STA 302 f2014 Quiz 11A

1. The training to be an astronaut is very demanding; most candidates who enter the programme do not finish successfully. Trainers at the Space Agency were able to pre-test a large sample of candidates and then observe whether they completed the training programme successfully. They combined the assessments into a single number called **Pretest** in the output below. The model is $\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x$.

```
> spacemodel = glm(Success ~ Pretest, family=binomial)
> summary(spacemodel)


Call:
glm(formula = Success ~ Pretest, family = binomial)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-1.6823  -1.0544  -0.5418   1.0624   1.9458

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.74279    1.42083  -4.042 5.30e-05 ***
Pretest      0.05733    0.01420   4.037 5.42e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 200.84  on 144  degrees of freedom
Residual deviance: 181.16  on 143  degrees of freedom
AIC: 185.16

Number of Fisher Scoring iterations: 4
```

(a) (2 points) An extra ten points on the Pretest (ten points, not one) multiplies the estimated odds of Success by ... **Circle your answer below.** The answer is a number.

$$e^{10 * 0.05733} = \boxed{1.77}$$

(b) (2 points) For a candidate who scores 100 on the Pretest, the estimated probability (not odds) of success is ... **Circle your answer below.** The answer is a number.

$$\hat{\pi} = \frac{e^{-5.7429 + 100*0.05733}}{1 + e^{-0.0099}} = \frac{0.99}{1.99} = \boxed{0.497}$$

2. Consider a logistic regression with two independent variables. The model is $\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

(a) (3 points) If the value of x_2 is increased by c units, the odds of $Y = 1$ are ____ times as great. Prove your answer.

$$\frac{e^{\beta_0 + \beta_1 x_1 + \beta_2(x_2+c)}}{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}} = \frac{e^{\beta_0 + \beta_1 x_1} \, e^{\beta_2 x_2} \, e^{c\beta_2}}{e^{\beta_0 + \beta_1 x_1} \, e^{\beta_2 x_2}}$$

$$= e^{c\beta_2}$$

(b) (3 points) For the logistic regression model above, express $\pi = P(Y = 1)$ in terms of the β values. Show your work.

$$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 = x'\beta$$

$$\implies \frac{\pi}{1-\pi} = e^{x'\beta} \implies \pi = e^{x'\beta} - \pi e^{x'\beta}$$

$$\implies \pi + \pi e^{x'\beta} = e^{x'\beta} \implies \pi(1 + e^{x'\beta}) = e^{x'\beta}$$

$$\implies \pi = \frac{e^{x'\beta}}{1 + e^{x'\beta}} = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}$$

STA 302 F 2014 Quiz 11C

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 sudents on the UTM campus, compared to students on the UTSC campus. Write the answer in the space below in terms of β quantities.

$$e^{\beta_1 - \beta_2}$$

(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 graduation within 5 years are less for students on the UTSC campus than for students on the St. George campus.

Page 1 of 2

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. **Circle your answer.**

$$e^{1.36259} = \boxed{3.91}$$

(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. **Circle your answer.**

$$\hat{\pi} = \frac{e^{x'\beta}}{1+e^{x'\beta}} \quad , \text{ with } \quad x'\beta = -1.93736 - 30 * 0.03976$$
$$= -3.13 \text{ , so}$$

$$\hat{\pi} = \frac{e^{-3.13}}{1+e^{-3.13}} = \frac{.0437}{1.0437} = \boxed{0.042}$$

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