset

Tasks:

(a) *Load the dataset and perform an exploratory analysis of the meat consumption for each type of meat (missing values, distribution plots, ACF & PACF, KPSS test, etc.). Are the time series stationary? Which ARIMA parameters would you suggest based on the KPSS test and ACF / PACF?*

(b) *Divide the dataset into training (20 years) and testing (9 years). For each meat type, train 4 baseline models (average, drift, naive, seasonal naive) and compute the RMSE on the test set. Further, build an exponential smoothing and an ARIMA model and plot the forecasts. Describe how you select the type of exponential smoothing (`model` parameter in the `ets()` function). Provide a matrix showing the RMSE on the test set for all models (see Tab. 1).*

(c) *Describe the benefits and drawbacks of the models applied in task (b). Take the properties of the time series into account. Which model performs best?*

(d) *According to Hyndman and Athanasopoulos (2021, chapter 9.10, `https://otexts.com/fpp3/arima-ets.html`), certain types of exponential smoothing models are special cases of ARIMA models. Considering only beef, compute the ARIMA parameters for the exponential smoothing model used in (b) using the formula in the book. Use the computed parameters in R to perform a forecast. The parameters can be used in the `fixed` argument of the `forecast::arima()` function. Compare the results with models in (b).*

Exercise 2 (Univariate forecasting using Exponential Smoothing & ARIMA)

The `US accident deaths` dataset describes the monthly total number of accidental deaths in the USA between 1973 and 1978.

Material provided:

- `USAccidentDeaths.csv`: dataset

Tasks:

(a) *Load and explore the dataset. Check for missing values, plot the data and interpret what you see. Argue whether the time series is stationary based on ACF, PACF, and KPSS test. Which type of ARIMA model would you suggest?*

(b) *Implement cross-validation for the dataset. For this purpose, use 24 months as an initial fold for training, and forecast $h = 12$ months ahead.*

(c) *Use the cross-validation function implemented in (b) to train an exponential smoothing model, as well as a SARIMA model with order $(p = 1, d = 0, q = 0)(P = 1, D = 0, Q = 0)$ and period 12. Try alternative hyperparameter setups. Provide the mean and standard deviation of the test RMSE across the folds.*

(d) *Compare the models in (c) and explain your choice of the model parameters/settings. Use the models from the last cross-validation fold to investigate the model parameters, as well as the residuals.*

Exercise 3 (Univariate forecasting with exogenous inputs)

The dataset `traffic volume` contains the hourly traffic volume of Interstate 94 between Minneapolis and St Paul, MN, USA, between 2012 and 2018. In addition to the hourly traffic volume, environmental aspects like `rain`, `snow`, `clouds` and `temperature` are presumed to influence the traffic volume.

Material provided:

- `taffic_volume.csv`: dataset

Tasks:

(a) *Load and explore the dataset. Check for missing values, and plot the data. Investigate whether the traffic time series is stationary (ACF & PACF, KPSS test). Interpret the results. Restrict the data to 01.01.2013 - 31.12.2015. Hint: you may work with daily averaged values. Is a seasonal model useful?*

(b) *Split the data into training and testing (training: 2013-14, testing: 2015) and train a (S)ARIMA model (without covariates) with* `auto.arima`*. Plot the forecast and compute the RMSE. Check the residuals and argue whether your model is a good choice for this type of data.*

(c) *Train a (S)ARIMAX (dynamic regression) model with climate measurements as covariates. Plot the forecast and compute the RMSE.*

(d) *Compare the two models in (b) and (c). What are the benefits/drawbacks of taking covariates into account?*

Exercise 4 (Stationarity in ARIMA models)

In this exercise, you will simulate time series following Autoregressive models (AR) and Moving-Average models (MA). Both models require stationarity, which is guaranteed by side constraints on the regression parameters.

Tasks:

(a) *To simulate time series following an $AR(2)$ or $MA(2)$ model without intercept, consider the following instructions:*

(i) sample a vector ε_t of $t_{\max}$ i.i.d. errors from a Gaussian distribution (`rnorm(n=t_max + 100, mean=0, sd=sigma)`)

(ii) for $AR(2)$ models, implement the formula

$$x_t = \begin{cases} \varepsilon_t & t \leq 2 \\ \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \varepsilon_t & t > 2 \end{cases}$$

(iii) for $MA(2)$ models, implement the formula

$$x_t = \begin{cases} \varepsilon_t & t \leq 2 \\ \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \varepsilon_t & t > 2 \end{cases}$$

(iv) finally, cut off the first 100 values $x_1, \ldots, x_{100}$ ("burn-in")

Implement the given simulation procedure as an R function. Set a random seed (e.g., `set.seed(10)`) and use the function to simulate 3 time series for each model parameter combination specified in Tab. 2. Set the length of the time series to $t_{\max} = 1000$ time points and the error standard deviation to $\sigma = 0.1$. Which of the parameter combinations in Tab. 2 describe proper $AR(2)$ or $MA(2)$ models?

(b) *Plot and compare all time series simulated in (a), as well as their ACFs and PACFs. Perform a KPSS tests - are the time series stationary?*

(c) *For all time series simulated in (a), fit the parameters of an ARIMA model using `auto.arima` (package `forecast`). Compare the estimated model parameters to the ground truth parameters. Which model parameter combinations can be restored?*

(d) *One parameter combination in Tab. 2 obviously violates the side conditions stated for $AR(2)$ models. Show that first-order differences lead to a proper $AR(2)$ model (i.e. fulfill the side conditions).*

	BEEF	PIG	POULTRY	SHEEP
avg				
drift				
naive				
snaive				
ETS				
ARIMA				

Table 1: RMSE table.

model no	type	φ_1	φ_2	θ_1	θ_2
1	$AR(2)$	0.6	-0.3	-	-
2	$AR(2)$	0.8	0.2	-	-
3	$MA(2)$	-	-	0.6	-0.3
4	$MA(2)$	-	-	0.8	0.2

Table 2: Model parameters for $ARMA(k, q)$ model simulation.