PSTAT225 Homework 1: Due 11/13/2022

1. A researcher studied the sodium content in lager beer by selecting a random six brands from the large number of brands of U.S. and Canadian beers sold in a metropolitan area. She then chose eight 12-ounce cans or bottles of each selected brand at random from retail outlets in the area, and measured the sodium content (in milligrams) of each can or bottle. The data set is in the same folder.

   (a) Test whether or not the mean sodium content is the same in all brands sold in the metropolitan area.

   (b) Estimate the mean sodium content for all brands and construct a 99% confidence interval.

2. An automobile manufacture wished to study the effects of differences between drivers and differences between cars on gasoline consumption. Four drivers were selected at random; also five cars of the same model with manual transmission were random selected from the assembly line. Each driver drove each car twice over a 40-mile test course and the miles per gallon were recorded. The data set is in the same folder.

   (a) Conduct an appropriate analysis.

   (b) Suppose these four drivers were not randomly selected from a population. They were actually four members in a family and we were interested in this family only. Conduct an appropriate analysis and compare these four drivers.

3. The Oxboys data in `nlme` consist of the heights of 26 boys from Oxford, England, each measured on nine different occasions.

   (a) Plot the data (using plot(Oxboys)) and verify that a simple linear regression model gives a suitable representation of the boys' growth patterns. Do there appear to be significant differences in the individual growth patterns?

   (b) Fit a simple linear regression model to height versus age using the `lm` function, ignoring the Subject effects. Obtain the boxplots of the residuals by Subject with

   ```
   bwplot(Subject ~ resid(object), Oxboys)
   ```

   where object should be replaced with the name of the fitted lm object. Explain the observed pattern.

   (c) Use the `lmList` function to fit separate simple linear regression models for each Subject. Compare the boxplots of the residuals by Subject for the lmList fit (obtained with `plot(object, Subject ~ resid(.))`, with object replaced with the name of the lmList object) to those obtained for the lm fit. Compare also the residual standard errors from the two fits and comment.

   (d) Plot the individual confidence intervals on the parameters estimated in the lmList fit and verify that both the intercept and the slope vary significantly with Subject.

   (e) Use the `lme` function to fit an LME model to the data with random effects for both the intercept and the slope. Examine the boxplots of the residuals by Subject, comparing them to those obtained for the lm and lmList fits.

(f) Produce the plot of the standardized residuals versus fitted values (`plot(object)`) and the normal plot of the standardized residuals (`qqnorm(object)`). (In both cases object should be replaced with the name of the lme object.) Can you identify any departures from the model's assumptions?

(g) Plot the augmented predictions for the lme fit (obtained with `plot(augPred(object))`). Do the linear models for each subject appear adequate?

(h) Another way of assessing the linear models for each subject is to plot the residuals versus age by Subject (use `plot(object, resid(.) ~ age | Subject)`, replacing object with the name of the lme object). Several subjects have a noticeable "scooping" pattern in their residuals, indicating the need for a model with curvature.

(i) Use the lmList function to fit separate quadratic models for each subject (`lmList(height ~ age + I(age**2), Oxboys)`).

(j) Examine a plot of the confidence intervals on coefficients from this second lmList fit. Are there indications that the coefficients differ between subjects? Are the quadratic coefficients significantly different from zero for some subjects?

(k) Fit the full mixed-effects model corresponding to the last lmList fit. The model will have linear and quadratic terms for age in the fixed-effects and the random effects. A simple way to describe this model is `lme(object)` replacing object with the name of the lmList fit.

(l) Check residual plots and numerical summaries for this lme model. Do there appear to be deficiencies in the fit? Do there appear to be terms in the model that could be eliminated?