STAR 513: HW 9

YOUR NAME HERE

Total points: 30
Questions are worth **2 pts** each, except where noted.
See Canvas calendar for due date.

Homework should be submitted as a pdf, doc or docx file via Canvas.
Use of R markdown HW template is strongly encouraged.
Add or delete code chunks as needed.
Knit frequently to avoid last minute problems!
Your submitted assignment should be neatly formatted and organized.

We continue with the Bike Share data from the previous assignment. The data contains daily observations for n = 551 days. This data is available from Canvas as `bike_sharing.csv`.

Q1 - Q6 (Polynomial and Splines)

For this group of questions use:

- registered_users (y)
- temp (x): average ambient temperature in degrees Celsius

Q1

Use the poly() function to model a quadratic, cubic, and quartic (**degree = 2, 3, 4**, respectively) relationship between temperature and the number of registered users.

Please NOT show the output, just save the results for later use.

Q2 (4 pts)

Using your polynomial models from Q1, choose a model based on (manual) "backwards elimination" based on p-values. We will follow the "Principle of Hierarchy".

What is the full model? From that full model, give the relevant p-value to decide whether to stick with the full model or simplify.

Depending on your answer to the question above, consider further simplifications to the model.

Report the degree of "final" selected model.

Full Model: Degree = ???
From this full model, we consider the test of ???.
Since p = ???, we . . . .

Additional steps as needed.

Selected Model: Degree = ???

Q3

Using your polynomial models from Q1, choose a model based on AIC. We will follow the "Principle of Hierarchy".
Give the AIC values for each of the models.
Report the selected model.

Selected Model: Degree = ???

Q4

Use the ns() function from the splines package to model the relationship between temperature and the number of registered users for **df = 3, 4, 5**.

Please do NOT show the output, just save the results for later use.

Q5

Using your spline models from Q4, choose a model based on AIC.
Give the AIC values for each of the models.
Report the selected model.

Selected Model: df = ???

Q6 (4 pts)

Create a scatterplot of registered users (y) vs temperature (x). Overlay the selected polynomial model (from Q3) and the selected spline model (from Q5) on a single plot using different color lines. Include an informative legend.

Q7 - Q12 (AIC)

We continue using the bike data with registered users as the response (y). But now we will consider additional predictors:

- **as.factor(year)**: 2011 or 2012
- season: winter, spring, summer, or fall
- weather: Clear, Light Precip or Mist
- **poly(temp,3)**
- humidity: percent humidity (0-100)
- windspeed: peak windspeed in kilometers per hour.

Q7 (4 pts)

Use MuMIn::dredge() to perform all subsets selection using AIC criteria. Consider **additive models** only (no interactions). For the selected model, report the (multiple) R2 value and the selected predictors.

Note: By default, dredge() will rank models by AICc. Since this question specifically asks for AIC, specify rank = "AIC".

R2 = ???
Predictors: ????

Q8

Consider the diagnostic plots: Resids vs Fitted and QQplot of Residuals. You do NOT need to show the graphs, just discuss any concerns you might have.

Response

Q9

A colleague suggests that you consider the variable `tempfeel` as a potential predictor. Explain why we do NOT include `tempfeel` as a potential predictor. Reference specific evidence.

Response

Q10

Looking at the MuMIn::dregde() results, a colleague asks why the + sign appears in the output. Briefly explain.

Response

Q11

A colleague suggests that you try stepwise AIC model selection. Do you expect this would change the selected model? Briefly discuss.

Response

Q12

Propose at least one thing we could try to improve the model fit. You don't need to actually do this, just propose an idea.

Response

Appendix

```r
#Retain this code chunk!!!
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(message = FALSE)
library(knitr)
library(tidyverse)
library(broom)
library(kableExtra)

#Q1 - Q6: Polynomial and Splines
#Q1

#Q2

#Q3

#Q4

#Q5

#Q6

#Q7 - Q12: AIC
#Q7

#Q8

#Q9
```