ECOM40006/ECOM90013 Econometrics 3
Department of Economics
University of Melbourne

Week 12 Tutorial Exercise Solutions

Semester 1, 2025

The primary function of this week's tutorial is revision and to provide you with the opportunity to ask your tutor questions that you may have from anywhere in the course. Challenge yourselves to challenge your tutor!

In the event that there is time remaining then you should attempt the following exercises.

1. The file `loanapp.dta` contains information on mortgage loan applicants in the US. The variables in the data set are described in the file. The pool of loan applicants is comprised of white, black and Hispanic people. Whether or not the application was approved is recorded in the variable *approve* which is equal to one if the application was successful and is zero otherwise.

   Load the package `haven`, and any dependencies, so that R can read Stata files and then read in `loanapp.dta`.

   (a) To test for discrimination in the mortgage loan market, estimate a probit model with *approve* as the dependent variable and a list of regressors which includes *hrat*, *obrat*, *loanprc*, *unem*, *male*, *married*, *dep*, *sch*, *cosign*, *chist*, *pubrec*, *mortlat1*, *mortlat2*, *vr*, and *white*. Do you find evidence of racial discrimination in the success of loan applications? Justify your answer.

   Solution:
   The R code to do the various bits and pieces can be found in Week_12.R. Here it is sufficient to look at the t-statistic on the variable *white*, which takes a value of 5.993. The corresponding p-value is extremely small meaning that the probability of obtaining a coefficient estimate as large as this when the true value is zero is essential zero. We conclude that race does seem important in determining probability of success of a loan application in these data.

   (b) Suppose that you add a new regressor that is an interaction (the product of the two variables) between *obrat* and *white*. Call the new variable *wobrat*. Does your conclusion about discrimination in the loan market changes? Justify your answer.

   Solution:
   A priori you wouldn't expect the conclusion to change given how significant *white* was in the previous model. However, it is no longer enough to just look

at a t-ratio as we have a joint test. Let's answer this question a bit more formally than above. (Like it was an exam question!) In particular, we have

$$\begin{aligned}&H_0 : \beta_{white} = \beta_{wobrat} = 0\\&\qquad \text{versus}\\&H_1 : \text{At least one of the two coefficients differs from zero.}\end{aligned}$$

An appropriate test statistic is the likelihood ratio:

$$LR = -2(\ln \mathcal{L}_0 - \ln \mathcal{L}_1),$$

where $\ln \mathcal{L}_0$ and $\ln \mathcal{L}_1$ denote the maximized values of the log-likelihood function in the restricted (null) and unrestricted models, respectively. Under the null hypothesis the likelihood ratio test will be χ^2_2. If testing at a 5% level then the appropriate critical value is $\chi^2_{2,0.05} = 5.9915$. (You can do this in R but it is also a good opportunity to practice using the statistical tables that you will have avaialable during the exam.) The decision rule is then to reject H_0 for all values of $LR > \chi^2_{2,0.05}$.
The R code that l provided calculates the LR test in two vauely different ways. The statistic called `LRtest` uses the maximized values of the log-likelihood function, as define above. The statistic `LR2test` the values of the Residual Deviance that comes as part of the standard output from the `summary` command when applied to output of the `glm` command. Note that in this latter case we calculate

$$LR = \text{Residual Deviance}_0 - \text{Residual Deviance}_1,$$

in an analogous notation to that above. In both cases we obtain a value for the test statistic of 37.08271 which far exceeds the critical value and so we conclude that there is evidence of racial discrimination for the probability of success of loan approvals in the data.

(c) Explain how you might construct prediction intervals for the probability of success of a given applicant.

Solution:

The predicted probability of success is a function of the form $\Phi(x_i'\hat{\beta})$, where $\Phi(\cdot)$ denotes the standard Normal cdf and $\hat{\beta} \underset{a}{\sim} \mathrm{N}(\beta, \mathcal{I}^{-1})$ is the maximum likelihood estimator and $\mathcal{I}^{-1}$ is the inverse of the information matrix. We just apply the the Delta mathod which tells us that

$$g(\hat{\beta}) \underset{a}{\sim} \mathrm{N}(g(\beta), G'\mathcal{I}^{-1}G), \qquad \text{where } G = \left.\frac{\partial g(\beta)'}{\partial \beta}\right|_{\beta=\hat{\beta}}.$$

Here

$$G = \left.\frac{\partial \Phi(x_i'\beta)}{\partial \beta}\right|_{\beta=\hat{\beta}} = \left.\frac{\mathrm{d}\Phi(x_i'\beta)}{\mathrm{d}x_i'\beta}\frac{\partial x_i'\beta}{\partial \beta}\right|_{\beta=\hat{\beta}} = \phi(x_i'\hat{\beta})x_i.$$

A really complete answer might see you derive the the information matrix, along the lines of the derivation given on pages 7–8 of the Binary Respnse Models handout, but, as you were only asked to explain how you would do it, I think that this is probably enough. (You should carefully check to make sure that I don't have my transposes messed up and let me know if I do.)

2. Deb and Trivedi (1997) analyze data on 4406 individuals, aged 66 and over, who are covered by Medicare, a public insurance program. Originally obtained from the US National Medical Expenditure Survey (NMES) for 1987/88, the data are available from the data archive of the Journal of Applied Econometrics at `http://qed.econ.queensu.ca/jae/1997-v12.3/deb-trivedi/`. The objective is to model the demand for medical care — as captured by the number of physician/non-physician office and hospital outpatient visits — by the covariates available for the patients. Here, we adopt the number of physician office visits *ofp* as the dependent variable and use the health status variables *hosp* (number of hospital stays), *health* (self-perceived health status), *numchron* (number of chronic conditions), as well as the socioeconomic variables *gender*, *school* (number of years of education), and *privins* (private insurance indicator) as regressors.

   (a) Create a subset of the data comprised only of columns 1, 6, 7, 8, 13, 15, 18. Call this dataset dt.

   Solution:
   See Week_12.R.

   (b) Use the following plotting command and describe the features of the resulting plot.

   plot(table(dt$ofp), main="*mtext*", xlab="*xtext*", ylab="Frequency")

   Note: so that the command would fit on a single line of this document I have factored out *main text* and *xtext*. You should replace these bits (between the quotes) by the following text in the command that you provide to R:

   mtext = Frequency distribution for number of physician office visits.
   xtext = Number of physician office visits

   Solution:
   See Week_12.R. We observe that there is substantial variation across individuals in the number of visits to the doctors and that there is a very substantial number of people who made no visits (may want to explore a zero-inflated model) or very few visits to the doctor.

   (c) Use the following command to fit a Poisson regression model to this data:

   PReg=glm(ofp ∼ ., data = dt, family = poisson)
   summary(PReg)

   Note that the .dot in the glm command simply says: use all the remaining variables in dt as regressors.

   Solution:
   See Week_12.R. All regressors appear significant, with those related to health appearing more significant than those of socio-economic variables.

   (d) Interpret the estimated coefficient on *school*.

   Solution:

   (e) Explain how you would construct standard errors on predicted values from this model.

   Solution:

Final Remarks

I should highlight the role that tutorials have played this semester. Although sometimes they have been used to reinforce ideas introduced during the lectures, they have frequently been used to expose you to material and ideas that we weren't going to get to during the regular lectures. As such your exam revision should probably include material covered during the tutorials.

Finally, although this is clearly your most important subject, your other subjects are important too. Good luck with all of your exams.

Cheers, Chris.