HH/HLST 2300: Statistical Methods in Health Studies
Winter Term Assignment 5
Assigned: Friday March 19, 2021; Due 5PM Friday April 2 9, 2021
Submit 1 file for Assignment 5: PDF
PDF document name: LASTNAME_FIRSTNAME_WTAssignment5
Submit via eClass

As noted in my eClass announcement on Nov 5, 2020, you will be deducted marks if you submit a file other than a pdf file and if that file is named incorrectly. **For WT Assignment 5, the deduction is 4 marks for incorrect file name. The submission file type has been set up in eClass such that the only accepted file type is PDF.**
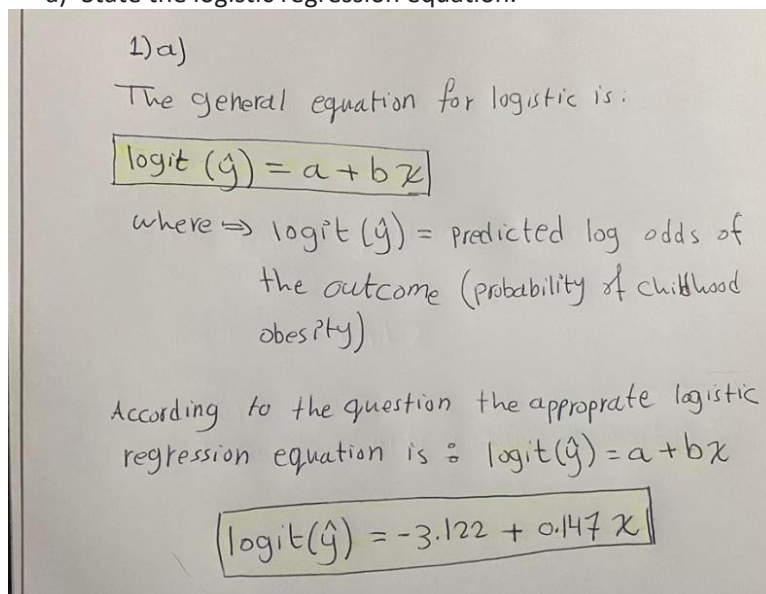
Other reminders that you should take care to ensure while completing your assignment:
- Questions involving a data file must be answered using SPSS
- HLST 2300 rounding rules apply unless otherwise stated
- Screenshots of any hand-written work and SPSS must be of high resolution and be pasted upright (not sideways) so that they can be easily read and graded
- Answers to questions must directly follow the question asked – do not change the order of the questions
- If you fail to include the SPSS output instructed of you, you will receive zero for any subsequent questions that rely on that output

1. The probability of childhood obesity (1 = obese; 0 = normal weight) was analyzed as a function of iron intake (mg) using logistic regression. Use the below output to answer the following questions:

Variables in the Equation				
	B	Wald	df	Sig
iron	.147	97.018	1	.000
Constant	-3.122	117.261	1	.000

a) State the logistic regression equation.

1)a)

The general equation for logistic is:

$$logit(\hat{y}) = a + bx$$

where $\Rightarrow$ logit($\hat{y}$) = predicted log odds of the outcome (probability of childhood obesity)

According to the question the appropriate logistic regression equation is: logit($\hat{y}$) = a + bx

$$logit(\hat{y}) = -3.122 + 0.147 x$$

Logit (obesity) = -3.122 + 0.147 (iron)

b) Interpret the coefficient for iron.

The coefficient for Iron is also knows as the log-odds. Log Odds can take on any value either positive or negative

The question here is reported as log odds (childhood obesity) = -3.122 + 0.147X

Therefore, the coefficient here is 0.147. as with the linear regression, the slope of the regression line is the key statistic. in this question the b coefficient is positive, showing that the log odds of childhood obesity increase by 0.147 for each one unit increase in Iron intake (in mg) also known as the predictor.

c) Calculate the odds ratio for iron. Show your work.

Now to introduce the concept of an odds ratio

The odds ratio (OR) is simply the odds in one group (exposed) divided by the odds in another (unexposed or control or reference)

We can now proceed to calculate the odds ratio (OR)

Since the logit (childhood obesity) is 0.147, the odds of children experiencing obesity is 1.16 (rounded to tow decimal places)

Logit (childhood obesity) = 0.147

$E^{logit(childhood\ obesity)} = e^{0.147}$

Odds (childhood obesity) = 1.15835

$e^{0.147}$ = EPX (0.147) = 1.16

d) Interpret the odds ratio for iron.

If we take the invers of the natural logarithm of the coefficient b=0.147 we get 1.16

When we compare the odd in terms of ratio, we are saying that the odds of the outcome is 1.16 times bigger one unit increase in the predictor. That means the odds of childhood obesity are multiplied by 1.16 for every extra mg of iron intake

2. Excel file: 2300 WT Assignment5 provides the age, comorbidity level (levels range from 0 – 4 where level 0 represents no significant comorbidity and level 4 represents the group with the largest number of comorbidities), ED triage level (resuscitation = 1; emergency = 2; urgent = 3), hospital length of stay (LOS) in days and visit disposition (1 = died in hospital; 0 = discharged back to LTC) for 152 patients admitted to hospital from long-term care (LTC).   Conduct a logistic regression analysis to predict death in hospital with the predictor variables described and set comorbidity level 0 and ED triage level emergency as the reference categories for their respective predictor variables, to answer the following questions.   HINT: you will need to recode ED triage level into a different variable to make emergency the reference category.

a) Is there any evidence that the logistic regression model is useful to predict death in hospital compared to the baseline model? Copy and paste the relevant output table(s) when reporting results.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	37.041	8	.000
	Block	37.041	8	.000
	Model	37.041	8	.000

The test is significant, X^2 (8) = 37.0.4, p < .001, where it's showing that the overall model is significantly better fit to the data than the baseline

b) Interpret the coefficient for comorbidity level 4. Copy and paste the relevant output table(s) when reporting results.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	triage_recoded			.701	2	.704			
	triage_recoded(1)	.191	.880	.047	1	.828	1.210	.216	6.792
	triage_recoded(2)	-.376	.493	.583	1	.445	.686	.261	1.804
	LOS (days)	-.041	.021	4.000	1	.045	.960	.922	.999
	Comorbidity Level			23.508	4	.000			
	Comorbidity Level(1)	1.305	1.197	1.189	1	.276	3.689	.353	38.551
	Comorbidity Level(2)	2.161	1.121	3.716	1	.054	8.682	.965	78.157
	Comorbidity Level(3)	2.550	1.115	5.224	1	.022	12.802	1.438	113.957
	Comorbidity Level(4)	3.960	1.103	12.891	1	.000	52.453	6.039	455.572
	Age (years)	.034	.035	.991	1	.320	1.035	.967	1.108
	Constant	-5.690	3.154	3.254	1	.071	.003		

a. Variable(s) entered on step 1: triage_recoded, LOS (days), Comorbidity Level, Age (years).

We are comparing the reference category of comorbidity level 0, the comorbidity level is higher which is 3.960 therefore, the odds ratio for comorbidity level 4 is higher by 52.453 holding the other predictors constant or adjust for the other predicators. The Wald test shows this this effect is significant Wald X^2 (1) = 1289, p< .001

c) If an 80 year old patient with comorbidity level 1 and ED triage level emergency has a hospital LOS of 10 days, what are the log odds of death? Show your work.

Logit(death) = -5.690 + .191 (triage recoded (1)) -0.376(triage recoded (2)) -0.041(LOS)+ 1.305 (comorbiditylevel(1)) +2.161 (comorbilitylevel(2))+ 2.550 (comorbilitylevel(3))+ 3.960(comorbidirylevel(4)) + 0.034 (age)

Logit(death) = -5.690 + .191 (0) -0.376(0) -0.041(10) + 1.305 (1) +2.161 (0) + 2.550 (0) + 3.960(0) + 0.034 (80)

= 18.715

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	triage_recoded			.701	2	.704			
	triage_recoded(1)	.191	.880	.047	1	.828	1.210	.216	6.792
	triage_recoded(2)	-.376	.493	.583	1	.445	.686	.261	1.804
	LOS (days)	-.041	.021	4.000	1	.045	.960	.922	.999
	Comorbidity Level			23.508	4	.000			
	Comorbidity Level(1)	1.305	1.197	1.189	1	.276	3.689	.353	38.551
	Comorbidity Level(2)	2.161	1.121	3.716	1	.054	8.682	.965	78.157
	Comorbidity Level(3)	2.550	1.115	5.224	1	.022	12.802	1.438	113.957
	Comorbidity Level(4)	3.960	1.103	12.891	1	.000	52.453	6.039	455.572
	Age (years)	.034	.035	.991	1	.320	1.035	.967	1.108
	Constant	-5.690	3.154	3.254	1	.071	.003		

a. Variable(s) entered on step 1: triage_recoded, LOS (days), Comorbidity Level, Age (years).

d) If an 80 year old patient with comorbidity level 1 and ED triage level emergency has a hospital LOS of 10 days, what are the odds of death? Show your work.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	triage_recoded			.701	2	.704			
	triage_recoded(1)	.191	.880	.047	1	.828	1.210	.216	6.792
	triage_recoded(2)	-.376	.493	.583	1	.445	.686	.261	1.804
	LOS (days)	-.041	.021	4.000	1	.045	.960	.922	.999
	Comorbidity Level			23.508	4	.000			
	Comorbidity Level(1)	1.305	1.197	1.189	1	.276	3.689	.353	38.551
	Comorbidity Level(2)	2.161	1.121	3.716	1	.054	8.682	.965	78.157
	Comorbidity Level(3)	2.550	1.115	5.224	1	.022	12.802	1.438	113.957
	Comorbidity Level(4)	3.960	1.103	12.891	1	.000	52.453	6.039	455.572
	Age (years)	.034	.035	.991	1	.320	1.035	.967	1.108
	Constant	-5.690	3.154	3.254	1	.071	.003		

a. Variable(s) entered on step 1: triage_recoded, LOS (days), Comorbidity Level, Age (years).

Rerun the logistic regression analysis using standardized values for the scalar variables.

Logit (death)= -5.69+0.34(80) -0.41(10) +1.305(1) =18.715

$e^{18.715}$ = 134,221,234.47 = EXP (18.715) Odds

e) Of the two scalar variables, which one is more important? Copy and paste the relevant output table(s) when reporting results.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	triage_recoded			.701	2	.704			
	triage_recoded(1)	.191	.880	.047	1	.828	1.210	.216	6.792
	triage_recoded(2)	-.376	.493	.583	1	.445	.686	.261	1.804
	Comorbidity Level			23.508	4	.000			
	Comorbidity Level(1)	1.305	1.197	1.189	1	.276	3.689	.353	38.551
	Comorbidity Level(2)	2.161	1.121	3.716	1	.054	8.682	.965	78.157
	Comorbidity Level(3)	2.550	1.115	5.224	1	.022	12.802	1.438	113.957
	Comorbidity Level(4)	3.960	1.103	12.891	1	.000	52.453	6.039	455.572
	Zscore: Age (years)	.218	.219	.991	1	.320	1.244	.809	1.912
	Zscore: LOS (days)	-.511	.256	4.000	1	.045	.600	.364	.990
	Constant	-3.398	1.058	10.316	1	.001	.033		

a. Variable(s) entered on step 1: triage_recoded, Comorbidity Level, Zscore: Age (years), Zscore: LOS (days).

In terms of the standardized output length of stay the LOS (day) which it's -. 511 is more important than age years which is .218

School of Health Policy and Management

Assignment Attachment Form

Student Name: ███████████

Student Number: ████████

Course Code: HHLST 2300

Assignment Title: assignment 5

Due Date: April 9, 2021

Tutorial Leader (if applicable):
Please check each box after reading, to acknowledge agreement with each statement.

☑ I have read and understand the Senate Policy on Academic Honesty found on website at the following <u>York Secretariat website on Academic Honesty</u>.

☑ I have read and understood the assignment submission described in the course outline (syllabus)

☑ I have read and understood the criteria used for assessment in this assignment

☑ I have read and understood and followed the referencing guidelines required for assignments submitted at York University

☑ This assignment is entirely my own work, except where I have given documented references to work of others

☑ This assignment or substantial parts of it has not previously been submitted for assessment in any formal course of study, unless acknowledged in the assignment and previously agreed to by my Tutorial Leader and Course Director

☑ I understand that this assignment may undergo electronic detection for plagiarism and a copy of the assignment may be retained on the database and used to make comparisons with other assignments in the future.

Signature: ████████████ Date: April 9, 2021