The purpose of this exam is to test your working knowledge of the models we've studied by giving you a chance to analyze some real data. Unlike the homework exercises I DO NOT want you to collaborate. I can do nothing more than trust you on this, so please honor that trust. A completed exam will consist of:

- A clearly-written and properly formatted document describing your analysis and findings. This should include any tables you present, sequentially numbered and placed at the end of the document.
- A .zip file holding any programs and m-functions you use.

The document should be no longer than 12 pages of text (excluding tables), and can be shorter. Please turn in the document and .zip files to me via email by the due date.

I will judge your work based on your use of the models and methods, analysis results, and clarity of presentation. Please take note of the latter – make sure you are taking care to describe your findings and how you arrived at them. When post-estimation computations are needed make sure you describe how you've gone about completing them. I will be reading the document and not the code or any computer output you might include (I am requesting programs only for checking process). In completing the exercises you may use code you wrote during the semester, the code from the HW keys, the workshop code I shared with you, and/or purpose-written code.

Data
Posted on the class website you will find the Matlab data file *carin.mat*. This holds data on automobile purchases made by a cross section of households in 2001. There are $n=3216$ observed purchases from among $j=55$ vehicle categories. The choice elements are manufacturer/class combinations so that elements are defined as ‘Ford midsize’, ‘Honda minivan’, etc. The data include various combinations among 7 manufacturers and 10 vehicle classes (not all manufacturers sell all types). The website contains an Excel file *choice set.xls* describing the specific elements and listing the manufacturers and classes.

When you load the Matlab data file you will see several matrices, defined as follows.

**Y**: A 3216 by 55 matrix of zeros and ones indicating the chosen alternatives

**X**: A 3216 by 715 matrix of vehicle attribute variables that includes

- price – retail price in $1000’s (constant over people)
- wheel base – distance in inches from front to rear axle (constant over people)
- operating cost – cost per mile driven in US cents (varies over people and alternatives)
- compact – vehicle class indicator
- luxury compact – vehicle class indicator
- midsize – vehicle class indicator
- full size – vehicle class indicator
- luxury full size – vehicle class indicator
- small SUV – vehicle class indicator
- large SUV – vehicle class indicator
- small truck – vehicle class indicator
- large truck – vehicle class indicator
- minivan – vehicle class indicator

---

1 Text double spaced, 1 inch margins, Times New Roman 11 point font, first line of paragraphs indented, equations centered.
I: A 3216 by X matrix of household characteristics that includes
   low income – a dummy variable set to one if the household is in the low income quartile
   size – number of people living in the household

X_names A character matrix holding variable names for the matrix I
I_names: A character matrix holding variable names for the matrix I

The variable operating cost is constructed using the each vehicle’s miles per gallon and dividing it by the average price per gallon of fuel in the person’s home state. The variable wheel base is a good proxy for the size of the vehicle (e.g. the number of passengers it can accommodate, gear it can haul, etc.).

Conditions of use: this data set is NOT publically available and can be used only for the purposes of this exercise. I will remove it from the website after you have downloaded. Please do not use it for other projects, exercises, etc.

Exercises
1) Within the context of a multinomial logit model complete and discuss the following.
   a) Estimate a model including price, wheel base, and operating costs as explanatory variables.
      Interpret your results. If you find any odd signs discuss why this might be.
   b) Estimate the model with price, wheel base, operating costs, and the ten class indicator variables.
      Explain how this changes your findings. Once again comment on any unexpected findings and their explanations.
   c) Estimate a model (without class dummies) that includes interactions between operating costs and
      low income, and wheel base and household size along with the level effects of price, operating
      costs, and wheel base. Interpret your findings and compare them to your estimates from (a) and
      (b).
   d) Estimate a model with a full set of alternative specific constants and the interactions noted above.
      Explain what you can and cannot estimate in this specification, and interpret your findings.
   e) Return to the specification from (b). Predict how a $1500 subsidy for the purchase of a compact
      class vehicle changes the sample average probabilities for the 55 alternatives. List out both the
      baseline and counterfactual predictions by vehicle class (i.e. aggregate over the manufacturers).

2) Within the context of a classical mixed logit model complete and discuss the following.
   a) Estimate a model that includes the price, wheel base, and operating cost variables such that price
      has a fixed coefficient, and wheel base and operating costs have normal coefficients. Describe
      your findings and compare them to 1(a) above.
   b) Consider a specification including price, wheel base, and operating costs with fixed coefficients.
      Suppose intuition suggests that aggregate classes of vehicles such as compact cars, other cars,
      trucks, and minivans/SUVs should have correlated unobserved utility components. Estimate a
      model with error components that can accommodate this intuition, making sure to properly
      constrain the relevant parameters to zero. Describe any evidence you find for such an error
      structure. Compare your findings to those from 1(a), 1(b), and 2(a).
   c) Recall that probabilities in the mixed logit model are given by
   \[
   p_{aj} = \frac{\exp(\beta_j X_{aj})}{\sum_{k=1}^J \exp(\beta_k X_{ak})} f(\beta_j) d \beta_j,
   \]
   where \( f(\cdot) \) is the mixing distribution. Using the model from 2(b) predict again how a $1500
   subsidy for the purchase of a compact class vehicle changes the sample average probabilities for
the 55 alternatives, and report them by class aggregates. Compare your predictions to those from 1(e) and discuss what drives any differences.

3) Within the context of a Bayesian mixed logit model complete and discuss the following.
   a) Estimate a model including price, wheel base, and operating costs. Specify independent (i.e. block diagonal variance/covariance) normal random parameters for all variables. Discuss how the parameter distributions for wheel base and operating costs compare to those found in the classical mixed logit model.
   b) Using the workshop code I distributed it is possible to estimate models with a subset of the parameters fixed. Estimate a model with the price coefficient fixed, and the wheel base and operating cost coefficients independent normal. This will match the specification from 2(a). Discuss the similarities and differences. Would you expect the estimates to be similar? Why? If they differ, why might this be?

4) Suppose we are interested in measuring consumer’s marginal willingness to pay in purchase price for lower operating costs. More specifically, suppose you are charged with estimating the average of this in the population. Use any approach you best see fit to provide the most defensible estimate of this quantity that you can. Make sure you:
   • motivate your choice of model and specification, and how your specification can be used to estimate the desired quantity; and
   • explain any potential for bias that might exist in the estimate that you report.