HSE and University of London Double Degree Programme in Data Science and Business Analytics

Elements of Econometrics, 2023-2024

Classteacher: Ksenia Kasianova.

Class 20: SEM. Bianary choice models

Problem 1

Theory recap:

a) What are the properties of ML estimator?

b) What is Fishers information? How is it related to estimator variance?

c) How Wald, LM and LR statistics are related to likelihood function?

Problem 2

Consider a sample: $X_1 \dots X_{100} \sim N(\mu, \sigma^2)$, where $\sum_{i=1}^{100} X_i = 20$, $\sum_{i=1}^{100} \left(X_i - \bar{X}\right)^2 = 400$. Conduct test $LR, LM, W-?$ 95% for the following hypothesis:

$$\begin{cases} H_0 : \begin{pmatrix} \mu \\ \sigma \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ H_a : \begin{pmatrix} \mu \\ \sigma \end{pmatrix} \neq \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{cases}$$

Hint:

$l = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$

Problem 3

Pablo estimates two unknown parameters: a - where the rockets are located, $b-$ where they sell candy.

Pablo estimated the parameters using the maximum likelihood method and obtained estimates $\hat{a} = 1.5, \hat{b} = 2.5$. Then Pablo decided to test the hypothesis $H_0 : a = 1$ and $b = 2$.

The values of the likelihood function, gradient and estimated Fisher information at two points are partially shown in the table:

	$\ell(a,b)$	(ℓ'_a, ℓ'_b)	$\hat{I}_F$
$a = 1.5, b = 2.5$	-200	?	$\begin{pmatrix} 16 & -1 \\ -1 & 20 \end{pmatrix}$
$a = 1, b = 2$	-250	$(2, -1)$	$\begin{pmatrix} 10 & -1 \\ -1 & 15 \end{pmatrix}$

a) Fill in the gaps

b) Test H_0 using LR, LM, W

Problem 4

More than a century ago on April 14, 1912, the unthinkable happened when the "unsinkable" Royal Mail Ship (RMS) Titanic crashed into an iceberg and sunk into the Atlantic Ocean. The 20 lifeboats aboard the ship were not enough to save a majority of the passengers, leaving over 1500 passengers and crew members aboard the sinking yessel. A total of 705 passengers escaped onto lifeboats and to safety. But not all passengers had an equal chance of getting onto a lifeboat and surviving the disaster. A student has found in internet the data on 2201 passengers and crew members of RMS Titanic during his last voyage

across the ocean. She is trying now, using regression analysis, to estimate the impact of passenger class ($\boldsymbol{CLASS} = 0$ for the crew members, $\boldsymbol{CLASS} = 1$ for the first class passengers, $\boldsymbol{CLASS} = 2$ for the second class passengers, $\boldsymbol{CLASS} = 3$ for the third class passengers), gender ($\boldsymbol{MALE} = 1$ for male and 0 for female) and age ($\boldsymbol{AGE} = 1$ for adults and 0 for children) on a person's likelihood of surviving the shipwreck $SURV$. In the data $SURV = 1$ if the person survived the shipwreck and $SURV = 0$ if not. She runs different regressions with the following results (dependent variable $\boldsymbol{SURV}$, regression (i) uses full sample, other regressions use subsample without crew members; (asymptotical) standard errors in parentheses).

	(i) (OLS)	(ii) (OLS)	(iii) (LOGIT)	(iv) (LOGIT)	(v) (PROBIT)
CLASS	-0.0515	-0.1515	-0.489	-0.875	-0.502
	(0.0072)	(0.0134)	(0.043)	(0.085)	(0.048)
AGE	-0.165	-0.181		-1.056	-0.580
	(0.007)	(0.040)	-	(0.243)	(0.138)
MALE	-0.552	-0.478		-2.367	-1.415
	(0.022)	(0.023)		(0.145)	(0.084)
Constant	0.985	1.208	0.793	3.895	2.257
	(0.047)	(0.053)	(0.103)	(0.347)	(0.193)
N	2,201	1,316	1,316	1,316	1,316
R^2	0.228	0.331			
McFadden R^2			0.0758	0.269	0.267
RSS	371.59	207.31	278.52	204.30	205.21
LogL	-1165.43	-651.45	-807.15	-638.40	-639.93
LR stat,			132.45	469.95	466.90

(a) Give interpretation to the regressions (i) and (ii). Are they significant? What are the advantages and the disadvantages of these models? Why in models (iii)-(iv)-(v) these shortcomings are absent or mitigated?

(b) Using regression (iv) help the student to evaluate the marginal effect of passenger class, taking as the initial values an adult female person traveling in a first class cabin (use method of derivatives). Compare your result with the corresponding result for regression (ii)

(c) What are McFadden R^2 and LR statistic? How they can be used for evaluation of the statistical quality of the regression? Comparing regressions (iii) and (iv) evaluate whether two dummy variables $\boldsymbol{AGE}$ and GENDER are significant both separately and taken together?

(d) What is the difference between regressions (iv) and (v)? Help the student to understand whether for female passengers of Titanic the chance to survive is significantly greater. To do this for regression (v) evaluate the marginal effect of gender, taking as the initial values an adult male person traveling in a first class cabin (use direct calculation without derivatives). Compare your result with the corresponding result for regression (ii).