

STA441 Assignment 11

Quiz in Tutorial on Tuesday April 3d

A car rental company randomly assigned automobiles to one of three maintenance programs. Outcome was whether the car needed repairs, specifically repairs not required because of an accident or because of superficial damage like dings and chips to the windshield. Data were recorded in each of 12 successive months (the company only keeps the cars for one year, then sells them). There are 438 cars in the sample. For each month, variables are

- Car identification number
- Maintenance program (1 2 3)
- Month
- Cumulative number of customers who have rented the car
- Cumulative number of kilometers driven
- Repair (0=No, 1=Yes)

The data are available in an Excel spreadsheet: [AutoRepair.xlsx](#)

1. First, just make tables of Maintenance program by Repair and Month by Repair, with no tests. Take a look. For example, what percent of cars had a repair in month 1? Month 12?
2. Fit a mixed model in which the response variable is repair, and there is just one explanatory variable: Maintenance program. Make Program 3 the reference category.
 - a. Be able to locate all the parameter estimates. For example, what is the MLE of σ^2 ?
 - b. Test whether Maintenance program affects repair rate. This is one test.
 - c. Carry out pairwise comparisons of the maintenance programs. With a Bonferroni correction, what do you conclude?
 - d. The odds of a repair are ____ times as great for Program 3 as for Program 1. Give an estimate of this odds ratio, along with a 95% confidence interval.

Once you can do this question, you have shown that you understand the ideas and you know how to use the software. That's good, but we should proceed to analyze the data.

3. Now fit a model with Maintenance program and Month, treating Month as categorical. Make month 1 the reference category for Month, and keep 3 the reference category for Maintenance program. Do not include interaction terms for now, because I tried it. It was very time-consuming and a numerical disaster. My -2 log likelihood is 3894.2. This is really encouraging, because R's glmer function fails to converge, stopping at a place in the parameter space that is fairly close to the SAS MLE, but clearly not a minimum. Test month controlling for maintenance program and maintenance program controlling for month.
4. The next step is to see whether month is still significantly related to repair rate once we control for number of customers and km (as well as Maintenance program). If not, we will be happy to get rid of all those dummy variables. If you try adding TotCust and TotKm to the model, you will see a scary warning about negative eigenvalues and use of the Moore-Penrose inverse. It's a symptom of fatal numerical problems in the numerical search for the MLE. In this case, it happens because total number of customers and especially total kilometers are very big compared to the zeros and ones of the dummy variables. So re-scale the quantitative variables, dividing TotCust by 100 and TotKm by 1000. Thus, total customers is in hundreds, and total kilometers is in thousands. You could not be expected to think of trying this. Test Month controlling for the other variables.
5. Now fit the model without Month. The explanatory variables are maintenance program and rescaled cumulative number of customers and cumulative kilometers.
 - a. This is the restricted model for the likelihood ratio test of month. The test statistic is the difference of the two -2 log likelihoods. Use proc iml to calculate the test statistic and p-

value. How does this compare to the Wald test in Question Four? To get the Wald statistic, multiply the so-called F statistics by the degrees of freedom.

- b. I hope you like the model without month now. I do. For this model, look at the test of total customers controlling for maintenance program and total kilometers. What do you conclude?
- 6. Since we seem to be in the business of model selection today, fit the model with just maintenance program and rescaled cumulative kilometers.
 - a. Controlling for kilometers on the car, test whether Maintenance program affects repair rate. This is one test.
 - b. Carry out pairwise comparisons of the maintenance programs controlling for kilometers. With a Bonferroni correction, what do you conclude? Did controlling for total kilometers change the conclusions (compared to Question 2)?
 - c. Now produce estimated probabilities of repair for each maintenance program, holding kilometers driven constant at the common mean number of kilometers at 6 months. You can get this number from proc means. Accompany your estimates with 95% confidence intervals. In practice, you would probably do this for each month.

Please **bring your log file and results file to the quiz**. As usual, answers to the questions are not to be handed in. They are just practice for the quiz.