

Economics 403A: Homework 3
Fall 2018, UCLA
Instructor: Dr. Rojas

Due Date: Nov 24, 2018

All problems require detailed worked out solutions to receive full credit, and are all worth the same.

1. The dataset 401KSUBS contains information on net financial wealth (*nettfa*), age of the survey (*age*), annual family income (*inc*), family size (*fsize*), and a binary variable for eligibility in a 401(k) plan (*e401k*) among other variables. The wealth and income variables are both recorded in thousands of dollars. Our response variable for this problem is *nettfa*. Note: The complete the dataset includes 10 predictors (please refer to the file description for details).
 - (a) Provide a descriptive analysis of your variables. This should include, histograms and fitted distributions, quantile plots, correlation plot, boxplots, scatterplots, and statistical summaries (e.g., the five-number summary). All figures must include comments.
 - (b) For each variable, test if a transformation to linearity is appropriate, and if so, apply the respective transformation. We will use these results later in part (m).
 - (c) Estimate a multiple linear regression model for *nettfa* that includes *income*, *age*, and *e401k* as explanatory variables. We will use this model as a baseline. Comment on the statistical and economic significance of your estimates. Also, make sure to provide an interpretation of your estimates. If there are any outliers worth removing, remove them before proceeding with the next steps.
 - (d) For your model in part (c) plot the respective residuals vs. \hat{y} , and y vs. \hat{y} , and comment on your results.
 - (e) For a more economically realistic model, the *income* and *age* variables should appear as quadratics. Re-estimate your model from part (c) including these two quadratic terms. Now, what is the estimated dollar effect of 401(k) eligibility?
 - (f) For the model estimated in part (e), add the interactions $e401k(age-41)$ and $e401k(age-41)^2$. Note that the average age in the sample is about 41, so that in the new model, the coefficient on *e401k* is the estimated effect of 401(k) eligibility at the average age. Are the interaction terms significant? Would you suggest keeping one of the interactions (or both)? Explain.
 - (g) Comparing the estimates from parts (c) and (e), do the estimated effects of 401(k) eligibility at age 41 differ much? Explain.
 - (h) Now, drop the interaction terms from the model in part (f), but define five family size dummy variables: *fsize1*, *fsize2*, *fsize3*, *fsize4*, and *fsize5*. The variable *fsize5* is unity for families with five or more members. Include the family size dummies in the model estimated in part (e) and make sure to choose the base group as *fsize2*. Comment on your estimates.

- (i) Now, do a Chow test for the model

$$netffa = \beta_0 + \beta_1 inc + \beta_2 inc^2 + \beta_3 age + \beta_4 age^2 + \beta_5 e401k + e$$

across the five family size categories, allowing for intercept differences. The restricted sum of squared residuals, SSR_r , is obtained from part (g) because that regression assumes all slopes are the same. The unrestricted sum of squared residuals is, $SSR_{ur} = SSR_1 + \dots + SSR_5$, where SSR_f is the sum of squared residuals for the equation estimated using only family size f . You should convince yourself that there are 30 parameters in the unrestricted model (5 intercepts plus 25 slopes) and 10 parameters in the restricted model (5 intercepts plus 5 slopes). Therefore, the number of restrictions being tested is $q = 20$, and the df for the unrestricted model is $9,275 - 30 = 9,245$.

- (j) Based on your model in part (f) plot and discuss the marginal effects plots across your predictors.
- (k) Is there an optimal level of net financial wealth based on income and age? If so, compute this level and show the respective perspective and image plots.
- (l) For each predictor, plot the predictor effect plot by family size.
- (m) Estimate a multiple regression model using your transformed predictor variables from part (b) using the same model as in part (f) and compare the two models. Which one do you prefer?
- (n) Based on all the available predictors, estimate a model with additive and interactions terms, and compare it to your model in part (f).
- (o) Lastly, choose your favorite model from all the ones estimated and perform a five-fold cross validation test on it.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

2. Assume a healthcare insurance company hired you as a consultant to develop an econometric model to estimate the number of doctor visits a patient has over a 3 month period. The rational behind this study is that patients with a higher number of doctors visits wold pose a higher liability in terms of insurance expenses, and therefore, this may be mitigated via a higher insurance premium. The panel data are from the *German Health Care Usage Dataset*, and consist of 7,293 Individuals across varying numbers of periods with a total of 27, 326 observations.
- (a) Build a multiple regression model with a subset of 10 predictors (at most), including interaction and non-linear transformations if appropriate. For this part you only need to briefly discuss a justification for the model chosen, and discuss the respective regression output.
 - (b) *Differences in Differences*: In 1987 the German Government passed a series of legislations to improve healthcare access for unemployed people and women.
 - i. Determine whether or not the policy worked for women.
 - ii. Determine whether or not the policy worked for unemployed.
 - (c) Test the hypothesis that the number of doctor visits a patient has over a 3 month period is greater for women than for men.
 - (d) Based on your findings propose and test your own hypothesis of interest using the linear functional form: $\lambda = c_1\beta_1 + c_2\beta_2 + \dots$.

<https://powcoder.com>
Data Description (For Problem 2)

This is a large dataset. There are altogether 17,326 observations. (The number of observations ranges from 1 to 7. (Frequencies are: 1=1525, 2=2158, 3=825, 4=926, 5=1051, 6=1000, 7=987). Note, the variable NUMOBS below tells how many observations there are for each person. This variable is repeated in each row of the data for the person. Below are the variables definitions. Note: You can ignore the variables TI and INCOME (this one is just a copy of HHINC).

ID = person - identification number

FEMALE = female = 1; male = 0

YEAR = calendar year of the observation

AGE = age in years

HSAT = health satisfaction, coded 0 (low) - 10 (high) Note, this variable has 40 coding errors. Variable NEWHSAT below fixes them.

HANDDUM = handicapped = 1; otherwise = 0

HEALTHY = self reported to be healthy = 1; otherwise = 0

ALC = average alcohol consumption in the last 3 months

TRAVEL = traveled in the last 3 months abroad = 1; otherwise = 0

HANDPER = degree of handicap in percent (0 - 100)
HHNINC = household nominal monthly net income in German marks / 10000
LOGINC = Natural log (ln) of household nominal monthly net income in German marks / 10000
HHKIDS = children under age 16 in the household = 1; otherwise = 0
EDUC = years of schooling
MARRIED = married = 1; otherwise = 0
HAUPTS = highest schooling degree is Hauptschul degree = 1; otherwise = 0
REALS = highest schooling degree is Realschul degree = 1; otherwise = 0
FACHHS = highest schooling degree is Polytechnical degree = 1; otherwise = 0
ABITUR = highest schooling degree is Abitur = 1; otherwise = 0
UNIV = highest schooling degree is university degree = 1; otherwise = 0
WORKING = employed = 1; otherwise = 0
BLUEC = blue collar employee = 1; otherwise = 0
WHITEC = white collar employee = 1; otherwise = 0
SELF = self employed = 1; otherwise = 0
BEAMT = civil servant = 1; otherwise = 0
DOCVIS = number of doctor visits in last three months
HOSPVIS = number of hospital visits in last calendar year
UNEMPLOY = unemployed = 1; otherwise = 0
DOCTOR = dummy variable = 1 if DOCVIS > 0, 0 otherwise.
HOSPITAL = dummy variable = 1 if HOSPVIS > 0, 0 otherwise.
PUBLIC = insured in public health insurance = 1; otherwise = 0
ADDON = insured by add-on insurance = 1; otherwise = 0
NUMOBS = number of observations for this person. Repeated in each row of data.
NEWSAT = recoded value of HSAT with coding errors corrected.
PRESCRIP = number of prescriptions in last three months

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder