

Name Jerry

Student Number _____

STA 442/2101 F 2014 Quiz 9

1. U of T administration is very interested in whether the chances of success are different on the three campuses for undergraduate students with similar performance in High School. So, the Statistical Consulting Service carried out a logistic regression analysis in which

$$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3,$$

where π is the probability of graduating within five years of admission, x_1 and x_2 are dummy variables for campus, and x_3 is High School Grade Point Average.

- (a) (3 points) The table below shows how the dummy variables are defined. Write the odds of graduating within 5 years for each campus.

	x_1	x_2	Odds of Graduating
UTM	1	0	$e^{\beta_0 + \beta_1 + \beta_3 x_3}$
UTSC	0	1	$e^{\beta_0 + \beta_2 + \beta_3 x_3}$
St. George	0	0	$e^{\beta_0 + \beta_3 x_3}$

- (b) (1 point) Controlling for High School Grade Point Average, the odds of graduating within five years are ____ times as great for students on the UTM campus, compared to students on the UTSC campus. Write the answer in the space below in terms of β quantities.

$$\frac{e^{\beta_0 + \beta_1 + \beta_3 x_3}}{e^{\beta_0 + \beta_2 + \beta_3 x_3}} = e^{\beta_1 - \beta_2}$$

- (c) (2 points) Suppose you concluded $\beta_2 < 0$. How would you express this in plain, non-statistical language? Use the word "chances" instead of "odds" or "probability," and begin with "Allowing for High School marks ..."

Allowing for High School marks, the chances of graduating within 5 years are less for students on the UTSC campus than for students on the St. George campus

2. In your analysis of the Bird-keeping data data, you fit a model in which the response variable was whether they got lung cancer (1=Yes, 0=No), and the explanatory variables were Gender (0=M, 1=F), Socioeconomic Status (0=Low, 1=High), Whether they are birdkeepers (1=Yes, 0=No) Age, How many years they have been smoking (including zero), and Cigarettes per day. Please base your answers on this full model.

- (a) (2 Points) Controlling for all the other variables in the model, being a bird-keeper multiplies the estimated odds of cancer by ...? Write the number in the space below.

$$e^{1.36259} = 3.9$$

- (b) (2 Points) Estimate the probability of lung cancer for a 30 year old male of low socioeconomic status who does not smoke and is not a bird-keeper. The answer is a number. Show a little work.

$$\hat{\pi} = \frac{e^{-1.93736 - (0.03976)30}}{1 + e^{-1.93736 - 1.1928}}$$

$$= \frac{0.0437}{1.0437} = 0.042$$

Please attach your R printout. You don't need to write anything on the printout this time except your name and student number.