STAT 425 — Spring 2016 — Sections 1UG, 1GR

Homework 7

Please submit your assignment *on paper*, following the Formatting Guidelines for Homework Submission. (Even if correct, answers might not receive credit if they are too difficult to read.) Remember to include relevant computer output.

1. File `videoviews.dat` contains view counts (`Views`) for five randomly-selected videos uploaded on each of six randomly-selected user channels (`Channel`) on a popular video-sharing website. Answer the following using a one-way random-effects ANOVA model with the Y variable being `log(Views)`. (All view counts are positive.) Assume all of the distributional conditions on the random effects and errors that were introduced in lecture (even though they are not exactly correct for this data).

   (a) [2 pts] Form an ordinary one-way ANOVA table, as if `Channel` were a fixed factor. What is the p-value for `Channel` effects?

   (b) [2 pts] Estimate the (random-effects) model using the REML method. (Provide a summary of the `lmer` output.)

   (c) [2 pts] Using the REML estimates, compute an estimate of the intra-class correlation coefficient.

   (d) [2 pts] Find the *maximum likelihood* estimate of the variance component associated with channels. (Use `lmer` with `REML=FALSE`.)

   (e) [2 pts] Determine the log-likelihood ratio test statistic value for the null hypothesis of no `Channel` effects. Also, determine its p-value (based on the usual chi-square reference distribution, *not* a bootstrap distribution).

   (f) [2 pts] Use the *parametric bootstrap* based on the log-likelihood ratio test statistic to compute a p-value for the null hypothesis of no `Channel` effects. (Note: Using `10000` samples is recommended for accuracy, even though the computations may take several minutes to finish. You may safely ignore any warnings produced.)

   (g) [2 pts] Predict the values of the random effects, based on the fit using REML.

2. Data set `choccake` in package `faraway` is from an experiment to examine the effects of three different recipes (`recipe`) and six different baking temperatures (`temp`) on the "breaking angle" (`breakang`) of a cake. (See the `help` file for details.) For each recipe, 15 batches of cake mix were prepared (presumably in a completely random order over all 45 batches). Each batch was then split into six parts, and each part was baked at a different temperature (assigned at random, independently between batches) to create a cake.

   (a) [2 pts] What is the whole-plot factor? What are the whole plots?

   (b) [2 pts] What is the split-plot factor? What are the split plots?

   (c) [2 pts] Notice that the batches are numbered *within* each recipe — two batches from different recipes can have the same number. You will thus need to create a new variable for batch that gives distinct labels to all batches. Create the variable using the factor interaction operator "`:`" by forming `choccake$recipe:choccake$batch` and naming it "`batches`." Display the values of this variable.

   Use this variable (instead of `batch`) in all of the following analyses.

(d) [2 pts] Using the function `lmer` (from the `lme4` package), fit an analysis model appropriate for this experiment using REML. (Note: You will need to convert `temp` to a factor variable.) Apply function `anova` to your fitted model, and display the results.

(e) [2 pts] By applying the `drop1` function to your fitted model from the previous part, perform a likelihood ratio (chi-squared) test for interaction effects between recipe and temperature. (Give your test statistic value, p-value, and conclusion.)

(f) [2 pts] *If it is appropriate*, test for main effects of recipe and temperature (by applying `drop1` to an updated model). *If this is not appropriate*, explain why not.

(g) [2 pts] Using the function `aov`, fit an analysis model appropriate for this experiment. (Note: You will need to convert `temp` to a factor variable.) Display a `summary` of your fitted model.

(h) [2 pts] Using the summary from the previous part, draw conclusions about recipe and temperature interactions and (if appropriate) main effects.

3. Data set `shocks` in package `alr4` records the results of "a small experiment to learn about the effects of small electric shocks on dairy cows." For each of six quantitative `Intensity` values (mA), the number `Y` of "positive reactions" out of a total of `m` trials was observed.

(a) [2 pts] Fit a *logistic* regression with `Intensity` as the predictor. (Let "positive reaction" be the event.) Produce a summary of your results.

(b) [2 pts] Create a single plot of "positive reaction" probability versus intensity that has (i) for each intensity, the fraction of "positive reactions," and (ii) the fitted logistic regression curve.

(c) [2 pts] Compute the *predicted values* (probabilities) and the *deviance* residuals.

(d) [2 pts] Estimate the probability that an intensity of 3.5 mA will cause a "positive reaction." Also, estimate the *odds* of this.

(e) [2 pts] Perform a lack-of-fit test based on the *deviance*. Compute a p-value, and state your conclusion.

(f) [2 pts] Compute Pearson's chi-square statistic. Then use it to perform another lack-of-fit test. Compute a p-value, and state your conclusion.

(g) [4 pts] Fit a *probit* regression with `Intensity` as the predictor. Produce a summary of your results. Then perform a lack-of-fit test, based on *deviance*, as in part (e).

4. [GRADUATE SECTION ONLY] Consider the one-way random effect model

$$Y_{ij} \quad = \quad \mu \; + \; \alpha_i \; + \; e_{ij}$$

where μ is fixed, $\alpha_i \sim N(0, \sigma_\alpha^2)$, $e_{ij} \sim N(0, \sigma^2)$, and all of the α_is and e_{ij}s are independent.

(a) [2 pts] Derive the expected value and variance of Y_{ij}.

(b) [2 pts] Derive the covariance of Y_{ij} and $Y_{ij'}$ for $j \neq j'$.

(c) [2 pts] Derive the *correlation* of Y_{ij} and $Y_{ij'}$ for $j \neq j'$.

Some reminders:

- Unless otherwise stated, all data sets are either automatically available or can be found in either the `alr4` package or the `faraway` package in R.

- Unless otherwise stated, use a 5% level ($\alpha = 0.05$) in all tests.